Diversity Enhancement via Magnitude

Steve Huntsman[✉]

STR, Arlington, VA 22203, USA
`steve.huntsman@str.us`

Abstract. Promoting and maintaining diversity of candidate solutions is a key requirement of evolutionary algorithms. In this paper, we use the recently developed theory of *magnitude* to construct a gradient flow that systematically manipulates finite subsets of Euclidean space to enhance their diversity, and we apply the ideas in service of multi-objective evolutionary algorithms. We demonstrate diversity enhancement on benchmark problems using leading algorithms.

1 Introduction

Promoting and maintaining diversity of candidate solutions is a key requirement of *evolutionary algorithms* (EAs) in general and *multi-objective EAs* (MOEAs) in particular [1,2]. Many ways of measuring diversity have been considered, and many shortcomings identified [3]. Perhaps the most theoretically attractive diversity measure, used by [4,5], is the *Solow-Polasky diversity* [6]. It turns out that a recently systematized theory of diversity in generalized metric spaces [7] singles out the Solow-Polasky diversity or *magnitude* of a (certain frequently total subset of a) finite metric space as equal to the maximum value of the "correct" definition (1) of diversity that uniquely satisfies various natural desiderata. While the notion of magnitude was implicit in the mathematical ecology literature over 25 years ago, an underlying notion of a diversity-maximizing probability distribution is much more recent and has not yet been applied to EAs.

In the context of MOEAs, a practical shortcoming associated with magnitude is its $O(n^3)$ algorithmic cost. To avoid this, [4,5] use an efficient approximation to merely *measure* diversity rather than attempting to *enhance* it.

However, it can be profitable to incur the marginal cost of computing a so-called weighting *en route* to the magnitude, since we can use it to enhance diversity near the boundary of the image of the candidate solution set under the objective functions. The nondominated part of this image is the current approximation to the Pareto front; the ability of weightings to couple both diversity and convergence to the Pareto front dovetails with recent indicator-based EA approaches to Pareto-dominance based MOEAs [8,9]. Moreover, the agnosticism of weightings to dimension further enhances their suitability for such applications.

In this paper, we construct a gradient flow that systematically manipulates finite subsets of Euclidean space to enhance their diversity, which provides a

useful primitive for quality diversity [10]. We then apply this primitive in service of MOEAs by diversifying solution data through local mutations. For the sake of illustration, we only perform these mutations on the results already obtained by a MOEA, though they can be performed during evolution.

The paper is organized as follows. In Sect. 2, we sketch the concepts of weightings, magnitude, and diversity, and describe an efficiently computable scale above which a weighting is guaranteed to be proportional to the unique diversity-maximizing distribution. In Sect. 3, we develop a notion of a weighting gradient (estimate) and an associated flow. In Sect. 4, we use this gradient flow to demonstrate diversity enhancement on a toy problem before turning to benchmark problems in Sect. 5. Finally, we discuss algorithmic extensions in Sect. 6 before remarks in Sect. 7.

2   Weightings, Magnitude, and Diversity

For details on the ideas in this section, see §6 of [7] and also [16,30].

Call a square matrix $Z \geq 0$ a *similarity matrix* if $\mathrm{diag}(Z) > 0$. A motivating class of examples is $Z = \exp[-td]$ where square brackets indicate entrywise function application, $t \in (0, \infty)$, and d is a square *dissimilarity matrix* (e.g., the matrix encoding a finite metric space). A *weighting* w is a column vector satisfying $Zw = 1$, where the vector of all ones is indicated on the right. A *coweighting* is the transpose of a weighting for Z^T. If Z admits both a weighting w and a coweighting, then its *magnitude* is defined via $1^T w = \sum_j w_j$, which also turns out to equal the sum of the coweighting components.
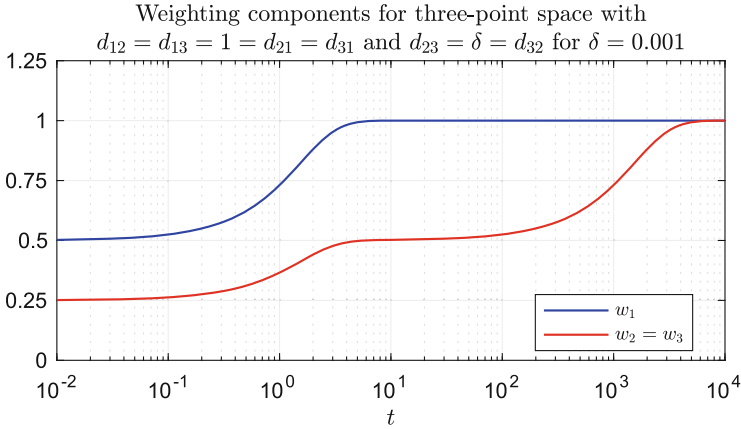
In the case $Z = \exp[-td]$ and d is the distance matrix corresponding to a finite subset of Euclidean space, Z is positive definite [11], hence invertible, and so its weighting and magnitude are well-defined and unique. More generally, if Z is invertible then its magnitude is $1^T Z^{-1} 1$. For d as specified above, the *magnitude function* is defined as the map $t \mapsto 1^T (\exp[-td])^{-1} 1$.

Weightings are excellent scale-dependent boundary detectors in Euclidean space (see, e.g., Fig. 2 and [12,13]). Meanwhile, magnitude is a very general notion of size that encompasses rich scale-dependent geometrical data [14].

Example 1. Consider a three-point space with $d_{12} = d_{13} = 1 = d_{21} = d_{31}$ and $d_{23} = \delta = d_{32}$. A routine calculation yields that

$$w_1 = \frac{e^{(\delta+2)t} - 2e^{(\delta+1)t} + e^{2t}}{e^{(\delta+2)t} - 2e^{\delta t} + e^{2t}}; \quad w_2 = w_3 = \frac{e^{(\delta+2)t} - e^{(\delta+1)t}}{e^{(\delta+2)t} - 2e^{\delta t} + e^{2t}}.$$

For $\delta = 10^{-3}$, Fig. 1 shows that at $t = 10^{-2}$, the "effective size" of the nearby points is ≈ 0.25; that of the distal point is ≈ 0.5, so the "effective number of points" is ≈ 1. At $t = 10$, these effective sizes are respectively ≈ 0.5 and ≈ 1, so the effective number of points is ≈ 2. Finally, at $t = 10^4$, the effective sizes are all ≈ 1, so the effective number of points is ≈ 3.

Weighting components for three-point space with
$d_{12} = d_{13} = 1 = d_{21} = d_{31}$ and $d_{23} = \delta = d_{32}$ for $\delta = 0.001$

Fig. 1. Weighting components for an "isoceles" metric space. The magnitude function $w_1 + w_2 + w_3$ gives a scale-dependent "effective number of points."

For a probability distribution p and similarity matrix Z, the *diversity* $D_q^Z(p)$ is defined for $1 < q < \infty$ (and via limits for $q = 1, \infty$) via

$$\log D_q^Z(p) := \frac{1}{1-q} \log \sum_{j : p_j > 0} p_j (Zp)_j^{q-1}. \tag{1}$$

This is the "correct" measure of diversity in essentially the same way that Shannon entropy is the "correct" measure of information [7]. (In fact, the expression (1) is a generalization of the Rényi entropy of order q. In the event $Z = I$, the usual Rényi entropy is recovered, with Shannon entropy as the case $q = 1$.) We therefore restrict our attention to it versus other measures such as those discussed in [2,3].

Recent mathematical developments [7,15] have clarified the role of magnitude in *maximizing* (1) versus merely computing it. Specifically, if $Z = \exp[-td]$ is positive definite with d symmetric, nonnegative, and with zero diagonal, and if Z admits a positive weighting $w = Z^{-1}1$, then this (unique) weighting is proportional to the diversity-maximizing distribution. This situation holds automatically if d is the distance matrix of a finite subset of Euclidean space and if Z is diagonally dominant (i.e., $Z_{jj} > \sum_{k \neq j} Z_{jk}$).

For d with zero diagonal and all other entries positive, there is a least $t_d > 0$ such that $\exp[-td]$ is diagonally dominant for any $t > t_d$. Because $\exp[-td]$ is diagonally dominant iff $1 > \max_j \sum_{k \neq j} \exp(-td_{jk})$, we can efficiently estimate t_d using the following elementary bounds and a binary search:

Lemma 1. *For $d \in M_n$ as above,* $\frac{\log(n-1)}{\min_j \max_k d_{jk}} \leq t_d \leq \frac{\log(n-1)}{\min_j \min_{k \neq j} d_{jk}}$.

More importantly, we can also use Lemma 1 to find the least $t_+ < t_d$ such that $\exp[-td]$ admits a positive weighting for $t > t_+$.

3   The Weighting Gradient Flow

We now define a gradient flow that (for $t \geq t_+$) increases the diversity of finite subsets of Euclidean space and thereby provides a useful primitive for EAs. Although there are various sophisticated approaches to estimating gradients on point clouds (see, e.g., [17]), a reasonable heuristic estimate for the specific case of the gradient of a weighting w on $\{x_j\}_j$ in Euclidean space is

$$(\hat{\nabla}w)_j := \sum_{k \neq j} \frac{Z_{jk}}{\sum_{k' \neq j} Z_{jk'}} \frac{w_k - w_j}{d_{jk}} e_{jk}, \tag{2}$$

where $e_{jk} := \frac{x_k - x_j}{d_{jk}}$. The *weighting gradient flow* induced by (2) is

$$\dot{x} = \hat{\nabla}w. \tag{3}$$

Example 2. Figure 2 illustrates how weightings identify boundaries at various scales, and the corresponding weighting gradient estimates (2).
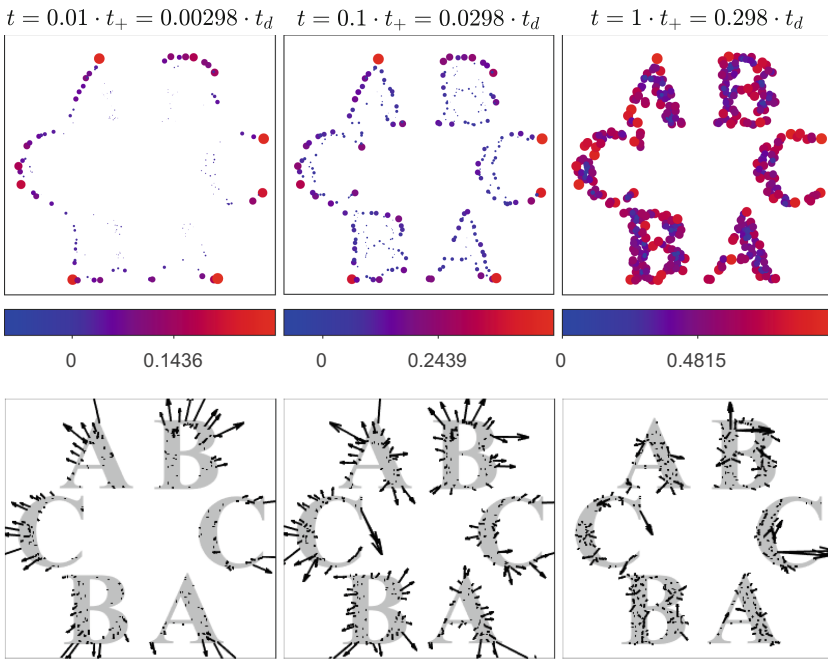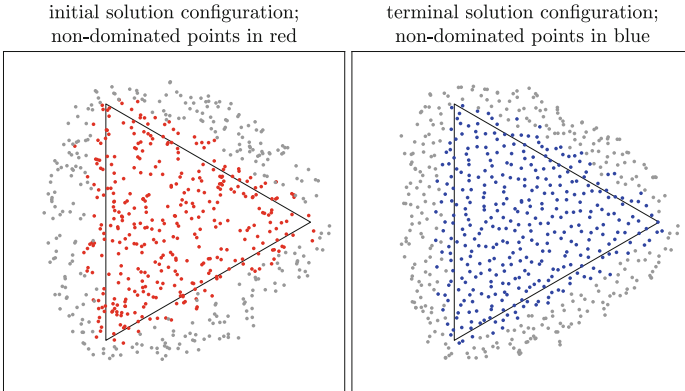

Fig. 2. (Top) Weighting components for 500 points sampled without replacement from a probability distribution on $\mathbb{Z}^2$ that is approximately uniform on its support. From left to right, various scale factors t defining $Z = \exp[-td]$ (with d = Euclidean distance) are shown in terms of the intrinsic scales t_d and t_+. Both the color and size of a point indicate its weighting component; the nonzero color axis tick mark is at half maximum. (Bottom) Weighting gradient estimate (2) for the data above. The gradient vectors are scaled uniformly in each panel for visualization purposes. Note that for the largest value of t the large gradient vectors have basepoints near other large gradient vectors.

4   Enhancing Diversity

Following [18], we apply the ideas sketched above to a toy problem where the objective function f has three components, each measuring the distance to a vertex of a regular triangle with vertices in S^1. The application is mostly conceptually straightforward, but we mention a few implementation details:

– We begin with a uniformly distributed sample of $n_0 = 10^3$ points in the disk of radius 1.25, and retain n points that are dominated by $\leq \delta = 0.1$;
– Replace misbehaving points (e.g., out of bounds or `NaN`s) with predecessors;
– Set $S_j := 1 - 2\frac{\mathrm{dom}_j}{\max_k \mathrm{dom}_k}$, where $\mathrm{dom}_j = |\{\text{points dominating the } j\text{th point}\}|$;
– Evolve the n points under a modulated version of (3) *on the objective space* with $t = t_+$ as $dy_j = ds \cdot S_j(\hat{\nabla}w)_j$ for only $N = 10$ steps and step size $ds = \sqrt{\langle \min_{k \neq j}(d_f)_{jk} \rangle / n}$, where the pullback metric is $d_f(x, x') := d(f(x), f(x'))$;
– Pull back the weighting gradient flow from objective to solution space using the Jacobian's pseudoinverse, then recompute points in objective space.

The result of this experiment is depicted in Figs. 3 and 4. The salutary effect on diversity in objective (and solution) space is apparent. This can be quantified via the objective space magnitude functions, as shown in Fig. 5.

initial solution configuration;
non-dominated points in red

terminal solution configuration;
non-dominated points in blue

Fig. 3. Comparision of initial (red; left) and terminal (blue; right) locations of points in the solution space. The weighting gradient flow produces more evenly distributed terminal points. The triangle defining objective components (by distance to vertices) is shown. The actual Pareto front is the interior of the triangle; the area displayed is $[-1, 1]^2$. Bottom: comparision of initial (red; left) and terminal (blue; right) points in the objective space. The terminal points are more evenly distributed. (Color figure online)
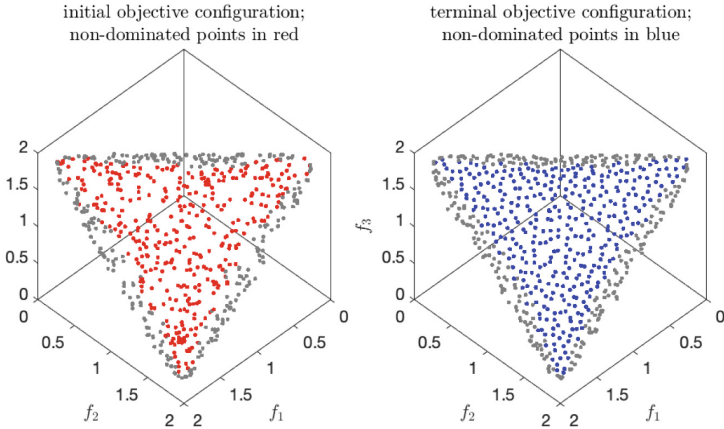
Fig. 4. Comparision of initial (red; left) and terminal (blue; right) points in the objective space. The terminal points are more evenly distributed. (Color figure online)
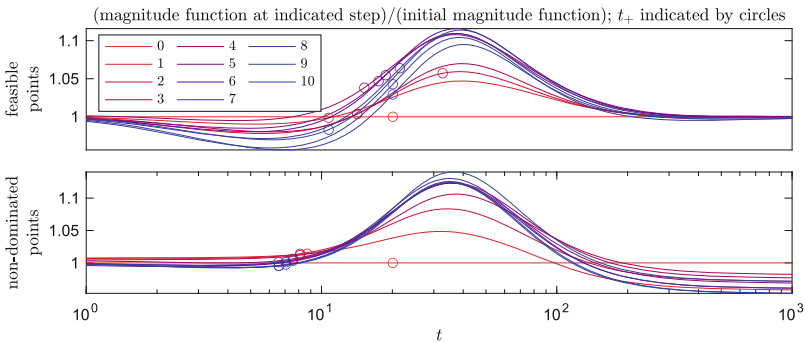

Fig. 5. Magnitude increases for the experiment of Sect. 4 at scales above t_+, where magnitude equals diversity. (Top) Magnitude function quotients at various timesteps for feasible points under the evolution of the (modulated) weighting gradient flow. The horizontal axis t indicates the scale parameter; timesteps of numerators are indicated via color, going from red at the initial timestep (0) to blue at the final timestep (10): the denominator is the function at the initial timestep. Circles indicate the scales t_+. (Bottom) As above, but for non-dominated points. (Color figure online)

5   Performance on Benchmarks

The effectiveness of the (modulated) weighting gradient flow approach hinges on the ability to cover and thereby "keep pressure on" the Pareto front. A straightforward way to do this is to use a MOEA to produce an initial over-approximation of the Pareto front as in [19], and then improve the diversity of the overapproximation via the weighting gradient flow. We proceed to detail the

results of an experiment along these lines. For the experiment we considered two leading MOEAs (NSGA-II [20] and SPEA2 [21]) and two leading benchmark problem sets (DTLZ [22][1] and WFG [24]), all implemented in PlatEMO version 2.9 [25]. For each problem, we used 10 decision variables, three objectives (to enable visualization), and performed 10 runs (which appears quite adequate for characterization purposes) with population size 250 and 10^4 fitness evaluations. We then took $N = 10$ timesteps for the weighting gradient flow as before.

Figure 6 (cf. Fig. 5) shows magnitude functions at various timesteps of the (modulated) weighting gradient flow applied to the results of NSGA-II on the WFG2 benchmark. Feasible points show a diversity (as measured by magnitude at scale t_+ for the feasible objective points) increase of about 10%, whereas non-dominated points show a diversity increase of several percent as well, even as the total number of non-dominated points decreases by about 15%.
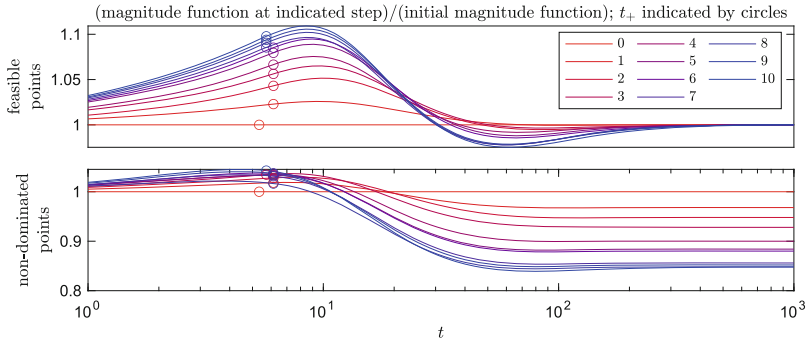

Fig. 6. As in Fig. 5, but for a solution of the WFG2 benchmark via NSGA-II.

We produce an ensemble characterization in Fig. 7. The figure shows that the number of non-dominated points decreases since the weighting gradient flow pushes some points a short distance away from the Pareto front (as illustrated in Fig. 9) before they are halted or reversed. The figure also shows that the diversity of non-dominated points generally increases slightly, and the diversity of feasible points increases significantly. As a consequence, the diversity contributions of individual solutions (as measured by the average weighting, i.e., the magnitude of non-dominated points divided by their cardinality) also increases significantly. For less challenging problems such as in Fig. 3, the number of non-dominated points will decrease less, and the diversity gains will be enhanced.

---

[1] For DTLZ, we considered only the two most relevant problems, viz. DTLZ4 and DTLZ7. DTLZ4 was formulated "to investigate an MOEA's ability to maintain a good distribution of solutions" and DTLZ7 was formulated to "test an algorithm's ability to maintain subpopulation in different Pareto-optimal regions" [22]. (NB. One approach for the latter, not pursued here, is to resample points so that the diversity per point in each connected component of the Pareto front is approximately equal. For the application of topological data analysis to Pareto fronts, see [23].).

On the other hand, the effects of the weighting gradient flow are considerably reduced in the case of SPEA2, which produces a visibly more uniform distribution in objective space than NSGA-II: see Fig. 11. The weighting gradient flow appears to *decrease* this uniformity; the formation of a gap just behind the boundary along with a slight increase in the population near the boundary are the main visible indicators that something useful (at least for DTLZ4, WFG2, WFG3, WFG6, and WFG8, per Fig. 7) is actually happening.
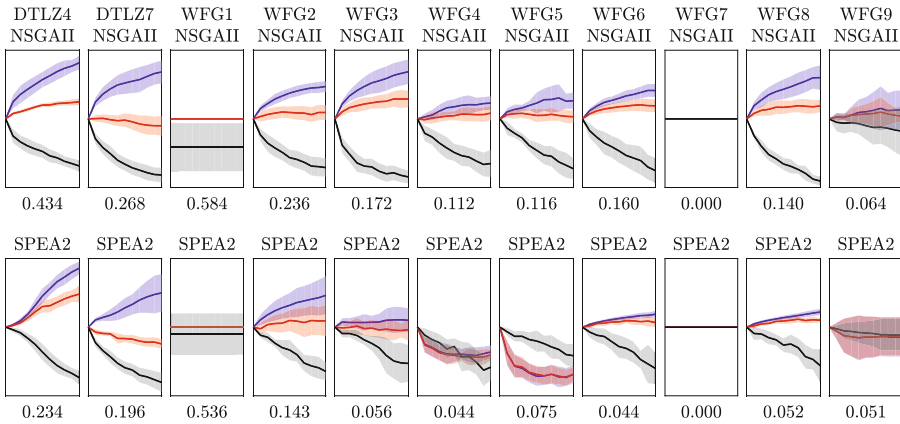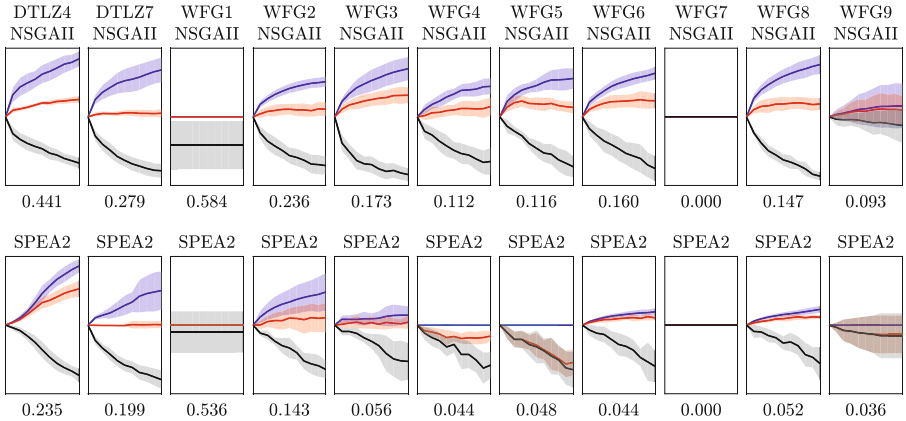

Fig. 7. Diversity of solutions increases markedly under the weighting gradient flow, even as some points become slightly dominated. (Top) Average diversity quotients of feasible (blue) and non-dominated (red) points under the weighting gradient flow along with proportion of population that remains non-dominated (black). Here the diversity is the magnitude at scale t_+. Shaded bands indicate one standard deviation. All panels have the same horizontal axis, viz., the number of timesteps (from 0 to $N = 10$). The vertical axes are $[1 - \Delta, 1 + \Delta]$, with Δ shown below each panel. Not shown explicitly is the average weighting of non-dominated points, i.e., the red curve divided by the black one, but so long as the colored bands already shown are visibly separate, this consistently lies above the blue band. (Bottom) As for the top panels, but for SPEA2. (Color figure online)

Although Figs. 7 and 8 shows that the weighting gradient flow causes a significant proportion of points to become dominated, Fig. 9 uses the *inverted generational distance* (IGD) relative to uniformly distributed reference points on Pareto fronts [26] to show that this qualitative change in dominance is belied by only minor quantitative changes in the distance to Pareto fronts.[2] (Note that the relatively large increases in IGD for DTLZ4 and DTLZ7 are consequences of starting from a low baseline.) That is, feasible points give a better quantitative sense of diversification performance than nondominated points, especially in light of use cases in which the weighting gradient flow is not limited to postprocessing.

---

[2] Recall that the IGD for X relative to reference set R is $\frac{1}{|R|} \sum_{r \in R} \min_{x \in X} d(x, r)$.

DTLZ4 NSGAII	DTLZ7 NSGAII	WFG1 NSGAII	WFG2 NSGAII	WFG3 NSGAII	WFG4 NSGAII	WFG5 NSGAII	WFG6 NSGAII	WFG7 NSGAII	WFG8 NSGAII	WFG9 NSGAII
0.441	0.279	0.584	0.236	0.173	0.112	0.116	0.160	0.000	0.147	0.093

SPEA2	SPEA2	SPEA2	SPEA2	SPEA2	SPEA2	SPEA2	SPEA2	SPEA2	SPEA2	SPEA2
0.235	0.199	0.536	0.143	0.056	0.044	0.048	0.044	0.000	0.052	0.036

Fig. 8. As in Fig. 7, but for diversity taken as the magnitude at the scale maximizing the quotient by the initial timestep.

Rather than relying solely on a delicate characterization of diversity, we also visualize some of the results directly: this is the rationale for three-objective problems. Figure 10 shows how diversity in objective space is promoted for WFG2-3. Figure 11 shows analogous results for SPEA2.

Careful inspection reveals that the weighting gradient flow tends to induce a gap between the boundary of the non-dominated region and its interior, which is consistent with the generally observed phenomenon that the largest weights in finite subsets of Euclidean space tend to occur on boundaries and the smallest weights immediately "behind" the boundary. Meanwhile, the boundary region tends to become slightly more populated.[3] From the perspective of a MOEA, this is frequently a benefit, since extremal and non-extremal points on the non-dominated approximation of the Pareto front differ in practical significance.[4]

6   Algorithmic Extensions

6.1   Multi-objective Weighting Gradient Flow

We can combine the weighting gradient flow with a multi-gradient descent strategy in a way somewhat akin to [28]. The basic additional ideas are:

– Introduce variable regularizing terms λ_w and λ_f for the weighting and function gradient flows, respectively;

---

[3] This highlights the need to distinguish between diversity and uniformity. The maximally diverse probability distribution on the interval $[0, L]$ is $\frac{1}{2+L}(\delta_0 + \lambda|_{[0,L]} + \delta_L)$, where Dirac and restricted Lebesgue measures are indicated on the right hand side [27]. Only in a suitable limit can boundary effects be ignored in relation to diversity.

[4] Using a scale $t > t_+$ for the weighting gradient flow would tend to diminish the distinction between uniformity (which is not a function of scale) and diversity (which is). That is, our experiments make this distinction to the greatest possible extent.
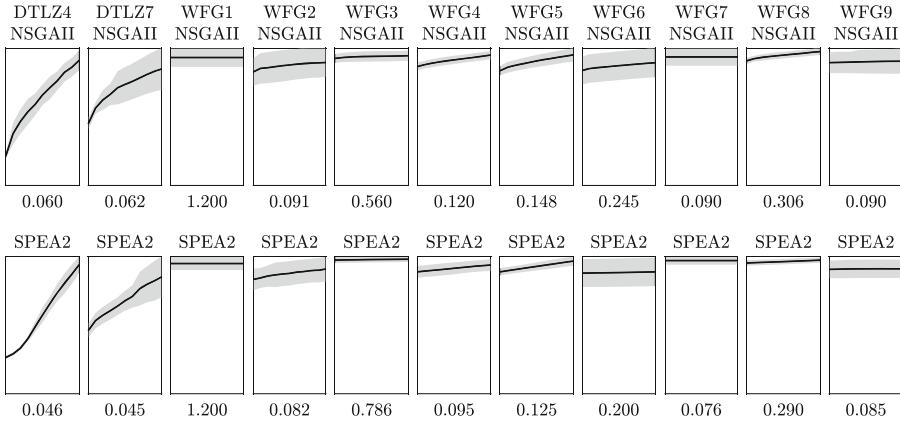
Fig. 9. The weighting gradient flow only slightly affects the quantitative dominance behavior of points, as measured by IGD. (Top) IGD under the weighting gradient flow starting from the results of NSGA-II runs, using uniformly distributed reference points on Pareto fronts. Shaded bands indicate one standard deviation. All panels have the same horizontal axis, viz., the number of timesteps (from 0 to $N = 10$). The vertical axes are $[0, y]$, where y is shown below each panel. (Bottom) As above, but for SPEA2.

– Form the objective-space differentials $dy_j = ds \cdot [\lambda_w S_j(\hat{\nabla} w)_j + \lambda_f \sum_\ell (\hat{\nabla} f)_\ell]$, where the sum is over ℓ such that $\langle (\hat{\nabla} w)_j, (\hat{\nabla} f_\ell)_j \rangle > 0$.

While we have tried this technique in isolation on MOEA benchmarks, the results are poor. However, this is unsurprising: the benchmarks are designed to frustrate MOEAs, much less multi-objective techniques relying on gradients.

6.2   Recycling Function Evaluations

In our experiments with post-processing the output of MOEAs, the weighting gradient flow evolution took time comparable to (and in the case of NSGA-II, slightly more than) the MOEA itself. Most of the time is spent evaluating the fitness function: apart from an initialization step, the evaluations are performed to compute Jacobians in service of pullback operations, and a lesser number are performed to compute pushforwards to maintain consistency.

However, our motivating problems require significant time (on the order of a second) for function evaluations. This demands a more efficient pullback scheme that minimizes or avoids function evaluations, even if the results are substantially worse. A reasonable idea is "recycling" in a sense similar to that employed in some modern Monte Carlo algorithms [29]. Specifically, rather than computing a good approximation to the Jacobian by evaluating functions afresh at very close points along coordinate axes, we settle instead for an approximation of lesser quality that exploits existing function evaluations. We have implemented this in concert with a *de novo* computation of the Jacobian in the event that this
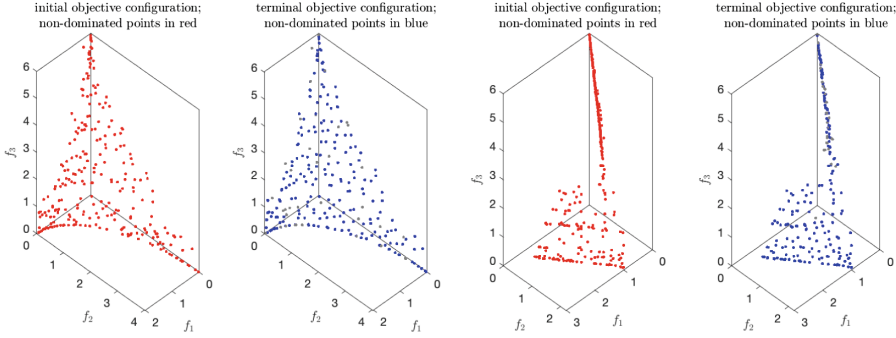
Fig. 10. (Far left) Initial configuration in objective space for WFG2 after a NSGA-II run. (Center left) Configuration after subsequently evolving under the weighting gradient flow. Dominated points are gray. (Right panels) As on the left, but for WFG3.
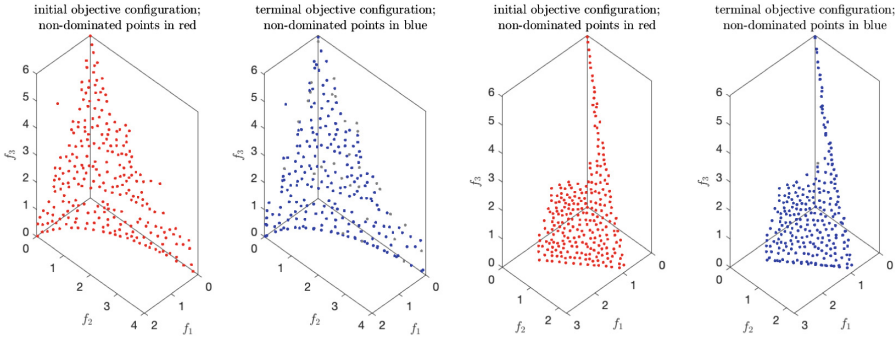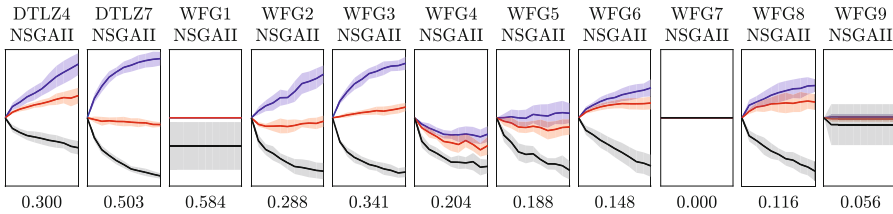

Fig. 11. As in Fig. 10, but for SPEA2. Note the formation of gaps behind the boundary.

initial Jacobian estimate does not have full rank. Our experiments suggest that this works reasonably well: for a typical run from Sect. 5, the number of function evaluations is reduced from 30250 to 2750, and the actual results are broadly comparable (sometimes better, sometimes worse): see Figs. 12 and 13.
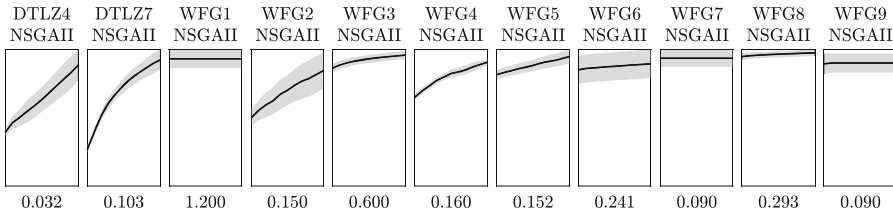
This strategy will work poorly if evaluation points lie on a manifold of nonzero codimension or low curvature, because in such cases a matrix that transforms vectors from a base point to evaluation points into (a small multiple of) the standard basis will have a large condition number. However, these situations are relatively unlikely to present major problems in practice, and the recycling approach is likely to be useful when function evaluations are expensive.

7   Remarks

Although our experiments have focused on the results of applying the weighting gradient flow and related constructions after a MOEA has been applied, the more natural application is in the course of a MOEA. As mentioned in Sect. 6, there

DTLZ4 NSGAII	DTLZ7 NSGAII	WFG1 NSGAII	WFG2 NSGAII	WFG3 NSGAII	WFG4 NSGAII	WFG5 NSGAII	WFG6 NSGAII	WFG7 NSGAII	WFG8 NSGAII	WFG9 NSGAII
0.300	0.503	0.584	0.288	0.341	0.204	0.188	0.148	0.000	0.116	0.056

Fig. 12. As in the top panel of Fig. 7, but for a Jacobian approximation that uses existing function evaluations, increasing speed at the cost of accuracy.

DTLZ4 NSGAII	DTLZ7 NSGAII	WFG1 NSGAII	WFG2 NSGAII	WFG3 NSGAII	WFG4 NSGAII	WFG5 NSGAII	WFG6 NSGAII	WFG7 NSGAII	WFG8 NSGAII	WFG9 NSGAII
0.032	0.103	1.200	0.150	0.600	0.160	0.152	0.241	0.090	0.293	0.090

Fig. 13. As in Fig. 9, but for a Jacobian approximation that uses existing function evaluations, increasing speed at the cost of accuracy.

is ample scope to refine and build on ideas for increasing weighting components in specific contexts. It is nevertheless clear that the theory of magnitude informs principled and practical diversity-promoting mechanisms that can already be usefully applied to benchmark multi-objective problems.

References

1. Eiben, A.E., Smith, J.E.: Introduction to Evolutionary Computing. Springer, Heidelberg (2015). https://doi.org/10.1007/978-3-662-44874-8
2. Basto-Fernandes, V., Yevseyeva, I., Deutz, A., Emmerich, M.: A survey of diversity oriented optimization: problems, indicators, and algorithms. In: Emmerich, M., Deutz, A., Schütze, O., Legrand, P., Tantar, E., Tantar, A.-A. (eds.) EVOLVE – A Bridge between Probability, Set Oriented Numerics and Evolutionary Computation VII. SCI, vol. 662, pp. 3–23. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-49325-1_1
3. Yan, J., Li, C., Wang, Z., Deng, L., Sun, D.: Diversity metrics in multi-objective optimization: review and perspective. In: 2007 IEEE International Conference on Integration Technology, pp. 553–557. IEEE (2007)

4. Ulrich, T., Bader, J., Thiele, L.: Defining and optimizing indicator-based diversity measures in multiobjective search. In: Schaefer, R., Cotta, C., Kołodziej, J., Rudolph, G. (eds.) PPSN 2010. LNCS, vol. 6238, pp. 707–717. Springer, Heidelberg (2010). https://doi.org/10.1007/978-3-642-15844-5_71

5. Ulrich, T., Thiele, L.: Maximizing population diversity in single-objective optimization. In: Proceedings of the 13th Annual Conference on Genetic and Evolutionary Computation, pp. 641–648 (2011)

6. Solow, A.R., Polasky, S.: Measuring biological diversity. Environ. Ecol. Stat. **1**(2), 95–103 (1994)

7. Leinster, T.: Entropy and Diversity: the Axiomatic Approach. Cambridge (2021)

8. Zitzler, E., Künzli, S.: Indicator-based selection in multiobjective search. In: Yao, X., et al. (eds.) PPSN 2004. LNCS, vol. 3242, pp. 832–842. Springer, Heidelberg (2004). https://doi.org/10.1007/978-3-540-30217-9_84

9. Wang, Y., Emmerich, M., Deutz, A., Bäck, T.: Diversity-indicator based multiobjective evolutionary algorithm: DI-MOEA. In: Deb, K., et al. (eds.) EMO 2019. LNCS, vol. 11411, pp. 346–358. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-12598-1_28

10. Pugh, J.K., Soros, L.B., Stanley, K.O.: Quality diversity: a new frontier for evolutionary computation. Front. Robot. AI **3**, 40 (2016)

11. Steinwart, I., Christmann, A.: Support Vector Machines. Springer, Heidelberg (2008). https://doi.org/10.1007/978-0-387-77242-4

12. Willerton, S.: Heuristic and computer calculations for the magnitude of metric spaces. arXiv preprint arXiv:0910.5500 (2009)

13. Bunch, E., Dickinson, D., Kline, J., Fung, G.: Practical applications of metric space magnitude and weighting vectors. arXiv preprint arXiv:2006.14063 (2020)

14. Leinster, T., Meckes, M.W.: The magnitude of a metric space: from category theory to geometric measure theory. In: Gigli, N. (ed.) Measure Theory in Non-Smooth Spaces (2017)

15. Leinster, T., Meckes, M.W.: Maximizing diversity in biology and beyond. Entropy **18**(3), 88 (2016)

16. Leinster, T.: The magnitude of metric spaces. Doc. Math. **18**, 857–905 (2013)

17. Luo, C., Safa, I., Wang, Y.: Approximating gradients for meshes and point clouds via diffusion metric. In: Computer Graphics Forum, vol. 28, no. 5, pp. 1497–1508. Wiley Online Library (2009)

18. Ishibuchi, H., Yamane, M., Akedo, N., Nojima, Y.: Two-objective solution set optimization to maximize hypervolume and decision space diversity in multiobjective optimization. In: The 6th International Conference on Soft Computing and Intelligent Systems, and The 13th International Symposium on Advanced Intelligence Systems, pp. 1871–1876. IEEE (2012)

19. Guariso, G., Sangiorgio, M.: Improving the performance of multiobjective genetic algorithms: an elitism-based approach. Information **11**(12), 587 (2020)

20. Deb, K., Pratap, A., Agarwal, S., Meyarivan, T.: A fast and elitist multiobjective genetic algorithm: NSGA-II. IEEE Trans. Evol. Comput. **6**(2), 182–197 (2002)

21. Zitzler, E., Laumanns, M., Thiele, L.: SPEA2: improving the strength Pareto evolutionary algorithm. TIK-report, vol. 103 (2001)

22. Deb, K., Thiele, L., Laumanns, M., Zitzler, E.: Scalable test problems for evolutionary multi-objective optimization. TIK-report, vol. 112 (2001)

23. Hamada, N., Goto, K.: Data-driven analysis of Pareto set topology. In: Proceedings of the Genetic and Evolutionary Computation Conference, pp. 657–664 (2018)

24. Huband, S., Hingston, P., Barone, L., While, L.: A review of multiobjective test problems and a scalable test problem toolkit. IEEE Trans. Evol. Comput. **10**(5), 477–506 (2006)
25. Tian, Y., Cheng, R., Zhang, X., Jin, Y.: PlatEMO: a MATLAB platform for evolutionary multi-objective optimization [educational forum]. IEEE Comput. Intell. Mag. **12**(4), 73–87 (2017)
26. Tian, Y., Xiang, X., Zhang, X., Cheng, R., Jin, Y.: Sampling reference points on the Pareto fronts of benchmark multi-objective optimization problems. In: 2018 IEEE congress on evolutionary computation (CEC), pp. 1–6. IEEE (2018)
27. Leinster, T., Roff, E.: The maximum entropy of a metric space. arXiv preprint arXiv:1908.11184 (2019)
28. Désidéri, J.-A.: Multiple-gradient descent algorithm (MGDA) for multiobjective optimization. C.R. Math. **350**(5–6), 313–318 (2012)
29. Frenkel, D.: Speed-up of Monte Carlo simulations by sampling of rejected states. Proc. Natl. Acad. Sci. **101**(51), 17 571–17 575 (2004)
30. Leinster, T., Cobbold, C.A.: Measuring diversity: the importance of species similarity. Ecology **93**(3), 477–489 (2012)