SURF: Steering the Scalarization Weight to Uniformly Traverse the Pareto Front

Liuyuan Jiang, Chentong Huang, Lisha Chen
Department of Electrical and Computer Engineering
University of Rochester, Rochester, NY 14627
{ljiang24, chuang80}@ur.rochester.edu, lisha.chen@rochester.edu

Abstract

Scalarization is widely used in multi-objective optimization owing to its simplicity and scalability. In many applications, the goal is to generate solutions that represent diverse user preferences, ideally with uniform coverage of the Pareto front (PF). However, uniformly sampling scalarization weights usually induces non-uniform coverage of the PF. We explain this mismatch through the geometry of the scalarization path. As the weight varies, the corresponding solutions traverse the PF at a generally nonuniform speed. Integrating this speed gives cumulative arc length. After normalization by the total length, it defines a cumulative distribution function (CDF) on weights whose inverse selects points at equal arc-length intervals. Building on this insight, we propose **SURF** (**S**ampling **U**niformly along the Par**e**to **F**ront). For structured problems, including bi-objective bandits, we derive analytical expressions for this CDF map and the resulting PF-aware weight sampling rule. For general problems, SURF alternates between CDF reconstruction and weight sampling. Theoretically, we show that under provable conditions for bi-objective LS, SURF's CDF reconstruction error converges linearly to an $\mathcal{O}(N^{-2})$ discretization floor. Empirically, experiments on bandits, multi-objective-gymnasium, and multi-objective LLM alignment demonstrate that SURF efficiently achieves more uniform PF coverage than baselines. The code is available at `https://github.com/Liuyuan999/MOO_Uniform_PF`.

1 Introduction

Multi-objective optimization (MOO) is increasingly central to modern machine learning and large language model (LLM) alignment, where one must balance competing criteria such as accuracy, robustness, fairness, and safety rather than optimize a single scalar reward [86, 102, 104, 111]. Formally, let $\mathbf{u}$ denote the decision variable in the feasible set $\mathcal{U} \subseteq \mathbb{R}^d$. MOO seeks to minimize M objectives $h_m : \mathbb{R}^d \to \mathbb{R}$:

$$\min_{\mathbf{u}\in\mathcal{U}} \left\{ \mathbf{h}(\mathbf{u}) := [h_1(\mathbf{u}), \cdots, h_M(\mathbf{u})]^\top \right\}. \tag{1}$$

We consider optimizing the above vector-valued objective to Pareto optimality. Intuitively, a solution $\mathbf{u}^*$ is *Pareto optimal* if no feasible point improves one objective without worsening another. The set of all Pareto-optimal solutions is the *Pareto set* $\mathcal{U}_{\mathrm{PF}}$, and its image $\mathcal{F}_{\mathrm{PF}}$ under $\mathbf{h}$ is the Pareto front (PF). Formal definitions are provided in Section 2. The PF thus provides a compact description of trade-offs via M objective values, typically with $M \ll d$. In *a priori* decision making, known preferences guide optimization toward a desired trade-off. In *a posteriori* decision making, users choose after inspecting a representative set of PF

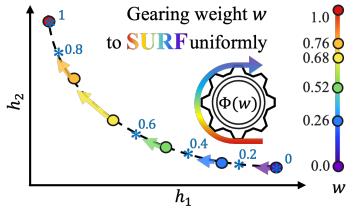

Figure 1: Multi-armed bandit LS results (details in Appendix F.1). Uniform weights produce uneven points on PF (stars), while PF-aware sampling yields uniformly spaced colored circles.

40th Conference on Neural Information Processing Systems (NeurIPS 2026).

solutions, motivating uniform coverage. A common practice to solve MOO is via *scalarization*. For example, in the bi-objective case ($M = 2$), given $w \in [0, 1]$, linear scalarization (LS) solves

$$\mathbf{u}_w^* \in \arg\min_{\mathbf{u}\in\mathcal{U}} \left\{ \mathrm{LS}_{\mathbf{h}}(\mathbf{u}; w) := wh_1(\mathbf{u}) + (1-w)h_2(\mathbf{u}) \right\}. \tag{2}$$

Under certain conditions (see Section 2), sweeping w over $[0, 1]$ traces out the entire PF, so the standard baseline uses evenly spaced weights [76]. Yet, as Figure 1 shows, this baseline rarely yields evenly spaced PF points: solutions could clump in some regions of the front and leave others almost untouched. Existing remedies are either substantially more involved than scalarization [24, 59, 62, 90, 92, 110] or offer no coverage guarantee in limited-computation-budget regimes [5, 16, 67, 110]. This motivates the question:

*Can we **efficiently** find solutions with objectives **uniformly distributed along the PF**?*

We answer this by proposing **S**ampling **U**niformly along the Pa**r**eto **F**ront (SURF; Figure 2), a plug-in wrapper around scalarization that leaves the inner solver untouched with preserved efficiency, and only adjusts *how the weights are sampled*. The key insight is the relation between the weight w and the PF geometry. Picture the scalarization weight w as a dial that drags a point along the PF: as w slides from 0 to 1, the LS solution $\mathbf{u}_w^*$ traces the entire front via $\mathbf{f}_{\mathrm{PF}}(w) := \mathbf{h}(\mathbf{u}_w^*)$. The catch is that the dial turns the front at uneven speed, i.e., in some regions a small change in w sweeps a long stretch of PF, while in others a large change barely budges the point (cf. Figure 1). To precisely characterize this relation, define the *traversal speed* $v(w) := \left\| \frac{\partial}{\partial w} \mathbf{f}_{\mathrm{PF}}(w) \right\|$ that captures how fast the PF geometry changes at w, the *PF arc length* $s(w) := \int_0^w v(p)\, dp$ that captures the traversal distance along the PF when turning the weight from 0 to w, and the *normalized cumulative arc-length* $\Phi(w) := s(w)/s(1)$ that specifies the fraction of the PF reached at w, which is the CDF of the weight parameter induced by the uniform probability measure on the PF with respect to arc length. The Euclidean norm measures PF traversal in (h_1, h_2), using the compact objective-space representation with $M \ll d$. The cure mirrors the diagnosis: instead of turning the dial uniformly, SURF turns it through Φ^{-1}. This yields the following PF-geometry-aware weight sampling rule:

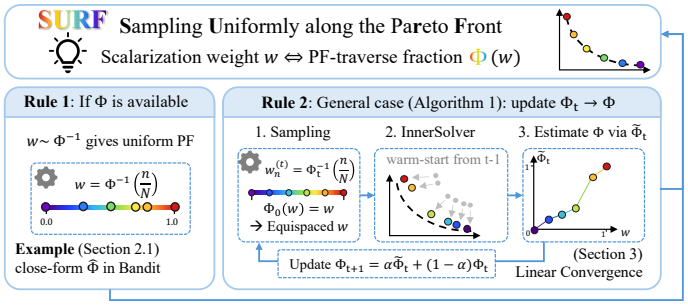

Figure 2: SURF overview: bridging w and PF points via CDF Φ yields two rules. Rule 1 directly samples $\{w_n\}_{n=0}^N$ when Φ is known (Section 2). Rule 2 handles the general case by iteratively approximating Φ (Algorithm 1).

Rule 1. *If Φ is available, SURF chooses $\{w_n\}_{n=0}^N$ given by $w_n = \Phi^{-1}\left(\frac{n}{N}\right)$.*

A priori preferences between objectives can similarly be expressed as desired normalized arc-length positions $q_n \in [0, 1]$ on the PF, giving $w_n = \Phi^{-1}(q_n)$. Section 2.1 carries this out in structured settings: for bi-objective reinforcement learning, we derive a closed-form $v(w)$ in terms of $\mathbf{u}_w^*$, and an analytically available Φ in the bandit specialization. For general problems where Φ is not analytically available, Section 3 introduces SURF (Algorithm 1), which alternates between solving the scalarized problem at the current set of weights and refining an estimate of Φ that drives the next set of refined weights. This yields the second joint estimation and sampling rule:

Rule 2. *If Φ is not available, SURF estimates Φ and the weights $\{w_n\}_{n=0}^N$ jointly.*

SURF estimates $s(w)$ by accumulating neighboring objective-space distances, leaving parameter-space optimization to the inner solver. Our contributions are summarized as follows:

C1) PF geometry characterization. In Section 2, we characterize the scalarization-induced PF geometry through the traversal speed $v(w)$ and CDF Φ, leading to the PF-aware weight sampling principle in Rule 1, and use the bandit setting as a concrete instantiation of this principle with analytical expressions (Section 2.1).

C2) SURF algorithm with convergence guarantee. In Section 3, we propose SURF (Algorithm 1), which realizes Rule 2 by reconstructing the unknown CDF Φ from finitely many inexact, warm-

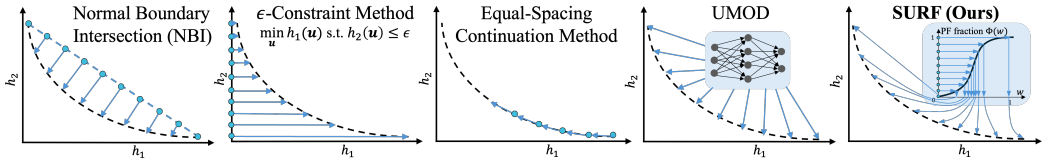

Figure 3: Comparison of approaches for obtaining diverse PF solutions, including NBI method [24], ϵ-constraint method [14], continuation/equal-spacing methods [71, 59], and UMOD [110]. These methods promote coverage through auxiliary constraints, directions, path tracing, or learned mappings, whereas SURF directly matches the scalarization weights to the uniform PF distribution.

started scalarized solves. Theorem 1 shows that the CDF reconstruction error converges to an unavoidable $\mathcal{O}(N^{-2})$ finite-sampling floor.

C3) Empirical verification. In Section 4, we empirically verify SURF on MO-Gymnasium benchmarks and LLM alignment tasks using different inner solvers, with additional Tchebycheff scalarization tests in Appendix F.2. SURF improves uniform PF coverage while maintaining or improving PF approximation quality over representative baselines under matched training budgets.

Related Work. Many practical applications involving multiple criteria decision making or multi-task learning, such as multi-objective reinforcement learning and LLM alignment, require adaptation to or representation of different user preferences or objective trade-offs rather than a single preferred solution on the PF [77, 34, 54, 86, 42]. A classical approach to solving MOO is scalarization, which converts the original vector-valued objective into a scalar-valued objective through scalarization. This includes LS, Tchebycheff scalarization [48, 55], ϵ-constraint method [14], etc. Among these, LS is a widely used baseline [113, 103, 51]. A separate line of work studies how to obtain diverse or well-distributed solutions along the PF. Representative approaches include adaptive scalarization methods [5, 1], which adjust scalarization weights to discover diverse trade-offs; preference-guided methods [24, 56, 60, 18, 110], which steer the search using prescribed reference directions or preference vectors; and continuation or equal-spacing methods [109, 71, 59], which explicitly trace the PF and attempt to maintain approximately even spacing between neighboring solutions. These methods typically enforce uniformity indirectly through weight selection, reference-point design, or local spacing control, as presented in Figure 3. OLS targets a convex coverage set [5, 1], while NBI introduces auxiliary constrained subproblems [24]. SURF instead corrects sampling density through distribution matching while retaining the scalarized solver. The map from scalarization weight to Pareto point induces a pushforward measure on the front. If we sample scalarization weights from a given distribution, the resulting Pareto points follow a distribution on the PF [41, 46, 94]. Uniform coverage can then be interpreted as matching this induced measure to a uniform arc-length target. An extended discussion of related work is deferred to Appendix A.

Notation. We denote $\mathbb{R}_+ = [0, \infty)$ and $\mathbb{R}_- = (-\infty, 0]$. For a set $\mathcal{X}$, $\mathrm{relint}(\mathcal{X})$ is its relative interior and $\mathrm{Proj}_{\mathcal{X}}(\mathbf{x})$ is the Euclidean projection. For $n \in \mathbb{N}$, write $[n] = \{1, \ldots, n\}$, $\Delta_n = \{\mathbf{w} \in \mathbb{R}^n_+ : \sum_{i=1}^n w_i = 1\}$, and $\Delta_n^\circ = \mathrm{relint}(\Delta_n)$. For $\mathbf{p}, \mathbf{q} \in \Delta_n$, we write $\mathrm{KL}(\mathbf{p} \| \mathbf{q}) := \sum_i p_i \log(p_i/q_i)$ for the Kullback-Leibler (KL) divergence, with the analogous integral form for continuous distributions. For vector $\mathbf{x}$, $\|\mathbf{x}\|$ is the Euclidean norm, $\mathrm{Diag}(\mathbf{x})$ is its diagonal matrix, and $\mathrm{SoftMax}(\mathbf{x})_i = \exp(x_i)/\sum_j \exp(x_j)$, with binary special case $\sigma(x) = (1 + e^{-x})^{-1}$. For matrix H, $\|H\|$ is the spectral norm, $\mathrm{vec}(H)$ is its row-wise vectorization, and $\ker(H)$ is its null space. Detailed problem-specific notation is summarized in Appendix B.

2 Pareto Front Geometry Analysis and Geometry-Aware Scalarization

We use LS as a concrete example for its tractable geometry and full PF reachability under hidden convexity. To make the dial-speed picture precise, we now formalize the regularity conditions under which $\mathbf{u}_w^*$ is well-defined and the PF traversal speed is smooth. We first provide basic definitions for MOO, then discuss additional assumptions for our problem setup.

Definition 1 (Pareto optimality and Pareto front)**.** *A feasible point $\mathbf{u}^* \in \mathcal{U}$ is* Pareto optimal *if there does not exist $\mathbf{u} \in \mathcal{U}$ such that $h_m(\mathbf{u}) \leq h_m(\mathbf{u}^*)$ for all $m \in [M]$ and $h_j(\mathbf{u}) < h_j(\mathbf{u}^*)$ for at least*

one $j \in [M]$. The collection of all Pareto-optimal solutions forms the Pareto set *(PS), denoted as $\mathcal{U}_{\mathrm{PF}}$, and its image in objective space forms the* Pareto front *(PF), $\mathcal{F}_{\mathrm{PF}} := \{\mathbf{h}(\mathbf{u}^*) : \mathbf{u}^* \in \mathcal{U}_{\mathrm{PF}}\}$.*

Modern machine-learning objectives are highly non-convex in their original parameterization. Nevertheless, many problems admit an equivalent hidden representation where the objective becomes convex or strongly convex [72, 32]. Based on this viewpoint, we impose the following assumption.

Assumption 1 (Hidden-space regularity)**.** *Assume there exists a bijection $g : \mathcal{U} \to \mathcal{X}$ such that $h_m(u) = f_m(g(u))$ for each $m \in [M]$. Each f_m is twice continuously differentiable on an open set containing $\mathcal{X}$ and is μ-strongly convex on $\mathcal{X}$. The hidden PS $\mathcal{X}_{\mathrm{PF}} = g(\mathcal{U}_{\mathrm{PF}})$ lies in a compact subset $\mathcal{K} \subset \mathcal{X}$ such that each ∇f_m is $l_{f,1}$-Lipschitz-continuous on $\mathcal{K}$.*

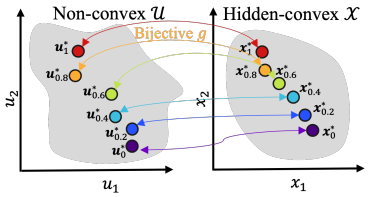

Figure 4: Original parameter space $\mathcal{U}$ and hidden space $\mathcal{X} = g(\mathcal{U})$. Bijection g maps non-convex structures in $\mathcal{U}$ to a convex space $\mathcal{X}$.

For local sensitivity along a regular solution branch, strong convexity can be replaced by a positive-definite reduced Lagrangian Hessian [36]. This assumption is provable in a range of settings, including nonlinear least squares [68, 28], power control and optimal doping profile problems [11], system-level synthesis in control [3], revenue management and inventory control [20, 21], and reinforcement learning [107, 81], which we revisit as a concrete example in Section 2.1.

Denote $\mathbf{x} = g(\mathbf{u})$, $\mathcal{X} = g(\mathcal{U})$, and $\mathbf{f}(\mathbf{x}) := [f_1(\mathbf{x}), \dots, f_M(\mathbf{x})]^\top$. Then MOO in (1) admits the hidden-space form $\min_{\mathbf{x}\in\mathcal{X}} \mathbf{f}(\mathbf{x})$. Hidden strong convexity [32] yields that for any weight $\mathbf{w} \in \Delta_M$, there is $\mathbf{x}^*_{\mathbf{w}} = g(\mathbf{u}^*_{\mathbf{w}})$ where

$$\mathbf{x}^*_{\mathbf{w}} \in \arg\min_{\mathbf{x}\in\mathcal{X}} \left\{ \mathrm{LS}_{\mathbf{f}}(\mathbf{x}; \mathbf{w}) := \textstyle\sum_{m=1}^{M} w_m f_m(\mathbf{x}) \right\}. \tag{3}$$

Thus, the original problem (1) and the hidden formulation $\min_{\mathbf{x}\in\mathcal{X}} \mathbf{f}(\mathbf{x})$ are equivalent in the sense of PS and PF. Formally, $\mathcal{X}_{\mathrm{PF}} = g(\mathcal{U}_{\mathrm{PF}})$ and $\mathcal{F}_{\mathrm{PF}} = \{\mathbf{f}(\mathbf{x}) : \mathbf{x} \in \mathcal{X}_{\mathrm{PF}}\}$. Furthermore, the hidden (strong) convexity in Assumption 1 ensures unique scalarized solutions, full exploration of the Pareto front, and a well-defined Pareto-front path. This is summarized in Proposition 1.

Proposition 1 ([29, Theorem 3.5, Proposition 3.9], [63, Theorem 3.1.3])**.** *Under Assumption 1, given $\mathbf{w} \in \Delta_M$, solution* (3) *is unique. Every such minimizer $\mathbf{x}^*_{\mathbf{w}}$ given $\mathbf{w} \in \Delta_M$ is Pareto optimal for $\mathbf{f}$, and every Pareto-optimal point can be obtained by* (3) *with some $\mathbf{w} \in \Delta_M$.*

Proposition 1 implies that, under Assumption 1, *steering the scalarization weight $\mathbf{w}$ steers the solution along the PF*. Through the bijection $\mathbf{x} = g(\mathbf{u})$, the unique global minimizer $\mathbf{x}^*_{\mathbf{w}}$ in the hidden convex formulation corresponds to the unique global minimizer $\mathbf{u}^*_{\mathbf{w}} = g^{-1}(\mathbf{x}^*_{\mathbf{w}})$ in the original parameterization, with the same objective vector. Since every Pareto-optimal point is obtained for some $\mathbf{w} \in \Delta_M$, sweeping the weights reaches the entire PF.

In the bi-objective case, writing $\mathbf{w} = [w, 1-w]^\top$ recovers (2). The induced PF map is

$$\mathbf{f}_{\mathrm{PF}}(\mathbf{w}) := \mathbf{h}(\mathbf{u}^*_{\mathbf{w}}) = \mathbf{f}(\mathbf{x}^*_{\mathbf{w}}) \in \mathbb{R}^M, \quad \mathbf{w} \in \Delta_M. \tag{4}$$

Below we provide a concrete example in the bi-objective case.

2.1 Instantiation: Bi-Objective Reinforcement Learning and Bandits

In the bi-objective case, we identify $\mathbf{w} = [w, 1-w]^\top \in \Delta_2$ with the scalar weight $w \in [0, 1]$, and write $\mathbf{f}_{\mathrm{PF}}(w) := \mathbf{f}_{\mathrm{PF}}([w, 1-w]^\top)$, which is a one-dimensional path. To ensure a differentiable and smooth PF path, we impose a standard condition in PF-tracing analysis [83, 6].

Assumption 2 (Constraint Regularity)**.** *Assume $\mathcal{X} = \{\mathbf{x} : \mathbf{c}_1(\mathbf{x}) \le 0, \mathbf{c}_2(\mathbf{x}) = 0\}$, where $\mathbf{c}_1(\mathbf{x})$ is twice differentiable, Lipschitz-smooth and convex and $\mathbf{c}_2(\mathbf{x})$ is linear. For $\min_{\mathbf{x}\in\mathcal{X}} \mathrm{LS}_f(\mathbf{x}; w)$ with any $w \in [0, 1]$, assume that the active constraints are fixed, satisfy the Linear Independence Constraint Qualification (LICQ), and the Lagrangian Hessian is Lipschitz-continuous along $\{\mathbf{x}^*_w : w \in [0, 1]\}$.*

Under Assumptions 1 and 2, the path $w \mapsto \mathbf{f}_{\mathrm{PF}}(w)$ is differentiable, with its first coordinate nonincreasing and its second nondecreasing (cf. Appendix C.1). Its geometry is described by

$$\underbrace{v(w) := \left\| \frac{\partial}{\partial w} \mathbf{f}_{\mathrm{PF}}(w) \right\|}_{\text{traversal speed}}, \quad \underbrace{s(w) := \int_0^w v(p)\, \mathrm{d}p}_{\text{cumulative arc length}}, \quad \underbrace{\Phi(w) := \frac{s(w)}{s(1)}}_{\text{normalized arc-length}}, \quad w \in [0, 1]. \tag{5}$$

The speed measures the local motion of the PF point, and $s(w)$ accumulates the distance traveled from weight 0 to w. When $s(1) > 0$, $\Phi(w)$ is the fraction of the total PF arc length traversed by weight w. Since Φ is nondecreasing, continuous, and satisfies $\Phi(0) = 0$ and $\Phi(1) = 1$, it defines a valid CDF on $[0, 1]$. Its corresponding weight-sampling density is $\Phi'(w) = v(w)/s(1)$, which is proportional to the PF traversal speed. In contrast, uniform weight sampling induces a PF arc-length density inversely proportional to $v(w)$, generally resulting in nonuniform PF coverage. Moreover, to ensure non-vanishing PF traversal speed, we impose the following nondegeneracy assumption.

Assumption 3 (Nonzero trade-off)**.** *For each $w \in [0, 1]$, let J_w denote the Jacobian matrix of the active constraints of the scalarized problem $\min_{\mathbf{x}\in\mathcal{X}} \mathrm{LS}_f(\mathbf{x}; w)$ at $\mathbf{x}^*_w$. Assume there exists some constant $c > 0$ such that $\left\| \mathrm{Proj}_{\ker(J_w)} \left(\nabla f_1(\mathbf{x}^*_w) - \nabla f_2(\mathbf{x}^*_w)\right)\right\| \geq c$ for all $w \in [0, 1]$.*

Assumption 3 rules out zero trade-offs along feasible directions, which would prevent changes in the scalarization weight from moving the solution along the PF. Such nondegeneracy-type conditions are standard in trade-off analysis and PF continuation analyses [30, 10]. This condition ensures that Φ is invertible. More discussion is deferred to Appendix C.2.

Lemma 1 (Bounded PF traversal speed)**.** *Suppose Assumptions 1, 2, 3 hold. Then, there exist constants $0 < v_{\min} \leq v_{\max} < \infty$ such that $v(w) \in [v_{\min}, v_{\max}]$ for all $w \in [0, 1]$. Consequently, $s(1) = \int_0^1 v(w)dw \in [v_{\min}, v_{\max}]$, and $s(w)$ and $\Phi(w)$ are strictly increasing on $[0, 1]$.*

The proof of Lemma 1 is given in Appendix C.2. Therefore, given a budget of $(N + 1)$ PF points, we can choose weights $w_n = \Phi^{-1}\left(\frac{n}{N}\right)$ for $n = 0, \ldots, N$, so that the induced points $\{\mathbf{f}(\mathbf{x}^*_{w_n})\}_{n=0}^N$ are uniformly spaced along the PF, i.e., $s(w_{n+1}) - s(w_n) = \frac{s(1)}{N}$. This enables Rule 1 when the PF geometry is explicit, covering analytic examples such as quadratic problems in Appendix C.4, as well as practical multi-objective reinforcement learning (MORL) problems [79, 113].

In bi-objective RL, decision problems are commonly modeled as Markov Decision Processes (MDPs) $\mathcal{M} = (\mathcal{S}, \mathcal{A}, \mathcal{P}, \mathbf{r}, \gamma, \rho, \tau)$, where $\mathcal{S}$ and $\mathcal{A}$ are finite state and action spaces, $\mathcal{P}(s'|s, a)$ is the transition kernel, $\gamma \in (0, 1)$ is the discount factor, $\rho \in \Delta_{|\mathcal{S}|}$ is the initial-state distribution, $\mathbf{r}(s, a) = [r_1(s, a), r_2(s, a)]^\top$ is the bi-objective reward, and τ is a regularization functional. We instantiate a KL-regularized MDP with $\tau(\pi(\cdot|s)) = \beta \mathrm{KL}(\pi(\cdot|s)\|\pi_{\mathrm{ref}}(\cdot|s))$ applied to the policy π with $\beta > 0$ and some reference policy $\pi_{\mathrm{ref}}(\cdot|s)$. This yields a generally non-convex instance of (1), with $\mathbf{u} = \pi \in \mathcal{U} = \Delta_{|\mathcal{A}|}^{|\mathcal{S}|}$ and $h_m(\mathbf{u}) = -\mathbb{E}_\pi\left[\sum_{t=0}^\infty \gamma^t \big(r_m(s_t, a_t) - \tau(\pi(\cdot|s_t))\big)\right]$, covering LLM alignment settings with competing rewards [23, 86] and RLHF/DPO-style objectives [75, 23, 98].

This KL-regularized MORL problem is generally nonconvex in policy space but strongly convex in normalized occupancy measures. Let

$$\mathbf{x} = \mathrm{vec}(\mu_\pi) \in \Delta_{|\mathcal{S}||\mathcal{A}|}, \quad \text{where} \quad \mu_\pi(s, a) := (1 - \gamma)\, \mathbb{E}_{\rho,\pi}\left[\sum_{t=0}^\infty \gamma^t \mathbf{1}\{s_t = s, a_t = a\}\right]. \tag{6}$$

Denote by E the per-state summation matrix with $E_{(s,a),s'} = \mathbf{1}\{s' = s\}$, and by P the transition matrix with $P_{(s,a),s'} = \mathcal{P}(s'|s, a)$. Then the occupancy space is a convex feasible set [107, 32]:

$$\mathcal{X} = \{\mathrm{vec}(\mu_\pi) : \pi \in \Delta_{|\mathcal{A}|}^{|\mathcal{S}|}\} = \left\{\mathbf{x} \in \mathbb{R}_+^{|\mathcal{S}||\mathcal{A}|} : (E - \gamma P)^\top \mathbf{x} = (1 - \gamma)\rho\right\}. \tag{7}$$

Although illustrated on finite tabular MDPs, this viewpoint extends to continuous state or action spaces, where the finite vector $\mathbf{x}$ is replaced by an infinite-dimensional density [7, 2]. Moreover, when $E^\top \mathbf{x} \geq d_{\min} > 0$, a standard sufficient exploration condition, the policy-to-occupancy map (6) is bijective [81, 107]. Hence, we can write $h_m(\mathbf{u}) = f_m(g(\mathbf{u}))$ for the bijection g. Writing $R_m = \mathrm{vec}(r_m(\mathcal{S}, \mathcal{A}))$ and $R_0 = \mathrm{vec}(\log \pi_{\mathrm{ref}}(\mathcal{A}|\mathcal{S}))$ gives the strongly convex formulation [81]

$$f_m(\mathbf{x}) = \beta\left(\langle \mathbf{x}, \log \mathbf{x}\rangle - \langle E^\top \mathbf{x}, \log(E^\top \mathbf{x})\rangle - \langle R_0, \mathbf{x}\rangle\right) - \langle R_m, \mathbf{x}\rangle, \quad m \in \{1, 2\}. \tag{8}$$

For any $w \in [0, 1]$, the LS problem associated with (8) admits a unique interior solution $\mathbf{x}^*_w \in \Delta^\circ_{|\mathcal{S}|\times|\mathcal{A}|}$ [66] that is Pareto optimal, and $w \mapsto \mathbf{x}^*_w$ is Lipschitz-continuous [47, Theorem 2.16]. Hence $\{\mathbf{x}^*_w : w \in [0, 1]\} \subset \mathrm{relint}(\mathcal{X})$ is compact, so the objective in (8) has uniformly bounded and Lipschitz-continuous Hessian on this set. Therefore, the MDP instance satisfies Assumption 1. Moreover, since the LS solution is interior, only the equality constraint in (7) is active. Its Jacobian is $J = (E - \gamma P)^\top$, which has full row rank as detailed in Appendix D.1, and hence Assumption 2 holds. Additionally, Assumption 3 reduces to $\| \mathrm{Proj}_{\ker((E-\gamma P)^\top)}(R_1 - R_2)\| \geq c$, a mild condition that typically holds unless the two objectives coincide along feasible directions. The norm of the objective-vector derivative along the optimal occupancy path gives $v(w)$, whose integral and normalization yield Φ for Rule 1.

Proposition 2 (PF speed in bi-objective RL)**.** *Under the sufficient exploration condition [81], the bi-objective RL in* (8) *with constraint in* (7) *yields*

$$v(w) = \sqrt{(1-w)^2+w^2}\Big|(R_1-R_2)^\top\big(\widetilde{H}_w^{-1} - \widetilde{H}_w^{-1}J^\top(J\widetilde{H}_w^{-1}J^\top)^{-1}J\widetilde{H}_w^{-1}\big)(R_1-R_2)\Big|,$$

where $\widetilde{H}_w = \beta\big(\operatorname{Diag}(\frac{1}{\mathbf{x}_w^*}) - E\operatorname{Diag}(\frac{1}{E^\top\mathbf{x}_w^*})E^\top\big) + cJ^\top J$ *is the constraint-augmented Hessian for all* $c > 0$*. When* $|\mathcal{S}| = 1$*, this reduces to the entropic bandit case where the LS solution admits the closed form* $\mathbf{u}_w^* = \mathbf{x}_w^* = \operatorname{SoftMax}\big(R_0 + \beta^{-1}(wR_1 + (1-w)R_2)\big) \in \Delta_A^\circ$*, and the speed becomes*

$$v(w) = \beta^{-1}\sqrt{(1-w)^2+w^2}(R_1-R_2)^\top\big(\operatorname{Diag}(\mathbf{u}_w^*) - \mathbf{u}_w^*(\mathbf{u}_w^*)^\top\big)(R_1-R_2). \tag{9}$$

The proof of Proposition 2 is deferred to Appendix D.3. Proposition 2 shows that Φ can be established once the solutions $\{\mathbf{x}_w^*\}_{w\in[0,1]}$ are known, while constructing Φ is meant to identify a finite set of weights $\{w_n\}_{n=0}^N$ to save the budget for finding the corresponding solutions. This creates a *coupled dependence* between geometry estimation on the PF and scalarized optimization over $\mathbf{x}_w^*$. Fortunately, this issue disappears when $|\mathcal{S}| = 1$, where $\mathbf{x}_w^*$ and $v(w)$ are available in closed form and $\Phi(w)$ is analytically available. Examples include the best-arm identification (BAI) problem in the bandit setting [12, 49], where each arm $a \in \mathcal{A}$ has an unknown mean reward vector $[R_1(a), R_2(a)]^\top$. While BAI for scalarized rewards is well studied [4, 101, 22], PF coverage remains less explored [38]. Proposition 2 then yields the following instantiation of the plug-in PF-aware weighting Rule 1.

Rule 1 example. *For the bandit problem, estimate reward vectors* $\hat{R}_1, \hat{R}_2$ *to construct* $\hat{v}(\cdot; \hat{R}_1, \hat{R}_2)$ *by replacing* (R_1, R_2) *with* $(\hat{R}_1, \hat{R}_2)$ *in* (9) *and similarly*

$$\hat{s}(w; \hat{R}_1, \hat{R}_2) = \int_0^w \hat{v}(u; \hat{R}_1, \hat{R}_2)du, \quad \hat{\Phi}(w; \hat{R}_1, \hat{R}_2) = \frac{\hat{s}(w; \hat{R}_1, \hat{R}_2)}{\hat{s}(1; \hat{R}_1, \hat{R}_2)}. \tag{10}$$

Then, apply Rule 1 with Φ *replaced by* $\hat{\Phi}$ *when choosing* $\{w_n\}_{n=0}^N$*.*

This analytical construction makes the PF geometry explicit and provides a tractable testbed for the geometry behind SURF, as visualized in Figure 1. Since the reward vector (R_1, R_2) is unknown in practice, we estimate it from samples and plug the estimate into the closed-form v. The resulting $\hat{\Phi}$ is a consistent estimator of Φ. The estimation error of $\hat{\Phi}$ is driven solely by reward-estimation error and is therefore controlled by standard statistical bounds [4, 27, 99]. Specifically, $\|\hat{\Phi} - \Phi\|_\infty$ decays at rate $\mathcal{O}(T^{-1/2})$, where T is the number of pulls used to estimate the reward vectors. This result is formalized in Appendix D.4 and validated empirically in Figure 5. This example also suggests a natural extension to scalarization-based MO-bandit frameworks [26, 101, 22]: rather than drawing weights w uniformly, one can sample w according to Rule 1 with the estimated $\hat{\Phi}$ in place of Φ.

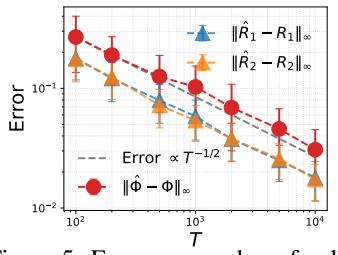

Figure 5: Error vs. number of pulls T in bandit. Error bars show standard deviations across 100 trials.

3 SURF: Algorithm for Sampling Uniformly along the Pareto Front

We use LS throughout this section, with SURF refining coverage along the PF path induced by LS. The previous section shows that Rule 1 achieves uniform PF coverage when Φ or its estimate (cf. Rule 1 example) is available. In general, however, Φ is not explicitly known. To apply Rule 2 in this setting, we propose SURF, an iterative refinement algorithm for **S**ampling **U**niformly along the Pa**r**eto **F**ront (SURF). At a high level, SURF refines the estimate of Φ while improving the scalarized solutions associated with the target weights $w_n = \Phi^{-1}(n/N)$. Specifically, at outer iteration t, SURF uses the current CDF estimate Φ_t to allocate weights $w_n^{(t)} = \Phi_t^{-1}\left(\frac{n}{N}\right)$ with initialization $\Phi_0(w) = w$ to start with the standard uniform weighting baseline. If $\Phi_t = \Phi$, then $w_n^{(t)}$ would produce uniform PF coverage. For each sampled weight $w_n^{(t)}$, SURF then approximately solves:

$$\mathbf{u}_n^{(t)} \approx \arg\min_{\mathbf{u}\in\mathcal{U}} \mathrm{LS}_h(\mathbf{u}; w_n^{(t)}) \quad \text{or equivalently } \mathbf{x}_n^{(t)} \approx \arg\min_{\mathbf{x}\in\mathcal{X}} \mathrm{LS}_f(\mathbf{x}; w_n^{(t)}). \tag{11}$$

Algorithm 1 SURF: **S**ampling **U**niformly along the Par**e**to **F**ront

Require: Number of segments N, initial CDF $\Phi_0(w) = w$, target quantiles $\{n/N\}_{n=0}^N$, damping factor α, a finite K-step inner solver InnerSolver for $\min_{\mathbf{u}\in\mathcal{U}} \mathrm{LS}_h(\mathbf{u}; w)$ with initial iterates $\{\mathbf{u}_n^{(-1)}\}_{n=0}^N$ or for $\min_{\mathbf{x}\in\mathcal{X}} \mathrm{LS}_f(\mathbf{x}; w)$ with initial iterates $\{\mathbf{x}_n^{(-1)}\}_{n=0}^N$.

1: **for** $t = 0, 1, 2, \cdots, T$ **do**
2: Find $w_n^{(t)} = \Phi_t^{-1}(\frac{n}{N})$, for all $n = 0, 1, \cdots, N$
3: **for** $n = 0, 1, \cdots, N$ **do**
4: Update $\mathbf{u}_n^{(t)} = \textsc{InnerSolver}\left(w_n^{(t)}; \mathbf{u}_{\text{init}} = \mathbf{u}_n^{(t-1)}\right)$ or $\mathbf{x}_n^{(t)}$ accordingly.
5: Save $\mathbf{h}(\mathbf{u}_n^{(t)})$ or $\mathbf{f}(\mathbf{x}_n^{(t)})$ equivalently.
6: **end for**
7: Build $\tilde{s}^{(t)}(w_n^{(t)})$ via (12) and $\tilde{s}^{(t)} : [0, 1] \to \mathbb{R}_+$ via interpolation (e.g., Algorithm 2).
8: Construct $\tilde{\Phi}_t(w) = \frac{\tilde{s}^{(t)}(w)}{\tilde{s}^{(t)}(1)}$ and update $\Phi_{t+1}(w) = \alpha\tilde{\Phi}_t(w) + (1-\alpha)\Phi_t(w), \forall w \in [0, 1]$.
9: **end for**

Ensure: Refined CDF Φ_{T+1} and PF samples $\{(\mathbf{u}_n^{(T)}, \mathbf{h}(\mathbf{u}_n^{(T)}))\}_{n=0}^N$.

SURF is compatible with off-the-shelf solvers, e.g., Adam for loss minimization [50] and PPO for RL [82]. The InnerSolver runs for K steps, warm-started from the previous iterate $\mathbf{u}_n^{(t-1)}$. The resulting PF estimates $\{\mathbf{h}(\mathbf{u}_n^{(t)})\}_{n=0}^N$ or equivalently $\{\mathbf{f}(\mathbf{x}_n^{(t)})\}_{n=0}^N$ are then used to build estimates for s at the sampled weights $\{w_n^{(t)}\}_{n=0}^N$ via cumulative chord length

$$\tilde{s}^{(t)}(w_0^{(t)}) := 0,\ \tilde{s}^{(t)}(w_{n+1}^{(t)}) := \tilde{s}^{(t)}(w_n^{(t)}) + \left\|\mathbf{h}(\mathbf{u}_{n+1}^{(t)}) - \mathbf{h}(\mathbf{u}_n^{(t)})\right\|, \quad n = 0, \cdots, N-1. \tag{12}$$

Then, SURF interpolates these values to construct $\tilde{s}^{(t)}$ on $[0, 1]$, for example, Piecewise Cubic Hermite Interpolating Polynomial (PCHIP) [37]. Compared with naive piecewise-linear interpolation, PCHIP remains simple to implement, is available in standard numerical libraries, preserves monotonicity, and provides more favorable Lipschitz-smoothness properties, as discussed in the next section. These features are useful both in practice and in the convergence analysis. In this way, an empirical estimate of Φ is established via $\tilde{\Phi}_t(w) := \frac{\tilde{s}^{(t)}(w)}{\tilde{s}^{(t)}(1)}$ and SURF then updates the CDF estimate via

$$\Phi_{t+1}(w) = \alpha\tilde{\Phi}_t(w) + (1-\alpha)\Phi_t(w), \tag{13}$$

where a damping factor α is especially useful for inexact InnerSolver to stabilize the update.

For a priori preferences among objectives, the uniform positions $\{\frac{n}{N}\}_{n=0}^N$ can be replaced by desired trade-off positions $\{q_n\}_{n=0}^N \in [0, 1]^{N+1}$, yielding $w_n^{(t)} = \Phi_t^{-1}(q_n)$ for each $n = 0, \cdots, N$.

Convergence Analysis. We analyze Algorithm 1 by separating two sources of error: the *discretization error* from reconstructing s using only $(N + 1)$ PF samples, and the *optimization error* from solving the scalarized subproblems (11) only approximately. We start with the exact-InnerSolver case that isolates the discretization effect: even when the scalarized solutions are exact, $\mathbf{u}_n^{(t)} = \mathbf{u}^*_{w_n^{(t)}}$ or equivalently $\mathbf{x}_n^{(t)} = \mathbf{x}^*_{w_n^{(t)}}$, estimating the PF geometry from only $(N+1)$ sampled points introduces an inevitable finite-sampling floor. We then consider the optimization error induced by running InnerSolver for only K steps. To quantify this perturbation, we characterize the convergence rate of InnerSolver via η_K, a K-step contraction factor such that, for any weight $w \in [0, 1]$ and initialization $\tilde{\mathbf{x}}_0$, its output $\tilde{\mathbf{x}}_K$ satisfies $\|\tilde{\mathbf{x}}_K - \mathbf{x}_w^*\| \le \eta_K\|\tilde{\mathbf{x}}_0 - \mathbf{x}_w^*\|$, where $\eta_K \downarrow 0$ as $K \to \infty$. For linearly convergent InnerSolver, e.g., K-step projected gradient descent [32], one may take $\eta_K = \left(1 - \mu/l_{f,1}\right)^K$, where $l_{f,1}$ is valid along the entire projected gradient descent trajectory; see Appendix D.2. For sublinearly convergent methods, hidden strong convexity converts objective-gap or stationarity guarantees into hidden-space distance bounds [32], so for sufficiently large K the same characterization can be instantiated with $\eta_K = \mathcal{O}(K^{-p})$ for some $p > 0$. The following theorem shows the convergence of SURF to the finite-sampling floor, with an additional coupled term accounting for inner optimization error. Let $C(N)$ be a scalar depending on N. We write

$\|\Phi_t - \Phi\|_\infty = \sup_{w\in[0,1]} |\Phi_t(w) - \Phi(w)|$ and define the convergence measure as

$$\mathcal{E}_t\big(\Phi_t, \{w_n^{(t)}\}, \{\mathbf{x}_n^{(t)}\}\big) := \underbrace{\|\Phi_t - \Phi\|_\infty}_{\text{CDF estimation error}} + C(N) \underbrace{\max_{0\le n\le N} \|\mathbf{x}_n^{(t)} - \mathbf{x}^*_{w_n^{(t)}}\|}_{\text{inner optimization error}}. \tag{14}$$

Theorem 1 (Convergence of SURF)**.** *Suppose Assumptions 1, 2, 3 hold. Then, Algorithm 1 generates $\{\Phi_t\}$ and $\{\tilde{\Phi}_t\}$ that are uniformly $l_{\Phi_{\text{iter}},1}$-Lipschitz-smooth for all $t \ge 0$. Let $\kappa := \frac{4\sqrt{2}v_{\max}}{v_{\min}}\sqrt{l_{\Phi_{\text{iter}},1} + l_{\Phi,1}}$, and $N = \Omega(\kappa^{1.5})$. For general* InnerSolver *with budget specified in Remark 1, SURF with $\alpha \le 0.5$ admits $C(N) = \Theta(N)$ such that, for all $t \ge 0$,*

$$\mathcal{E}_t\big(\Phi_t, \{w_n^{(t)}\}, \{\mathbf{x}_n^{(t)}\}\big) \le \Big(1 - \frac{\alpha}{4}\Big)^t \mathcal{E}_0\big(\Phi_0, \{w_n^{(0)}\}, \{\mathbf{x}_n^{(0)}\}\big) + \underbrace{\mathcal{O}(N^{-2})}_{\textit{discretization floor}}. \tag{15}$$

Remark 1 (Inner-solver budget)**.** *If* InnerSolver *returns exact solutions, then the inner optimization error times $C(N)$ is zero. Otherwise, consider an* InnerSolver *with K-step contraction factor η_K. We choose K so that $\eta_K = \mathcal{O}\big(s(1)/N\big)$, that is $K = \Omega\big(\log(N v_{\max} v_{\min}^{-2})\big)$ for a linearly convergent solver and $K = \Omega\big(\big(N/s(1)\big)^{1/p}\big)$ for a sublinearly convergent solver with $\eta_K = \mathcal{O}(K^{-p})$ for some $p > 0$. Then the inner-solver perturbs the CDF reconstruction by a sum of N chord-length errors, so $C(N) = \Theta(N)$ is the natural amplification factor. Alternatively, one may use a small outer stepsize $\alpha = \mathcal{O}(N^{-1})$, in which case $\eta_K = \mathcal{O}(1)$ is enough after a warmup that ensures $\max_n \|\mathbf{x}_n^{(0)} - \mathbf{x}^*_{w_n^{(0)}}\| = \mathcal{O}\left(v_{\min}^2 v_{\max}^{-1} N^{-1}\right)$, as detailed in Theorem 4 in Appendix E.2.*

The proof bounds CDF reconstruction through chord and interpolation errors, then couples CDF-driven grid motion with warm-started tracking of the scalarized optima to contract the combined error to a finite-sampling floor. The geometric factor κ combines traversal-speed variation and CDF smoothness. The proof of Theorem 1 and the constants in $\mathcal{O}(\cdot)$ and $\Omega(\cdot)$ are made explicit in Appendix E. Theorem 1 guarantees that SURF drives the CDF estimation error to an $\mathcal{O}(N^{-2})$ finite-sampling discretization floor, matching the standard second-order accuracy of polyline-based reconstruction of smooth curves [88]. In contrast, even with an exact InnerSolver, the first-step reconstruction $\tilde{\Phi}_0$ can carry an $\mathcal{O}(\kappa^3 N^{-2})$ error, where the κ factor comes from uneven clustering on the PF, as theoretically discussed in Lemma 10 in Appendix E.1 and empirically verified in Appendix F.1. In Appendix E.3, we further relate CDF reconstruction error to a PF uniformity metric measuring the relative deviation of sampled PF segments from equal spacing, *Coefficient of Variation* (CV) [31], as

$$\text{CV} := \frac{\text{std}\Big(\{\|\mathbf{h}(\mathbf{u}_{n+1}^{(t)}) - \mathbf{h}(\mathbf{u}_n^{(t)})\|\}_{n=0}^{N-1}\Big)}{\text{mean}\Big(\{\|\mathbf{h}(\mathbf{u}_{n+1}^{(t)}) - \mathbf{h}(\mathbf{u}_n^{(t)})\|\}_{n=0}^{N-1}\Big)} = \mathcal{O}\big(N^{-1} + N\|\Phi_t - \Phi\|_\infty\big) = \mathcal{O}(N^{-1}), \tag{16}$$

at sufficiently large outer iteration $t = \Omega(\log N)$. Equation (16) links CDF refinement accuracy to PF spacing uniformity, showing that SURF improves uniform PF coverage.

Remark 1 illustrates that SURF does not require each scalarized subproblem to be solved to high precision at every outer iteration. To achieve $\|\Phi_T - \Phi\|_\infty = \mathcal{O}(N^{-2})$ with a linearly convergent inner solver [32], SURF requires $T = \mathcal{O}(\log N)$ outer iterations and per-iteration inner complexity $\mathcal{O}(N \log N)$, giving total complexity $\mathcal{O}(N(\log N)^2)$. At the same time, SURF yields $\max_n \|\mathbf{x}_n^{(T)} - \mathbf{x}^*_{w_n^{(T)}}\| = \mathcal{O}(N^{-3})$. In contrast, uniform weighting requires $\mathcal{O}(N \log(N^3)) = \mathcal{O}(N \log N)$ total complexity for N to reach the same inner accuracy, while providing no uniform PF-coverage guarantee. Alternatively, the small-stepsize regime $\alpha = \mathcal{O}(N^{-1})$ in Remark 1 relaxes the required inner-solver contraction, making it especially suitable for sublinearly convergent InnerSolver. In practice, with a moderate choice of $N = 5$ and $\alpha = 0.3$ in the LLM alignment task in Section 4, SURF achieves efficiency comparable to the uniform scalarization baseline under the same training budget, while significantly improving PF uniformity.

4 Experiments

In this section, we empirically verify the effectiveness of SURF against **baselines** including *uniform weighting* with $w_n = \frac{n}{N}$, *Optimistic Linear Support (OLS)* [5] which aims to identify supported PF so-

Tasks	Metric	Soup	Uniform-w	UMOD	OLS	Ours
DST	HV ($\times 10^{-3}$) ↑	3.10 ± 0	1.83 ± 0	3.85 ± 0.00	3.69 ± 0	$\mathbf{3.87 \pm 0}$
	IGD ($\times 10^{-2}$) ↓	0.68 ± 0	4.51 ± 0	0.23 ± 0.00	0.41 ± 0	$\mathbf{0.21 \pm 0}$
	CV ($\times 10^{-1}$) ↓	*0.09 ± 0*	10.91 ± 0	3.17 ± 0.05	9.87 ± 0	$\mathbf{0.81 \pm 0}$
	Gap Ratio ↓	*1.04 ± 0*	17.16 ± 0	2.41 ± 0.13	26.85 ± 0	$\mathbf{1.45 \pm 0}$
Fishwood	HV ($\times 10^{-1}$) ↑	2.39 ± 0	$\mathbf{2.40 \pm 0}$	2.39 ± 0.00	2.39 ± 0	2.39 ± 0
	IGD ($\times 10^{-2}$) ↓	0.91 ± 0	0.91 ± 0	0.92 ± 0.00	0.96 ± 0	$\mathbf{0.90 \pm 0}$
	CV ($\times 10^{-1}$) ↓	2.00 ± 0	1.94 ± 0	2.33 ± 0.17	3.57 ± 0	$\mathbf{0.26 \pm 0}$
	Gap Ratio ↓	1.90 ± 0	1.75 ± 0	2.14 ± 0.14	2.49 ± 0	$\mathbf{1.12 \pm 0}$
MO-Mountaincar	HV ($\times 10^{3}$) ↑	0.33 ± 0.20	0.89 ± 0.27	0.99 ± 0.17	0.35 ± 0.30	$\mathbf{1.11 \pm 0.23}$
	IGD ↓	9.45 ± 1.02	6.41 ± 0.88	17.28 ± 10.56	23.77 ± 6.63	$\mathbf{6.00 \pm 0.72}$
	CV ($\times 10^{-1}$) ↓	12.32 ± 5.72	7.25 ± 1.33	10.06 ± 1.57	6.54 ± 3.61	$\mathbf{4.41 \pm 2.03}$
	Gap Ratio ↓	489.94 ± 401.29	74.96 ± 97.16	10.28 ± 3.89	9.47 ± 10.66	$\mathbf{3.75 \pm 3.75}$

Table 1: MO-Gymnasium [33] results comparing PF approximation quality and uniform-coverage metrics. We use $N = 15$ for DST, $N = 10$ for Fishwood, and $N = 5$ for MO-Mountaincar. Results are reported as mean±standard deviation over 8 realizations. Zero standard deviation represents deterministic experiments without randomness.

lutions, *Uniform Multi-Objective optimization based on Decomposition (UMOD)* [110], which targets uniform PF coverage, and inference-time weighted model merging named *reward soup (Soup)* [76]. For budget fairness, all optimization-based methods are given the same total number of inner iterations. We use **evaluation metrics** including *Hypervolume* (HV) [115] and *Inverted Generational Distance* (IGD) [9] for PF approximation quality, and *Gap Ratio* $= \frac{\max_n \|\mathbf{h}(\mathbf{u}_{n+1}^{(t)}) - \mathbf{h}(\mathbf{u}_n^{(t)})\|}{\min_n \|\mathbf{h}(\mathbf{u}_{n+1}^{(t)}) - \mathbf{h}(\mathbf{u}_n^{(t)})\|}$ [79] and CV [31] for spacing uniformity. Spacing is assessed alongside approximation quality, since evenly spaced points can lie off the PF, as Soup illustrates. Reference points, reference-front construction, and objective scales are specified in Appendix F. We conduct experiments on MO Gymnasium benchmarks [33] in Section 4.1 and an LLM alignment task for Reddit summarization [91] in Section 4.2. Additional details for the experiments and toy examples are provided in Appendix F. Beyond LS, Appendix F.2 shows that SURF improves spacing on the nonconvex ZDT2 and quarter-circle fronts using weighted Tchebycheff scalarization.

4.1 MO-Gymnasium Problems

Table 1 summarizes the results on three MO-Gymnasium [33] tasks. Appendix F.3 provides details on the environment, condition checks of occupancy-measure regularity, and implementation details.

DST. DST [89, 101] is a classical bi-objective RL benchmark, where the agent trades off treasure value and travel time. We evaluate SURF Algorithm 1 with 30 outer iterations and $\alpha = 0.3$. Figure 6 shows that SURF produces more uniform PF coverage than the optimization-based baselines. Compared with uniform weighting, SURF improves HV by $2.11\times$ and reduces IGD, CV, and Gap Ratio by $21.5\times$, $13.5\times$, and $11.8\times$, respectively. SURF also outperforms UMOD, another baseline explicitly designed for uniform PF coverage, with slightly higher HV, lower IGD, and substantially better spacing, reducing the Gap Ratio from 2.41 to 1.45. Although the Soup method yields uniform spacing, the solutions are suboptimal and not on the PF, as indicated by the visualization in Figure 6. In Appendix F.3.1, we additionally visualize the DST policies learned by different methods through their induced trajectories, and provide ablations on N, α, and K, together with convergence-time analysis showing that UMOD is slower than SURF.

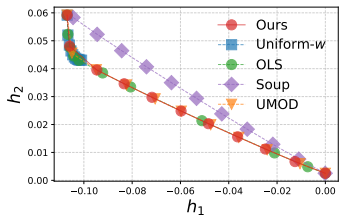

Figure 6: PFs obtained by SURF and baselines on DST.

Fishwood. The Fishwood task trades off fishing and wood-collection rewards [78]. We apply SURF with $\alpha = 0.3$ for 15 outer iterations. Additional setup details are provided in Appendix F.3.2. As shown in Table 1, all methods achieve nearly identical HV and IGD, indicating similar PF approximation quality. However, SURF substantially improves spacing uniformity, reducing CV from the best baseline value 0.194 to 0.026 and Gap Ratio from 1.75 to 1.12, corresponding to $7.5\times$ and $1.6\times$ reductions, respectively.

MO-Mountaincar. The MO-Mountaincar problem is a multi-objective variant of the classical Mountaincar control task, where an underpowered car must build momentum by moving back and forth in a valley to reach the goal near the top of the right hill [65]. In this setting, the agent faces a trade-off between fast completion and action efficiency, induced by a per-step time penalty and an action penalty. Additional details are provided in Appendix F.3.3. As shown in Table 1, SURF achieves the best overall performance, improving HV from the best baseline value 0.99 to 1.11 and reducing IGD from 6.41 to 6.00. It also improves spacing uniformity, reducing CV from the best baseline value 0.654 to 0.441 and Gap Ratio from 9.47 to 3.75. These results indicate that SURF improves both PF approximation quality and uniform coverage in this task.

4.2 LLM Alignment on the Reddit Summarization Task

To test SURF beyond standard MORL benchmarks, we apply it to a stochastic LLM alignment problem on Reddit summarization [91, 87]. We fine-tune `Qwen/Qwen2.5-0.5B-Instruct` [74] with PPO [82], while keeping the pretrained backbone frozen and updating only parameter-efficient finetuning (PEFT) adapter parameters via LoRA [45]. The two objectives are induced by a summarization-quality reward model and a factual-faithfulness reward model, together with a fixed KL penalty to the reference policy with $\beta = 0.05$. Each scalarized PPO job uses one 40GB A100 GPU, 4 CPU cores, and 16GB host memory, with both SURF and Uniform-w receiving 12 PPO epochs per slot after a shared 10-epoch warm start. Full training details are deferred to Appendix F.4.

Table 2 shows that SURF improves both performance and PF coverage over Uniform-w under the same per-slot PPO budget. Compared with Uniform-w, SURF achieves $1.07\times$ higher HV, $2.2\times$ lower CV, and $2.0\times$ lower Gap Ratio. The relatively large standard deviation of Gap Ratio reflects the stochasticity of LLM evaluation: each estimate is computed on batches of only 64 prompts, so a small number of unusually short or long objective-space segments can substantially affect the max/min spacing ratio. Soup is less effective in our setting, suggesting that endpoint-only interpolation between PEFT adapters may not reliably yield well-spaced PF coverage for these objectives, especially with a smaller backbone and limited training budget. We omit UMOD and OLS because they require substantially heavier computation in this LLM setting, and IGD because constructing a sufficiently dense reference PF is beyond our available computational budget.

Metric	Soup	Uniform-w	Ours
HV ↑	0.75 ± 0.30 (0.95 ± 0.30)	3.76 ± 0.31 (4.00 ± 0.30)	**4.03 ± 0.35** (**4.24 ± 0.35**)
CV ↓	1.69 ± 0.26 (1.69 ± 0.26)	0.65 ± 0.12 (0.65 ± 0.12)	**0.30 ± 0.17** (**0.29 ± 0.17**)
Gap Ratio ↓	161.20 ± 15.27 (159.83 ± 15.16)	5.89 ± 1.15 (6.01 ± 1.21)	**2.93 ± 5.35** (**2.89 ± 5.19**)

Table 2: LLM alignment results on Reddit summarization. Main entries are evaluated using KL-regularized objectives, while values in parentheses are evaluated using reward-only objectives without the KL term. All metrics are evaluated over 10 batches with batch size 64, reported as mean $\pm$ standard deviation.

5 Conclusion

This paper introduced SURF, a scalarization wrapper for improving uniform coverage along the reachable Pareto front while retaining the chosen inner solver. One of our key contributions is a geometric view of scalarization, where scalarization not only produces individual Pareto points but also induces a traversal of the PF with generally non-uniform speed. By characterizing this weight-to-PF map through the arc-length CDF Φ, we obtain a principled correction from uniform weights to uniform PF coverage. For structured bandit settings, we derive explicit forms of this geometry, and for general problems, we propose an iterative CDF-refinement algorithm compatible with standard InnerSolver. We establish convergence to a discretization floor for bi-objective LS under the stated assumptions and inner-solver conditions. Empirically, SURF improves PF coverage in the tested optimization, reinforcement learning, and LLM alignment settings, suggesting that arc-length-aware scalarization is a practical tool for constructing more uniformly distributed Pareto front approximations. Future directions include extending geometry-aware sampling to higher-dimensional Pareto fronts and larger-scale LLM alignment tasks.

References

[1] Lucas Nunes Alegre, Ana Bazzan, and Bruno C Da Silva. Optimistic linear support and successor features as a basis for optimal policy transfer. In *International conference on machine learning*, pages 394–413. PMLR, 2022.

[2] Eitan Altman. *Constrained Markov decision processes*. Routledge, 2021.

[3] James Anderson, John C Doyle, Steven H Low, and Nikolai Matni. System level synthesis. *Annual Reviews in Control*, 47:364–393, 2019.

[4] Peter Auer, Nicolo Cesa-Bianchi, and Paul Fischer. Finite-time analysis of the multiarmed bandit problem. *Machine learning*, 47(2):235–256, 2002.

[5] Leon Barrett and Srini Narayanan. Learning all optimal policies with multiple criteria. In *Proceedings of the 25th international conference on Machine learning*, pages 41–47, 2008.

[6] Fernanda Beltrán, Oliver Cuate, and Oliver Schütze. The pareto tracer for general inequality constrained multi-objective optimization problems. *Mathematical and Computational Applications*, 25(4):80, 2020.

[7] Abhay G Bhatt and Vivek S Borkar. Occupation measures for controlled markov processes: Characterization and optimality. *The Annals of Probability*, pages 1531–1562, 1996.

[8] Patrick Billingsley. *Probability and measure*. John Wiley & Sons, 2017.

[9] José M Bogoya, Andrés Vargas, and Oliver Schütze. The averaged hausdorff distances in multi-objective optimization: A review. *Mathematics*, 7(10):894, 2019.

[10] Matthias Bolten, Onur Tanil Doganay, Hanno Gottschalk, and Kathrin Klamroth. Tracing locally pareto-optimal points by numerical integration. *SIAM Journal on Control and Optimization*, 59(5):3302–3328, 2021.

[11] Stephen Boyd, Seung-Jean Kim, Lieven Vandenberghe, and Arash Hassibi. A tutorial on geometric programming. *Optimization and engineering*, 8(1):67–127, 2007.

[12] Sébastien Bubeck, Rémi Munos, and Gilles Stoltz. Pure exploration in multi-armed bandits problems. In *International conference on Algorithmic learning theory*, pages 23–37. Springer, 2009.

[13] Chris J Budd and JF Williams. Moving mesh generation using the parabolic monge–ampère equation. *SIAM Journal on Scientific Computing*, 31(5):3438–3465, 2009.

[14] Vira Chankong and Yacov Y Haimes. *Multiobjective decision making: theory and methodology*. Courier Dover Publications, 2008.

[15] Lisha Chen, Heshan Fernando, Yiming Ying, and Tianyi Chen. Three-way trade-off in multi-objective learning: Optimization, generalization and conflict-avoidance. *J. Machine Learn. Res.*, 25(193):1–53, 5 2024.

[16] Lisha Chen, AFM Saif, Yanning Shen, and Tianyi Chen. Ferero: A flexible framework for preference-guided multi-objective learning. *Advances in Neural Information Processing Systems*, 37:18758–18805, 2024.

[17] Lisha Chen, Quan Xiao, Ellen Hidemi Fukuda, Xinyi Chen, Kun Yuan, and Tianyi Chen. Efficient first-order optimization on the Pareto set for multi-objective learning under preference guidance. In *International Conference on Machine Learning (ICML), as spotlight*, 2025.

[18] Weiyu Chen and James Kwok. Pareto merging: Multi-objective optimization for preference-aware model merging. *arXiv preprint arXiv:2408.12105*, 2024.

[19] Weiyu Chen, Xiaoyuan Zhang, Baijiong Lin, Xi Lin, Han Zhao, Qingfu Zhang, and James T Kwok. Gradient-based multi-objective deep learning: Algorithms, theories, applications, and beyond. *arXiv preprint arXiv:2501.10945*, 2025.

[20] Xin Chen, Niao He, Yifan Hu, and Zikun Ye. Efficient algorithms for a class of stochastic hidden convex optimization and its applications in network revenue management. *Operations Research*, 73(2):704–719, 2025.

[21] Yiwei Chen and Cong Shi. Network revenue management with online inverse batch gradient descent method. *Production and Operations Management*, 32(7):2123–2137, 2023.

[22] Ji Cheng, Bo Xue, Chengyu Lu, Ziqiang Cui, and Qingfu Zhang. Multi-objective neural bandits with random scalarization. In *Proceedings of the Thirty-Fourth International Joint Conference on Artificial Intelligence*, pages 4914–4922, 2025.

[23] Josef Dai, Xuehai Pan, Ruiyang Sun, Jiaming Ji, Xinbo Xu, Mickel Liu, Yizhou Wang, and Yaodong Yang. Safe rlhf: Safe reinforcement learning from human feedback. *arXiv preprint arXiv:2310.12773*, 2023.

[24] Indraneel Das and John E Dennis. Normal-boundary intersection: A new method for generating the pareto surface in nonlinear multicriteria optimization problems. *SIAM journal on optimization*, 8(3):631–657, 1998.

[25] Kalyanmoy Deb, Lothar Thiele, Marco Laumanns, and Eckart Zitzler. Scalable multi-objective optimization test problems. In *Proceedings of the 2002 congress on evolutionary computation. CEC'02 (Cat. No. 02TH8600)*, volume 1, pages 825–830. IEEE, 2002.

[26] Madalina M Drugan and Ann Nowe. Designing multi-objective multi-armed bandits algorithms: A study. In *The 2013 international joint conference on neural networks (IJCNN)*, pages 1–8. IEEE, 2013.

[27] Madalina M Drugan and Ann Nowé. Scalarization based pareto optimal set of arms identification algorithms. In *2014 International Joint Conference on Neural Networks (IJCNN)*, pages 2690–2697. IEEE, 2014.

[28] Dmitriy Drusvyatskiy and Courtney Paquette. Efficiency of minimizing compositions of convex functions and smooth maps. *Mathematical Programming*, 178(1):503–558, 2019.

[29] Matthias Ehrgott. *Multicriteria optimization*. Springer, 2005.

[30] Petri Eskelinen and Kaisa Miettinen. Trade-off analysis approach for interactive nonlinear multiobjective optimization. *OR spectrum*, 34(4):803–816, 2012.

[31] Brian S Everitt and Anders Skrondal. *The Cambridge dictionary of statistics*, volume 140. Cambridge university press Cambridge, UK, 2002.

[32] Ilyas Fatkhullin, Niao He, and Yifan Hu. Stochastic optimization under hidden convexity. *SIAM Journal on Optimization*, 35(4):2544–2571, 2025.

[33] Florian Felten, Lucas N. Alegre, Ann Nowé, Ana L. C. Bazzan, El Ghazali Talbi, Grégoire Danoy, and Bruno C. da Silva. A toolkit for reliable benchmarking and research in multi-objective reinforcement learning. In *Proceedings of the 37th Conference on Neural Information Processing Systems (NeurIPS 2023)*, 2023.

[34] Florian Felten, El-Ghazali Talbi, and Grégoire Danoy. Multi-objective reinforcement learning based on decomposition: A taxonomy and framework. *Journal of Artificial Intelligence Research*, 79:679–723, 2024.

[35] Heshan Fernando, Han Shen, Miao Liu, Subhajit Chaudhury, Keerthiram Murugesan, and Tianyi Chen. Mitigating gradient bias in multi-objective learning: A provably convergent stochastic approach. In *Proc. Int. Conf. Learn. Representations*, Kigali, Rwanda, May 2023.

[36] Anthony V Fiacco. Sensitivity analysis for nonlinear programming using penalty methods. *Mathematical programming*, 10(1):287–311, 1976.

[37] Frederick N Fritsch and Ralph E Carlson. Monotone piecewise cubic interpolation. *SIAM Journal on Numerical Analysis*, 17(2):238–246, 1980.

[38] Aurélien Garivier and Wouter M Koolen. Sequential learning of the pareto front for multi-objective bandits. In *International Conference on Artificial Intelligence and Statistics*, pages 3583–3591. PMLR, 2024.

[39] Matthieu Geist, Bruno Scherrer, and Olivier Pietquin. A theory of regularized markov decision processes. In *International conference on machine learning*, pages 2160–2169. PMLR, 2019.

[40] Philip E Gill, Walter Murray, and Michael A Saunders. User's guide for snopt version 7: software for large-scale nonlinear programming. Technical report, Systems Optimization Laboratory, Stanford University, 2006.

[41] Rafael C Gonzalez. *Digital image processing*. Pearson education india, 2009.

[42] Yiju Guo, Ganqu Cui, Lifan Yuan, Ning Ding, Zexu Sun, Bowen Sun, Huimin Chen, Ruobing Xie, Jie Zhou, Yankai Lin, et al. Controllable preference optimization: Toward controllable multi-objective alignment. In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, pages 1437–1454, 2024.

[43] Naoki Hamada, Kenta Hayano, Shunsuke Ichiki, Yutaro Kabata, and Hiroshi Teramoto. Topology of pareto sets of strongly convex problems. *SIAM Journal on Optimization*, 30(3):2659–2686, 2020.

[44] Godfrey Harold Hardy, John Edensor Littlewood, and George Pólya. *Inequalities*. Cambridge university press, 1952.

[45] Edward J Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Liang Wang, Weizhu Chen, et al. Lora: Low-rank adaptation of large language models. *Iclr*, 1(2):3, 2022.

[46] Weizhang Huang and Robert D Russell. *Adaptive moving mesh methods*, volume 174. Springer Science & Business Media, 2010.

[47] Kazufumi Ito and Karl Kunisch. *Lagrange multiplier approach to variational problems and applications*. SIAM, 2008.

[48] Johannes Jahn. Scalarization in multi objective optimization. In *Mathematics of multi objective optimization*, pages 45–88. Springer, 1985.

[49] Emilie Kaufmann, Olivier Cappé, and Aurélien Garivier. On the complexity of best-arm identification in multi-armed bandit models. *The Journal of Machine Learning Research*, 17(1):1–42, 2016.

[50] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

[51] Chengao Li, Hanyu Zhang, Yunkun Xu, Hongyan Xue, Xiang Ao, and Qing He. Gradient-adaptive policy optimization: Towards multi-objective alignment of large language models. In *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 11214–11232, 2025.

[52] Zhuo Li, Guodong Du, Weiyang Guo, Yigeng Zhou, Xiucheng Li, Wenya Wang, Fangming Liu, Yequan Wang, Deheng Ye, Min Zhang, et al. Multi-objective large language model alignment with hierarchical experts. *arXiv preprint arXiv:2505.20925*, 2025.

[53] Baijiong Lin, Weisen Jiang, Yuancheng Xu, Hao Chen, and Ying-Cong Chen. Parm: Multi-objective test-time alignment via preference-aware autoregressive reward model. *arXiv preprint arXiv:2505.06274*, 2025.

[54] Qian Lin, Chao Yu, Zongkai Liu, and Zifan Wu. Policy-regularized offline multi-objective reinforcement learning. *arXiv preprint arXiv:2401.02244*, 2024.

[55] Xi Lin, Xiaoyuan Zhang, Zhiyuan Yang, Fei Liu, Zhenkun Wang, and Qingfu Zhang. Smooth Tchebycheff scalarization for multi-objective optimization. *arXiv preprint arXiv:2402.19078*, 2024.

[56] Xi Lin, Hui-Ling Zhen, Zhenhua Li, Qing-Fu Zhang, and Sam Kwong. Pareto multi-task learning. In *Proc. Advances Neural Inf. Process. Syst.*, Vancouver, Canada, December 2019.

[57] Bo Liu, Xingchao Liu, Xiaojie Jin, Peter Stone, and Qiang Liu. Conflict-Averse Gradient Descent for Multi-task Learning. In *Proc. Advances Neural Inf. Process. Syst.*, virtual, December 2021.

[58] Suyun Liu and Luis Nunes Vicente. The stochastic multi-gradient algorithm for multi-objective optimization and its application to supervised machine learning. *Annals of Operations Research*, pages 1–30, 2021.

[59] Alberto Lovison. Singular continuation: generating piecewise linear approximations to pareto sets via global analysis. *SIAM Journal on Optimization*, 21(2):463–490, 2011.

[60] Debabrata Mahapatra and Vaibhav Rajan. Multi-task learning with user preferences: Gradient descent with controlled ascent in Pareto optimization. In *Proc. Int. Conf. Machine Learn.*, virtual, 2020.

[61] R Timothy Marler and Jasbir S Arora. The weighted sum method for multi-objective optimization: new insights. *Structural and multidisciplinary optimization*, 41(6):853–862, 2010.

[62] Brice Martin, Alexandre Goldsztejn, Laurent Granvilliers, and Christophe Jermann. On continuation methods for non-linear bi-objective optimization: towards a certified interval-based approach. *Journal of Global Optimization*, 64(1):3–16, 2016.

[63] Kaisa Miettinen. *Nonlinear Multiobjective Optimization*, volume 12. Springer US, Boston, MA, 1998.

[64] Michinari Momma, Chaosheng Dong, and Jia Liu. A multi-objective/multi-task learning framework induced by Pareto stationarity. In *Proc. Int. Conf. Machine Learn.*, Baltimore, MD, USA, 2022.

[65] Andrew William Moore. Efficient memory-based learning for robot control. Technical report, University of Cambridge, 1990.

[66] Johannes Müller and Semih Cayci. Optimal rates of convergence for entropy regularization in discounted markov decision processes. *Information and Inference: A Journal of the IMA*, 15(1), 2026.

[67] Aviv Navon, Aviv Shamsian, Ethan Fetaya, and Gal Chechik. Learning the pareto front with hypernetworks. In *Proc. Int. Conf. Learn. Representations*, virtual, April 2020.

[68] Yurii Nesterov and Boris T Polyak. Cubic regularization of newton method and its global performance. *Mathematical programming*, 108(1):177–205, 2006.

[69] Victor Pereyra. Asynchronous distributed solution of large scale nonlinear inversion problems. *Applied numerical mathematics*, 30(1):31–40, 1999.

[70] Victor Pereyra. Fast computation of equispaced pareto manifolds and pareto fronts for multiobjective optimization problems. *Mathematics and Computers in Simulation*, 79(6):1935–1947, 2009.

[71] Victor Pereyra, Michael Saunders, and José Castillo. Equispaced pareto front construction for constrained bi-objective optimization. *Mathematical and Computer Modelling*, 57(9-10):2122–2131, 2013.

[72] Ieva Petrulionyte, Julien Mairal, and Michael Arbel. Functional bilevel optimization for machine learning. *Advances in Neural Information Processing Systems*, 37:14016–14065, 2024.

[73] Shuang Qiu, Dake Zhang, Rui Yang, Boxiang Lyu, and Tong Zhang. Traversing pareto optimal policies: Provably efficient multi-objective reinforcement learning. *arXiv preprint arXiv:2407.17466*, 2024.

[74] Qwen Team, An Yang, Baosong Yang, Beichen Zhang, Binyuan Hui, Bo Zheng, Bowen Yu, Chengyuan Li, Dayiheng Liu, Fei Huang, Haoran Wei, Huan Lin, Jian Yang, Jianhong Tu, Jianwei Zhang, Jianxin Yang, Jiaxi Yang, Jingren Zhou, Junyang Lin, Kai Dang, Keming Lu, Keqin Bao, Kexin Yang, Le Yu, Mei Li, Mingfeng Xue, Pei Zhang, Qin Zhu, Rui Men, Runji Lin, Tianhao Li, Tianyi Tang, Tingyu Xia, Xingzhang Ren, Xuancheng Ren, Yang Fan, Yang Su, Yichang Zhang, Yu Wan, Yuqiong Liu, Zeyu Cui, Zhenru Zhang, and Zihan Qiu. Qwen2.5 technical report, 2025.

[75] Rafael Rafailov, Archit Sharma, Eric Mitchell, Christopher D Manning, Stefano Ermon, and Chelsea Finn. Direct preference optimization: Your language model is secretly a reward model. *Advances in neural information processing systems*, 36:53728–53741, 2023.

[76] Alexandre Rame, Guillaume Couairon, Corentin Dancette, Jean-Baptiste Gaya, Mustafa Shukor, Laure Soulier, and Matthieu Cord. Rewarded soups: towards pareto-optimal alignment by interpolating weights fine-tuned on diverse rewards. *Advances in Neural Information Processing Systems*, 36:71095–71134, 2023.

[77] Mathieu Reymond, Eugenio Bargiacchi, and Ann Nowé. Pareto conditioned networks. *arXiv preprint arXiv:2204.05036*, 2022.

[78] Diederik M Roijers, Denis Steckelmacher, and Ann Nowé. Multi-objective reinforcement learning for the expected utility of the return. In *Proceedings of the Adaptive and Learning Agents workshop at FAIM*, volume 2018, 2018.

[79] Diederik M Roijers, Peter Vamplew, Shimon Whiteson, and Richard Dazeley. A survey of multi-objective sequential decision-making. *Journal of Artificial Intelligence Research*, 48:67–113, 2013.

[80] Abhishek Roy, Geelon So, and Yi-An Ma. Optimization on pareto sets: On a theory of multi-objective optimization. *arXiv preprint arXiv:2308.02145*, 2023.

[81] Andreas Schlaginhaufen and Maryam Kamgarpour. Towards the transferability of rewards recovered via regularized inverse reinforcement learning. *Advances in Neural Information Processing Systems*, 37:21461–21501, 2024.

[82] John Schulman, Filip Wolski, Prafulla Dhariwal, Alec Radford, and Oleg Klimov. Proximal policy optimization algorithms. *arXiv preprint arXiv:1707.06347*, 2017.

[83] Oliver Schütze, Alessandro Dell'Aere, and Michael Dellnitz. On continuation methods for the numerical treatment of multi-objective optimization problems. In *Practical Approaches to Multi-Objective Optimization*, 2005.

[84] Ozan Sener and Vladlen Koltun. Multi-task learning as multi-objective optimization. In *Proc. Advances Neural Inf. Process. Syst.*, Montreal, Canada, December 2018.

[85] Yiran Shen, Yu Xia, Jonathan Chang, and Prithviraj Ammanabrolu. Simultaneous multi-objective alignment across verifiable and non-verifiable rewards. *arXiv preprint arXiv:2510.01167*, 2025.

[86] Ruizhe Shi, Yifang Chen, Yushi Hu, Alisa Liu, Hanna Hajishirzi, Noah A Smith, and Simon S Du. Decoding-time language model alignment with multiple objectives. *Advances in Neural Information Processing Systems*, 37:48875–48920, 2024.

[87] Nisan Stiennon, Long Ouyang, Jeffrey Wu, Daniel Ziegler, Ryan Lowe, Chelsea Voss, Alec Radford, Dario Amodei, and Paul F Christiano. Learning to summarize with human feedback. *Advances in neural information processing systems*, 33:3008–3021, 2020.

[88] Endre Süli and David F Mayers. *An introduction to numerical analysis*. Cambridge university press, 2003.

[89] Peter Vamplew, Richard Dazeley, Adam Berry, Rustam Issabekov, and Evan Dekker. Empirical evaluation methods for multi-objective reinforcement learning algorithms. *Machine learning*, 84(1):51–80, 2011.

[90] Diogo A Vieira, Ricardo HC Takahashi, and Rodney R Saldanha. Multicriteria optimization with a multiobjective golden section line search. *Mathematical Programming*, 131(1):131–161, 2012.

[91] Michael Völske, Martin Potthast, Shahbaz Syed, and Benno Stein. Tl; dr: Mining reddit to learn automatic summarization. In *Proceedings of the workshop on new frontiers in summarization*, pages 59–63, 2017.

[92] Hao Wang, Antonio E Rodriguez-Fernandez, Luis Uribe, Andre Deutz, Oliver Cortés-Piña, and Oliver Schütze. A newton method for hausdorff approximations of the pareto front within multi-objective evolutionary algorithms. *IEEE Transactions on Evolutionary Computation*, 2024.

[93] Kaiwen Wang, Rahul Kidambi, Ryan Sullivan, Alekh Agarwal, Christoph Dann, Andrea Michi, Marco Gelmi, Yunxuan Li, Raghav Gupta, Kumar Avinava Dubey, et al. Conditional language policy: A general framework for steerable multi-objective finetuning. In *Findings of the Association for Computational Linguistics: EMNLP 2024*, pages 2153–2186, 2024.

[94] Hilary Weller, Philip Browne, Chris Budd, and Mike Cullen. Mesh adaptation on the sphere using optimal transport and the numerical solution of a monge–ampère type equation. *Journal of Computational Physics*, 308:102–123, 2016.

[95] Andrzej P Wierzbicki. Reference point methods in vector optimization and decision support. Technical Report IR-98-017, International Institute for Applied Systems Analysis (IIASA), Laxenburg, Austria, 1998.

[96] Marcus Williams. Multi-objective reinforcement learning from ai feedback. *arXiv preprint arXiv:2406.07295*, 2024.

[97] Peiyao Xiao, Hao Ban, and Kaiyi Ji. Direction-oriented multi-objective learning: Simple and provable stochastic algorithms. In *Proc. Advances Neural Inf. Process. Syst.*, New Orleans, LA, USA, 2023.

[98] Nuoya Xiong and Aarti Singh. Projection optimization: A general framework for multi-objective and multi-group rlhf. *arXiv preprint arXiv:2502.15145*, 2025.

[99] Mengfan Xu and Diego Klabjan. Pareto regret analyses in multi-objective multi-armed bandit. In *International Conference on Machine Learning*, pages 38499–38517. PMLR, 2023.

[100] X Xu, Weizhang Huang, RD Russell, and JF Williams. Convergence of de boor's algorithm for the generation of equidistributing meshes. *IMA Journal of Numerical Analysis*, 31(2):580–596, 2011.

[101] Saba Yahyaa and Bernard Manderick. Thompson sampling for multi-objective multi-armed bandits problem. In *Proceedings of the European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning (ESANN 2015)*, 2015.

[102] Kailai Yang, Zhiwei Liu, Qianqian Xie, Jimin Huang, Tianlin Zhang, and Sophia Ananiadou. Metaaligner: Towards generalizable multi-objective alignment of language models. *Advances in Neural Information Processing Systems*, 37:34453–34486, 2024.

[103] Qinwei Yang, Xueqing Liu, Yan Zeng, Ruocheng Guo, Yang Liu, and Peng Wu. Learning the optimal policy for balancing short-term and long-term rewards. *Advances in Neural Information Processing Systems*, 37:36514–36540, 2024.

[104] Rui Yang, Xiaoman Pan, Feng Luo, Shuang Qiu, Han Zhong, Dong Yu, and Jianshu Chen. Rewards-in-context: Multi-objective alignment of foundation models with dynamic preference adjustment. *arXiv preprint arXiv:2402.10207*, 2024.

[105] Mao Ye and Qiang Liu. Pareto navigation gradient descent: a first-order algorithm for optimization in pareto set. In *Uncertainty in artificial intelligence*, pages 2246–2255, 2022.

[106] Bianca Zadrozny and Charles Elkan. Transforming classifier scores into accurate multiclass probability estimates. In *Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 694–699, 2002.

[107] Junyu Zhang, Alec Koppel, Amrit Singh Bedi, Csaba Szepesvari, and Mengdi Wang. Variational policy gradient method for reinforcement learning with general utilities. *Advances in Neural Information Processing Systems*, 33:4572–4583, 2020.

[108] Qiuyi Richard Zhang. Optimal scalarizations for sublinear hypervolume regret. *Advances in Neural Information Processing Systems*, 37:39963–39999, 2024.

[109] Wei Hong Zhang and Tong Gao. A min–max method with adaptive weightings for uniformly spaced pareto optimum points. *Computers & structures*, 84(28):1760–1769, 2006.

[110] Xiaoyuan Zhang, Genghui Li, Xi Lin, Yichi Zhang, Yifan Chen, and Qingfu Zhang. Gliding over the pareto front with uniform designs. *Advances in Neural Information Processing Systems*, 37:2215–2245, 2024.

[111] Xiaoyuan Zhang, Liang Zhao, Yingying Yu, Xi Lin, Yifan Chen, Han Zhao, and Qingfu Zhang. Libmoon: A gradient-based multiobjective optimization library in pytorch. *Advances in Neural Information Processing Systems*, 37:2026–2044, 2024.

[112] Shiji Zhou, Wenpeng Zhang, Jiyan Jiang, Wenliang Zhong, Jinjie Gu, and Wenwu Zhu. On the convergence of stochastic multi-objective gradient manipulation and beyond. In *Proc. Advances Neural Inf. Process. Syst.*, volume 35, pages 38103–38115, New Orleans, LA, USA, December 2022.

[113] Zhanhui Zhou, Jie Liu, Jing Shao, Xiangyu Yue, Chao Yang, Wanli Ouyang, and Yu Qiao. Beyond one-preference-fits-all alignment: Multi-objective direct preference optimization. In *Findings of the Association for Computational Linguistics: ACL 2024*, pages 10586–10613, 2024.

[114] Eckart Zitzler, Kalyanmoy Deb, and Lothar Thiele. Comparison of multiobjective evolutionary algorithms: Empirical results. *Evolutionary computation*, 8(2):173–195, 2000.

[115] Eckart Zitzler and Lothar Thiele. Multiobjective evolutionary algorithms: a comparative case study and the strength pareto approach. *IEEE transactions on Evolutionary Computation*, 3(4):257–271, 1999.

Supplementary Material: Contents

A Additional Related Work **17**

B Notation Table **19**

C Preliminaries and Geometry Analysis Details **19**

C.1 Smoothness of the PF 19

C.2 Proof of Lemma 1 22

C.3 Extension to the Multi-Objective Setting 24

C.4 Analytical Expressions for Φ in Quadratic Problems 25

D PF Geometry in Bi-Objective Reinforcement Learning **26**

D.1 Provable Regularity Condition for Discounted Occupancy Constraints 26

D.2 Inner Optimization Trajectories Stay in the Subset $\mathcal{K}$ in the Relative Interior of $\mathcal{X}$. 26

D.3 Proof of Proposition 2: PF Speed in Bi-Objective RL 27

D.4 CDF Estimation Error in Bandit Problems 28

E Convergence Analysis of SURF **29**

E.1 Proof of the First Part of Theorem 1: Discretization Error under Exact InnerSolver 30

E.2 Proof of the Second Part of Theorem 1: Optimization Error Control under Inexact InnerSolver Solutions 40

E.3 Connection between SURF Convergence and PF Uniformity 46

F Experimental Details **47**

F.1 Toy Examples 47

F.2 SURF with Tchebycheff Scalarization on Nonconvex Fronts 48

F.3 Details of Experiments on MO-Gymnasium Benchmarks 48

F.3.1 Details of DST Problem 48

F.3.2 Details of Fishwood Problem 51

F.3.3 Details of MO-Mountaincar Problem 52

F.4 LLM Multi-Objective Fine-Tuning 52

G Limitations and Broader Impacts **53**

A Additional Related Work

MORL and LLM Alignment. Multi-objective reinforcement learning provides a natural application domain because one often seeks not a single policy, but a family of policies spanning different trade-offs among objectives or diverse preferences of users [77, 34, 54]. In language-model alignment, the same need arises when balancing criteria such as helpfulness, faithfulness, harmlessness, or style control [86, 102, 42, 85]. Recent work on multi-objective reinforcement learning from AI feedback and multi-objective decoding highlights both the importance of this setting and the practical appeal of scalarization-based [73, 113, 96, 93, 51] or interpolation-based pipelines [76, 86, 52, 53]. Our goal is complementary: rather than replacing scalarization, we characterize and correct its

Symbol	Description
$\mathbf{u} \in \mathcal{U}$	Decision variable and feasible set in the original space
$\mathbf{h}(\mathbf{u}) = [h_1(\mathbf{u}), \cdots, h_M(\mathbf{u})]^\top$	Objective vector in the original space
$\min_{\mathbf{u}\in\mathcal{U}} \mathbf{h}(\mathbf{u})$	Original multi-objective problem
$\mathcal{U}_{\mathrm{PF}}$	Pareto set in the original space
$\mathbf{x} = g(\mathbf{u}) \in \mathcal{X} = g(\mathcal{U})$	Hidden-space variable and feasible space induced by g
$\mathbf{f}(\mathbf{x}) = [f_1(\mathbf{x}), \cdots, f_M(\mathbf{x})]^\top$	Objectives in the hidden space, with $\mathbf{f}(\mathbf{x}) = \mathbf{h}(g^{-1}(\mathbf{x}))$
$\min_{\mathbf{x}\in\mathcal{X}} \mathbf{f}(\mathbf{x})$	Hidden-space multi-objective problem
$\mathcal{X}_{\mathrm{PF}} = g(\mathcal{U}_{\mathrm{PF}})$	Pareto set in the hidden space
$\mathcal{F}_{\mathrm{PF}} = \mathbf{h}(\mathcal{U}_{\mathrm{PF}}) = f(\mathcal{X}_{\mathrm{PF}})$	Pareto front
$\mathbf{w} \in \Delta_M$	LS weight vector
$\mathrm{LS}_h(\mathbf{u}; \mathbf{w}) = \sum_{m=1}^M w_m h_m(\mathbf{u})$	LS objective in the original space
$\mathbf{u}^*_{\mathbf{w}} \in \arg\min_{\mathbf{u}\in\mathcal{U}} \mathrm{LS}_h(\mathbf{u}; \mathbf{w})$	LS optimizer in the original space
$\mathrm{LS}_f(\mathbf{x}; \mathbf{w}) = \sum_{m=1}^M w_m f_m(\mathbf{x})$	LS objective in the hidden space
$\mathbf{x}^*_{\mathbf{w}} \in \arg\min_{\mathbf{x}\in\mathcal{X}} \mathrm{LS}_f(\mathbf{x}; \mathbf{w})$	LS optimizer in the hidden space
$\mathbf{f}_{\mathrm{PF}}(\mathbf{w}) = \mathbf{f}(\mathbf{x}^*_{\mathbf{w}}) = \mathbf{h}(\mathbf{u}^*_{\mathbf{w}})$	Multi-objective PF map induced by LS
$w \in [0, 1]$	Scalar LS weight for $M = 2$, with $\mathbf{w} = [w, 1 - w]$
$\mathrm{LS}_h(\mathbf{u}; w), \mathbf{u}^*_w, \mathrm{LS}_f(\mathbf{x}; w), \mathbf{x}^*_w, \mathbf{f}_{\mathrm{PF}}(w)$	Shorthand for the $M = 2$ case with $\mathbf{w} = [w, 1 - w]$
$v(w) = \left\|\frac{\partial}{\partial w}\mathbf{f}_{\mathrm{PF}}(w)\right\|$	Bi-objective PF traversal speed
$s(w) = \int_0^w v(p)dp$	Bi-objective PF arc length up to weight w
$\Phi(w) = \frac{s(w)}{s(1)}$	Normalized arc-length CDF

Table 3: Summary of notation. The first block lists the original-space MOO problem and hidden-space reformulation through the bijection $\mathbf{x} = g(\mathbf{u})$, the second block defines LS weights, scalarized objectives, optimizers, and PF maps, and the final block summarizes the induced PF geometry for $M = 2$, where the scalar weight $w \in [0, 1]$ induces the path $\mathbf{f}_{\mathrm{PF}}(w)$, whose traversal speed, arc length, and normalized arc-length CDF are given by v, s, and Φ, respectively.

geometric sampling bias so that repeated scalarized solves yield a uniformly representative set of Pareto solutions.

Gradient-Based MOO and Scalarization. Scalarization methods for reducing MOO to single-objective subproblems are widely used in practice [61], including LS, Tchebycheff scalarization [48, 55], ϵ-constraint method and its variants [14, 64, 60, 105, 18, 80, 17], and reference-point methods [95]. Among them, LS is a commonly used baseline [113, 103, 108, 19]. Additionally, gradient-based multi-objective methods such as MGDA [84, 57, 58, 112, 35, 15, 97] typically return a single Pareto-stationary solution per run, rather than a preference-specified or uniformly distributed set of Pareto points.

PF Approximation. A variety of methods have been proposed to achieve some coverage of the PF. Classical Normal Boundary Intersection (NBI) method [24] generates well-distributed points via structured subproblems. Optimistic Linear Support (OLS) adaptively selects scalarization weights to construct the convex coverage set using repeated single-objective solves, making it a standard adaptive scalarization baseline in MORL [5, 1]. Adaptive weighting strategies, such as min–max formulations with local arc-length control, promote uniform spacing but typically rely on sequential PF tracing and differential-geometry ingredients [109]. Another line of work enforces equal-spacing constraints explicitly within an optimization framework, solved sequentially using off-the-shelf optimizers [40] or in parallel via second-order techniques [69, 70, 71], which can be memory intensive and less attractive when only first-order or noisy oracle access is available. Another line of work employs continuation or PF-tracking methods, tracing the Pareto manifold via differential equations and predictor-corrector schemes [59, 62, 90, 10]. Recent studies assess uniformity via the averaged Hausdorff distance [9], and develop second-order methods to promote uniform PF spacing [92, 110], which are expensive due to their second-order nature and the need for additional operations, e.g., sorting or clustering.

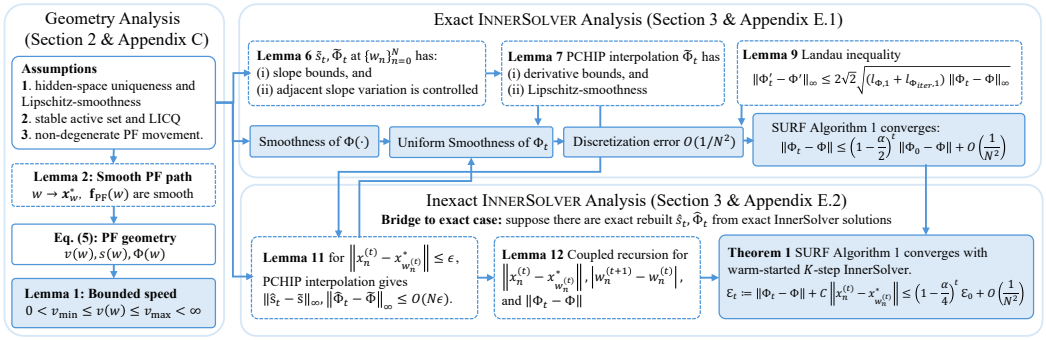

Figure 7: Theoretical analysis roadmap for SURF. The solid-line blocks summarize the main results presented in the main paper, while the dashed-line blocks indicate supporting intermediate results developed in the appendix.

Pushforward Measures and Distribution Matching. Histogram equalization provides a classical example of distribution matching by transforming an empirical distribution to a uniform target [41, 8]. In our setting, treating the scalarization weight w as a random variable induces a pushforward measure on the PF through the map $\mathbf{f}_{\mathrm{PF}}(w)$. Uniform arc-length sampling can therefore be viewed as matching this induced measure to a uniform target on the front. Related perspectives arise in monotone calibration [106], adaptive moving-mesh methods [46, 100], and optimal-transport-based mesh adaptation [13, 94].

B Notation Table

Table 3 summarizes the main problem-specific notation used throughout the paper, grouped by original-space quantities, hidden-space quantities, and PF geometry.

C Preliminaries and Geometry Analysis Details

In this appendix, we develop the geometric ingredients shown in the left part of Figure 7. Specifically, Appendix C.1 proves smoothness of the scalarization-induced PF path, Appendix C.2 establishes bounds on the traversal speed, Appendix C.3 discusses the extension to the multiple-objective setting, and Appendix C.4 provides an analytical quadratic example.

C.1 Smoothness of the PF

The following lemma shows that LS induces a Lipschitz-smooth PF path, and that the PF point moves in a monotone direction as the scalarization weight varies.

Lemma 2 (Smoothness and monotonicity of the scalarization path)**.** *Suppose Assumptions 1, 2 hold. Then, for every $w \in [0,1]$, the scalarized problem* (3) *has a unique minimizer $\mathbf{x}^*_w$. Moreover, the mapping $w \mapsto \mathbf{x}^*_w$ is Lipschitz-continuous on $[0,1]$, continuously differentiable on $(0,1)$, and admits one-sided derivatives at $w = 0$ and $w = 1$ that are the one-sided limits of $\partial_w \mathbf{x}^*_w$. The mapping*

$$w \mapsto \mathbf{f}_{\mathrm{PF}}(w) = \mathbf{f}(\mathbf{x}^*_w) = [f_{\mathrm{PF}_1}(w), f_{\mathrm{PF}_2}(w)]^\top \tag{17}$$

is continuously differentiable on $(0,1)$ and admits one-sided derivatives at $w = 0$ and $w = 1$ that are the one-sided limits of $\partial_w \mathbf{f}_{\mathrm{PF}}(w)$. Additionally, f_{PF_1} is nonincreasing and f_{PF_2} is nondecreasing on $[0,1]$. In particular,

$$\frac{\partial}{\partial w}\mathbf{f}_{\mathrm{PF}}(w) \in \mathbb{R}_- \times \mathbb{R}_+, \quad \forall w \in [0,1], \tag{18}$$

that is, $\frac{\partial}{\partial w} f_{\mathrm{PF}_1}(w) \le 0$ and $\frac{\partial}{\partial w} f_{\mathrm{PF}_2}(w) \ge 0$ for all $w \in [0,1]$.

Additionally, if Assumption 3 holds, $\mathbf{f}_{\mathrm{PF}}(\cdot)$, $\Phi(\cdot)$, and $\mathbf{f}_{\mathrm{PF}}(s^{-1}(\cdot))$ are respectively $l_{\mathbf{f}_{\mathrm{PF}},1}$-, $l_{\Phi,1}$- and $l_{\tilde{\mathbf{f}}_{\mathrm{PF}},1}$-Lipschitz-smooth, and v is $l_{v,0}$-Lipschitz-continuous.

Proof. Fix any $w \in [0,1]$. By Assumption 1, the scalarized objective

$$\mathrm{LS}_f(\mathbf{x}; w) = w f_1(\mathbf{x}) + (1-w) f_2(\mathbf{x}) \tag{19}$$

is μ-strongly convex on the convex set $\mathcal{X}$. Therefore, the minimizer $\mathbf{x}_w^* = \arg\min_{\mathbf{x}\in\mathcal{X}} \mathrm{LS}_f(\mathbf{x}; w)$ is unique.

We next study how this solution varies with w. Under Assumptions 1, 2, let $\boldsymbol{\nu}_w^*$ denote the corresponding optimal Lagrange multiplier vector for the scalarized problem (3). By the standard sensitivity result for nonlinear programming, the primal-dual solution mapping $w \mapsto (\mathbf{x}_w^*, \boldsymbol{\nu}_w^*)$ is locally continuously differentiable in w [36, Theorem 2.1]. Moreover, the feasible set $\mathcal{X}$ does not depend on w, and the scalarized objective $\mathrm{LS}_f(\mathbf{x}; w)$ is affine, hence Lipschitz-continuous, in w. Therefore, $w \mapsto (\mathbf{x}_w^*, \boldsymbol{\nu}_w^*)$ is also Lipschitz-continuous in w [47, Theorem 2.16]. In particular, $w \mapsto \mathbf{x}_w^*$ is Lipschitz-continuous on $[0,1]$ and continuously differentiable on $(0,1)$, with one-sided derivatives at $w = 0$ and $w = 1$. Since f_1 and f_2 are continuously differentiable, the composition $\mathbf{f}_{\mathrm{PF}}(w) = \mathbf{f}(\mathbf{x}_w^*)$ is continuously differentiable on $(0,1)$, with one-sided derivatives at the endpoints.

It remains to show the monotonicity of the PF coordinates. Fix $0 \le w_1 < w_2 \le 1$. By optimality of $\mathbf{x}_{w_1}^*$ for weight w_1,

$$w_1 f_1(\mathbf{x}_{w_1}^*) + (1-w_1) f_2(\mathbf{x}_{w_1}^*) \le w_1 f_1(\mathbf{x}_{w_2}^*) + (1-w_1) f_2(\mathbf{x}_{w_2}^*). \tag{20}$$

Similarly, by optimality of $\mathbf{x}_{w_2}^*$ for weight w_2,

$$w_2 f_1(\mathbf{x}_{w_2}^*) + (1-w_2) f_2(\mathbf{x}_{w_2}^*) \le w_2 f_1(\mathbf{x}_{w_1}^*) + (1-w_2) f_2(\mathbf{x}_{w_1}^*). \tag{21}$$

Multiplying (20) by $1-w_2$ and (21) by $1-w_1$, then adding, gives $(w_2-w_1)\big(f_1(\mathbf{x}_{w_2}^*) - f_1(\mathbf{x}_{w_1}^*)\big) \le 0$. Since $w_2 - w_1 > 0$, $f_{\mathrm{PF}\,1}$ is nonincreasing on $[0,1]$, as follows from

$$f_{\mathrm{PF}\,1}(w_2) = f_1(\mathbf{x}_{w_2}^*) \le f_1(\mathbf{x}_{w_1}^*) = f_{\mathrm{PF}\,1}(w_1). \tag{22}$$

Substituting this back into (20), we obtain

$$(1-w_1)\big(f_2(\mathbf{x}_{w_1}^*) - f_2(\mathbf{x}_{w_2}^*)\big) \le w_1\big(f_1(\mathbf{x}_{w_2}^*) - f_1(\mathbf{x}_{w_1}^*)\big) \le 0, \tag{23}$$

hence $f_{\mathrm{PF}\,2}(w_1) = f_2(\mathbf{x}_{w_1}^*) \le f_2(\mathbf{x}_{w_2}^*) = f_{\mathrm{PF}\,2}(w_2)$ and $f_{\mathrm{PF}\,2}$ is nondecreasing on $[0,1]$.

Since $\mathbf{f}_{\mathrm{PF}}$ is continuously differentiable on $[0,1]$, the above monotonicity implies $\frac{\partial}{\partial w} f_{\mathrm{PF}\,1}(w) \le 0$ and $\frac{\partial}{\partial w} f_{\mathrm{PF}\,2}(w) \ge 0$ for all $w \in [0,1]$. Equivalently, $\frac{\partial}{\partial w}\mathbf{f}_{\mathrm{PF}}(w) \in \mathbb{R}_- \times \mathbb{R}_+$. Let $\mathbf{c}(\mathbf{x})$ denote the fixed active constraint vector along the solution path, and let $\boldsymbol{\nu}_w^*$ be the associated Lagrange multiplier vector at $\mathbf{x}_w^*$, so

$$\mathcal{L}(\mathbf{x}, \boldsymbol{\nu}; w) := w f_1(\mathbf{x}) + (1-w) f_2(\mathbf{x}) + \boldsymbol{\nu}^\top \mathbf{c}(\mathbf{x}). \tag{24}$$

By Assumption 2, the local KKT system is

$$w\nabla f_1(\mathbf{x}_w^*) + (1-w)\nabla f_2(\mathbf{x}_w^*) + J_w^\top \boldsymbol{\nu}_w^* = 0, \quad \mathbf{c}(\mathbf{x}_w^*) = 0, \tag{25}$$

where $J_w := \nabla \mathbf{c}(\mathbf{x}_w^*)$. As discussed above, $w \mapsto (\mathbf{x}_w^*, \boldsymbol{\nu}_w^*)$ is Lipschitz-continuous on $[0,1]$, continuously differentiable on $(0,1)$, and admits one-sided derivatives at the endpoints.

Differentiating (25) with respect to w gives

$$\underbrace{\begin{bmatrix} \nabla^2_{\mathbf{xx}}\mathcal{L}(\mathbf{x}_w^*, \boldsymbol{\nu}_w^*; w) & J_w^\top \\ J_w & 0 \end{bmatrix}}_{K(w):=} \begin{bmatrix} \frac{\partial}{\partial w}\mathbf{x}_w^* \\ \frac{\partial}{\partial w}\boldsymbol{\nu}_w^* \end{bmatrix} = -\begin{bmatrix} \nabla f_1(\mathbf{x}_w^*) - \nabla f_2(\mathbf{x}_w^*) \\ 0 \end{bmatrix}. \tag{26}$$

By Assumption 1, strong convexity on $\mathcal{X}$ implies that $\nabla^2_{\mathbf{xx}}\mathcal{L}(\mathbf{x}_w^*, \boldsymbol{\nu}_w^*; w)$ is positive definite on the active feasible tangent space $\ker(J_w)$; together with LICQ, this implies that $K(w)$ is nonsingular for every $w \in [0,1]$. To use a full-space inverse, choose a sufficiently large fixed $c > 0$ and define

$$\widetilde{\nabla}^2_{\mathbf{xx}}\mathcal{L}(\mathbf{x}_w^*, \boldsymbol{\nu}_w^*; w) := \nabla^2_{\mathbf{xx}}\mathcal{L}(\mathbf{x}_w^*, \boldsymbol{\nu}_w^*; w) + c J_w^\top J_w. \tag{27}$$

Since $\nabla^2_{\mathbf{xx}}\mathcal{L}(\mathbf{x}_w^*, \boldsymbol{\nu}_w^*; w)$ is uniformly positive definite on $\ker(J_w)$ and J_w has full row rank by LICQ, this choice gives $\widetilde{\nabla}^2_{\mathbf{xx}}\mathcal{L}(\mathbf{x}_w^*, \boldsymbol{\nu}_w^*; w) \succ 0$ uniformly along the compact path. Moreover,

$J_w \frac{\partial}{\partial w}\mathbf{x}_w^* = 0$, so replacing $\nabla^2_{\mathbf{xx}}\mathcal{L}(\mathbf{x}_w^*, \boldsymbol{\nu}_w^*; w)$ by $\widetilde{\nabla}^2_{\mathbf{xx}}\mathcal{L}(\mathbf{x}_w^*, \boldsymbol{\nu}_w^*; w)$ does not change the primal sensitivity equation. Since $\widetilde{\nabla}^2_{\mathbf{xx}}\mathcal{L}(\mathbf{x}_w^*, \boldsymbol{\nu}_w^*; w) \succ 0$ and J_w has full row rank, the Schur-complement inverse formula applies to the augmented KKT system. On the compact path $\{\mathbf{x}_w^* : w \in [0,1]\}$, the uniform restricted strong convexity and LICQ constants imply that $\widetilde{\nabla}^2_{\mathbf{xx}}\mathcal{L}(\mathbf{x}_w^*, \boldsymbol{\nu}_w^*; w)^{-1}$ and $\left(J_w \widetilde{\nabla}^2_{\mathbf{xx}}\mathcal{L}(\mathbf{x}_w^*, \boldsymbol{\nu}_w^*; w)^{-1} J_w^\top\right)^{-1}$ are uniformly bounded. Moreover, by Assumption 2, the Lagrangian Hessian is Lipschitz-continuous along the solution path; together with the fixed active set and LICQ, this implies that the Schur-complement factor mapping the right-hand side of (26) to $\frac{\partial}{\partial w}\mathbf{x}_w^*$ is Lipschitz-continuous in w. Since $w \mapsto \mathbf{x}_w^*$ is Lipschitz-continuous and ∇f_m is Lipschitz-continuous on $\mathcal{K}$, the vector $\nabla f_1(\mathbf{x}_w^*) - \nabla f_2(\mathbf{x}_w^*)$ is Lipschitz-continuous in w. Therefore, there exists $L_x < \infty$ such that

$$\left\| \frac{\partial}{\partial w}\mathbf{x}_{w_1}^* - \frac{\partial}{\partial w}\mathbf{x}_{w_2}^* \right\| \le L_x |w_1 - w_2|. \tag{28}$$

By the chain rule, $\frac{\partial}{\partial w}\mathbf{f}_{\mathrm{PF}\,m}(w) = \nabla f_m(\mathbf{x}_w^*)^\top \frac{\partial}{\partial w}\mathbf{x}_w^*$, for $m \in \{1,2\}$. Since ∇f_m is Lipschitz-continuous and bounded on the compact set containing $\{\mathbf{x}_w^* : w \in [0,1]\}$, while $w \mapsto \mathbf{x}_w^*$ is Lipschitz-continuous and $\frac{\partial}{\partial w}\mathbf{x}_w^*$ is bounded and Lipschitz-continuous by (28), we have

$$\begin{aligned}
&\left\| \frac{\partial}{\partial w}\mathbf{f}_{\mathrm{PF}}(w_1) - \frac{\partial}{\partial w}\mathbf{f}_{\mathrm{PF}}(w_2) \right\| \\
\le& \sum_{m=1}^{2} \left| \nabla f_m(\mathbf{x}_{w_1}^*)^\top \frac{\partial}{\partial w}\mathbf{x}_{w_1}^* - \nabla f_m(\mathbf{x}_{w_2}^*)^\top \frac{\partial}{\partial w}\mathbf{x}_{w_2}^* \right| \\
\le& \sum_{m=1}^{2} \left\| \nabla f_m(\mathbf{x}_{w_1}^*) - \nabla f_m(\mathbf{x}_{w_2}^*) \right\| \left\| \frac{\partial}{\partial w}\mathbf{x}_{w_1}^* \right\| + \sum_{m=1}^{2} \left\| \nabla f_m(\mathbf{x}_{w_2}^*) \right\| \left\| \frac{\partial}{\partial w}\mathbf{x}_{w_1}^* - \frac{\partial}{\partial w}\mathbf{x}_{w_2}^* \right\| \\
\le& l_{\mathbf{f}_{\mathrm{PF}},1} |w_1 - w_2|
\end{aligned} \tag{29}$$

for some constant $l_{\mathbf{f}_{\mathrm{PF}},1} < \infty$. Now, since $v(w) = \left\| \frac{\partial}{\partial w}\mathbf{f}_{\mathrm{PF}}(w) \right\|$, the triangle inequality yields

$$|v(w_1) - v(w_2)| \le \left\| \frac{\partial}{\partial w}\mathbf{f}_{\mathrm{PF}}(w_1) - \frac{\partial}{\partial w}\mathbf{f}_{\mathrm{PF}}(w_2) \right\| \le l_{\mathbf{f}_{\mathrm{PF}},1} |w_1 - w_2|. \tag{30}$$

Thus v is Lipschitz-continuous on $[0,1]$ with modulus $l_{v,0} \le l_{\mathbf{f}_{\mathrm{PF}},1}$. For $\mathbf{f}_{\mathrm{PF}\,s}(s) := \mathbf{f}_{\mathrm{PF}}(s^{-1}(s))$, by Lemma 1,

$$v(w) = s'(w) \ge v_{\min} > 0, \quad \forall w \in [0,1]. \tag{31}$$

Hence s is strictly increasing, so s^{-1} is well defined and Lipschitz-continuous with constant at most $1/v_{\min}$. By the chain rule, there is $\mathbf{f}_{\mathrm{PF}}{}'_s(s) = \frac{\frac{\partial}{\partial w}\mathbf{f}_{\mathrm{PF}}(w)}{v(w)}\big|_{w=s^{-1}(s)}$. For any $w_1, w_2 \in [0,1]$,

$$\begin{aligned}
0 \le& \left\| \frac{\frac{\partial}{\partial w}\mathbf{f}_{\mathrm{PF}}(w_1)}{v(w_1)} - \frac{\frac{\partial}{\partial w}\mathbf{f}_{\mathrm{PF}}(w_2)}{v(w_2)} \right\| \\
\le& \frac{\left\| \frac{\partial}{\partial w}\mathbf{f}_{\mathrm{PF}}(w_1) - \frac{\partial}{\partial w}\mathbf{f}_{\mathrm{PF}}(w_2) \right\|}{v_{\min}} + \left\| \frac{\partial}{\partial w}\mathbf{f}_{\mathrm{PF}}(w_2) \right\| \left| \frac{1}{v(w_1)} - \frac{1}{v(w_2)} \right| \\
\le& \frac{l_{\mathbf{f}_{\mathrm{PF}},1}}{v_{\min}} |w_1 - w_2| + \frac{v_{\max} l_{v,0}}{v_{\min}^2} |w_1 - w_2|.
\end{aligned} \tag{32}$$

Now let $s_1, s_2 \in [0, s(1)]$, and write $w_i = s^{-1}(s_i)$. Then there exists $l_{\mathbf{f}_{\mathrm{PF}\,s},1} < \infty$ such that

$$\|\mathbf{f}_{\mathrm{PF}}{}'_s(s_1) - \mathbf{f}_{\mathrm{PF}}{}'_s(s_2)\| \le l_{\mathbf{f}_{\mathrm{PF}\,s},1} |s_1 - s_2|. \tag{33}$$

Finally, by the fundamental theorem of calculus, s is differentiable with $s'(w) = v(w)$. Since $\Phi = s/s(1)$, there is $\Phi'(w) = \frac{s'(w)}{s(1)} = \frac{v(w)}{s(1)}$. Therefore, for any $w_1, w_2 \in [0,1]$,

$$|\Phi'(w_1) - \Phi'(w_2)| = \frac{|v(w_1) - v(w_2)|}{s(1)} \le \frac{l_{v,0}}{s(1)} |w_1 - w_2|. \tag{34}$$

Thus Φ' is Lipschitz-continuous on $[0,1]$ with modulus $l_{\Phi,1} \le \frac{l_{v,0}}{s(1)}$. This completes the proof. $\square$

Lemma 2 shows that LS induces a Lipschitz-continuous and Lipschitz-smooth PF path. A one-to-one parametrization by w requires an additional nondegeneracy condition. This is established later in Lemma 1 under Assumption 3, which implies strictly positive PF speed.

C.2 Proof of Lemma 1

Before proving Lemma 1, we first clarify the geometric meaning of Assumption 3.

Remark 2. *Under Assumption 2, the active set is fixed for all $w \in [0,1]$ so J_w is continuous with respect to w. Moreover, by KKT, there exists Lagrangian multiplier $\boldsymbol{\nu}_w^*$ such that $\mathrm{Proj}_{\ker(J_w)}\left(w\nabla f_1(\mathbf{x}_w^*) + (1-w)\nabla f_2(\mathbf{x}_w^*) + J_w^\top \boldsymbol{\nu}_w^*\right) = 0$. Here, $\mathrm{Proj}_{\ker(J_w)}$ is a linear operation as $\ker(J_w)$ is a linear subspace. Then, the range of $J_w^\top$ is orthogonal to $\ker(J_w)$ and hence, for $w \in (0,1)$,*

$$\mathrm{Proj}_{\ker(J_w)}\left(\nabla f_1(\mathbf{x}_w^*) - \nabla f_2(\mathbf{x}_w^*)\right) = \frac{1}{1-w}\,\mathrm{Proj}_{\ker(J_w)}\,\nabla f_1(\mathbf{x}_w^*) = -\frac{1}{w}\,\mathrm{Proj}_{\ker(J_w)}\,\nabla f_2(\mathbf{x}_w^*).$$

Therefore, Assumption 3 is equivalent to requiring $\|\,\mathrm{Proj}_{\ker(J_w)}\,\nabla f_1(\mathbf{x}_w^*)\| \geq c(1-w)$, or $\|\,\mathrm{Proj}_{\ker(J_w)}\,\nabla f_2(\mathbf{x}_w^*)\| \geq cw$, with endpoint cases interpreted by one-sided limits.

Geometrically, Assumption 3 rules out points where the two objectives are first-order identical along feasible directions, which would make the PF traversal speed vanish. Such degeneracy is non-generic and typically arises only from special symmetry or gradient alignment. Similar regularity conditions are standard in sensitivity and geometric analyses of PSs/fronts, where Lipschitz-smooth structure is derived from multiplier regularity, rank conditions, or Lipschitz-smooth path-following assumptions [43, 10]. Our condition serves the same purpose, but is tailored to arc-length analysis by directly enforcing a uniformly nonzero tangent trade-off.

We now use this tangent-space separation to prove the uniform lower and upper bounds on the PF traversal speed.

Proof of Lemma 1. Under Assumptions 1, 2, Lemma 2 gives that $w \mapsto \mathbf{x}_w^*$ and $\mathbf{f}_{\mathrm{PF}}(w)$ are differentiable on $w \in [0,1]$. By Assumption 1, the solution path $\{\mathbf{x}_w^* : w \in [0,1]\}$ lies in the compact set $\mathcal{K} \subset \mathrm{relint}(\mathcal{X})$, and $\nabla f_1, \nabla f_2$ are $l_{f,1}$-Lipschitz-continuous on $\mathcal{K}$. Together with the fixed active set and LICQ in Assumption 2, standard sensitivity results imply that $w \mapsto \mathbf{x}_w^*$ is continuously differentiable on $[0,1]$ up to one-sided derivatives at the endpoints. Therefore

$$\frac{\partial}{\partial w}\mathbf{f}_{\mathrm{PF}}(w) = J_f(\mathbf{x}_w^*)\frac{\partial \mathbf{x}_w^*}{\partial w} \tag{35}$$

is continuous on $[0,1]$. Since $[0,1]$ is compact, $v_{\max} := \sup_{w\in[0,1]} \left\|\frac{\partial}{\partial w}\mathbf{f}_{\mathrm{PF}}(w)\right\| < \infty$.

To derive the lower bound, fix any $w \in [0,1]$. Let $\boldsymbol{\nu}_w^*$ denote the optimal Lagrange multipliers of the scalarized problem $\min_{x\in\mathcal{X}} \mathrm{LS}_f(x;w)$, and let $\mathbf{c}$ denote the active constraint functions associated with the locally fixed active index set under Assumption 2. The local KKT system at $\mathbf{x}_w^*$ is

$$\underbrace{w\nabla f_1(\mathbf{x}_w^*) + (1-w)\nabla f_2(\mathbf{x}_w^*)}_{\mathrm{I}(w)} + \underbrace{J_w^\top \boldsymbol{\nu}_w^*}_{\mathrm{II}(w)} = 0, \quad \underbrace{\mathbf{c}(\mathbf{x}_w^*)}_{\mathrm{III}(w)} = 0, \tag{36}$$

where $J_w := \nabla \mathbf{c}(\mathbf{x}_w^*)$ is the Jacobian of the active constraints at $\mathbf{x}_w^*$.

Next, differentiate the stationarity condition in (36). For the first term $\mathrm{I}(w)$, there is

$$\frac{\partial}{\partial w}\mathrm{I}(w) = \nabla f_1(\mathbf{x}_w^*) - \nabla f_2(\mathbf{x}_w^*) + \left(w\nabla^2 f_1(\mathbf{x}_w^*) + (1-w)\nabla^2 f_2(\mathbf{x}_w^*)\right)\frac{\partial}{\partial w}\mathbf{x}_w^*. \tag{37}$$

For the second term $\mathrm{II}(w)$, using $J_w = \nabla \mathbf{c}(\mathbf{x}_w^*)$, there is

$$\frac{\partial}{\partial w}\mathrm{II}(w) = \frac{\partial}{\partial w}\left(\nabla \mathbf{c}(\mathbf{x}_w^*)^\top \boldsymbol{\nu}_w^*\right) = \nabla_{xx}^2\left((\boldsymbol{\nu}_w^*)^\top \mathbf{c}(\mathbf{x}_w^*)\right)\frac{\partial}{\partial w}\mathbf{x}_w^* + J_w^\top \frac{\partial}{\partial w}\boldsymbol{\nu}_w^*. \tag{38}$$

Since the active index set is locally fixed, the constraint map $\mathbf{c}$ itself does not depend on w explicitly, but only through $\mathbf{x}_w^*$. Differentiating the feasibility condition $\mathbf{c}(\mathbf{x}_w^*) = 0$ with respect to w therefore

$$\frac{\partial}{\partial w}\mathrm{III}(w) = J_w \frac{\partial}{\partial w}\mathbf{x}_w^* = 0. \tag{39}$$

Hence, $\frac{\partial}{\partial w}\mathbf{x}_w^* \in \ker(J_w)$. Combining the above yields

$$\mathcal{H}_w \frac{\partial}{\partial w}\mathbf{x}_w^* + (\nabla f_1(\mathbf{x}_w^*) - \nabla f_2(\mathbf{x}_w^*)) + J_w^\top \frac{\partial}{\partial w}\boldsymbol{\nu}_w^* = 0, \quad J_w \frac{\partial}{\partial w}\mathbf{x}_w^* = 0, \tag{40}$$

$$\text{where } \mathcal{H}_w = \nabla_{xx}^2 \mathcal{L}_w(\mathbf{x}_w^*, \boldsymbol{\nu}_w^*) = w\nabla^2 f_1(\mathbf{x}_w^*) + (1-w)\nabla^2 f_2(\mathbf{x}_w^*) + \nabla_{xx}^2\big((\boldsymbol{\nu}_w^*)^\top \mathbf{c}(\mathbf{x}_w^*)\big). \tag{41}$$

Let $\Pi_w \in \mathbb{R}^{d\times d}$ denote the orthogonal projection matrix onto the tangent space $\ker(J_w)$. That is, for every $u \in \mathbb{R}^d$, $\mathrm{Proj}_{\ker(J_w)}(u) = \Pi_w u$. Since $\ker(J_w)$ is a linear subspace, such matrix Π_w exists and is uniquely defined. Moreover, by the nature of projection, $\|\Pi_w u\| \le \|u\|$ for all $u \in \mathbb{R}^d$. Additionally, let $Q_w \in \mathbb{R}^{d\times r_w}$ be a matrix whose columns form an orthonormal basis of the tangent space $\ker(J_w)$, i.e., $\Pi_w = Q_w Q_w^\top$. Here, denote $r_w = \dim(\ker(J_w))$, there is $Q_w^\top Q_w = I_{r_w}$.

Since (40) implies $J_w \frac{\partial}{\partial w}\mathbf{x}_w^* = 0$, we have $\frac{\partial}{\partial w}\mathbf{x}_w^* \in \ker(J_w)$. Therefore, there exists a vector $p_w \in \mathbb{R}^{r_w}$ such that $\frac{\partial}{\partial w}\mathbf{x}_w^* = Q_w p_w$.

Premultiplying the first equation in (40) by $Q_w^\top$ gives

$$Q_w^\top \mathcal{H}_w \frac{\partial}{\partial w}\mathbf{x}_w^* + Q_w^\top(\nabla f_1(\mathbf{x}_w^*) - \nabla f_2(\mathbf{x}_w^*)) + Q_w^\top J_w^\top \frac{\partial}{\partial w}\boldsymbol{\nu}_w^* = 0. \tag{42}$$

Since the columns of Q_w span $\ker(J_w)$, we have $J_w Q_w = 0$, and hence $Q_w^\top J_w^\top = (J_w Q_w)^\top = 0$. Substituting $\frac{\partial}{\partial w}\mathbf{x}_w^* = Q_w p_w$ yields

$$Q_w^\top \mathcal{H}_w Q_w p_w = -Q_w^\top(\nabla f_1(\mathbf{x}_w^*) - \nabla f_2(\mathbf{x}_w^*)). \tag{43}$$

Denote the reduced Hessian $R_w := Q_w^\top \mathcal{H}_w Q_w$. Then $p_w = -R_w^{-1} Q_w^\top(\nabla f_1(\mathbf{x}_w^*) - \nabla f_2(\mathbf{x}_w^*))$, and

$$\frac{\partial}{\partial w}\mathbf{x}_w^* = -Q_w R_w^{-1} Q_w^\top(\nabla f_1(\mathbf{x}_w^*) - \nabla f_2(\mathbf{x}_w^*)). \tag{44}$$

The inverse above is well defined. Indeed, by Assumption 1, both f_1 and f_2 are μ-strongly convex, and by Assumption 2 the active inequality constraints are convex with nonnegative multipliers. Hence the Lagrangian Hessian satisfies $\mathcal{H}_w \succeq \mu I$. Therefore, for every $q \in \mathbb{R}^{r_w}$,

$$q^\top R_w q = q^\top Q_w^\top \mathcal{H}_w Q_w q = (Q_w q)^\top \mathcal{H}_w (Q_w q) \ge \mu \|Q_w q\|^2 = \mu \|q\|^2. \tag{45}$$

So $R_w \succeq \mu I$, and in particular R_w is positive definite and invertible.

Now let $g_w := \nabla f_1(\mathbf{x}_w^*) - \nabla f_2(\mathbf{x}_w^*)$. Then

$$\left|\frac{\partial}{\partial w}\big(f_1(\mathbf{x}_w^*) - f_2(\mathbf{x}_w^*)\big)\right| = \left|g_w^\top \frac{\partial}{\partial w}\mathbf{x}_w^*\right| = \left|-g_w^\top Q_w R_w^{-1} Q_w^\top g_w\right| = (Q_w^\top g_w)^\top R_w^{-1}(Q_w^\top g_w). \tag{46}$$

Since the path lies in the compact set $\mathcal{K}$ and ∇f_m is $l_{f,1}$-Lipschitz-continuous on $\mathcal{K}$, the objective Hessians are uniformly bounded along the path. Together with the fixed active set and Lipschitz-smooth active constraints, there exists $L_H < \infty$ such that $\lambda_{\max}(R_w) \le L_H$ for all $w \in [0,1]$. Hence, $R_w^{-1} \succeq \frac{1}{L_H} I$, and (46) implies

$$\left|\frac{\partial}{\partial w}\big(f_1(\mathbf{x}_w^*) - f_2(\mathbf{x}_w^*)\big)\right| \ge \frac{1}{L_H}\|Q_w^\top g_w\|^2. \tag{47}$$

Finally, since $Q_w Q_w^\top = \Pi_w$ is the orthogonal projection matrix onto $\ker(J_w)$,

$$\|Q_w^\top g_w\|^2 = g_w^\top Q_w Q_w^\top g_w = g_w^\top \Pi_w g_w = \|\Pi_w g_w\|^2 = \|\,\mathrm{Proj}_{\ker(J_w)}(g_w)\|^2. \tag{48}$$

By Assumption 3, (47) gives $\left|\frac{\partial}{\partial w}\big(f_1(\mathbf{x}_w^*) - f_2(\mathbf{x}_w^*)\big)\right| \ge \frac{c^2}{L_H}$. Therefore, using $\sqrt{a^2+b^2} \ge |a-b|/\sqrt{2}$, there is

$$\begin{aligned}\left\|\frac{\partial}{\partial w}\mathbf{f}_{\mathrm{PF}}(w)\right\| =& \sqrt{\left(\frac{\partial}{\partial w} f_1(\mathbf{x}_w^*)\right)^2 + \left(\frac{\partial}{\partial w} f_2(\mathbf{x}_w^*)\right)^2} \\ \ge& \frac{1}{\sqrt{2}}\left|\frac{\partial}{\partial w}\big(f_1(\mathbf{x}_w^*) - f_2(\mathbf{x}_w^*)\big)\right| \ge \frac{c^2}{\sqrt{2}L_H}.\end{aligned} \tag{49}$$

In this way, the invertibility part follows directly from the fundamental theorem of calculus. □

C.3 Extension to the Multi-Objective Setting

Smoothness of the PF. We next note that the Lipschitz-smoothness result extends naturally to the general M-objective setting, where the scalarization weight lies on the simplex Δ_M. The following lemma establishes a Lipschitz-smooth PF map and a relative trade-off inequality along feasible pairwise directions of the simplex.

Lemma 3 (Generalized version of Lemma 2)**.** *Suppose Assumptions 1, 2 hold. Then, for every $\mathbf{w} \in \Delta_M$, the scalarized problem* (3) *has a unique minimizer $\mathbf{x}^*_{\mathbf{w}}$. Moreover, the mappings*

$$\mathbf{w} \mapsto \mathbf{x}^*_{\mathbf{w}}, \quad \mathbf{w} \mapsto \mathbf{f}_{\mathrm{PF}}(\mathbf{w}) = \mathbf{f}(\mathbf{x}^*_{\mathbf{w}}) = [f_{\mathrm{PF}\,1}(\mathbf{w}), \dots, f_{\mathrm{PF}\,M}(\mathbf{w})]^\top \tag{50}$$

are continuously differentiable on Δ_M in the relative sense, with one-sided directional derivatives on the boundary. In addition, $\mathbf{w} \mapsto \mathbf{x}^*_{\mathbf{w}}$ is Lipschitz-continuous, and $\mathbf{f}_{\mathrm{PF}}$ is $l_{\mathbf{f}_{\mathrm{PF}},1}$-Lipschitz-smooth on Δ_M. Additionally, fix any $i, j \in \{1, \dots, M\}$ with $i \neq j$, and any $\mathbf{w} \in \Delta_M$ such that $\mathbf{w} + t(\mathbf{e}_i - \mathbf{e}_j) \in \Delta_M$ for all sufficiently small $t \geq 0$. Then, for all such $t \geq 0$,

$$f_{\mathrm{PF}\,i}\big(\mathbf{w} + t(\mathbf{e}_i - \mathbf{e}_j)\big) - f_{\mathrm{PF}\,i}(\mathbf{w}) \leq f_{\mathrm{PF}\,j}\big(\mathbf{w} + t(\mathbf{e}_i - \mathbf{e}_j)\big) - f_{\mathrm{PF}\,j}(\mathbf{w}). \tag{51}$$

Consequently, $\big[D\mathbf{f}_{\mathrm{PF}}(\mathbf{w})(\mathbf{e}_i - \mathbf{e}_j)\big]_i \leq \big[D\mathbf{f}_{\mathrm{PF}}(\mathbf{w})(\mathbf{e}_i - \mathbf{e}_j)\big]_j$.

Proof. The differentiability of $\mathbf{w} \mapsto \mathbf{x}^*_{\mathbf{w}}$, $\mathbf{w} \mapsto \mathbf{f}_{\mathrm{PF}}(\mathbf{w}) = \mathbf{f}(\mathbf{x}^*_{\mathbf{w}}) = [f_{\mathrm{PF}\,1}(\mathbf{w}), \dots, f_{\mathrm{PF}\,M}(\mathbf{w})]^\top$, and the Lipschitz-smoothness of $\mathbf{f}_{\mathrm{PF}}$ follows the same arguments as in the proof of Lemma 2. It remains to show the directional trade-off inequality. Fix $i \neq j$, $\mathbf{w} \in \Delta_M$, and $t > 0$ such that $\tilde{\mathbf{w}} := \mathbf{w} + t(\mathbf{e}_i - \mathbf{e}_j) \in \Delta_M$. Then $\tilde{w}_i = w_i + t$, $\tilde{w}_j = w_j - t$, and $\tilde{w}_k = w_k$ for $k \neq i, j$. By optimality of $\mathbf{x}^*_{\mathbf{w}}$ for weight $\mathbf{w}$ and optimality of $\mathbf{x}^*_{\tilde{\mathbf{w}}}$ for weight $\tilde{\mathbf{w}}$

$$\sum_{m=1}^{M} w_m f_m(\mathbf{x}^*_{\mathbf{w}}) \leq \sum_{m=1}^{M} w_m f_m(\mathbf{x}^*_{\tilde{\mathbf{w}}}), \quad \sum_{m=1}^{M} \tilde{w}_m f_m(\mathbf{x}^*_{\tilde{\mathbf{w}}}) \leq \sum_{m=1}^{M} \tilde{w}_m f_m(\mathbf{x}^*_{\mathbf{w}}). \tag{52}$$

This yields $t\Big(f_i(\mathbf{x}^*_{\tilde{\mathbf{w}}}) - f_i(\mathbf{x}^*_{\mathbf{w}})\Big) - t\Big(f_j(\mathbf{x}^*_{\tilde{\mathbf{w}}}) - f_j(\mathbf{x}^*_{\mathbf{w}})\Big) \leq 0$ and thus

$$f_{\mathrm{PF}\,i}(\tilde{\mathbf{w}}) - f_{\mathrm{PF}\,i}(\mathbf{w}) \leq f_{\mathrm{PF}\,j}(\tilde{\mathbf{w}}) - f_{\mathrm{PF}\,j}(\mathbf{w}). \tag{53}$$

Finally, dividing by t and letting $t \downarrow 0$ gives $\big[D\mathbf{f}_{\mathrm{PF}}(\mathbf{w})(\mathbf{e}_i - \mathbf{e}_j)\big]_i \leq \big[D\mathbf{f}_{\mathrm{PF}}(\mathbf{w})(\mathbf{e}_i - \mathbf{e}_j)\big]_j$. This completes the proof. □

Lemma 3 shows that the scalarization-induced PF map remains smooth in the multiple-objective case. We next discuss how the bi-objective speed and CDF generalize to a surface-area density on the simplex.

Generalization of v, s, and Φ. In the general M-objective setting, the weight $\mathbf{w}$ lies on the $(M-1)$-dimensional simplex Δ_M, so LS induces an $(M-1)$-dimensional surface on the PF rather than a curve. Let $D\mathbf{f}_{\mathrm{PF}}(\mathbf{w})$ denote the Jacobian of $\mathbf{f}_{\mathrm{PF}}$ with respect to local coordinates on Δ_M. The natural geometric analogue of the bi-objective speed is the Jacobian-induced surface-area density

$$v(\mathbf{w}) := \sqrt{\det\big(D\mathbf{f}_{\mathrm{PF}}(\mathbf{w})^\top D\mathbf{f}_{\mathrm{PF}}(\mathbf{w})\big)}. \tag{54}$$

When $M = 2$, Δ_2 is one-dimensional and (54) reduces to the usual PF traversal speed in (5). For $M > 2$, $v(\mathbf{w})$ plays the role of a surface-area density on the simplex, and its normalization induces uniform surface-area coverage of the PF.

This also gives a natural multivariate analogue of the arc-length CDF. Since $\mathbf{w} \in \Delta_M$ has only $M-1$ degrees of freedom, we use the first $M-1$ coordinates and write $w_M = 1 - \sum_{m=1}^{M-1} w_m$. For $(w_1, \dots, w_{M-1})$ satisfying $w_m \geq 0$ and $\sum_{m=1}^{M-1} w_m \leq 1$, define

$$\Phi(w_1, \dots, w_{M-1}) := \frac{\int_{\{\mathbf{z} \in \Delta_M : 0 \leq z_m \leq w_m,\ m=1,\dots,M-1\}} v(\mathbf{z})\, d\mathbf{z}}{\int_{\Delta_M} v(\mathbf{z})\, d\mathbf{z}}. \tag{55}$$

Equivalently, Φ is the multivariate CDF associated with the surface-area density $v(\mathbf{w})$ under the coordinate chart $(w_1, \dots, w_{M-1})$. Thus, the bi-objective CDF refinement principle extends conceptually to the multi-objective case by replacing arc-length density with surface-area density on the simplex. Developing an efficient multi-dimensional refinement procedure, for example, through multivariate CDF estimation, transport maps, or resampling on Δ_M, is beyond the scope of this paper and left for future work.

C.4 Analytical Expressions for Φ in Quadratic Problems

We next give a simple quadratic example where the scalarization path and the corresponding arc-length CDF admit analytical expressions. This example illustrates the geometric mechanism behind SURF through integral and antiderivative representations.

n-Dimensional Quadratic Bi-Objective Problem. Consider the unconstrained bi-objective problem on $\mathbf{x} \in \mathbb{R}^n$:

$$\mathbf{f}(\mathbf{x}) = \begin{bmatrix} f_1(\mathbf{x}) \\ f_2(\mathbf{x}) \end{bmatrix} = \begin{bmatrix} (\mathbf{x}-\mathbf{b}_1)^\top Q_1 (\mathbf{x}-\mathbf{b}_1) \\ (\mathbf{x}-\mathbf{b}_2)^\top Q_2 (\mathbf{x}-\mathbf{b}_2) \end{bmatrix}, \quad Q_1, Q_2 \succ 0, \quad \mathbf{b}_1 \neq \mathbf{b}_2. \tag{56}$$

For each scalarization weight $w \in [0,1]$, $\mathrm{LS}_f(\mathbf{x}; w) = w f_1(\mathbf{x}) + (1-w) f_2(\mathbf{x})$ is strongly convex and admits the unique minimizer

$$\mathbf{x}_w^* = \big(wQ_1 + (1-w)Q_2\big)^{-1}\big(wQ_1\mathbf{b}_1 + (1-w)Q_2\mathbf{b}_2\big), \tag{57}$$

and the PS is $\mathcal{X}_{\mathrm{PF}} = \{\mathbf{x}_w^* : w \in [0,1]\}$. Let

$$\Sigma(w) := wQ_1 + (1-w)Q_2, \quad P(w) := \Sigma(w)^{-1} w Q_1, \quad \mathbf{d} := \mathbf{b}_1 - \mathbf{b}_2. \tag{58}$$

Then (57) can be written as $\mathbf{x}_w^* = P(w)\mathbf{b}_1 + \big(I - P(w)\big)\mathbf{b}_2 = \mathbf{b}_2 + P(w)\mathbf{d}$. Therefore, $\mathbf{x}_w^* - \mathbf{b}_1 = -\big(I - P(w)\big)\mathbf{d}$ and $\mathbf{x}_w^* - \mathbf{b}_2 = P(w)\mathbf{d}$, and the induced PF manifold is

$$\mathbf{f}_{\mathrm{PF}}(w) := \mathbf{f}(\mathbf{x}_w^*) = \begin{bmatrix} \mathbf{d}^\top \big(I - P(w)\big)^\top Q_1 \big(I - P(w)\big)\mathbf{d} \\ \mathbf{d}^\top P(w)^\top Q_2 P(w)\mathbf{d} \end{bmatrix}, \quad w \in [0,1]. \tag{59}$$

Differentiating $P(w) = \Sigma(w)^{-1} w Q_1$ gives $P'(w) = \Sigma(w)^{-1}\Big(Q_1\big(I - P(w)\big) + Q_2 P(w)\Big)$ and

$$\frac{\partial}{\partial w}\mathbf{f}_{\mathrm{PF}}(w) = \begin{bmatrix} -2\mathbf{d}^\top \big(I - P(w)\big)^\top Q_1 P'(w)\mathbf{d} \\ 2\mathbf{d}^\top P(w)^\top Q_2 P'(w)\mathbf{d} \end{bmatrix}. \tag{60}$$

In this way, the PF traversal speed satisfies

$$\begin{aligned} v(w) &:= \left\| \frac{\partial}{\partial w}\mathbf{f}_{\mathrm{PF}}(w) \right\| = 2 \left\| \begin{bmatrix} -\mathbf{d}^\top \big(I - P(w)\big)^\top Q_1 \Sigma(w)^{-1}\big(Q_1(I - P(w)) + Q_2 P(w)\big)\mathbf{d} \\ \mathbf{d}^\top P(w)^\top Q_2 \Sigma(w)^{-1}\big(Q_1(I - P(w)) + Q_2 P(w)\big)\mathbf{d} \end{bmatrix} \right\| \\ &\propto \sqrt{(m_1(w) + m_3(w))^2 + (m_2(w) + m_3(w))^2}, \end{aligned} \tag{61}$$

where

$$\begin{aligned} m_1(w) &:= \mathbf{d}^\top (I - P(w))^\top Q_1 \Sigma(w)^{-1} Q_1 (I - P(w))\mathbf{d}, \\ m_2(w) &:= \mathbf{d}^\top P(w)^\top Q_2 \Sigma(w)^{-1} Q_2 P(w)\mathbf{d}, \\ m_3(w) &:= \mathbf{d}^\top P(w)^\top Q_2 \Sigma(w)^{-1} Q_1 (I - P(w))\mathbf{d}. \end{aligned}$$

Thus, in general, $v(w)$ is non-constant, so uniform sampling of w does not produce uniform PF coverage. Accordingly,

$$\Phi(w) = \frac{\int_0^w \sqrt{(m_1(u) + m_3(u))^2 + (m_2(u) + m_3(u))^2}\, du}{\int_0^1 \sqrt{(m_1(u) + m_3(u))^2 + (m_2(u) + m_3(u))^2}\, du}. \tag{62}$$

Thus, Rule 1 applies directly in this example. Φ and its inverse can be computed numerically in a straightforward manner, and the PF-aware weights are then given by $w_n = \Phi^{-1}(n/N)$.

One-Dimensional Quadratic Bi-Objective Problem. In the one-dimensional case $x \in \mathbb{R}$, consider

$$\mathbf{f}(x) = \begin{bmatrix} f_1(x) \\ f_2(x) \end{bmatrix} = \begin{bmatrix} q_1 (x - b_1)^2 \\ q_2 (x - b_2)^2 \end{bmatrix}, \qquad q_1, q_2 > 0, \quad b_1 \neq b_2. \tag{63}$$

For each $w \in [0,1]$, the scalarized objective $\mathrm{LS}_f(x; w) = w f_1(x) + (1-w) f_2(x)$ is strongly convex and admits the unique minimizer $x_w^* = \frac{w q_1 b_1 + (1-w) q_2 b_2}{w q_1 + (1-w) q_2}$. Thus, the PS is $\mathcal{X}_{\mathrm{PF}} = \{x_w^* : w \in [0,1]\} = [\min\{b_1, b_2\}, \max\{b_1, b_2\}]$, and the induced PF is

$$\mathbf{f}_{\mathrm{PF}}(w) = \mathbf{f}(x_w^*) = \frac{(b_2 - b_1)^2}{\big(wq_1 + (1-w)q_2\big)^2} \begin{bmatrix} q_1 q_2^2 (1-w)^2 \\ q_2 q_1^2 w^2 \end{bmatrix}. \tag{64}$$

Differentiating yields the PF traversal speed

$$v(w) := \left\| \frac{d}{dw} \mathbf{f}_{\mathrm{PF}}(w) \right\| = \frac{2q_1^2 q_2^2 (b_2 - b_1)^2 \sqrt{(1-w)^2 + w^2}}{\big(wq_1 + (1-w)q_2\big)^3}. \tag{65}$$

Thus, uniformly spaced scalarization weights generally do not yield uniform PF coverage.

Since multiplicative constants cancel after normalization, Φ depends on (q_1, q_2) only through $r := q_2/q_1$. To evaluate the arc-length integral, set $\zeta(w) := \frac{rw-(1-w)}{w+(1-w)r}$. Define $H(\zeta) := \zeta\sqrt{1+\zeta^2} + \operatorname{asinh}(\zeta)$. Since $\zeta(0) = -1/r$, integrating the traversal speed gives

$$s(w) = \int_0^w v(u)\, du = \frac{q_1^2 q_2^2 (b_2 - b_1)^2}{(q_1^2 + q_2^2)^{3/2}} \big[H(\zeta(w)) + H(1/r)\big]. \tag{66}$$

Using $\zeta(1) = r$ and normalizing by $s(1)$ yields $\Phi(w) = \frac{H(\zeta(w))+H(1/r)}{H(r)+H(1/r)}$.

D PF Geometry in Bi-Objective Reinforcement Learning

This appendix specializes the PF-geometry analysis to multi-objective reinforcement learning. We first verify the regularity of the discounted occupancy-measure constraints in Appendix D.1, then derive the PF speed formula used in the main paper in Appendix D.3, and finally analyze the arc-length CDF estimation error in the singleton-state bandit case in Appendix D.4.

D.1 Provable Regularity Condition for Discounted Occupancy Constraints

Lemma 4 (Provable verification of Assumption 2 for RL)**.** *Let $E_{(s,a),s'} = \mathbf{1}\{s' = s\}$, $P_{(s,a),s'} = \mathcal{P}(s' \mid s, a)$, and suppose $\gamma \in (0,1)$. Then $E - \gamma P$ has full column rank. Equivalently, $(E - \gamma P)^\top$ has full row rank.*

Proof. It suffices to show that $\ker(E - \gamma P) = \{0\}$. Let $y \in \mathbb{R}^{|\mathcal{S}|}$ satisfy $(E - \gamma P)y = 0$. Then, for every $(s,a) \in \mathcal{S} \times \mathcal{A}$, $y(s) = \gamma \sum_{s' \in \mathcal{S}} \mathcal{P}(s' \mid s, a) y(s')$. Choose $\bar{s} \in \arg\max_{s \in \mathcal{S}} |y(s)|$. Since $\mathcal{P}(\cdot \mid \bar{s}, a)$ is a probability distribution, for any $a \in \mathcal{A}$, $\|y\|_\infty = |y(\bar{s})| \le \gamma \sum_{s' \in \mathcal{S}} \mathcal{P}(s' \mid \bar{s}, a)|y(s')| \le \gamma \|y\|_\infty$. Because $0 < \gamma < 1$, this implies $\|y\|_\infty = 0$, hence $y = 0$. Therefore $\ker(E - \gamma P) = \{0\}$, so $E - \gamma P$ has full column rank. □

D.2 Inner Optimization Trajectories Stay in the Subset $\mathcal{K}$ in the Relative Interior of $\mathcal{X}$

For the negative conditional-entropy formulation (8), the gradient magnitude diverges as any element of $\mathbf{x}$ goes to 0. Fix some $\mathbf{x}^{(0)} \in \operatorname{relint}(\mathcal{X}) = \{\mathbf{x} \in \mathbb{R}_{++}^{|\mathcal{S}||\mathcal{A}|} : (E - \gamma P)^\top \mathbf{x} = (1-\gamma)\rho\}$ as the initialization. Then, for each $w \in [0,1]$, the sublevel set $\{\mathbf{x} \in \mathcal{X} : \mathrm{LS}_{\mathbf{f}}(\mathbf{x}; w) \le \mathrm{LS}_{\mathbf{f}}(\mathbf{x}^{(0)}; w)\}$ remains strictly bounded away from the boundary of $\mathcal{X}$. Therefore, we can directly adopt

$$\mathcal{K} := \bigcup_{w \in [0,1]} \{\mathbf{x} \in \mathcal{X} : \mathrm{LS}_{\mathbf{f}}(\mathbf{x}; w) \le \mathrm{LS}_{\mathbf{f}}(\mathbf{x}^{(0)}; w)\} \tag{67}$$

Since $[0,1]$ is compact and $\mathrm{LS}_{\mathbf{f}}(\mathbf{x}; w)$ is continuous in $(\mathbf{x}, w)$, it follows that $\mathcal{K}$ is a compact subset of $\operatorname{relint}(\mathcal{X})$. Moreover, for any fixed $w \in [0,1]$, projected gradient descent on the policy or occupancy simplex with stepsize $l_{f,1}^{-1}$ on the μ-strongly convex and locally $l_{f,1}$-smooth objective at $\mathbf{x}^{(t)}$ satisfies

$$\mathrm{LS}_{\mathbf{f}}(\mathbf{x}^{(t+1)}; w) - \mathrm{LS}_{\mathbf{f}}(\mathbf{x}_w^*; w) \le \big(1 - \mu l_{f,1}^{-1}\big)\big(\mathrm{LS}_{\mathbf{f}}(\mathbf{x}^{(t)}; w) - \mathrm{LS}_{\mathbf{f}}(\mathbf{x}_w^*; w)\big), \tag{68}$$

where $0 < 1 - \mu l_{f,1}^{-1} < 1$ [32]. Hence $\mathrm{LS}_{\mathbf{f}}(\mathbf{x}^{(t+1)}; w) \le \mathrm{LS}_{\mathbf{f}}(\mathbf{x}^{(t)}; w) \le \mathrm{LS}_{\mathbf{f}}(\mathbf{x}^{(0)}; w)$, so $\mathbf{x}^{(t)} \in \mathcal{K}$ for all t. By the bijection $h_m(\mathbf{u}) = f_m(g(\mathbf{u}))$ and the equivalence between policy-space and occupancy-space optimization [81], the same sublevel-set invariance also holds for the corresponding policy-space update [32].

For general problems under Assumption 1, fix $\mathbf{x}^{(0)} \in \mathcal{K}$, and we can similarly obtain $\cup_{w \in [0,1]} \{\mathbf{x} \in \mathcal{X} : \mathrm{LS}_{\mathbf{f}}(\mathbf{x}; w) \le \mathrm{LS}_{\mathbf{f}}(\mathbf{x}^{(0)}; w)\} \subseteq \mathcal{K}$. Since $\mathrm{LS}_{\mathbf{f}}(\mathbf{x}; w)$ is continuous in $(\mathbf{x}, w)$ and $[0,1]$ is compact,

this level-set union is closed. Moreover, $\mathrm{LS}_{\mathbf{f}}(\cdot; w)$ is μ-strongly convex for every $w \in [0,1]$, so its sublevel sets are bounded. By compactness of $[0,1]$, the bound is uniform over w. Hence $\cup_{w\in[0,1]}\{\mathbf{x} \in \mathcal{X} : \mathrm{LS}_{\mathbf{f}}(\mathbf{x}; w) \le \mathrm{LS}_{\mathbf{f}}(\mathbf{x}^{(0)}; w)\}$ is compact. Therefore, we can obtain $l_{f,1} = \sup_{\mathbf{x}\in\mathcal{K},\ w\in[0,1]} \|\nabla^2 \mathrm{LS}_{\mathbf{f}}(\mathbf{x}; w)\|_2 < \infty$. Following a similar analysis to (68), we can show that $\mathbf{x}^{(t)}$ stays in the level-set union that belongs to $\mathcal{K}$, where each f_m remains $l_{f,1}$-smooth on $\mathcal{K}$, for all t and all $w \in [0,1]$.

D.3 Proof of Proposition 2: PF Speed in Bi-Objective RL

Proof. The unique minimizer satisfies $\mathbf{x}^*_w \in \Delta^\circ_{S\times A}$ [66], so only the equality constraint is active. Hence the Lagrangian for $\min_{\mathbf{x}\in\mathcal{X}} \mathrm{LS}_f(\mathbf{x}; w)$ is

$$\mathcal{L}(\mathbf{x}, \boldsymbol{\nu}; w) = \mathrm{LS}_f(\mathbf{x}; w) + \left\langle \boldsymbol{\nu}, (E - \gamma P)^\top \mathbf{x} - (1-\gamma)\rho \right\rangle. \tag{69}$$

The KKT conditions are

$$\begin{aligned} 0 =& \nabla_{\boldsymbol{\nu}} \mathcal{L}(\mathbf{x}^*_w, \boldsymbol{\nu}^*_w; w) = (E - \gamma P)^\top \mathbf{x}^*_w - (1-\gamma)\rho, && (70)\\ 0 =& \nabla_{\mathbf{x}} \mathcal{L}(\mathbf{x}^*_w, \boldsymbol{\nu}^*_w; w) \\ =& \underbrace{\beta \log \mathbf{x}^*_w - \beta E \log(E^\top \mathbf{x}^*_w) - \beta R_0}_{=\nabla_{\mathbf{x}} \bar{\tau}(\mathbf{x}^*_w)} - \big(wR_1 + (1-w)R_2\big) + (E - \gamma P)\boldsymbol{\nu}^*_w. && (71) \end{aligned}$$

Since H_w is positive definite on $\ker(J)$ and $J = (E - \gamma P)^\top$ has full row rank by Lemma 4, standard sensitivity results imply that $w \mapsto (\mathbf{x}^*_w, \boldsymbol{\nu}^*_w)$ is Lipschitz-continuous in w [47, Theorem 2.16] and locally continuously differentiable [36, Theorem 2.1].

Differentiating (70)-(71) with respect to w yields

$$J \nabla_w \mathbf{x}^*_w = 0, \quad H_w \nabla_w \mathbf{x}^*_w + J^\top \nabla_w \boldsymbol{\nu}^*_w = R_1 - R_2, \tag{72}$$

where $J = (E - \gamma P)^\top$ and $H_w = \beta\big(\mathrm{Diag}(1/\mathbf{x}^*_w) - E\,\mathrm{Diag}(1/(E^\top \mathbf{x}^*_w))E^\top\big)$. Although H_w is singular on the full space, it is positive definite on the feasible tangent space $\ker(J)$; we therefore use the augmented Hessian $\widetilde{H}_w := H_w + cJ^\top J$ with any fixed $c > 0$.

For any z, the conditional-entropy Hessian is positive semidefinite:

$$z^\top H_w z = \beta \sum_s \left(\sum_a \frac{z(s,a)^2}{\mathbf{x}^*_w(s,a)} - \frac{\left(\sum_a z(s,a)\right)^2}{\sum_a \mathbf{x}^*_w(s,a)} \right) \ge 0. \tag{73}$$

Equality in the Cauchy-Schwarz bound requires $z(s,a) = \lambda_s \mathbf{x}^*_w(s,a)$ for all a, i.e., only a state-marginal perturbation; by sufficient exploration and the bijective policy-occupancy correspondence, no nonzero such direction lies in $\ker(J)$, so H_w is positive definite on $\ker(J)$. For every $z \neq 0$,

$$z^\top \widetilde{H}_w z = z^\top H_w z + c\|Jz\|^2 > 0, \tag{74}$$

where the case $Jz \neq 0$ follows from $c\|Jz\|^2 > 0$, and the case $Jz = 0$ follows from positive definiteness of H_w on $\ker(J)$.

Thus $\widetilde{H}_w \succ 0$, and since J has full row rank, $\widetilde{S}_w := J\widetilde{H}_w^{-1} J^\top$ is positive definite and invertible.

Because $J\nabla_w \mathbf{x}^*_w = 0$, replacing H_w by $\widetilde{H}_w$ does not change the primal sensitivity equation:

$$\begin{aligned} \widetilde{H}_w \nabla_w \mathbf{x}^*_w + J^\top \nabla_w \boldsymbol{\nu}^*_w =& H_w \nabla_w \mathbf{x}^*_w + cJ^\top J \nabla_w \mathbf{x}^*_w + J^\top \nabla_w \boldsymbol{\nu}^*_w \\ =& H_w \nabla_w \mathbf{x}^*_w + J^\top \nabla_w \boldsymbol{\nu}^*_w \\ =& R_1 - R_2. \end{aligned} \tag{75}$$

Therefore, the same $\nabla_w \mathbf{x}^*_w$ solves the augmented KKT system

$$\begin{bmatrix} \widetilde{H}_w & J^\top \\ J & 0 \end{bmatrix} \begin{bmatrix} \nabla_w \mathbf{x}^*_w \\ \nabla_w \boldsymbol{\nu}^*_w \end{bmatrix} = \begin{bmatrix} R_1 - R_2 \\ 0 \end{bmatrix}. \tag{76}$$

Since $\widetilde{H}_w \succ 0$ and $\widetilde{S}_w = J\widetilde{H}_w^{-1}J^\top \succ 0$, the standard Schur-complement inverse formula gives

$$\begin{bmatrix} \widetilde{H}_w & J^\top \\ J & 0 \end{bmatrix}^{-1} = \begin{bmatrix} \widetilde{H}_w^{-1} - \widetilde{H}_w^{-1}J^\top \widetilde{S}_w^{-1} J\widetilde{H}_w^{-1} & \widetilde{H}_w^{-1}J^\top \widetilde{S}_w^{-1} \\ \widetilde{S}_w^{-1}J\widetilde{H}_w^{-1} & -\widetilde{S}_w^{-1} \end{bmatrix}, \tag{77}$$

$$\Rightarrow \quad \nabla_w \mathbf{x}_w^* = \left(\widetilde{H}_w^{-1} - \widetilde{H}_w^{-1}J^\top \widetilde{S}_w^{-1} J\widetilde{H}_w^{-1}\right)(R_1 - R_2). \tag{78}$$

Next, for the bi-objective PF map $\mathbf{f}_{\mathrm{PF}}(w) = [f_{\mathrm{PF}1}(w), f_{\mathrm{PF}2}(w)]^\top$, we have

$$\begin{aligned} \frac{\partial}{\partial w} f_{\mathrm{PF}1}(w) &= \langle \nabla_{\mathbf{x}} f_1(\mathbf{x}_w^*), \nabla_w \mathbf{x}_w^* \rangle = \langle \nabla_{\mathbf{x}} \bar{\tau}(\mathbf{x}_w^*) - R_1, \nabla_w \mathbf{x}_w^* \rangle \\ &= \langle (1-w)(R_2 - R_1) - (E - \gamma P)\boldsymbol{\nu}_w^*, \nabla_w \mathbf{x}_w^* \rangle \\ &= (1-w)\left\langle R_2 - R_1, \nabla_w \mathbf{x}_w^* \right\rangle, \end{aligned} \tag{79}$$

where the third equality uses (71), and the last uses (72). Similar analysis for $\frac{\partial}{\partial w} f_{\mathrm{PF}2}(w)$ gives

$$\begin{aligned} v(w) &= \left\| \frac{\partial}{\partial w} \mathbf{f}_{\mathrm{PF}}(w) \right\| = \sqrt{\left(\frac{\partial}{\partial w} f_{\mathrm{PF}1}(w)\right)^2 + \left(\frac{\partial}{\partial w} f_{\mathrm{PF}2}(w)\right)^2} \\ &= \sqrt{(1-w)^2 + w^2}\left|(R_1 - R_2)^\top \nabla_w \mathbf{x}_w^*\right|. \end{aligned} \tag{80}$$

Substituting (78) yields the PF speed for bi-objective RL.

Now consider the case $|\mathcal{S}| = 1$. Then the discounted occupancy measure coincides with the action distribution, so the feasible set reduces to $\Delta_A = \{\mathbf{x} \in \mathbb{R}_+^A : \langle \mathbf{1}, \mathbf{x} \rangle = 1\}$. Since $\mathbf{x} \mapsto \beta\langle \mathbf{x}, \log \mathbf{x} \rangle$ is strictly convex on Δ_A° and the remaining terms are linear, $\mathrm{LS}_f(\cdot; w)$ is strictly convex on Δ_A and has a unique minimizer. The Lagrangian is

$$\mathcal{L}(\mathbf{x}, \nu; w) = \mathrm{LS}_f(\mathbf{x}; w) + \nu(\langle \mathbf{1}, \mathbf{x} \rangle - 1).$$

Stationarity gives $\beta \log \mathbf{x} - \beta R_0 - \big(wR_1 + (1-w)R_2\big) + \nu \mathbf{1} = 0$, so $\mathbf{x}$ is proportional to $\exp\big(R_0 + \beta^{-1}\big(wR_1 + (1-w)R_2\big)\big)$. Enforcing $\langle \mathbf{1}, \mathbf{x} \rangle = 1$ yields

$$\mathbf{u}_w^* = \mathbf{x}_w^* = \mathrm{SoftMax}\left(R_0 + \beta^{-1}\big(wR_1 + (1-w)R_2\big)\right) \in \Delta_A^\circ. \tag{81}$$

Differentiating $\mathbf{x}_w^* = \mathrm{SoftMax}(z(w))$ with $z(w) = R_0 + \beta^{-1}\big(wR_1 + (1-w)R_2\big)$ gives $z'(w) = \beta^{-1}(R_1 - R_2)$. Using the standard SoftMax Jacobian $D\,\mathrm{SoftMax}(z) = \mathrm{Diag}(\mathbf{x}_w^*) - \mathbf{x}_w^*(\mathbf{x}_w^*)^\top$,

$$\frac{\partial}{\partial w}\mathbf{x}_w^* = \beta^{-1}\Big(\mathrm{Diag}(\mathbf{x}_w^*) - \mathbf{x}_w^*(\mathbf{x}_w^*)^\top\Big)(R_1 - R_2). \tag{82}$$

Substituting this into $v(w) = \sqrt{(1-w)^2 + w^2}\left|(R_1 - R_2)^\top \frac{\partial}{\partial w}\mathbf{x}_w^*\right|$ and using symmetry of $\mathrm{Diag}(\mathbf{x}_w^*) - \mathbf{x}_w^*(\mathbf{x}_w^*)^\top$ gives (9). □

D.4 CDF Estimation Error in Bandit Problems

Proposition 3 (Arc-length CDF estimation error in bandits)**.** *Consider MDP with singleton state space $|\mathcal{S}| = 1$ and $|\mathcal{A}| = A \geq 2$ actions. Suppose the bounded speed condition in Lemma 1 holds. Given a budget of T pulls, suppose that each pull of arm a returns a bounded bi-objective reward vector $[r_1(a), r_2(a)]$ with $\mathbb{E}[r_m(a)] = R_m(a)$ for $m \in \{1, 2\}$. We estimate $R_m(a)$ by the empirical mean $\hat{R}_m(a)$. Then, with probability at least $1 - \delta$, there is*

$$\|\hat{\Phi} - \Phi\|_\infty = \mathcal{O}\left(s(1)^{-1}\|\hat{R}_m - R_m\|_\infty\right) = \mathcal{O}\left(s(1)^{-1}\sqrt{A\log(A/\delta)T^{-1}}\right). \tag{83}$$

Proof. We first control the reward estimation error. Under a total budget of T pulls, suppose each arm is sampled at least $\lfloor T/A \rfloor$ times. Since each reward component lies in $[-R_{\max}, R_{\max}]$, Hoeffding's inequality yields that for any $\epsilon > 0$ and each $m \in \{1, 2\}$, $\Pr\left(\max_{a \in [A]} |\hat{R}_m(a) - R_m(a)| \geq \epsilon\right) \leq 2A\exp\left(-\frac{\lfloor T/A \rfloor \epsilon^2}{2R_{\max}^2}\right)$. Hence, by choosing $\epsilon_R := 2R_{\max}\sqrt{\frac{A\log(4A/\delta)}{2T}}$, we obtain, with probability at least $1 - \delta$,

$$\|\hat{R}_m - R_m\|_\infty \leq \epsilon_R, \quad m \in \{1, 2\}. \tag{84}$$

We now bound $\|\hat{\Phi} - \Phi\|_\infty$. For any $w \in [0,1]$,

$$\left|\hat{\Phi}(w) - \Phi(w)\right| = \left|\frac{\hat{s}(w)}{\hat{s}(1)} - \frac{s(w)}{s(1)}\right| \le \frac{|\hat{s}(w) - s(w)|}{s(1)} + \frac{s(w)}{s(1)}\frac{|\hat{s}(1) - s(1)|}{\hat{s}(1)}.$$

Since $0 \le s(w) \le s(1)$, we have

$$|\hat{s}(w) - s(w)| \le \int_0^w |\hat{v}_u(u) - v_u(u)|du \le \sup_{u\in[0,1]} |\hat{v}_u(u) - v_u(u)|,$$

$$|\hat{s}(1) - s(1)| \le \int_0^1 |\hat{v}_u(u) - v_u(u)|du \le \sup_{u\in[0,1]} |\hat{v}_u(u) - v_u(u)|.$$

Therefore,

$$\left|\hat{\Phi}(w) - \Phi(w)\right| \le \frac{1}{s(1)} \sup_{u\in[0,1]} |\hat{v}_u(u) - v_u(u)| + \frac{1}{\hat{s}(1)} \sup_{u\in[0,1]} |\hat{v}_u(u) - v_u(u)|. \tag{85}$$

It remains to control the speed perturbation. Since $|R_m(a)| \le R_{\max}$ for all $a \in [A]$ and $m \in \{1,2\}$, the logits $\beta^{-1}(wR_1 + (1-w)R_2)$ remain uniformly bounded in ℓ_∞ for all $w \in [0,1]$. Hence the softmax vector $\mathbf{x}_w^*$ stays in a compact subset of Δ_A°, uniformly over $w \in [0,1]$, and the map $(R_1, R_2) \mapsto v(w)$ is uniformly Lipschitz-continuous over $w \in [0,1]$. Therefore, there exists a constant $L_R < \infty$, depending only on $(A, \beta, R_{\max})$, such that

$$\sup_{w\in[0,1]} |\hat{v}(w) - v(w)| \le L_R\big(\|\hat{R}_1 - R_1\|_\infty + \|\hat{R}_2 - R_2\|_\infty\big). \tag{86}$$

On $\left\{\sup_{u\in[0,1]} |\hat{v}_u(u) - v_u(u)| \le \frac{s(1)}{2}\right\}$, we have $\hat{s}(1) \ge s(1) - |\hat{s}(1) - s(1)| \ge s(1) - \sup_{u\in[0,1]} |\hat{v}_u(u) - v_u(u)| \ge \frac{s(1)}{2}$. Substituting this into (85) gives, for all $w \in [0,1]$, $\left|\hat{\Phi}(w) - \Phi(w)\right| \le \frac{1}{s(1)} \sup_u |\hat{v}_u(u) - v_u(u)| + \frac{2}{s(1)} \sup_u |\hat{v}_u(u) - v_u(u)| = \frac{3}{s(1)} \sup_{u\in[0,1]} |\hat{v}_u(u) - v_u(u)|$. Hence, on $\left\{\sup_{u\in[0,1]} |\hat{v}_u(u) - v_u(u)| \le \frac{s(1)}{2}\right\}$,

$$\|\hat{\Phi} - \Phi\|_\infty \le \frac{3}{s(1)} \sup_{u\in[0,1]} |\hat{v}_u(u) - v_u(u)|. \tag{87}$$

Moreover, by (86), the event $\left\{\sup_{u\in[0,1]} |\hat{v}_u(u) - v_u(u)| \le \frac{s(1)}{2}\right\}$ is implied by $L_R\big(\|\hat{R}_1 - R_1\|_\infty + \|\hat{R}_2 - R_2\|_\infty\big) \le \frac{s(1)}{2}$. In particular, if $2L_R\epsilon_R \le \frac{s(1)}{2}$, then (84) implies $\left\{\sup_{u\in[0,1]} |\hat{v}_u(u) - v_u(u)| \le \frac{s(1)}{2}\right\}$. Equivalently, it suffices that $\epsilon_R \le \frac{s(1)}{4L_R}$, which holds whenever T is large enough, namely $T \gtrsim \frac{AR_{\max}^2 L_R^2}{s(1)^2} \log \frac{A}{\delta}$.

Finally, on the event (84), combining (87) and (86) yields $\|\hat{\Phi} - \Phi\|_\infty \le \frac{3L_R}{s(1)}\big(\|\hat{R}_1 - R_1\|_\infty + \|\hat{R}_2 - R_2\|_\infty\big) \le \frac{6L_R}{s(1)}\epsilon_R$. Therefore, with probability at least $1-\delta$, $\|\hat{\Phi} - \Phi\|_\infty = O\left(s(1)^{-1}\sqrt{\frac{A\log(A/\delta)}{T}}\right)$, where the hidden constant depends only on $(A, \beta, R_{\max})$ through L_R. This completes the proof. □

E Convergence Analysis of SURF

Building on the geometric results summarized in Figure 7, this appendix proves the convergence guarantees for SURF. We first analyze the exact-InnerSolver setting and then extend the argument to the practical warm-started finite-step InnerSolver used in Algorithm 1.

Algorithm 1 can be implemented in the parameter space $\mathcal{U}$ and returns iterates $\mathbf{u}_n^{(t)}$, or in the hidden space $\mathcal{X}$ to return $\mathbf{x}_n^{(t)}$. Throughout this section, we work with their images in the convex hidden space $\mathbf{x}_n^{(t)} := g(\mathbf{u}_n^{(t)}), \mathbf{x}_w^* := g(\mathbf{u}_w^*)$ under Assumption 1. Since $h_m(\mathbf{u}) = f_m(g(\mathbf{u}))$, we equivalently have $\mathbf{h}(\mathbf{u}_n^{(t)}) = \mathbf{f}(\mathbf{x}_n^{(t)}), \mathbf{h}(\mathbf{u}_w^*) = \mathbf{f}(\mathbf{x}_w^*)$. Hence, all approximation and chord-length bounds below are stated in the $\mathbf{x}$-space.

E.1 Proof of the First Part of Theorem 1: Discretization Error under Exact InnerSolver

In this subsection, we first consider the exact-InnerSolver setting. We begin by establishing Lipschitz-smoothness properties of $\tilde{\Phi}_t$ and Φ_t through Lemmas 6-8. We then use these properties to control the discretization error in Lemma 10, which completes the proof of the first part of Theorem 1.

Smoothness of Φ_t under Exact InnerSolver Solutions. To prove the Lipschitz-smoothness of Φ_t, the key issue is that the update in Algorithm 1 depends on an interpolated CDF constructed from finitely many sampled Pareto points. Thus, to control the outer-loop dynamics, we must show that this interpolated CDF remains monotone and sufficiently Lipschitz-smooth. To do so, we first consider the exact InnerSolver solutions. In this case, $\tilde{s}^{(t)}$ and $\tilde{\Phi}_t$ are built from exact sampled solutions $\mathbf{u}_n^{(t)} = \mathbf{u}^*_{w_n^{(t)}}$ or equivalently $\mathbf{x}_n^{(t)} = \mathbf{x}^*_{w_n^{(t)}}$.

We establish this in three steps. First, we show that the rebuilt CDF increases at a controlled rate across the sampled weights, and that these local growth rates do not change too abruptly between neighboring intervals in Lemma 6. Next, we show that the PCHIP interpolation rule lifts these discrete controls into global derivative and Lipschitz-smoothness bounds for the reconstructed CDF in Lemma 7. Finally, we prove by induction that these properties are preserved by the SURF outer update throughout all iterations in Lemma 8.

We begin with the discrete CDF values at the sampled weights. Before turning to the rebuilt CDF itself, we first present the following geometric lemma, which relates PF arc-length increments to chord lengths along the PF.

Lemma 5 (Arc-chord bound)**.** *Suppose Assumptions 1, 2 hold. Moreover, for any $0 \le w_1 < w_2 \le 1$, there is $s(w_2) - s(w_1) \le \sqrt{2}\,\|\mathbf{f}_{\mathrm{PF}}(w_2) - \mathbf{f}_{\mathrm{PF}}(w_1)\|$.*

Proof. Denote $\tilde{\mathbf{f}}_{\mathrm{PF}}(s) := \mathbf{f}_{\mathrm{PF}}(s^{-1}(s))$ for $s \in [0, s(1)]$ and write $\mathbf{f}_{\mathrm{PF}s}(s) = \begin{bmatrix} f_{\mathrm{PF}s,1}(s) \\ f_{\mathrm{PF}s,2}(s) \end{bmatrix}$.

Under Assumptions 1 and 2, Lemma 2 shows that $f_{\mathrm{PF}s,1}$ is monotone nonincreasing and $f_{\mathrm{PF}s,2}$ is monotone nondecreasing. Thus, along the front, the first coordinate is nonincreasing and the second coordinate is nondecreasing as w increases. Since $s(\cdot)$ is strictly increasing, the same monotonicity carries over to the PF parametrization. Therefore, on the interval $[s(w_1), s(w_2)]$,

$$\int_{s(w_1)}^{s(w_2)} \left|f_{\mathrm{PF}s,1}'(s)\right| ds = \left|f_{\mathrm{PF}s,1}(s(w_2)) - f_{\mathrm{PF}s,1}(s(w_1))\right|, \tag{88}$$

$$\int_{s(w_1)}^{s(w_2)} \left|f_{\mathrm{PF}s,2}'(s)\right| ds = \left|f_{\mathrm{PF}s,2}(s(w_2)) - f_{\mathrm{PF}s,2}(s(w_1))\right|. \tag{89}$$

Using $\|u\| \le \|u\|_1$ and then $\|u\|_1 \le \sqrt{2}\|u\|$ for all $u \in \mathbb{R}^2$, we obtain

$$\begin{aligned}
s(w_2) - s(w_1) &= \int_{s(w_1)}^{s(w_2)} \|\mathbf{f}'_{\mathrm{PF},s}(s)\| ds \\
&\le \int_{s(w_1)}^{s(w_2)} \left(|f_{\mathrm{PF}s,1}'(s)| + |f_{\mathrm{PF}s,2}'(s)|\right) ds \\
&= \left|f_{\mathrm{PF}s,1}(s(w_2)) - f_{\mathrm{PF}s,1}(s(w_1))\right| + \left|f_{\mathrm{PF}s,2}(s(w_2)) - f_{\mathrm{PF}s,2}(s(w_1))\right| \\
&= \|\mathbf{f}_{\mathrm{PF},s}(s(w_2)) - \mathbf{f}_{\mathrm{PF},s}(s(w_1))\|_1 \\
&\le \sqrt{2}\,\|\mathbf{f}_{\mathrm{PF}s}(s(w_2)) - \mathbf{f}_{\mathrm{PF}s}(s(w_1))\|\,.
\end{aligned}$$

Finally, since $\mathbf{f}_{\mathrm{PF}s}(s(w_i)) = \mathbf{f}_{\mathrm{PF}}(w_i)$ for $i \in \{1, 2\}$, we conclude the proof. □

This geometric relation ensures that the cumulative chord-length construction provides a faithful approximation to the true arc-length growth along the front. In particular, it allows us to control how fast the rebuilt CDF increases across sampled weights, and how much this local growth can vary from one interval to the next. We formalize this in the next lemma.

Lemma 6 (Discrete growth control of the rebuilt CDF)**.** *Suppose Assumptions 1, 2, 3 hold. For any sampled weights* $\{w_n^{(t)}\}_{n=0}^N$*, let* $\tilde{\Phi}_t(w_n^{(t)}) = \frac{\tilde{s}^{(t)}(w_n^{(t)})}{\tilde{s}^{(t)}(1)}$ *where* $\tilde{s}^{(t)}(w_n^{(t)})$ *is generated via* (12) *with exact* INNERSOLVER *solutions. Then:*

(i) for every $n \in \{0, \cdots, N-1\}$*,*

$$\frac{1}{\sqrt{2}}\frac{v_{\min}}{v_{\max}} \le \frac{\tilde{\Phi}_t(w_{n+1}^{(t)}) - \tilde{\Phi}_t(w_n^{(t)})}{w_{n+1}^{(t)} - w_n^{(t)}} \le \sqrt{2}\frac{v_{\max}}{v_{\min}}; \tag{90}$$

(ii) for every $n \in \{0, \cdots, N-2\}$*, there exists* $l_{\tilde{\Phi}_t,n,1} \le \frac{1}{\tilde{s}^{(t)}(1)}\Big(l_{v,0} + \frac{1}{2}l_{\mathbf{f}_{\mathrm{PF}},1}\Big)$ *such that*

$$\left|\frac{\tilde{\Phi}_t(w_{n+2}^{(t)}) - \tilde{\Phi}_t(w_{n+1}^{(t)})}{w_{n+2}^{(t)} - w_{n+1}^{(t)}} - \frac{\tilde{\Phi}_t(w_{n+1}^{(t)}) - \tilde{\Phi}_t(w_n^{(t)})}{w_{n+1}^{(t)} - w_n^{(t)}}\right| \le l_{\tilde{\Phi}_t,n,1}\big(w_{n+2}^{(t)} - w_n^{(t)}\big). \tag{91}$$

Proof. Since $\tilde{\Phi}_t = \tilde{s}^{(t)}/\tilde{s}^{(t)}(1)$, we have

$$\frac{\tilde{\Phi}_t(w_{n+1}^{(t)}) - \tilde{\Phi}_t(w_n^{(t)})}{w_{n+1}^{(t)} - w_n^{(t)}} = \frac{1}{\tilde{s}^{(t)}(1)} \cdot \frac{\|\mathbf{f}_{\mathrm{PF}}(w_{n+1}^{(t)}) - \mathbf{f}_{\mathrm{PF}}(w_n^{(t)})\|}{w_{n+1}^{(t)} - w_n^{(t)}}. \tag{92}$$

Thus it suffices to study the chord secant slopes of $\mathbf{f}_{\mathrm{PF}}$.

We first prove (90). For any $0 \le a < b \le 1$, Lemma 5 gives

$$\|\mathbf{f}_{\mathrm{PF}}(b) - \mathbf{f}_{\mathrm{PF}}(a)\| \ge \frac{1}{\sqrt{2}}\big(s(b) - s(a)\big). \tag{93}$$

Using the bounded speed condition in Lemma 1, we obtain

$$\|\mathbf{f}_{\mathrm{PF}}(w_{n+1}^{(t)}) - \mathbf{f}_{\mathrm{PF}}(w_n^{(t)})\| \ge \frac{1}{\sqrt{2}}\int_{w_n^{(t)}}^{w_{n+1}^{(t)}} \|\mathbf{f}_{\mathrm{PF}}{}'(w)\|dw \ge \frac{1}{\sqrt{2}}v_{\min}(w_{n+1}^{(t)} - w_n^{(t)}), \tag{94}$$

$$\|\mathbf{f}_{\mathrm{PF}}(w_{n+1}^{(t)}) - \mathbf{f}_{\mathrm{PF}}(w_n^{(t)})\| \le \int_{w_n^{(t)}}^{w_{n+1}^{(t)}} \|\mathbf{f}_{\mathrm{PF}}{}'(w)\|dw \le v_{\max}(w_{n+1}^{(t)} - w_n^{(t)}). \tag{95}$$

Hence, on each interval $[w_n^{(t)}, w_{n+1}^{(t)}]$, the secant slope of $\tilde{s}^{(t)}$ lies in $\Big[\frac{1}{\sqrt{2}}v_{\min}, v_{\max}\Big]$.

Applying Lemma 5 to each sampled segment gives $\tilde{s}^{(t)}(1) = \sum_{n=0}^{N-1}\|\mathbf{f}_{\mathrm{PF}}(w_{n+1}^{(t)}) - \mathbf{f}_{\mathrm{PF}}(w_n^{(t)})\| \in \Big[\frac{1}{\sqrt{2}}s(1), s(1)\Big]$, while $s(1) = \int_0^1 v(u)du \in [v_{\min}, v_{\max}]$. Therefore the secant slope of $\tilde{\Phi}_t$ on each interval $[w_n^{(t)}, w_{n+1}^{(t)}]$ lies in $\left[\frac{\frac{1}{\sqrt{2}}v_{\min}}{v_{\max}}, \frac{v_{\max}}{\frac{1}{\sqrt{2}}v_{\min}}\right] = \Big[\frac{1}{\sqrt{2}}\frac{v_{\min}}{v_{\max}}, \sqrt{2}\frac{v_{\max}}{v_{\min}}\Big]$, which proves (90).

Next we prove (91). By the integral mean value theorem, there exist $\xi_n \in (w_n^{(t)}, w_{n+1}^{(t)}), \quad \xi_{n+1} \in (w_{n+1}^{(t)}, w_{n+2}^{(t)})$ such that

$$\frac{1}{w_{n+1}^{(t)} - w_n^{(t)}}\int_{w_n^{(t)}}^{w_{n+1}^{(t)}} v(u)du = v(\xi_n). \tag{96}$$

Hence, by Lemma 2,

$$\begin{aligned} 0 \le &\left|\frac{1}{w_{n+2}^{(t)} - w_{n+1}^{(t)}}\int_{w_{n+1}^{(t)}}^{w_{n+2}^{(t)}} v(u)du - \frac{1}{w_{n+1}^{(t)} - w_n^{(t)}}\int_{w_n^{(t)}}^{w_{n+1}^{(t)}} v(u)du\right| \\ =&|v(\xi_{n+1}) - v(\xi_n)| \le l_{v,0}|\xi_{n+1} - \xi_n| \le l_{v,0}\big(w_{n+2}^{(t)} - w_n^{(t)}\big). \end{aligned} \tag{97}$$

Next, we bound the chord-arc gap on each interval using Lipschitzness of $\mathbf{f}_{\mathrm{PF}}{}'$. Using $|\|x\| - \|y\|| \le \|x - y\|$ and Jensen/triangle inequalities,

$$\begin{aligned} 0 \le & \frac{1}{w_{n+1}^{(t)} - w_n^{(t)}} \int_{w_n^{(t)}}^{w_{n+1}^{(t)}} \|\mathbf{f}_{\mathrm{PF}}{}'(u)\| du - \left\| \frac{1}{w_{n+1}^{(t)} - w_n^{(t)}} \int_{w_n^{(t)}}^{w_{n+1}^{(t)}} \mathbf{f}_{\mathrm{PF}}{}'(u) du \right\| \\ \le & \frac{1}{w_{n+1}^{(t)} - w_n^{(t)}} \int_{w_n^{(t)}}^{w_{n+1}^{(t)}} \left\| \mathbf{f}_{\mathrm{PF}}{}'(u) - \frac{1}{w_{n+1}^{(t)} - w_n^{(t)}} \int_{w_n^{(t)}}^{w_{n+1}^{(t)}} \mathbf{f}_{\mathrm{PF}}{}'(r) dr \right\| du. \end{aligned} \tag{98}$$

For any fixed $u \in [w_n^{(t)}, w_{n+1}^{(t)}]$, by Lemma 2 ,

$$\begin{aligned} 0 \le & \left\| \mathbf{f}_{\mathrm{PF}}{}'(u) - \frac{1}{w_{n+1}^{(t)} - w_n^{(t)}} \int_{w_n^{(t)}}^{w_{n+1}^{(t)}} \mathbf{f}_{\mathrm{PF}}{}'(r) dr \right\| \\ \le & \frac{1}{w_{n+1}^{(t)} - w_n^{(t)}} \int_{w_n^{(t)}}^{w_{n+1}^{(t)}} \|\mathbf{f}_{\mathrm{PF}}{}'(u) - \mathbf{f}_{\mathrm{PF}}{}'(r)\| dr \\ \le & \frac{1}{w_{n+1}^{(t)} - w_n^{(t)}} \int_{w_n^{(t)}}^{w_{n+1}^{(t)}} l_{\mathbf{f}_{\mathrm{PF}},1} |u - r| dr \le \frac{l_{\mathbf{f}_{\mathrm{PF}},1}}{2} (w_{n+1}^{(t)} - w_n^{(t)}). \end{aligned} \tag{99}$$

Averaging over u and combining with (98) yields

$$\left| \frac{\|\mathbf{f}_{\mathrm{PF}}(w_{n+1}^{(t)}) - \mathbf{f}_{\mathrm{PF}}(w_n^{(t)})\|}{w_{n+1}^{(t)} - w_n^{(t)}} - \frac{1}{w_{n+1}^{(t)} - w_n^{(t)}} \int_{w_n^{(t)}}^{w_{n+1}^{(t)}} v(u) du \right| \le \frac{l_{\mathbf{f}_{\mathrm{PF}},1}}{2} (w_{n+1}^{(t)} - w_n^{(t)}). \tag{100}$$

Combining (97) and (100) on the two consecutive intervals $[w_n^{(t)}, w_{n+1}^{(t)}]$ and $[w_{n+1}^{(t)}, w_{n+2}^{(t)}]$ by the triangle inequality gives

$$\begin{aligned} 0 \le & \left| \frac{\|\mathbf{f}_{\mathrm{PF}}(w_{n+2}^{(t)}) - \mathbf{f}_{\mathrm{PF}}(w_{n+1}^{(t)})\|}{w_{n+2}^{(t)} - w_{n+1}^{(t)}} - \frac{\|\mathbf{f}_{\mathrm{PF}}(w_{n+1}^{(t)}) - \mathbf{f}_{\mathrm{PF}}(w_n^{(t)})\|}{w_{n+1}^{(t)} - w_n^{(t)}} \right| \\ \le & l_{v,0} \big(w_{n+2}^{(t)} - w_n^{(t)}\big) + \frac{l_{\mathbf{f}_{\mathrm{PF}},1}}{2} \Big((w_{n+2}^{(t)} - w_{n+1}^{(t)}) + (w_{n+1}^{(t)} - w_n^{(t)}) \Big) \\ = & \Big(l_{v,0} + \frac{1}{2} l_{\mathbf{f}_{\mathrm{PF}},1} \Big) \big(w_{n+2}^{(t)} - w_n^{(t)}\big). \end{aligned} \tag{101}$$

Multiplying (101) by $1/\tilde{s}^{(t)}(1)$ and using (92) yields (91). □

Lemma 6 provides the two discrete properties needed for the interpolation analysis. The bound in (90) shows that the rebuilt CDF increases at a controlled rate at the sampled weights, so it is neither too flat nor too steep. The bound in (91) further shows that these local growth rates do not change too abruptly between neighboring intervals. Together, these properties ensure that the PCHIP reconstruction preserves a stable global shape and admits uniform derivative and Lipschitz-smoothness bounds.

Lemma 7 (Derivative bounds for the PCHIP rebuild)**.** *Suppose Assumptions 1, 2, 3 hold. Consider a CDF Φ_t satisfying $|\Phi_t'| \in \left[\frac{1}{2\sqrt{2}} \frac{v_{\min}}{v_{\max}}, \frac{5}{2\sqrt{2}} \frac{v_{\max}}{v_{\min}} \right]$. Let $\{w_n^{(t)}\}_{n=0}^N$ be the sampled weights generated via $w_n^{(t)} = \Phi_t^{-1}(n/N)$, let $\tilde{s}^{(t)}(w_n^{(t)})$ be generated via* (12) *with exact InnerSolver solutions, let $\tilde{s}^{(t)}$ be the PCHIP interpolant built via Algorithm 2, and define $\tilde{\Phi}_t(w) = \frac{\tilde{s}^{(t)}(w)}{\tilde{s}^{(t)}(1)}$. Denote some $l_{\tilde{\Phi}_t,n,1} \le \frac{1}{\tilde{s}^{(t)}(1)} \Big(l_{v,0} + \frac{1}{2} l_{\mathbf{f}_{\mathrm{PF}},1} \Big)$. Then,*

(i) sampled weights $w_{n+1}^{(t)} - w_n^{(t)} \in \left[\frac{2\sqrt{2}}{5} \frac{v_{\min}}{v_{\max}} \frac{1}{N}, 2\sqrt{2} \frac{v_{\max}}{v_{\min}} \frac{1}{N} \right]$ for all $n \in \{0, 1, \cdots, N-1\}$.

(ii) $\tilde{\Phi}_t$ is strictly increasing with $\tilde{\Phi}_t' \in \left[\frac{1}{\sqrt{2}} \frac{v_{\min}}{v_{\max}} - \frac{4\sqrt{2} \frac{v_{\max}}{v_{\min}} l_{\tilde{\Phi}_t,n,1}}{N}, \sqrt{2} \frac{v_{\max}}{v_{\min}} + \frac{4\sqrt{2} \frac{v_{\max}}{v_{\min}} l_{\tilde{\Phi}_t,n,1}}{N} \right]$. If $N \ge 16 \frac{v_{\max}^2}{v_{\min}^2} \frac{1}{\tilde{s}^{(t)}(1)} \Big(l_{v,0} + \frac{1}{2} l_{\mathbf{f}_{\mathrm{PF}},1} \Big)$, then $\tilde{\Phi}_t' \in \left[\frac{1}{2\sqrt{2}} \frac{v_{\min}}{v_{\max}}, \frac{5}{2\sqrt{2}} \frac{v_{\max}}{v_{\min}} \right]$.

(iii) $\tilde{\Phi}_t$ *is Lipschitz-smooth with modulus* $l_{\Phi_{\mathrm{iter}},1} = \mathcal{O}\left(\frac{v_{\max}^2}{s(1)v_{\min}^2}\left(l_{v,0} + \frac{1}{2}l_{\mathbf{f}_{\mathrm{PF}},1}\right)\right)$.

Proof. Since Φ_t is strictly increasing and $\Phi_t'(w) \in \left[\frac{1}{2\sqrt{2}}\frac{v_{\min}}{v_{\max}}, \frac{5}{2\sqrt{2}}\frac{v_{\max}}{v_{\min}}\right]$, $w \in [0,1]$, the inverse Φ_t^{-1} is well-defined on $[0,1]$ and satisfies $(\Phi_t^{-1})'(q) = \frac{1}{\Phi_t'(\Phi_t^{-1}(q))} \in \left[\frac{2\sqrt{2}}{5}\frac{v_{\min}}{v_{\max}}, 2\sqrt{2}\frac{v_{\max}}{v_{\min}}\right]$. Now, since $w_n^{(t)} = \Phi_t^{-1}(n/N)$ and $w_{n+1}^{(t)} - w_n^{(t)} = \Phi_t^{-1}\left(\frac{n+1}{N}\right) - \Phi_t^{-1}\left(\frac{n}{N}\right)$, by the mean value theorem, there exists $\xi_n \in (n/N, (n+1)/N)$ such that $w_{n+1}^{(t)} - w_n^{(t)} = (\Phi_t^{-1})'(\xi_n)\frac{1}{N} \in \left[\frac{2\sqrt{2}}{5}\frac{v_{\min}}{v_{\max}}\frac{1}{N}, 2\sqrt{2}\frac{v_{\max}}{v_{\min}}\frac{1}{N}\right]$. This proves claim (i).

Suppose Assumptions 1, 2, 3 hold. Lemma 6.(i) yields

$$s_n := \frac{\tilde{\Phi}_t(w_{n+1}^{(t)}) - \tilde{\Phi}_t(w_n^{(t)})}{w_{n+1}^{(t)} - w_n^{(t)}} \in \left[\frac{1}{\sqrt{2}}\frac{v_{\min}}{v_{\max}}, \sqrt{2}\frac{v_{\max}}{v_{\min}}\right], \quad n = 0, \cdots, N-1, \tag{102}$$

and Lemma 6.(ii) ensures that the adjacent secants satisfy

$$|s_{n+1} - s_n| \le l_{\tilde{\Phi}_t,n,1}(w_{n+2}^{(t)} - w_n^{(t)}), \quad n = 0, \cdots, N-2, \tag{103}$$

where $l_{\tilde{\Phi}_t,n,1} \le \frac{1}{\tilde{s}^{(t)}(1)}\left(l_{v,0} + \frac{1}{2}l_{\mathbf{f}_{\mathrm{PF}},1}\right)$ for all $n \in \{0, \cdots, N-2\}$.

Let $d_n := \tilde{\Phi}_t'(w_n^{(t)})$, and $c_n := 3s_n - d_n - d_{n+1}$ for $n = 0, \cdots, N-1$. According to Algorithm 2, on each interval $[w_n^{(t)}, w_{n+1}^{(t)}]$, writing $\theta = \frac{w - w_n^{(t)}}{w_{n+1}^{(t)} - w_n^{(t)}} \in [0,1]$, there is

$$\tilde{\Phi}_t'(w) = (1-\theta)^2 d_n + 2\theta(1-\theta)c_n + \theta^2 d_{n+1}. \tag{104}$$

Here, the coefficients $(1-\theta)^2$, $2\theta(1-\theta)$, θ^2 are nonnegative and sum to one, so $\tilde{\Phi}_t'(w)$ is a convex combination of d_n, c_n, d_{n+1}. Therefore it suffices to bound these three quantities uniformly.

For interior sampled weights, i.e., $n = 1, \cdots, N-1$, the PCHIP slope rule gives

$$d_n \in [\min(s_{n-1}, s_n), \max(s_{n-1}, s_n)] \subseteq \left[\frac{1}{\sqrt{2}}\frac{v_{\min}}{v_{\max}}, \sqrt{2}\frac{v_{\max}}{v_{\min}}\right]. \tag{105}$$

It remains to control the endpoints d_0, d_N and the interval control values c_n. By the endpoint PCHIP formula, for some $\lambda \in (0,1)$, one has $d_0 = s_0 + \lambda(s_0 - s_1)$. In this way, (103) yields

$$|d_0 - s_0| \le \lambda|s_1 - s_0| \le l_{\tilde{\Phi}_t,0,1}(w_2^{(t)} - w_0^{(t)}) \le \frac{4\sqrt{2}\frac{v_{\max}}{v_{\min}}l_{\tilde{\Phi}_t,n,1}}{N}. \tag{106}$$

Hence the endpoints satisfy

$$d_0 \in \left[\frac{1}{\sqrt{2}}\frac{v_{\min}}{v_{\max}} - \frac{4\sqrt{2}\frac{v_{\max}}{v_{\min}}l_{\tilde{\Phi}_t,n,1}}{N}, \sqrt{2}\frac{v_{\max}}{v_{\min}} + \frac{4\sqrt{2}\frac{v_{\max}}{v_{\min}}l_{\tilde{\Phi}_t,n,1}}{N}\right], \tag{107}$$

$$d_N \in \left[\frac{1}{\sqrt{2}}\frac{v_{\min}}{v_{\max}} - \frac{4\sqrt{2}\frac{v_{\max}}{v_{\min}}l_{\tilde{\Phi}_t,n,1}}{N}, \sqrt{2}\frac{v_{\max}}{v_{\min}} + \frac{4\sqrt{2}\frac{v_{\max}}{v_{\min}}l_{\tilde{\Phi}_t,n,1}}{N}\right]. \tag{108}$$

Next consider $c_n = 3s_n - d_n - d_{n+1}$. For interior intervals $n = 1, \cdots, N-2$, we obtain

$$|c_n - s_n| \le |s_n - d_n| + |s_n - d_{n+1}| \le |s_n - s_{n-1}| + |s_{n+1} - s_n| \le \frac{4\sqrt{2}\frac{v_{\max}}{v_{\min}}l_{\tilde{\Phi}_t,n,1}}{N}. \tag{109}$$

Since $s_n \in \left[\frac{1}{\sqrt{2}}\frac{v_{\min}}{v_{\max}}, \sqrt{2}\frac{v_{\max}}{v_{\min}}\right]$, it follows that $\forall n \in \{1, \cdots, N-2\}$,

$$c_n \in \left[\frac{1}{\sqrt{2}}\frac{v_{\min}}{v_{\max}} - \frac{4\sqrt{2}\frac{v_{\max}}{v_{\min}}l_{\tilde{\Phi}_t,n,1}}{N}, \sqrt{2}\frac{v_{\max}}{v_{\min}} + \frac{4\sqrt{2}\frac{v_{\max}}{v_{\min}}l_{\tilde{\Phi}_t,n,1}}{N}\right]. \tag{110}$$

For the left endpoint interval, similarly we have $|c_0 - s_0| \le |d_0 - s_0| + |d_1 - s_0| \le \frac{4\sqrt{2}\frac{v_{\max}}{v_{\min}} l_{\tilde{\Phi}_t,n,1}}{N}$. Hence,

$$c_0 \in \Big[\frac{1}{\sqrt{2}}\frac{v_{\min}}{v_{\max}} - \frac{4\sqrt{2}\frac{v_{\max}}{v_{\min}} l_{\tilde{\Phi}_t,n,1}}{N}, \sqrt{2}\frac{v_{\max}}{v_{\min}} + \frac{4\sqrt{2}\frac{v_{\max}}{v_{\min}} l_{\tilde{\Phi}_t,n,1}}{N}\Big], \quad \text{and}$$
$$c_{N-1} \in \Big[\frac{1}{\sqrt{2}}\frac{v_{\min}}{v_{\max}} - \frac{4\sqrt{2}\frac{v_{\max}}{v_{\min}} l_{\tilde{\Phi}_t,n,1}}{N}, \sqrt{2}\frac{v_{\max}}{v_{\min}} + \frac{4\sqrt{2}\frac{v_{\max}}{v_{\min}} l_{\tilde{\Phi}_t,n,1}}{N}\Big].$$

Combining the above bounds, we have

$$d_n, c_n \in \Big[\frac{1}{\sqrt{2}}\frac{v_{\min}}{v_{\max}} - \frac{4\sqrt{2}\frac{v_{\max}}{v_{\min}} l_{\tilde{\Phi}_t,n,1}}{N}, \sqrt{2}\frac{v_{\max}}{v_{\min}} + \frac{4\sqrt{2}\frac{v_{\max}}{v_{\min}} l_{\tilde{\Phi}_t,n,1}}{N}\Big], \tag{111}$$

for all relevant n. Since $\tilde{\Phi}'_t(w)$ is a convex combination of d_n, c_n, d_{n+1} on each interval, it follows:

$$\frac{1}{\sqrt{2}}\frac{v_{\min}}{v_{\max}} - \frac{4\sqrt{2}\frac{v_{\max}}{v_{\min}} l_{\tilde{\Phi}_t,n,1}}{N} \le \tilde{\Phi}'_t(w) \le \sqrt{2}\frac{v_{\max}}{v_{\min}} + \frac{4\sqrt{2}\frac{v_{\max}}{v_{\min}} l_{\tilde{\Phi}_t,n,1}}{N}, \quad \forall w \in [0,1]. \tag{112}$$

Finally, if $N \ge 16\frac{v_{\max}^2}{v_{\min}^2}\frac{1}{\tilde{s}^{(t)}(1)}\Big(l_{v,0} + \frac{1}{2}l_{\mathbf{f}_{\mathrm{PF}},1}\Big)$, then $l_{\tilde{\Phi}_t,n,1} \le \frac{1}{\tilde{s}^{(t)}(1)}\Big(l_{v,0} + \frac{1}{2}l_{\mathbf{f}_{\mathrm{PF}},1}\Big)$ implies $\frac{4\sqrt{2}\frac{v_{\max}}{v_{\min}} l_{\tilde{\Phi}_t,n,1}}{N} \le \frac{\frac{1}{\sqrt{2}}\frac{v_{\min}}{v_{\max}}}{2}$. Therefore, it proves claim (ii):

$$\frac{1}{2\sqrt{2}}\frac{v_{\min}}{v_{\max}} \le \tilde{\Phi}'_t(w) \le \frac{5}{2\sqrt{2}}\frac{v_{\max}}{v_{\min}}, \quad \forall w \in [0,1]. \tag{113}$$

To prove claim (iii), it remains to bound $\tilde{\Phi}''_t$ uniformly on each interval. For $w \in [w_n^{(t)}, w_{n+1}^{(t)}]$,

$$\tilde{\Phi}''_t(w) = \frac{1}{w_{n+1}^{(t)} - w_n^{(t)}}\Big[(6d_n + 6d_{n+1} - 12s_n)\theta + (6s_n - 4d_n - 2d_{n+1})\Big]. \tag{114}$$

Since $\tilde{\Phi}''_t$ is affine in θ, its maximum absolute value on $[w_n^{(t)}, w_{n+1}^{(t)}]$ is attained at an endpoint. Hence

$$|\tilde{\Phi}''_t(w)| \le \frac{1}{w_{n+1}^{(t)} - w_n^{(t)}} \max\Big\{|4(d_n - s_n) + 2(d_{n+1} - s_n)|, |2(d_n - s_n) + 4(d_{n+1} - s_n)|\Big\}. \tag{115}$$

Using $s_n \in \Big[\frac{1}{\sqrt{2}}\frac{v_{\min}}{v_{\max}}, \sqrt{2}\frac{v_{\max}}{v_{\min}}\Big]$ and claim (i), the exact PCHIP (harmonic-mean) formula in Algorithm 2 yields

$$|d_n - s_n| \le \frac{5\frac{v_{\max}^2}{v_{\min}^2}\Big(2 + 5\frac{v_{\max}^2}{v_{\min}^2}\Big)}{\Big(2 + 5\frac{v_{\max}^2}{v_{\min}^2}\Big) + \Big(1 + 10\frac{v_{\max}^2}{v_{\min}^2}\Big)\frac{1}{2}\frac{v_{\min}^2}{v_{\max}^2}}|s_n - s_{n-1}|, \quad \text{and} \tag{116}$$

$$|d_{n+1} - s_n| \le \frac{2 + 5\frac{v_{\max}^2}{v_{\min}^2}}{\Big(1 + 10\frac{v_{\max}^2}{v_{\min}^2}\Big)\frac{1}{2}\frac{v_{\min}^2}{v_{\max}^2} + \Big(2 + 5\frac{v_{\max}^2}{v_{\min}^2}\Big)}|s_{n+1} - s_n|. \tag{117}$$

By (103), $|s_n - s_{n-1}| \le l_{\tilde{\Phi}_t,n,1}(w_n^{(t)} - w_{n-1}), \quad |s_{n+1} - s_n| \le l_{\tilde{\Phi}_t,n,1}(w_{n+1}^{(t)} - w_n^{(t)})$. Substituting these into (115), and again using claim (i) to bound $\frac{w_n^{(t)} - w_{n-1}}{w_{n+1}^{(t)} - w_n^{(t)}}$, we obtain

$$\sup_{w \in [w_n^{(t)}, w_{n+1}^{(t)}]} |\tilde{\Phi}''_t(w)| \le C_{\mathrm{pchip}} l_{\tilde{\Phi}_t,n,1}, \quad n \in \{1, \cdots, N-2\}, \tag{118}$$

where $C_{\mathrm{pchip}} = \mathcal{O}\Big(\frac{v_{\max}^2}{v_{\min}^2}\Big)$. For the first and last intervals, the endpoint PCHIP formula gives $\sup_{w \in [w_0^{(t)}, w_1^{(t)}]} |\tilde{\Phi}''_t(w)|, \sup_{w \in [w_{N-1}^{(t)}, w_N^{(t)}]} |\tilde{\Phi}''_t(w)| \le 6 l_{\tilde{\Phi}_t,n,1}$.

Combining the interior and endpoint bounds, $\tilde{\Phi}_t''$ is uniformly bounded on every open interval, and since the PCHIP interpolant is C^1, it follows that $\tilde{\Phi}_t'$ is globally Lipschitz-continuous on $[0,1]$. Therefore $\tilde{\Phi}_t$ is Lipschitz-smooth on $[0,1]$ with modulus

$$l_{\tilde{\Phi}_t,1} := \max\{6, C_{\text{pchip}}\} \max_{0\le n\le N-2} l_{\tilde{\Phi}_t,n,1}. \tag{119}$$

Using $l_{\tilde{\Phi}_t,n,1} \le \frac{1}{\tilde{s}^{(t)}(1)}\Big(l_{v,0} + \frac{1}{2} l_{\mathbf{f}_{\text{PF}},1}\Big)$, we may take

$$l_{\Phi_{\text{iter}},1} = \max\{6, C_{\text{pchip}}\} \frac{1}{\tilde{s}^{(t)}(1)}\Big(l_{v,0} + \frac{1}{2} l_{\mathbf{f}_{\text{PF}},1}\Big). \tag{120}$$

This proves claim (iii). □

The previous lemma analyzes a single interpolation step assuming the input CDF already satisfies suitable derivative bounds. To complete the argument, it remains to show that these bounds are preserved throughout the SURF iterations. The next lemma establishes this by induction on the outer-loop index t. This is an extension to the statement that Φ_t is uniformly $l_{\Phi_{\text{iter}},1}$-Lipschitz-smooth in Theorem 1.

Lemma 8. *Suppose Assumptions 1, 2, 3 hold. Consequently, Algorithm 1 with exact* InnerSolver *solutions and the interpolation rule in Algorithm 2 generates iterations that satisfy*

(i) *the sampled weight* $w_{n+1}^{(t)} - w_n^{(t)} \in \left[\frac{2\sqrt{2}}{5}\frac{v_{\min}}{v_{\max}}\frac{1}{N}, 2\sqrt{2}\frac{v_{\max}}{v_{\min}}\frac{1}{N}\right]$ *for all* $n \in \{0,1,\cdots,N-1\}$;

(ii) Φ_t *is strictly increasing with* $\Phi_t' \in \left[\frac{1}{\sqrt{2}}\frac{v_{\min}}{v_{\max}} - \frac{2\sqrt{2}\frac{v_{\max}}{v_{\min}} l_{\Phi_{\text{iter}},1}}{N}, \sqrt{2}\frac{v_{\max}}{v_{\min}} + \frac{2\sqrt{2}\frac{v_{\max}}{v_{\min}} l_{\Phi_{\text{iter}},1}}{N}\right]$. *If* N *is chosen such that* $N \ge 16\frac{v_{\max}^2}{v_{\min}^2}\frac{1}{\tilde{s}^{(t)}(1)}\Big(l_{v,0} + \frac{1}{2} l_{\mathbf{f}_{\text{PF}},1}\Big)$, *then* $\Phi_t' \in \left[\frac{1}{2\sqrt{2}}\frac{v_{\min}}{v_{\max}}, \frac{5}{2\sqrt{2}}\frac{v_{\max}}{v_{\min}}\right]$;

(iii) Φ_t *is Lipschitz-smooth with modulus* $l_{\Phi_{\text{iter}},1} = \mathcal{O}\left(\frac{v_{\max}^2}{s(1)v_{\min}^2}\Big(l_{v,0} + \frac{1}{2} l_{\mathbf{f}_{\text{PF}},1}\Big)\right)$.

Proof. We use induction on $t \ge 0$.

For the base case, $\Phi_0(w) = w$. Therefore, $w_{n+1}^{(0)} - w_n^{(0)} = \frac{1}{N}$, so claim (i) holds. Additionally, Φ_0 is differentiable with $\Phi_0'(w) = 1$, hence Φ_0 satisfies the derivative bounds in claim (ii). Moreover, Φ_0' is Lipschitz-continuous with modulus $l_{\Phi_0,1} = 0 \le l_{\Phi_{\text{iter}},1}$, which verifies claim (iii). Hence, all claims hold for $t = 0$.

Now assume the claims hold for some $t \ge 0$. We prove that they also hold at $t+1$.

By Lemma 7, we know that $\tilde{\Phi}_t$ satisfies the same derivative and Lipschitz-smoothness properties. The update rule in Algorithm 1, $\Phi_{t+1} = (1-\alpha)\Phi_t + \alpha\tilde{\Phi}_t$, gives

$$\begin{aligned}\Phi_{t+1}'(w) =& (1-\alpha)\Phi_t'(w) + \alpha\tilde{\Phi}_t'(w) \\ &\in \left[\frac{1}{\sqrt{2}}\frac{v_{\min}}{v_{\max}} - \frac{2\sqrt{2}\frac{v_{\max}}{v_{\min}} l_{\Phi_{\text{iter}},1}}{N}, \sqrt{2}\frac{v_{\max}}{v_{\min}} + \frac{2\sqrt{2}\frac{v_{\max}}{v_{\min}} l_{\Phi_{\text{iter}},1}}{N}\right],\end{aligned} \tag{121}$$

since both Φ_t' and $\tilde{\Phi}_t'$ lie in this interval. This proves claim (ii).

Additionally, Φ_t' and $\tilde{\Phi}_t'$ are Lipschitz-continuous with moduli at most $l_{\Phi_{\text{iter}},1}$. Therefore, for all $u, w \in [0,1]$,

$$\begin{aligned}|\Phi_{t+1}'(w) - \Phi_{t+1}'(u)| \le& (1-\alpha)|\Phi_t'(w) - \Phi_t'(u)| + \alpha|\tilde{\Phi}_t'(w) - \tilde{\Phi}_t'(u)| \\ \le& \big((1-\alpha) l_{\Phi_{\text{iter}},1} + \alpha l_{\Phi_{\text{iter}},1}\big)|w-u| = l_{\Phi_{\text{iter}},1}|w-u|.\end{aligned}$$

Thus Φ_{t+1}' is Lipschitz-continuous with constant at most $l_{\Phi_{\text{iter}},1}$, proving claim (iii).

Next, since Φ_{t+1} satisfies the derivative bounds in claim (ii), Lemma 7((i)) directly gives $w_{n+1}^{(t+1)} - w_n^{(t+1)} \in \left[\frac{2\sqrt{2}}{5}\frac{v_{\min}}{v_{\max}}\frac{1}{N}, 2\sqrt{2}\frac{v_{\max}}{v_{\min}}\frac{1}{N}\right]$, which proves claim (i). The induction is complete. □

Algorithm 2 Piecewise Cubic Hermite Interpolating Polynomial [37] for $\{(w_n^{(t)}, \tilde{s}^{(t)}(w_n^{(t)}))\}_{n=0}^N$

Require: Samples $\{(w_n^{(t)}, \tilde{s}^{(t)}(w_n^{(t)}))\}_{n=0}^N$ at strictly increasing weights. Hermite basis functions H, given by $H_{00}(\theta) = 2\theta^3 - 3\theta^2 + 1$, $H_{10}(\theta) = \theta^3 - 2\theta^2 + \theta$, $H_{01}(\theta) = -2\theta^3 + 3\theta^2$, and $H_{11}(\theta) = \theta^3 - \theta^2$.

if $N = 1$ **then**
  $m_0, m_1 = \dfrac{\tilde{s}^{(t)}(w_1^{(t)}) - \tilde{s}^{(t)}(w_0^{(t)})}{w_1^{(t)} - w_0^{(t)}}$
else
  for $n = 0, \cdots, N-1$ **do**
    $\delta_n = \dfrac{\tilde{s}^{(t)}(w_{n+1}^{(t)}) - \tilde{s}^{(t)}(w_n^{(t)})}{w_{n+1}^{(t)} - w_n^{(t)}}$
  end for
  for $n = 1, \cdots, N-1$ **do**
    if $\delta_{n-1}\delta_n \le 0$ **then**
      $m_n = 0$
    else
      $\omega_1 = 2(w_{n+1}^{(t)} - w_n^{(t)}) + (w_n^{(t)} - w_{n-1}^{(t)})$
      $\omega_2 = (w_{n+1}^{(t)} - w_n^{(t)}) + 2(w_n^{(t)} - w_{n-1}^{(t)})$
      $m_n = (\omega_1 + \omega_2)/\big(\frac{\omega_1}{\delta_{n-1}} + \frac{\omega_2}{\delta_n}\big)$ ▷ weighted harmonic mean
    end if
  end for
  $m_0 = \dfrac{\big(2(w_1^{(t)} - w_0^{(t)}) + (w_2^{(t)} - w_1^{(t)})\big)\delta_0 - (w_1^{(t)} - w_0^{(t)})\delta_1}{(w_1^{(t)} - w_0^{(t)}) + (w_2^{(t)} - w_1^{(t)})}$ ▷ Endpoint
  if $m_0\delta_0 \le 0$ **then**
    $m_0 = 0$
  else if $|m_0| > 3|\delta_0|$ **then**
    $m_0 = 3\delta_0$
  end if
  $m_N = \dfrac{\big(2(w_N^{(t)} - w_{N-1}^{(t)}) + (w_{N-1}^{(t)} - w_{N-2}^{(t)})\big)\delta_{N-1} - (w_N^{(t)} - w_{N-1}^{(t)})\delta_{N-2}}{(w_N^{(t)} - w_{N-1}^{(t)}) + (w_{N-1}^{(t)} - w_{N-2}^{(t)})}$ ▷ Endpoint
  if $m_N\delta_{N-1} \le 0$ **then**
    $m_N = 0$
  else if $|m_N| > 3|\delta_{N-1}|$ **then**
    $m_N = 3\delta_{N-1}$
  end if
end if
for $n = 0, \cdots, N-1$ **do**
  Build the normalized coordinate $\theta = \frac{w - w_n^{(t)}}{w_{n+1}^{(t)} - w_n^{(t)}}$ for all $w \in [w_n^{(t)}, w_{n+1}^{(t)}]$ and $\tilde{s}^{(t)}(w) = H_{00}(\theta)\tilde{s}^{(t)}(w_n^{(t)}) + H_{10}(\theta)(w_{n+1}^{(t)} - w_n^{(t)})m_n + H_{01}(\theta)\tilde{s}^{(t)}(w_{n+1}^{(t)}) + H_{11}(\theta)(w_{n+1}^{(t)} - w_n^{(t)})m_{n+1}$
end for
return $\tilde{s}^{(t)}(\cdot)$

Discretization Error under Exact InnerSolver Solutions. The Lipschitz-smoothness results established earlier are essential for applying Landau inequality [44], summarized in Lemma 9.

Lemma 9 (Landau inequality [44])**.** *Suppose $\Phi, \Phi_t : [0,1] \to \mathbb{R}$ are respectively $l_{\Phi,1}$- and $l_{\Phi_{\text{iter}},1}$-Lipschitz-smooth on $[0,1]$. Then,*

$$\|\Phi' - \Phi_t'\|_\infty \le 2\sqrt{2}\sqrt{(l_{\Phi,1} + l_{\Phi_{\text{iter}},1})\|\Phi - \Phi_t\|_\infty}. \tag{122}$$

Moreover, if Φ, Φ_t are both second-order Lipschitz-continuous, then $\|\tilde{\Phi}_t'' - \Phi''\|_\infty = \mathcal{O}(\|\tilde{\Phi}_t - \Phi\|_\infty^{1/3})$.

This tool allows us to convert function-value error in the CDF approximation into derivative error, which is the key step in controlling the discretization error induced by rebuilding the polyline

approximation $\tilde{s}^{(t)}$ from finitely many sampled PF points. We first bound the uniform discrepancy between s and its rebuilt approximation $\tilde{s}$ in terms of the sampled polyline error $\max_n(s(w_n)-\tilde{s}(w_n))$. In this way, we can proceed to the proof of the convergence of SURF using exact-InnerSolver solutions in Theorem 1.

Lemma 10 (Discretization error)**.** *Suppose Assumptions 1, 2, 3 hold. Let* $\kappa := 4\sqrt{2}\frac{v_{\max}}{v_{\min}}\sqrt{l_{\Phi_{\text{iter}},1}+l_{\Phi,1}}$. *Then*

$$\|s-\tilde{s}^{(t)}\|_\infty \leq \frac{\tilde{s}^{(t)}(1)(l_{\tilde{\Phi}_t\circ\Phi_t^{-1},1}+l_{\Phi\circ\Phi_t^{-1},1})}{8}\frac{1}{N^2}+\max_{0\leq n\leq N}(s(w_n^{(t)})-\tilde{s}(w_n^{(t)})). \tag{123}$$

where $l_{\tilde{\Phi}_t\circ\Phi_t^{-1},1} \leq \left(\frac{s(1)}{v_{\min}}\right)^2\|\tilde{\Phi}_t''-\Phi_t''\|_\infty + 2\sqrt{2}l_{\Phi_{\text{iter}},1}\left(\frac{s(1)}{v_{\min}}\right)^3\sqrt{2l_{\Phi_{\text{iter}},1}\|\tilde{\Phi}_t-\Phi_t\|_\infty}$, *and* $l_{\Phi\circ\Phi_t^{-1},1} \leq \left(\frac{s(1)}{v_{\min}}\right)^2\|\Phi_t''-\Phi''\|_\infty + 2\sqrt{2}l_{\Phi_{\text{iter}},1}\left(\frac{s(1)}{v_{\min}}\right)^3\sqrt{(l_{\Phi_{\text{iter}},1}+l_{\Phi,1})\|\Phi_t-\Phi\|_\infty}$. *Define* $e_t(q) := \Phi(\Phi_t^{-1}(q))-q$. *Then*

$$\begin{aligned}&\max_{0\leq n\leq N}(s(w_n^{(t)})-\tilde{s}(w_n^{(t)})) \\ \leq&\frac{l_{\tilde{\mathbf{f}}_{\text{PF}},1}^2 s(1)^3}{6}\left(\frac{1}{N^2}+3\max_{0\leq n\leq N-1}\left|e_t\left(\tfrac{n+1}{N}\right)-e_t\left(\tfrac{n}{N}\right)\right|^2+N\max_{0\leq n\leq N-1}\left|e_t\left(\tfrac{n+1}{N}\right)-e_t\left(\tfrac{n}{N}\right)\right|^3\right).\end{aligned} \tag{124}$$

Moreover, $\|\tilde{\Phi}_t-\Phi\|_\infty \leq \frac{2\sqrt{2}}{s(1)}\|\tilde{s}^{(t)}-s\|_\infty$ *and* $\max_{0\leq n\leq N-1}\left|e_t\left(\frac{n+1}{N}\right)-e_t\left(\frac{n}{N}\right)\right| \leq \frac{\kappa}{N}\|\Phi_t-\Phi\|_\infty^{1/2}$.

Proof. Fix $w\in[0,1]$ and choose $n\in\{0,\dots,N-1\}$ such that $w\in[w_n^{(t)},w_{n+1}^{(t)}]$. Let $q=\Phi_t(w)$ and $q_n=\Phi_t(w_n^{(t)})$. Since the knots are generated by $w_n^{(t)}=\Phi_t^{-1}(n/N)$, we have $q_n=n/N$. Let I_n^q denote the linear interpolant in the q-coordinate, pulled back to the w-coordinate. Then

$$\begin{aligned}|s(w)-\tilde{s}^{(t)}(w)| \leq&|s(w)-I_n^q s(w)|+|I_n^q s(w)-I_n^q\tilde{s}^{(t)}(w)|+|I_n^q\tilde{s}^{(t)}(w)-\tilde{s}^{(t)}(w)| \\ \leq&|s(w)-I_n^q s(w)|+|I_n^q\tilde{s}^{(t)}(w)-\tilde{s}^{(t)}(w)|+\max_{0\leq n\leq N}|s(w_n^{(t)})-\tilde{s}^{(t)}(w_n^{(t)})|.\end{aligned} \tag{125}$$

Here, the second inequality follows because $I_n^q s(w)-I_n^q\tilde{s}^{(t)}(w)$ is a convex combination of the endpoint errors. Since $s(w)=s(1)\Phi(w)$, we have $s(\Phi_t^{-1}(q))=s(1)\Phi(\Phi_t^{-1}(q))$. Hence, the standard linear interpolation analysis [88] gives

$$|s(w)-I_n^q s(w)| \leq \frac{s(1)l_{\Phi\circ\Phi_t^{-1},1}}{8N^2}, \quad \text{and} \quad |\tilde{s}^{(t)}(w)-I_n^q\tilde{s}^{(t)}(w)| \leq \frac{s(1)l_{\tilde{\Phi}_t\circ\Phi_t^{-1},1}}{8N^2}. \tag{126}$$

In this way, $\frac{\partial}{\partial q}\tilde{\Phi}_t(\Phi_t^{-1}(q))=\frac{\tilde{\Phi}_t'(\Phi_t^{-1}(q))}{\Phi_t'(\Phi_t^{-1}(q))}$ and $\frac{\partial}{\partial q}\Phi(\Phi_t^{-1}(q))=\frac{\Phi'(\Phi_t^{-1}(q))}{\Phi_t'(\Phi_t^{-1}(q))}$ gives

$$\begin{aligned}l_{\tilde{\Phi}_t\circ\Phi_t^{-1},1} \leq&\left(\frac{s(1)}{v_{\min}}\right)^2\|\tilde{\Phi}_t''-\Phi_t''\|_\infty+2\sqrt{2}l_{\Phi_{\text{iter}},1}\left(\frac{s(1)}{v_{\min}}\right)^3\sqrt{2l_{\Phi_{\text{iter}},1}\|\tilde{\Phi}_t-\Phi_t\|_\infty}, \quad \text{and} \\ l_{\Phi\circ\Phi_t^{-1},1} \leq&\left(\frac{s(1)}{v_{\min}}\right)^2\|\Phi_t''-\Phi''\|_\infty+2\sqrt{2}l_{\Phi_{\text{iter}},1}\left(\frac{s(1)}{v_{\min}}\right)^3\sqrt{(l_{\Phi_{\text{iter}},1}+l_{\Phi,1})\|\Phi_t-\Phi\|_\infty}.\end{aligned}$$

By uniform Lipschitz-smoothness of $\tilde{\Phi}_t$, Φ_t, and Φ (see Lemma 2, 7, 8), we have $\|\tilde{\Phi}_t''-\Phi_t''\|_\infty \leq 2l_{\Phi_{\text{iter}},1}$ and $\|\Phi_t''-\Phi''\|_\infty \leq l_{\Phi_{\text{iter}},1}+l_{\Phi,1}$. Plugging these in (125) proves the first inequality.

Fix any $n\in\{0,\dots,N-1\}$ and write $a=s(w_n^{(t)})$, $b=s(w_{n+1}^{(t)})$. Then $\tilde{\mathbf{f}}_{\text{PF}}(b)-\tilde{\mathbf{f}}_{\text{PF}}(a)=\int_a^b\tilde{\mathbf{f}}_{\text{PF}}'(r)dr$. Using $\|\tilde{\mathbf{f}}_{\text{PF}}'(a)\|=1$, we obtain

$$\begin{aligned}\|\tilde{\mathbf{f}}_{\text{PF}}(b)-\tilde{\mathbf{f}}_{\text{PF}}(a)\| \geq&\left\langle\tilde{\mathbf{f}}_{\text{PF}}(b)-\tilde{\mathbf{f}}_{\text{PF}}(a),\tilde{\mathbf{f}}_{\text{PF}}'(a)\right\rangle=\int_a^b\left\langle\tilde{\mathbf{f}}_{\text{PF}}'(r),\tilde{\mathbf{f}}_{\text{PF}}'(a)\right\rangle dr \\ =&\int_a^b\left(1-\frac{1}{2}\|\tilde{\mathbf{f}}_{\text{PF}}'(r)-\tilde{\mathbf{f}}_{\text{PF}}'(a)\|^2\right)dr \geq (b-a)-\frac{l_{\tilde{\mathbf{f}}_{\text{PF}},1}^2}{6}(b-a)^3.\end{aligned} \tag{127}$$

Also, $\|\tilde{\mathbf{f}}_{\mathrm{PF}}(b) - \tilde{\mathbf{f}}_{\mathrm{PF}}(a)\| \le \int_a^b \|\tilde{\mathbf{f}}'_{\mathrm{PF}}(r)\| dr = b - a$. Combining this with (127)

$$0 \le \big(s(w_{n+1}^{(t)}) - s(w_n^{(t)})\big) - \left\|\tilde{\mathbf{f}}_{\mathrm{PF}}(s(w_{n+1}^{(t)})) - \tilde{\mathbf{f}}_{\mathrm{PF}}(s(w_n^{(t)}))\right\| \le \frac{l^2_{\tilde{\mathbf{f}}_{\mathrm{PF}},1}}{6}\big(s(w_{n+1}^{(t)}) - s(w_n^{(t)})\big)^3. \tag{128}$$

By definition of $\tilde{s}^{(t)}$ in (12), each summand is nonnegative by (128). Hence $0 \le s(w_0^{(t)}) - \tilde{s}^{(t)}(w_0^{(t)}) \le \cdots \le s(w_N^{(t)}) - \tilde{s}^{(t)}(w_N^{(t)})$, so it suffices to bound the terminal error. Summing (128) over $n = 0, \ldots, N-1$ gives

$$0 \le s(w_N^{(t)}) - \tilde{s}^{(t)}(w_N^{(t)}) \le \frac{l^2_{\tilde{\mathbf{f}}_{\mathrm{PF}},1}}{6} \sum_{n=0}^{N-1} \big(s(w_{n+1}^{(t)}) - s(w_n^{(t)})\big)^3. \tag{129}$$

Moreover, by definition of $e_t(q) = \Phi(\Phi_t^{-1}(q)) - q$, and since $w_n^{(t)} = \Phi_t^{-1}(n/N)$,

$$s(w_{n+1}^{(t)}) - s(w_n^{(t)}) = s(1)\left(\Phi(w_{n+1}^{(t)}) - \Phi(w_n^{(t)})\right) = s(1)\left(\frac{1}{N} + e_t\left(\frac{n+1}{N}\right) - e_t\left(\frac{n}{N}\right)\right). \tag{130}$$

Substituting (130) into (129) and expanding the cube gives

$$\begin{aligned} s(w_N^{(t)}) - \tilde{s}^{(t)}(w_N^{(t)}) \le & \frac{l^2_{\tilde{\mathbf{f}}_{\mathrm{PF}},1} s(1)^3}{6}\Bigg(\frac{1}{N^2} + 3 \max_{0\le n\le N-1} \left|e_t\left(\tfrac{n+1}{N}\right) - e_t\left(\tfrac{n}{N}\right)\right|^2 \\ & + N \max_{0\le n\le N-1} \left|e_t\left(\tfrac{n+1}{N}\right) - e_t\left(\tfrac{n}{N}\right)\right|^3\Bigg), \end{aligned} \tag{131}$$

where we used $\sum_{n=0}^{N-1}\left(e_t\left(\frac{n+1}{N}\right) - e_t\left(\frac{n}{N}\right)\right) = e_t(1) - e_t(0) = 0$. Together with monotonicity of the prefix errors, this proves (124). It remains to bound $\|\tilde{\Phi}_t - \Phi\|_\infty$. By definition,

$$\begin{aligned} \|\tilde{\Phi}_t - \Phi\|_\infty &= \left\|\frac{1}{\tilde{s}^{(t)}(1)}\tilde{s}^{(t)} - \frac{1}{s(1)}s\right\|_\infty \\ &\le \left\|\frac{1}{\tilde{s}^{(t)}(1)}\big(\tilde{s}^{(t)} - s\big)\right\|_\infty + \left|\frac{1}{\tilde{s}^{(t)}(1)} - \frac{1}{s(1)}\right| \|s\|_\infty. \end{aligned} \tag{132}$$

Since $\|s\|_\infty = s(1)$ and $\left|\frac{1}{\tilde{s}^{(t)}(1)} - \frac{1}{s(1)}\right| = \frac{|s(1) - \tilde{s}^{(t)}(1)|}{\tilde{s}^{(t)}(1)s(1)}$, we obtain

$$\|\tilde{\Phi}_t - \Phi\|_\infty \le \frac{\|\tilde{s}^{(t)} - s\|_\infty}{\tilde{s}^{(t)}(1)} + \frac{|s(1) - \tilde{s}^{(t)}(1)|}{\tilde{s}^{(t)}(1)} \le \frac{2\|\tilde{s}^{(t)} - s\|_\infty}{\tilde{s}^{(t)}(1)}, \tag{133}$$

where the last inequality uses $|s(1) - \tilde{s}^{(t)}(1)| \le \|\tilde{s}^{(t)} - s\|_\infty$. By Lemma 5, there is $\tilde{s}^{(t)}(1) \ge \frac{1}{\sqrt{2}}s(1)$. Substituting into (133) gives $\|\tilde{\Phi}_t - \Phi\|_\infty \le \frac{2\sqrt{2}}{s(1)}\|\tilde{s}^{(t)} - s\|_\infty$. Finally, for each n,

$$\begin{aligned} \left|e_t\left(\tfrac{n+1}{N}\right) - e_t\left(\tfrac{n}{N}\right)\right| &= \left|\Phi(w_{n+1}^{(t)}) - \Phi(w_n^{(t)}) - \frac{1}{N}\right| \\ &= \left|\int_{n/N}^{(n+1)/N} \frac{\Phi'(\Phi_t^{-1}(q)) - \Phi_t'(\Phi_t^{-1}(q))}{\Phi_t'(\Phi_t^{-1}(q))} dq\right|. \end{aligned} \tag{134}$$

By Lemma 8, $\Phi_t'(w) \ge \frac{1}{2\sqrt{2}}\frac{v_{\min}}{v_{\max}}$, $\forall w \in [0,1]$. Therefore, $\left|e_t\left(\frac{n+1}{N}\right) - e_t\left(\frac{n}{N}\right)\right| \le \frac{1}{N}\frac{2\sqrt{2}v_{\max}}{v_{\min}}\|\Phi' - \Phi_t'\|_\infty$. By Lemma 9, $\|\Phi' - \Phi_t'\|_\infty \le 2\sqrt{2}\sqrt{(l_{\Phi_{\mathrm{iter}},1} + l_{\Phi,1})\|\Phi_t - \Phi\|_\infty}$. This finishes the proof. □

Convergence of Algorithm 1 with Exact InnerSolver Solutions. The earlier analysis therefore directly renders the convergence of SURF with exact InnerSolver solutions, which is a special case of Theorem 1.

Theorem 2. *Suppose* InnerSolver *is exact. Suppose Assumptions 1, 2, 3 hold. Denote* $\kappa := 4\sqrt{2}\frac{v_{\max}}{v_{\min}}\sqrt{l_{\Phi_{\text{iter}},1}+l_{\Phi,1}}$. *Let* $\alpha \in (0,1]$ *and choose* $N \geq 2\sqrt{\frac{\sqrt{2}l^2_{\tilde{\mathbf{f}}_{\text{PF}},1}s(1)^2}{3}(3\kappa^2+\kappa^3)} = \Theta(\kappa^{3/2})$. *Then Algorithm 1 satisfies the convergence bound for all* $t \geq 0$,

$$\|\Phi_t - \Phi\|_\infty \leq \left(1-\frac{\alpha}{2}\right)^t \|\Phi_0 - \Phi\|_\infty + \mathcal{O}\left(\frac{1}{N^2}\right). \tag{135}$$

Remark 3. *In this case, we can choose* $\alpha = 1$ *to maximize the convergence.*

Proof. By the Φ_{t+1} update rule and the triangle inequality,

$$\|\Phi_{t+1} - \Phi\|_\infty \leq (1-\alpha)\|\Phi_t - \Phi\|_\infty + \alpha\|\tilde{\Phi}_t - \Phi\|_\infty. \tag{136}$$

By Lemma 10, there is

$$\begin{aligned}
&\|\tilde{\Phi}_t - \Phi\|_\infty \\
\leq&\frac{\sqrt{2}}{4N^2}\left(l_{\tilde{\Phi}_t\circ\Phi_t^{-1},1} + l_{\Phi\circ\Phi_t^{-1},1}\right) + \frac{\sqrt{2}l^2_{\tilde{\mathbf{f}}_{\text{PF}},1}s(1)^2}{3N^2}\left(1+3\kappa^2\|\Phi_t-\Phi\|_\infty + \kappa^3\|\Phi_t-\Phi\|_\infty^{3/2}\right) \\
\leq&\frac{\sqrt{2}l^2_{\tilde{\mathbf{f}}_{\text{PF}},1}s(1)^2}{3N^2}\left(3\kappa^2+\kappa^3\right)\|\Phi_t-\Phi\|_\infty + \frac{\frac{\sqrt{2}}{4}\left(l_{\tilde{\Phi}_t\circ\Phi_t^{-1},1} + l_{\Phi\circ\Phi_t^{-1},1}\right) + \frac{\sqrt{2}}{3}l^2_{\tilde{\mathbf{f}}_{\text{PF}},1}s(1)^2}{N^2} \\
\leq&\frac{1}{4}\|\Phi_t-\Phi\|_\infty + \frac{\frac{\sqrt{2}}{4}\left(l_{\tilde{\Phi}_t\circ\Phi_t^{-1},1} + l_{\Phi\circ\Phi_t^{-1},1}\right) + \frac{\sqrt{2}}{3}l^2_{\tilde{\mathbf{f}}_{\text{PF}},1}s(1)^2}{N^2} \\
\leq&\frac{1}{2}\|\Phi_t-\Phi\|_\infty + \frac{\frac{\sqrt{2}}{4}\left(\frac{s(1)}{v_{\min}}\right)^2\left(\|\tilde{\Phi}_t''-\Phi_t''\|_\infty + \|\Phi_t''-\Phi''\|_\infty\right) + \frac{\sqrt{2}}{3}l^2_{\tilde{\mathbf{f}}_{\text{PF}},1}s(1)^2}{N^2} + \mathcal{O}\left(\frac{1}{N^4}\right)
\end{aligned}$$

where the second inequality holds because Φ_t and Φ are CDFs and $\|\Phi_t-\Phi\|_\infty \leq 1$, the third inequality follows $\frac{\sqrt{2}l^2_{\tilde{\mathbf{f}}_{\text{PF}},1}s(1)^2}{3N^2}\left(\kappa^2+\frac{\sqrt{2}}{3}\kappa^3\right) \leq \frac{1}{4}$ from the $N = 2\sqrt{\frac{\sqrt{2}l^2_{\tilde{\mathbf{f}}_{\text{PF}},1}s(1)^2}{3N^2}\left(\kappa^2+\frac{\sqrt{2}}{3}\kappa^3\right)} = \Omega(\kappa^{1.5})$ choice, and the fourth is from the upper bound for $l_{\tilde{\Phi}_t\circ\Phi^{-1},1}$ in Lemma 10 and Young's inequality for the term $\frac{2\sqrt{2}l_{\Phi,1}\left(\frac{s(1)}{v_{\min}}\right)^3\sqrt{(l_{\Phi_{\text{iter}},1}+l_{\Phi,1})\|\tilde{\Phi}_t-\Phi\|_\infty}}{N^2}$ in $l_{\Phi\circ\Phi_t}$ and its counterpart. Substituting this into (136) yields

$$\begin{aligned}
\|\Phi_{t+1}-\Phi\|_\infty \leq&\left(1-\frac{\alpha}{2}\right)\|\Phi_t-\Phi\|_\infty + \alpha\mathcal{O}\left(\frac{1}{N^4}\right) \\
&+\alpha\frac{\frac{\sqrt{2}}{4}\left(\frac{s(1)}{v_{\min}}\right)^2\left(\|\tilde{\Phi}_t''-\Phi_t''\|_\infty + \|\Phi_t''-\Phi''\|_\infty\right) + \frac{\sqrt{2}}{3}l^2_{\tilde{\mathbf{f}}_{\text{PF}},1}s(1)^2}{N^2}.
\end{aligned} \tag{137}$$

Here, $\left(\|\tilde{\Phi}_t''-\Phi_t''\|_\infty + \|\Phi_t''-\Phi''\|_\infty\right) \leq 3l_{\Phi_{\text{iter}},1} + l_{\Phi,1}$ by the Lipschitz-smoothness of Φ_t and Φ (see Lemma 2, 7, 8). Thus, unrolling gives its convergence to the floor of $\mathcal{O}(1/N^2)$:

$$\begin{aligned}
&\|\Phi_t-\Phi\|_\infty \\
\leq&\left(1-\frac{\alpha}{2}\right)^t\|\Phi_0-\Phi\|_\infty + \frac{\frac{\sqrt{2}}{2}\left(\frac{s(1)}{v_{\min}}\right)^2(3l_{\Phi_{\text{iter}},1}+l_{\Phi,1}) + \frac{2\sqrt{2}}{3}l^2_{\tilde{\mathbf{f}}_{\text{PF}},1}s(1)^2}{N^2} + \mathcal{O}\left(\frac{1}{N^4}\right).
\end{aligned} \tag{138}$$

Moreover, if Φ_t and Φ are 2nd-order Lipschitz-continuous, Landau inequality [44] gives $\|\tilde{\Phi}_t'' - \Phi''\|_\infty = \mathcal{O}(\|\tilde{\Phi}_t-\Phi\|_\infty^{1/3})$ and we can similarly obtain that $\|\Phi_t-\Phi\|_\infty$ converges linearly to $\frac{\sqrt{2}l^2_{\tilde{\mathbf{f}}_{\text{PF}},1}s(1)^2}{3N^2} + \mathcal{O}\left(\frac{1}{N^4}\right)$. □

The exact InnerSolver analysis characterizes the error when each scalarized subproblem is solved exactly, leaving only the finite-sampling discretization floor. We next extend the analysis to the practical setting where the InnerSolver is run for finitely many steps and the returned scalarized solutions are inexact.

E.2 Proof of the Second Part of Theorem 1: Optimization Error Control under Inexact InnerSolver Solutions

In this section, we prove the convergence of SURF with inexact InnerSolver in Theorem 1. In this case, $\tilde{s}$ and $\tilde{\Phi}$ are built via inexact estimation $\mathbf{u}_n^{(t)}$ or $\mathbf{x}_n^{(t)}$. In contrast, we use $\hat{s}_t$ to denote the unknown estimate that can be built up by the result from exact InnerSolver solutions. That is, for $n = 0, \cdots, N-1$,

$$\hat{s}^{(t)}(w_0^{(t)}) := 0, \quad \hat{s}^{(t)}(w_{n+1}^{(t)}) := \hat{s}^{(t)}(w_n^{(t)}) + \left\| \mathbf{h}\left(\mathbf{u}^*_{w_{n+1}^{(t)}}\right) - \mathbf{h}\left(\mathbf{u}^*_{w_n^{(t)}}\right) \right\| \tag{139}$$

or equivalently, $\hat{s}^{(t)}(w_{n+1}^{(t)}) := \hat{s}^{(t)}(w_n^{(t)}) + \left\| \mathbf{f}\left(\mathbf{x}^*_{w_{n+1}^{(t)}}\right) - \mathbf{f}\left(\mathbf{x}^*_{w_n^{(t)}}\right) \right\|$, then build interpolation for $w \in [0,1]$. Consequently, we denote $\hat{\Phi}(w) = \hat{s}(w)/\hat{s}(1)$.

Algorithm 1 may be implemented either in the original space $\mathcal{U}$, returning iterates $\mathbf{u}_n^{(t)}$, or directly in the hidden space $\mathcal{X}$, returning iterates $\mathbf{x}_n^{(t)}$ when g and f_m are explicitly available. Under Assumption 1, these two viewpoints are equivalent, and standard InnerSolver such as gradient-based methods can be used efficiently in either formulation [32]. In this section, we analyze the algorithm in the convex hidden space $\mathcal{X}$. Accordingly, when the InnerSolver returns iterates $\mathbf{u}_n^{(t)}$, we define their hidden-space images by $\mathbf{x}_n^{(t)} := g(\mathbf{u}_n^{(t)})$ and $\mathbf{x}_w^* := g(\mathbf{u}_w^*)$. If the algorithm is formulated directly in $\mathcal{X}$, then $\mathbf{x}_n^{(t)}$ simply denotes the returned iterate. Under Assumption 1, since $h_m(\mathbf{u}) = f_m(g(\mathbf{u}))$, we equivalently have $\mathbf{h}(\mathbf{u}_n^{(t)}) = \mathbf{f}(\mathbf{x}_n^{(t)}), \mathbf{h}(\mathbf{u}_w^*) = \mathbf{f}(\mathbf{x}_w^*)$. Therefore, all approximation and chord-length bounds below are stated in the $\mathbf{x}$-space.

We first prepare the following lemma that quantifies the optimization error and establishes the uniform smoothness claim for $\tilde{\Phi}_t$ in Theorem 1.

Lemma 11 (Inexact InnerSolver perturbation and Lipschitz-smoothness)**.** *Suppose Assumptions 1, 2, 3 hold. Consider Algorithm 1 with inexact InnerSolver solutions that return iterates $\mathbf{u}_n^{(t)}$, and define their convex-space images by $\mathbf{x}_n^{(t)} := g(\mathbf{u}_n^{(t)})$ or simply let $\mathbf{x}_n^{(t)}$ denote the returned iterate when the algorithm is formulated directly in the convex space $\mathcal{X}$. Assume $\max_n \|\mathbf{x}_n^{(t)} - \mathbf{x}^*_{w_n^{(t)}}\| \le \epsilon$ for all $t \ge 0$. Then,*

$$\left|\tilde{s}_t(w_n^{(t)}) - \hat{s}_t(w_n^{(t)})\right| = \mathcal{O}\left(N\epsilon\right), \tag{140}$$

where $\tilde{s}_t(w_n^{(t)})$ and $\hat{s}_t(w_n^{(t)})$ are generated via the inexact construction (12) *and the exact construction* (139)*, respectively. Additionally, suppose $\epsilon = \mathcal{O}\left(\frac{v_{\min}^2}{v_{\max}} N^{-1}\right)$. Then, $\tilde{\Phi}_t$ and Φ_t are uniformly $l_{\Phi_{\text{iter}},1}$-Lipschitz-smooth and*

$$\|\tilde{\Phi}_t - \hat{\Phi}_t\|_\infty \le \mathcal{O}\left(\frac{1}{s(1)} N\epsilon\right). \tag{141}$$

Proof. Fix t and n. Under Assumptions 1 and 2, the exact solution map $w \mapsto \mathbf{x}_w^*$ is continuous on $[0,1]$. Therefore $\mathcal{X}_{\text{PF}} := \{\mathbf{x}_w^* : w \in [0,1]\}$ is compact. Its ϵ-tube $\mathcal{X}_{\text{PF}_\epsilon} := \{\mathbf{x} : \text{dist}(\mathbf{x}, \mathcal{X}_{\text{PF}}) \le \epsilon\}$ is also compact. Thus, for each $m \in \{1,2\}$, $G_m := \sup_{\mathbf{x} \in \mathcal{X}_{\text{PF}_\epsilon}} \|\nabla f_m(\mathbf{x})\| < \infty$. Let $G_f := \sqrt{G_1^2 + G_2^2}$. By the mean value theorem and $\max_n \|\mathbf{x}_n^{(t)} - \mathbf{x}^*_{w_n^{(t)}}\| \le \epsilon$,

$$\|\mathbf{f}(\mathbf{x}_n^{(t)}) - \mathbf{f}_{\text{PF}}(w_n^{(t)})\| = \|\mathbf{f}(\mathbf{x}_n^{(t)}) - \mathbf{f}(\mathbf{x}^*_{w_n^{(t)}})\| \le G_f \|\mathbf{x}_n^{(t)} - \mathbf{x}^*_{w_n^{(t)}}\| \le G_f \epsilon. \tag{142}$$

Now apply the reverse triangle inequality to each segment:

$$\begin{aligned} &\left| \|\mathbf{f}(\mathbf{x}_{n+1}^{(t)}) - \mathbf{f}(\mathbf{x}_n^{(t)})\| - \|\mathbf{f}_{\text{PF}}(w_{n+1}^{(t)}) - \mathbf{f}_{\text{PF}}(w_n^{(t)})\| \right| \\ \le &\|\mathbf{f}(\mathbf{x}_{n+1}^{(t)}) - \mathbf{f}_{\text{PF}}(w_{n+1}^{(t)})\| + \|\mathbf{f}(\mathbf{x}_n^{(t)}) - \mathbf{f}_{\text{PF}}(w_n^{(t)})\| \\ \le &2G_f\epsilon = \mathcal{O}(\epsilon). \end{aligned} \tag{143}$$

Summing (143) over the first m segments yields $\left|\tilde{s}_t(w_m^{(t)}) - \hat{s}_t(w_m^{(t)})\right| \le m \cdot \mathcal{O}(\epsilon) = \mathcal{O}(N\epsilon)$ for all $m = 0, \dots, N$, proving (140).

We view Algorithm 2 as applied on the uniform coordinate $\Phi_t(w)$, i.e., on the knots $\{\Phi_t(w_n^{(t)})\}_{n=0}^N = \{n/N\}_{n=0}^N$, and then pulled back to w through Φ_t. Since $\hat{s}_t$ is built from exact sampled cumulative chord lengths, Lemma 6 and the derivative bounds on Φ_t give

$$\frac{2}{5}\frac{v_{\min}^2}{v_{\max}} \le \frac{\hat{s}_t(w_{n+1}^{(t)}) - \hat{s}_t(w_n^{(t)})}{1/N} \le 2\sqrt{2}\frac{v_{\max}^2}{v_{\min}}, \quad n = 0, \dots, N-1. \tag{144}$$

For every $n = 0, \dots, N-1$, by the reverse triangle inequality and (142),

$$\begin{aligned}&\left|\frac{\tilde{s}_t(w_{n+1}^{(t)}) - \tilde{s}_t(w_n^{(t)})}{1/N} - \frac{\hat{s}_t(w_{n+1}^{(t)}) - \hat{s}_t(w_n^{(t)})}{1/N}\right| \\ \le& N\Big(\|\mathbf{f}(\mathbf{x}_{n+1}^{(t)}) - \mathbf{f}_{\mathrm{PF}}(w_{n+1}^{(t)})\| + \|\mathbf{f}(\mathbf{x}_n^{(t)}) - \mathbf{f}_{\mathrm{PF}}(w_n^{(t)})\|\Big) = \mathcal{O}(N\epsilon).\end{aligned} \tag{145}$$

Choosing some $\epsilon = \mathcal{O}\left(\frac{v_{\min}^2}{v_{\max}}N^{-1}\right)$ gives $\left|\frac{\tilde{s}_t(w_{n+1}^{(t)}) - \tilde{s}_t(w_n^{(t)})}{1/N} - \frac{\hat{s}_t(w_{n+1}^{(t)}) - \hat{s}_t(w_n^{(t)})}{1/N}\right| \le \frac{1}{5}\frac{v_{\min}^2}{v_{\max}}$ so combining this with (144) yields

$$\frac{1}{5}\frac{v_{\min}^2}{v_{\max}} \le \frac{\tilde{s}_t(w_{n+1}^{(t)}) - \tilde{s}_t(w_n^{(t)})}{1/N} \le 3\sqrt{2}\frac{v_{\max}^2}{v_{\min}}, \quad n = 0, \dots, N-1. \tag{146}$$

Thus the exact and inexact nodal data both stay in the same strictly increasing PCHIP regime. Let $\hat{m}_n$ and $\tilde{m}_n$ be the PCHIP slopes generated by Algorithm 2 from the exact and inexact nodal data in the coordinate $\Phi_t(w)$. On the positive-slope branch, the interior PCHIP slope is a smooth function of neighboring positive segment slopes, and the endpoint formulas with monotonicity clipping are Lipschitz. Hence there exists $C_m = \mathcal{O}(1)$ such that

$$\max_{0\le n\le N}|\tilde{m}_n - \hat{m}_n| \le C_m \max_{0\le n\le N-1}\left|\frac{\tilde{s}_t(w_{n+1}^{(t)}) - \tilde{s}_t(w_n^{(t)})}{1/N} - \frac{\hat{s}_t(w_{n+1}^{(t)}) - \hat{s}_t(w_n^{(t)})}{1/N}\right|. \tag{147}$$

Now fix any $w \in [0,1]$ and choose n such that $\Phi_t(w) \in [n/N, (n+1)/N]$. Let $q = \Phi_t(w)$ and $\theta = N(q - n/N) \in [0,1]$. By Algorithm 2, applied in the coordinate $\Phi_t(w)$ and pulled back through Φ_t, we have

$$\tilde{s}_t(w) = H_{00}(\theta)\tilde{s}_t(w_n^{(t)}) + H_{10}(\theta)\frac{1}{N}\tilde{m}_n + H_{01}(\theta)\tilde{s}_t(w_{n+1}^{(t)}) + H_{11}(\theta)\frac{1}{N}\tilde{m}_{n+1}, \tag{148}$$

$$\hat{s}_t(w) = H_{00}(\theta)\hat{s}_t(w_n^{(t)}) + H_{10}(\theta)\frac{1}{N}\hat{m}_n + H_{01}(\theta)\hat{s}_t(w_{n+1}^{(t)}) + H_{11}(\theta)\frac{1}{N}\hat{m}_{n+1}. \tag{149}$$

Since the Hermite basis functions in Algorithm 2 are uniformly bounded on $[0,1]$, there exists $C_H = \mathcal{O}(1)$ such that

$$|\tilde{s}_t(w) - \hat{s}_t(w)| \le C_H\left(\max_{0\le j\le N}|\tilde{s}_t(w_j^{(t)}) - \hat{s}_t(w_j^{(t)})| + \frac{1}{N}\max_{0\le j\le N}|\tilde{m}_j - \hat{m}_j|\right). \tag{150}$$

Moreover, by the definition of the segment slopes used in Algorithm 2,

$$\left|\frac{\tilde{s}_t(w_{j+1}^{(t)}) - \tilde{s}_t(w_j^{(t)})}{1/N} - \frac{\hat{s}_t(w_{j+1}^{(t)}) - \hat{s}_t(w_j^{(t)})}{1/N}\right| \le 2N\max_{0\le k\le N}|\tilde{s}_t(w_k^{(t)}) - \hat{s}_t(w_k^{(t)})|. \tag{151}$$

Substituting (147) and (151) into (150), and taking the supremum over $w \in [0,1]$, yields

$$\|\tilde{s}_t - \hat{s}_t\|_\infty \le C_{\mathrm{int}}\max_{0\le j\le N}|\tilde{s}_t(w_j^{(t)}) - \hat{s}_t(w_j^{(t)})|, \tag{152}$$

where $C_{\mathrm{int}} = \mathcal{O}(1)$. By Lemma 11, the sampled cumulative chord-length perturbation satisfies $\max_{0\le j\le N}|\tilde{s}_t(w_j^{(t)}) - \hat{s}_t(w_j^{(t)})| = \mathcal{O}(N\epsilon)$, and therefore $\|\tilde{s}_t - \hat{s}_t\|_\infty = \mathcal{O}(N\epsilon)$.

It remains to normalize. Since $\hat{s}_t(1)$ is the exact cumulative chord length, Lemma 5 gives $\hat{s}_t(1) \geq s(1)/\sqrt{2}$. Also, by Lemma 11 and the local choice of ϵ, $\tilde{s}_t(1) \geq s(1)/(2\sqrt{2})$. For any $w \in [0,1]$,

$$\begin{aligned}|\tilde{\Phi}_t(w) - \hat{\Phi}_t(w)| = \left|\frac{\tilde{s}_t(w)}{\tilde{s}_t(1)} - \frac{\hat{s}_t(w)}{\hat{s}_t(1)}\right| \leq & \frac{|\tilde{s}_t(w) - \hat{s}_t(w)|}{\tilde{s}_t(1)} + \hat{s}_t(w)\left|\frac{1}{\tilde{s}_t(1)} - \frac{1}{\hat{s}_t(1)}\right| \\ \leq & \frac{C}{s(1)}\|\tilde{s}_t - \hat{s}_t\|_\infty = \mathcal{O}\left(\frac{N\epsilon}{s(1)}\right). \end{aligned} \tag{153}$$

Taking the supremum over $w \in [0,1]$ proves (141).

Finally, for the Lipschitz-smoothness claim, the bounds (146) show that the inexact data remain in the strictly positive-slope regime, and it remains to verify the discrete slope-variation condition required by Lemma 7. The exact segment slopes satisfy this condition by Lemma 6. For the inexact segment slopes, (145) gives an $\mathcal{O}(N\epsilon)$ perturbation, and after normalization by $\tilde{s}_t(1)$ this contributes $\mathcal{O}(N\epsilon/s(1))$ to the normalized segment slopes. Hence, under the local condition $\epsilon = \mathcal{O}\left(\frac{v_{\min}^2}{v_{\max}}N^{-1}\right)$ with a sufficiently small hidden constant, the positive-slope and discrete slope-variation conditions of Lemma 7 continue to hold with uniform constants. Therefore, the PCHIP interpolant remains strictly increasing and uniformly $l_{\Phi_{\text{iter}},1}$-Lipschitz-smooth. Pulling the interpolant back through the known map Φ_t preserves the stated derivative bounds, which completes the proof. □

In this way, we first give an idealized inexact-solver result: assuming the inner iterates remain uniformly within an ϵ-tube around the exact scalarized solutions, which practically can be achieved with large K choice for InnerSolver, SURF retains the exact-solver contraction up to an additional perturbation floor.

Theorem 3 (SURF convergence with inexact InnerSolver)**.** *Suppose Assumptions 1, 2, 3 hold. Suppose InnerSolver returns* $\mathbf{x}_n^{(t)}$ *satisfying* $\max_n \|\mathbf{x}_n^{(t)} - \mathbf{x}^*_{w_n^{(t)}}\| \leq \epsilon = \mathcal{O}\left(\frac{v_{\min}^2}{v_{\max}}N^{-1}\right)$ *for all* t *and it chooses* $N = \Omega(\kappa^{1.5})$*, with* $\kappa := \frac{4\sqrt{2}v_{\max}}{v_{\min}}\sqrt{l_{\Phi_{\text{iter}},1} + l_{\Phi,1}}$*. Then Algorithm 1 converges to a floor, i.e.,*

$$\|\Phi_t - \Phi\|_\infty \leq \left(1 - \frac{\alpha}{2}\right)^t \|\Phi_0 - \Phi\|_\infty + \mathcal{O}\left(\frac{1}{N^2} + \frac{N\epsilon}{s(1)}\right), \quad \forall t \geq 1. \tag{154}$$

Proof. From the update rule $\Phi_{t+1} = (1-\alpha)\Phi_t + \alpha\tilde{\Phi}_t$, we have

$$\begin{aligned}\|\Phi_{t+1} - \Phi\|_\infty \leq & (1-\alpha)\|\Phi_t - \Phi\|_\infty + \alpha\|\tilde{\Phi}_t - \Phi\|_\infty \\ \leq & (1-\alpha)\|\Phi_t - \Phi\|_\infty + \alpha\|\hat{\Phi}_t - \Phi\|_\infty + \alpha\|\tilde{\Phi}_t - \hat{\Phi}_t\|_\infty. \end{aligned} \tag{155}$$

By Lemma 11, when $\epsilon = \mathcal{O}\left(\frac{v_{\min}^2}{v_{\max}}N^{-1}\right)$, we have $\|\tilde{\Phi}_t - \hat{\Phi}_t\|_\infty \leq \mathcal{O}\left(\frac{N\epsilon}{s(1)}\right)$. Substituting this into (155) gives

$$\|\Phi_{t+1} - \Phi\|_\infty \leq (1-\alpha)\|\Phi_t - \Phi\|_\infty + \alpha\|\hat{\Phi}_t - \Phi\|_\infty + \alpha\mathcal{O}\left(\frac{N\epsilon}{s(1)}\right). \tag{156}$$

Now, $\hat{\Phi}_t$ is the rebuilt CDF obtained from the exact PF points on the same sampled weights $\{w_n^{(t)}\}_{n=0}^N$. Hence the exact-case argument from the proof of (135) applies verbatim. In particular, under the same condition on N as required there, we have

$$\begin{aligned}&(1-\alpha)\|\Phi_t - \Phi\|_\infty + \alpha\|\hat{\Phi}_t - \Phi\|_\infty \\ \leq & \left(1 - \frac{\alpha}{2}\right)\|\Phi_t - \Phi\|_\infty + \alpha\frac{\frac{\sqrt{2}}{4}\left(\frac{s(1)}{v_{\min}}\right)^2(3l_{\Phi_{\text{iter}},1} + l_{\Phi,1}) + \frac{\sqrt{2}}{3}l_{\tilde{\mathbf{f}}_{\text{PF}},1}^2 s(1)^2}{N^2} + \alpha\mathcal{O}\left(\frac{1}{N^4}\right). \end{aligned} \tag{157}$$

Combining (156) and (157), we obtain

$$\begin{aligned}\|\Phi_{t+1} - \Phi\|_\infty \leq & \left(1 - \frac{\alpha}{2}\right)\|\Phi_t - \Phi\|_\infty + \alpha\frac{\frac{\sqrt{2}}{4}\left(\frac{s(1)}{v_{\min}}\right)^2(3l_{\Phi_{\text{iter}},1} + l_{\Phi,1}) + \frac{\sqrt{2}}{3}l_{\tilde{\mathbf{f}}_{\text{PF}},1}^2 s(1)^2}{N^2} \\ & + \alpha\mathcal{O}\left(\frac{N\epsilon}{s(1)} + \frac{1}{N^4}\right). \end{aligned} \tag{158}$$

Iterating (158) for $t \geq 1$ proves the claim. $\square$

We now turn to the proof of Theorem 1, where the InnerSolver is warm-started and run for a finite number of steps. The next lemma couples the outer CDF recursion with the optimization error induced by the inexact InnerSolver.

Lemma 12 (Coupled recursion)**.** *Suppose Assumptions 1, 2, 3 hold. Consider Algorithm 1 with an inexact InnerSolver that returns iterates $\mathbf{u}_n^{(t)}$, and define their hidden-space images by $\mathbf{x}_n^{(t)} := g(\mathbf{u}_n^{(t)})$ or simply let $\mathbf{x}_n^{(t)}$ denote the returned iterate when the algorithm is formulated directly in the convex space $\mathcal{X}$. Assume moreover that the inner solver is warm-started from the previous outer iterate and that, after K inner steps, it satisfies*

$$\|\mathbf{x}_n^{(t+1)} - \mathbf{x}^*_{w_n^{(t+1)}}\| \leq \eta_K \|\mathbf{x}_n^{(t)} - \mathbf{x}^*_{w_n^{(t+1)}}\|, \quad \forall t, n, \tag{159}$$

where $\{\eta_K\}_{K\geq 1}$ satisfies $\eta_K \downarrow 0$ as $K \to \infty$. Then there exist positive constants $C_1, C_2, C_3, C_4 < \infty$, independent of t, N, and K, such that for all $t \geq 0$,

$$\|\Phi_{t+1} - \Phi\|_\infty \leq \left(1 - \tfrac{\alpha}{2}\right)\|\Phi_t - \Phi\|_\infty + C_1\alpha\frac{1}{N^2} + C_2\alpha\frac{N}{s(1)}\max_n \|\mathbf{x}_n^{(t)} - \mathbf{x}^*_{w_n^{(t)}}\|, \tag{160}$$

$$\max_n \|\mathbf{x}_n^{(t+1)} - \mathbf{x}^*_{w_n^{(t+1)}}\| \leq \eta_K \max_n \|\mathbf{x}_n^{(t)} - \mathbf{x}^*_{w_n^{(t)}}\| + C_3\eta_K \max_n |w_n^{(t+1)} - w_n^{(t)}|, \tag{161}$$

$$\max_n |w_n^{(t+1)} - w_n^{(t)}| \leq C_4\alpha\|\Phi_t - \Phi\|_\infty + C_4\alpha\frac{1}{N^2} + C_4\alpha\frac{N}{s(1)}\max_n \|\mathbf{x}_n^{(t)} - \mathbf{x}^*_{w_n^{(t)}}\|. \tag{162}$$

Proof. We first prove (160). Let $\hat{\Phi}_t$ denote the rebuilt CDF obtained from the exact PF points $\{\mathbf{f}_{\mathrm{PF}}(w_n^{(t)})\}_{n=0}^N$, and let $\tilde{\Phi}_t$ denote the rebuilt CDF obtained from the inexact points $\{\mathbf{f}(\mathbf{x}_n^{(t)})\}_{n=0}^N$.

By the same argument as in Lemma 11, for every n, $\|\mathbf{f}(\mathbf{x}_n^{(t)}) - \mathbf{f}_{\mathrm{PF}}(w_n^{(t)})\| \leq G_f\|\mathbf{x}_n^{(t)} - \mathbf{x}^*_{w_n^{(t)}}\|$. Repeating the proofs of Lemma 11 with ϵ replaced by $\max_n \|\mathbf{x}_n^{(t)} - \mathbf{x}^*_{w_n^{(t)}}\|$, and using that PCHIP is applied in the coordinate $q = \Phi_t(w)$, gives

$$\|\tilde{\Phi}_t - \hat{\Phi}_t\|_\infty \leq C_2\frac{N}{s(1)}\max_n \|\mathbf{x}_n^{(t)} - \mathbf{x}^*_{w_n^{(t)}}\|. \tag{163}$$

Now, using the update rule $\Phi_{t+1} = (1-\alpha)\Phi_t + \alpha\tilde{\Phi}_t$, we have

$$\begin{aligned}\|\Phi_{t+1} - \Phi\|_\infty &\leq (1-\alpha)\|\Phi_t - \Phi\|_\infty + \alpha\|\tilde{\Phi}_t - \Phi\|_\infty \\ &\leq (1-\alpha)\|\Phi_t - \Phi\|_\infty + \alpha\|\hat{\Phi}_t - \Phi\|_\infty + \alpha\|\tilde{\Phi}_t - \hat{\Phi}_t\|_\infty. \end{aligned}\tag{164}$$

Substituting (163), and using the same exact-rebuild estimate as in the proof of (135) proves (160):

$$(1-\alpha)\|\Phi_t - \Phi\|_\infty + \alpha\|\hat{\Phi}_t - \Phi\|_\infty \leq \left(1 - \frac{\alpha}{2}\right)\|\Phi_t - \Phi\|_\infty + C_1\alpha\frac{1}{N^2}. \tag{165}$$

Next we prove (161). By the assumed warm-started K-step contraction (159), $\|\mathbf{x}_n^{(t+1)} - \mathbf{x}^*_{w_n^{(t+1)}}\| \leq \eta_K\|\mathbf{x}_n^{(t)} - \mathbf{x}^*_{w_n^{(t+1)}}\|$. Adding and subtracting $\mathbf{x}^*_{w_n^{(t)}}$ gives

$$\|\mathbf{x}_n^{(t+1)} - \mathbf{x}^*_{w_n^{(t+1)}}\| \leq \eta_K\|\mathbf{x}_n^{(t)} - \mathbf{x}^*_{w_n^{(t)}}\| + \eta_K\|\mathbf{x}^*_{w_n^{(t)}} - \mathbf{x}^*_{w_n^{(t+1)}}\|. \tag{166}$$

By Lemma 2, the map $w \mapsto \mathbf{x}^*_w$ is continuously differentiable on the compact interval $[0,1]$, hence Lipschitz. Therefore there exists a constant $C_3 > 0$ such that $\|\mathbf{x}^*_{w_n^{(t)}} - \mathbf{x}^*_{w_n^{(t+1)}}\| \leq C_3|w_n^{(t+1)} - w_n^{(t)}|$. Taking maximum over n proves (161).

Finally we prove (162). Since $w_n^{(t)} = \Phi_t^{-1}(n/N)$, the inverse stability bound from the derivative lower bound on Φ_t implies $\max_n |w_n^{(t+1)} - w_n^{(t)}| \leq C\|\Phi_{t+1} - \Phi_t\|_\infty$ for some constant $C > 0$. Also, $\Phi_{t+1} - \Phi_t = \alpha(\tilde{\Phi}_t - \Phi_t)$, hence

$$\begin{aligned}\|\Phi_{t+1} - \Phi_t\|_\infty &\leq \alpha\|\tilde{\Phi}_t - \Phi\|_\infty + \alpha\|\Phi_t - \Phi\|_\infty \\ &\leq \alpha\|\hat{\Phi}_t - \Phi\|_\infty + \alpha\|\tilde{\Phi}_t - \hat{\Phi}_t\|_\infty + \alpha\|\Phi_t - \Phi\|_\infty. \end{aligned}\tag{167}$$

From the proof of (135), the exact rebuilt CDF satisfies $\|\hat{\Phi}_t - \Phi\|_\infty \le \frac{1}{2}\|\Phi_t - \Phi\|_\infty + C\frac{1}{N^2}$, where the exact-case constants are absorbed into C. Substituting this and (163) into (167), we obtain

$$\|\Phi_{t+1} - \Phi_t\|_\infty \le C\alpha\|\Phi_t - \Phi\|_\infty + C\alpha\frac{1}{N^2} + C\alpha\frac{N}{s(1)}\max_n \|\mathbf{x}_n^{(t)} - \mathbf{x}^*_{w_n^{(t)}}\|. \tag{168}$$

Absorbing the inverse-stability constant into C_4 proves (162). □

We are now ready to finish proving Theorem 1. Combining the preceding inexact-CDF perturbation bound with the coupled outer-inner recursion yields the following more general inexact convergence result. Theorem 1 follows directly from the regime (i) while regime (ii) provides additional analysis.

Theorem 4 (SURF convergence under inexact InnerSolver)**.** *Suppose Assumptions 1, 2, 3 hold. Use the notation of Lemma 12. Choose $N = \Omega(\kappa^{1.5})$ with $\kappa := \frac{4\sqrt{2}v_{\max}}{v_{\min}}\sqrt{l_{\Phi_{\text{iter}},1} + l_{\Phi,1}}$. The following two regimes apply.*

(i) Stronger inner contraction, allowing a constant outer step. *Choose $\alpha \le 1/2$ and K such that*

$$\eta_K = \mathcal{O}\left(\min\left\{\frac{v_{\min}^2}{v_{\max}}N^{-1}, \frac{s(1)}{\alpha N}\right\}\right) = \mathcal{O}(N^{-1}), \tag{169}$$

then there exists $C = \left(\frac{4C_2}{s(1)} + \frac{v_{\max}}{v_{\min}^2}\right)N = \Theta(N)$ such that

$$\begin{aligned}&\|\Phi_t - \Phi\|_\infty + C\max_n \|\mathbf{x}_n^{(t)} - \mathbf{x}^*_{w_n^{(t)}}\| \\ &\le \left(1 - \frac{\alpha}{4}\right)^t \left(\|\Phi_0 - \Phi\|_\infty + C\max_n \|\mathbf{x}_n^{(0)} - \mathbf{x}^*_{w_n^{(0)}}\|\right) + \mathcal{O}\left(\frac{1}{N^2}\right).\end{aligned} \tag{170}$$

Consequently, $\max_n \|\mathbf{x}_n^{(t)} - \mathbf{x}^*_{w_n^{(t)}}\| = \mathcal{O}\left(\frac{v_{\min}^2}{v_{\max}}N^{-1}\right)$ for all $t \ge 0$, and after the transient term is below the floor, $\|\Phi_t - \Phi\|_\infty = \mathcal{O}(N^{-2})$ and $\max_n \|\mathbf{x}_n^{(t)} - \mathbf{x}^*_{w_n^{(t)}}\| = \mathcal{O}(N^{-3})$.

(ii) Smaller outer step with warmup. *Use warmup initialization to ensure $\max_n \|\mathbf{x}_n^{(0)} - \mathbf{x}^*_{w_n^{(0)}}\| = \mathcal{O}\left(\frac{v_{\min}^2}{v_{\max}}N^{-1}\right)$, choose $\alpha = \Theta(N^{-1})$ and K such that*

$$\eta_K \le \min\left\{\frac{1}{2}\left(1 + C_3C_4\alpha\frac{N}{s(1)}\right)^{-1}, \frac{1}{4}\left(\left(1 + \frac{4C_2\alpha N}{s(1)}\right)C_3C_4\right)^{-1}\right\} = \mathcal{O}(1), \tag{171}$$

then there exists $C = 1 + \frac{4C_2\alpha N}{s(1)}$ such that

$$\begin{aligned}&\|\Phi_t - \Phi\|_\infty + C\max_n \|\mathbf{x}_n^{(t)} - \mathbf{x}^*_{w_n^{(t)}}\| \\ &\le \left(1 - \frac{\alpha}{4}\right)^t \left(\|\Phi_0 - \Phi\|_\infty + C\max_n \|\mathbf{x}_n^{(0)} - \mathbf{x}^*_{w_n^{(0)}}\|\right) + \mathcal{O}\left(\frac{1}{N^2}\right).\end{aligned} \tag{172}$$

Consequently, once the transient term is below the discretization floor, $\|\Phi_t - \Phi\|_\infty = \mathcal{O}(N^{-2})$. The inner tracking recursion then contracts to the floor $\mathcal{O}(\alpha\eta_K N^{-2})$, so after $\mathcal{O}(\log N)$ additional outer iterations, $\max_n \|\mathbf{x}_n^{(t)} - \mathbf{x}^*_{w_n^{(t)}}\| = \mathcal{O}(\alpha\eta_K N^{-2}) = \mathcal{O}(N^{-3})$.

Proof. By Lemma 12, substituting (162) into (161) gives

$$\begin{aligned}&\max_n \|\mathbf{x}_n^{(t+1)} - \mathbf{x}^*_{w_n^{(t+1)}}\| \\ \le&\eta_K\left(1 + C_3C_4\alpha\frac{N}{s(1)}\right)\max_n \|\mathbf{x}_n^{(t)} - \mathbf{x}^*_{w_n^{(t)}}\| + C_3C_4\alpha\eta_K\|\Phi_t - \Phi\|_\infty + C_3C_4\alpha\eta_K\frac{1}{N^2}.\end{aligned} \tag{173}$$

For regime (i), choose $C \geq \left(\frac{4C_2}{s(1)} + \frac{v_{\max}}{v_{\min}^2}\right) N$, so $C = \Theta(N)$. Multiplying (173) by C and adding (160) gives

$$\begin{aligned}
&\|\Phi_{t+1} - \Phi\|_\infty + C \max_n \|\mathbf{x}_n^{(t+1)} - \mathbf{x}^*_{w_n^{(t+1)}}\| \\
&\leq \left(1 - \frac{\alpha}{2} + CC_3C_4\alpha\eta_K\right) \|\Phi_t - \Phi\|_\infty \\
&\quad + \left(C_2\alpha\frac{N}{s(1)} + C\eta_K\left(1 + C_3C_4\alpha\frac{N}{s(1)}\right)\right) \max_n \|\mathbf{x}_n^{(t)} - \mathbf{x}^*_{w_n^{(t)}}\| \\
&\quad + \alpha\left(C_1 + CC_3C_4\eta_K\right)\frac{1}{N^2}.
\end{aligned} \tag{174}$$

The condition (169) can be chosen with a sufficiently small hidden constant so that $CC_3C_4\eta_K \leq 1/4$ and $\eta_K\left(1 + C_3C_4\alpha\frac{N}{s(1)}\right) \leq 1 - \alpha/2$. Since $C \geq 4C_2N/s(1)$, we also have $C_2\alpha N/s(1) \leq \alpha C/4$. Therefore,

$$\begin{aligned}
&\|\Phi_{t+1} - \Phi\|_\infty + C \max_n \|\mathbf{x}_n^{(t+1)} - \mathbf{x}^*_{w_n^{(t+1)}}\| \\
&\leq \left(1 - \frac{\alpha}{4}\right) \|\Phi_t - \Phi\|_\infty + \left(\frac{\alpha}{4} + 1 - \frac{\alpha}{2}\right) C \max_n \|\mathbf{x}_n^{(t)} - \mathbf{x}^*_{w_n^{(t)}}\| + \alpha\mathcal{O}\left(\frac{1}{N^2}\right) \\
&= \left(1 - \frac{\alpha}{4}\right) \|\Phi_t - \Phi\|_\infty + \left(1 - \frac{\alpha}{4}\right) C \max_n \|\mathbf{x}_n^{(t)} - \mathbf{x}^*_{w_n^{(t)}}\| + \alpha\mathcal{O}\left(\frac{1}{N^2}\right) \\
&= \left(1 - \frac{\alpha}{4}\right)\left(\|\Phi_t - \Phi\|_\infty + C \max_n \|\mathbf{x}_n^{(t)} - \mathbf{x}^*_{w_n^{(t)}}\|\right) + \alpha\mathcal{O}\left(\frac{1}{N^2}\right).
\end{aligned} \tag{175}$$

Here the first inequality uses the two bounds on η_K and the choice $C \geq 4C_2N/s(1)$, and the last two lines only collect terms. Iterating (175) gives (170). Since $\|\Phi_0 - \Phi\|_\infty = \mathcal{O}(1)$ and $C\max_n \|\mathbf{x}_n^{(0)} - \mathbf{x}^*_{w_n^{(0)}}\| = \mathcal{O}(1)$, the bound implies $\max_n \|\mathbf{x}_n^{(t)} - \mathbf{x}^*_{w_n^{(t)}}\| = \mathcal{O}\left(\frac{v_{\min}^2}{v_{\max}}N^{-1}\right)$ for all $t \geq 0$. After the transient term is below $\mathcal{O}(N^{-2})$, (170) gives $\|\Phi_t - \Phi\|_\infty = \mathcal{O}(N^{-2})$ and $C\max_n \|\mathbf{x}_n^{(t)} - \mathbf{x}^*_{w_n^{(t)}}\| = \mathcal{O}(N^{-2})$, hence $\max_n \|\mathbf{x}_n^{(t)} - \mathbf{x}^*_{w_n^{(t)}}\| = \mathcal{O}(N^{-3})$.

For regime (ii), choose $C \geq 1 + \frac{4C_2\alpha N}{s(1)}$, so $C = \Theta(1)$ because $\alpha = \mathcal{O}(N^{-1})$. Multiplying (173) by C and adding (160) gives

$$\begin{aligned}
&\|\Phi_{t+1} - \Phi\|_\infty + C \max_n \|\mathbf{x}_n^{(t+1)} - \mathbf{x}^*_{w_n^{(t+1)}}\| \\
&\leq \left(1 - \frac{\alpha}{2} + CC_3C_4\alpha\eta_K\right) \|\Phi_t - \Phi\|_\infty + \alpha\left(C_1 + CC_3C_4\eta_K\right)\frac{1}{N^2} \\
&\quad + \left(C_2\alpha\frac{N}{s(1)} + C\eta_K\left(1 + C_3C_4\alpha\frac{N}{s(1)}\right)\right) \max_n \|\mathbf{x}_n^{(t)} - \mathbf{x}^*_{w_n^{(t)}}\|
\end{aligned} \tag{176}$$

The condition (171) gives $CC_3C_4\eta_K \leq 1/4$ and $\eta_K\left(1 + C_3C_4\alpha\frac{N}{s(1)}\right) \leq 1/2$. Since $C \geq 4C_2\alpha N/s(1)$, we have $C_2\alpha N/s(1) \leq C/4$. Therefore,

$$\begin{aligned}
&\|\Phi_{t+1} - \Phi\|_\infty + C \max_n \|\mathbf{x}_n^{(t+1)} - \mathbf{x}^*_{w_n^{(t+1)}}\| \\
&\leq \left(1 - \frac{\alpha}{4}\right) \|\Phi_t - \Phi\|_\infty + \frac{3}{4}C \max_n \|\mathbf{x}_n^{(t)} - \mathbf{x}^*_{w_n^{(t)}}\| + \alpha\mathcal{O}\left(\frac{1}{N^2}\right) \\
&\leq \left(1 - \frac{\alpha}{4}\right) \|\Phi_t - \Phi\|_\infty + \left(1 - \frac{\alpha}{4}\right) C \max_n \|\mathbf{x}_n^{(t)} - \mathbf{x}^*_{w_n^{(t)}}\| + \alpha\mathcal{O}\left(\frac{1}{N^2}\right) \\
&= \left(1 - \frac{\alpha}{4}\right)\left(\|\Phi_t - \Phi\|_\infty + C \max_n \|\mathbf{x}_n^{(t)} - \mathbf{x}^*_{w_n^{(t)}}\|\right) + \alpha\mathcal{O}\left(\frac{1}{N^2}\right).
\end{aligned} \tag{177}$$

Here the first inequality uses (171) and $C \geq 4C_2\alpha N/s(1)$, the second uses $\alpha \leq 1$, and the last line collects terms. Iterating (177) gives (172). Once the transient term in (172) is below $\mathcal{O}(N^{-2})$, we have $\|\Phi_t - \Phi\|_\infty = \mathcal{O}(N^{-2})$. Substituting this into (173) and using (171) gives

$$\max_n \|\mathbf{x}_n^{(t+1)} - \mathbf{x}^*_{w_n^{(t+1)}}\| \leq \frac{1}{2} \max_n \|\mathbf{x}_n^{(t)} - \mathbf{x}^*_{w_n^{(t)}}\| + \mathcal{O}(\alpha\eta_K N^{-2}). \tag{178}$$

Thus, after an additional logarithmic transient, $\max_n \|\mathbf{x}_n^{(t)} - \mathbf{x}^*_{w_n^{(t)}}\| = \mathcal{O}(\alpha\eta_K N^{-2}) = \mathcal{O}(N^{-3})$, where the last equality uses $\alpha = \mathcal{O}(N^{-1})$ and $\eta_K = \mathcal{O}(1)$. This completes the proof. □

Additionally, we provide the connection between CDF convergence and PF uniformity metrics CV.

E.3 Connection between SURF Convergence and PF Uniformity

Lemma 13 (CDF error controls PF spacing CV)**.** *Let $w_n^{(t)} = \Phi_t^{-1}(n/N)$ for $n = 0, \dots, N$, and let $\tilde{\Phi}_t$ be the normalized cumulative chord-length CDF constructed from the sampled PF points $\{\mathbf{h}(\mathbf{u}_n^{(t)})\}_{n=0}^N$ as in* (12)*. Then, for all $t \geq 0$,*

$$\mathrm{CV} := \frac{\mathrm{std}\big(\{\|\mathbf{h}(\mathbf{u}_{n+1}^{(t)}) - \mathbf{h}(\mathbf{u}_n^{(t)})\|\}_{n=0}^{N-1}\big)}{\mathrm{mean}\big(\{\|\mathbf{h}(\mathbf{u}_{n+1}^{(t)}) - \mathbf{h}(\mathbf{u}_n^{(t)})\|\}_{n=0}^{N-1}\big)} \leq 2N\|\tilde{\Phi}_t - \Phi_t\|_\infty. \tag{179}$$

Moreover, under the discretization and inexact-InnerSolver perturbation bounds used in Theorem 1, if $\max_{0\leq n\leq N} \|\mathbf{x}_n^{(t)} - \mathbf{x}^*_{w_n^{(t)}}\| = \mathcal{O}\big(N^{-1}\|\Phi_t - \Phi\|_\infty\big)$, then

$$\mathrm{CV} = \mathcal{O}\left(N^{-1} + N\|\Phi_t - \Phi\|_\infty\right). \tag{180}$$

Consequently, under the conditions of Theorem 1, for large enough t we have $\mathrm{CV} = \mathcal{O}(N^{-1})$.

Proof. By construction of $\tilde{\Phi}_t$ from normalized cumulative chord length, each CDF increment equals the corresponding normalized PF segment length:

$$\tilde{\Phi}_t(w_{n+1}^{(t)}) - \tilde{\Phi}_t(w_n^{(t)}) = \frac{\|\mathbf{h}(\mathbf{u}_{n+1}^{(t)}) - \mathbf{h}(\mathbf{u}_n^{(t)})\|}{\sum_{m=0}^{N-1} \|\mathbf{h}(\mathbf{u}_{m+1}^{(t)}) - \mathbf{h}(\mathbf{u}_m^{(t)})\|}. \tag{181}$$

Since the mean segment length is $\frac{1}{N}\sum_{m=0}^{N-1} \|\mathbf{h}(\mathbf{u}_{m+1}^{(t)}) - \mathbf{h}(\mathbf{u}_m^{(t)})\|$, the CV can be rewritten as

$$\begin{aligned}
\mathrm{CV} &= \frac{\sqrt{\frac{1}{N}\sum_{n=0}^{N-1}\left(\|\mathbf{h}(\mathbf{u}_{n+1}^{(t)}) - \mathbf{h}(\mathbf{u}_n^{(t)})\| - \frac{1}{N}\sum_{m=0}^{N-1}\|\mathbf{h}(\mathbf{u}_{m+1}^{(t)}) - \mathbf{h}(\mathbf{u}_m^{(t)})\|\right)^2}}{\frac{1}{N}\sum_{m=0}^{N-1}\|\mathbf{h}(\mathbf{u}_{m+1}^{(t)}) - \mathbf{h}(\mathbf{u}_m^{(t)})\|} \\
&= \sqrt{N\sum_{n=0}^{N-1}\left(\frac{\|\mathbf{h}(\mathbf{u}_{n+1}^{(t)}) - \mathbf{h}(\mathbf{u}_n^{(t)})\|}{\sum_{m=0}^{N-1}\|\mathbf{h}(\mathbf{u}_{m+1}^{(t)}) - \mathbf{h}(\mathbf{u}_m^{(t)})\|} - \frac{1}{N}\right)^2} \\
&= \sqrt{N\sum_{n=0}^{N-1}\left(\tilde{\Phi}_t(w_{n+1}^{(t)}) - \tilde{\Phi}_t(w_n^{(t)}) - \frac{1}{N}\right)^2}.
\end{aligned} \tag{182}$$

Because $w_n^{(t)} = \Phi_t^{-1}(n/N)$, we have $\Phi_t(w_{n+1}^{(t)}) - \Phi_t(w_n^{(t)}) = 1/N$, and hence

$$\begin{aligned}
\left|\tilde{\Phi}_t(w_{n+1}^{(t)}) - \tilde{\Phi}_t(w_n^{(t)}) - \frac{1}{N}\right| &= \left|\tilde{\Phi}_t(w_{n+1}^{(t)}) - \tilde{\Phi}_t(w_n^{(t)}) - \big(\Phi_t(w_{n+1}^{(t)}) - \Phi_t(w_n^{(t)})\big)\right| \\
&= \left|(\tilde{\Phi}_t - \Phi_t)(w_{n+1}^{(t)}) - (\tilde{\Phi}_t - \Phi_t)(w_n^{(t)})\right| \leq 2\|\tilde{\Phi}_t - \Phi_t\|_\infty.
\end{aligned} \tag{183}$$

Substituting the above bound into the CV expression gives

$$\mathrm{CV} \leq \sqrt{N\sum_{n=0}^{N-1} 4\|\tilde{\Phi}_t - \Phi_t\|_\infty^2} = 2N\|\tilde{\Phi}_t - \Phi_t\|_\infty. \tag{184}$$

It remains to control $\|\tilde{\Phi}_t - \Phi_t\|_\infty$. By Lemma 10, after substituting the increment bound

$$\max_{0\leq n\leq N-1}\left|e_t\left(\frac{n+1}{N}\right) - e_t\left(\frac{n}{N}\right)\right| \leq \frac{\kappa}{N}\|\Phi_t - \Phi\|_\infty^{1/2}, \tag{185}$$

the exact chord-length reconstruction satisfies

$$\|\hat{\Phi}_t - \Phi\|_\infty \leq \mathcal{O}\left(N^{-2} + \frac{\kappa^2}{N^2}\|\Phi_t - \Phi\|_\infty + \frac{\kappa^3}{N^2}\|\Phi_t - \Phi\|_\infty^{3/2}\right). \tag{186}$$

Since $N = \Omega(\kappa^{1.5})$, we have $\kappa^2/N^2 = \mathcal{O}(1)$ and $\kappa^3/N^2 = \mathcal{O}(1)$, and since $\|\Phi_t - \Phi\|_\infty \leq \mathcal{O}(1)$, this gives $\|\hat{\Phi}_t - \Phi\|_\infty \leq \mathcal{O}\left(N^{-2} + \|\Phi_t - \Phi\|_\infty\right)$. By Lemma 11, the inexact inner-solver perturbation satisfies $\|\tilde{\Phi}_t - \hat{\Phi}_t\|_\infty \leq \mathcal{O}\left(\frac{1}{s(1)} N \max_{0\leq n\leq N} \|\mathbf{x}_n^{(t)} - \mathbf{x}^*_{w_n^{(t)}}\|\right)$. Therefore,

$$\begin{aligned}\|\tilde{\Phi}_t - \Phi_t\|_\infty \leq& \|\tilde{\Phi}_t - \hat{\Phi}_t\|_\infty + \|\hat{\Phi}_t - \Phi\|_\infty + \|\Phi_t - \Phi\|_\infty \\ \leq& \mathcal{O}\left(N^{-2} + \|\Phi_t - \Phi\|_\infty + N \max_{0\leq n\leq N} \|\mathbf{x}_n^{(t)} - \mathbf{x}^*_{w_n^{(t)}}\|\right).\end{aligned} \tag{187}$$

Combining this with $\mathrm{CV} \leq 2N\|\tilde{\Phi}_t - \Phi_t\|_\infty$ gives

$$\mathrm{CV} = \mathcal{O}\left(N^{-1} + N\|\Phi_t - \Phi\|_\infty + N^2 \max_{0\leq n\leq N} \|\mathbf{x}_n^{(t)} - \mathbf{x}^*_{w_n^{(t)}}\|\right). \tag{188}$$

Under the conditions of Theorem 1, for large enough t, $\|\Phi_t - \Phi\|_\infty = \mathcal{O}(N^{-2})$ and $C(N)\max_{0\leq n\leq N} \|\mathbf{x}_n^{(t)} - \mathbf{x}^*_{w_n^{(t)}}\| = \mathcal{O}(N^{-2})$ with $C(N) = \Theta(N)$. Hence $\max_{0\leq n\leq N} \|\mathbf{x}_n^{(t)} - \mathbf{x}^*_{w_n^{(t)}}\| = \mathcal{O}(N^{-3})$, and therefore $\mathrm{CV} = \mathcal{O}(N^{-1})$. □

F Experimental Details

This appendix provides additional experimental details for the toy examples in Appendix F.1, the MO-Gymnasium benchmarks in Appendix F.3, and the LLM fine-tuning experiments in Appendix F.4. We describe the problem setups, optimization protocols, baselines, and hyperparameters used to reproduce the reported results.

Metric Settings. Metrics use the task's objective coordinates without per-objective rescaling. DST uses KL-regularized normalized discounted returns, Fishwood uses normalized discounted rewards, and MO-Mountaincar uses KL-regularized discounted episode returns. In minimization coordinates, the common HV reference point is the componentwise worst value among compared points plus 10^{-6} for DST or 1 for MO-Mountaincar, and $(0.2, 0.2)$ for Fishwood. IGD uses a dense 101-point SURF reference for DST and 201-point uniform-weight references for Fishwood and MO-Mountaincar. LLM metrics use the regularized or reward-only objectives indicated in Table 2, with HV reference point $(2.1, 1.4)$.

F.1 Toy Examples

We first describe the controlled toy examples used to visualize how scalarization weights map to Pareto-front coverage and the behavior of SURF.

Offline Bandit. Consider a single-state bandit with $|\mathcal{S}| = 1$ and a finite action set $\mathcal{A}$. The policy $\pi \in \Delta_{|\mathcal{A}|}$ coincides with the normalized occupancy measure. Let $x_a \in [0, 1]$ be evenly spaced over $[0, 1]$ for each action $a \in \mathcal{A}$, and define the true mean rewards

$$R_1(a) = x_a, \quad \text{and} \quad R_2(a) = 1 - x_a^4. \tag{189}$$

We use a uniform reference policy $\pi_{\mathrm{ref}} = \mathbf{1}/|\mathcal{A}|$ and define $R_0 = \log \pi_{\mathrm{ref}}$ elementwise. A fixed balanced offline dataset $\mathcal{D} = \{(a_t, r_{1,t}, r_{2,t})\}_{t=1}^T$ is collected, where every action appears roughly equally often and rewards are observed with Gaussian noise with standard deviation 0.5.

Quadratic Gear Toy. To isolate scalarization-induced traversal from changes in PF geometry, we construct a controlled gear toy in which the reported PF trajectory is fixed but the LS traversal speed is tunable. The fixed reporting map is $\bar{\mathbf{h}}(x) = ((1-x)^2, x^2)$ for $x \in [0, 1]$. We induce different traversal speeds using raw quadratic objectives $h_1^p(x) = \frac{\sqrt{p}}{2}(x-1)^2$ and $h_2(x) = \frac{1}{2}x^2$, for which

the LS solution is $x_w^* = \frac{\sqrt{p}\,w}{1+(\sqrt{p}-1)w}$. The raw objectives determine x_w^*, while SURF computes chord lengths and the CDF from $\bar{\mathbf{h}}(x_w^*)$. Thus, changing the gear parameter p changes how uniform weights traverse the same reported PF, while its curve and arc length remain fixed.

We run SURF with exact inner solutions and PCHIP interpolation for $p \in \{1, \dots, 10\}$ and $N \in \{25, \cdots, 40\}$, and compute the induced condition number $\kappa_p = \frac{v_{\max}}{v_{\min}}\sqrt{l_{\Phi,1}}$ from the true CDF. Empirically fitting $\|\tilde{\Phi}_0 - \Phi\|_\infty$ to $C\kappa_p^q/N^2$ gives q decreasing from 1.134 to 0.907 as N increases, showing a clear condition-number-dependent first reconstruction error. In contrast, fitting the final error $\|\Phi_T - \Phi\|_\infty$ gives q decreasing from 0.042 to 0.019, showing that after refinement the error is essentially independent of κ_p and reaches the expected N^{-2} discretization floor. This supports Lemma 10 and Theorem 1: the first CDF reconstruction may carry a large condition-number-dependent constant, while SURF converges to the final $\mathcal{O}(N^{-2})$ floor.

For an analytical interpretation, consider sufficiently large N for which the uniform initial grid $w_m = m/N$ contains $w_n = \frac{1}{\sqrt{p}+1}$, i.e., $n = \frac{N}{\sqrt{p}+1}$ is an integer. Then $x_{w_n}^* = \frac{1}{2}$. Therefore, $\tilde{\Phi}_0(w_n) - \Phi(w_n) = \frac{\frac{1}{2}(s(1)-\tilde{s}^{(0)}(1))-(s(w_n)-\tilde{s}^{(0)}(w_n))}{\tilde{s}^{(0)}(1)}$ and the exact chord-arc expansion gives $s(w_{m+1}) - s(w_m) - \|\bar{\mathbf{h}}(x_{w_{m+1}}^*) - \bar{\mathbf{h}}(x_{w_m}^*)\| = \frac{(x_{w_{m+1}}^* - x_{w_m}^*)^3}{12((1-x_{w_m}^*)^2+(x_{w_m}^*)^2)^{3/2}} + o((x_{w_{m+1}}^* - x_{w_m}^*)^3)$. Substituting $x_{w_m}^* = \frac{\sqrt{p}\,m}{N+(\sqrt{p}-1)m}$ and integrating gives $s(w_n) - \tilde{s}^{(0)}(w_n) - \frac{1}{2}(s(1) - \tilde{s}^{(0)}(1)) = \Theta(\frac{p}{N^2})$. Hence $\|\tilde{\Phi}_0 - \Phi\|_\infty = \Omega(\frac{p}{N^2}) = \Omega(\frac{v_{\max}/v_{\min}}{N^2})$. In contrast, $s(1) = \int_0^1 2\sqrt{(1-x)^2+x^2}\,dx$ is independent of p, and $l_{\tilde{\mathbf{f}}_{\mathrm{PF}},1} = \sup_{x\in[0,1]} \frac{1}{2((1-x)^2+x^2)^{3/2}} = \sqrt{2}$. Therefore, $\|\Phi_T - \Phi\|_\infty = \mathcal{O}\left(\frac{l_{\tilde{\mathbf{f}}_{\mathrm{PF}},1}^2 s(1)^3}{6N^2}\right) = \mathcal{O}(N^{-2})$, with both $s(1)$ and $l_{\tilde{\mathbf{f}}_{\mathrm{PF}},1}$ independent of p.

F.2 SURF with Tchebycheff Scalarization on Nonconvex Fronts

We evaluate weight refinement beyond LS on two connected nonconvex fronts: the ZDT2 front [114], parameterized by $(t, 1-t^2)$, and a quarter-circle front, parameterized by $(\cos(\pi t/2), \sin(\pi t/2))$, with $t \in [0,1]$. The latter coincides with the bi-objective DTLZ2 front [25]. We minimize the weighted Tchebycheff objective $\max\{w(f_1 - z_1^*), (1-w)(f_2 - z_2^*)\}$ with ideal point $\mathbf{z}^* = (0,0)$, using a 6001-point search over t followed by golden-section refinement to tolerance 10^{-10}. We use $N = 15$ segments, weights in $[10^{-3}, 1-10^{-3}]$, and 30 SURF outer updates with $\alpha = 0.3$ and PCHIP reconstruction. IGD uses 4001 uniformly spaced values of t, with no objective rescaling.

Uniform LS yields only two distinct endpoints on both fronts, whereas Tchebycheff scalarization yields 16 distinct points. Compared with uniform Tchebycheff weights, SURF reduces CV from 0.152 on ZDT2 and 0.202 on the circle to near zero, and Gap Ratio from 1.657 and 1.863 to 1.000, respectively. IGD decreases from 0.0250 to 0.0246 on ZDT2 and from 0.0272 to 0.0261 on the circle. These diagnostics demonstrate empirical compatibility with Tchebycheff scalarization outside the LS setting of Theorem 1.

F.3 Details of Experiments on MO-Gymnasium Benchmarks

We next describe the three MO-Gymnasium environments used in Table 1. For each task, we provide the reward structure, optimization setup, and relevant implementation details.

F.3.1 Details of DST Problem

This section describes the experimental setup used to evaluate the proposed PF sampling method on the Deep Sea Treasure (DST) benchmark [89, 101], provided in MO-Gymnasium [33].

Problem Setup. As illustrated in Figure 8, the environment is a grid world containing obstacle cells (black), treasure cells (gold), and water (light blue). This problem can be modeled via $\mathcal{M} = (\mathcal{S}, \mathcal{A}, \mathcal{P}, \mathbf{r}, \gamma, \rho, \tau)$, a regularized MDP where $\mathcal{S}$ denotes the set of feasible states (all non-obstacle cells in Figure 8), $\mathcal{A}$ is the action set that consists of four actions (up, down, left, and right), and $\mathcal{P}(s'|s,a)$ denotes the deterministic transition dynamics where the action moves to the adjacent cell in the chosen direction or stays at the current cell if the action would move it outside the grid

or into an obstacle. The reward vector $\mathbf{r}(s,a) = [r_1(s,a), r_2(s,a)]$ consists of r_1 representing the per-step time reward and r_2 representing the treasure reward obtained upon reaching a treasure cell. $\gamma = 0.999$ is the discount factor, chosen close to and smaller than 1. ρ is the initial state distribution, which is deterministic at the starting location. To encourage stochasticity [39], we adopt τ as the KL regularizer with a moderate choice of $\beta = 1.5$ and π_{ref} as the uniform policy.

The goal is to find a policy $\pi \in \Delta_{|\mathcal{A}|}^{|\mathcal{S}|}$ minimizing two regularized discounted objectives. Specifically, for $m = 1, 2$,

$$h_m(\pi) := \mathbb{E}_\pi \left[\sum_{t=0}^{\infty} \gamma^t \big(-r_m(s_t, a_t) + \tau(\pi(\cdot|s_t)) \big) \right], \qquad (190)$$

where $r_1(s,a) = -1$ is the per-step time reward, so that $-r_1(s,a) = 1$ corresponds to the time cost incurred at each step until a treasure is reached, and $r_2(s,a)$ is the treasure reward obtained upon reaching a treasure cell and is zero otherwise. Thus, h_1 favors policies with smaller travel time, while h_2 favors policies reaching higher-value treasures, which tend to take longer time as they are deeper (cf. Figure 8). The resulting trade-off defines the PF studied in our experiments.

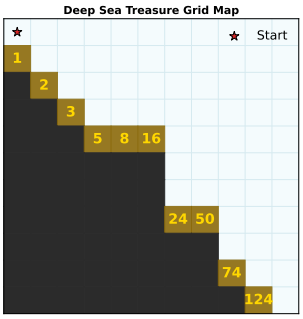

Figure 8: DST environment.

Occupancy Measure Representation. In this problem, the active state space consists of the 72 non-obstacle, non-terminal grid states in Figure 8, each with 4 actions. Let $\mu_\pi(s,a)$ denote the discounted occupancy measure distribution induced by policy π. For the active state-action pairs, the set of valid occupancy distributions forms a convex polytope (cf. (6)) and the policy optimization problem is equivalent to the occupancy distribution optimization described in Section 2.1. Specifically, the sufficient exploration condition [107], given by $\sum_a \mu_\pi(s,a) \geq d_{\min}$ for every active state s, and Assumption 3 have been numerically verified on a dense interior grid of scalarization weights, using 1001 uniformly spaced values in $[0,1]$. In particular, with a moderate regularization parameter $\beta = 1.5$, there is $d_{\min} \geq 1.02 \times 10^{-8}$, with minimum attained near $w \approx 0.964$. Increasing the regularization to $\beta = 10$ further improves it to $d_{\min} \geq 1.19 \times 10^{-6}$. Additionally, the active-state flow matrix has shape 72×288 and full row rank, verifying Assumption 2. Furthermore, the projected gradient-difference norm in Assumption 3 remains uniformly positive over the entire weight grid. In the experiment with $\beta = 1.5$, the empirical lower bound is 40.3, attained near $w \approx 0.45$, confirming that the tangent-gap condition holds throughout the tested weight range. In this way, according to the theorem of hidden convexity [32], optimizing over policies is equivalent to optimizing over occupancy measures, where the objectives become strongly convex in μ and have the properties needed to establish the CDF Φ and enable the use of the proposed SURF Algorithm 1 to obtain Pareto solutions that uniformly cover the Pareto Front.

Baselines. We compare our proposed arc-length-uniform Pareto sampling method with three scalarization-based baselines. *Uniform weighting* chooses equispaced scalarization weights on $[0,1]$. *Optimistic Linear Support (OLS)* [5] adaptively selects scalarization weights based on the convex coverage set discovered so far. At each step, OLS chooses a new weight vector that is optimistic with respect to the currently known solutions and solves the corresponding scalarized optimization problem to expand the supported portion of the PF. *UMOD* [110] is a preference-learning approach that iteratively adapts scalarization weights using a surrogate model of the PF. It fits a parametric model mapping preferences to objective vectors from existing solutions, then selects new preferences to maximize the minimum chord length between predicted Pareto points. These preferences are converted into scalarization weights and used to query the optimization oracle for additional solutions. This approach implicitly relies on the surrogate geometry to reflect meaningful separation between solutions. However, when the mapping from scalarization weights to Pareto solutions produces strongly clustered solutions, such that many nearby weights yield nearly identical objective vectors, the surrogate spacing signal becomes weak, potentially reducing the effectiveness of the preference updates.

Results. As shown in Figure 6, the PF produced by the proposed SURF Algorithm 1 method exhibits more evenly distributed solutions compared with the baselines. This improvement is reflected quantitatively by substantially lower CV, Gap Ratio, and IGD relative to uniform weight sampling and

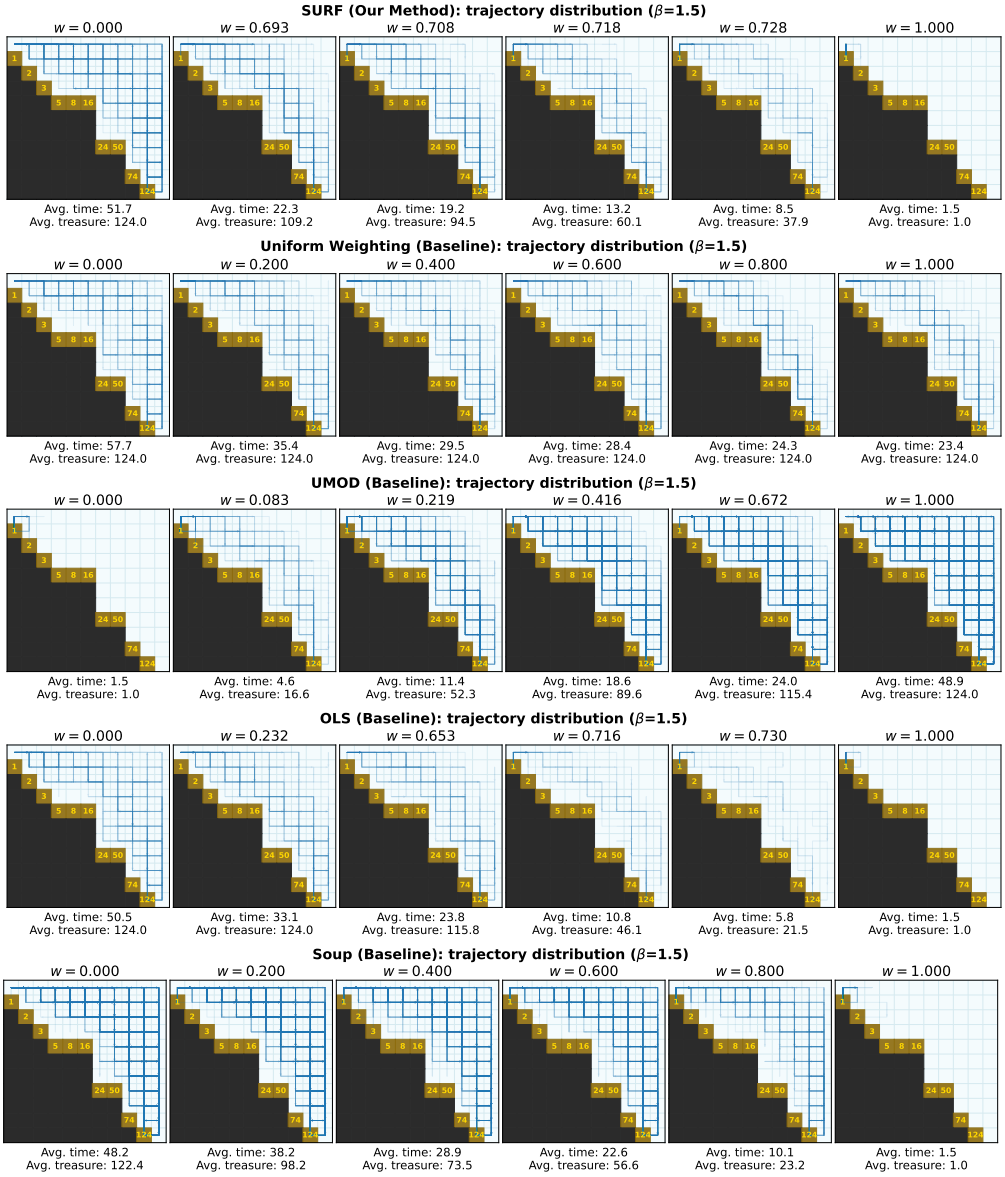

Figure 9: Trajectory distributions induced by policies obtained via our SURF Algorithm 1, *Uniform Weighting*, *UMOD*, *Optimistic Linear Support*, and *Soup* baselines with $(N + 1) = 6$. Each panel visualizes state visitation from 50 rollouts at the indicated scalarization weight w, except for Soup, where w is the endpoint interpolation coefficient. Blue trajectories show the paths taken by the agent with darker or more opaque segments indicating higher visitation frequency, while lighter segments indicate rarer transitions. The grid background represents the environment. Statistics below each panel report the average travel time and collected treasure.

optimistic linear support. These results indicate that the learned scalarization weights better adapt to the geometry of the PF.

Additionally, we provide visualization of the trajectory distributions induced by policies obtained by our SURF Algorithm 1 and the baselines. Figure 9 illustrates that the policies obtained by our proposed method exhibit a gradual transition from favoring high-value treasure to prioritizing shorter travel time. In contrast, uniform weight sampling generates policies that are largely similar across many weights, often leading to nearly identical trajectory patterns and thus failing to reveal intermediate trade-offs. OLS produces somewhat more diverse policies, but several solutions still

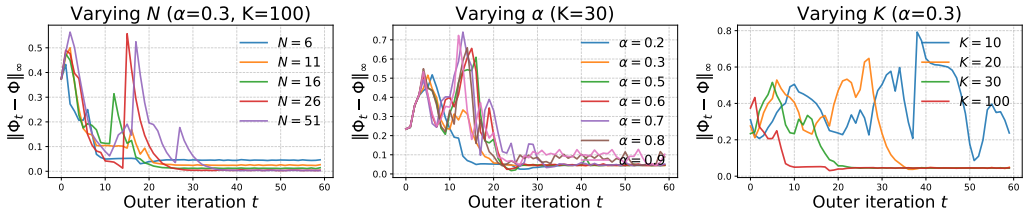

Figure 10: Convergence behavior of the SURF Algorithm 1 under different parameter choices. Left: varying the sampling budget. Middle: varying the outer update parameter α. Right: varying the number of inner optimization steps K.

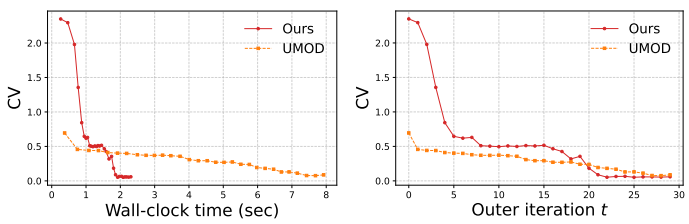

Figure 11: Convergence comparison between our SURF and UMOD on the DST problem. Left: CV versus wall-clock time. Right: CV versus outer iteration.

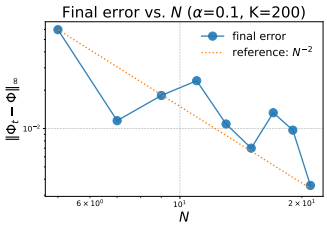

Figure 12: $\|\Phi_t - \Phi\|_\infty$ vs. $N \in \{4, 6, \ldots, 20\}$ on DST.

cluster around similar behaviors, resulting in uneven coverage of the trade-off spectrum. Overall, the SURF Algorithm 1 method provides a more gradual progression of policies along the Pareto trade-off.

Figure 10 empirically corroborates Theorem 1. With a safe choice of $\alpha = 0.3$ and sufficiently many inner iterations ($K = 100$), the convergence curves in the left panel show decay toward an error floor in $\|\Phi_t - \Phi\|_\infty$. The initial fluctuations arise because the contraction in Theorem 1 applies to the combined quantity $\mathcal{E}_t$ defined in (14), so early iterations primarily reduce the inner optimization error before the outer CDF error dominates. The left panel also shows that a larger N leads to a smaller final error floor, consistent with the $\mathcal{O}(N^{-2})$ discretization term predicted by the theorem. This is further corroborated in Figure 12. The middle and right panels in Figure 10 illustrate the parameter trade-offs suggested by the theorem: α should be chosen moderately small to avoid overly abrupt updates of Φ_t, while the number of inner iterations K must be sufficiently large to ensure that the inner optimization keeps pace with the outer refinement. Although warm-starting mitigates the effect of small K in practice, larger K still improves stability and convergence.

F.3.2 Details of Fishwood Problem

Fishwood Problem Setup. This problem can be modeled via $\mathcal{M} = (\mathcal{S}, \mathcal{A}, \mathcal{P}, \mathbf{r}, \gamma, \rho, \tau)$, a regularized MDP where $\mathcal{S} = \{0, 1\}$ denotes the state space with 0 representing the fishing state and 1 representing the woods, and $\mathcal{A} = \{0, 1\}$ is the action set consisting of two actions: fishing and collecting wood. The transition dynamics $\mathcal{P}(s'|s, a)$ are deterministic in the sense that the next state is determined by the chosen action, i.e., taking action 0 leads to the fishing state and action 1 leads to the woods. The reward vector $\mathbf{r}(s, a) = [r_1(s, a), r_2(s, a)]$ is stochastic and consists of two components: r_1 represents the wood reward and r_2 represents the fish reward. Specifically, when the agent is in the woods, it receives a reward of $+1$ in r_1 with probability p_{wood} and 0 otherwise, while when the agent is fishing, it receives a reward of $+1$ in r_2 with probability p_{fish} and 0 otherwise. This induces a trade-off between maximizing wood collection and maximizing fish collection, since the agent can only occupy one state at a time.

The discount factor is chosen as $\gamma = 0.995$, and ρ denotes the initial state distribution, which is deterministic at the woods state. As in the DST setup, we adopt the same discounted infinite-horizon formulation (190) and KL regularizer $\tau(\pi(\cdot|s)) = \beta \text{KL}(\pi(\cdot|s)\|\pi_{\text{ref}})$ with a choice of $\beta = 0.5$ and π_{ref} as the uniform policy. To approximate the PF, we use $N = 10$ segments, hence $N + 1 = 11$

preference weights. Each scalarized problem is solved with $K = 100$ steps in the inner loop using learning rate 0.05. Our proposed algorithm is run for 15 outer iterations with $\alpha = 0.3$.

F.3.3 Details of MO-Mountaincar Problem

MO-Mountaincar Problem Setup. MO-Mountaincar is a multi-objective variant of the classical Mountaincar control problem, in which an underpowered car must build momentum by moving back and forth in a valley in order to reach the goal near the top of the right hill. In the multi-objective setting, the agent is penalized at every time step and also incurs additional penalties for control behavior, thereby inducing a trade-off between fast completion and action efficiency.

This problem can be modeled by a multi-objective Markov decision process $\mathcal{M} = (\mathcal{S}, \mathcal{A}, \mathcal{P}, \mathbf{r}, \gamma, \rho, \tau)$, where the state space is continuous, $\mathcal{S} \subseteq [-1.2, 0.6] \times [-0.07, 0.07]$, and each state $s_t = (x_t, v_t)$ consists of the car position x_t and velocity v_t. The action space is $\mathcal{A} = \{0, 1, 2\}$, corresponding to accelerating to the left, taking no push, and accelerating to the right, respectively. Equivalently, the control entering the dynamics is $\tilde{a}_t = a_t - 1 \in \{-1, 0, 1\}$. The transition dynamics are deterministic with $v_{t+1} = v_t + 0.001\tilde{a}_t - 0.0025\cos(3x_t)$, $x_{t+1} = x_t + v_{t+1}$, followed by clipping x_{t+1} to $[-1.2, 0.6]$ and v_{t+1} to $[-0.07, 0.07]$; if the car hits the left boundary, the velocity is reset to zero. In our implementation, we use a two-dimensional reward vector $\mathbf{r}(s_t, a_t) = [r_1, r_2]$, where $r_1 = -1$ is the per-step time penalty, and $r_2 = -1$ if $a_t = 2$ and 0 otherwise is the reverse-action penalty. This induces a trade-off between reaching the goal quickly and reducing reverse control effort. The discount factor is chosen as $\gamma = 0.995$, and ρ denotes the initial-state distribution, where the initial position is sampled uniformly from $[-0.6, -0.4]$ and the initial velocity is fixed at zero. An episode terminates once the car reaches the goal position, i.e., when $x_t \geq 0.5$, and is truncated if the trajectory length reaches 200 steps. We use the KL regularizer $\tau(\pi(\cdot \mid s)) = \beta\,\mathrm{KL}\big(\pi(\cdot \mid s) \,\|\, \pi_{\mathrm{ref}}\big)$, with $\beta = 0.05$ and π_{ref} taken as the uniform policy over $\mathcal{A}$. For the inner solver, we use a DQN trained with Adam optimizer and learning rate 10^{-3}. In the reported comparison, we use $N + 1 = 6$ sparse preference weights. Our method uses 10 outer iterations with update parameter $\alpha = 0.3$ and trains each scalarized subproblem for $K = 100$ steps per outer iteration.

F.4 LLM Multi-Objective Fine-Tuning

This section describes the experimental setup used to evaluate SURF on a stochastic LLM fine-tuning problem, where each scalarized subproblem is solved approximately by PPO and the PF is estimated from reward-model feedback.

Problem Setup. We study bi-objective alignment for Reddit summarization [91] using prompts from the `openai/summarize_from_feedback` dataset [87]. Following the experimental configuration, we use the training split and restrict PPO training to a subset of 4096 training prompts. The policy is initialized from `Qwen/Qwen2.5-0.5B-Instruct` [74] and generates summaries conditioned on Reddit posts. To make repeated scalarized fine-tuning feasible, we keep the pretrained backbone fixed and train only LoRA adapter parameters [45].

We view LLM summarization as a contextual RL problem in which s denotes the prompt and a denotes the generated summary. For each prompt-summary pair (s, a), we evaluate two reward signals: $r_1(s, a)$ is given by `Tristan/gpt2_reward_summarization`, and $r_2(s, a)$ is given by `CogComp/bart-faithful-summary-detector`. The corresponding reward outputs are parsed using the configured reward formats `0` and `1-0`, respectively. These reward models induce a trade-off between a summarization-quality reward and a factual-faithfulness reward. Following KL- or divergence-regularized multi-objective alignment formulations [93, 86], we instantiate $\tau(\pi(\cdot|s)) = \beta\mathrm{KL}(\pi(\cdot|s)\|\pi_{\mathrm{ref}}(\cdot|s))$ where π_{ref} is the same base model before PEFT adaptation and $\beta = 0.05$ is the KL coefficient. We define the minimization objectives

$$h_m(\pi) := -\mathbb{E}_{s\sim\mathcal{D}, a\sim\pi(\cdot|s)}\left[r_m(s, a) - \tau(\pi(\cdot|s))\right], \quad m \in \{1, 2\}. \tag{191}$$

For weight $w \in [0, 1]$, the scalarized minimization objective follows the LS convention $\mathrm{LS}_h(\pi; w) = wh_1(\pi) + (1 - w)h_2(\pi)$. Equivalently, PPO maximizes the scalarized regularized reward

$$wr_1(s, a) + (1 - w)r_2(s, a) - \tau(\pi(\cdot|s)). \tag{192}$$

All reported objective values use the minimization coordinates (h_1, h_2), where smaller values are better.

Training Protocol. We solve each scalarized subproblem using PPO [82], updating only the PEFT adapter parameters while keeping the Qwen backbone frozen. The PPO runs use a fixed KL coefficient $\beta = 0.05$ with adaptive KL disabled, learning rate 5×10^{-5}, PPO batch size 64, mini-batch size 16, and generated response length in $[16, 32]$. We first train a $w = 0.5$ scalarized PPO model for 10 epochs as a common warm start. All runs are conducted on NVIDIA A100-PCIE GPUs with 40GB memory. Each scalarized PPO job uses one A100 GPU, 4 CPU cores, and 16GB host memory.

We compare SURF Algorithm 1 with the uniform weighting baseline using $N = 5$ PF segments, hence $N + 1 = 6$ scalarization weights. For uniform weighting, we train one PPO run for each fixed weight in $\{0, 0.2, 0.4, 0.6, 0.8, 1\}$ for 12 epochs, with each run initialized from the shared $w = 0.5$ warm start and trained on one A100 GPU. For SURF, we use the epoch-5 checkpoint from the corresponding uniform weighting run as the initial iterate $\mathbf{u}_n^{(-1)}$ for each slot $n = 0, \ldots, 5$, which serves as a good warm start. Then, SURF performs 7 outer updates, using one PPO epoch as the finite inner-solver budget at each outer iteration. SURF reuses the endpoint slots $w = 0$ and $w = 1$ from the subsequent uniform weighting runs and selects and updates the four middle scalarization points. When reconstructing the PF points for the SURF CDF update, we apply an exponential moving average with rate 0.2 to the per-batch reward logs from the inner PPO training, which stabilizes the stochastic estimates of $\mathbf{h}(\mathbf{u}_n^{(t)})$ used to build $\tilde{s}^{(t)}$ in (12). Independent middle-slot jobs are run in parallel, each on one A100 GPU. Thus, both methods have a per-slot budget of 12 PPO epochs after the shared 10-epoch warm start, with SURF using 5 initial epochs plus 7 refinement epochs. Shared warm-start and endpoint training is counted once in actual compute. The comparison uses per-slot PPO epochs as the training-budget measure. Reported PF points and metrics use separate post-training evaluations over 10 batches of 64 prompts from the training split.

We also consider Soup, an inference baseline [76], which constructs intermediate models by interpolating the PEFT adapter parameters of the two endpoint policies. That is, for $\mathbf{u}_0^{(12)}$ and $\mathbf{u}_1^{(12)}$ corresponding to the epoch-12 outputs from PPO training on weights 0 and 1, respectively, it directly merges the PEFT parameters as $(1 - w)\mathbf{u}_0^{(12)} + w\mathbf{u}_1^{(12)}$ for $w \in \{0, 0.2, 0.4, 0.6, 0.8, 1\}$.

G Limitations and Broader Impacts

Our current theory and algorithm focus primarily on the bi-objective setting. Extending SURF to problems with more than two objectives is a natural next step that calls for techniques to characterize and uniformly sample higher-dimensional PF geometry. Further directions include extending the analysis to scalarizers that reach unsupported nonconvex regions and paths with active-set changes or vanishing traversal speed. Component-wise reconstruction could also extend SURF to disconnected fronts by measuring arc length separately within each component. Future work should evaluate SURF on larger multi-objective benchmarks, more complex reinforcement-learning environments, and larger-scale language-model alignment problems to better understand its scalability and robustness.

More broadly, SURF offers a general framework for analyzing trade-offs among competing objectives. By promoting more uniform coverage of the Pareto front, it can provide a more comprehensive view of the trade-off landscape and reduce the risk of drawing conclusions from a narrowly sampled region. This can be beneficial in applications where transparent and reliable trade-off analysis is important, such as reinforcement learning, model alignment, and resource allocation. A future direction is to combine this representative menu with user preferences and explicit safety, fairness, or faithfulness requirements to guide the selection of deployable trade-offs.

NeurIPS Paper Checklist

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: The abstract and introduction (Section 1) clearly state the main contributions, including characterizing scalarization-induced PF geometry, proposing SURF as a CDF-refinement method, and validating it on bandits, MO-Gymnasium benchmarks, and LLM

alignment. The claims are supported by theory and experiments showing improved uniform PF coverage while preserving the simplicity of linear scalarization.

Guidelines:

- The answer [N/A] means that the abstract and introduction do not include the claims made in the paper.
- The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A [No] or [N/A] answer to this question will not be perceived well by the reviewers.
- The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
- It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

Question: Does the paper discuss the limitations of the work performed by the authors?

Answer: [Yes]

Justification: The paper discusses its limitations in Appendix G, including the focus on the bi-objective setting and the need to extend SURF to the multiple-objective case. It also notes that future work should scale the experiments to larger multi-objective benchmarks, more complex RL environments, and larger language-model alignment settings.

Guidelines:

- The answer [N/A] means that the paper has no limitation while the answer [No] means that the paper has limitations, but those are not discussed in the paper.
- The authors are encouraged to create a separate "Limitations" section in their paper.
- The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
- The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
- The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
- The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
- If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
- While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory assumptions and proofs**

Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

Answer: [Yes]

Justification: The paper provides a complete set of assumptions and detailed proofs for each theoretical result, with clear numbering and cross-referencing in both the main text and appendix. The assumptions are stated alongside the theoretical results in Sections 2–3, while full proofs are provided in Appendix C–E.

Guidelines:

- The answer [N/A] means that the paper does not include theoretical results.
- All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
- All assumptions should be clearly stated or referenced in the statement of any theorems.
- The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
- Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
- Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental result reproducibility**

   Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

   Answer: [Yes]

   Justification: The paper provides sufficient details to reproduce the main experimental results, including benchmark environments, model architectures, hyperparameter settings, and implementation details. It clearly describes the baselines, evaluation metrics, and training procedures, with the relevant settings reported in Section 4 and Appendix F. These details ensure that the main claims and conclusions can be independently verified.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - If the paper includes experiments, a [No] answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
   - If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
   - Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
   - While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
     (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
     (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
     (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).
     (d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. **Open access to data and code**

Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

Answer: [Yes]

Justification: The code, data and instructions required to reproduce the main experimental results are provided in the supplementary material.

Guidelines:

Guidelines:

- The answer [N/A] means that paper does not include experiments requiring code.
- Please see the NeurIPS code and data submission guidelines (`https://neurips.cc/public/guides/CodeSubmissionPolicy`) for more details.
- While we encourage the release of code and data, we understand that this might not be possible, so [No] is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
- The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (`https://neurips.cc/public/guides/CodeSubmissionPolicy`) for more details.
- The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
- The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
- At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
- Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental setting/details**

   Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer) necessary to understand the results?

   Answer: [Yes]

   Justification: Section 4 and Appendix F specify the the training and testing settings for all experiments. Detailed implementation is included in the supplementary material.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
   - The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment statistical significance**

   Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

   Answer: [Yes]

   Justification: See Section 4. This paper provides results over multiple random seeds for the stochastic experiments and reports error bars or standard deviation.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The authors should answer [Yes] if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.

- The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
- The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
- The assumptions made should be given (e.g., Normally distributed errors).
- It should be clear whether the error bar is the standard deviation or the standard error of the mean.
- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g., negative error rates).
- If error bars are reported in tables or plots, the authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments compute resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

   Answer: [Yes]

   Justification: See Appendix F. The paper reports the main computational resources used for the experiments, including CPU/GPU type, memory requirements, and implementation details needed to reproduce the reported results.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
   - The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
   - The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code of ethics**

   Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics `https://neurips.cc/public/EthicsGuidelines`?

   Answer: [Yes]

   Justification: The authors obey the NeurIPS Code of Ethics.

   Guidelines:

   - The answer [N/A] means that the authors have not reviewed the NeurIPS Code of Ethics.
   - If the authors answer [No], they should explain the special circumstances that require a deviation from the Code of Ethics.
   - The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader impacts**

    Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

    Answer: [Yes]

    Justification: See Appendix G. The paper discusses the positive impact of SURF as a general framework for trade-off analysis, where more uniform Pareto-front coverage can support more transparent and reliable decision making. It also notes limitations related to

scalability and objective design, especially when applying the method to more complex RL, model-alignment, and resource-allocation settings.

Guidelines:

- The answer [N/A] means that there is no societal impact of the work performed.
- If the authors answer [N/A] or [No], they should explain why their work has no societal impact or why the paper does not address societal impact.
- Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
- The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate Deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
- The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
- If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pre-trained language models, image generators, or scraped datasets)?

Answer: [N/A]

Justification: The paper does not release a new high-risk pretrained model, image generator, or scraped dataset. The released materials are primarily algorithmic code and experimental scripts for reproducing SURF and the reported benchmark evaluations.

Guidelines:

- The answer [N/A] means that the paper poses no such risks.
- Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
- Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
- We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

Answer: [Yes]

Justification: The paper cites the original sources for the benchmark environments, datasets, models, and baseline algorithms used in the experiments in Section 4.

Guidelines:

- The answer [N/A] means that the paper does not use existing assets.

- The authors should cite the original paper that produced the code package or dataset.
- The authors should state which version of the asset is used and, if possible, include a URL.
- The name of the license (e.g., CC-BY 4.0) should be included for each asset.
- For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
- If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, `paperswithcode.com/datasets` has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
- For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
- If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New assets**

    Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

    Answer: [Yes]

    Justification: We release the code through an anonymized zip file and provide all training details in Section 4 and Appendix F.

    Guidelines:

    - The answer [N/A] means that the paper does not release new assets.
    - Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
    - The paper should discuss whether and how consent was obtained from people whose asset is used.
    - At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and research with human subjects**

    Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

    Answer: [N/A]

    Justification: This paper does not involve crowdsourcing experiments or research with human subjects.

    Guidelines:

    - The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
    - Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
    - According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional review board (IRB) approvals or equivalent for research with human subjects**

    Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

    Answer: [N/A]

Justification: This paper does not involve crowdsourcing experiments or research with human subjects.

Guidelines:

- The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
- Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
- We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
- For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

16. **Declaration of LLM usage**

Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does *not* impact the core methodology, scientific rigor, or originality of the research, declaration is not required.

Answer: [N/A]

Justification: The core method developed in this work is original and does not involve the use of LLMs.

Guidelines:

- The answer [N/A] means that the core method development in this research does not involve LLMs as any important, original, or non-standard components.
- Please refer to our LLM policy in the NeurIPS handbook for what should or should not be described.