SPREAD : Sampling-based Pareto front Refinement via Efficient Adaptive Diffusion

Sedjro Salomon Hotegni[1,2], Sebastian Peitz[1,2]

TU Dortmund[1], Lamarr Institute[2]

SPREAD at a Glance

Goal: Efficiently approximate the Pareto front of multi-objective optimization problems by iteratively refining candidate solutions with a diffusion-based generative process.

Core claim: A conditional diffusion model guided by multi-gradient descent directions and diversity-promoting repulsion can efficiently generate well-distributed, high-quality Pareto optimal solutions across diverse optimization settings.

Problem: Multi-Objective Optimization (MOO)

$$\min_{\mathbf{x} \in \mathcal{X}} \mathbf{F}(\mathbf{x}) = \begin{pmatrix} f_1(\mathbf{x}) \\ \vdots \\ f_m(\mathbf{x}) \end{pmatrix}$$

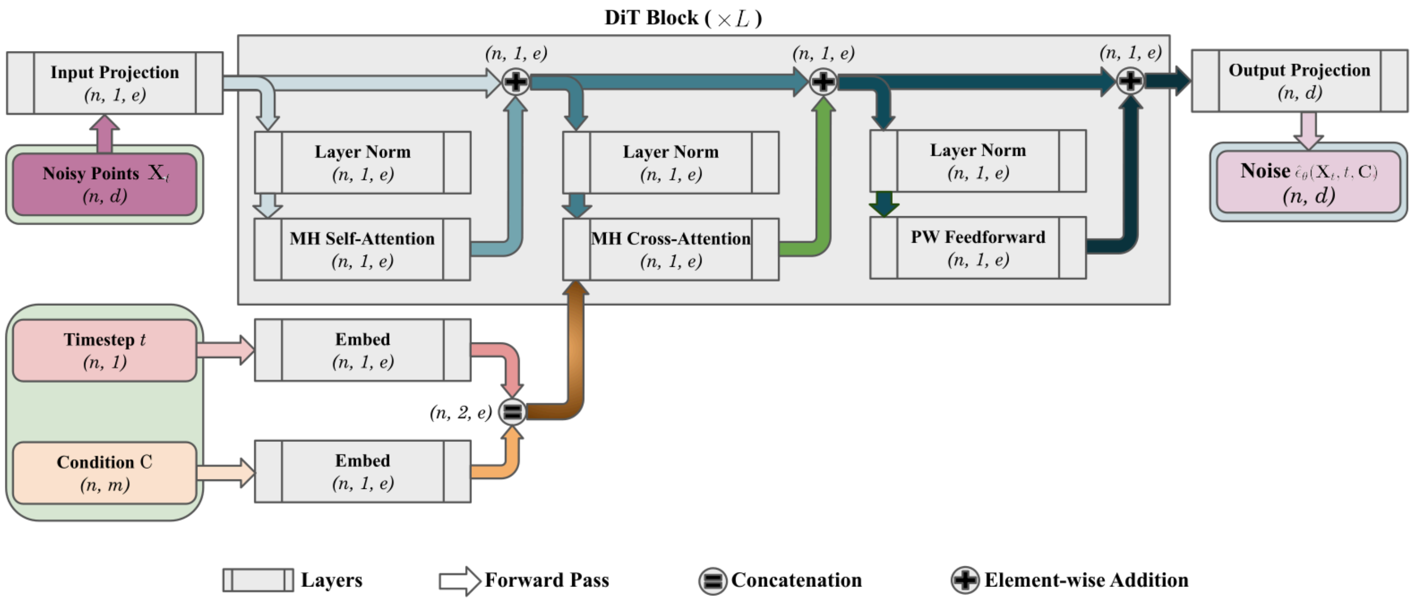
- We seek a **set** of trade-off solutions (Pareto front), not a single optimum.
- Two requirements must be met jointly:
  - **Convergence:** solutions should approach Pareto optimality.
  - **Diversity:** solutions should cover the front well.
- Existing gradient-based methods often converge but lose diversity.

Why Conditional Diffusion for MOO?

- Diffusion models are strong at learning complex multimodal distributions.
- Here, data are vectors in $\mathbb{R}^d$, and conditions encode objective values.
- Conditioning enables targeted sampling toward better objective regions.

Why Conditional Diffusion for MOO?

- Diffusion models are strong at learning complex multimodal distributions.
- Here, data are vectors in $\mathbb{R}^d$, and conditions encode objective values.
- Conditioning enables targeted sampling toward better objective regions.
- This motivates **DiT-MOO** as a generative backbone for Pareto front refinement.



DiT-MOO architecture

SPREAD: Method Overview

Training

- Sample points in $\mathcal{X}$, evaluate objectives, and train conditional DiT denoiser.
- Use positively shifted labels $\mathbf{C} = \mathbf{F}(\mathbf{X}) + \Xi$ to improve robustness.

Sampling (Refinement)

- Start from random population $\mathbf{X}_T$ and run reverse diffusion.
- At each step, add adaptive guidance to denoising update $(\mathbf{X}_t')$:

$$\mathbf{X}_{t-1} \leftarrow \mathbf{X}_t' - \eta_t \tilde{\mathbf{h}}_t(\mathbf{X}_t').$$

- Keep top non-dominated points with crowding distance.

Adaptive Guidance in SPREAD

$$\tilde{\mathbf{h}}_t = \mathbf{h}_t + \gamma_t^\top \delta_t$$

- $\mathbf{h}_t$: aligns with MGD-style descent directions to improve objectives and encourage spread in objective space.
- $\gamma_t^\top \delta_t$: adds random perturbation δ_t to avoid collapse and maintain exploration; the adaptive scale γ_t controls perturbation strength while preserving the effect of $\mathbf{h}_t$.

Theory (Key Guarantees)

Objective Improvement (Theorem 1):

- If the conditional sampler approximates the true conditional distribution within total-variation error τ, then
- sampled points dominate the initialization with probability at least $1 - \tau$.

Convergence Direction (Theorem 2):

- Under smooth objectives and suitable γ_t scaling, $-\tilde{\mathbf{h}}_t$ is a common descent direction for all objectives.

Experimental Protocol

- Three settings:
  - Online MOO
  - Offline MOO (no access to the true objective functions)
  - Bayesian MOO (limited expensive evaluations)
- Strong baselines include PMGDA, MOO-SVGD, HVGrad, ParetoFlow, PGD-MOO, CDM-PSL.
- Metrics:
  - Hypervolume (↑), Δ-spread (↓), and LHD in BO (↓).

Main Results

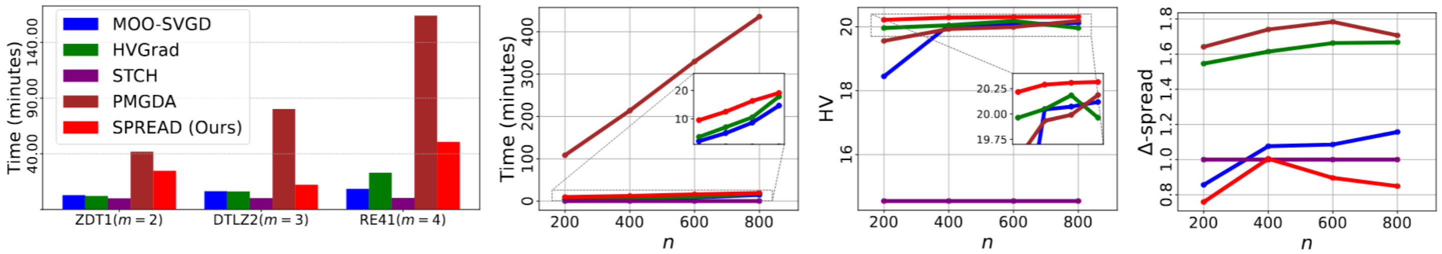
Table 1: Hypervolume results averaged over 5 independent runs. The best values are **bold**.

HV (↑)	$m=2$				$m=3$						$m=4$
Method	**ZDT1**	**ZDT2**	**ZDT3**	**RE21**	**DTLZ2**	**DTLZ4**	**DTLZ7**	**RE33**	**RE34**	**RE37**	**RE41**
PMGDA	**5.72±0.00**	**6.22±0.00**	5.85±0.00	48.14±0.00	**22.97±0.00**	19.69±0.20	17.82±0.00	43.06±0.00	210.07±0.00	1.18±0.00	901.90±3.36
MOO-SVGD	5.70±0.00	6.21±0.00	6.08±0.02	20.43±0.32	22.61±0.02	19.69±0.62	13.57±0.03	16.26±0.17	156.20±0.57	1.05±0.09	579.53±6.42
STCH	5.71±0.00	5.89±0.00	5.44±0.13	19.07±0.00	22.92±0.01	14.55±0.00	17.46±0.00	12.14±0.00	156.72±0.00	1.31±0.02	506.33±2.86
HVGrad	**5.72±0.00**	**6.22±0.00**	**6.10±0.00**	43.65±0.00	22.93±0.00	19.98±0.04	17.48±0.05	36.13±0.00	156.72±0.00	**1.44±0.00**	936.17±8.91
SPREAD	**5.72±0.00**	**6.22±0.00**	**6.10±0.00**	**70.10±0.01**	22.91±0.00	**20.22±0.01**	**18.07±0.01**	**133.76±1.72**	**243.15±0.49**	1.42±0.00	**1008.75±6.30**

Table 2: Results of the Δ-spread diversity measure. The best value, along with those whose mean falls within one standard deviation of it, are shown in **bold**.

Δ-spread (↓)	$m=2$				$m=3$						$m=4$
Method	**ZDT1**	**ZDT2**	**ZDT3**	**RE21**	**DTLZ2**	**DTLZ4**	**DTLZ7**	**RE33**	**RE34**	**RE37**	**RE41**
PMGDA	0.42±0.17	**0.23±0.01**	1.57±0.02	1.53±0.00	**0.66±0.02**	1.71±0.07	1.02±0.08	1.11±0.00	1.46±0.00	0.59±0.01	1.46±0.01
MOO-SVGD	0.78±0.20	1.16±0.11	0.90±0.08	1.01±0.00	1.31±0.01	1.02±0.09	0.71±0.03	1.00±0.00	1.20±0.17	0.58±0.07	1.13±0.04
STCH	1.01±0.04	1.00±0.00	1.05±0.03	1.00±0.00	1.00±0.04	1.00±0.00	1.06±0.05	1.00±0.00	1.00±0.00	0.80±0.04	1.38±0.02
HVGrad	0.36±0.05	1.07±0.05	1.08±0.10	1.00±0.00	1.18±0.05	1.56±0.06	**0.66±0.03**	1.00±0.00	1.00±0.00	**0.51±0.01**	1.00±0.02
SPREAD	**0.32±0.01**	0.29±0.02	**0.53±0.01**	**0.44±0.02**	0.93±0.05	**0.80±0.06**	**0.69±0.05**	**0.97±0.02**	**0.88±0.03**	0.80±0.01	**0.92±0.03**

Online MOO



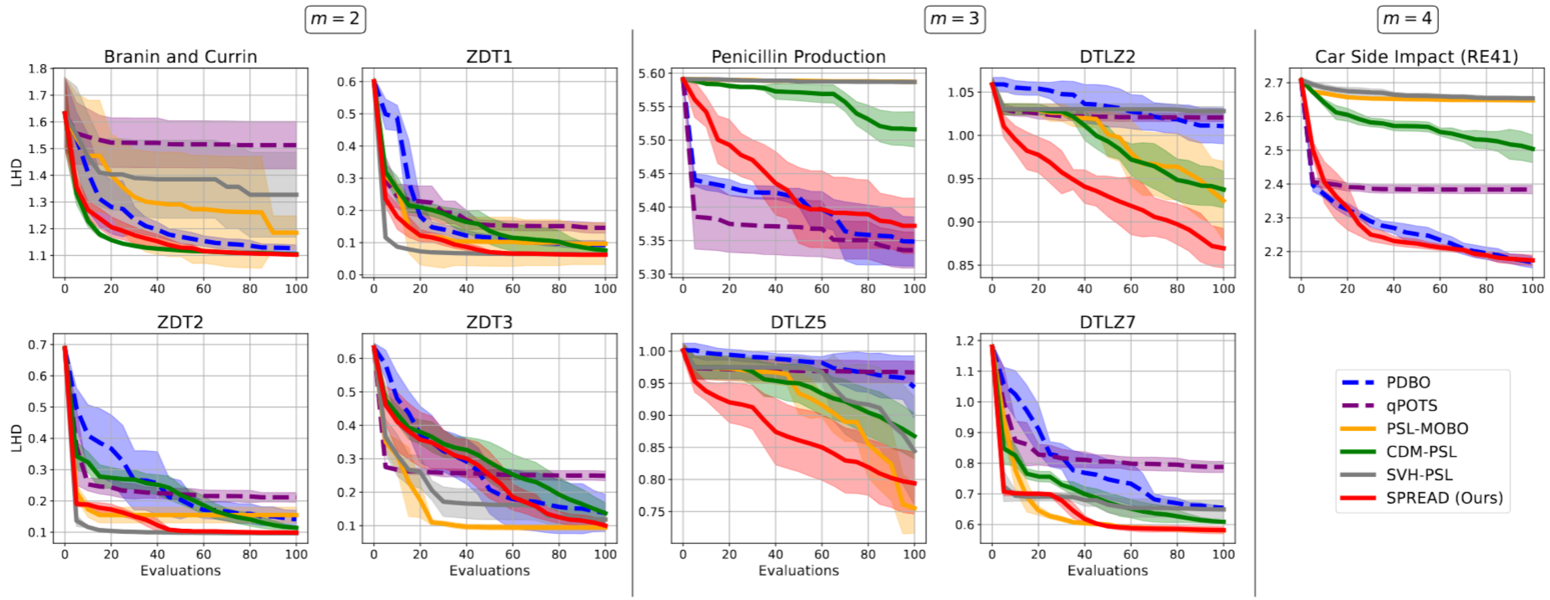
(a) (b) (c) (d)

SPREAD remains competitive while scaling in objectives and sample size.

Main Results

Method	Synthetic	RE
$\mathcal{D}$(best)	9.08	11.83
MM	5.92	3.92
MM-COM	8.00	7.42
MM-IOM	5.67	4.33
MM-ICT	6.50	3.83
MM-RoMA	6.08	7.75
MM-TriMentoring	8.17	4.58
MH	7.58	5.67
MH-PcGrad	5.75	6.92
MH-GradNorm	9.58	11.50
ParetoFlow	7.50	6.83
PGD-MOO	4.58	8.75
SPREAD	**3.50**	**1.83**

Offline MOO (avg rank)



Bayesian MOO (LHD)

Ablation study

Table 3: **Ablation study** on the diversity-promoting mechanisms in SPREAD. Best values are highlighted in **bold**. For Δ-spread, any mean value within one standard deviation of the best is also shown in **bold**. Worst values are shown in red, while best values are shown in blue (HV) and green (Δ-spread).

Problem	SPREAD		SPREAD (w/o diversity)		SPREAD (w/o perturbation)		SPREAD (w/o repulsion)	
	HV	Δ-spread	HV	Δ-spread	HV	Δ-spread	HV	Δ-spread
ZDT1	**5.72±0.00**	**0.32±0.01**	5.06±0.00	+∞	**5.72±0.00**	**0.32±0.02**	4.25±0.08	0.88±0.05
ZDT2	**6.22±0.00**	**0.29±0.02**	5.89±0.00	+∞	**6.22±0.00**	**0.28±0.02**	4.40±0.14	+∞
ZDT3	**6.10±0.00**	0.53±0.01	5.06±0.00	0.66±0.00	**6.10±0.00**	**0.51±0.01**	4.34±0.07	0.84±0.05
RE21	**70.10±0.01**	0.44±0.02	70.03±0.01	**0.41±0.02**	69.01±0.14	0.84±0.05	70.03±0.03	0.51±0.03
DTLZ2	22.91±0.00	0.93±0.05	**22.94±0.00**	**0.73±0.03**	22.8±0.01	0.91±0.07	22.79±0.04	1.06±0.08
DTLZ4	20.22±0.01	**0.80±0.06**	**20.36±0.01**	0.89±0.11	20.01±0.02	0.89±0.2	20.34±0.02	0.97±0.15
DTLZ7	**18.07±0.01**	**0.69±0.05**	16.7±0.00	+∞	18.05±0.01	0.80±0.03	12.84±0.33	0.87±0.04
RE33	**133.76±1.72**	**0.97±0.02**	8.72±0.65	**0.99±0.00**	125.06±0.46	**0.97±0.04**	99.89±9.7	1.04±0.17
RE34	**243.15±0.49**	0.88±0.03	236.86±0.94	0.97±0.03	242.47±0.22	0.99±0.02	237.34±0.77	**0.82±0.05**
RE37	**1.42±0.00**	0.80±0.01	1.32±0.00	0.98±0.03	**1.42±0.00**	**0.75±0.03**	1.40±0.00	0.79±0.05
RE41	1008.75±6.3	0.92±0.03	950.45±7.32	**0.81±0.10**	969.43±6.44	0.93±0.03	1011.03±7.52	**0.78±0.06**

Conclusion

- SPREAD combines conditional diffusion and adaptive multi-objective guidance.
- It improves both convergence and diversity of approximate Pareto sets.
- Empirically strong across online, offline, and Bayesian MOO benchmarks.

Project page

Thank You!