Obtaining Approximately Optimal and Diverse Solutions via Dispersion

Jie Gao*, Mayank Goswami†, Karthik C.S.*, Meng-Tsung Tsai‡, Shih-Yu Tsai§, Hao-Tsung Yang¶
*Department of Computer Science, Rutgers University, USA
†Department of Computer Science, Queens College, City University of New York, USA
‡Institute of Information Science, Academia Sinica, Taiwan
§Department of Computer Science, Stony Brook University, USA
¶ Department of Computer Science and Information Engineering, National Central University, Taiwan

***Abstract*—There has been a long-standing interest in computing diverse solutions to optimization problems. In 1995 J. Krarup [28] posed the problem of finding k-edge disjoint Hamiltonian Circuits of minimum total weight, called the peripatetic salesman problem (PSP). Since then researchers have investigated the complexity of finding diverse solutions to spanning trees, paths, vertex covers, matchings, and more. Unlike the PSP that has a constraint on the total weight of the solutions, recent work has involved finding diverse solutions that are all optimal.**

However, sometimes the space of exact solutions may be too small to achieve sufficient diversity. Motivated by this, we initiate the study of obtaining sufficiently-diverse, yet approximately-optimal solutions to optimization problems. Formally, given an integer k, an approximation factor c, and an instance I of an optimization problem, we aim to obtain a set of k solutions to I that a) are all c approximately-optimal for I and b) maximize the diversity of the k solutions. Finding such solutions, therefore, requires a better understanding of the global landscape of the optimization function.

Given a metric on the space of solutions, and the diversity measure as the sum of pairwise distances between solutions, we first provide a general reduction to an associated budget-constrained optimization (BCO) problem, where one objective function is to optimized subject to a bound on the second objective function. We then prove that bi-approximations to the BCO can be used to give bi-approximations to the diverse approximately optimal solutions problem.

As applications of our result, we present polynomial time approximation algorithms for several problems such as diverse c-approximate *maximum matchings, $s-t$ shortest paths, global min-cut, and minimum weight bases of a matroid*. The last result gives us *diverse c-approximate minimum spanning trees*, advancing a step towards achieving diverse c-approximate TSP tours.

We also explore the connection to the field of multiobjective optimization and show that the class of problems to which our result applies includes those for which the associated DUAL-RESTRICT problem defined by Papadimitriou and Yannakakis [35], and recently explored by Herzel et al. [26] can be solved in polynomial time.

INTRODUCTION

Techniques for optimization problems focus on obtaining optimal solutions to an objective function and have widespread applications ranging from machine learning, operations research, computational biology, networks, to geophysics, economics, and finance. However, in many scenarios, the optimal solution is not only computationally difficult to obtain, but can also render the system built upon its utilization vulnerable to adversarial attacks. Consider a patrolling agent tasked with monitoring n sites in the plane. The most efficient solution (i.e., maximizing the frequency of visiting each of the n sites) would naturally be to patrol along the tour of shortest length[1] (the solution to TSP - the Traveling Salesman Problem). However, an adversary who wants to avoid the patroller can also compute the shortest TSP tour and can design its actions strategically [39]. Similarly, applications utilizing the minimum spanning tree (MST) on a communication network may be affected if an adversary gains knowledge of the network [13]; systems using solutions to a linear program (LP) would be vulnerable if an adversary gains knowledge of the program's function and constraints.

One way to address the vulnerability is to use a set of approximately optimal solutions and randomize among them. However, this may not help much to mitigate the problem, if these approximate solutions are combinatorially too "similar" to the optimal solution. For example, all points in a sufficiently small neighborhood of the optimal solution on the LP polytope will be approximately optimal, but these solutions are not too much different and the adversaries can still effectively carry out their attacks. Similarly one may use another tree instead of the MST, but if the new tree shares many edges with the MST the same vulnerability persists. Thus k-best enumeration algorithms ([18], [24], [30], [31], [33]) developed for a variety of problems fall short in this regard.

One of the oldest known formulations is the Peripatetic Salesman problem (PSP) by Krarup [28], which asks for k-edge disjoint Hamiltonian circuits of minimum total weight in a network. Since then, several researchers have tried to compute diverse solutions for several optimization problems [4], [5], [16], [23]. Most of these works are on graph problems, and diversity usually corresponds to the size of the symmetric difference of the edge sets in the solutions. Crucially, almost all of the aforementioned work demands either every solution individually be optimal, or the set of solutions in totality (as in the case of the PSP) be optimal. Nevertheless, **the space of optimal solutions may be too small to achieve sufficient diversity**, and it may just be singular (unique solution). In addition, for NP-complete problems finding just one optimal

[1] We assume without loss of generality that the optimal TSP is combinatorially unique by a slight perturbation of the distances.

solution is already difficult. While there is some research that takes the route of developing FPT algorithms for this setting [5], [17], to us it seems practical to also consider the relaxation to approximately-optimal solutions.

This motivates the problem of finding a set of diverse and *approximately optimal* solutions, which is the problem considered in this article. The number of solutions k and the desired approximation factor $c > 1$ is provided by the user as input. Working in the larger class gives one more hope of finding diverse solutions, yet every solution has a guarantee on its quality.

A. Our Contributions

We develop approximation algorithms for finding k solutions to the given optimization problem: for every solution, the quality is bounded by a user-given approximation ratio $c > 1$ to the optimal solution and the diversity of these k solutions is maximized. Given a metric on the space of solutions to the problem, we consider the diversity measure given by the sum (or average) of pairwise distances between the k solutions. Combining ideas from the well-studied problem on dispersion (which we describe next), we reduce the above problem to a budget constrained optimization (BCO) program.

B. Dispersion

Generally speaking, if the optimization problem itself is $\mathcal{NP}$-hard, finding diverse solutions for that problem is also $\mathcal{NP}$-hard. On the other hand, interestingly, even if the original problem is not $\mathcal{NP}$-hard, finding diverse and approximately optimal solutions can still be $\mathcal{NP}$-hard. This is due to the connection of the diversity maximization objective with the general family of problems that consider selecting k elements from the given input set with maximum "dispersion", defined as max-min distance, max-average distance, and so on.

The dispersion problem has a long history, with many variants both in the metric setting and the geometric setting [15], [29], [38]. For example, finding a subset of size k from an input set of n points in a metric space that maximizes the distance between closest pairs or the sum of distances of the k selected points are both $\mathcal{NP}$-hard [1], [37]. For the max-sum dispersion problem, the best known approximation factor is 2 for general metrics [7], [25], although PTAS are available for Euclidean metrics or more generally, metrics of negative type, even with matroid constraints [10], [11].

Dispersion in exponentially-sized space We make use of the general framework of the 2-approximation algorithm [8], [37] to the max-sum k-dispersion problem, a greedy algorithm where the $i+1$th solution is chosen to be the most distant/diverse one from the first i solutions. Notice that in our setting, there is an important additional challenge to understand the space within which the approximate solutions stay. In all of the problems we study, the total number of solutions can be *exponential in the input size*. Thus we need to have a non-trivial way of navigating within this large space and carry furthest insertion without considering all points in the space. This is where our reduction to budget constrained problem comes in.

Self avoiding dispersion Furthermore, even after implicitly defining the $i+1$th furthest point insertion via some optimization problem, one needs to take care that the (farthest, in terms of sum of distances) solution does not turn out to equal one of the previously found i solutions, as this is a requirement for the furthest point insertion algorithm. This is an issue one faces because of the implicit nature of the furthest point procedure in the exponential-sized space of solutions: in the metric k-dispersion problem, it was easy to guarantee distinctness as one only considered the $n-i$ points not yet selected.

C. Reduction to Budget Constrained Optimization

Combining with dispersion, we reduce the diversity computational problem to a budget constrained optimization (BCO) problem where the budget is an upper (resp. lower) bound if the quality of solution is described by a minimization (resp. maximization) problem. Intuitively the budget guarantees the quality of the solution, and the objective function maximizes diversity. Recall that the number of solutions k and the approximation factor c is input by the user; a larger c allows for more diversity.

We show how using an (a, b) bi-approximation algorithm for the BCO problem provides a set of $O(a)$-diverse, bc approximately-optimal solutions to the diversity computational problem (the hidden constant is at most 4). This main reduction is described in our full version paper [19].

The main challenge in transferring the bi-approximation results because of a technicality that we describe next. Let $\mathcal{S}(c)$ be the space of c approximate solutions. A $(*, b)$ bi-approximation algorithm to the BCO relaxes the budget constraint by a factor b, and hence only promises to return a faraway point in the larger space $\mathcal{S}(b\cdot c)$. Thus bi-approximation of BCO do not simply give a farthest point insertion in the space of solutions, and instead *return a point in a larger space*. Nevertheless, we prove that in most cases, one loses a factor of at most 4 in the approximation factor for the diversity.

Once the reduction to BCOs is complete, for diverse approximate matchings, spanning trees and shortest paths we exploit the special characteristics of the corresponding BCO to solve it optimally ($a = b = 1$). For other problems such as global min-cut, diverse approximate minimum weight spanning trees, and the more general minimum weight bases of a matroid, we utilize known bi-approximations to the BCO to obtain bi-approximations for the diversity problem. For all problems except diverse (unweighted) spanning trees[2], our algorithms are the first polynomial time bi-approximations for these problems.

We also connect to the wide literature on multicriteria optimization and show that our result applies to the entire class of problems for which the associated DUALRESTRICT problem (defined by Papadimitriou and Yannakakis [35], and recently studied by Herzel et al. [26]) has a polynomial time solution. We discuss this in more detail after presenting our reduction.

[2]While an exact algorithm for diverse unweighted spanning trees is known [23], we give a faster (by a factor $\Omega(n^{1.5}k^{1.5}/\alpha(n, m))$ where $\alpha(\cdot)$ denotes the inverse of the Ackermann function), 2-approximation here.

This work has been accepted in LATIN 2022. Due to space constraints, all proofs can be found in the publicly available full version of this paper at [19].

References

[1] Abbar, S., Amer-Yahia, S., Indyk, P., Mahabadi, S., Varadarajan, K.R.: Diverse near neighbor problem. In: Symposium on Computational Geometry (SoCG). pp. 207–214. ACM (2013)

[2] Armon, A., Zwick, U.: Multicriteria global minimum cuts. Algorithmica **46**(1), 15–26 (2006)

[3] Ausiello, G., Marchetti-Spaccamela, A., Crescenzi, P., Gambosi, G., Protasi, M., Kann, V.: Complexity and approximation: combinatorial optimization problems and their approximability properties. Springer (1999)

[4] Baste, J., Fellows, M.R., Jaffke, L., Masařík, T., de Oliveira Oliveira, M., Philip, G., Rosamond, F.A.: Diversity of solutions: An exploration through the lens of fixed-parameter tractability theory. Artificial Intelligence **303**, 103644 (2022)

[5] Baste, J., Jaffke, L., Masařík, T., Philip, G., Rote, G.: Fpt algorithms for diverse collections of hitting sets. Algorithms **12**(12), 254 (2019)

[6] Berger, A., Bonifaci, V., Grandoni, F., Schäfer, G.: Budgeted matching and budgeted matroid intersection via the gasoline puzzle. Math. Program. **128**(1-2), 355–372 (2011)

[7] Birnbaum, B.E., Goldman, K.J.: An improved analysis for a greedy remote-clique algorithm using factor-revealing LPs. Algorithmica **55**(1), 42–59 (2009)

[8] Borodin, A., Jain, A., Lee, H.C., Ye, Y.: Max-sum diversification, monotone submodular functions, and dynamic updates. ACM Trans. Algorithms **13**(3), 41:1–41:25 (2017)

[9] Camerini, P.M., Galbiati, G., Maffioli, F.: Random pseudo-polynomial algorithms for exact matroid problems. J. Algorithms **13**(2), 258–273 (1992)

[10] Cevallos, A., Eisenbrand, F., Zenklusen, R.: Max-sum diversity via convex programming. In: Fekete, S.P., Lubiw, A. (eds.) 32nd International Symposium on Computational Geometry, SoCG 2016, June 14-18, 2016, Boston, MA, USA. LIPIcs, vol. 51, pp. 26:1–26:14. Schloss Dagstuhl - Leibniz-Zentrum für Informatik (2016)

[11] Cevallos, A., Eisenbrand, F., Zenklusen, R.: Local search for max-sum diversification. In: Proceedings of the Twenty-Eighth Annual ACM-SIAM Symposium on Discrete Algorithms. pp. 130–142. SIAM (2017)

[12] Clausen, J., Hansen, L.A.: Finding k edge-disjoint spanning trees of minimum total weight in a network: an application of matroid theory. In: Combinatorial Optimization II, pp. 88–101. Springer (1980)

[13] Commander, C.W., Pardalos, P.M., Ryabchenko, V., Uryasev, S., Zrazhevsky, G.: The wireless network jamming problem. Journal of Combinatorial Optimization **14**(4), 481–498 (2007)

[14] Diakonikolas, I., Yannakakis, M.: Small approximate pareto sets for biobjective shortest paths and other problems. SIAM Journal on Computing **39**(4), 1340–1371 (2010)

[15] Erkut, E.: The discrete p-dispersion problem. European Journal of Operational Research **46**(1), 48–60 (1990). https://doi.org/https://doi.org/10.1016/0377-2217(90)90297-O, https://www.sciencedirect.com/science/article/pii/037722179090297O

[16] Fomin, F.V., Golovach, P.A., Jaffke, L., Philip, G., Sagunov, D.: Diverse pairs of matchings. In: 31st International Symposium on Algorithms and Computation (ISAAC). LIPIcs, vol. 181, pp. 26:1–26:12. Schloss Dagstuhl - Leibniz-Zentrum für Informatik (2020)

[17] Fomin, F.V., Golovach, P.A., Panolan, F., Philip, G., Saurabh, S.: Diverse collections in matroids and graphs. In: Bläser, M., Monmege, B. (eds.) 38th International Symposium on Theoretical Aspects of Computer Science, STACS 2021, March 16-19, 2021, Saarbrücken, Germany (Virtual Conference). LIPIcs, vol. 187, pp. 31:1–31:14. Schloss Dagstuhl - Leibniz-Zentrum für Informatik (2021)

[18] Gabow, H.N.: Two algorithms for generating weighted spanning trees in order. SIAM Journal on Computing **6**(1), 139–150 (1977)

[19] Gao, J., Goswami, M., S., K.C., Tsai, M.T., Tsai, S.Y., Yang, H.T.: Obtaining approximately optimal and diverse solutions via dispersion (2022). https://doi.org/10.48550/ARXIV.2202.10028, https://arxiv.org/abs/2202.10028

[20] Goldenberg, E., Karthik C. S.: Hardness amplification of optimization problems. In: Vidick, T. (ed.) 11th Innovations in Theoretical Computer Science Conference, ITCS 2020, January 12-14, 2020, Seattle, Washington, USA. LIPIcs, vol. 151, pp. 1:1–1:13. Schloss Dagstuhl - Leibniz-Zentrum für Informatik (2020). https://doi.org/10.4230/LIPIcs.ITCS.2020.1, https://doi.org/10.4230/LIPIcs.ITCS.2020.1

[21] Hanaka, T., Kiyomi, M., Kobayashi, Y., Kobayashi, Y., Kurita, K., Otachi, Y.: A framework to design approximation algorithms for finding diverse solutions in combinatorial problems. CoRR **abs/2201.08940** (2022)

[22] Hanaka, T., Kobayashi, Y., Kurita, K., Lee, S.W., Otachi, Y.: Computing diverse shortest paths efficiently: A theoretical and experimental study. CoRR **abs/2112.05403** (2021)

[23] Hanaka, T., Kobayashi, Y., Kurita, K., Otachi, Y.: Finding diverse trees, paths, and more. In: Thirty-Fifth AAAI Conference on Artificial Intelligence (AAAI). pp. 3778–3786. AAAI Press (2021)

[24] Hara, S., Maehara, T.: Enumerate lasso solutions for feature selection. In: Singh, S.P., Markovitch, S. (eds.) Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence (AAAI). pp. 1985–1991. AAAI Press (2017)

[25] Hassin, R., Rubinstein, S., Tamir, A.: Approximation algorithms for maximum dispersion. Oper. Res. Lett. **21**(3), 133–137 (Oct 1997)

[26] Herzel, A., Bazgan, C., Ruzika, S., Thielen, C., Vanderpooten, D.: One-exact approximate pareto sets. Journal of Global Optimization **80**(1), 87–115 (2021)

[27] Herzel, A., Ruzika, S., Thielen, C.: Approximation methods for multiobjective optimization problems: A survey. INFORMS Journal on Computing **33**(4), 1284–1299 (2021)

[28] Krarup, J.: The peripatetic salesman and some related unsolved problems. In: Combinatorial programming: methods and applications, pp. 173–178. Springer (1995)

[29] Kuby, M.: Programming models for facility dispersion: The p-dispersion and maxisum dispersion problems. Geographical Analysis **19**(4), 315–329 (Oct 1987). https://doi.org/10.1111/j.1538-4632.1987.tb00133.x

[30] Lawler, E.L.: A procedure for computing the k best solutions to discrete optimization problems and its application to the shortest path problem. Management Science **18**(7), 401–405 (1972)

[31] Lindgren, E.M., Dimakis, A.G., Klivans, A.: Exact map inference by avoiding fractional vertices. In: International Conference on Machine Learning. pp. 2120–2129. PMLR (2017)

[32] Mulmuley, K., Vazirani, U.V., Vazirani, V.V.: Matching is as easy as matrix inversion. Comb. **7**(1), 105–113 (1987)

[33] Murty, K.G.: An algorithm for ranking all the assignments in order of increasing cost. Operations research **16**(3), 682–687 (1968)

[34] Namorado Climaco, J.C., Queirós Vieira Martins, E.: A bicriterion shortest path algorithm. European Journal of Operational Research **11**(4), 399–404 (1982)

[35] Papadimitriou, C.H., Yannakakis, M.: On the approximability of trade-offs and optimal access of web sources. In: Proceedings 41st annual symposium on foundations of computer science. pp. 86–92. IEEE (2000)

[36] Ravi, R., Goemans, M.X.: The constrained minimum spanning tree problem. In: Scandinavian Workshop on Algorithm Theory. pp. 66–75. Springer (1996)

[37] Ravi, S.S., Rosenkrantz, D.J., Tayi, G.K.: Heuristic and special case algorithms for dispersion problems. Operations Research **42**(2), 299–310 (1994)

[38] Wang, D., Kuo, Y.S.: A study on two geometric location problems. Information Processing Letters **28**(6), 281–286 (1988). https://doi.org/https://doi.org/10.1016/0020-0190(88)90174-3, https://www.sciencedirect.com/science/article/pii/0020019088901743

[39] Yang, H.T., Tsai, S.Y., Liu, K.S., Lin, S., Gao, J.: Patrol scheduling against adversaries with varying attack durations. In: Proceedings of the 18th International Conference on Autonomous Agents and MultiAgent Systems. pp. 1179–1188 (2019)