Markov Chain from Human Feedback

Takuya Koriyama
The University of Chicago

Tengyuan Liang
The University of Chicago

Abstract

We propose Markov Chain from Human Feedback (MCHF), an elementary approach for aligning generative models from pairwise human preferences. Unlike Reinforcement Learning from Human Feedback (RLHF), which reduces comparisons to a scalar reward, and Nash Learning from Human Feedback (NLHF), which preserves pairwise utilities through a KL-regularized minimax optimization, MCHF uses pairwise preferences directly to define a transition mechanism over model outputs. Given a pairwise utility $U(x, y)$, which quantifies human preference for y over x, and a reference probability distribution μ_{ref}, we define a Markov kernel $\mathsf{P}(x, dy) \propto \exp(U(x, y))\mu_{\mathsf{ref}}(dy)$, and take the Markov chain starting from μ_{ref} as an iterative alignment procedure. We show that MCHF converges geometrically fast to the stationary distribution, with a convergence rate governed by the seminorm $\|U\|_{\oplus} = \inf_{g,f \in L^\infty(\mu_{\mathsf{ref}})} \|U - g \oplus f\|_\infty$, which quantifies the non-transitive structure of the pairwise utility. We further show that a mirror-descent algorithm for NLHF satisfies an analogous structure-adaptive convergence guarantee. Finally, we prove that when $\|U\|_{\oplus}$ is small, MCHF and NLHF agree up to first order around an RLHF solution, which yields a unified view of reward-based, game-theoretic, and Markovian approaches to alignment. In particular, for two natural algorithms that converge to the MCHF/NLHF equilibria, we show that the first step of MCHF and NLHF recovers the RLHF solution based on the column-sum reward $\hat{f}(y) = \int \mu_{\mathsf{ref}}(dx)U(x, y)$, and starting from the second iteration, both algorithms incorporate the same linear functional of the residual $U - (-\hat{f}) \oplus \hat{f}$, which captures the non-transitive structure of the pairwise utility U.

1. Introduction

Reward based approach: Reinforcement Learning from Human Feedback (RLHF) Alignment from human feedback often begins with pairwise comparisons. RLHF reduces such data to a scalar reward R, typically through a BTL model [2, 14]:

$$\mathcal{P}(x \prec y) = \sigma_{\text{sigmoid}}(R(y) - R(x)) \quad \text{where} \quad \sigma_{\text{sigmoid}}(t) \equiv (1 + e^{-t})^{-1} \tag{1}$$

and then performs KL-regularized policy optimization [1, 4, 18]:

$$\mu_{\mathsf{RL}} = \operatorname*{argmax}_{\mu} \int \mu(dy)R(y) - D_{\mathsf{KL}}(\mu|\mu_{\mathsf{ref}}) \tag{2}$$

Game theoretic approach: Nash Learning from Human Feedback (NLHF) More recently, several variants to the standard RLHF have been proposed [7, 8, 10, 15, 20]. However, this reduction to scalar rewards cannot directly represent cyclic or pair-specific preferences. Nash Learning from Human Feedback (NLHF) [16] avoids this reduction by computing the minimax optimization

© T. Koriyama & T. Liang.

problem:

$$(\mu_{\mathsf{NL}}, \nu_{\mathsf{NL}}) \in \operatorname*{argmax}_{\mu} \operatorname*{argmin}_{\nu} \iint \nu(dx)\mu(dy)U(x,y) - D_{\mathsf{KL}}(\mu \mid \mu_{\mathsf{ref}}) + D_{\mathsf{KL}}(\nu \mid \mu_{\mathsf{ref}}) \tag{3}$$

where $U(x, y)$ is a pairwise utility, typically chosen as an increasing function of human's preference model $\mathcal{P}(x \prec y)$. More broadly, game-theoretic perspectives on alignment have led to formulations based on other concepts, including Stackelberg games [5, 19] and multiplayer games [25].

The original paper [16] argues that NLHF is attractive because it can yield a more diverse distribution than reward-based alignment methods. More refined analyses have been developed through the lens of social choice theory [9, 23].

Nevertheless, because NLHF is formulated as a worst-case equilibrium against an adversarial opponent, it is intrinsically more pessimistic than reward-based alignment, and it is not always clear that such a pessimistic objective is the most appropriate one for fine-tuning generative models. Moreover, NLHF treats alignment as a simultaneous game, in which the two players, μ and ν, optimize a regularized zero-sum objective at the same time. In contrast, preference refinement can be more naturally viewed as a non-simultaneous, leader-follower procedure: given a current output, one samples a new output conditionally on it, favoring alternatives that are preferred to the current one.

Contribution: a Markov chain approach Motivated by this viewpoint, we propose Markov Chain from Human Feedback (MCHF), a Markovian alignment approach that uses pairwise preference utilities directly through such a non-simultaneous formulation. Concretely, given a utility $U(x, y)$, where larger values indicate that y is preferred to x, and a reference model μ_{ref}, we define the Markov kernel P by

$$\mathsf{P}(x, dy) = \frac{1}{Z(x)} \exp(U(x,y))\mu_{\mathsf{ref}}(dy), \quad Z(x) \equiv \int \exp(U(x,y))\mu_{\mathsf{ref}}(dy) \tag{4}$$

where $Z(x)$ is the normalizing constant. Then, given a current distribution μ, one step of MCHF updates it according to

$$\mu_{\text{new}}(dy) = \mu\mathsf{P}(dy) = \int \mu(dx)\mathsf{P}(x, dy).$$

Here, the kernel P favors transitions from a current output x to outputs y that are preferred relative to x, while the reference distribution μ_{ref} preserves the structure, especially the support, of the reference distribution μ_{ref}. Thus, in contrast to the simultaneous-game interpretation underlying NLHF, MCHF implements a non-simultaneous/leader-follower alignment procedure via the conditional sampling from the Markov kernel P.

This one-step update can also be iterated, starting from μ_{ref}:

$$\mu_{\mathsf{ref}} \mapsto \mu_{\mathsf{ref}}\mathsf{P} \mapsto \mu_{\mathsf{ref}}\mathsf{P}^2 \mapsto \cdots \to \mu_{\mathsf{ref}}\mathsf{P}^{\infty} = \mu_{\mathsf{MC}}$$

Then, MCHF induces an interpolation path from the reference distribution to the stationary distribution μ_{MC}, which fully incorporates the pairwise preference utility U. This iterative alignment procedure is inspired by classical spectral rank-aggregation methods [3, 17], where one constructs a directed graph from pairwise comparisons and represents an aggregate ranking as the stationary distribution of a Markov chain on the graph.

Beyond introducing MCHF, our main theoretical contribution is to clarify its relationship with existing alignment methods such as NLHF and RLHF. A key observation is that, when the pairwise

utility is written by an additive form $U(x, y) = g(x) + f(y)$ for some functions g and f, then the MCHF and NLHF equilibria based on such U both reduce to the RLHF solution with reward f: $\mu_{\mathsf{RL}}(f)(dy) \propto \exp(f(y))\mu_{\mathsf{ref}}(dy)$. This motivates us to compare MCHF and NLHF via a common perturbation framework: what are their first-order behaviors when U is close to an additive pairwise utility $g \oplus f$?

We address this question by introducing a seminorm $\|\cdot\|_{\oplus}$ on $L^{\infty}(\mu_{\mathsf{ref}} \otimes \mu_{\mathsf{ref}})$ (see Definition 1) which quantifies the non-additive or non-transitive structure of the pairwise preference utility U. We then show that, when $\|U\|_{\oplus}$ is small, RLHF provides a natural baseline corresponding to the transitive component of the pairwise utility, and around this baseline, NLHF and MCHF exhibit the same first-order structure, despite arising from different motivations: NLHF from a game-theoretic equilibrium and MCHF from a stationary distribution of a preference-driven Markov chain. This allows us to compare the three methods, RLHF, NLHF, and MCHF, within a common perturbative framework and to quantify how their differences depend on the non-transitive component of the preference utility U. In this sense, we do not merely propose MCHF as an additional alignment algorithm, but also provide a unified view of reward-based, game-theoretic, and Markovian approaches to preference alignment.

Organization A comprehensive list of notation is provided at the beginning of the appendix. A longer version of this paper is available on arXiv [11], which includes a more detailed discussion of each theorem and a more refined comparison between MCHF and NLHF.

2. Main result

2.1. Definition of seminorm: measuring non-additive structure of utility

Let $U(x, y)$ be a pairwise utility function that encodes preference such that "$U(x, y)$ is large" if and only if "a human is more likely to prefer y to x". Throughout this paper, we assume $U \in L^{\infty}(\mu_{\mathsf{ref}} \otimes \mu_{\mathsf{ref}})$, i.e., its L^{∞} norm with respect to the product measure is bounded.

We now define a seminorm on $L^{\infty}(\mu_{\mathsf{ref}} \otimes \mu_{\mathsf{ref}})$, which serves as a fundamental complexity measure for the alignment dynamics of MCHF and NLHF throughout this paper.

Definition 1 *Define the seminorm $\|\cdot\|_{\oplus}$ on $L^{\infty}(\mu_{\mathsf{ref}} \otimes \mu_{\mathsf{ref}})$ by*

$$\|U\|_{\oplus} = \inf_{g,f \in L^{\infty}(\mu_{\mathsf{ref}})} \|U - g \oplus f\|_{\infty}, \quad \textit{where} \quad (g \oplus f)(x, y) = g(x) + f(y).$$

This seminorm $\|U\|_{\oplus}$ captures the *additive defect* of the utility; the larger $\|U\|_{\oplus}$ is, the more U has a non-additive structure. Note that $\|U\|_{\oplus} \leq \|U\|_{\infty}$ by the definition. In practice, the utility is often taken to be a possibly nonlinear transformation of the preference model $\mathcal{P}(x \prec y)$, so the value of $\|U\|_{\oplus}$ depends both on the nonlinear transformation and on the structure of $\mathcal{P}(x \prec y)$. Importantly, the utility function need not be exactly additive, but there are natural situations in which it is approximately additive (see Section A).

Note that the exact computation of the seminorm $\|U\|_{\oplus}$ can be challenging, especially when the state space is non-discrete or when it is discrete but very large, since the problem is essentially an L^{∞} linear program. However, we can estimate $\|U\|_{\oplus}$ up to a multiplicative order (see Proposition 9): $\|U\|_{\oplus} \asymp \|U - \hat{g} \oplus \hat{f}\|_{\infty}$, where $(\hat{g}, \hat{f})$ is an L^2-solution $(\hat{g}, \hat{f}) \in \inf_{g,f} \|U - g \oplus f\|_2$

2.2. Structure-Adaptive Convergence of MCHF and NLHF

Based on the utility function $U \in L^\infty(\mu_{\mathsf{ref}} \otimes \mu_{\mathsf{ref}})$, let P be the Markov kernel defined by (4). Let us consider the Markov chain $\mu_{\mathsf{ref}} \mapsto \mu_{\mathsf{ref}}\mathsf{P} \to \mu_{\mathsf{ref}}\mathsf{P}^2 \cdots$ on the complete metric space $(\mathscr{P}_{\mu_{\mathsf{ref}}}, d_{\mathsf{TV}})$ where $\mathscr{P}_{\mu_{\mathsf{ref}}}$ is the set of probability measures on $\mathcal{X}$ that are absolutely continuous with respect to μ_{ref}, and d_{TV} is the total variation distance. Now we claim that the map $\mu \mapsto \mu\mathsf{P}$ is a contraction mapping on this space.

Theorem 2 *Let P be the Markov kernel defined by (4) with the utility $U \in L^\infty(\mu_{\mathsf{ref}} \otimes \mu_{\mathsf{ref}})$. Then,*

$$\forall \mu, \nu \in \mathscr{P}_{\mu_{\mathsf{ref}}}, \quad d_{\mathsf{TV}}(\mu\mathsf{P}, \nu\mathsf{P}) \le c \cdot d_{\mathsf{TV}}(\mu, \nu),$$

where the contraction rate c is given by

$$c = c(\|U\|_\oplus) = \min(1 - e^{-2\|U\|_\oplus}, \|U\|_\oplus) \in [0, 1). \tag{5}$$

Notice that the contraction rate $c(\|U\|_\oplus)$ is an increasing function of $\|U\|_\oplus$ taking values in $[0, 1)$. Thus, by the Banach fixed-point theorem, we obtain the following result.

Corollary 3 *Let c be the contraction rate given by (5). Then, there exists a unique stationary distribution $\mu_{\mathsf{MC}} \in \mathscr{P}_{\mu_{\mathsf{ref}}}$ satisfying $\mu_{\mathsf{MC}} = \mu_{\mathsf{MC}}\mathsf{P}$ and*

$$d_{\mathsf{TV}}(\mu_{\mathsf{ref}}\mathsf{P}^t, \mu_{\mathsf{MC}}) \le d_{\mathsf{TV}}(\mu_{\mathsf{ref}}, \mu_{\mathsf{MC}}) \cdot c^t,$$

Corollary 3 implies that the stationary distribution μ_{MC} can be approximated by the t-step Markov chain $\mu_{\mathsf{ref}}\mathsf{P}^t$ as $t \to +\infty$. Importantly, the contraction rate c is bounded in terms of the seminorm $\|U\|_\oplus$, rather than the naive L^∞-norm $\|U\|_\infty$. Since the construction of the Markov kernel and the operation $\mu \mapsto \mu\mathsf{P}$ do not require prior knowledge of the value of $\|U\|_\oplus$, Corollary 3 suggests that MCHF *implicitly* adapts to the additive structure of U.

Next, we discuss algorithms for NLHF. Note that there are various algorithms for computing μ_{NL}; see, for example, [16, 22, 24]. Existing algorithms either provide convergence guarantees for the average iterate $T^{-1}\sum_{t=1}^T \mu_t$, or obtain last-iterate convergence by adding a local regularization term $D_{\mathsf{KL}}(\cdot \mid \mu_t)$ in the update from μ_t to μ_{t+1}. Here, we propose a mirror descent algorithm for NLHF, showing that if the step size is chosen carefully, the mirror descent enjoys the last-iterate convergence guarantee with a similar structure-adaptive convergence rate to MCHF.

Theorem 4 *Define the iteration (ν_t, μ_t) by*

$$\begin{aligned} \nu_t &= \operatorname*{argmin}_\nu \int \nu(dx)\mu_t(dy)U(x,y) + D_{\mathsf{KL}}(\nu \mid \mu_{\mathsf{ref}}), \\ \mu_{t+1} &= \operatorname*{argmax}_\mu \int \nu_t(dx)\mu(dy)U(x,y) - D_{\mathsf{KL}}(\mu \mid \mu_{\mathsf{ref}}) - \eta^{-1}D_{\mathsf{KL}}(\mu \mid \mu_t), \end{aligned} \tag{6}$$

initialized at $\mu_0 = \mu_{\mathsf{ref}}$. Then, for any step size η such that $0 < \eta \le \|U\|_\oplus^{-2}$, it holds that

$$D_{\mathsf{KL}}(\mu_{\mathsf{NL}} \mid \mu_t) \le (1+\eta)^{-t} \cdot D_{\mathsf{KL}}(\mu_{\mathsf{NL}} \mid \mu_{\mathsf{ref}}).$$

If we take the step size $\eta = \|U\|_{\oplus}^{-2}$, we obtain $D_{\mathsf{KL}}(\mu_{\mathsf{NL}} \mid \mu_t) \leq (1 + \|U\|_{\oplus}^{-2})^{-t} \cdot D_{\mathsf{KL}}(\mu_{\mathsf{NL}} \mid \mu_{\mathsf{ref}})$. A similar algorithm to (6) with a geometric convergence guarantee of the last iterate is given by previous works, such as [24], for the specific form $U(x, y) = \lambda^{-1}(\mathcal{P}(x \prec y) - 1/2)$. However, the convergence guarantee in previous work depends on $\|U\|_{\infty}$ rather than on $\|U\|_{\oplus}$. Since $\|U\|_{\oplus} \leq \|U\|_{\infty}$ always holds, and since $\|U\|_{\oplus}$ can be much smaller than $\|U\|_{\infty}$ when U has a strong additive component as in Example 1 and 2, Theorem 4 provides a sharper convergence rate for NLHF than previous works.

However, to achieve this refined convergence rate for the last iterate, we still need to choose the step size as $\eta = \|U\|_{\oplus}^{-2}$. Although $\|U\|_{\oplus}$ can be estimated explicitly up to a multiplicative constant as in Proposition 9 and 10, its exact calculation may be challenging in practice. Moreover, our algorithm still contains a local regularization term, which makes sampling from μ_t difficult. Indeed, letting $p_t = d\mu_t/d\mu_{\mathsf{ref}}$, the mirror descent update (6) can be written as

$$\nu_t(dx) \propto \exp\Big(-\int \mu_t(dy)U(x, y)\Big)\mu_{\mathsf{ref}}(dx), \quad \mu_{t+1}(dy) \propto \exp\Big(\frac{\eta}{\eta+1}\int \nu_t(dx)U(x, y)\Big)\tilde{\mu}_t^{\eta}(dy),$$

where $\tilde{\mu}_t^{\eta}(dy) \equiv (\frac{d\mu_t}{d\mu_{\mathsf{ref}}}(y))^{\frac{1}{1+\eta}}\mu_{\mathsf{ref}}(dy)$. The update of ν_t from μ_t is a simple exponential tilt of the reference distribution. In contrast, the update of μ_{t+1} requires computing the geometric mixture $\tilde{\mu}_t^{\eta}$ of μ_{ref} and μ_t, due to the local regularization term $D_{\mathsf{KL}}(\mu \mid \mu_t)$. Exact sampling from such a geometric mixture of probability distributions is challenging in practice; see [16, Section F.1] for a practical algorithm to approximately sample from such mixtures.

Finally, we show that when $\|U\|_{\oplus}$ is smaller than 1, the same algorithm (6) without the local regularization term $D_{\mathsf{KL}}(\mu|\mu_t)$ (i.e., $\eta = +\infty$) converges to the equilibrium at a geometric convergence rate in TV distance.

Theorem 5 *The mirror descent algorithm* (6) *with* $\eta = +\infty$ *satisfies*

$$d_{\mathsf{TV}}(\nu_t, \nu_{\mathsf{NL}}) \leq \|U\|_{\oplus} d_{\mathsf{TV}}(\mu_t, \mu_{\mathsf{NL}}), \quad d_{\mathsf{TV}}(\mu_{t+1}, \mu_{\mathsf{NL}}) \leq \|U\|_{\oplus} d_{\mathsf{TV}}(\nu_t, \nu_{\mathsf{NL}}).$$

Thus, if $\|U\|_{\oplus} < 1$*, we have* $d_{\mathsf{TV}}(\mu_t, \mu_{\mathsf{NL}}) \leq \|U\|_{\oplus}^{2t} \cdot d_{\mathsf{TV}}(\mu_{\mathsf{ref}}, \mu_{\mathsf{NL}}) \to 0$.

We emphasize that the convergence rate is bounded in terms of $\|U\|_{\oplus}$, rather than $\|U\|_{\infty}$. Thus, similarly to the MCHF update in Corollary 3, this algorithm implicitly adapts to the additive structure of U, whenever $\|U\|_{\oplus} < 1$.

2.3. Comparison between MCHF and NLHF

Computational perspective There is a computational trade-off between the two approaches. MCHF requires sampling from the conditional distribution $\mathsf{P}(x, \cdot)$, whereas NLHF repeatedly solves KL-regularized optimization problems.

As discussed in Section 2.2, last-iterate convergence of NLHF typically requires adding a local regularization term $\eta^{-1}D_{\mathsf{KL}}(\mu \mid \mu_{t-1})$ to the update, and because of this, each iteration requires both updating the reward function and sampling from a distribution involving a geometric mixture of μ_{t-1} and μ_{ref}. This additional sampling step can be practically challenging.

By contrast, MCHF requires implementing the conditional sampler $\mathsf{P}(x, dy)$, but this sampler needs to be implemented only once. Once such a sampling block is available, the alignment procedure simply consists of repeatedly applying the same Markov kernel. In this sense, MCHF can be viewed as attaching a preference-guided conditional sampling module on top of the existing architecture that generates samples from μ_{ref}.

Coupling perspective Another key difference between NLHF and MCHF lies in their algorithms. For the sake of simplicity, we focus on the antisymmetric utility U. Suppose μ^0 is a current distribution, and we update it via the following two different rules:

$$\mu^1_{\mathsf{MC}} = \mu^0 \mathsf{P}, \qquad \mu^1_{\mathsf{NL}} \in \operatorname*{argmax}_{\mu} \int \mu^0(dx)\mu(dy)U(x,y) - D_{\mathsf{KL}}(\mu \mid \mu_{\mathsf{ref}})$$

Note that μ_{MC} is the stationary point of the first update, while μ_{NL} is the equilibrium of the second update since we are assuming that U is antisymmetric now.

To compare these two updates, consider the couplings induced by MCHF and NLHF:

$$\pi_{\mathsf{MC}}(dx, dy) = \mu^0(dx)\mathsf{P}(x, dy), \quad \pi_{\mathsf{NL}}(dx, dy) = \mu^0(dx)\mu^1_{\mathsf{NL}}(dy).$$

The second marginal of these couplings is precisely μ^1_{MC} and μ^1_{NL}, respectively. Then, by the definition of Markov kernel P and the update rule of μ^1_{NL}, π_{MC} and π_{NL} maximize the same objective function but over different feasible spaces of couplings:

$$\pi_{\mathsf{MC}} = \operatorname*{argmax}_{\int \pi(\cdot, dy) = \mu^0(\cdot)} F(\pi), \quad \pi_{\mathsf{NL}} = \operatorname*{argmax}_{\exists \mu, \pi = \mu^0 \otimes \mu} F(\pi), \quad F(\pi) = \iint U(x,y)\pi(dx,dy) - D_{\mathsf{KL}}(\pi || \mu^0 \otimes \mu_{\mathsf{ref}})$$

Thus, the MCHF coupling optimizes over a larger class of couplings than that of NLHF, and in this sense the coupling induced by MCHF can extract more utility from the preference structure than the product coupling induced by NLHF.

This coupling perspective also highlights a practical difference between the two methods. Since MCHF produces a conditional distribution $\mathsf{P}(x, \cdot)$ given the current output x, it naturally supports an *inference-time refinement* procedure: one may repeatedly refine a model output by sampling from $\mathsf{P}(x, \cdot)$ until the user is satisfied. This is reminiscent of the inference-time refinement mechanism discussed in Stackelberg game-based alignment [19]. By contrast, the NLHF update directly produces a marginal distribution and therefore does not provide the same conditional refinement mechanism. For more details, see Section 4.2 of [11].

Alignment dynamics in the small $\|U\|_\oplus$ asymptotics By Proposition 9, $\|U\|_\oplus = 0$ is equivalent to $U(x,y) = \hat{g}(x) + \hat{f}(y)$, where $(\hat{g}, \hat{f})$ is a solution to the L^2 minimization problem $\min_{g,f} \|U - g \oplus f\|_2$, which can be written explicitly as the weighted row/column sum, as in (8). Moreover, by the definitions of MCHF and NLHF, if $U = g \oplus f$, then MCHF and NLHF equilibria both collapse to the RLHF solution with reward f, i.e., $\mu_{\mathsf{RL}}(f)(dy) \propto \exp(f(y))\mu_{\mathsf{ref}}(dy)$. We summarize the above discussion in the proposition below.

Proposition 6 *Let $\mu_{\mathsf{MC}}(U)$ and $\mu_{\mathsf{NL}}(U)$ be MCHF and NLHF based on utility function $U \in L^\infty(\mu_{\mathsf{ref}} \otimes \mu_{\mathsf{ref}})$ respectively. Then it holds that*

$$\|U\|_\oplus = 0 \quad \Leftrightarrow \quad U = \hat{g} \oplus \hat{f} \quad \Rightarrow \quad \mu_{\mathsf{MC}}(U) = \mu_{\mathsf{NL}}(U) = \mu_{\mathsf{RL}}(\hat{f})$$

where $\mu_{\mathsf{RL}}(\hat{f})(dy) \propto \exp(\hat{f}(y))\mu_{\mathsf{ref}}(dy)$ and $\hat{f}(y) = \int \mu_{\mathsf{ref}}(dx)U(x,y)$.

We aim to provide a quantitative version of this claim: if $\|U\|_\oplus$ is small but not exactly zero, how far are $\mu_{\mathsf{MC}}(U)$ and $\mu_{\mathsf{NL}}(U)$ from $\mu_{\mathsf{RL}}(\hat{f})$? This question is natural since $\|U\|_\oplus$ can be significantly small (relative to $\|U\|_\infty$) in natural human preference profiles (see Section A).

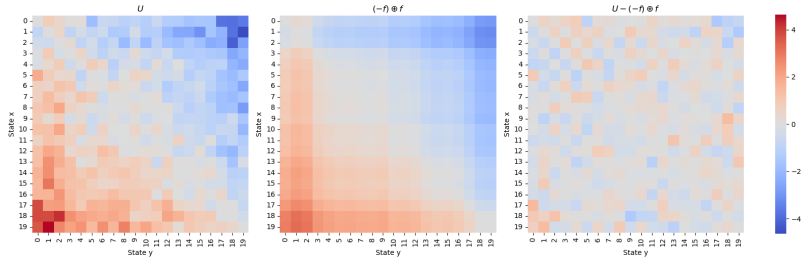

Figure 1: Comparison of antisymmetric utility U generated by (7), $(-\hat{f}) \oplus \hat{f}$ with $\hat{f}(y) = \int \mu_{\mathsf{ref}}(dx)U(x,y)$, and residual $U - (-\hat{f}) \oplus \hat{f}$.

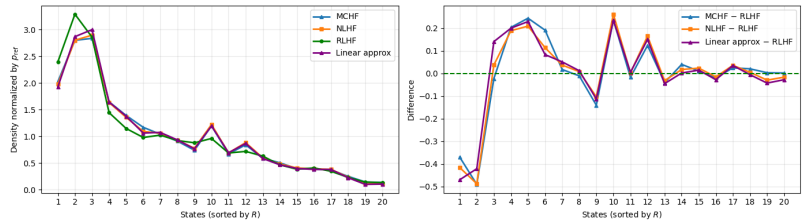

Figure 2: Comparison of MCHF, NLHF, RLHF, and the first-order approximation in Theorem 7.

Theorem 7 *For any sequence of antisymmetric utilities $U \in L^\infty(\mu_{\mathsf{ref}} \otimes \mu_{\mathsf{ref}})$ such that $\|U\|_\oplus \to 0$, letting $\hat{f}(y) = \int \mu_{\mathsf{ref}}(dx)U(x,y)$, for each $* \in \{\mathsf{MC}, \mathsf{NL}\}$,*

$$\left\|p_*(U) - p_{\mathsf{RL}}(\hat{f}) - p_{\mathsf{RL}}(\hat{f}) \odot \left(U - (-\hat{f}) \oplus \hat{f}\right)^\star p_{\mathsf{RL}}(\hat{f})\right\|_{L^1(\mu_{\mathsf{ref}})} = o(\|U\|_\oplus),$$

This implies that MCHF and NLHF have the same leading term in the small-$\|U\|_\oplus$ asymptotics. Combined with the triangle inequality, we also get $d_{\mathsf{TV}}(\mu_{\mathsf{MC}}(U), \mu_{\mathsf{NL}}(U)) = o(\|U\|_\oplus)$.

Let us verify Theorem 7 by numerical simulation. We set μ_{ref} to be the uniform distribution on a discrete space with $|\mathcal{X}| = 20$, and generate utility U as

$$U(x,y) = \frac{1}{2}\log\frac{\mathcal{P}(x \prec y)}{1 - \mathcal{P}(x \prec y)}, \qquad \mathcal{P}(x \prec y) = \Phi\Big(R(y) - R(x) + \frac{1}{2}E(x,y)\Big), \tag{7}$$

where Φ is the CDF of the standard normal distribution. We generate $R \sim N(0, I_n)$, and take $E \in \mathbb{R}^{n\times n}$ as an antisymmetric matrix whose off-diagonal entries are drawn from $N(0,1)$. The resulting pairwise utility U is visualized in Figure 1, which suggests that $\|U\|_\oplus \ll \|U\|_\infty$. Figure 2 plots MCHF, NLHF, RLHF, and the first-order approximation given by Theorem 7. We observe that MCHF and NLHF nearly coincide, and that their deviation from RLHF is accurately captured by the first-order approximation.

Next, we claim that MCHF and NLHF agree up to first order in their corresponding iterative dynamics as well.

Theorem 8 *For any antisymmetric utility $U \in L^\infty(\mu_{\mathsf{ref}} \otimes \mu_{\mathsf{ref}})$, consider the following iterations:*

$$\mu^t_{\mathsf{MC}} = \mu^{t-1}_{\mathsf{MC}}\mathsf{P}, \quad \mu^t_{\mathsf{NL}} = \operatorname*{argmax}_\mu \int \mu^{t-1}_{\mathsf{NL}}(dx)\mu(dy)U(x,y) - D_{\mathsf{KL}}(\mu||\mu_{\mathsf{ref}})$$

initialized at $\mu^0_{\mathsf{MC}} = \mu^0_{\mathsf{NL}} = \mu_{\mathsf{ref}}$*, where* P *is the Markov kernel based on* U*. Let* $p^t_*(U) = d\mu^t_*/d\mu_{\mathsf{ref}}$ *be the corresponding density for* $t \geq 1$ *and* $* \in \{\mathsf{MC}, \mathsf{NL}\}$*. Then, for any sequence of antisymmetric utilities* U *such that* $\|U\|_\oplus \to 0$*, we have that for each* $* \in \{\mathsf{MC}, \mathsf{NL}\}$*,*

$$\left\|p^1_*(U) - p_{\mathsf{RL}}(\hat{f})\right\|_{L^1(\mu_{\mathsf{ref}})} = o(\|U\|_\oplus),$$
$$\sup_{t\geq 2}\left\|p^t_*(U) - p_{\mathsf{RL}}(\hat{f}) - p_{\mathsf{RL}}(\hat{f}) \odot \left(U - (-\hat{f}) \oplus \hat{f}\right)^\star p_{\mathsf{RL}}(\hat{f})\right\|_{L^1(\mu_{\mathsf{ref}})} = o(\|U\|_\oplus).$$

From this theorem, we see that MCHF and NLHF agree at the iteration level up to first order. Moreover, this theorem sheds light on the dynamics of alignment. At the first iteration $t = 1$, both algorithms recover the RLHF update with reward $\hat{f}$, which is the weighted column sum of U. Then for $t \geq 2$, they incorporate the non-additive structure $U - (-\hat{f}) \oplus \hat{f}$ as a first-order correction to the RLHF update. Furthermore, comparing this correction with the leading term of the equilibrium in Theorem 7, we see that the two algorithms capture the leading term within the first two iterations.

References

[1] Yuntao Bai, Andy Jones, Kamal Ndousse, Amanda Askell, Anna Chen, Nova DasSarma, Dawn Drain, Stanislav Fort, Deep Ganguli, Tom Henighan, et al. Training a helpful and harmless assistant with reinforcement learning from human feedback. *arXiv preprint arXiv:2204.05862*, 2022.

[2] Ralph Allan Bradley and Milton E. Terry. Rank Analysis of Incomplete Block Designs: I. The Method of Paired Comparisons. *Biometrika*, 39(3/4):324–345, 1952.

[3] Sergey Brin and Lawrence Page. The anatomy of a large-scale hypertextual Web search engine. *Computer Networks and ISDN Systems*, 30(1):107–117, 1998. ISSN 0169-7552. Proceedings of the Seventh International World Wide Web Conference.

[4] Paul F. Christiano, Jan Leike, Tom B. Brown, Miljan Martic, Shane Legg, and Dario Amodei. Deep reinforcement learning from human preferences. In *Advances in Neural Information Processing Systems*, volume 30, 2017.

[5] Xu Chu, Zhixin Zhang, Tianyu Jia, and Yujie Jin. Stackelberg Self-Annotation: A Robust Approach to Data-Efficient LLM Alignment. In *The Thirty-ninth Annual Conference on Neural Information Processing Systems*, 2026.

[6] R. L. Dobrushin. Central Limit Theorem for Nonstationary Markov Chains. I. *Theory of Probability & Its Applications*, 1(1):65–80, 1956.

[7] Kawin Ethayarajh, Winnie Xu, Niklas Muennighoff, Dan Jurafsky, and Douwe Kiela. KTO: Model Alignment as Prospect Theoretic Optimization. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pages 12634–12651. PMLR, 2024.

[8] Mohammad Gheshlaghi Azar, Zhaohan Daniel Guo, Bilal Piot, Remi Munos, Mark Rowland, Michal Valko, and Daniele Calandriello. A General Theoretical Paradigm to Understand Learning from Human Preferences. In Sanjoy Dasgupta, Stephan Mandt, and Yingzhen

Li, editors, *Proceedings of The 27th International Conference on Artificial Intelligence and Statistics*, volume 238 of *Proceedings of Machine Learning Research*, pages 4447–4455. PMLR, 2024.

[9] Paul Gölz, Nika Haghtalab, and Kunhe Yang. Distortion of AI Alignment: Does Preference Optimization Optimize for Preferences? In *The Thirty-ninth Annual Conference on Neural Information Processing Systems*, 2026.

[10] Audrey Huang, Wenhao Zhan, Tengyang Xie, Jason Lee, Wen Sun, Akshay Krishnamurthy, and Dylan Foster. Correcting the Mythos of KL-Regularization: Direct Alignment without Overoptimization via Chi-Squared Preference Optimization. In *International Conference on Learning Representations*, volume 2025, pages 92647–92701, 2025.

[11] Takuya Koriyama and Tengyuan Liang. A Markov Chain Approach to Preference Alignment. *arXiv preprint arXiv:2606.22652*, 2026.

[12] Steven G. Krantz and Harold R. Parks. *The Implicit Function Theorem: History, Theory, and Applications*. Modern Birkhäuser Classics. Birkhäuser New York, New York, NY, 2013. doi: 10.1007/978-1-4614-5981-1.

[13] Haihao Lu, Robert M. Freund, and Yurii Nesterov. Relatively smooth convex optimization by first-order methods, and applications. *SIAM Journal on Optimization*, 28(1):333–354, 2018.

[14] R. Duncan Luce. *Individual Choice Behavior: A Theoretical Analysis*. Wiley, 1959.

[15] Yu Meng, Mengzhou Xia, and Danqi Chen. SimPO: Simple Preference Optimization with a Reference-Free Reward. In *Advances in Neural Information Processing Systems*, volume 37, pages 124198–124235. Curran Associates, Inc., 2024.

[16] Rémi Munos, Michal Valko, Daniele Calandriello, Mohammad Gheshlaghi Azar, Mark Rowland, Zhaohan Daniel Guo, Yunhao Tang, Matthieu Geist, Thomas Mesnard, Côme Fiegel, et al. Nash Learning from Human Feedback. In *Forty-first International Conference on Machine Learning*, 2024.

[17] Sahand Negahban, Sewoong Oh, and Devavrat Shah. Rank centrality: Ranking from pairwise comparisons. *Operations Research*, 65(1):266–287, 2017.

[18] Long Ouyang, Jeffrey Wu, Xu Jiang, Diogo Almeida, Carroll L. Wainwright, Pamela Mishkin, Chong Zhang, Sandhini Agarwal, Katarina Slama, Alex Ray, et al. Training language models to follow instructions with human feedback. In *Advances in Neural Information Processing Systems*, volume 35, pages 27730–27744, 2022.

[19] Barna Pásztor, Thomas Kleine Buening, and Andreas Krause. Stackelberg Learning from Human Feedback: Preference Optimization as a Sequential Game. In *The Fourteenth International Conference on Learning Representations*, 2026.

[20] 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.

[21] Jeffrey S. Rosenthal. Minorization Conditions and Convergence Rates for Markov Chain Monte Carlo. *Journal of the American Statistical Association*, 90(430):558–566, 1995.

[22] Corby Rosset, Ching-An Cheng, Arindam Mitra, Michael Santacroce, Ahmed Awadallah, and Tengyang Xie. Direct Nash Optimization: Teaching Language Models to Self-Improve with General Preferences. *arXiv preprint arXiv:2404.03715*, 2024.

[23] Zhekun Shi, Kaizhao Liu, Qi Long, Weijie J Su, and Jiancong Xiao. Fundamental Limits of Game-Theoretic LLM Alignment: Smith Consistency and Preference Matching. *arXiv preprint arXiv:2505.20627*, 2025.

[24] Daniil Tiapkin, Daniele Calandriello, Denis Belomestny, Eric Moulines, Alexey Naumov, Kashif Rasul, Michal Valko, and Pierre Menard. Accelerating Nash Learning from Human Feedback via Mirror Prox. *arXiv preprint arXiv:2505.19731*, 2025.

[25] Fang Wu, Xu Huang, Weihao Xuan, Zhiwei Zhang, Yijia Xiao, Guancheng Wan, Xiaomin Li, Bing Hu, Peng Xia, Jure Leskovec, and Yejin Choi. Multiplayer Nash Preference Optimization. In *The Fourteenth International Conference on Learning Representations*, 2026.

Contents

1 Introduction **1**

2 Main result **3**
2.1 Definition of seminorm: measuring non-additive structure of utility . . . 3
2.2 Structure-Adaptive Convergence of MCHF and NLHF . . . 4
2.3 Comparison between MCHF and NLHF . . . 5

Notation **12**

A Examples of pairwise utilities and the corresponding seminorm $\|U\|_\oplus$ **12**

B Properties of seminorm $\|\cdot\|_\oplus$ **13**
B.1 Proof of Proposition 9 . . . 14
B.2 Proof of Proposition 10 . . . 14

C Proofs for Section 2.2 **15**
C.1 Proof of Theorem 2 . . . 15
C.1.1 Proof of $d_{\mathsf{TV}}(\mu\mathsf{P}, \nu\mathsf{P}) \leq (1 - e^{-2\|U\|_\oplus}) d_{\mathsf{TV}}(\mu, \nu)$ . . . 15
C.1.2 Proof of $d_{\mathsf{TV}}(\mu\mathsf{P}, \nu\mathsf{P}) \leq \|U\|_\oplus d_{\mathsf{TV}}(\mu, \nu)$ . . . 16
C.1.3 Proof of Lemma 12 . . . 17
C.1.4 Proof of Lemma 13 . . . 19
C.2 Proof of Theorem 4 . . . 19
C.2.1 Proof of Lemma 15 . . . 21
C.2.2 Proof of Lemma 16 . . . 21
C.3 Proof of Theorem 5 . . . 22

D Fréchet Differentiability and Stability **23**
D.1 Fréchet differentiability . . . 23
D.1.1 Fréchet differentiability of MCHF . . . 23
D.1.2 Fréchet differentiability of NLHF . . . 25
D.2 Stability under perturbations of utility . . . 27
D.2.1 Stability of MCHF . . . 27
D.2.2 Stability of NLHF . . . 29
D.2.3 Proof of Lemma 24 . . . 31
D.3 Uniform Fréchet differentiability . . . 33
D.3.1 MCHF . . . 34
D.3.2 NLHF . . . 36

E Alignment dynamics of MCHF and NLHF **38**
E.1 Proof of Theorem 7 . . . 38
E.2 Proof of Proposition 31 . . . 38
E.3 Proof of Theorem 8 . . . 40

Notation

Notation	Meaning
μ_{ref}	Reference distribution, e.g. the pre-trained model.
$U(x, y)$	Pairwise utility; larger values mean that y is preferred to x.
$g \oplus f$	Additive function defined by $(g \oplus f)(x, y) = g(x) + f(y)$.
$\|U\|_\infty$	Essential supremum norm on $L^\infty(\mu_{\mathsf{ref}} \otimes \mu_{\mathsf{ref}})$.
$\|U\|_\oplus$	Additive-defect seminorm: $\inf_{g,f \in L^\infty(\mu_{\mathsf{ref}})} \|U - g \oplus f\|_\infty$.
$\hat{f}$	Column-sum projection: $\hat{f}(y) = \int U(x, y)\mu_{\mathsf{ref}}(dx)$.
$\hat{g}$	Row-sum projection: $\hat{g}(x) = \int U(x, y)\mu_{\mathsf{ref}}(dy) - m$, where $m = \int U(x, y)\mu_{\mathsf{ref}}(dx)\mu_{\mathsf{ref}}(dy)$.
$\mathsf{P}(x, dy)$	MCHF Markov kernel: $\mathsf{P}(x, dy) = Z(x)^{-1} \exp(U(x, y))\mu_{\mathsf{ref}}(dy)$.
$\mu\mathsf{P}$	One-step update of μ by the Markov kernel P.
$\mu_{\mathsf{MC}}(U)$	Stationary distribution of the MCHF Markov kernel based on U.
$\mu_{\mathsf{NL}}(U)$	NLHF equilibrium distribution based on U.
$\nu_{\mathsf{NL}}(U)$	Adversarial/opponent equilibrium distribution in NLHF based on U.
$\mu_{\mathsf{RL}}(f)$	RLHF distribution with reward f: $\mu_{\mathsf{RL}}(f)(dy) \propto \exp(f(y))\mu_{\mathsf{ref}}(dy)$.
$p_*(U)$	Density $d\mu_*(U)/d\mu_{\mathsf{ref}}$ for $* \in \{\mathsf{MC}, \mathsf{NL}\}$.
$p_{\mathsf{RL}}(f)$	Density $d\mu_{\mathsf{RL}}(f)/d\mu_{\mathsf{ref}}$.
Kh	Integral operator $Kh(x) = \int K(x, y)h(y)\mu_{\mathsf{ref}}(dy)$.
$K^\star h$	Adjoint operator: $K^\star h(y) = \int K(x, y)h(x)\mu_{\mathsf{ref}}(dx)$.
$\odot$	Entrywise product.
L^p	Set of $L^p(\mu_{\mathsf{ref}})$-measurable functions.
L^p_0	Set of $L^p(\mu_{\mathsf{ref}})$-measurable functions with mean zero: $L^p_0 = \{f \in L^p : \int \mu_{\mathsf{ref}}(dy)f(y) = 0\}$.
Δ^∞	Density simplex: $\Delta^\infty = \{p \in L^\infty : p(\cdot) \geq 0, \int \mu_{\mathsf{ref}}(dz)p(z) = 1\}$.
s	Softmax operator $\mathsf{s} : L^\infty \to \Delta^\infty$ defined by $\mathsf{s}(f)(x) = \frac{\exp(f(x))}{\int \mu_{\mathsf{ref}}(dx')\exp(f(x'))}$.
J_p	Demean operator $J_p : L^1 \to L^1_0$ for each $p \in \Delta^\infty$: $J_p f = p \odot \big(f - \int \mu_{\mathsf{ref}}(dz')p(z')f(z')\big)$.

Appendix A. Examples of pairwise utilities and the corresponding seminorm $\|U\|_\oplus$

Example 1 (Logit transform) *Suppose U is the logit transform of the preference probability:*

$$U(x, y) = \sigma^{-1}_{sigmoid} \circ \mathcal{P}(x \prec y) \quad \textit{where} \quad \sigma^{-1}_{\text{sigmoid}}(t) = \log \frac{t}{1-t}$$

and the preference model $\mathcal{P}(x \prec y)$ follows

$$\mathcal{P}(x \prec y) = \Phi(R(y) - R(x) + E(x, y)),$$

where Φ is a nonlinear link function and $E(x, y)$ is a pairwise interaction term. In particular, if $\mathcal{P}(x \prec y)$ follows the BTL model (i.e., $\Phi = \sigma_{\text{sigmoid}}$ and $E = 0$), we have $U(x, y) = R(y) - R(x)$ and hence $\|U\|_\oplus = 0$. More generally, whenever Φ is close to σ_{sigmoid} and the pairwise interaction term E is small, $U(x, y)$ is close to $R(y) - R(x)$, and hence $\|U\|_\oplus$ is small (see Figure 1).

Example 2 *(Linear transform) Suppose*

$$U(x, y) = \mathcal{P}(x \prec y) - 1/2,$$

where $\mathcal{P}(x \prec y)$ follows the BTL model $\mathcal{P}(x \prec y) = \sigma_{\text{sigmoid}}(R(y) - R(x))$ for a reward R, and assume that $\|R\|_\infty$ is small. Then, using the linear approximation of the sigmoid, $\sigma_{\text{sigmoid}}(x) = \frac{1}{2} + \frac{x}{4} + O(x^3)$ as $x \to 0$, one can show

$$\|U\|_\oplus = O(\|(-R) \oplus R\|_\infty^3) \quad \textit{and} \quad \|U\|_\infty = 4^{-1}\|(-R) \oplus R\|_\infty + O(\|(-R) \oplus R\|_\infty^3)$$

so that the ratio is significantly small: $\|U\|_\oplus / \|U\|_\infty = O(\|(-R) \oplus R\|_\infty^2) = O(\|R\|_\infty^2) \to 0$.

Appendix B. Properties of seminorm $\|\cdot\|_\oplus$

Proposition 9 *For any $U \in L^\infty(\mu_{\mathsf{ref}} \otimes \mu_{\mathsf{ref}})$, let $\square(U)$ be the rectangle defect defined as $\square(U) = \operatorname{esssup}_{x,x',y,y'} |U(x', y') - U(x, y') - U(x', y) + U(x, y)|$, and let $(\hat{g}, \hat{f})$ be any solution to the L^2 minimization $\min_{g,f} \|U - g \oplus f\|_2$. Then*

$$\square(U) \asymp \|U\|_\oplus \asymp \|U - \hat{g} \oplus \hat{f}\|_\infty,$$

or more precisely $\frac{1}{4}\square(U) \le \|U\|_\oplus \le \|U - \hat{g} \oplus \hat{f}\|_\infty \le \square(U)$.

We see that the rectangle defect $\square(U)$ is an explicit quantity by its definition, measuring the *intransitivity* of the utility for any cycle consisting of 4 points. The plug-in estimate $\|U - \hat{g} \oplus \hat{f}\|_\infty$ can also be computed explicitly, since $(\hat{g}, \hat{f})$ is the minimizer of $\min_{g,f} \|U - g \oplus f\|_2$ and one of the solutions is given by the weighted row/column sums:

$$\hat{g}(x) = \int U(x, y)\mu_{\mathsf{ref}}(dy) - m, \qquad \hat{f}(y) = \int U(x, y)\mu_{\mathsf{ref}}(dx), \tag{8}$$

where $m = \iint U(x, y)\mu_{\mathsf{ref}}(dx)\mu_{\mathsf{ref}}(dy)$. In particular, if $\|U\|_\oplus = 0$, then we have $U = \hat{g} \oplus \hat{f}$.

Finally, when U is antisymmetric, it holds that $\hat{g} = -\hat{f}$, and we obtain a sharper characterization of the seminorm $\|U\|_\oplus$.

Proposition 10 *Suppose $U \in L^\infty(\mu_{\mathsf{ref}} \otimes \mu_{\mathsf{ref}})$ is antisymmetric, i.e., $U(x, y) = -U(y, x)$. Let $\Delta(U)$ be the triangle defect defined by $\Delta(U) = \operatorname{esssup}_{x,y,z} |U(x, y) + U(y, z) + U(z, x)|$ and let $\hat{f}(y) = \int \mu_{\mathsf{ref}}(dx)U(x, y)$ be the weighted column sum. Then, we have*

$$\|U\|_\oplus \asymp \Delta(U) \asymp \|U - (-\hat{f}) \oplus \hat{f}\|_\infty,$$

or more precisely, $\frac{1}{3}\Delta(U) \le \|U\|_\oplus \le \|U - (-\hat{f}) \oplus \hat{f}\|_\infty \le \Delta(U)$.

We see that $\Delta(U)$ measures the intransitivity of U for any cycle consisting of 3 points. Furthermore, if $\|U\|_\oplus = 0$, then we have $U = (-\hat{f}) \oplus \hat{f}$ where $\hat{f}(y) = \int \mu_{\mathsf{ref}}(dx)U(x, y)$ is the weighted column sum.

B.1. Proof of Proposition 9

Take $(\hat{g}, \hat{f}) \in \mathrm{argmin}_{g,f \in L^2} \|U - g \oplus f\|_2$ as

$$\hat{g}(x) = \int U(x, y')\mu_{\mathsf{ref}}(dy') - m, \quad \hat{f}(y) = \int U(x', y)\mu_{\mathsf{ref}}(dx'), \quad m = \int U(x', y)\mu_{\mathsf{ref}}(dx')\mu_{\mathsf{ref}}(dy')$$

Here, we have $\hat{g}, \hat{f} \in L^\infty$ by $U \in L^\infty(\mu_{\mathsf{ref}} \otimes \mu_{\mathsf{ref}})$, and hence $\|U\|_\oplus \le \|U - \hat{g} \oplus \hat{f}\|_\infty$ holds by the definition of $\|U\|_\oplus$. Thus, it suffices to show $4^{-1}\square(U) \le \|U\|_\oplus$ and $\|U - \hat{g} \oplus \hat{f}\|_\infty \le \square(U)$.

Proof of $4^{-1}\square(U) \le \|U\|_\oplus$ Fix arbitrary $g, f \in L^\infty$ and let $U^{g\oplus f} = U - g \oplus f$. For every x, x', y, y', noting that the g- and f-terms cancel,

$$U(x', y') - U(x, y') - U(x', y) + U(x, y) = U^{g\oplus f}(x', y') - U^{g\oplus f}(x, y') - U^{g\oplus f}(x', y) + U^{g\oplus f}(x, y),$$

Taking $\mathrm{esssup}_{x,y,x',y'}$ on both sides, the LHS becomes $\square(U)$, while using the triangle inequality for the RHS, we get $\square(U) \le 4\|U^{g\oplus f}\|_\infty$. Taking $\inf_{g,f\in L^\infty}$ on the RHS, we get $\square(U) \le 4\|U\|_\oplus$ and complete the proof.

Proof of $\|U - \hat{g} \oplus \hat{f}\|_\infty \le \square(U)$ By the definition of $\hat{g}$ and $\hat{f}$, for all (x, y),

$$(U - \hat{g} \oplus \hat{f})(x, y) = \int \Big(U(x, y) - U(x', y) - U(x, y') + U(x', y')\Big)\mu_{\mathsf{ref}}(dx')\mu_{\mathsf{ref}}(dy').$$

Hence, taking $\mathrm{esssup}_{x,y} |\cdot|$ on both sides,

$$\|U - \hat{g} \oplus \hat{f}\|_\infty \le \operatorname*{esssup}_{x,y,x',y'} \big|U(x, y) - U(x', y) - U(x, y') + U(x', y')\big| = \square(U),$$

so the proof is complete.

B.2. Proof of Proposition 10

We first claim $\|U\|_\oplus = \inf_f \|U - (-f) \oplus f\|_\infty$. To prove this, it suffices to show $\inf_{g,f} \|U - g \oplus f\|_\infty \ge \inf_f \|U - (-f) \oplus f\|_\infty$. Let us fix $g, f \in L^\infty$ and let $h = 2^{-1}(-g + f)$. Note

$$\big((-h) \oplus h\big)(x, y) = 2^{-1}\big(g(x) - f(x) - g(y) + f(y)\big) = 2^{-1}\big(g \oplus f - f \oplus g\big)(x, y)$$

and the $\|U + f \oplus g\|_\infty = \|-U + g \oplus f\|_\infty$ since U is antisymmetric. Therefore, by the triangle inequality,

$$\inf_f \|U - (-f) \oplus f\|_\infty \le \|U - (-h) \oplus h\|_\infty \le 2^{-1}\|U - g \oplus f\|_\infty + 2^{-1}\|U + f \oplus g\|_\infty = \|U - g \oplus f\|_\infty.$$

Taking $\inf_{g,f}$ on the RHS, we obtain the claim.

By $\|U\|_\oplus = \inf_f \|U - (-f) \oplus f\|_\infty$, we have $\|U\|_\oplus \le \|U - (-\hat{f}) \oplus \hat{f}\|_\infty$ for $\hat{f}(y) = \int \mu_{\mathsf{ref}}(dx)U(x, y)$. Furthermore, using $\|U\|_\oplus = \inf_f \|U - (-f) \oplus f\|_\infty$ and the same argument as that of Section B.1, one can show $3^{-1}\Delta(U) \le \|U\|_\oplus$ and $\|U - (-\hat{f}) \oplus \hat{f}\|_\infty \le \Delta(U)$. Thus, we complete the proof.

Appendix C. Proofs for Section 2.2

C.1. Proof of Theorem 2

We prove $d_{\mathsf{TV}}(\mu\mathsf{P}, \nu\mathsf{P}) \le (1 - e^{-2\|U\|_\oplus}) d_{\mathsf{TV}}(\mu, \nu)$ and $d_{\mathsf{TV}}(\mu\mathsf{P}, \nu\mathsf{P}) \le \|U\|_\oplus d_{\mathsf{TV}}(\mu, \nu)$ separately.

C.1.1. PROOF OF $d_{\mathsf{TV}}(\mu\mathsf{P}, \nu\mathsf{P}) \le (1 - e^{-2\|U\|_\oplus}) d_{\mathsf{TV}}(\mu, \nu)$

Let $L^1 = L^1(\mu_{\mathsf{ref}})$ and $L^1_0 = \{h \in L^1 : \int \mu_{\mathsf{ref}}(dy) h(y) = 0\}$. We denote the L_p norm as $\|f\|_p = \|f\|_{L^p}$. We also write $L^\infty = L^\infty(\mu_{\mathsf{ref}} \otimes \mu_{\mathsf{ref}})$ and $\|U\|_\infty = \|U\|_{L^\infty(\mu_{\mathsf{ref}} \otimes \mu_{\mathsf{ref}})}$ for any $U \in L^\infty(\mu_{\mathsf{ref}} \otimes \mu_{\mathsf{ref}})$ when the context is clear. For any $U \in L^\infty(\mu_{\mathsf{ref}} \otimes \mu_{\mathsf{ref}})$, define the Markov transition density $P \in L^\infty(\mu_{\mathsf{ref}} \otimes \mu_{\mathsf{ref}})$ as

$$P(x, y) = \frac{\exp(U(x, y))}{\int \mu_{\mathsf{ref}}(dy') \exp(U(x, y'))}. \tag{9}$$

For all $\mu, \nu \in \mathscr{P}_{\mu_{\mathsf{ref}}}$, we can write the TV distance as

$$d_{\mathsf{TV}}(\mu, \nu) = \frac{1}{2}\Big\|\frac{d\mu}{d\mu_{\mathsf{ref}}} - \frac{d\nu}{d\mu_{\mathsf{ref}}}\Big\|_1, \quad d_{\mathsf{TV}}(\mu\mathsf{P}, \nu\mathsf{P}) = \frac{1}{2}\Big\||P^\star\Big(\frac{d\mu}{d\mu_{\mathsf{ref}}} - \frac{d\nu}{d\mu_{\mathsf{ref}}}\Big)\Big\|_1$$

where $P^\star h(y) = \int \mu_{\mathsf{ref}}(dx) h(x) P(x, y)$ is the adjoint integral operator. Since $\frac{d\mu}{d\mu_{\mathsf{ref}}} - \frac{d\nu}{d\mu_{\mathsf{ref}}} \in L^1_0 = \{h \in L^1 : \int \mu_{\mathsf{ref}}(dy) h(y) = 0\}$, it suffices to show the contraction rate of $P^\star$ on the Banach space $(L^1_0, \|\cdot\|_1)$.

Lemma 11 *Let P be the Markov density given by (9). Then, we have*

$$\sup_{h \in L^1_0 : \|h\|_1 \le 1} \|P^\star h\|_1 \le 1 - \exp(-2\|U\|_\oplus)$$

where $L^1_0 = \{h \in L^1 : \int \mu_{\mathsf{ref}}(dy) h(y) = 0\}$.

Proof We prove this lemma using the standard minorization condition [21]. Let us fix $g, f \in L^\infty$ and let $U^{g \oplus f} = U - g \oplus f$. Define $p_f \in L^\infty$ as

$$p_f(y) = \frac{\exp(f(y))}{\int \mu_{\mathsf{ref}}(dy') \exp(f(y'))}$$

so that p_f is a valid density with respect to μ_{ref}, i.e., $p_f(\cdot) \ge 0$ and $\int \mu_{\mathsf{ref}}(dy) p_f(y) = 1$. Since the factor $\exp(g(x))$ cancels between the numerator and the denominator in the definition of P (see (9)), the Markov density P can be rewritten as

$$P(x, y) = \frac{\exp(U(x, y) - g(x) - f(y)) e^{f(y)}}{\int \mu_{\mathsf{ref}}(dy') \exp(U(x, y') - g(x) - f(y')) e^{f(y')}} = \frac{\exp(U^{g \oplus f}(x, y)) p_f(y)}{\int \mu_{\mathsf{ref}}(dy') \exp(U^{g \oplus f}(x, y')) p_f(y')} \tag{10}$$

Then, for almost every (x, y) with respect to the product measure $\mu_{\mathsf{ref}} \otimes \mu_{\mathsf{ref}}$, using $\int \mu_{\mathsf{ref}}(dy') p_f(y') = 1$,

$$P(x, y) \ge \frac{\exp(-\|U^{g \oplus f}\|_\infty)}{\exp(\|U^{g \oplus f}\|_\infty)} \cdot \frac{p_f(y)}{\int \mu_{\mathsf{ref}}(dy') p_f(y')} = \delta^{g \oplus f} \cdot p_f(y)$$

where $\delta^{g\oplus f} = \exp(-2\|U^{g\oplus f}\|_\infty) \in (0,1]$. If we define $\tilde{P}$ as

$$\tilde{P}(x,y) = \frac{P(x,y) - \delta^{g\oplus f} \cdot p_f(y)}{1 - \delta^{g\oplus f}}$$

then $\tilde{P}$ satisfies $\tilde{P}(x,\cdot) \geq 0$ and $\int \mu_{\mathsf{ref}}(dy)\tilde{P}(x,y) = 1$, that is, $\tilde{P}$ is a valid Markov transition density with respect to μ_{ref}. Rearranging this,

$$P(x,y) = \delta^{g\oplus f} \cdot p_f(y) + (1 - \delta^{g\oplus f})\tilde{P}(x,y).$$

For any $h \in L^1_0$, noting $\int \mu_{\mathsf{ref}}(dx)h(x) = 0$,

$$P^\star h(y) = \int \mu_{\mathsf{ref}}(dx)P(x,y)h(x) = (1 - \delta^{g\oplus f})\tilde{P}^\star h(y)$$

Using the Data Processing Inequality: $\|P^\star h\|_1 \leq \|h\|_1$, which follows by

$$\|\tilde{P}^\star h\|_1 = \int \mu_{\mathsf{ref}}(dy)\Big|\int \mu_{\mathsf{ref}}(dx)\tilde{P}(x,y)h(x)\Big| \leq \int \mu_{\mathsf{ref}}(dx)\mu_{\mathsf{ref}}(dy)\tilde{P}(x,y)|h(x)| = \int \mu_{\mathsf{ref}}(dx)|h(x)|$$

where we have used the Fubini's theorem and the fact that $\tilde{P}$ is the Markov transition probability, we get

$$\sup_{h\in L^1_0:\|h\|_1\leq 1} \|P^\star h\|_1 \leq (1 - \delta^{g\oplus f}) \sup_{h\in L^1_0:\|h\|_1\leq 1} \|\tilde{P}^\star h\|_1 \leq (1 - \delta^{g\oplus f}).$$

Since $\delta^{g\oplus f} = \exp(-2\|U - g \oplus f\|_\infty)$, taking $\inf_{g,f}$ on the RHS, we complete the proof. ■

C.1.2. PROOF OF $d_{\mathsf{TV}}(\mu\mathsf{P}, \nu\mathsf{P}) \leq \|U\|_\oplus d_{\mathsf{TV}}(\mu,\nu)$

It suffices to show $\sup_{h\in L^1_0:\|h\|_1\leq 1} \|P^\star h\|_1 \leq \|U\|_\oplus$. We prove this using the following two standard lemmas.

Lemma 12 *For any $P \in L^\infty(\mu_{\mathsf{ref}} \otimes \mu_{\mathsf{ref}})$,*

$$\sup_{h\in L^1_0:\|h\|_1\leq 1} \|P^\star h\|_1 = \operatorname*{esssup}_{x,x'} \frac{1}{2}\|P(x,\cdot) - P(x',\cdot)\|_1$$

where $\operatorname{esssup}_{x,x'}$ is taken with respect to the product measure $\mu_{\mathsf{ref}} \otimes \mu_{\mathsf{ref}}$.

Proof This identity, when μ_{ref} is a simplex over finite space, is known as classical Dobrushin ergodicity coefficient [6]. We prove this result for general probability measure μ_{ref} using a standard argument; see Section C.1.3. ■

Lemma 13 *Fix a probability measure μ and define the map $\mathsf{s} : L^\infty(\mu) \to L^\infty(\mu)$ as*

$$\mathsf{s}(g) \equiv \frac{\exp(g(y))}{\int \mu(dy')\exp(g(y'))}.$$

Then, for all $g, \tilde{g} \in L^\infty(\mu)$, we have $\|\mathsf{s}(g) - \mathsf{s}(\tilde{g})\|_{L^1(\mu)} \leq \|g - \tilde{g}\|_{L^\infty(\mu)}$.

Proof This is essentially the Lipschitz continuity of the softmax function. See Section C.1.4 for the proof. ■

Applying Lemma 12, we have

$$\sup_{h\in L_0^1:\|h\|_1\le 1}\|P^\star h\|_1 = \operatorname*{esssup}_{x,x'}\frac{1}{2}\|P(x,\cdot)-P(x',\cdot)\|_1.$$

By (10) $P(x,y)$ can be written as

$$P(x,y)=\frac{\exp(U^{g\oplus f}(x,y))p_f(y)}{\int \mu_{\mathsf{ref}}(dy')\exp(U^{g\oplus f}(x,y'))p_f(y')},\quad p_f(y)=\frac{\exp(f(y))}{\int \mu_{\mathsf{ref}}(dy')\exp(f(y'))}$$

for any $g,f\in L^\infty$, where $U^{g\oplus f}=U-g\oplus f$. Letting $\mu_f(dy)=\mu_{\mathsf{ref}}(dy)p_f(y)$ be the exponentially tilted probability measure by f, using Lemma 13 $\mu=\mu_f$ and $g(y)=U^{g\oplus f}(x,y)$ and $\tilde{g}(y)=U^{g\oplus f}(x',y)$, for each x,x',

$$\|P(x,\cdot)-P(x',\cdot)\|_1\le\|U^{g\oplus f}(x,\cdot)-U^{g\oplus f}(x',\cdot)\|_{L^\infty(\mu_f)}=\|U^{g\oplus f}(x,\cdot)-U^{g\oplus f}(x',\cdot)\|_{L^\infty(\mu_{\mathsf{ref}})},$$

where the second equation follows from the fact that μ_f and μ_{ref} are absolutely continuous with respect to each other. Thus, by the definition: $\operatorname{esssup}_y|f(y)|=\|f\|_{L^\infty(\mu_{\mathsf{ref}})}$, combining all together,

$$\sup_{h\in L_0^1:\|h\|_1\le 1}\|P^\star h\|_1\le\operatorname*{esssup}_{x,x'}\frac{1}{2}\operatorname*{esssup}_y|U^{g\oplus f}(x,y)-U^{g\oplus f}(x',y)|\le\|U^{g\oplus f}\|_\infty$$

where we used the triangle inequality for the last inequality. Since $U^{g\oplus f}=U-g\oplus f$, taking $\inf_{g,f}$ on the RHS, we obtain $\sup_{h\in L_0^1:\|h\|_1\le 1}\|P^\star h\|_1\le\|U\|_\oplus$ and the proof is complete. Below, we prove the intermediate lemmas we used.

C.1.3. PROOF OF LEMMA 12

Now we claim the following standard identity:

Lemma 14 *For all $g\in L^\infty$, we have*

$$\sup_{h\in L_0^1:\|h\|_1\le 1}\left|\int h(x)g(x)\mu_{\mathsf{ref}}(dx)\right|=\frac{1}{2}\operatorname{essosc}(g)$$

where $\operatorname{essosc}(g)=\operatorname{esssup}_x g(x)-\operatorname{essinf}_x g(x)=\operatorname{esssup}_{x,y}|g(x)-g(y)|$

Proof Fix $g\in L^\infty$ and a constant $c\in\mathbb{R}$. Note $\int h(x)g(x)\mu_{\mathsf{ref}}(dx)=\int h(x)(g(x)-c)\mu_{\mathsf{ref}}(dx)$ for $h\in L_0^1$. Then,

$$\sup_{h\in L_0^1:\|h\|_1\le 1}\left|\int h(x)g(x)\mu_{\mathsf{ref}}(dx)\right|=\sup_{h\in L_0^1:\|h\|_1\le 1}\left|\int h(x)(g(x)-c)\mu_{\mathsf{ref}}(dx)\right|\le\|g-c\mathbf{1}\|_\infty$$

where $\|g-c\mathbf{1}\|_\infty=\operatorname{esssup}_x|g(x)-c|$. If we take $c=c_g=2^{-1}(\operatorname{esssup}_x g(x)+\operatorname{essinf}_x g(x))$, we have

$$\|g-c\mathbf{1}\|_\infty=2^{-1}(\operatorname*{esssup}_x g(x)-\operatorname*{essinf}_x g(x))=2^{-1}\operatorname{essosc}(g)$$

which completes the proof of the upper bound.

To show a lower bound, let $M = \operatorname{esssup}_x g(x)$ and $m = \operatorname{essinf}_x g(x)$ so that $\operatorname{essosc} g = M - m$. By the definition of esssup and essinf, for any $\epsilon > 0$, the following two events Ω_ϵ^+ and Ω_ϵ^- have a positive measure under μ_{ref}:

$$\Omega_\epsilon^+ = \{x : g(x) > M - \epsilon\}, \quad \Omega_\epsilon^- = \{x : g(x) < m + \epsilon\}.$$

Now we define the measurable function h_ϵ as $h_\epsilon = \frac{\mathbf{1}\{\Omega_\epsilon^+\}}{2\mu_{\mathsf{ref}}(\Omega_\epsilon^+)} - \frac{\mathbf{1}\{\Omega_\epsilon^-\}}{2\mu_{\mathsf{ref}}(\Omega_\epsilon^-)}$. Notice that this is well-defined since $\mu_{\mathsf{ref}}(\Omega_\epsilon^\pm) > 0$, and satisfies $h_\epsilon \in L_0^1$, $\|h_\epsilon\|_1 \le 1$. Then,

$$\begin{aligned}\int h_\epsilon(x) g(x) \mu_{\mathsf{ref}}(dx) &= \frac{1}{2\mu_{\mathsf{ref}}(\Omega_\epsilon^+)} \int \mathbf{1}\{\Omega_\epsilon^+\} g(x) \mu_{\mathsf{ref}}(dx) - \frac{1}{2\mu_{\mathsf{ref}}(\Omega_\epsilon^-)} \int \mathbf{1}\{\Omega_\epsilon^-\} g(x) \mu_{\mathsf{ref}}(dx) \\ &\ge 2^{-1}(M - \epsilon) - 2^{-1}(m + \epsilon) \\ &= 2^{-1}(M - m) - \epsilon.\end{aligned}$$

Taking $\epsilon \to 0$, we get the lower bound. ■

Let us prove Lemma 12 using Lemma 14. By the duality of L^1-L^∞ norms,

$$\|P^\star h\|_1 = \sup_{\|f\|_\infty \le 1} \Big| \int \mu_{\mathsf{ref}}(dy) f(y) P^\star h(y) \Big|.$$

Using Fubini's theorem, for all $f \in L^\infty$,

$$\int \mu_{\mathsf{ref}}(dy) f(y) P^\star h(y) = \int \mu_{\mathsf{ref}}(dy) f(y) \int \mu_{\mathsf{ref}}(dx) P(x, y) h(x) = \int \mu_{\mathsf{ref}}(dx) h(x) P f(x)$$

where $Pf(x) \equiv \int \mu_{\mathsf{ref}}(dy) P(x, y) f(y)$. Then, applying Lemma 14 with $g = Pf$, swapping the order of $\sup_{h \in L_0^1 : \|h\|_1 = 1}$ and $\sup_{\|f\|_\infty \le 1}$,

$$\sup_{h \in L_0^1 : \|h\|_1 \le 1} \|P^\star h\|_1 = \sup_{\|f\|_\infty \le 1} \sup_{h \in L_0^1 : \|h\|_1 \le 1} \Big| \int \mu_{\mathsf{ref}}(dy) f(y) P^\star h(y) \Big| = \frac{1}{2} \sup_{\|f\|_\infty \le 1} \operatorname{essosc}(Pf)$$

where

$$\sup_{\|f\|_\infty \le 1} \operatorname{essosc}(Pf) = \sup_{\|f\|_\infty \le 1} \operatorname*{esssup}_{x, x'} \Big| \int \mu_{\mathsf{ref}}(dy) f(y) \big(P(x, y) - P(x', y) \big) \Big|$$

Swapping the order of $\sup_{\|f\|_\infty \le 1}$ and $\operatorname{esssup}_{x, x'}$, using the duality of L^1-L^∞ norms, we get

$$\sup_{\|f\|_\infty \le 1} \operatorname{essosc}(Pf) = \operatorname*{esssup}_{x, x'} \|P(x, \cdot) - P(x', \cdot)\|_1,$$

so the proof is complete.

C.1.4. PROOF OF LEMMA 13

Let $h(x) = g(x) - \tilde{g}(x)$ and define the interpolator $g_t \in L^\infty$ for $t \in [0,1]$ as $g_t \equiv \tilde{g} + th$ so that $g_0 = \tilde{g}$ and $g_1 = g$. Now we define the tilted probability measure μ_t as

$$\mu_t(dy) \equiv \mathsf{s}(g_t)(y)\mu(dy).$$

Note that μ_t and μ are absolutely continuous with respect to each other. Then,

$$\|\mathsf{s}(g) - \mathsf{s}(\tilde{g})\|_1 = \sup_{\|f\|_{L^\infty(\mu)} \le 1} \Big| \int \mu_{\mathsf{ref}}(dx) f(x)(\mathsf{s}(g_1)(x) - \mathsf{s}(g_0)(x)) \Big| = \sup_{\|f\|_{L^\infty(\mu)} \le 1} \Big| \phi_f(1) - \phi_f(0) \Big|$$

where

$$\forall t \in [0,1], \quad \phi_f(t) \equiv \int f(x)\mu_t(dx) \equiv \int \mu(dx) f(x) \frac{\exp(g_t(x))}{\int \mu(dx') \exp(g_t(x'))}.$$

Here, since $g_t = \tilde{g} + th$ with $\tilde{g}, h \in L^\infty$, ϕ_f is C^1 on $[0,1]$ and the derivative is given by

$$\dot{\phi}_f(t) = \int \mu_t(dx) f(x) h(x) - \Big(\int \mu_t(dx) f(x) \Big) \cdot \Big(\int \mu_t(dx) h(x) \Big) = \mathrm{Cov}_{X \sim \mu_t}(f(X), h(X))$$

Using the Cauchy–Schwarz inequality $|\mathrm{Cov}(X,Y)| \le \sqrt{\mathrm{Var}(X)\mathrm{Var}(Y)} \le \|X\|_\infty \|Y\|_\infty$, noting that μ_t and μ are absolutely continuous with respect to each other, we obtain the uniform upper bound:

$$|\dot{\phi}_f(t)| \le \|f\|_{L^\infty(\mu_t)} \cdot \|h\|_{L^\infty(\mu_t)} = \|f\|_{L^\infty(\mu)} \|h\|_{L^\infty(\mu)}.$$

Therefore, by the fundamental theorem of calculus,

$$\|\mathsf{s}(g) - \mathsf{s}(\tilde{g})\|_1 = \sup_{\|f\|_{L^\infty(\mu)} \le 1} \Big| \phi_f(1) - \phi_f(0) \Big| \le \sup_{\|f\|_{L^\infty(\mu)} \le 1} \int_0^1 |\dot{\phi}_f(t)| dt \le \|h\|_{L^\infty(\mu)}.$$

This completes the proof.

C.2. Proof of Theorem 4

We use the following standard convergence guarantee for proximal mirror descents.

Lemma 15 *Let X be a convex subset of a Hilbert space $(\mathcal{H}, \langle\rangle)$ and consider the minimization problem*

$$\min_{p \in X} \ell(p) + w(p),$$

where both ℓ and w are differentiable and convex, and furthermore ℓ is relatively L-smooth with respect to w in the sense that

$$\forall p, q \in X, \qquad D_\ell(p\|q) \le L D_w(p\|q),$$

where $D_h(p\|q) = h(p) - h(q) - \langle \nabla h(q), p - q \rangle$ denotes the Bregman divergence associated with a convex function $h \in \{\ell, w\}$. Now we consider the (proximal) mirror descent iteration

$$p_{t+1} = \operatorname*{argmin}_{p \in X} \Big\{ \langle \nabla \ell(p_t), p \rangle + w(p) + \eta^{-1} D_w(p\|p_t) \Big\}.$$

Then for any $\eta \le 1/L$, letting $p_* \in \operatorname{argmin} \ell(p) + w(p)$ be the minimizer, we have

$$D_w(p_*\|p_t) \le (1+\eta)^{-t} D_w(p_*\|p_0).$$

Proof The proof relies on the relative smoothness framework introduced in [13]. To keep our paper self-contained, we prove this lemma in Section C.2.1. ■

Let us show Theorem 4 using Lemma 15. Let $p_{\mathsf{NL}} = \frac{d\mu_{\mathsf{NL}}}{d\mu_{\mathsf{ref}}}$ and $q_{\mathsf{NL}} = \frac{d\nu_{\mathsf{NL}}}{d\mu_{\mathsf{ref}}}$ be the densities of the NLHF solutions. Consider the density simplex $\Delta^\infty = \{f \in L^\infty : f(\cdot) \geq 0,\ \int \mu_{\mathsf{ref}}(dy) f(y) = 1\}$, which is a convex subset of the Hilbert space L^2 with the usual inner product $\langle f, g\rangle = \int \mu_{\mathsf{ref}}(dx) g(x) f(x)$, and define the linear operator $h \mapsto Uh$ by $Uh(x) = \int \mu_{\mathsf{ref}}(dy) U(x,y) h(y)$, and let $U^\star$ be its adjoint operator. Then, $(p_{\mathsf{NL}}, q_{\mathsf{NL}})$ solves the following minimax optimization problem:

$$\operatorname*{argmax}_{p\in\Delta^\infty} \min_{q\in\Delta^\infty} \langle q, Up\rangle + w(q) - w(p) \quad \text{where} \quad w(p) = \int \mu_{\mathsf{ref}}(dy) p(y) \log p(y)$$

If we define the convex function $\ell : \Delta^\infty \to \mathbb{R}$ as $\ell(p) = -(\min_{q\in\Delta^\infty} \langle q, Up\rangle + w(q))$ then p_{NL} is the solution to the following convex optimization problem:

$$p_{\mathsf{NL}} \in \operatorname*{argmin}_{p\in\Delta^\infty} \ell(p) + w(p).$$

Now we claim that ℓ is relatively $\|U\|_\oplus^2$-smooth with respect to w.

Lemma 16 *Fix $U \in L^\infty(\mu_{\mathsf{ref}} \otimes \mu_{\mathsf{ref}})$, and for each $p \in \Delta^\infty$, let $\ell(p) = -(\min_{q\in\Delta^\infty} \langle q, Up\rangle + w(q))$ and $q_p \in \operatorname{argmin}_{q\in\Delta^\infty} \langle q, Up\rangle + w(q)$. Then,*

1. *$\|q_p - q_{p'}\|_1 \leq \|U\|_\oplus \|p - p'\|_1$ for all $p, p' \in \Delta^\infty$.*
2. *ℓ is $\|U\|_\oplus^2$-smooth with respect to w, i.e., $D_\ell(p\|q) \leq \|U\|_\oplus^2 D_w(p\|q)$*

where $\|U\|_\oplus = \inf_{g,f\in L^\infty} \|U - g \oplus f\|_\infty$.

Proof See Section C.2.2 for the proof. ■

Thus, combined with Lemma 16-(2), we may apply Lemma 15 with

$$L = \|U\|_\oplus^2, \qquad w(p) = \int \mu_{\mathsf{ref}}(dy) p(y) \log p(y), \qquad \ell(p) = -\Big(\min_{q\in\Delta} \langle q, Up\rangle + w(q)\Big)$$

and conclude that for any $\eta \leq 1/\|U\|_\oplus^2$, the iteration p_t defined by

$$p_{t+1} = \operatorname*{argmin}_p \Big\{ \langle \nabla\ell(p_t), p\rangle + w(p) + \eta^{-1} D_w(p\|p_t) \Big\}$$

satisfies $D_w(p_*\|p_t) \leq (1+\eta)^{-t} \cdot D_w(p_*\|p_0)$. Here, by the envelope theorem, $\nabla\ell(p_t) = -U^\star q_t$ where $q_t = \operatorname{argmin}_q \langle q, Up_t\rangle + w(q)$. Thus, the total algorithm can be written as

$$q_t = \operatorname*{argmin}_q \Big\{ \langle q, Up_t\rangle + w(q) \Big\}, \quad p_{t+1} = \operatorname*{argmin}_p \Big\{ -\langle U^\star q_t, p\rangle + w(p) + \eta^{-1} D_w(p\|p_t) \Big\}$$

Finally, by the change of variable $(p, q) \mapsto (\mu, \nu)$ where $\mu(dy) = p(y) d\mu_{\mathsf{ref}}(dy)$ and $\nu(dx) = q(x)\nu_{\mathsf{ref}}(dx)$, noting that $w(p) = D_{\mathsf{KL}}(\mu|\mu_{\mathsf{ref}})$ and $D_w(p|q) = D_{\mathsf{KL}}(\mu|\nu)$, we obtain Theorem 4.

C.2.1. PROOF OF LEMMA 15

Let $G(p) = \ell(p) + w(p)$ and define the surrogate objective

$$G_t(p) = \ell(p_t) + \langle \nabla \ell(p_t), p - p_t \rangle + \eta^{-1} D_w(p \| p_t) + w(p).$$

Then $p_{t+1} \in \operatorname{argmin}_{p \in X} G_t(p)$. By the definition of Bregman divergence,

$$\begin{aligned} G(p) &= \ell(p) + w(p) \\ &= \ell(p_t) + \langle \nabla \ell(p_t), p - p_t \rangle + D_\ell(p \| p_t) + w(p) \quad \text{by } D_\ell(p \| p_t) = \ell(p) - \ell(p_t) - \langle \nabla \ell(p_t), p - p_t \rangle \\ &= G_t(p) + D_\ell(p \| p_t) - \eta^{-1} D_w(p \| p_t). \end{aligned}$$

Using the assumption $D_\ell(p \| p_t) \le L D_w(p \| p_t)$ and the non-negativity of the Bregman divergence $D_\ell(p \| p_t) \ge 0$, the approximation error $G(p) - G_t(p)$ can be controlled as

$$-\eta^{-1} D_w(p \| p_t) \le G(p) - G_t(p) \le -(\eta^{-1} - L) D_w(p \| p_t). \tag{11}$$

Now, since $p_{t+1} \in \operatorname{argmin} G_t(p)$ and G_t is convex and differentiable, we have

$$\langle \nabla G_t(p_{t+1}), q - p_{t+1} \rangle \ge 0 \qquad \forall q \in X.$$

Equivalently, by the definition of Bregman divergence,

$$D_{G_t}(q \| p_{t+1}) \le G_t(q) - G_t(p_{t+1}), \qquad \forall q \in X.$$

Applying this with $q = p_*$, we get

$$\begin{aligned} (\eta^{-1} + 1) D_w(p_* \| p_{t+1}) &= D_{G_t}(p_* \| p_{t+1}) \qquad \text{since } G_t(p) = (\eta^{-1} + 1) w(p) + \text{linear terms} \\ &\le G_t(p_*) - G_t(p_{t+1}) \\ &\le \Big(F(p_*) + \eta^{-1} D_w(p_* \| p_t) \Big) - \Big(F(p_{t+1}) + (\eta^{-1} - L) D_w(p_{t+1} \| p_t) \Big) \qquad \text{by (11)} \\ &\le \eta^{-1} D_w(p_* \| p_t) - (\eta^{-1} - L) D_w(p_{t+1} \| p_t) \qquad \text{since } F(p_*) \le F(p_{t+1}) \\ &\le \eta^{-1} D_w(p_* \| p_t), \end{aligned}$$

where the last step uses $\eta^{-1} \ge L$ and $D_w(\cdot|\cdot) \ge 0$. Dividing both sides by $\eta^{-1} + 1$ and iterating completes the proof.

C.2.2. PROOF OF LEMMA 16

Fix $g, f \in L^\infty$ and let $U^{g \oplus f} = U - g \oplus f$. Then, we can rewrite $\ell(p)$ as

$$\ell(p) = -\Big(\min_{q \in \Delta^\infty} \langle q, U^{g \oplus f} p \rangle + \langle q, g \rangle + \langle p, f \rangle + w(q) \Big).$$

By the envelope theorem, ℓ is differentiable with its derivative given by

$$\nabla \ell(p) = -(U^{g \oplus f})^\star q_p + f, \tag{12}$$

Here, q_p is the solution of $\min_q \langle q, U^{g\oplus f} p\rangle + \langle q, g\rangle + \langle p, f\rangle + w(q)$, which can be written explicitly as

$$q_p(x) = \frac{\exp(-U^{g\oplus f} p(x)) e^{-g(x)}}{\int \mu_{\mathsf{ref}}(dx') \exp(-U^{g\oplus f} p(x')) e^{-g(x')}}. \tag{13}$$

Then, for any $p, p' \in \Delta^\infty$, using $\|Uf\|_\infty \le \|U\|_\infty \|f\|_1$ by the Cauchy–Schwarz inequality and Lemma 13 with $\mu_{-g}(dx) \propto e^{-g(x)} \mu_{\mathsf{ref}}(dx)$, noting that μ_{-g} and μ_{ref} are absolutely continuous with respect to each other,

$$\begin{aligned}
\|q_p - q_{p'}\|_1 &\le \| - U^{g\oplus f} p + U^{g\oplus f} p'\|_{L^\infty(\mu_{-g})} && \text{by (13) and Lemma 13 with } \mu_{-g}(dx) \propto e^{-g(x)}\mu_{\mathsf{ref}}(dx) \\
&= \| - U^{g\oplus f} p + U^{g\oplus f} p'\|_\infty && \|\cdot\|_{L^\infty(\mu_{-g})} = \|\cdot\|_{L^\infty(\mu_{\mathsf{ref}})} \\
&\le \|U^{g\oplus f}\|_\infty \|p - p'\|_1 && \text{by Hölder's inequality.}
\end{aligned}$$

Since $g, f \in L^\infty$ are arbitrary, taking $\inf_{g,f\in L^\infty}$ on the RHS, we complete the proof of $\|q_p - q_{p'}\|_1 \le \|U\|_\oplus \|p - p'\|_1$.

Now, combined with the derivative form (12),

$$\begin{aligned}
\|\nabla \ell(p) - \nabla \ell(p')\|_\infty &= \| - (U^{g\oplus f})^\star q_p + (U^{g\oplus f})^\star q_{p'}\|_\infty && \text{by (12)} \\
&\le \|(U^{g\oplus f})^\star\|_\infty \|q_p - q_{p'}\|_1 && \text{by Hölder's inequality} \\
&\le \|U^{g\oplus f}\|_\infty \|U\|_\oplus \|p - p'\|_1
\end{aligned}$$

Again, since $g, f \in L^\infty$ are arbitrary, taking $\inf_{g,f\in L^\infty}$ on the RHS, we get

$$\forall p, p' \in \Delta^\infty, \quad \|\nabla \ell(p) - \nabla \ell(p')\|_\infty \le \|U\|_\oplus^2 \|p - p'\|_1. \tag{14}$$

Therefore, by the fundamental theorem of calculus, the Bregman divergence $D_\ell(p\|q) = \ell(p) - \ell(q) - \langle \nabla \ell(q), p - q\rangle$ is bounded from above as

$$\begin{aligned}
D_\ell(p\|q) &= \int_0^1 \langle \nabla \ell(tp + (1-t)q) - \nabla \ell(q), p - q\rangle dt \\
&\le \int_0^1 \|U\|_\oplus^2 \|t(p-q)\|_1 \|p - q\|_1 dt && \text{by Hölder's inequality and (14)} \\
&= \frac{1}{2} \|U\|_\oplus^2 \|p - q\|_1^2.
\end{aligned}$$

Finally, by Pinsker's inequality $\mathsf{TV} \le \sqrt{D_{\mathsf{KL}}/2}$, which is translated to $2^{-1}\|p - q\|_1 \le \sqrt{D_w(p\|q)/2}$ in our setup, the last term is bounded from above by $\|U\|_\oplus^2 D_w(p\|q)$. Thus, we obtain $D_\ell(p\|q) \le \|U\|_\oplus^2 D_w(p\|q)$ and complete the proof.

C.3. Proof of Theorem 5

Let $p_t = \frac{d\mu_{\mathsf{NL}}^t}{d\mu_{\mathsf{ref}}}$, $q_t = \frac{d\nu_{\mathsf{NL}}^t}{d\mu_{\mathsf{ref}}}$, $p_{\mathsf{NL}} = \frac{d\mu_{\mathsf{NL}}}{d\mu_{\mathsf{ref}}}$, and $q_{\mathsf{NL}} = \frac{d\nu_{\mathsf{NL}}}{d\mu_{\mathsf{ref}}}$ so that

$$\begin{aligned}
q_t &= \operatorname*{argmin}_{q\in\Delta^\infty} \langle q, Up_t\rangle + w(q), & q_{\mathsf{NL}} &= \operatorname*{argmin}_{q\in\Delta^\infty} \langle q, Up_{\mathsf{NL}}\rangle + w(q), \\
p_{t+1} &= \operatorname*{argmax}_{p\in\Delta^\infty} \langle q_t, Up\rangle - w(p), & p_{\mathsf{NL}} &= \operatorname*{argmax}_{p\in\Delta^\infty} \langle q_{\mathsf{NL}}, Up\rangle - w(p)
\end{aligned}$$

By Lemma 16-(1) and the invariance $\|U^\star\|_\oplus = \|U\|_\oplus$, we have $\|q_t - q_{\mathsf{NL}}\|_1 \le \|U\|_\oplus \|p_t - p_{\mathsf{NL}}\|_1$ and $\|p_{t+1} - p_{\mathsf{NL}}\|_1 \le \|U\|_\oplus \|q_t - q_{\mathsf{NL}}\|_1$. Hence, the proof is complete.

Appendix D. Fréchet Differentiability and Stability

This section provides the technical ingredients needed for the analysis of the alignment dynamics, in particular Theorems 7 and 8. Our main goal is to establish the uniform Fréchet differentiability of MCHF and NLHF, together with an explicit derivative formula. To this end, we organize this section as follows:

- Section D.1: Fréchet differentiability and derivative formulas for MCHF and NLHF.
- Section D.2: Stability estimate
- Section D.3: Uniform Fréchet differentiability

Readers primarily interested in the consequences for the alignment dynamics may read the statements of Theorems 17 and 19, skip the other theorems and technical proofs, and proceed directly to Section E.

D.1. Fréchet differentiability

D.1.1. FRÉCHET DIFFERENTIABILITY OF MCHF

For each $U \in L^\infty(\mu_{\mathsf{ref}} \otimes \mu_{\mathsf{ref}})$, define $P_U \in L^\infty(\mu_{\mathsf{ref}} \otimes \mu_{\mathsf{ref}})$ as

$$P_U(x,y) = \frac{\exp(U(x,y))}{\int \mu_{\mathsf{ref}}(dy') \exp(U(x,y'))}$$

Let $P_U^\star : L^1(\mu_{\mathsf{ref}}) \to L^1(\mu_{\mathsf{ref}})$ be the adjoint operator of P_U, that is, $P_U^\star f(y) = \int \mu_{\mathsf{ref}}(dx) f(x) P_U(x,y)$. Let $\mu_{\mathsf{MC}}(U)$ be the unique stationary distribution of the Markov kernel $\mathsf{P}_U(x,dy) = P_U(x,y)\mu_{\mathsf{ref}}(dy)$ (the existence and uniqueness follow from Corollary 3) and let $p_{\mathsf{MC}}(U) = \frac{d\mu_{\mathsf{MC}}(U)}{d\mu_{\mathsf{ref}}}$ be its density. Then, for each $U \in L^\infty$, $p = p_{\mathsf{MC}}(U)$ is the unique solution to the system:

$$p \in L^1, \qquad P_U^\star p = p, \qquad \int \mu_{\mathsf{ref}}(dy) p(y) = 1.$$

Now we claim that $U \mapsto p_{\mathsf{MC}}(U)$ is Fréchet differentiable.

Theorem 17 (Fréchet differentiability of MCHF) *The map $U \mapsto p_{\mathsf{MC}}(U)$ is continuously Fréchet differentiable as a map from $L^\infty(\mu_{\mathsf{ref}} \otimes \mu_{\mathsf{ref}})$ to L^1. More precisely, for each $U \in L^\infty(\mu_{\mathsf{ref}} \otimes \mu_{\mathsf{ref}})$, for any sequence of $E \in L^\infty(\mu_{\mathsf{ref}} \otimes \mu_{\mathsf{ref}})$ such that $\|E\|_\infty \to 0$,*

$$\lim_{\|E\|_\infty \to 0} \frac{\|p_{\mathsf{MC}}(U+E) - p_{\mathsf{MC}}(U) - Dp_{\mathsf{MC}}(U)[E]\|_1}{\|E\|_\infty} = 0,$$

where the Fréchet derivative $Dp_{\mathsf{MC}}(U)[E] \in L_0^1$ is given by

$$\begin{aligned} &Dp_{\mathsf{MC}}(U)[E] \\ &= \left((I - P_U^\star)\,|_{L_0^1}\right)^{-1} \Big(y \mapsto \int \mu_{\mathsf{ref}}(dx) p_{\mathsf{MC}}(U)(x) P_U(x,y) \big[E(x,y) - \int \mu_{\mathsf{ref}}(dy') P_U(x,y') E(x,y')\big] \Big), \end{aligned}$$

where $L_0^1 \equiv \{f \in L^1(\mu_{\mathsf{ref}}) : \int \mu_{\mathsf{ref}}(dy) f(y) = 0\}$.

Proof Denote $p(U) = p_{\mathsf{MC}}(U)$, and define $f(U) = p(U) - \mathbf{1}$. Then $f(U) \in L^1_0$ and $f = f(U)$ is the unique solution to the system

$$f \in L^1_0, \quad F(f, U) = 0$$

where $F : L^1_0 \times L^\infty(\mu_{\mathsf{ref}} \otimes \mu_{\mathsf{ref}}) \to L^1_0$ is defined by

$$F(f, U) = (I - P^\star_U)(f + \mathbf{1}).$$

Here, the image of F is included in L^1_0 since $\int \mu_{\mathsf{ref}}(dy) P_U(x, y) = 1$ for every x.

We first show that F is continuously Fréchet differentiable. The key point is that the softmax map $U \mapsto P_U$ is continuously Fréchet differentiable from $L^\infty(\mu_{\mathsf{ref}} \otimes \mu_{\mathsf{ref}})$ to $L^\infty(\mu_{\mathsf{ref}} \otimes \mu_{\mathsf{ref}})$. Its derivative at U in the direction $E \in L^\infty(\mu_{\mathsf{ref}} \otimes \mu_{\mathsf{ref}})$ is

$$DP_U[E](x, y) = P_U(x, y)\Big[E(x, y) - \int \mu_{\mathsf{ref}}(dy') P_U(x, y') E(x, y')\Big]. \tag{15}$$

Indeed, this follows by differentiating the numerator and denominator in the normalized exponential map. Consequently, F is continuously Fréchet differentiable with derivatives given by

$$\begin{aligned} \forall h \in L^1_0, \quad D_f F(f, U)[h] &= (I - P^\star_U) h \\ \forall E \in L^\infty(\mu_{\mathsf{ref}} \otimes \mu_{\mathsf{ref}}), \quad D_U F(f, U)[E] &= -(DP_U[E])^\star (f + \mathbf{1}). \end{aligned} \tag{16}$$

Next we show that $D_f F(f(U), U)$ has a bounded inverse on L^1_0. By Lemma 11,

$$\forall h \in L^1_0, \quad \|P^\star_U h\|_1 \le (1 - e^{-2\|U\|_\oplus}) \cdot \|h\|_1$$

Since $(1 - e^{-2\|U\|_\oplus}) \in [0, 1)$, $(I - P^\star_U) \mid_{L^1_0}$ has a bounded inverse given by the Neumann series $((I - P^\star_U) \mid_{L^1_0})^{-1} = \sum_{t=0}^\infty (P^\star_U)^t$, with its operator norm bounded as

$$\sup_{h \in L^1_0 : \|h\|_1 \le 1} \Big\| \big((I - P^\star_U) \mid_{L^1_0}\big)^{-1} h \Big\|_1 \le \sum_{t=0}^\infty (1 - e^{-2\|U\|_\oplus})^t = e^{2\|U\|_\oplus} < +\infty. \tag{17}$$

Therefore, by the implicit function theorem for the Banach space (cf. [12, Theorem 3.4.10]), there exists a neighborhood of U in $L^\infty(\mu_{\mathsf{ref}} \otimes \mu_{\mathsf{ref}})$ such that $U \mapsto f(U)$ is continuously Fréchet differentiable on the neighborhood. Since $f(U) = p(U) - 1$, $U \mapsto p(U)$ is also continuously Fréchet differentiable at U.

It remains to compute the derivative. Differentiating the identity $F(f(U), U) = 0$ in the direction E gives

$$Df(U)[E] = -\Big(D_f F(f(U), U) \mid_{L^1_0}\Big)^{-1} D_U F(f(U), U)[E].$$

Substituting (16) and (15) to the above equation, noting $f(U) + \mathbf{1} = p(U)$, we have

$$\begin{aligned} &Df(U)[E] \\ &= \Big((I - P^\star_U) \mid_{L^1_0}\Big)^{-1} \Big(y \mapsto \int \mu_{\mathsf{ref}}(dx) p(U)(x) P_U(x, y) \Big[E(x, y) - \int \mu_{\mathsf{ref}}(dy') P_U(x, y') E(x, y')\Big]\Big). \end{aligned}$$

Finally, since $p(U) = f(U) + \mathbf{1}$, we have $Dp(U)[E] = Df(U)[E]$. This proves the claimed formula for the Fréchet derivative of $p_{\mathsf{MC}}(U)$. ■

D.1.2. FRÉCHET DIFFERENTIABILITY OF NLHF

Let $\Delta^\infty \subset L^\infty$ be the density simplex defined as

$$\Delta^\infty = \{p \in L^\infty : p(\cdot) \geq 0, \int \mu_{\mathsf{ref}}(dz)p(z) = 1\}.$$

Notice that for any $p \in \Delta^\infty$, it induces a tilted probability measure $\mu(dy) = p(y)\mu_{\mathsf{ref}}(dy)$. We define the operator s taking values in Δ^∞ as

$$\mathsf{s} : L^\infty \to \Delta^\infty, \quad f \mapsto \mathsf{s}(f)(x) = \frac{\exp(f(x))}{\int \mu_{\mathsf{ref}}(dx')\exp(f(x'))}$$

For each $p \in \Delta^\infty$, define the linear operator $J_p : L^2 \to L^2_0$ as $J_p = \mathrm{Diag}(p) - pp^\star$ or more precisely

$$J_p : L^2 \to L^2_0, \qquad f \mapsto J_p f(z) = p(z)f(z) - p(z)\int \mu_{\mathsf{ref}}(dz')p(z')f(z')$$

where $L^2_0 = \{h \in L^2 : \int \mu_{\mathsf{ref}}(dz)h(z) = 0\} \subset L^2$

Lemma 18 *For any $p \in \Delta^\infty$, J_p is bounded, self-adjoint, and positive semidefinite operator on L^2 such that*

$$\|J_p\|_{2\to 2} = \lim_{\|h\|_2 \leq 2} \|J_p h\|_2 \leq \|p\|_\infty$$

Proof For any $f, g \in L^2$, letting $\mu_p(dz) = p(z)\mu_{\mathsf{ref}}(dz)$ be the tilted probability measure by the density $p \in \Delta^\infty$, one can show $\langle f, J_p g\rangle = \mathrm{Cov}_{Z\sim\mu_p}(f(Z), g(Z))$. This implies J_p is self-adjoint and positive semidefinite operator. Moreover, the operator norm $\|J_p\|_{L^2\to L^2}$ is bounded by $\|p\|_\infty$ because the Cauchy–Schwarz inequality yields

$$|\langle f, J_p g\rangle| \leq \sqrt{\mathrm{Var}_{Z\sim\mu_p} f(Z)} \cdot \sqrt{\mathrm{Var}_{Z\sim\mu_p} g(Z)} \leq \|f\|_{L^2(\mu_p)} \cdot \|g\|_{L^2(\mu_p)} \leq \|p\|_\infty \cdot \|f\|_{L^2(\mu_{\mathsf{ref}})} \cdot \|g\|_{L^2(\mu_{\mathsf{ref}})}$$

where the last inequality follows from $\|f\|_{L^2(\mu_p)} \leq \|\frac{d\mu_p}{dp_{\mathsf{ref}}}\|^{1/2}_{L^\infty(\mu_{\mathsf{ref}})} \cdot \|f\|_{L^2(\mu_{\mathsf{ref}})} = \|p\|^{1/2}_\infty \|f\|_2$. ■

Therefore, by the spectral theorem, there exists a bounded self-adjoint positive semidefinite operator T such that $T \cdot T = J_p$. We denote such T by $J_p^{1/2}$.

Theorem 19 (Fréchet differentiability of NLHF) *Let $U \in L^\infty(\mu_{\mathsf{ref}} \otimes \mu_{\mathsf{ref}})$, and let $(p_{\mathsf{NL}}(U), q_{\mathsf{NL}}(U))$ be the density pair, which is the unique solution to the NLHF fixed-point equations*

$$p, q \in \Delta^\infty, \quad p = \mathsf{s}(U^\star q), \qquad q = \mathsf{s}(-Up).$$

Then the map $U \mapsto p_{\mathsf{NL}}(U)$ is continuously Fréchet differentiable as a map from $L^\infty(\mu_{\mathsf{ref}} \otimes \mu_{\mathsf{ref}})$ to $L^2(\mu_{\mathsf{ref}})$. More precisely, for each $U \in L^\infty(\mu_{\mathsf{ref}} \otimes \mu_{\mathsf{ref}})$, and for any sequence of $E \in L^\infty(\mu_{\mathsf{ref}} \otimes \mu_{\mathsf{ref}})$ such that $\|E\|_\infty \to 0$, we have

$$\lim_{\|E\|_\infty \to 0} \frac{\|p_{\mathsf{NL}}(U+E) - p_{\mathsf{NL}}(U) - Dp_{\mathsf{NL}}(U)[E]\|_2}{\|E\|_\infty} = 0$$

where the Fréchet derivative $Dp_{\mathsf{NL}}(U)[E]$ belongs to L^2_0 and is given by

$$Dp_{\mathsf{NL}}(U)[E] = \big(I + J_{p(U)}U^\star J_{q(U)}U\big)^{-1}\big(J_{p(U)}E^\star q(U) - J_{p(U)}U^\star J_{q(U)}Ep(U)\big),$$

where $(p(U), q(U)) = (p_{\mathsf{NL}}(U), q_{\mathsf{NL}}(U))$ on the RHS. Here $(I + J_{p(U)}U^\star J_{q(U)}U)$ has a bounded inverse on L^2.

Proof Let $x(U) = p(U) - \mathbf{1}$ and $y(U) = q(U) - \mathbf{1}$. Then $x(U), y(U) \in L_0^2$. Define the Implicit function $F : L_0^2 \times L_0^2 \times L^\infty(\mu_{\mathsf{ref}} \otimes \mu_{\mathsf{ref}}) \to L_0^2 \times L_0^2$ by

$$F(x, y, U) = \begin{pmatrix} F_1(x, y, U) \\ F_2(x, y, U) \end{pmatrix} \equiv \begin{pmatrix} (x + \mathbf{1}) - \mathsf{s}(U^\star(y + \mathbf{1})) \\ (y + \mathbf{1}) - \mathsf{s}(-U(x + \mathbf{1})) \end{pmatrix}$$

so that $F(x(U), y(U), U) = 0$. Moreover, F indeed takes values in $L_0^2 \times L_0^2$, since s maps into the density simplex Δ^∞.

We first verify that F is continuously Fréchet differentiable. Note that the map $\mathsf{s} : L^\infty(\mu_{\mathsf{ref}}) \to L^\infty(\mu_{\mathsf{ref}})$ is continuously Fréchet differentiable, with derivative $D\mathsf{s}(f)[h] = J_{\mathsf{s}(f)} h$. Furthermore, for $U \in L^\infty(\mu_{\mathsf{ref}} \otimes \mu_{\mathsf{ref}})$ and $v \in L^2$, $\|Uv\|_\infty \le \|U\|_\infty \|v\|_1 \le \|U\|_\infty \|v\|_2$, and similarly $\|U^\star v\|_\infty \le \|U\|_\infty \|v\|_2$. Thus, the two bilinear maps, $(U, v) \mapsto Uv$ and $(U, v) \mapsto U^\star v$, are bounded from $L^\infty(\mu_{\mathsf{ref}} \otimes \mu_{\mathsf{ref}}) \times L^2(\mu_{\mathsf{ref}})$ to $L^\infty(\mu_{\mathsf{ref}})$. Therefore, by the chain rule, F is continuously Fréchet differentiable.

Let us compute its derivatives. For perturbations $(\delta x, \delta y) \in L_0^2 \times L_0^2$, we have

$$\begin{aligned} D_{x,y} F_1(x, y, U)[\delta x, \delta y] &= \delta x - J_{\mathsf{s}(U^\star(y+\mathbf{1}))} U^\star \delta y, \\ D_{x,y} F_2(x, y, U)[\delta x, \delta y] &= \delta y + J_{\mathsf{s}(-U(x+\mathbf{1}))} U \delta x. \end{aligned}$$

In particular, at the solution $(x(U), y(U))$ to $F(x, y, U) = 0$, we obtain

$$D_{x,y} F(x(U), y(U), U) = \begin{bmatrix} I & -J_p U^\star \\ J_q U & I \end{bmatrix} \tag{18}$$

as a bounded operator on $L_0^2 \times L_0^2$, where we denote $J_p = J_{p(U)}$ and $J_q = J_{q(U)}$ for simplicity.

Next, for a perturbation $E \in L^\infty(\mu_{\mathsf{ref}} \otimes \mu_{\mathsf{ref}})$, the Fréchet derivative of F with respect to U is

$$\begin{aligned} D_U F_1(x, y, U)[E] &= -J_{\mathsf{s}(U^\star(y+\mathbf{1}))} E^\star (y + \mathbf{1}), \\ D_U F_2(x, y, U)[E] &= J_{\mathsf{s}(-U(x+\mathbf{1}))} E (x + \mathbf{1}). \end{aligned}$$

Therefore, at the solution,

$$D_U F(x(U), y(U), U)[E] = \begin{pmatrix} -J_p E^\star q(U) \\ J_q E p(U) \end{pmatrix} \tag{19}$$

where $J_p = J_{p(U)}$ and $J_q = J_{q(U)}$.

We now show that $D_{x,y} F(x(U), y(U), U)$ given by (18) is a bounded isomorphism on $L_0^2 \times L_0^2$. For any $(r_1, r_2) \in L_0^2 \times L_0^2$, solving

$$\begin{bmatrix} I & -J_p U^\star \\ J_q U & I \end{bmatrix} \begin{pmatrix} \delta x \\ \delta y \end{pmatrix} = \begin{pmatrix} r_1 \\ r_2 \end{pmatrix},$$

eliminating δy first, we have

$$\delta y = r_2 - J_q U \delta x.$$

Substituting into the first equation yields

$$\big(I + J_p U^\star J_q U\big) \delta x = r_1 + J_p U^\star r_2.$$

Since J_qU and $J_pU^\star$ are both bounded linear operators on L^2, it is enough to show that $I+J_pU^\star J_qU$ has a bounded inverse on L_0^2. On L_0^2, write

$$I + J_pU^\star J_qU = I + AB, \qquad A = J_p^{1/2}, \qquad B = J_p^{1/2}U^\star J_qU.$$

Then

$$BA = J_p^{1/2}U^\star J_qU J_p^{1/2}.$$

This operator is self-adjoint and positive semidefinite on L^2 since J_q is positive semidefinite. Thus, by the Lax–Milgram theorem, $I + BA$ has a bounded inverse on L^2, with

$$\sup_{h\in L^2:\|h\|_2=1} \|(I + BA)^{-1}h\|_2 \le 1.$$

Thus, $I + J_pU^\star J_qU = I + AB$ is also invertible on $L_0^2 \subset L^2$, with inverse $(I + AB)^{-1} = I - A(I + BA)^{-1}B$. Hence, $D_{x,y}F(x(U), y(U), U)$ has a bounded inverse on $L_0^2 \times L_0^2$.

Therefore, by the Implicit Function Theorem on Banach spaces, there exists a neighborhood of U in $L^\infty(\mu_{\mathsf{ref}} \otimes \mu_{\mathsf{ref}})$ such that $U \mapsto (x(U), y(U))$ is continuously Fréchet differentiable on the neighborhood. By $x(U) = p(U) - \mathbf{1}$ and $y(U) = q(U) - \mathbf{1}$, the map $U \mapsto (p(U), q(U))$ is also continuously Fréchet differentiable at U.

It remains to compute the derivative. Differentiating $F(x(U), y(U), U) = 0$ in the direction E gives

$$D_{x,y}F(x(U), y(U), U)\begin{pmatrix} Dx(U)[E] \\ Dy(U)[E] \end{pmatrix} + D_UF(x(U), y(U), U)[E] = 0.$$

Since $p(U) = x(U) + \mathbf{1}$ and $q(U) = y(U) + \mathbf{1}$, we have $Dp(U)[E] = Dx(U)[E]$ and $Dq(U)[E] = Dy(U)[E]$. Substituting (19) and (18), we get

$$\begin{bmatrix} I & -J_pU^\star \\ J_qU & I \end{bmatrix}\begin{pmatrix} Dp(U)[E] \\ Dq(U)[E] \end{pmatrix} = \begin{pmatrix} J_pE^\star q \\ -J_qEp \end{pmatrix}$$

where $J_p = J_{p(U)}$ and $J_q = J_{q(U)}$. Eliminating $Dq(U)[E]$ gives

$$Dp(U)[E] = \big(I + J_pU^\star J_qU\big)^{-1}\big(J_pE^\star q - J_pU^\star J_qEp\big).$$

This completes the proof. ■

D.2. Stability under perturbations of utility

D.2.1. STABILITY OF MCHF

Let $c(\|U\|_\oplus) = \min(1 - e^{-2\|U\|_\oplus}, \|U\|_\oplus) < 1$ be the contraction rate of the map $\mu \mapsto \mu\mathsf{P}$ given by Theorem 2. The next theorem shows how the perturbation error of U propagates along the Markov chain.

Theorem 20 *For any $\hat{U}, U \in L^\infty(\mu_{\mathsf{ref}} \otimes \mu_{\mathsf{ref}})$, let P and $\hat{\mathsf{P}}$ be the Markov kernels (4) constructed from U and $\hat{U}$, respectively. Then,*

$$\forall t \ge 1, \quad d_{\mathsf{TV}}(\mu_{\mathsf{ref}}\mathsf{P}^t, \mu_{\mathsf{ref}}\hat{\mathsf{P}}^t) \le \frac{1}{2}\|U - \hat{U}\|_\infty \cdot \frac{1 - \big(c(\|U\|_\oplus) \wedge c(\|\hat{U}\|_\oplus)\big)^t}{1 - c(\|U\|_\oplus) \wedge c(\|\hat{U}\|_\oplus)}$$

where $c(\|U\|_\oplus) = \min(1 - e^{-2\|U\|_\oplus}, \|U\|_\oplus) < 1$.

Proof Let c and $\hat{c}$ be the contraction coefficient given by Theorem 2 for $\mu \mapsto \mathsf{P}\mu$ and $\mu \mapsto \hat{\mathsf{P}}\mu$, respectively. Consider the two iterations, $\mu_t = \mu_{t-1}\mathsf{P}$ and $\hat{\mu}_t = \hat{\mu}_{t-1}\hat{\mathsf{P}}$, initialized with $\mu_0 = \hat{\mu}_0 \in \mathscr{P}_{\mu_{\mathsf{ref}}}$. Using the triangle inequality, we have

$$d_{\mathsf{TV}}(\mu_t, \hat{\mu}_t) \le d_{\mathsf{TV}}(\mu_{t-1}\mathsf{P}, \hat{\mu}_{t-1}\mathsf{P}) + d_{\mathsf{TV}}(\hat{\mu}_{t-1}\mathsf{P}, \hat{\mu}_{t-1}\hat{\mathsf{P}}) \le c \cdot d_{\mathsf{TV}}(\mu_{t-1}, \hat{\mu}_{t-1}) + d_{\mathsf{TV}}(\hat{\mu}_{t-1}\mathsf{P}, \hat{\mu}_{t-1}\hat{\mathsf{P}}).$$

Let us bound the second term. By the definition of TV distance, using Fubini's lemma,

$$\begin{aligned}
d_{\mathsf{TV}}(\hat{\mu}_{t-1}\mathsf{P}, \hat{\mu}_{t-1}\hat{\mathsf{P}}) &= \frac{1}{2}\int \mu_{\mathsf{ref}}(dy)\Big|\int \hat{\mu}_{t-1}(dx)\big(P(x,y) - \hat{P}(x,y)\big)\Big| \\
&\le \frac{1}{2}\int \mu_{\mathsf{ref}}(dy)\hat{\mu}_{t-1}(dx)|P(x,y) - \hat{P}(x,y)| \\
&= \frac{1}{2}\int \hat{\mu}_{t-1}(dx)\|P(x,\cdot) - \hat{P}(x,\cdot)\|_1 && \text{Fubini's lemma} \\
&\le \frac{1}{2}\int \hat{\mu}_{t-1}(dx)\big\|U(x,\cdot) - \hat{U}(x,\cdot)\big\|_\infty && \text{by Lemma 13} \\
&\le \frac{1}{2}\|U - \hat{U}\|_\infty
\end{aligned}$$

where we used Lemma 13 with $\mu = \mu_{\mathsf{ref}}, g(\cdot) = U(x,\cdot)$ and $\tilde{g}(\cdot) = \hat{U}(x,\cdot)$. The last inequality follows from the fact that $\hat{\mu}_{t-1}$ is absolutely continuous with respect to μ_{ref}.

Thus, we obtain

$$d_{\mathsf{TV}}(\mu_t, \hat{\mu}_t) \le c \cdot d_{\mathsf{TV}}(\mu_{t-1}, \hat{\mu}_{t-1}) + \frac{1}{2}\|U - \hat{U}\|_\infty.$$

Iterating this inequality and using the initial condition $\mu_0 = \hat{\mu}_0$, we get $d_{\mathsf{TV}}(\mu_t, \hat{\mu}_t) \le \frac{1}{2}\|U - \hat{U}\|_\infty \sum_{s=0}^{t-1} c^s$. By symmetry, the same upper bound also holds with c replaced by $\hat{c}$. Thus, we complete the proof. ■

Now, letting $t \to \infty$ in Theorem 20, noting that $\mu_{\mathsf{ref}}\mathsf{P}^t \to \mu_{\mathsf{MC}}$ as $t \to +\infty$ in TV distance (see Corollary 3), we obtain the following corollary.

Corollary 21 (Stability of MCHF equilibrium) *For any $U, \hat{U} \in L^\infty(\mu_{\mathsf{ref}} \otimes \mu_{\mathsf{ref}})$, letting $\mu_{\mathsf{MC}}(U)$ and $\mu_{\mathsf{MC}}(\hat{U})$ denote the stationary distributions of P and $\hat{\mathsf{P}}$, respectively,*

$$d_{\mathsf{TV}}(\mu_{\mathsf{MC}}(U), \mu_{\mathsf{MC}}(\hat{U})) \le \frac{1}{2} \cdot \frac{\|U - \hat{U}\|_\infty}{1 - c(\|U\|_\oplus) \wedge c(\|\hat{U}\|_\oplus)}$$

where $c(\|U\|_\oplus) = \min(1 - e^{-2\|U\|_\oplus}, \|U\|_\oplus) < 1$.

Since $c(\|U\|_\oplus)$ is always strictly less than 1, we conclude that the map $U \mapsto \mu_{\mathsf{MC}}(U)$ is locally Lipschitz over $L^\infty(\mu_{\mathsf{ref}} \otimes \mu_{\mathsf{ref}})$, and the local Lipschitz constant is controlled by the seminorm $\|U\|_\oplus$.

D.2.2. STABILITY OF NLHF

Next, we claim the stability of the NLHF solution μ_{NL}.

Theorem 22 (Stability of NLHF equilibrium) *There exists an absolute constant C such that, for any $U, \hat{U} \in L^\infty(\mu_{\mathsf{ref}} \otimes \mu_{\mathsf{ref}})$,*

$$d_{\mathsf{TV}}\big(\mu_{\mathsf{NL}}(U), \mu_{\mathsf{NL}}(\hat{U})\big) \le C\big(1 + \|U\|_\oplus^6 + \|\hat{U}\|_\oplus^6\big)\|U - \hat{U}\|_\infty.$$

Proof The proof of this theorem requires a different technique from that of Theorem 8 since the fixed point equation of the NLHF is not always contractive under the TV distance.

Fix $U, \hat{U} \in L^\infty(\mu_{\mathsf{ref}} \otimes \mu_{\mathsf{ref}})$. For each $t \in [0,1]$, let U_t be the linear interpolation

$$U_t = t\hat{U} + (1-t)U = U + tE \quad \text{where} \quad E = \hat{U} - U$$

and let (μ_t, ν_t) be the solution to the NLHF based on the pairwise utility U_t, and let $p_t = \frac{d\mu_t}{d\mu_{\mathsf{ref}}}$ and $q_t = \frac{d\nu_t}{d\mu_{\mathsf{ref}}}$ be the density for each $t \in [0,1]$, so that our goal is to bound $d_{\mathsf{TV}}(\mu_1, \mu_0) = \frac{1}{2}\|p_1 - p_0\|_1$.

By Theorem 19, $t \mapsto p_t$ is C^1 from $\mathbb{R}$ to L^2 in the sense of

$$\lim_{\epsilon \to 0} \|\epsilon^{-1}(p_{t+\epsilon} - p_t) - \dot{p}_t\|_2 = 0$$

where the derivative $\dot{p}_t$ is given by

$$\dot{p}_t = (I + J_{p_t}(U_t)^\star J_{q_t} U_t)^{-1}(J_{p_t} E^\star q_t - J_{p_t} U_t^\star J_{q_t} E p_t).$$

Here, since $\|\cdot\|_1 \le \|\cdot\|_2$, p_t is in particular C^1 from $\mathbb{R}$ to L^1. Thus, applying the fundamental theorem of calculus on the Banach space L^1 to $d_{\mathsf{TV}}(\mu_1, \mu_0) = \frac{1}{2}\|p_1 - p_0\|_1$, using Fubini's lemma,

$$d_{\mathsf{TV}}(\mu_1, \mu_0) = \frac{1}{2}\int \mu_{\mathsf{ref}}(dy)\Big|\int_0^1 \dot{p}_t(y)dt\Big| \le \frac{1}{2}\int \mu_{\mathsf{ref}}(dy)\int_0^1 |\dot{p}_t(y)|dt = \frac{1}{2}\int_0^1 dt\|\dot{p}_t\|_1.$$

Next, we claim that there exists an absolute constant C such that

$$\forall t \in [0,1], \quad \|\dot{p}_t\|_1 \le C \cdot (1 + \|U_t\|_\oplus^6) \cdot \|E\|_\infty. \tag{20}$$

If this claim holds, by the triangle inequality for the seminorm $\|U_t\|_\oplus = \|t\hat{U} + (1-t)U\|_\oplus$, one can show $\int_0^1 dt\|U_t\|_\oplus^6 \le C'(\|U\|_\oplus^6 + \|\hat{U}\|_\oplus^6)$ for an absolute constant C'. Consequently, we obtain

$$d_{\mathsf{TV}}(p_{\mathsf{NL}}(U), p_{\mathsf{NL}}(\hat{U})) \le \frac{1}{2}\int_0^1 C(1 + \|U_t\|_\oplus^6)\|E\|_\infty \le \frac{1}{2}CC' \cdot \big(1 + \|\hat{U}\|_\oplus^6 + \|U\|_\oplus^6\big) \cdot \|E\|_\infty,$$

and the proof of Theorem 22 is complete. Thus, the rest of the goal is to show (20). Below, we prove it using the following two lemmas.

Lemma 23 *For any $p \in \Delta^\infty$ and any $f \in L^\infty$, $\|J_p f\|_1 \le \|f\|_\infty$.*

Proof Letting $\mu(dz) = p(z)\mu_{\mathsf{ref}}(dz)$ be the tilted measure induced by density $p \in \Delta^\infty$, for any $f \in L^\infty$, we have $\|J_p f\|_1 = \mathbb{E}_\mu[|f - \mathbb{E}_\mu[f]|]$ where $\mathbb{E}_\mu[h] = \int \mu(dz)h(z)$. Using

$$\mathbb{E}_\mu[|f - \mathbb{E}_\mu[f]|] \le \sqrt{\mathrm{Var}_\mu(f)} \le \|f\|_{L^2(\mu)} \le \|f\|_{L^\infty(\mu)}$$

and $\|f\|_{L^\infty(\mu)} \le \|f\|_{L^\infty(\mu_{\mathsf{ref}})}$ since μ is absolutely continous with respect to μ_{ref}, we complete the proof. ■

Lemma 24 *There exists a universal constant C such that for any $U \in L^\infty(\mu_{\mathsf{ref}} \otimes \mu_{\mathsf{ref}})$, $g \in L^2$, and $p, q \in \Delta^\infty$,*

$$\|(I + J_p U^\star J_q U)^{-1} g\|_1 \le C(1 + \|U\|_\infty^5)\|g\|_1.$$

Proof The proof requires a careful analysis of the operator norm of $(I + J_p U^\star J_q U)^{-1}$ measured by L^1 norm (rather than L^2), which is of independent interest. See Section D.2.3. ■

Let us finish the proof of (20). Fix $f, g \in L^\infty$ and let $U_t^{g\oplus f} = U_t - g \oplus f$. For any $h \in L_0^2$, using $J_{q_t}\mathbf{1} = 0$ and $\langle \mathbf{1}, h\rangle = 0$, we have

$$J_{q_t} U_t h = J_{q_t}(U_t - g \oplus f)h + J_{q_t}\mathbf{1}\langle f, h\rangle + J_{q_t} g \langle \mathbf{1}, h\rangle = J_{q_t} U_t^{g\oplus f} h$$

and hence $J_{q_t} U_t$ operates on L_0^2 as

$$J_{q_t} U_t|_{L_0^2} = J_{q_t} U_t^{g\oplus f}.$$

By the same argument, we have

$$J_{p_t} U_t^\star|_{L_0^2} = J_{p_t}(U_t^{g\oplus f})^\star.$$

Since the image of J_{p_t} and J_{q_t} are both L_0^2, we can rewrite $\dot p_t$ as

$$\dot p_t = \big(I + J_{p_t}(U_t^{g\oplus f})^\star J_{q_t} U_t^{g\oplus f}\big)^{-1}\big(J_{p_t} E^\star q_t - J_{p_t}(U_t^{g\oplus f})^\star J_{q_t} E p_t\big).$$

Applying Lemma 23 with $\|p_t\|_1 = \|q_t\|_1 = 1$ and Hölder's inequality $\|Kh\|_\infty \le \|K\|_\infty \|h\|_1$, we get

$$\big\|J_{p_t} E^\star q_t - J_{p_t}(U_t^{g\oplus f})^\star J_{q_t} E p_t\big\|_1 \le \|E\|_\infty + \|U_t^{g\oplus f}\|_\infty \|E\|_\infty.$$

Applying Lemma 24 with $U_t = U_t^{g\oplus f}$ and $(p, q) = (p_t, q_t)$, combined with the above L^1-bound, we have

$$\|\dot p_t\|_1 \le C(1 + \|U_t^{g\oplus f}\|_\infty^5)(1 + \|U_t^{g\oplus f}\|_\infty)\|E\|_\infty \le C'(1 + \|U_t^{g\oplus f}\|_\infty^6)\|E\|_\infty$$

where C and C' are absolute constants. Recalling $U_t^{g\oplus f} = U_t - g \oplus f$ and $g, f \in L^\infty$ are arbitrary, taking $\inf_{g,f}$ on the RHS, with $\|U\|_\oplus = \inf_{g,f} \|U - g \oplus f\|_\infty$, we complete the proof of (20). ■

Remark 25 (Other stability bounds for NLHF) *Using the formulation of NLHF as the solution to a strongly convex optimization as in Section C.2, one can easily show the bound under KL:*

$$D_{\mathsf{KL}}\big(\mu_{\mathsf{NL}}(U) \| \mu_{\mathsf{NL}}(\hat U)\big) \le 2\|U - \hat U\|_\infty.$$

Combined with Pinsker's inequality, $d_{\mathsf{TV}} \le \sqrt{D_{\mathsf{KL}}/2}$, this yields global Hölder-$1/2$ continuity in TV.

On the other hand, using the fact that the Lipschitz constant of the map $p \mapsto \operatorname{argmin}_{q\in\Delta^\infty}\langle q, Up\rangle + w(q)$ is $\|U\|_\oplus$-Lipschitz under the $L^1(\mu_{\mathsf{ref}})$ metric (see Lemma 16-(1)), an argument similar to the proof of Theorem 20 yields

$$d_{\mathsf{TV}}\big(\mu_{\mathsf{NL}}(U), \mu_{\mathsf{NL}}(\hat U)\big) \le \frac{1}{2} \cdot \frac{\|U - \hat U\|_\infty}{1 - (\|U\|_\oplus \wedge \|\hat U\|_\oplus)}$$

whenever $\|U\|_\oplus \wedge \|\hat U\|_\oplus < 1$.

However, neither of these bounds is sufficient for the proof of uniform Fréchet differentiability for NLHF (Corollary 30), which requires Lipschitz continuity on every subset bounded in the seminorm $\|\cdot\|_\oplus$.

D.2.3. PROOF OF LEMMA 24

Let $f = (I + J_p U^\star J_q U)^{-1} g$. We aim to bound $\|f\|_1$ by $\|g\|_1$. If we define $h = J_q U f$, the triple (f, h, g) satisfies

$$f + J_p U^\star h = g, \qquad -J_q U f + h = 0.$$

Now, for any $p \in \Delta^\infty$, we rewrite the operator J_p as $J_p = D_{\sqrt{p}} \cdot D_{\sqrt{p}} - pp^\star$ where $D_{\sqrt{\nu}}$ and $\nu\nu^\star$ are bounded linear operators defined as $D_{\sqrt{p}} f(z) = \sqrt{p(z)} f(z)$ and $pp^\star f(z) = p(z) \cdot \langle p, f \rangle$. Substituting this into the previous system, letting $u_f = D_{\sqrt{q}} U f$, $u_h = D_{\sqrt{p}} U^\star h$, we are left with

$$f + D_{\sqrt{p}} u_h - p \cdot c_h = g, \quad -D_{\sqrt{q}} u_f + q \cdot c_f + h = 0 \tag{21}$$

where we defined the scalars c_h, c_f as

$$c_h = \langle p, U^\star h \rangle = \langle \sqrt{p}, D_{\sqrt{p}} U^\star h \rangle = \langle \sqrt{p}, u_h \rangle,$$
$$c_f = \langle q, U f \rangle = \langle \sqrt{q}, D_{\sqrt{q}} U f \rangle = \langle \sqrt{q}, u_f \rangle$$

Below, we repeatedly use

$$\forall p \in \Delta^\infty, \quad \|D_{\sqrt{p}} f\|_2 \le \|f\|_\infty, \quad \|D_{\sqrt{p}} f\|_1 \le \|f\|_2$$

which can be shown by using Hölder's inequality and $p \in \Delta^\infty$, i.e., $p() \ge 0$ and $\int \mu_{\mathsf{ref}}(dx) p(x) = 1$.

Here, from the first equation of (21), taking L^1 norm and using triangle inequality, we see

$$\|f\|_1 \le \|D_{\sqrt{p}} u_h\|_1 + c_h \|p\|_1 + \|g\|_1 \le \|u_h\|_2 + |c_h| + \|g\|_1 \tag{22}$$

Thus, it is sufficient to bound the $\|u_h\|_2$ and $|c_h| = |\langle \sqrt{p}, u_h \rangle|$.

Now, multiplying the first equation in (21) by $D_{\sqrt{q}} U$ and the second equation by $D_{\sqrt{p}} U^\star$, letting $A = D_{\sqrt{q}} U D_{\sqrt{p}}$, noting $D_{\sqrt{p}} \sqrt{p} = p$ and $D_{\sqrt{q}} \sqrt{q} = q$, we have

$$u_f + A u_h - c_h A \sqrt{p} = D_{\sqrt{q}} U g, \qquad -A^\star u_f + u_h + c_f A^\star \sqrt{q} = 0. \tag{23}$$

Now we claim $\|A\|_{2 \to 2} \le \|U\|_\infty$. Indeed,

$$\forall f \in L^2, \quad \|Af\|_2 = \|D_{\sqrt{q}} U D_{\sqrt{p}} f\|_2 \le \|U D_{\sqrt{p}} f\|_\infty \le \|U\|_\infty \cdot \|D_{\sqrt{p}} f\|_1 \le \|U\|_\infty \|f\|_2$$

where we used $\|D_{\sqrt{p}} f\|_2 \le \|f\|_\infty$, $\|Uf\|_\infty \le \|U\|_\infty \|f\|_1$, and $\|D_{\sqrt{p}} f\|_1 \le \|f\|_2$.

Below, we solve the system (23) for (u_f, u_h) and obtain explicit upper bounds of (c_f, c_h). Substituting the second equation $u_h = A^\star u_f - c_f A^\star \sqrt{q}$ in (23) to the first equation, rearranging it, we get

$$(I + AA^\star) u_f = c_f A A^\star \sqrt{q} + c_h A \sqrt{p} + D_{\sqrt{q}} U g.$$

Now we claim that $(I + AA^\star)$ has a bounded inverse on L^2. Indeed, $(I + AA^\star)$ is coercive since $AA^\star$ is a self adjoint positive semidefinite linear operator on L^2. Moreover, the operator norm is bounded as

$$\|AA^\star\|_{2 \to 2} \le \|A\|_{2 \to 2}^2 \le \|U\|_\infty^2.$$

Thus, by the Lax-Milgram theorem, $(I+AA^\star)$ has a bounded inverse, and letting $T=(I+AA^\star)^{-1}$, the spectrum of T is controlled as

$$\forall f \in L^2, \quad (1+\|U\|_\infty^2)^{-1}\|f\|_2^2 \le \langle f, Tf\rangle \le \|f\|_2^2. \tag{24}$$

Then, u_f can be solved as

$$u_f = T\Big(c_f AA^\star\sqrt{q} + c_h A\sqrt{p} + D_{\sqrt{q}}Ug\Big) = c_f(I-T)\sqrt{q} + c_h TA\sqrt{p} + TD_{\sqrt{q}}Ug$$

where we used $TAA^\star = 1-T$, which follow from the definition $T=(I+AA^\star)^{-1}$.

By the same argument, $(I+A^\star A)$ has a bounded inverse on L^2, and letting $S=(I+A^\star A)^{-1}$, the spectrum of S is controlled as

$$\forall f \in L^2, \quad (1+\|U\|_\infty^2)^{-1}\|f\|_2^2 \le \langle f, Sf\rangle \le \|f\|_2^2 \tag{25}$$

and u_h can be solved explicitly as

$$u_h = c_h(I-S)\sqrt{p} - c_f \cdot SA^\star\sqrt{q} + SA^\star D_{\sqrt{q}}Ug$$

Putting all together, we get

$$\begin{aligned} u_f &= c_f(I-T)\sqrt{q} + c_h TA\sqrt{p} + TD_{\sqrt{q}}Ug \\ u_h &= c_h(I-S)\sqrt{p} - c_f SA^\star\sqrt{q} + SA^\star D_{\sqrt{q}}Ug. \end{aligned} \tag{26}$$

Taking the inner products $\langle\sqrt{q},\cdot\rangle$ and $\langle\sqrt{p},\cdot\rangle$ on the first equation and the second equation respectively, recalling $c_f = \langle\sqrt{q}, u_f\rangle$ and $c_h = \langle\sqrt{p}, u_h\rangle$, with $\langle\sqrt{p},\sqrt{p}\rangle = \langle\sqrt{q},\sqrt{q}\rangle = 1$ since $p,q\in\Delta^\infty$, we are left with

$$\begin{aligned} c_f &= c_f(1-\langle\sqrt{q}, T\sqrt{q}\rangle) + c_h\langle\sqrt{q}, TA\sqrt{p}\rangle + \langle\sqrt{q}, TD_{\sqrt{q}}Ug\rangle \\ c_h &= c_h(1-\langle\sqrt{p}, S\sqrt{p}\rangle) - c_f\langle\sqrt{p}, SA^\star\sqrt{q}\rangle + \langle\sqrt{p}, SA^\star D_{\sqrt{q}}Ug\rangle \end{aligned}$$

Noting that c_f and c_h are cancelled out, using the identity $TA = AS$ from the definition of S and T, we are left with the linear system of (c_f, c_h):

$$\begin{bmatrix} \alpha & -\gamma \\ \gamma & \beta \end{bmatrix}\begin{bmatrix} c_f \\ c_h \end{bmatrix} = \begin{bmatrix} \langle\sqrt{q}, TD_{\sqrt{q}}Ug\rangle \\ \langle\sqrt{p}, SA^\star D_{\sqrt{q}}Ug\rangle \end{bmatrix} \quad \text{where} \quad \begin{cases} \alpha &= \langle\sqrt{q}, T\sqrt{q}\rangle \\ \beta &= \langle\sqrt{p}, S\sqrt{p}\rangle \\ \gamma &= \langle\sqrt{q}, TA\sqrt{p}\rangle \end{cases}$$

Since $\|\sqrt{p}\|_2 = \|\sqrt{q}\|_2 = 1$, using the established bound of the spectrum of S and T (24)-(25), we know

$$(1+\|U\|_\infty^2)^{-1} \le \alpha \le 1, \quad (1+\|U\|_\infty^2)^{-1} \le \beta \le 1, \quad |\gamma| \le 2^{-1} \tag{27}$$

Here $|\gamma| \le 2^{-1}$ follows from the fact that $\|TA\|_{2\to2} = \sqrt{\|(TA)(TA)^\star\|_{2\to2}}$ where $(TA)(TA)^\star = AA^\star(I+AA^\star)^{-2}$, whose spectrum is given by $\sigma(1+\sigma)^{-2}$ where σ is the spectrum of $AA^\star$. Since $\sup_{x\ge0} x/(1+x)^2 = 1/4$, taking square root, we get $\|TA\|_{2\to2} \le 1/2$. By the same argument, $\|SA^\star\|_{2\to2} \le 1/2$ holds.

By the lower bound of α and β in (27), the determinant of the matrix is strictly positive:

$$\det \begin{bmatrix} \alpha & -\gamma \\ \gamma & \beta \end{bmatrix} = \alpha\beta + \gamma^2 \geq \frac{1}{(1+\|U\|_\infty^2)^2}.$$

Therefore, this matrix is invertible, and we obtain

$$\begin{bmatrix} c_f \\ c_h \end{bmatrix} = \frac{1}{\alpha\beta+\gamma^2} \begin{bmatrix} \beta & \gamma \\ -\gamma & \alpha \end{bmatrix} \begin{bmatrix} \langle \sqrt{q}, TD_{\sqrt{q}}Ug \rangle \\ \langle \sqrt{p}, SA^\star D_{\sqrt{q}}Ug \rangle \end{bmatrix}$$

Taking the absolute value and using the estimate of $|\alpha|, |\beta|, |\gamma|$ from (27), noting that $\|T\|_{2\to 2} \leq 1$ and $\|SA^\star\|_{2\to 2} \leq 1/2$ give

$$|\langle \sqrt{q}, TD_{\sqrt{q}}Ug \rangle| \leq \|D_{\sqrt{q}}Ug\|_2, \qquad |\langle \sqrt{p}, SA^\star D_{\sqrt{q}}Ug \rangle| \leq 2^{-1}\|D_{\sqrt{q}}Ug\|_2,$$

where $\|D_{\sqrt{q}}Ug\|_2 \leq \|Ug\|_\infty \leq \|U\|_\infty \|g\|_1$, we obtain the following inequality for each coordinate:

$$\begin{bmatrix} |c_f| \\ |c_h| \end{bmatrix} \leq (1+\|U\|_\infty^2)^2 \begin{bmatrix} 1 & 2^{-1} \\ 2^{-1} & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 2^{-1} \end{bmatrix} \cdot \|U\|_\infty \|g\|_1 = (1+\|U\|_\infty^2)^2 \|U\|_\infty \|g\|_1 \begin{bmatrix} 5/4 \\ 1 \end{bmatrix} \quad (28)$$

Let us finish the proof. By the second equation of (26),

$$u_h = c_h(I-S)\sqrt{p} - c_f SA^\star \sqrt{q} + SA^\star D_{\sqrt{q}}Ug.$$

Taking $\|\cdot\|_2$, using $\|I-S\|_{2\to 2} \leq 1$ from (25), $\|SA^\star\|_{2\to 2} \leq 1/2$ and $\|D_{\sqrt{q}}Ug\|_2 \leq \|U\|_\infty \|g\|_1$, we obtain

$$\|u_h\|_2 \leq |c_h| + 2^{-1}|c_f| + 2^{-1}\|U\|_\infty \|g\|_1.$$

Combined with the upper bound $\|f\|_1 \leq \|u_h\|_2 + |c_h| + \|g\|_1$ from (22),

$$\|f\|_1 \leq |c_h| + 2^{-1}|c_f| + 2^{-1}\|U\|_\infty \|g\|_1 + |c_h| + \|g\|_1 = [2^{-1}, 2] \begin{bmatrix} |c_f| \\ |c_h| \end{bmatrix} + (2^{-1}\|U\|_\infty + 1)\|g\|_1$$

Finally, substituting the upper bounds of $[|c_f|, |c_h|]$ in (28), with $[2^{-1}, 2] \cdot [5/4, 1]^\top = 21/8$, we obtain

$$\|f\|_1 \leq \Big(\frac{21}{8}(1+\|U\|_\infty^2)^2 \|U\|_\infty + 1 + \frac{1}{2}\|U\|_\infty \Big) \|g\|_1.$$

Expanding the RHS, one can find an absolute constant C such that $\|f\|_1 \leq C(1+\|U\|_\infty^5)\|g\|_1$. Recalling $f = (I + J_p U^\star J_q U)^{-1} g$, the proof is complete.

D.3. Uniform Fréchet differentiability

We have seen in Section D.1 that MCHF and NLHF are both Fréchet differentiable; there exists derivatives $Dp_*(U)$ for $* \in \{\mathsf{MC}, \mathsf{NL}\}$ such that

$$\forall U \in L^\infty(\mu_{\mathsf{ref}} \otimes \mu_{\mathsf{ref}}), \qquad \lim_{\|E\|_\infty \to 0} \frac{\|p_*(U+E) - p_*(U) - Dp_*(U)[E]\|_1}{\|E\|_\infty} = 0.$$

Now we aim to make this convergence result uniform for all U such that $\|U\|_\oplus \leq R$ for a fixed constant R, i.e., we aim to show the above convergence with $\sup_{\|U\|_\oplus \leq R}$ inside $\lim_{\|E\|_\infty \to 0}$.

Lemma 26 $\|J_p - J_q\|_{\infty\to 1} = \sup_{\|f\|_\infty \le 1} \|(J_p - J_q)f\|_1 \le 3\|p-q\|_1$ *for any* $p, q \in \Delta^\infty$

Proof For any $f \in L^\infty$, writing $J_p f = p \odot f - p \cdot \langle p, f\rangle$,

$$(J_p - J_q)f = (p-q)\odot f - (p-q)\cdot\langle p, f\rangle - q\cdot\langle p-q, f\rangle$$

Using Hölder's inequality with $\|p\|_1 = \|q\|_1 = 1$, one can show that the L^1 norm of each term on RHS is bounded by $\|p-q\|_1\|f\|_\infty$. Thus, by the triangle inequality, complete the proof. ■

D.3.1. MCHF

Lemma 27 *The Fréchet derivative* $Dp_{\mathsf{MC}}(U)$ *is locally Lipschitz in the sense that for all* $U, V \in L^\infty$,

$$\sup_{\|E\|_\infty \le 1} \|Dp_{\mathsf{MC}}(U)[E] - Dp_{\mathsf{MC}}(V)[E]\|_1 \le 5\cdot\exp(2\|U\|_\oplus + 2\|V\|_\oplus)\cdot\|U-V\|_\infty$$

Proof Let A_U and h_U as

$$A_U = I - P_U^\star, \quad h_{U,E}(y) = \int \mu_{\mathsf{ref}}(dx) p_{\mathsf{MC}}(U)(x)[J_{P_U(x,\cdot)}E(x,\cdot)](y)$$

so that $Dp_{\mathsf{MC}}(U)[E] = \left(A_U\mid_{L_0^1}\right)^{-1} h_{U,E}$. We use the decomposition

$$Dp(U)[E] - Dp(V)[E] = \left((A_U|_{L_0^1})^{-1} - (A_V|_{L_0^1})^{-1}\right)h_{V,E} + (A_U|_{L_0^1})^{-1}(h_{U,E} - h_{V,E})$$

First, we note

$$(A_U|_{L_0^1})^{-1} - (A_V|_{L_0^1})^{-1} = (A_U|_{L_0^1})^{-1}(A_V - A_U)(A_V|_{L_0^1})^{-1}.$$

Recall $\|(A_U\mid_{L_0^1})^{-1}\|_{L_0^1\to L_0^1} \le e^{2\|U\|_\oplus}$ from (17), and $\|A_U - A_V\|_{1\to 1} = \|P_V^\star - P_U^\star\|_{1\to 1}$. Here, for all $h \in L^1$,

$$\|(P_V^\star - P_U^\star)h\|_1 = \|\int \mu_{\mathsf{ref}}(dx)h(x)(P_V(x,\cdot) - P_U(x,\cdot))\|_1 \le \|h\|_1 \operatorname*{esssup}_x \|P_V(x,\cdot) - P_U(x,\cdot)\|_1.$$

Then, using Lemma 13, we have

$$\operatorname*{esssup}_x \|P_V(x,\cdot) - P_U(x,\cdot)\|_1 \le \operatorname*{esssup}_x \|V(x,\cdot) - U(x,\cdot)\|_\infty = \|U-V\|_\infty$$

and hence $\|A_U - A_V\|_{1\to 1} = \|P_V^\star - P_U^\star\|_{1\to 1} \le \|U-V\|_\infty$. Therefore, we obtain

$$\left\|\left((A_U|_{L_0^1})^{-1} - (A_V|_{L_0^1})^{-1}\right)\mid_{L_0^1}\right\|_{1\to 1} \le \exp(2\|U\|_\oplus)\exp(2\|V\|_\oplus)\|U-V\|_\infty$$

Next, we bound $h_{V,E}$. Using $\|J_p\|_{\infty\to 1} \le 1$ from Lemma 23, whenever $\|E\|_\infty \le 1$,

$$\|h_{V,E}\|_1 \le \int \mu_{\mathsf{ref}}(dx) p_{\mathsf{MC}}(V)(x)\|J_{P_V(x,\cdot)}E(x,\cdot)\|_1 \le \operatorname*{esssup}_x \|J_{P_V(x,\cdot)}E(x,\cdot)\|_1 \le 1.$$

Hence,

$$\left\|\left((A_U|_{L_0^1})^{-1} - (A_V|_{L_0^1})^{-1}\right)h_{V,E}\right\|_1 \le \exp(2\|U\|_\oplus + 2\|V\|_\oplus)\|U - V\|_\infty.$$

It remains to bound $h_{U,E} - h_{V,E}$. We write

$$\begin{aligned} h_{U,E} - h_{V,E} = &\int \mu_{\mathsf{ref}}(dx)\{p_{\mathsf{MC}}(U)(x) - p_{\mathsf{MC}}(V)(x)\}J_{P_U(x,\cdot)}E(x,\cdot) \\ &+ \int \mu_{\mathsf{ref}}(dx)p_{\mathsf{MC}}(V)(x)\{J_{P_U(x,\cdot)} - J_{P_V(x,\cdot)}\}E(x,\cdot). \end{aligned}$$

For the first term, using again $\operatorname{esssup}_x \|J_{P_U(x,\cdot)}E(x,\cdot)\|_1 \le \|E\|_\infty \le 1$, we get

$$\left\|\int \mu_{\mathsf{ref}}(dx)\{p_{\mathsf{MC}}(U)(x) - p_{\mathsf{MC}}(V)(x)\}J_{P_U(x,\cdot)}E(x,\cdot)\right\|_1 \le \|p_{\mathsf{MC}}(U) - p_{\mathsf{MC}}(V)\|_1.$$

By Corollary 21,

$$\|p_{\mathsf{MC}}(U) - p_{\mathsf{MC}}(V)\|_1 \le \exp\{2(\|U\|_\oplus \wedge \|V\|_\oplus)\}\|U - V\|_\infty.$$

Thus, the first term is bounded by $\exp\{2(\|U\|_\oplus \wedge \|V\|_\oplus)\}\|U - V\|_\infty$. For the second term, using $\|J_p - J_q\|_{\infty\to 1} \le 3\|p - q\|_1$ in Lemma 26,

$$\left\|\int \mu_{\mathsf{ref}}(dx)p_{\mathsf{MC}}(V)(x)\{J_{P_U(x,\cdot)} - J_{P_V(x,\cdot)}\}E(x,\cdot)\right\|_1 \le 3 \operatorname*{esssup}_x \|P_U(x,\cdot) - P_V(x,\cdot)\|_1$$

Using again $\operatorname{esssup}_x \|P_U(x,\cdot) - P_V(x,\cdot)\|_1 \le \|U - V\|_\infty$, the second term is bounded by $3\|U - V\|_\infty$. Combining the two estimates gives

$$\|h_{U,E} - h_{V,E}\|_1 \le \exp\{2(\|U\|_\oplus \wedge \|V\|_\oplus)\}\|U - V\|_\infty + 3\|U - V\|_\infty.$$

Absorbing constants, we may write

$$\|h_{U,E} - h_{V,E}\|_1 \le 4\exp\{2(\|U\|_\oplus \wedge \|V\|_\oplus)\}\|U - V\|_\infty.$$

Therefore,

$$\|(A_U|_{L_0^1})^{-1}(h_{U,E} - h_{V,E})\|_1 \le \exp(2\|U\|_\oplus)\|h_{U,E} - h_{V,E}\|_1 \le 4\exp(2\|U\|_\oplus + 2\|V\|_\oplus)\|U - V\|_\infty.$$

Putting the two bounds together, for every $\|E\|_\infty \le 1$,

$$\|Dp_{\mathsf{MC}}(U)[E] - Dp_{\mathsf{MC}}(V)[E]\|_1 \le (1+4)\exp(2\|U\|_\oplus + 2\|V\|_\oplus)\|U - V\|_\infty.$$

Taking the supremum over $\|E\|_\infty \le 1$ completes the proof. ■

Corollary 28 (Uniform Fréchet differentiability for MCHF) *For any constant $R \ge 0$,*

$$\lim_{\|E\|_\infty \to 0} \sup_{\|U\|_\oplus \le R} \frac{\|p_{\mathsf{MC}}(U+E) - p_{\mathsf{MC}}(U) - Dp_{\mathsf{MC}}(U)[E]\|_1}{\|E\|_\infty} = 0.$$

Proof Fix $U \in L^\infty(\mu_{\mathsf{ref}} \otimes \mu_{\mathsf{ref}})$. Since $t \mapsto p_{\mathsf{MC}}(U + tE)$ is C^1 from $\mathbb{R}$ to L^1 by Theorem 17, the fundamental theorem of calculus in Banach spaces gives

$$p_{\mathsf{MC}}(U + E) - p_{\mathsf{MC}}(U) - Dp_{\mathsf{MC}}(U)[E] = \int_0^1 \{Dp_{\mathsf{MC}}(U + tE) - Dp_{\mathsf{MC}}(U)\}[E]dt.$$

Taking the L^1-norm yields

$$\|p_{\mathsf{MC}}(U + E) - p_{\mathsf{MC}}(U) - Dp_{\mathsf{MC}}(U)[E]\|_1 \le \int_0^1 \|Dp_{\mathsf{MC}}(U + tE)[E] - Dp_{\mathsf{MC}}(U)[E]\|_1 dt.$$

Applying Lemma 27 with $V = U + tE$, we obtain

$$\begin{aligned}\|Dp_{\mathsf{MC}}(U + tE)[E] - Dp_{\mathsf{MC}}(U)[E]\|_1 &\le 5\|E\|_\infty \exp\bigl(2\|U + tE\|_\oplus + 2\|U\|_\oplus\bigr)\|tE\|_\infty \\ &\le 5t \exp\bigl(4\|U\|_\oplus + 2\|E\|_\infty\bigr)\|E\|_\infty^2,\end{aligned}$$

for all $t \in [0, 1]$. Here, the second inequality follows from the triangle inequality for the seminorm $\|\cdot\|_\oplus$ and the property $\|\cdot\|_\oplus \le \|\cdot\|_\infty$. Hence, taking the integration $\int_0^1 dt$, using $\int_0^1 tdt = 1/2$, we get

$$\|p_{\mathsf{MC}}(U + E) - p_{\mathsf{MC}}(U) - Dp_{\mathsf{MC}}(U)[E]\|_1 \le (5/2) \cdot \exp\bigl(4\|U\|_\oplus + 2\|E\|_\infty\bigr)\|E\|_\infty^2.$$

Therefore,

$$\sup_{\|U\|_\oplus \le R} \frac{\|p_{\mathsf{MC}}(U + E) - p_{\mathsf{MC}}(U) - Dp_{\mathsf{MC}}(U)[E]\|_1}{\|E\|_\infty} \le \frac{5}{2} \exp\bigl(4R + 2\|E\|_\infty\bigr)\|E\|_\infty.$$

Letting $\|E\|_\infty \to 0$ proves the claim. ■

D.3.2. NLHF

Lemma 29 *There exists a universal constant $c > 0$ such that for any $U, V \in L^\infty(\mu_{\mathsf{ref}} \otimes \mu_{\mathsf{ref}})$,*

$$\sup_{\|E\|_\infty \le 1} \|Dp_{\mathsf{NL}}(U)[E] - Dp_{\mathsf{NL}}(V)[E]\|_1 \le c(1 + \|U\|_\oplus^c + \|V\|_\oplus^c)\|U - V\|_\infty.$$

Proof The proof follows the same structure as that of Lemma 27. Write $p_U = p_{\mathsf{NL}}(U)$, $q_U = q_{\mathsf{NL}}(U)$ and define

$$A_U = I + J_{p_U} U^\star J_{q_U} U, \qquad h_{U,E} = J_{p_U} E^\star q_U - J_{p_U} U^\star J_{q_U} E p_U.$$

By Theorem 19, $Dp_{\mathsf{NL}}(U)[E] = A_U^{-1} h_{U,E}$, so we decompose

$$Dp_{\mathsf{NL}}(U)[E] - Dp_{\mathsf{NL}}(V)[E] = A_U^{-1}(A_V - A_U)A_V^{-1} h_{V,E} + A_U^{-1}(h_{U,E} - h_{V,E}),$$

where we used the identity $A_U^{-1} - A_V^{-1} = A_U^{-1}(A_V - A_U)A_V^{-1}$.

We now collect the bounds needed for the two terms. First, by Lemma 23 and Lemma 26, for any $K \in L^\infty(\mu_{\mathsf{ref}} \otimes \mu_{\mathsf{ref}})$ and any $p, q \in \Delta^\infty$, we have

$$\begin{aligned}\forall h \in L^1, \quad &\|J_p K h\|_1 \le \|Kh\|_\infty \le \|K\|_\infty \|h\|_1 \\ \forall h \in L^1, \quad &\|(J_p - J_q)Kh\|_1 \le 3\|p - q\|_1 \|Kh\|_\infty \le 3\|p - q\|_1 \|K\|_\infty \|h\|_1.\end{aligned}$$

Moreover, if $h \in L_0^1$, with additive parts removed as in the proof of Theorem 22, the above bounds hold with $\|K\|_\infty$ replaced by $\|K\|_\oplus$:

$$\forall h \in L_0^1, \quad \|J_p K h\|_1 \le \|K\|_\oplus \|h\|_1, \quad \|(J_p - J_q)Kh\|_1 \le 3\|p-q\|_1 \|K\|_\oplus \|h\|_1,$$

Second, by Theorem 22, noting that the same bound applied to the dual variable μ_{NL} by replacing U by $-U^\star$ in the proof, there exists a universal constant $c > 0$ such that

$$\|p_U - p_V\|_1 + \|q_U - q_V\|_1 \le c\big(1 + \|U\|_\oplus^6 + \|V\|_\oplus^6\big)\|U - V\|_\infty.$$

Third, by the L^1-bound of the inverse operator from Lemma 24, with additive parts removed as in the proof of Theorem 22,

$$\forall h \in L_0^1, \quad \|A_U^{-1} h\|_1 \le c(1 + \|U\|_\oplus^5)\|h\|_1, \qquad \|A_V^{-1} h\|_1 \le c(1 + \|V\|_\oplus^5)\|h\|_1.$$

Notice that A_U maps L_0^1 into itself because J_{p_U} always outputs a mean-zero function. Also, $h_{U,E}$ and $h_{V,E}$ belong to L_0^1, since they are in the ranges of J_{p_U} and J_{p_V}, respectively.

Given these bounds, we next claim that

$$\|(A_U - A_V)\mid_{L_0^1}\|_{1\to 1} \le c\big(1 + \|U\|_\oplus^c + \|V\|_\oplus^c\big)\|U - V\|_\infty.$$

Indeed, for any $h \in L_0^1$,

$$(A_U - A_V)h = J_{p_U} U^\star J_{q_U} U h - J_{p_V} V^\star J_{q_V} V h.$$

Expanding this difference into the four natural terms and using the bounds mentioned above, we get

$$\forall h \in L_0^1, \quad \|(A_U - A_V)h\|_1 \le c\big(1 + \|U\|_\oplus^c + \|V\|_\oplus^c\big)\|U - V\|_\infty \|h\|_1.$$

Similarly,

$$\|h_{V,E}\|_1 \le \|J_{p_V} E^\star q_V\|_1 + \|J_{p_V} V^\star J_{q_V} E p_V\|_1 \le (1 + \|V\|_\oplus)\|E\|_\infty,$$

and

$$\|h_{U,E} - h_{V,E}\|_1 \le c\big(1 + \|U\|_\oplus^c + \|V\|_\oplus^c\big)\|U - V\|_\infty \|E\|_\infty.$$

The second bound follows by expanding

$$h_{U,E} - h_{V,E} = \{J_{p_U} E^\star q_U - J_{p_V} E^\star q_V\} - \{J_{p_U} U^\star J_{q_U} E p_U - J_{p_V} V^\star J_{q_V} E p_V\},$$

and applying the same ingredients used to prove the bound on $\|(A_U - A_V)\mid_{L_0^1}\|_{1\to 1}$.

Combining these estimates, we obtain

$$\begin{aligned}
&\|Dp_{\mathsf{NL}}(U)[E] - Dp_{\mathsf{NL}}(V)[E]\|_1 \\
&\le \|(A_U\mid_{L_0^1})^{-1}\|_{1\to 1}\Big(\|(A_U - A_V)\mid_{L_0^1}\|_{1\to 1}\|(A_V\mid_{L_0^1})^{-1}\|_{1\to 1}\|h_{V,E}\|_1 + \|h_{U,E} - h_{V,E}\|_1\Big) \\
&\le c\big(1 + \|U\|_\oplus^c + \|V\|_\oplus^c\big)\|U - V\|_\infty \|E\|_\infty.
\end{aligned}$$

Taking the supremum over $\|E\|_\infty \le 1$ completes the proof for Dp_{NL}. ■

Corollary 30 (Uniform Fréchet differentiability for NLHF) *For any fixed constant $R \ge 0$,*

$$\lim_{\|E\|_\infty \to 0} \sup_{\|U\|_\oplus \le R} \frac{\|p_{\mathsf{NL}}(U+E) - p_{\mathsf{NL}}(U) - Dp_{\mathsf{NL}}(U)[E]\|_1}{\|E\|_\infty} = 0$$

Proof This corollary follows by the exactly same argument as for the MCHF; apply the fundamental theorem of calculus with Lemma 29. Thus, we omit the proof. ■

Appendix E. Alignment dynamics of MCHF and NLHF

Throughout this section, let $p_*(U) = \frac{d\mu_*(U)}{d\mu_{\mathsf{ref}}}$ be the density for $* \in \{\mathsf{MC}, \mathsf{NL}\}$, where $\mu_{\mathsf{MC}}(U)$ and $\mu_{\mathsf{NL}}(U)$ are MCHF/NLHF equilibria based on pairwise utility U.

E.1. Proof of Theorem 7

By Corollaries 28 and 30, $U \mapsto p_*(U)$ is uniformly Fréchet differentiable in the following sense:

$$\forall * \in \{\mathsf{MC}, \mathsf{NL}\}, \quad \forall R \geq 0, \quad \lim_{\|E\|_\infty \to 0} \sup_{\|U\|_\oplus \leq R} \frac{\|p_*(U+E) - p_*(U) - Dp_*(U)[E]\|_1}{\|E\|_\infty} = 0 \quad (29)$$

where $Dp_*(U)[E]$ is a Fréchet derivative (see Theorems 17 and 19 for the explicit form for $* = \mathsf{MC}$ and $* = \mathsf{NL}$, respectively).

Note that the derivatives $Dp_{\mathsf{MC}}(U)$ and $Dp_{\mathsf{NL}}(U)$ differ for general U. However, when $U = -f \oplus f$ and E is antisymmetric, they collapse into the following simple form.

Proposition 31 *For any $f \in L^\infty$ and antisymmetric $E \in L^\infty(\mu_{\mathsf{ref}} \otimes \mu_{\mathsf{ref}})$, for each $* \in \{\mathsf{MC}, \mathsf{NL}\}$, we have*

$$Dp_*((-f) \oplus f)[E] = p_{\mathsf{RL}}(f) \odot E^\star p_{\mathsf{RL}}(f)$$

where $p_{\mathsf{RL}}(f)(y) = \frac{\exp(f(y))}{\int \mu_{\mathsf{ref}}(dy') \exp(f(y'))}$, $\odot$ is the entrywise product, and $E^\star h(\cdot) = \int \mu_{\mathsf{ref}}(dx) E(x, \cdot) h(x)$.

That is, when U takes the additive form $U = -f \oplus f$, and for antisymmetric perturbation E, the derivatives of MCHF and NLHF coincide.

Now, consider a sequence of antisymmetric utilities $U_1, U_2, \ldots$ such that $\|U_n\|_\oplus \to 0$. Let $\hat{f}_n(y) = \int \mu_{\mathsf{ref}}(dx) U_n(x, y)$ be the weighted column sum. Then, by Proposition 10, we know

$$\|U_n - (-\hat{f}_n) \oplus \hat{f}_n\|_\infty \asymp \|U_n\|_\oplus.$$

Applying (29) with $U = (-\hat{f}_n) \oplus \hat{f}_n$ and $E = U_n - (-\hat{f}_n) \oplus \hat{f}_n$, noting that $\|U\|_\oplus = 0$ in this case, using the derivative formula by Proposition 31, we complete the proof of Theorem 7.

E.2. Proof of Proposition 31

We prove a general version of Proposition 31 for general base point and perturbation (not necessarily antisymmetric):

Proposition 32 *For any $g, f \in L^\infty$ and $E \in L^\infty(\mu_{\mathsf{ref}} \otimes \mu_{\mathsf{ref}})$,*

$$Dp_{\mathsf{MC}}(g \oplus f)[E] = J_{p_{\mathsf{RL}}(f)} E^\star p_{\mathsf{RL}}(f)$$
$$Dp_{\mathsf{NL}}(g \oplus f)[E] = J_{p_{\mathsf{RL}}(f)} E^\star p_{\mathsf{RL}}(-g)$$

where $J_p h = p \odot h - p\langle p, h\rangle$.

Note that Proposition 31 immediately follows from Proposition 32. Indeed, if E is antisymmetric, we have that for any density $p \in \Delta^\infty$,

$$J_p E^\star p = p \odot E^\star p - p\langle p, E^\star p\rangle = p \odot E^\star p$$

since $\langle p, E^\star p\rangle = 0$ by the antisymmetry of E. Therefore, assuming the antisymmetry of E and substituting $g = -f$ in Proposition 32, we obtain $Dp_*((-f)\oplus f) = p_{\mathsf{RL}}(f)\odot E^\star p_{\mathsf{RL}}(f)$ for each $* \in \{\mathsf{MC}, \mathsf{NL}\}$, thereby completing the proof of Proposition 31.

Below we prove Proposition 32. Recall the derivative formula by Theorem 17:

$$
\begin{aligned}
&Dp_{\mathsf{MC}}(U)[E] \\
&= \Big((I - P_U^\star)\mid_{L_0^1}\Big)^{-1}\Big(y \mapsto \int \mu_{\mathsf{ref}}(dx)p_{\mathsf{MC}}(U)(x)P_U(x,y)\Big[E(x,y) - \int \mu_{\mathsf{ref}}(dy')P_U(x,y')E(x,y')\Big]\Big).
\end{aligned}
$$

When $U = g\oplus f$,

$$
P_{g\oplus f}(x,y) = \frac{\exp(g(x)+f(y))}{\int \mu_{\mathsf{ref}}(dy')\exp(g(x)+f(y'))} = \frac{\exp(f(y))}{\int \mu_{\mathsf{ref}}(dy')\exp(f(y'))} = p_{\mathsf{RL}}(f)(y)
$$

and $p_{\mathsf{MC}}(g\oplus f) = p_{\mathsf{RL}}(f)$. Note that $P^\star_{g\oplus f}h = 0$ for any $h \in L_0^1$ since

$$
P^\star_{g\oplus f}h(y) = \int \mu_{\mathsf{ref}}(dx)h(x)P_{g\oplus f}(x,y) = \Big(\int \mu_{\mathsf{ref}}(dx)h(x)\Big)p_{\mathsf{RL}}(f)(y) = 0.
$$

Thus, we have

$$
((I - P^\star_{g\oplus f})\mid_{L_0^1})^{-1} = I.
$$

On the other hand,

$$
\begin{aligned}
&\int \mu_{\mathsf{ref}}(dx)p_{\mathsf{MC}}(g\oplus f)(x)P_{g\oplus f}(x,y)\Big[E(x,y) - \int \mu_{\mathsf{ref}}(dy')P_{g\oplus f}(x,y')E(x,y')\Big] \\
&= \int \mu_{\mathsf{ref}}(dx)p_{\mathsf{RL}}(f)(x)p_{\mathsf{RL}}(f)(y)\Big[E(x,y) - \int \mu_{\mathsf{ref}}(dy')p_{\mathsf{RL}}(f)(y')E(x,y')\Big] \\
&= p_{\mathsf{RL}}(f)(y)\Big([E^\star p_{\mathsf{RL}}(f)](y) - \int \mu_{\mathsf{ref}}(dy')[E^\star p_{\mathsf{RL}}(f)](y')\Big) \\
&= [J_{p_{\mathsf{RL}}(f)}E^\star p_{\mathsf{RL}}(f)](y)
\end{aligned}
$$

Therefore, we get $Dp_{\mathsf{MC}}(U)[E] = J_{p_{\mathsf{RL}}(f)}E^\star p_{\mathsf{RL}}(f)$, and the proof is complete for MC.

For the NLHF, recall the derivative formula from Theorem 19:

$$
Dp_{\mathsf{NL}}(U)[E] = \big(I + J_{p(U)}U^\star J_{q(U)}U\big)^{-1}\big(J_{p(U)}E^\star q(U) - J_{p(U)}U^\star J_{q(U)}Ep(U)\big)
$$

where $p(U) = p_{\mathsf{NL}}(U)$ and $q(U) = q_{\mathsf{NL}}(U)$ on the RHS. Note $p_{\mathsf{NL}}(g\oplus f) = p_{\mathsf{RL}}(f)$ and $q_{\mathsf{NL}}(g\oplus f) = p_{\mathsf{RL}}(-g)$.

Now, for any $h \in L_0^2 = \{h \in L^2 : \int \mu_{\mathsf{ref}}(dy)h(y) = 0\}$,

$$
((f\oplus g)h)(x) = \int \mu_{\mathsf{ref}}(dy)(f\oplus g)(x,y)h(y) = \int \mu_{\mathsf{ref}}(dy)g(y)h(y).
$$

Thus, $(f\oplus g)h$ is a constant. Since $J_{p(U)}$ is a mean-subtracting operator, we get $J_{p_0}(f\oplus g)h = 0$ and hence

$$
J_{p(U)}(f\oplus g)\mid_{L_0^2} = 0.
$$

By the same argument, we also get $J_{q(U)}(g \oplus f) \mid_{L_0^2} = 0$. Since the image space of $J_{p(U)}$ and $J_{q(U)}$ are both L_0^2, when $U = g \oplus f$, we get

$$\left(I + J_{p(U)} U^\star J_{q(U)} U\right)^{-1} = I, \qquad J_{p(U)} E^\star q(U) - J_{p(U)} U^\star J_{q(U)} E p(U) = J_{p(U)} E^\star q(U).$$

Finally, substituting $p_{\mathsf{NL}}(g \oplus f) = p_{\mathsf{RL}}(f)$ and $q_{\mathsf{NL}}(g \oplus f) = p_{\mathsf{RL}}(-g)$, we obtain $Dp_{\mathsf{NL}}(g \oplus f)[E] = J_{p_{\mathsf{RL}}(f)} E^\star p_{\mathsf{RL}}(-g)$ and the proof is complete.

Remark 33 *Proposition 32 yields a generalization of Theorem 7 to a general $U \in L^\infty(\mu_{\mathsf{ref}} \otimes \mu_{\mathsf{ref}})$, not necessarily antisymmetric. Indeed, for any L^2 solution $(\hat{g}, \hat{f}) \in \operatorname{argmin}_{g,f} \|U - g \oplus f\|_2$, noting that $\|U - \hat{g} \oplus \hat{f}\|_\infty \asymp \|U\|_\oplus$ from Proposition 9, applying the uniform Fréchet differentiability results in Corollary 28 and 30, together with the derivative formula in Proposition 32, we obtain*

$$\|p_{\mathsf{MC}}(U) - p_{\mathsf{RL}}(\hat{f}) - J_{p_{\mathsf{RL}}(\hat{f})}(U - \hat{g} \oplus \hat{f})^\star p_{\mathsf{RL}}(\hat{f})\|_{L^1(\mu_{\mathsf{ref}})} = o(\|U\|_\oplus)$$
$$\|p_{\mathsf{NL}}(U) - p_{\mathsf{RL}}(\hat{f}) - J_{p_{\mathsf{RL}}(\hat{f})}(U - \hat{g} \oplus \hat{f})^\star p_{\mathsf{RL}}(-\hat{g})\|_{L^1(\mu_{\mathsf{ref}})} = o(\|U\|_\oplus)$$

for any sequence of $U \in L^\infty(\mu_{\mathsf{ref}} \otimes \mu_{\mathsf{ref}})$ such that $\|U\|_\oplus \to 0$.

Thus, MCHF and NLHF differ in their first order term by $p_{\mathsf{RL}}(\hat{f})$ and $p_{\mathsf{RL}}(-\hat{g})$. Notice that they coincide when U is antisymmetric, since $\hat{g} = -\hat{f}$. This recovers Theorem 7 in the antisymmetric case.

E.3. Proof of Theorem 8

Assume U is antisymmetric. Let us show the claim for $t = 1$ first. Noting that $d\mu_{\mathsf{ref}}/d\mu_{\mathsf{ref}} = \mathbf{1}$, using the notation in Section D.1, we can write $p^1_{\mathsf{MC}} = \frac{d\mu^1_{\mathsf{MC}}}{d\mu_{\mathsf{ref}}}$ and $p^1_{\mathsf{NL}} = \frac{d\mu^1_{\mathsf{NL}}}{d\mu_{\mathsf{ref}}}$ as

$$p^1_{\mathsf{MC}}(U) = P_U^\star \mathbf{1} = \int \mu_{\mathsf{ref}}(dx) \mathsf{s}(U(x, \cdot)), \quad p^1_{\mathsf{NL}}(U) = \mathsf{s}(U^\star \mathbf{1}).$$

Then, by the exactly same argument we used in Section D.1 and D.3, we can show that for any constant $R \geq 0$,

$$\lim_{\|E\|_\infty \to 0} \sup_{\|U\|_\oplus \leq R} \frac{\|p^1_*(U + E) - p^1_*(U) - Dp^1_*(U)[E]\|_1}{\|E\|_\infty} = 0, \tag{30}$$

where

$$Dp^1_{\mathsf{MC}}(U)[E] = \int \mu_{\mathsf{ref}}(dx) J_{\mathsf{s}(U(x,\cdot))} E(x, \cdot), \quad Dp^1_{\mathsf{NL}}(U)[E] = J_{\mathsf{s}(U^\star 1)} E^\star \mathbf{1}.$$

Here, when $U = g \oplus f$, we have $\mathsf{s}(U(x, \cdot)) = \mathsf{s}(U^\star \mathbf{1}) = p_{\mathsf{RL}}(f)$ and hence the Fréchet derivatives collapse into the following form for each $* \in \{\mathsf{MC}, \mathsf{NL}\}$:

$$Dp^1_*(g \oplus f)[E] = J_{p_{\mathsf{RL}}(f)} E^\star \mathbf{1}.$$

Now, substituting $E = U - (-\hat{f}) \oplus \hat{f}$ and $(g, f) = (-\hat{f}, \hat{f})$ with $\hat{f}(\cdot) = \int \mu_{\mathsf{ref}}(dx) U(x, \cdot) = U^\star \mathbf{1}$, we have

$$(U - (-\hat{f}) \oplus \hat{f})^\star \mathbf{1} = U^\star \mathbf{1} - ((-\hat{f}) \oplus \hat{f})^\star \mathbf{1} = \hat{f} - \hat{f} = 0$$

where the second equation follows from $((-\hat{f}) \oplus \hat{f})^\star \mathbf{1} = \hat{f}(\cdot) - \langle \mathbf{1}, \hat{f} \rangle$ and $\langle \mathbf{1}, \hat{f} \rangle = \langle \mathbf{1}, U^\star 1 \rangle = 0$ by the antisymmetry of U. Thus, when U is antisymmetric, the derivative at $(-\hat{f}) \oplus \hat{f}$ to the direction $U - (-\hat{f}) \oplus \hat{f}$ is 0:

$$Dp_*^1((-\hat{f}) \oplus \hat{f})[U - (-\hat{f}) \oplus \hat{f}] = J_{p_{\mathsf{RL}}(\hat{f})}(U - (-\hat{f}) \oplus \hat{f})^\star \mathbf{1} = J_{p_{\mathsf{RL}}(\hat{f})} 0 = 0.$$

Substituting this to (30), using $\|U - (-\hat{f}) \oplus \hat{f}\|_\infty \asymp \|U\|_\oplus$ from Proposition 10 and $p_*^1((-\hat{f}) \oplus \hat{f}) = p_{\mathsf{RL}}(\hat{f})$, we obtain

$$\|p_*^1(U) - p_{\mathsf{RL}}(\hat{f}) - 0\|_1 = o(\|U\|_\oplus).$$

This completes the proof for $t = 1$.

For $\sup_{t \geq 2}$, we compare the iterates with their equilibrium. By Corollary 3 and Theorem 5, for each $* \in \{\mathsf{MC}, \mathsf{NL}\}$, we have

$$\forall t \geq 1, \qquad \|p_*^t(U) - p_*(U)\|_1 \leq \|U\|_\oplus^t \|\mathbf{1} - p_*(U)\|_1 \leq 2\|U\|_\oplus^t.$$

Here the second inequality follows from the triangle inequality, $\|\mathbf{1} - p_*(U)\|_1 \leq \|\mathbf{1}\|_1 + \|p_*(U)\|_1 = 2$. For MCHF, this estimate follows directly from Corollary 3, since the contraction coefficient satisfies $c(U) \leq \|U\|_\oplus$. For NLHF, since U is antisymmetric, the two equilibrium marginals coincide, that is, $\mu_{\mathsf{NL}} = \nu_{\mathsf{NL}}$. Moreover, our single iterate p_{NL}^t can be identified with the two-stage iterates (μ_t, ν_t) in Theorem 5 through $p_{\mathsf{NL}}^{2t} = \frac{d\mu_t}{d\mu_{\mathsf{ref}}}$ and $p_{\mathsf{NL}}^{2t+1} = \frac{d\nu_t}{d\mu_{\mathsf{ref}}}$. Hence, the contraction estimate of Theorem 5 applies to the present iteration as well.

Therefore, for any sequence of antisymmetric utilities U such that $\|U\|_\oplus \to 0$,

$$\sup_{t \geq 2} \|p_*^t(U) - p_*(U)\|_1 = O(\|U\|_\oplus^2) = o(\|U\|_\oplus).$$

Combining this estimate with Theorem 7 and applying the triangle inequality completes the proof for $\sup_{t \geq 2}$.

Camera-Ready Edits Summary

- We have added an explanation of the motivation behind MCHF
- We have added a list of notations at the beginning of appendix
- We have polished the proofs in appendix.