International Conference on Hydroinformatics

2014

A Diagnostic Assessment Of Evolutionary Multiobjective Optimization For Water Resources Systems

Patrick M. Reed

A DIAGNOSTIC ASSESSMENT OF EVOLUTIONARY MULTIOBJECTIVE OPTIMIZATION FOR WATER RESOURCES SYSTEMS

PATRICK REED, JONATHAN HERMAN (1), JOSHUA KOLLAT (2), DAVID HADKA (3), AND JOSEPH KASPRZYK (4)

(1): School of Civil and Environmental Engineering, Cornell University, 211 Hollister Hall, Ithaca, NY 14853, USA

(2): DecisionVis, LLC, 315 S. Allen St., Suite 321, State College, PA 16801, USA

(3): Applied Research Laboratory, The Pennsylvania State University, State College, PA 16801, USA

(4): Civil, Env., Arch. Engineering Department, University of Colorado Boulder, ECOT 441, UCB 428 Boulder, CO 80309, USA

This study contributes a rigorous diagnostic assessment of state-of-the-art multiobjective evolutionary algorithms (MOEAs) and highlights key advances that the water resources field can exploit to better discover the critical tradeoffs constraining our systems. This study provides the most comprehensive diagnostic assessment of MOEAs for water resources to date, exploiting more than 100,000 MOEA runs and trillions of design evaluations. The diagnostic assessment measures the effectiveness, efficiency, reliability, and controllability of ten benchmark MOEAs for a representative suite of water resources applications addressing rainfall-runoff calibration, long-term groundwater monitoring (LTM), and risk-based water supply portfolio planning. The suite of problems encompasses a range of challenging problem properties including (1) many-objective formulations with 4 or more objectives, (2) multi-modality (or false optima), (3) nonlinearity, (4) discreteness, (5) severe constraints, (6) stochastic objectives, and (7) non-separability (also called epistasis). The applications are representative of the dominant problem classes that have shaped the history of MOEAs in water resources and that will be dominant foci in the future. Recommendations are provided for which modern MOEAs should serve as tools and benchmarks in the future water resources literature..

OVERVIEW OF KEY CONCLUSIONS

This study utilized more than 100,000 separate MOEA runs and trillions of function evaluations to carefully evaluate 10 MOEAs. Our results provide a clear state-of-the-practice for many-objective optimization for water resources applications. Given the breadth of the results, problem types, and algorithms tested in this study, this section will provide a very brief synopsis of the relative strengths and weaknesses for each of the MOEAs.

BORG. The BORG algorithm showed the best scalability on the DTLZ2 test problem with objective counts as high as 16. It was also a dominant contributor to all of the applications' best known Pareto fronts. The algorithm exploits adaptive operator selection and adaptive population sizing to attain both high probabilities of success ("attainment") and controllability ("sweet spots" on control maps). Generally, BORG was the only algorithm to consistently exploit small populations and limited NFE to attain competitive to superior ε-indicator and hypervolume results. Overall BORG represents a top performing MOEA that could strongly support many-objective water resources applications. It was the most consistent of the ten algorithms tested across the problem suite. BORG also satisfies the requirements for having both a theoretical proof of convergence and diversity maintenance.

AMALGAM. AMALGAM has the advantage of blending several algorithms. Unfortunately, in its current variant its component algorithms are weak in comparison to the existing operators and tools tested in this study. Although the algorithm is best applied to two or three objective unconstrained rainfall-runoff problems, users will have to use higher NFE and larger population sizes relative to those needed for the top performing MOEAs. Additionally, AMALGAM suffers from algorithmic inefficiencies that can dramatically increase the wall clock time of its evolution, limiting its usefulness.

GDE3. GDE3 exploits the Differential Evolution search operators and is consequently capable of rotationally invariant search (i.e., use on non-separable real-valued problems). It is one of simplest and most parsimonious MOEAs tested in this study. It exhibited a somewhat reduced scalability for large objective counts on the DTLZ2 test case and did not contribute to the HBV or LRGV best known reference Pareto fronts. The algorithm did have high success rates ("attainment") on the HBV calibration and the LTM problems. In the highly constrained and stochastic LRGV test case, the algorithm suffered from premature convergence and poor hypervolume attainment. Although GDE3 has a limited number of parameters, its control maps show complex and difficult-to-predict trends which would necessitate careful analysis to ensure an effective parameterization ("low controllability").

MOEA/D. MOEA/D is a top performing algorithm both in this study and in prior evolutionary computation competitions. Among the tested MOEAs it is unique in its use of a traditional re-formulation of multiobjective problems into a population of single objective problems using a neighborhood-based Chebyshev decomposition. The algorithm was shown to be one of the most scalable to large objective counts on the DTLZ2 test case. It struggled to contribute to the best known Pareto fronts for the three water applications. The algorithm's attainment success probabilities were competitive for the HBV calibration for both ε-indicator and hypervolume. The algorithm struggled to maintain its high attainments for the constrained LTM and LRGV applications. MOEA/D is sensitive to its population size both in terms of its search capabilities as well as it computational tractability. MOEA/D is a viable tool for unconstrained applications such as rainfall-runoff calibration.

IBEA. Although conceptually interesting as a metric-based search tool, IBEA struggled across the full suite of problems. IBEA's use of the hypervolume metric as part of its evolutionary search implicitly limits its tractability and use. The hypervolume metric has a dramatic growth rate in wall clock time as the number of objectives increases. IBEA had very high failure rates across the three water resources applications and should not be considered as a viable tool.

εMOEA. εMOEA is a steady-state MOEA which means that the selection, mating, and mutation loop executes for each individual in its search population (versus traditional generational mating pool methods). As a steady-state MOEA it is highly parallelizable because its evolution can proceed asynchronously, avoiding bottlenecks that constrain other MOEAs. Its highly efficient use of ε-dominance archiving as part of its evolutionary process inspired the core architecture of the BORG algorithm. Nonetheless, εMOEA exploits the traditional simulated binary crossover and polynomial mutation operators used in most MOEAs and consequently lacks adaptivity in its operators and key parameters such as population size. It was less reliable at very high objective counts for the DTLZ2 test case compared to other top performing MOEAs such as BORG, MOEA/D, and εNSGA-II. εMOEA was a significant contributor to the LTM and LRGV test cases' best known Pareto fronts. Its attainment success probabilities were weak across the water resources applications and its control maps general show the need for larger population size and increased NFE. Overall εMOEA's performance was in the mid-range between the best and worst algorithms. This algorithm does satisfy the conditions necessary for theoretical proofs of convergence and diversity.

εNSGA-II. The εNSGA-II was the first algorithm to utilize adaptive population sizing and ε-dominance archiving. The algorithm is massively parallel, and because of its injection of random solutions when adjusting its population size, it is capable of pre-conditioning search to improve computational efficiency. BORG improves upon many of these innovations in its search. Overall εNSGA-II was one of the most scalable algorithms and was major contributor to the LTM test case's best known Pareto front. It was also a top performer in terms of its ε-indicator and hypervolume success probability attainments for the HBV calibration and LTM test cases. Its controllability for the HBV and LTM test cases was competitive with other top-performing algorithms, although it generally required larger initial search populations and a higher NFE. Overall εNSGA-II performed slightly better than εMOEA but its performance would be expected to be exceeded by BORG. εNSGA-II does satisfy the conditions necessary for a theoretical proof of convergence and diversity.

OMOPSO. OMOPSO is unique among the algorithms as the only representative of the particle swarm heuristics. The algorithm has been shown to be promising on a range of problems in the evolutionary computation literature and benefits from exploiting an ε-dominance archiving strategy. OMOPSO had a reduced performance in terms of scalability for the DTLZ2 test problem relative to the other algorithms that include ε-dominance (BORG, εNSGA-II, and εMOEA). It also had limited contributions to the best known Pareto fronts for the three water resources applications. OMOPSO had variable attainment success rate probabilities across the applications. It struggled to reliably attain high hypervolumes for the HBV calibration. For the LTM test case the algorithm was competitive with BORG and GDE3 as top performers. OMOPSO was strongly superior on the highly constrained, stochastic LRGV test case in terms of both attainment and controllability. Its success is largely controlled by the NFE utilized and its control maps did show the potential for high quality hypervolume performance for relatively reduced NFE for the LRGV test case. OMOPSO satisfies the conditions necessary for a theoretical proof of convergence and diversity.

SPEA2. SPEA2 was in the lower third of the MOEAs in terms of its scalability on the DTLZ2 test case. It was only able to contribute significantly to the LTM problem's best known Pareto front. It generally had poor attainment probabilities across the three water resources applications. Its control map for the HBV test case had isolated islands of high hypervolume performance, which agrees with prior studies that SPEA2 can be challenging to parameterize. It largely failed on the LRGV test case. Additionally, the algorithm suffers from ranking inefficiencies that can make its use in practice unnecessarily expensive for many-objective water resources applications. SPEA2 has been featured prevalently in the water resources literature as a benchmark MOEA. This study shows that the algorithm has become dated and that the field should move to new tools for benchmarking many-objective water resources applications.

NSGA-II. NSGA-II performed reasonably well in its scalability for the DTLZ2 test case. In the water resources problem suite, NSGA-II only contributed to the LTM test case's best known Pareto front. Its attainment success rate probabilities were poor for the HBV and LRGV test cases. In terms of the LRGV test case, the NSGA-II had the worst performance overall. Its control maps show consistently that the NSGA-II needs large population sizes and high NFE to attain acceptable hypervolume performance. Although NSGA-II is the most prevalently used algorithm in the water resources literature, this study emphasizes the need for the field to transition to more robust MOEAs.

Overall the BORG, MOEA/D, and OMOPSO MOEAs represent the top performing modern tools. These algorithms have tremendous potential for advancing the size and scope of the many-objective water resources applications that are now feasible. These algorithms should be the focus for future benchmarking studies, which need to follow rigorously constructed statistical experiments such as the one contributed in this study.