Markov Chain from Human Feedback

author names withheld

Under Review for NExT-Game 2026

Abstract

We propose Markov Chain from Human Feedback (MCHF), an alternative approach for aligning generative models from pairwise human preferences. Unlike RLHF, which reduces comparisons to a scalar reward, and NLHF, which preserves pairwise utilities through a KL-regularized minimax objective, MCHF uses pairwise preferences directly to define a transition mechanism over model outputs. Given a pairwise utility $U(x, y)$ and a reference model μ_{ref}, we define a Markov kernel $\mathsf{P}(x, dy) \propto \exp(U(x, y))\mu_{\mathsf{ref}}(dy)$, and take its stationary distribution as the aligned model. We show that MCHF converges geometrically fast to the stationary distribution, with a convergence rate governed by the seminorm $\|U\|_{\oplus} = \inf_{g,f} \|U - g \oplus f\|_{\infty}$, which quantifies the non-transitive structure of the pairwise utility. We further show that a mirror-descent algorithm for NLHF, with an appropriate step size, satisfies an analogous structure-adaptive convergence guarantee. Finally, through a perturbation analysis, we prove that when $\|U\|_{\oplus}$ is small, MCHF and NLHF agree up to first order around a RLHF solution, which yields a unified view of reward-based, game-theoretic, and our Markovian alignment.

1. Introduction

Alignment from human feedback often begins with pairwise comparisons. RLHF reduces such data to a scalar reward R, typically through a BTL model [3, 9]:

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

and then performs KL-regularized policy optimization [2, 4, 13]:

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

More recently, several variants to the standard RLHF have been proposed [1, 6, 8, 10, 16]. However, this reduction to scalar rewards cannot directly represent cyclic or pair-specific preferences. Nash Learning from Human Feedback (NLHF) [11] avoids this reduction by computing the minimax optimization 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, 15] and multiplayer games [20].

The original paper [11] 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 [7, 18]. Nevertheless, it is not always clear that such a pessimistic objective is the right one for fine-tuning generative models. In addition, existing algorithms (cf. [11, 17, 19]) require repeatedly solving a KL-regularized optimization with an iteration dependent reward R_t and a local regularization term $D_{\mathsf{KL}}(\cdot|\mu_t)$, which is computationally difficult for large generative models, as discussed in [11] (see Section 2.3 for details).

At a structural level, however, pairwise comparisons naturally define a directed graph over the space of possible outputs, where edges encode relative preference information. This perspective is classical in ranking and social choice, and has also motivated spectral methods for rank aggregation, in which the stationary distribution of a Markov chain constructed from pairwise comparisons represents the aggregate ranking [12, 14].

Motivated by this perspective, we propose *Markov Chain from Human Feedback* (MCHF). We construct the Markov kernel P from the pairwise utility as $\mathsf{P}(x, dy) \propto \exp(U(x, y))\mu_{\mathsf{ref}}(dy)$. We then iteratively align μ_{ref} by $\mu_{\mathsf{ref}} \to \mu_{\mathsf{ref}}\mathsf{P} \to \mu_{\mathsf{ref}}\mathsf{P}^2 \cdots$ and take the stationary distribution as μ_{MC}. Our results can be summarized as follows:

- In Section 2.2, we prove the geometric convergence of MCHF to the stationary distribution μ_{MC}. In particular, the convergence rate is governed by the seminorm $\|U\|_\oplus = \inf_{g,f\in L^\infty(\mu_{\mathsf{ref}})} \|U - g \oplus f\|_\infty$ which measures the intransitive part of U. We show that the MCHF implicitly adapts to the additive structure of U.
- In Section 2.3, we provide a refined convergence rate analysis for NLHF, showing that the NLHF also adapts to this additive structure of U if step-size is appropriately chosen.
- In Section 2.4, we compare MCHF and NLHF through (1) computational perspective, (2) coupling perspective, and (3) alignment dynamics. Especially, as for (3), we develop a perturbation analysis for MCHF and NLHF, showing that when $\|U\|_\oplus$ is small, MCHF and NLHF agrees up to the first order around the RLHF with reward $f(y) = \int \mu_{\mathsf{ref}}(dx)U(x, y)$.

2. Main result

2.1. Definition of seminorm $\|\cdot\|_\oplus$: Measuring non-additive structure of Utility

Let $(\mathcal{X}, \mathcal{F}, \mu_{\mathsf{ref}})$ be a probability space. We write $\|X\|_p = \|X\|_{L^p(\mu_{\mathsf{ref}})}$ for the L^p norm. Let $U(\cdot, \cdot) : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ be a utility function that encodes preference information. Throughout this paper, we assume $U \in L^\infty(\mu_{\mathsf{ref}} \otimes \mu_{\mathsf{ref}})$.

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 *additive defect* of the utility; the larger $\|U\|_\oplus$ is, the more U has a non-additive structure. Furthermore, when U is antisymmetric, $\|U\|_\oplus$ quantifies the *non-transitivity* of U: $\|U\|_\oplus \asymp \operatorname{esssup}_{x,y,z} |U(x, y) + U(y, z) + U(z, x)|$ (see Proposition 13).

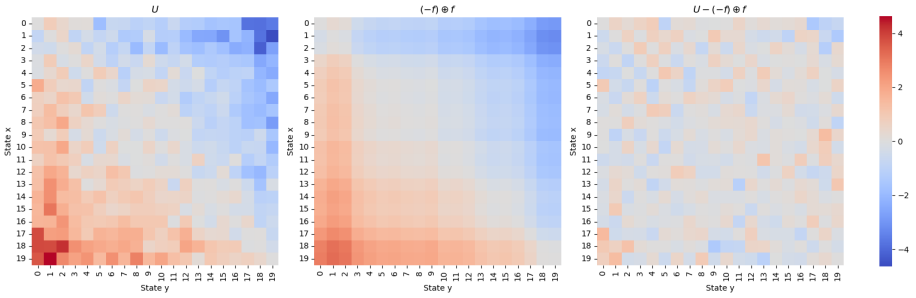

Figure 1: Comparison of antisymmetric utility U generated by (8), $(-\hat{f}) \oplus \hat{f}$ with $\hat{f}(y) = \int \mu_{\mathsf{ref}}(dx)U(x,y)$, and residual $U - (-\hat{f}) \oplus \hat{f}$.

Note that $\|U\|_\oplus \le \|U\|_\infty$ by the definition. In practice, the utility is often taken to be a possibly nonlinear transformation of the preference model $\mathcal{P}(x \prec y)$. Thus, the value of $\|U\|_\oplus$ depends both on the transformation and on the structure of $\mathcal{P}(x \prec y)$. Importantly, there are natural situations in which $\|U\|_\oplus \ll \|U\|_\infty$.

- Suppose $U(x,y) = \sigma_{\text{sigmoid}}^{-1} \circ \mathcal{P}(x \prec y)$ where $\sigma_{\text{sigmoid}}^{-1}(t) = \log t/(1-t)$ and $\mathcal{P}(x \prec y)$ takes the form $\mathcal{P}(x \prec y) = \Phi(R(y) - R(x) + E(x,y))$, where Φ is a nonlinear link function and $E(x,y)$ is a pairwise interaction term. In this case, $U(x,y)$ is close to $R(y) - R(x)$ (and hence $\|U\|_\oplus$ is small), whenever Φ is close to σ_{sigmoid} and the pairwise interaction term E is small. In contrast, $\|U\|_\infty$ can be large, depending on R and E (see Figure 1). 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 $\|U\|_\oplus = 0$.
- Suppose $U(x,y) = \mathcal{P}(x \prec y) - 1/2$ and $\mathcal{P}(x \prec y)$ follows the BTL model. If the reward R is small, using $\sigma_{\text{sigmoid}}(x) = \frac{1}{2} + \frac{x}{4} + O(x^3)$ as $x \to 0$, one can show $\|U\|_\oplus = O(\|R\|_\infty^3)$. Meanwhile, we can always construct R such that $\|U\|_\infty \asymp \|R\|_\infty$; for example, consider the discrete case $|\mathcal{X}| = 2$ with $R = (\gamma, -\gamma)$ and let $\gamma \to 0$. In this case, we have $\|U\|_\oplus / \|U\|_\infty = O(\|R\|_\infty^2) \to 0$.

The exact computation of the seminorm $\|U\|_\oplus$ can be challenging, since it is essentially an L^∞-Linear programming. However, we can estimate $\|U\|_\oplus$ up to a multiplicative order: $\|U\|_\oplus \asymp \|U - \hat{g} \oplus \hat{f}\|_\infty$, where $(\hat{g}, \hat{f})$ is a L^2-solution $(\hat{g}, \hat{f}) \in \inf_{g,f} \|U - g \oplus f\|_2$ (see Proposition 12).

2.2. Definition of MCHF, iterative convergence, and equilibrium

Based on the utility function $U \in L^\infty(\mu_{\mathsf{ref}} \otimes \mu_{\mathsf{ref}})$, we define the following Markov kernel P:

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

where $Z(x)$ is a normalizing constant. Here, the conditional distribution $Y|X \sim \mathsf{P}(X, \cdot)$ will move mass towards preferred directions (where $U(x, \cdot)$ takes large values) while preserving μ_{ref}.

Given the Markov kernel P defined by (4), we iteratively update μ_{ref} by the Markov chain:

$$\mu_{\mathsf{ref}} \mapsto \mu_{\mathsf{ref}}\mathsf{P} \to \mu_{\mathsf{ref}}\mathsf{P}^2 \cdots,$$

where $\mu\mathsf{P}(dy) = \int \mu(dx)\mathsf{P}(x, dy)$. Now we claim that the map $\mu \mapsto \mu\mathsf{P}$ is a contraction mapping on the complete metric space $(\mathscr{P}_{\mu_{\mathsf{ref}}}, d_{\mathsf{TV}})$, where $\mathscr{P}_{\mu_{\mathsf{ref}}}$ is the set of probability measure that is absolutely continuous with respect to μ_{ref}.

Theorem 2 *For any $U \in L^\infty(\mu_{\mathsf{ref}} \otimes \mu_{\mathsf{ref}})$, let P be the Markov Kernel defined by (4). 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 $c \in [0, 1)$ is given by

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

Note that $c(\|U\|_\oplus)$ is an increasing function of $\|U\|_\oplus$ and strictly less than 1. Thus, by the Banach fixed-point theorem, we obtain the following result.

Corollary 3 *There exists a unique stationary distribution $\mu_{\mathsf{MC}} = \mu_{\mathsf{MC}}\mathsf{P}$, and we have*

$$d_{\mathsf{TV}}(\mu_{\mathsf{ref}}\mathsf{P}^t, \mu_{\mathsf{MC}}) \le d_{\mathsf{TV}}(\mu_{\mathsf{ref}}, \mu_{\mathsf{MC}}) \cdot c(\|U\|_\oplus)^t.$$

We emphasize that the contraction rate $c(\|U\|_\oplus)$ is bounded in terms of the semi-norm $\|U\|_\oplus$, rather than the naive L^∞-norm $\|U\|_\infty$. Thus, the MCHF *implicitly* adapts to the additive structure of U.

2.3. Refined convergence rate analysis for NLHF

In this section, we propose a mirror descent algorithm for NLHF, which also adapts to the additive structure of U if the step size is chosen carefully.

Theorem 4 *For any step size $\eta > 0$, define the iteration (ν_t, μ_t) by*

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

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

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

Taking $\eta = \|U\|_\oplus^{-2}$ gives $D_{\mathsf{KL}}(\mu_{\mathsf{NL}} \mid \mu_t) \le (1 + \|U\|_\oplus^{-2})^{-t} D_{\mathsf{KL}}(\mu_{\mathsf{NL}} \mid \mu_{\mathsf{ref}})$. Related last-iterate convergence guarantees were obtained in, e.g., [19], for the special utility $U(x, y) = \lambda^{-1}(\mathcal{P}(x \succ y) - 1/2)$. However, their update also regularizes ν_t by $\eta^{-1} D_{\mathsf{KL}}(\nu \mid \nu_{t-1})$, and their rate depends on $\|U\|_\infty$ rather than $\|U\|_\oplus$. Since $\|U\|_\oplus \le \|U\|_\infty$, and $\|U\|_\oplus \ll \|U\|_\infty$ can occur when U has a strong additive component, Theorem 4 yields a sharper, structure-adaptive rate.

Note that our refined rate still requires choosing $\eta = \|U\|_\oplus^{-2}$, whose exact value may be hard to compute in practice, even though it can be estimated up to multiplicative constants by Proposition 12-13. A further computational issue is the local regularization $D_{\mathsf{KL}}(\mu \mid \mu_t)$. Indeed, the 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 $\tilde{\mu}_t^\eta(dy) = (d\mu_t/d\mu_{\mathsf{ref}})^{1/(1+\eta)}\mu_{\mathsf{ref}}(dy)$. Thus, while ν_t is obtained by an exponential tilt of μ_{ref}, updating μ_{t+1} requires sampling from the geometric mixture $\tilde{\mu}_t^\eta$ of μ_t and μ_{ref}, which is difficult in practice [11, Section F.1].

2.4. Comparison between MCHF and NLHF

Computational perspective As discussed above, for general U, last-iterate convergence of NLHF typically requires the local regularizer $\eta^{-1}D_{\mathsf{KL}}(\mu \mid \mu_{t-1})$. Consequently, each iteration must update the reward and sample from a geometric mixture of μ_{t-1} and μ_{ref}, which can be computationally challenging.

By contrast, MCHF requires implementing the conditional sampler $\mathsf{P}(x, dy)$, but we just need to implement it 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 The key difference between NLHF and MCHF lies in their update dynamics. For simplicity, consider an antisymmetric utility U and the iterations

$$\mu^t_{\mathsf{MC}} = \mu^{t-1}_{\mathsf{MC}}\mathsf{P}, \qquad \mu^t_{\mathsf{NL}} = \operatorname*{argmax}_{\mu} \int \mu^{t-1}_{\mathsf{NL}}(dx)\mu(dy)U(x,y) - D_{\mathsf{KL}}(\mu \mid \mu_{\mathsf{ref}}), \tag{7}$$

initialized at $\mu^0_{\mathsf{MC}} = \mu^0_{\mathsf{NL}} = \mu_{\mathsf{ref}}$. Note that when $\|U\|_{\oplus} < 1$ and U is antisymmetric, the NLHF iteration μ^t_{NL} also converges to μ_{NL} (see Theorem 18).

To compare the updates, define the induced couplings

$$\pi_{\mathsf{MC}}(dx, dy) = \mu_{t-1}(dx)\mathsf{P}(x, dy), \qquad \pi_{\mathsf{NL}}(dx, dy) = \mu_{t-1}(dx)\mu^t_{\mathsf{NL}}(dy),$$

whose second marginals are μ^t_{MC} and μ^t_{NL}, respectively. By the definition of the iterates and P, one can show that both couplings maximize the same objective $F(\pi)$ but over different feasible sets:

$$\pi_{\mathsf{MC}} = \operatorname*{argmax}_{\int \pi(\cdot,dy)=\mu_{t-1}(\cdot)} F(\pi), \quad \pi_{\mathsf{NL}} = \operatorname*{argmax}_{\exists \mu:\ \pi=\mu_{t-1}\otimes\mu} F(\pi), \quad F(\pi) = \iint U(x,y)\pi(dx,dy) - D_{\mathsf{KL}}(\pi \mid \mu_{t-1}\otimes\mu_{\mathsf{ref}}),$$

Thus, NLHF restricts the coupling to be independent, whereas MCHF optimizes over all couplings with first marginal μ_{t-1}. In this sense, MCHF can exploit more of the pairwise preference structure than the product coupling induced by NLHF.

This perspective also reveals a practical distinction. Since MCHF defines a conditional distribution $\mathsf{P}(x, \cdot)$, it naturally supports inference-time refinement: starting from an output x, one can repeatedly run the Markov chain by P until the user is satisfied. This resembles the inference-time alignment mechanism in Stackelberg game-based alignment [15]. In contrast, NLHF directly updates a marginal distribution and does not provide such a conditional refinement mechanism.

Alignment dynamics Let $\mu_{\mathsf{MC}}(U)$ be the stationary distribution of the Markov kernel $\mathsf{P}(x, dy) \propto \exp(U(x, y))\mu_{\mathsf{ref}}(dy)$, and let $\mu_{\mathsf{NL}}(U)$ be the NLHF solution (3) based on the utility function U.

Suppose $\|U\|_{\oplus} = 0$. Then by Proposition 12, we have $U(x, y) = \hat{g}(x) + \hat{f}(y)$ with $\hat{f}(y) = \int \mu_{\mathsf{ref}}(dx)U(x, y)$. Moreover, by the definitions of MCHF and NLHF, if $U = g \oplus f$, then both $\mu_{\mathsf{MC}}(U)$ and $\mu_{\mathsf{NL}}(U)$ collapse to the RLHF solution (2) with reward f. In summary,

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

where $\mu_{\mathsf{RL}}(\hat{f})(dz) \propto \exp(\hat{f}(z))\mu_{\mathsf{ref}}(dz)$.

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 $\mu_{\mathsf{RL}}(\hat{f})$? For simplicity, we focus on antisymmetric utilities satisfying $U(x, y) = -U(y, x)$. See Section E for the general utility case.

Theorem 5 *Suppose $U \in L^\infty(\mu_{\mathsf{ref}} \otimes \mu_{\mathsf{ref}})$ is antisymmetric, and let $\hat{f}(y) = \int \mu_{\mathsf{ref}}(dx)\, U(x,y)$. Let $p_*(U) = d\mu_*(U)/d\mu_{\mathsf{ref}}$ be the density for $* \in \{\mathsf{MC}, \mathsf{NL}\}$, and let $p_{\mathsf{RL}}(\hat{f}) = d\mu_{\mathsf{RL}}(\hat{f})/d\mu_{\mathsf{ref}}$. Then,*

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

where $\odot$ denotes entrywise multiplication and $K^\star h(\cdot) = \int \mu_{\mathsf{ref}}(dx)\, K(x,\cdot) h(x)$.

We can view $p_{\mathsf{RL}}(\hat{f}) + p_{\mathsf{RL}}(\hat{f}) \odot (U - (-\hat{f}) \oplus \hat{f})^\star p_{\mathsf{RL}}(\hat{f})$ as a first-order approximation of $p_*(U)$ around $U = (-\hat{f}) \oplus \hat{f}$. In particular, the second term captures the non-additive structure of the utility U. Interestingly, MCHF and NLHF agree up to the first-order term. Thus, by the triangle inequality,

$$d_{\mathsf{TV}}(\mu_{\mathsf{NL}}(U), \mu_{\mathsf{MC}}(U)) = 2^{-1} \|p_{\mathsf{NL}}(U) - p_{\mathsf{MC}}(U)\|_1 = o(\|U\|_\oplus).$$

We emphasize that the right-hand side is $o(\|U\|_\oplus)$, not merely $O(\|U\|_\oplus)$.

The next theorem claims that MCHF and NLHF coincide at the iteration level.

Theorem 6 *Let μ^t_{MC} and μ^t_{NL} be the iterations defined in (25), and let $p^t_*(U) = d\mu^t_*(U)/d\mu_{\mathsf{ref}}$ be the density for $* \in \{\mathsf{MC}, \mathsf{NL}\}$. Then, for each $* \in \{\mathsf{MC}, \mathsf{NL}\}$,*

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

Thus, the MCHF and NLHF iterations also agree up to first order. Moreover, since the first-order approximation is the same for all $t \geq 2$, this implies that most of the update is completed within the first two iterations.

Finally, we verify Theorem 5 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

$$U(x,y) = 2^{-1} \log \frac{\mathcal{P}(x \prec y)}{1 - \mathcal{P}(x \prec y)}, \qquad \mathcal{P}(x \prec y) = \Phi\big(R(y) - R(x) + 0.5E(x,y)\big), \tag{8}$$

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 $N(0,1)$. We visualize U 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 5. We observe that MCHF and NLHF nearly coincide, and that their deviation from RLHF is accurately captured by the first-order approximation.

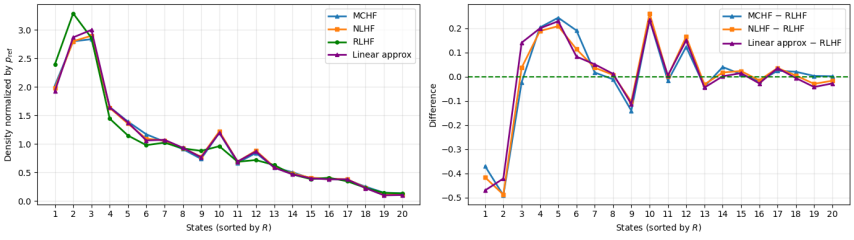

Figure 2: Comparison of MCHF, NLHF, RLHF, and the first-order approximation in Theorem 5.

References

[1] 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 *International Conference on Artificial Intelligence and Statistics*, pages 4447–4455. PMLR, 2024.

[2] 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.

[3] 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.

[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*, 2025.

[6] Kawin Ethayarajh, Winnie Xu, Niklas Muennighoff, Dan Jurafsky, and Douwe Kiela. Kto: Model alignment as prospect theoretic optimization. *arXiv preprint arXiv:2402.01306*, 2024.

[7] Paul Gölz, Nika Haghtalab, and Kunhe Yang. Distortion of ai alignment: Does preference optimization optimize for preferences? *arXiv preprint arXiv:2505.23749*, 2025.

[8] Audrey Huang, Wenhao Zhan, Tengyang Xie, Jason D Lee, Wen Sun, Akshay Krishnamurthy, and Dylan J Foster. Correcting the mythos of kl-regularization: Direct alignment without overoptimization via chi-squared preference optimization. *arXiv preprint arXiv:2407.13399*, 2024.

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

[10] Yu Meng, Mengzhou Xia, and Danqi Chen. Simpo: Simple preference optimization with a reference-free reward. *Advances in Neural Information Processing Systems*, 37:124198–124235, 2024.

[11] 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.

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

[13] 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.

[14] Lawrence Page, Sergey Brin, Rajeev Motwani, and Terry Winograd. The pagerank citation ranking: Bringing order to the web. Technical report, Stanford InfoLab, 1999.

[15] 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.

[16] 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.

[17] 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.

[18] 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.

[19] 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.

[20] 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.

Appendix A. Perturbation analysis

We analyze the sensitivity of the equilibrium distribution to perturbations in U. This is important because, in practice, the utility function U may be updated as new preference data become available. We may also consider a setting in which U is taken to be an increasing transformation of the preference model $\mathcal{P}(x \succ y)$, which is itself estimated, for example, by logistic regression as in the original paper [11]. In this case, estimation error in $\mathcal{P}$ induces a perturbation in U. It is therefore important to understand how such perturbations affect the resulting equilibrium.

We show that both MCHF and NLHF are locally Lipschitz with respect to perturbations of U under the L^∞ metric. The proofs of the theorem stated in this section are given in the subsequent sections.

A.1. MCHF

First, we discuss the sensitivity of MCHF. Let

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

be the constant given by Theorem 2, which upper bounds the Lipschitz constant of the map $\mu \mapsto \mu\mathsf{P}$ under the TV distance.

Theorem 7 *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, for all $t \geq 1$,*

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

See Section C.2 for the proof. Combining this result with Corollary 3, and letting $\mu_{\mathsf{MC}}(U)$ and $\mu_{\mathsf{MC}}(\hat{U})$ denote the stationary distributions of P and $\hat{\mathsf{P}}$, respectively, we obtain the following corollary.

Corollary 8 *For any $U, \hat{U} \in L^\infty(\mu_{\mathsf{ref}} \otimes \mu_{\mathsf{ref}})$,*

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

Since $c(\|U\|_\oplus)$ is always strictly less than 1, we conclude that the map $U \mapsto \mu_{\mathsf{MC}}(U)$ is locally Lipschitz.

A.2. NLHF

We now turn to NLHF and first establish the following stability result.

Theorem 9 *For any $U, \hat{U} \in L^\infty(\mu_{\mathsf{ref}} \otimes \mu_{\mathsf{ref}})$,*

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

This theorem follows from the strong convexity induced by the KL regularization term. Combined with Pinsker's inequality, $d_{\mathsf{TV}} \le \sqrt{D_{\mathsf{KL}}/2}$, it yields

$$d_{\mathsf{TV}}\big(\mu_{\mathsf{NL}}(U), \mu_{\mathsf{NL}}(\hat{U})\big) \le \|U - \hat{U}\|_\infty^{1/2}.$$

Thus, the map $U \mapsto \mu_{\mathsf{NL}}(U)$ is globally Hölder-$1/2$ continuous. The next theorem shows that this map becomes Lipschitz when the domain is suitably restricted.

Theorem 10 *If $\|U\|_\oplus \wedge \|\hat{U}\|_\oplus < 1$, then*

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

Compared with Corollary 8, this result shows that $U \mapsto \mu_{\mathsf{NL}}(U)$ is also locally Lipschitz as a map from $\|\cdot\|_\infty$ to d_{TV}, but the domain is restricted to the smaller set $\{U \in L^\infty(\mu_{\mathsf{ref}} \otimes \mu_{\mathsf{ref}}) : \|U\|_\oplus < 1\}$. We next show that μ_{NL} satisfies a global pseudo-Lipschitz bound.

Theorem 11 *There exists an absolute constant C such that, for any $U, \hat{U} \in L^\infty(\mu_{\mathsf{ref}} \otimes \mu_{\mathsf{ref}})$,*

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

The proof of this theorem requires a different technique from the preceding arguments. We first establish the differentiability of the map $\epsilon \mapsto \mu_{\mathsf{NL}}(U + \epsilon E)$ for fixed $U, E \in L^\infty(\mu_{\mathsf{ref}} \otimes \mu_{\mathsf{ref}})$ and then bound the L^1 norm of its derivative.

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

Proposition 12 *Fix $U \in L^\infty(\mu_{\mathsf{ref}} \otimes \mu_{\mathsf{ref}})$. Define $\square(U)$ as the rectangle defect:*

$$\square(U) = \operatorname*{esssup}_{x,x',y,y'} \big|U(x', y') - U(x, y') - U(x', y) + U(x, y)\big|.$$

Let $(\hat{g}, \hat{f})$ be a solution to $\min_{g,f} \|U - g \oplus f\|_2$. Then

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

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

The minimizer of the L^2 projection problem $\min_{g,f} \|U - g \oplus f\|_2$ is not unique, but one solution is given explicitly by

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

where $m = \iint U(x, y)\mu_{\mathsf{ref}}(dx)\mu_{\mathsf{ref}}(dy)$.

When U is antisymmetric, we obtain a sharper characterization.

Proposition 13 *Suppose $U \in L^\infty(\mu_{\mathsf{ref}} \otimes \mu_{\mathsf{ref}})$ is antisymmetric, i.e., $U(x, y) = -U(y, x)$. Define $\Delta(U)$ as the triangle defect:*

$$\Delta(U) = \operatorname*{esssup}_{x,y,z} |U(x, y) + U(y, z) + U(z, x)|.$$

Let $\hat{f}(y) = \int \mu_{\mathsf{ref}}(dx)U(x, y)$. Then

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

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

B.1. Proof of Proposition 12

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

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

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 set $U_{g,f} = U - g \oplus f$. Then for every x, x', y, y', noting that the g- and f-terms cancel,

$$U(x',y') - U(x,y') - U(x',y) + U(x,y) = U_{g,f}(x',y') - U_{g,f}(x,y') - U_{g,f}(x',y) + U_{g,f}(x,y),$$

Thus, taking $\operatorname{esssup}_{x,y,x',y'}$ on both sides, the LHS is $\square(U)$, while using the triangle inequality for the RHS, we get

$$\square(U) \le 4\|U_{g,f}\|_\infty$$

Since this holds for every f, g, taking $\inf_{g,f \in L^\infty}$ on the RHS, we 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),

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

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

$$\begin{aligned}\|U - \hat{g} \oplus \hat{f}\|_\infty &\le \operatorname*{esssup}_{x,y} \Big| \iint U(x,y) - U(x',y) - U(x,y') + U(x',y')\mu_{\mathsf{ref}}(dx')\mu_{\mathsf{ref}}(dy') \Big| \\ &\le \operatorname*{esssup}_{x,y,x',y'} \big| U(x,y) - U(x',y) - U(x,y') + U(x',y') \big| \\ &= \square(U),\end{aligned}$$

so the proof is complete.

B.2. Proof of Proposition 13

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

$$(-h) \oplus h = \frac{1}{2}g \oplus f - \frac{1}{2}f \oplus g$$

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

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

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

Therefore, 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. Proof of MCHF

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 $d_{\mathsf{TV}}(\mu\mathsf{P}, \nu\mathsf{P}) \le (1 - e^{-2\|U\|_\oplus}) d_{\mathsf{TV}}(\mu, \nu)$ VIA MINORIZATION LEMMA

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

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

and we define the bounded linear operator $P^\star : L^1 \to L^1$ as

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

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

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

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 following lemma.

Lemma 14 (Minoritation lemma) *For any $U \in L^\infty(\mu_{\mathsf{ref}} \otimes \mu_{\mathsf{ref}})$, let P be the Markov density given by (9). Then, the map $h \mapsto P^\star h$ is a contraction mapping on $L^1_0 = \{h \in L^1 : \int \mu_{\mathsf{ref}}(dy) h(y) = 0\}$ and with its contraction rate given by*

$$\sup_{h \in L^1_0 : h \neq 0} \frac{\|P^\star h\|_1}{\|h\|_1} \le 1 - \exp(-2\|U\|_\oplus)$$

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

Proof Let us fix $g, f \in L^\infty$ and let $U_{g,f} = U - g \oplus f$. Define $p_f \in L^\infty$ as

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

so that p_f is a valid density with respect to μ_{ref}, i.e., $p_f(\cdot) \ge 0$ and $\int \mu_{\mathsf{RL}}(dy) p_f(y) = 1$. Now, we can rewrite the Markov density P as

$$\begin{aligned} P(x, y) &= \frac{\exp(U(x, y))}{\int \mu_{\mathsf{ref}}(dy') \exp(U(x, y'))} \\ &= \frac{\exp(U(x, y) - g(x) - f(y)) \exp(f(y))}{\int \mu_{\mathsf{ref}}(dy') \exp(U(x, y') - g(x) - f(y')) \exp(f(y'))} \\ &= \frac{\exp(U_{g,f}(x, y)) p_f(y)}{\int \mu_{\mathsf{ref}}(dy') \exp(U_{g,f}(x, y')) p_f(y')} \end{aligned} \tag{10}$$

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

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

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

$$\tilde{P}(x,y) = \frac{P(x,y) - \delta p_f(y)}{1-\delta}$$

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,

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

For any $h \in L_0^1$ such that $h \neq 0$, noting $\int \mu_{\mathsf{ref}}(dx)h(x) = 0$, we get

$$\begin{aligned} P^\star h(y) &= \int \mu_{\mathsf{ref}}(dx)P(x,y)h(x) \\ &= \delta p_f(y) \int \mu_{\mathsf{ref}}(dx)h(x) + (1-\delta)\int \mu_{\mathsf{ref}}(dx)\tilde{P}(x,y)h(x) \\ &= (1-\delta)\tilde{P}^\star h(y). \end{aligned}$$

Taking the L_1-norm and using the fact (a.k.a. Data Processing Inequality):

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

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

$$\|P^\star h\|_1 \leq (1-\delta)\|\tilde{P}^\star h\|_1 \leq (1-\delta)\|h\|_1$$

so that

$$\sup_{h \in L_0^1 : h \neq 0} \frac{\|P^\star h\|_1}{\|h\|_1} \leq 1 - \delta.$$

Since $\delta = \exp(-2\|U_{g,f}\|_\infty)$ and $\|U_{g,f}\|_\infty = \|U - g \oplus f\|_\infty$, taking $\inf_{g,f}$ on the RHS, we complete the proof of (14). ■

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

$$\sup_{h \in L_0^1 : h \neq 0} \frac{\|P^\star h\|_1}{\|h\|_1} \leq \|U\|_\oplus.$$

We prove this using the following two lemmas.

Lemma 15 (Doeblin-Dobrushin characterization) *For any $P \in L^\infty(\mu_{\mathsf{ref}} \otimes \mu_{\mathsf{ref}})$,*

$$\sup_{h \in L_0^1 : h \neq 0} \frac{\|P^\star h\|_1}{\|h\|_1} \leq \operatorname*{esssup}_{x,x'} \frac{1}{2} \|P(x,\cdot) - P(x',\cdot)\|_1$$

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

Lemma 16 (Lipschitz continuity of Gibbs measure) *Fix a probability measure μ and define the map $\mathcal{G} : L^\infty(\mu) \to L^\infty(\mu)$ as*

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

Then, for all $g, \tilde{g} \in L^\infty(\mu)$, we have

$$\|\mathcal{G}(g) - \mathcal{G}(\tilde{g})\|_{L^1(\mu)} \leq \|g - \tilde{g}\|_{L^\infty(\mu)}$$

Applying Lemma 15, we have

$$\sup_{h \in L_0^1 : h \neq 0} \frac{\|P^\star h\|_1}{\|h\|_1} \leq \operatorname*{esssup}_{x,x'} \frac{1}{2} \|P(x,\cdot) - P(x',\cdot)\|_1.$$

Recall that for any $g, f \in L^\infty$, $P(x,y)$ can be written as follows (see (10) for the derivation):

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

where $U_{g,f} = U - g \oplus f$. Therefore, letting $\mu_f(dy) = \mu_{\mathsf{ref}}(dy) p_f(y)$ be the exponentially tilted probability measure by f, using Lemma 16 $\mu = \mu_f$ and $g(y) = U_{g,f}(x,y)$ and $\tilde{g}(y) = U_{g,f}(x',y)$, we have

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

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

$$\begin{aligned}
\sup_{h \in L_0^1 : h \neq 0} \frac{\|P^\star h\|_1}{\|h\|_1} &\leq \operatorname*{esssup}_{x,x'} \frac{1}{2} \operatorname*{esssup}_y |U_{g,f}(x,y) - U_{g,f}(x',y)| \\
&\leq \operatorname*{esssup}_{x,y} |U_{g,f}(x,y)| && \text{triangle inequality} \\
&= \|U_{g,f}\|_\infty
\end{aligned}$$

Since $U_{g,f} = U - g \oplus f$ and $g, f \in L^\infty(\mu_{\mathsf{ref}})$ are arbitrary, taking $\inf_{g,f}$ on the RHS, we complete the proof of $\sup_{h \in L_0^1 : h \neq 0} \frac{\|P^\star h\|_1}{\|h\|_1} \leq \|U\|_\oplus$. Below, we prove the intermediate lemmas we used.

C.1.3. PROOF OF LEMMA 15

Below we denote $\|X\|_p = \|X\|_{L^p(\mu_{\mathsf{ref}})}$ for simplicity. Note that $\|P^\star h\|_1$ can be represented as

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

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

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

where $Pf(x) \equiv \int \mu_{\mathsf{ref}}(dy) P(x,y) f(y)$. Since $\int \mu_{\mathsf{ref}}(dx) h(x) = 0$ by $h \in L^1_0$, for any constant $c \in \mathbb{R}$, we have

$$\int \mu_{\mathsf{ref}}(dx) h(x) P f(x) = \int \mu_{\mathsf{ref}}(dx) h(x) \big(P f(x) - c \big).$$

Now we let $c = 2^{-1}(\operatorname{esssup}_x Pf(x) + \operatorname{essinf}_x Pf(x)) \in \mathbb{R}$ so that

$$\operatorname*{esssup}_x |Pf(x) - c| \le \frac{1}{2} \Big(\operatorname*{esssup}_x Pf(x) - \operatorname*{essinf}_x Pf(x) \Big).$$

Combined with Hölder's inequality

$$\Big| \int \mu_{\mathsf{ref}}(dx) h(x) \big(Pf(x) - c \big) \Big| \le \|h\|_1 \cdot \frac{1}{2} \Big(\operatorname*{esssup}_x Pf(x) - \operatorname*{essinf}_x Pf(x) \Big).$$

Notice that for almost every x, x',

$$\begin{aligned} \frac{1}{2} \Big(Pf(x) - Pf(x') \Big) &= \frac{1}{2} \int \mu_{\mathsf{ref}}(dy) f(y) \big(P(x,y) - P(x',y) \big) \\ &\le \frac{1}{2} \|P(x,\cdot) - P(x',\cdot)\|_1 \cdot \|f\|_\infty \end{aligned}$$

Therefore, taking $\operatorname{esssup}_{x,x'}$, we are left with

$$\frac{1}{2} \Big(\operatorname*{esssup}_x Pf(x) - \operatorname*{essinf}_x Pf(x) \Big) \le \|f\|_\infty \cdot \operatorname*{esssup}_{x,x'} \frac{1}{2} \|P(x,\cdot) - P(x',\cdot)\|_1$$

Putting all together, we obtain

$$\Big| \int \mu_{\mathsf{ref}}(dy) f(y) P^\star h(y) \Big| \le \|f\|_\infty \|h\|_1 \operatorname*{esssup}_{x,x'} \frac{1}{2} \|P(x,\cdot) - P(x',\cdot)\|_1$$

for all $f \in L^\infty$ and $h \in L^1_0$. Taking supremum of f such that $\|f\|_\infty \le 1$, we complete the proof.

C.1.4. PROOF OF LEMMA 16

Let $h(x) = g(x) - \tilde{g}(x)$ and define the interpolator $g_t \in L^\infty$ for $t \in [0,1]$ as

$$\forall t \in [0,1], \quad g_t \equiv \tilde{g} + th$$

so that $g_0 = \tilde{g}$ and $g_1 = g$. Now we define the tilted probability measure μ_t as

$$\mu_t(dy) \equiv \mathcal{G}(g_t)(y)\mu(dy).$$

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

$$\|\mathcal{G}(g) - \mathcal{G}(\tilde{g})\|_1 = \sup_{\|f\|_{L^\infty(\mu)} \le 1} \left| \int f(x)\mu_1(dx) - f(x)\mu_0(dx) \right| = \sup_{\|f\|_{L^\infty(\mu)} \le 1} \left| \phi_f(1) - \phi_f(0) \right|$$

where

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

Taking the derivative of $\phi_f(t)$, noting $\frac{d}{dt} g_t(x) = h(x)$, and h is bounded,

$$\begin{aligned}
\frac{d}{dt}\phi_f(t) &= \int \mu(dx) f(x) \frac{\exp(g_t(x))h(x)}{\int \mu(dx') \exp(g_t(x'))} - \int \mu(dx) f(x) \exp(g_t(x)) \cdot \frac{\int \mu(dx') \exp(g_t(x'))h(x')}{\left(\int \mu(dx') \exp(g_t(x'))\right)^2} \\
&= \int \mu_t(dx) f(x) h(x) - \left(\int \mu_t(dx) f(x)\right) \cdot \left(\int \mu_t(dx) h(x)\right) \\
&= \mathrm{Cov}_{X \sim \mu_t}(f(X), h(X))
\end{aligned}$$

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:

$$|\frac{d}{dt}\phi_f(t)| \le \|f\|_{L^\infty(\mu_t)} \cdot \|h\|_{L^\infty(\mu_t)} = \|f\|_{L^\infty(\mu)} \|h\|_{L^\infty(\mu)}.$$

Therefore, we get

$$\|\mathcal{G}(g) - \mathcal{G}(\tilde{g})\|_1 = \sup_{\|f\|_{L^\infty} \le 1} \left| \phi_f(1) - \phi_f(0) \right| \le \sup_{\|f\|_{L^\infty(\mu)} \le 1} \int_0^1 |\frac{d}{dt}\phi_f(t)| dt \le \|h\|_{L^\infty(\mu)}.$$

This completes the proof.

C.2. Proof of Theorem 7

Let c and $\hat{c}$ be the Lipschitz constants (given by Theorem 2) of $\mu \mapsto \mathsf{P}\mu$ and $\mu \mapsto \hat{\mathsf{P}}\mu$, respectively. Consider the iterations

$$\mu_t = \mu_{t-1}\mathsf{P}, \qquad \hat{\mu}_t = \hat{\mu}_{t-1}\hat{\mathsf{P}},$$

initialized with $\mu_0 = \hat{\mu}_0 \in \mathscr{P}_{\mu_{\mathsf{ref}}}$. Using the triangle inequality, and the Lipschitz continuity of $\mu \mapsto \mu\mathsf{P}$,

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

By the definition of TV distance,

$$d_{\mathsf{TV}}(\hat{\mu}_{t-1}\mathsf{P}, \hat{\mu}_{t-1}\hat{\mathsf{P}}) = \frac{1}{2}\sup_{\|f\|_\infty \le 1}\Big|\int f(y)\big[(\hat{\mu}_{t-1}\mathsf{P})(dy) - (\hat{\mu}_{t-1}\hat{\mathsf{P}})(dy)\big]\Big|.$$

By Fubini's lemma, for any measurable function f with $\|f\|_\infty \le 1$,

$$\begin{aligned}
&\frac{1}{2}\Big|\int f(y)\big[(\hat{\mu}_{t-1}\mathsf{P})(dy) - (\hat{\mu}_{t-1}\hat{\mathsf{P}})(dy)\big]\Big| \\
&= \frac{1}{2}\Big|\int \mu_{\mathsf{ref}}(dy) f(y) \int \hat{\mu}_{t-1}(dx)\big(P(x,y) - \hat{P}(x,y)\big)\Big| \\
&= \frac{1}{2}\Big|\int \hat{\mu}_{t-1}(dx) \int \mu_{\mathsf{ref}}(dy) f(y)\big(P(x,y) - \hat{P}(x,y)\big)\Big| \\
&\le \frac{1}{2}\int \hat{\mu}_{t-1}(dx)\Big|\int \mu_{\mathsf{ref}}(dy) f(y)\big(P(x,y) - \hat{P}(x,y)\big)\Big| \\
&\le \frac{1}{2}\int \hat{\mu}_{t-1}(dx)\|P(x,\cdot) - \hat{P}(x,\cdot)\|_1 \qquad \|f\|_\infty \le 1 \\
&\le \frac{1}{2}\int \hat{\mu}_{t-1}(dx)\big\|U(x,\cdot) - \hat{U}(x,\cdot)\big\|_\infty \qquad \text{by Lemma 16 with } \mu = \mu_{\mathsf{ref}}, g(\cdot) = U(x,\cdot) \text{ and } \tilde{g}(\cdot) = \hat{U}(x,\cdot) \\
&\le \frac{1}{2}\|U - \hat{U}\|_\infty \qquad \hat{\mu}_{t-1} \ll \mu_{\mathsf{ref}}
\end{aligned}$$

Therefore,

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

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.

C.3. Differentiability of MCHF

Let us fix $U, E \in L^\infty(\mu_{\mathsf{ref}} \otimes \mu_{\mathsf{ref}})$. For each $\epsilon \in \mathbb{R}$, let p_ϵ be the unique stationary distribution of the Markov transition density P_ϵ defined by

$$P_\epsilon(x,y) = \frac{\exp(U_\epsilon(x,y))}{\int \mu_{\mathsf{ref}}(dy')\exp(U_\epsilon(x,y'))}, \quad U_\epsilon(x,y) \equiv U(x,y) + \epsilon E(x,y),$$

Now we define the linear operator $P_\epsilon^\star : L_1 \to L_1$ as

$$P_\epsilon^\star f(y) = \int \mu_{\mathsf{ref}}(dx) f(x) P_\epsilon(x,y).$$

Let μ_ϵ be the stationary distribution of the Markov kernel $\mathsf{P}_\epsilon(x,dy) = P_\epsilon(x,y)\mu_{\mathsf{ref}}(y)$ and let $p_\epsilon = \frac{d\mu_\epsilon}{d\mu_{\mathsf{ref}}}$ be its density. Note that this p_ϵ satisfies

$$P_\epsilon^\star p_\epsilon = p_\epsilon, \quad \int \mu_{\mathsf{ref}}(dy) p_\epsilon(y) = 1.$$

The next theorem calculates the derivative of p_ϵ in ϵ using the Implicit function theorem.

Theorem 17 *The map $\epsilon \mapsto p_\epsilon$ is differentiable at $\epsilon = 0$ in the sense that*

$$\lim_{\epsilon \to 0} \left\| \frac{p_\epsilon - p_0}{\epsilon} - \dot{p}_0 \right\|_1 = 0$$

where the derivative $\dot{p}_0 \in L_0^1$ is given by

$$\dot{p}_0 = \Big((I - P_0^\star) \mid_{L_0^1} \Big)^{-1} \left(y \mapsto \int \mu_{\mathsf{ref}}(dx) p_0(x) P_0(x,y) \Big(E(x,y) - \int \mu_{\mathsf{ref}}(dy') P_0(x,y') E(x,y') \Big) \right)$$

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

Proof Let $f_\epsilon = p_\epsilon - \mathbf{1}$, which lies in L_0^1 by $\int \mu_{\mathsf{ref}}(dy) p_\epsilon(y) = 1$. Define $F : L_0^1 \times \mathbb{R} \to L_0^1$ as

$$F(f, \epsilon) = (I - P_\epsilon^\star)(f + \mathbf{1})$$

so that $F(f_\epsilon, \epsilon) = 0$. Note that the image of F is included in L_0^1 by the fact that $\int \mu_{\mathsf{ref}}(dy) P_\epsilon(x,y) = 1$.

Below, we derive the derivative of f_ϵ in the Banach space $(L_0^1, \| \cdot \|_1)$. Notice that F is linear in f, especially the Jacobian $D_f F(f, \epsilon) : L_0^1 \to L_0^1$ with respect to f is given by

$$D_f F(f, \epsilon) = I - P_\epsilon^\star.$$

Now we claim that $I - P_0^\star$ has a bounded inverse on L_0^1. Noting that for all $h \in L_0^1$, Lemma 14 implies that there exists a constant $c = c(U) \in [0, 1)$ such that

$$\| P_0^\star h \|_1 \le c \cdot \| h \|_1.$$

Since $0 \le c < 1$, $(I - P_0^\star)$ has the inverse $(I - P_0^\star)^{-1} = \sum_{t=0}^\infty (P_0^\star)^t$ and its operator norm is bounded by $1/(1-c)$ since for all $h \in L_0^1$,

$$\| (I - P_0^\star)^{-1} h \|_1 = \| \sum_{t=0}^\infty (P_0^\star)^t h \|_1 \le \sum_{t=0}^\infty \| (P_0^\star)^t h \|_1 = \sum_{t=0}^\infty c^t \| h \|_1 = \frac{1}{1-c} \| h \|_1.$$

Thus, $\sum_{t=0}^\infty (P_0^\star)^t$ is the bounded inverse of $(I - P_0^\star)$ on L_0^1.

On the other hand, for the derivative with respect to ϵ, one can show that

$$D_\epsilon F(f, \epsilon)(y) = - \int \mu_{\mathsf{ref}}(dx) (f(x) + 1) P_\epsilon(x,y) \Big(E(x,y) - \int \mu_{\mathsf{ref}}(dy') P_\epsilon(x,y') E(x,y') \Big).$$

Therefore, by the Implicit function theorem for the Banach space $(L_0^1, \| \cdot \|_1)$, f_ϵ is differentiable at $\epsilon = 0$ in the sense that

$$\lim_{\epsilon \to 0} \| \frac{f_\epsilon - f_0}{\epsilon} - \dot{f}_0 \|_1 = 0$$

where the derivative f_0 is given by

$$\dot{f}_0 = \Big((I - P_0^\star) \mid_{L_0^1} \Big)^{-1} \left(y \mapsto \int \mu(dx) (f_0 + 1)(x) P_0(x,y) \Big(E(x,y) - \int \mu(dy') P_0(x,y') E(x,y') \Big) \right).$$

Substituting $f_\epsilon = p_\epsilon - 1$, we complete the proof. ■

Appendix D. Proof for NLHF

D.1. Geometric convergence under TV when $\|U\|_\oplus < 1$

Theorem 18 *The algorithm (6) with $\eta = \infty$ (i.e., no local regularization term) satisfies*

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

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

Here, the convergence rate is bounded by $\|U\|_\oplus$, rather than $\|U\|_\infty$. Thus, similarly to the MCHF update in Corollary 3, this algorithm implicitly adapts to the additive structure of U.

Proof 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

$$q_t = \operatorname*{argmin}_{q \in \Delta^\infty} \langle q, U p_t \rangle + w(q), \qquad q_{\mathsf{NL}} = \operatorname*{argmin}_{q \in \Delta^\infty} \langle q, U p_{\mathsf{NL}} \rangle + w(q),$$
$$p_{t+1} = \operatorname*{argmax}_{p \in \Delta^\infty} \langle q_t, U p \rangle - w(p), \qquad p_{\mathsf{NL}} = \operatorname*{argmax}_{p \in \Delta^\infty} \langle q_{\mathsf{NL}}, U p \rangle - w(p)$$

By Lemma 20-(1) and $\|U^\star\|_\oplus = \|U\|_\oplus$, we have

$$\|q_t - q_{\mathsf{NL}}\|_1 \le \|U\|_\oplus \|p_t - p_{\mathsf{NL}}\|_1, \quad \|p_{t+1} - p_{\mathsf{NL}}\|_1 \le \|U\|_\oplus \|q_t - q_{\mathsf{NL}}\|_1,$$

which completes the proof. ■

D.2. Proof of Theorem 4

Lemma 19 *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. We further assume that 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

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

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

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

We will prove Lemma 19 in Section D.2.1. Let us show Theorem 4 using Lemma 19. Define p_{NL} and q_{NL} as

$$p_{\mathsf{NL}} = \frac{d\mu_{\mathsf{NL}}}{d\mu_{\mathsf{ref}}}, \quad q_{\mathsf{NL}} = \frac{d\nu_{\mathsf{NL}}}{d\mu_{\mathsf{ref}}}.$$

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

$$\ell(p) = -\Big(\min_{q\in\Delta^\infty} \langle q, Up\rangle + w(q)\Big)$$

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

Thus, combined with Lemma 20-(2), we may apply Lemma 19 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 (q_t, p_t) defined by

$$\begin{aligned} q_t &= \operatorname*{argmin}_q \Big\{ \langle q, Up_t\rangle + w(q) \Big\} \\ p_{t+1} &= \operatorname*{argmin}_p \Big\{ \langle \nabla \ell(p_t), p\rangle + w(p) + \frac{1}{\eta} D_w(p\|p_t) \Big\} \\ &= \operatorname*{argmin}_p \Big\{ -\langle q_t, Up\rangle + w(p) + \frac{1}{\eta} D_w(p\|p_t) \Big\} \end{aligned}$$

satisfies that

$$D_w(p_*\|p_t) \leq \Big(\frac{1}{1+\eta}\Big)^t D_w(p_*\|p_0).$$

Finally, letting $\mu(dy) = p(y) d\mu_{\mathsf{ref}}(dy)$ 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.

D.2.1. PROOF OF LEMMA 19

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.

D.2.2. PROOF OF LEMMA 20

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

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

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

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

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

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

Then, for any $p, p' \in \Delta^\infty$, using $\|Uf\|_\infty \le \|U\|_\infty \|f\|_1$ by the Cauchy–Schwarz inequality and Lemma 16 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,

$$\begin{aligned}
\|q_p - q_{p'}\|_1 &\le \| - U_{g,f}p + U_{g,f}p'\|_{L^\infty(\mu_{-g})} && \text{by (13) and Lemma 16 with } \mu_{-g}(dx) \propto e^{-g(x)}\mu_{\mathsf{ref}}(dx)\\
&= \| - U_{g,f}p + U_{g,f}p'\|_\infty && \|\cdot\|_{L^\infty(\mu_{-g})} = \|\cdot\|_{L^\infty(\mu_{\mathsf{ref}})}\\
&\le \|U_{g,f}\|_\infty \|p - p'\|_1 && \text{Cauchy–Schwarz.}
\end{aligned}$$

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, using the derivative form (12),

$$\begin{aligned}
\|\nabla\ell(p) - \nabla\ell(p')\|_\infty &= \| - U^\star_{g,f}q_p + U^\star_{g,f}q_{p'}\|_\infty && \text{by (12)}\\
&\le \|U^\star_{g,f}\|_\infty \|q_p - q_{p'}\|_1 && \text{Cauchy–Schwarz}\\
&\le \|U_{g,f}\|_\infty \|U\|_\oplus \|p - p'\|_1 && \text{Cauchy–Schwarz.}
\end{aligned}$$

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

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

Therefore, letting $\phi(t) = \ell(tp + (1-t)q)$, we have

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

where the last inequality follows from Pinsker's inequality $D_w(p\|q) \ge \frac{1}{2}\|p - q\|_1^2$ for $w(p) = \int \mu_{\mathsf{ref}}(dy)p(y)\log p(y)$. Therefore, ℓ is relatively $\|U\|^2_\infty$-smooth with respect to w.

D.3. Proof of Theorem 9

Let $(\mu_{\mathsf{NL}}(U), \nu_{\mathsf{NL}}(U))$ be the solution to the NLHF with utility $U \in L^\infty(\mu_{\mathsf{ref}} \otimes \mu_{\mathsf{ref}})$ and let $p(U) = \frac{d\mu_{\mathsf{NL}}(U)}{d\mu_{\mathsf{ref}}}$. We showed in Section D.3 that $p(U)$ is the solution to the strongly convex optimization:

$$p(U) = \operatorname*{argmin}_{p \in \Delta^\infty} G_U(p) \equiv \ell_U(p) + w(p), \quad \ell_U(p) = -\Big(\min_{q \in \Delta^\infty} \langle q, Up\rangle + w(q)\Big)$$

where $w(p) = \int \mu_{\mathsf{ref}}(dz) p(z) \log p(z)$. Since G_U is strongly convex with respect to w, $p(U)$ also solves

$$p(U) \in \operatorname*{argmin}_{p \in \Delta^\infty} G_U(p) - D_w(p||p(U)).$$

Then, for any $U, \hat{U} \in L^\infty(\mu_{\mathsf{ref}} \otimes \mu_{\mathsf{ref}})$, using this optimality of $p(U)$ against $p(\hat{U})$, noting $D_w(p(U)||p(U)) = 0$, we have

$$G_U(p(U)) \le G_U(p(\hat{U})) - D_w(p(\hat{U})||p(U)).$$

Rearranging this inequality,

$$\begin{aligned}
D_w(p(\hat{U})||p(U)) &\le G_U(p(\hat{U})) - G_U(p(U)) \\
&= G_U(p(\hat{U})) - G_{\hat{U}}(p(\hat{U})) + G_{\hat{U}}(p(\hat{U})) - G_U(p(U)) \\
&\le G_U(p(\hat{U})) - G_{\hat{U}}(p(\hat{U})) + G_{\hat{U}}(p(U)) - G_U(p(U)) \quad p(\hat{U}) \in \operatorname*{argmin}_p G_{\hat{U}}(p) \\
&\le 2 \sup_{p \in \Delta^\infty} |G_U(p) - G_{\hat{U}}(p)| \qquad (15)
\end{aligned}$$

Here, letting $q_p \in \operatorname{argmin}_q \langle q, Up\rangle + w(q)$,

$$\begin{aligned}
G_U(p) &= -\langle q_p, Up\rangle - w(q_p) + w(p) \\
&= -\langle q_p, (U - \hat{U})p\rangle - \langle q_p, \hat{U}p\rangle - w(q_p) + w(p) \\
&\le -\langle q_p, (U - \hat{U})p\rangle - \min_q \Big(\langle q, \hat{U}p\rangle + w(q)\Big) + w(p) \\
&= -\langle q_p, (U - \hat{U})p\rangle + G_{\hat{U}}(p),
\end{aligned}$$

so that

$$G_U(p) - G_{\hat{U}}(p) \le |\langle q_p, (U - \hat{U})p\rangle| \le \|q_p\|_1 \|p\|_1 \|U - \hat{U}\|_\infty = \|U - \hat{U}\|_\infty$$

for all $p \in \Delta^\infty$. By symmetry, the above inequality also holds with U and $\hat{U}$ swapped. Thus, we get

$$\sup_{p \in \Delta^\infty} |G_U(p) - G_{\hat{U}}(p)| \le \|U - \hat{U}\|_\infty.$$

Combined with (15), we get

$$D_w(p(\hat{U})||p(U)) \le 2\|U - \hat{U}\|_\infty.$$

Substituting $p(U) = d\mu_{\mathsf{NL}}(U)/d\mu_{\mathsf{ref}}$ and $w(p) = \int \mu_{\mathsf{ref}}(dz) p(z) \log p(z)$, we have $D_w(p(\hat{U})||p(U)) = D_{\mathsf{KL}}(\mu_{\mathsf{NL}}(\hat{U})||\mu_{\mathsf{NL}}(U))$ and hence complete the proof.

D.4. Proof of Theorem 10

Let $(\mu_{\mathsf{NL}}(U), \nu_{\mathsf{NL}}(U))$ be the solution to the NLHF with utility $U \in L^\infty(\mu_{\mathsf{ref}} \otimes \mu_{\mathsf{ref}})$ and let $p(U) = \frac{d\mu_{\mathsf{NL}}(U)}{d\mu_{\mathsf{ref}}}$ and $q(U) = \frac{d\mu_{\mathsf{NL}}(U)}{d\mu_{\mathsf{ref}}}$. Note in passing that for any $U, \hat{U}$,

$$q(U) = \operatorname*{argmin}_{q \in \Delta^\infty} \langle q, Up(U) \rangle + w(q), \quad q(\hat{U}) = \operatorname*{argmin}_{q \in \Delta^\infty} \langle q, \hat{U}p(\hat{U}) \rangle + w(q).$$

Now we define an intermediate $\tilde{q}$ as

$$\tilde{q} = \operatorname*{argmin}_{q \in \Delta^\infty} \langle q, \hat{U}p(U) \rangle + w(q)$$

and decompose $\|q(U) - q(\hat{U})\|_1$ as

$$\|q(U) - q(\hat{U})\|_1 \le \|q(U) - \tilde{q}\|_1 + \|\tilde{q} - q(\hat{U})\|_1.$$

For the first term, we may apply Lemma 16 with $g = Up(U)$ and $\tilde{g} = \hat{U}p(U)$ and obtain

$$\|q(U) - \tilde{q}\|_1 \le \|Up(U) - \hat{U}p(U)\|_\infty \le \|U - \hat{U}\|_\infty \|p(U)\|_1 = \|U - \hat{U}\|_\infty.$$

For the second term, we apply Lemma 20-(1) and obtain

$$\|\tilde{q} - q(\hat{U})\|_1 \le \|\hat{U}\|_\oplus \|p(U) - p(\hat{U})\|_1.$$

Combining the above displays, we get

$$\|q(U) - q(\hat{U})\|_1 \le \|U - \hat{U}\|_\infty + \|\hat{U}\|_\oplus \|p(U) - p(\hat{U})\|_1.$$

Here, swapping U and $\hat{U}$, the above inequality also holds with $\|\hat{U}\|_\oplus$ replaced by $\|U\|_\oplus$. Thus, we get

$$\|q(U) - q(\hat{U})\|_1 \le \|U - \hat{U}\|_\infty + (\|U\|_\oplus \wedge \|\hat{U}\|_\oplus) \|p(U) - p(\hat{U})\|_1.$$

Applying the same argument to the pair $(p(U), p(\hat{U}))$ which satisfy

$$p(U) = \operatorname*{argmin}_{p \in \Delta^\infty} \langle p, -U^\star q(U) \rangle + w(p), \quad p(\hat{U}) = \operatorname*{argmin}_{p \in \Delta^\infty} \langle p, -\hat{U}^\star q(\hat{U}) \rangle + w(p),$$

noting $\| - U^\star\|_\oplus = \|U\|_\oplus$, it also holds that

$$\|p(U) - p(\hat{U})\|_1 \le \|U - \hat{U}\|_\infty + (\|U\|_\oplus \wedge \|\hat{U}\|_\oplus) \|q(U) - q(\hat{U})\|_1.$$

Combining the above two displays,

$$\|p(U) - p(\hat{U})\|_1 \le \|U - \hat{U}\|_\infty + (\|U\|_\oplus \wedge \|\hat{U}\|_\oplus) \Big(\|U - \hat{U}\|_\infty + (\|U\|_\oplus \wedge \|\hat{U}\|_\oplus) \|p(U) - p(\hat{U})\|_1 \Big).$$

Rearranging this,

$$\Big(1 - (\|U\|_\oplus \wedge \|\hat{U}\|_\oplus)^2\Big) \|p(U) - p(\hat{U})\|_1 \le (1 + \|U\|_\oplus \wedge \|\hat{U}\|_\oplus) \|U - \hat{U}\|_\infty.$$

By the assumption $\|U\|_\oplus \wedge \|\hat{U}\|_\oplus < 1$, we may divide by $(1 - (\|U\|_\oplus \wedge \|\hat{U}\|_\oplus)^2) = (1 - (\|U\|_\oplus \wedge \|\hat{U}\|_\oplus))(1 + (\|U\|_\oplus \wedge \|\hat{U}\|_\oplus))$ and complete the proof.

D.5. Proof of Theorem 11

We denote $\|\cdot\|_p = \|\cdot\|_{L^p(\mu_{\mathsf{ref}})}$ and $\|U\|_\infty = \|U\|_{L^\infty(\mu_{\mathsf{ref}}\otimes\mu_{\mathsf{ref}})}$ for simplicity. Now for each $t \in [0,1]$, let μ_t be the solution to the following minimax problem

$$\mu_t \in \operatorname*{argmax}_\mu \min_\nu \Big(\mathbb{E}_{y\sim\mu, x\sim\nu}\big[tU(x,y) + (1-t)\hat{U}(x,y)\big] - D_{\mathsf{KL}}(\mu|\mu_{\mathsf{ref}}) + D_{\mathsf{KL}}(\nu|\mu_{\mathsf{ref}}) \Big).$$

Our goal is to bound $d_{\mathsf{TV}}(\mu_1, \mu_0)$. Now we claim that the density $p_t(y) \equiv \frac{\mu_t(dy)}{\mu_{\mathsf{ref}}(dy)}$ is differentiable in t in the sense that there exists $\dot{p}_t \in L^1$ such that

$$\lim_{\epsilon\to 0} \|\frac{p_{t+\epsilon} - p_t}{\epsilon} - \dot{p}_t\|_1 = 0,$$

and there exists an absolute constant C such that the L1 norm of the derivative is bounded as

$$\|\dot{p}_t\|_1 = \int \mu_{\mathsf{ref}}(dy)|\dot{p}_t(y)| \le C \cdot (1 + \|tU + (1-t)\hat{U}\|_\oplus^6) \cdot \|U - \hat{U}\|_\infty \tag{16}$$

for all $t \in [0,1]$. Then, by Fubini's theorem,

$$\begin{aligned}
d_{\mathsf{TV}}(\mu_1, \mu_0) &= \frac{1}{2}\int \mu_{\mathsf{ref}}(dy)\Big|p_1(y) - p_0(y)\Big| \\
&= \frac{1}{2}\int \mu_{\mathsf{ref}}(dy)\Big|\int_0^1 \dot{p}_t(y)dt\Big| \\
&\le \frac{1}{2}\int \mu_{\mathsf{ref}}(dy)\int_0^1 |\dot{p}_t(y)|dt \\
&\le \frac{1}{2}\int_0^1 dt \int \mu_{\mathsf{ref}}(dy)|\dot{p}_t(y)| \\
&\le \frac{C}{2}\|U - \hat{U}\|_\infty \int_0^1 dt(1 + \|tU + (1-t)\hat{U}\|_\oplus^6) \qquad \text{by (16)}
\end{aligned}$$

Here, by the triangle inequality for the semi-norm $\|\cdot\|_\oplus$, one can show $\int_0^1 dt\|tU + (1-t)\hat{U}\|_\oplus^6 \le C'(\|U\|_\oplus^6 + \|\hat{U}\|_\oplus^6)$ for an absolute constant C'. Then we get

$$d_{\mathsf{TV}}(p_{\mathsf{NL}}(U), p_{\mathsf{NL}}(\hat{U})) \le 2^{-1}CC' \cdot \|U - \hat{U}\|_\infty \cdot \Big(1 + \|\hat{U}\|_\oplus^6 + \|U\|_\oplus^6\Big).$$

Since $2^{-1}CC'$ is an absolute constant independent of all other quantities, the proof of Theorem 11 is complete. Thus, the rest of the goal is to show (16). Notice that it suffices to prove the following lemma:

Lemma 21 *Fix $U, E \in L^\infty(\mu_{\mathsf{ref}} \otimes \mu_{\mathsf{ref}})$. For any $\epsilon \in \mathbb{R}$, let μ_ϵ be the solution to the following minimax problem*

$$\mu_\epsilon \in \operatorname*{argmax}_\mu \min_\nu \Big(\mathbb{E}_{y\sim\mu, x\sim\nu}\big[U(x,y) + \epsilon E(x,y)\big] - D_{\mathsf{KL}}(\mu|\mu_{\mathsf{ref}}) + D_{\mathsf{KL}}(\nu|\mu_{\mathsf{ref}}) \Big). \tag{17}$$

Then, the density $p_\epsilon(y) = \frac{p_\epsilon}{\mu_{\mathsf{ref}}}(dy)$ is differentiable at $\epsilon = 0$ in the sense of

$$\lim_{\epsilon\to 0} \|\frac{p_\epsilon - p_0}{\epsilon} - \dot{p}_0\|_1 = 0,$$

where the norm of its derivative bounded as

$$\int |\dot{p}_0(y)|\mu_{\mathsf{ref}}(dy) \le C \cdot (1 + \|U\|_\infty^6) \cdot \|E\|_\infty$$

where C is an absolute constant.

Indeed, if the above lemma holds, then replacing U by $tU + (1-t)U$ and E by $U - \hat{U}$, we obtain (16). Thus, it suffices to show Lemma 21.

D.5.1. PROOF OF LEMMA 21

Notation We first fix notation. Let $L^2 = L^2(\mu_{\mathsf{ref}})$ be the Hilbert space with usual inner product $\langle f, g\rangle = \int f(z)g(z)\mu_{\mathsf{ref}}(dz)$. For any $U \in L^\infty(\mu_{\mathsf{ref}} \otimes \mu_{\mathsf{ref}})$, let us write $\|U\|_\infty = \|U\|_{L^\infty(\mu_{\mathsf{ref}} \otimes \mu_{\mathsf{ref}})}$. Now we define the linear operators U as

$$U : L^2 \to L^\infty, \qquad f \mapsto Uf(\cdot) = \int U(\cdot, y)f(y)\mu_{\mathsf{ref}}(dy)$$

and let $U^\star$ be its adjoint operator such that $\langle g, Uf\rangle = \langle U^\star g, f\rangle$ for all $f, g \in L^2$. Note that U is a bounded operator with operator norm bounded as

$$\|U\|_{L^2 \to L^\infty} = \sup_{\|f\|_2 = 1} \|Uf\|_\infty \le \sup_{\|f\|_2=1} \|U\|_\infty \|f\|_1 \le \sup_{\|f\|_2=1} \|U\|_\infty \|f\|_2 = \|U\|_\infty.$$

Next, let $\Delta^\infty \subset L^\infty$ be the density simplex defined as

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

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 $\mathcal{G}$ taking value in Δ^∞ as

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

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 $L^2_0 = \{h \in L^2 : \int \mu_{\mathsf{ref}}(dz)h(z) = 0\} \subset L^2$. Notice that J_p is bounded, self-adjoint, and positive semidefinite operator on L^2. Indeed, 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

$$\langle f, J_p g\rangle = \mathrm{Cov}_{Z\sim\mu_p}(f(Z), g(Z))$$

from which we conclude that 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

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

where the last inequality follows from $\|f\|_{L^2(\mu_p)} \le \|\frac{d\mu_p}{d\mu_{\mathsf{ref}}}\|^{1/2}_{L^\infty(\mu_{\mathsf{ref}})} \cdot \|f\|_{L^2(\mu_{\mathsf{ref}})} = \|p\|^{1/2}_\infty \|f\|_\infty$ for any $f \in L^2$. 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 $J_p^{1/2}$.

Finally, the two operators $\mathcal{G}, J_p$ are related via Fréchet derivative: $D_f\mathcal{G} = J_{\mathcal{G}(f)}$. That is, for any $f, \delta f \in L^\infty$, by the quotient rule, it holds that

$$\mathcal{G}(f + \delta f) = \mathcal{G}(f) + J_{\mathcal{G}(f)}\delta f + o(\|\delta f\|_\infty).$$

We now claim that the density of NLHF, $p(U) = d\mu_{\mathsf{NL}}(U)/d\mu_{\mathsf{ref}}$, is Frèchet differentiable.

Theorem 22 *Fix $U, E \in L^\infty(\mu_{\mathsf{ref}} \otimes \mu_{\mathsf{ref}})$. For any $\epsilon \in \mathbb{R}$, let $(\mu_\epsilon, \nu_\epsilon)$ be the solution to the following minimax problem*

$$\max_\mu \min_\nu \Big(\mathbb{E}_{y\sim\mu, x\sim\nu}\big[U(x,y) + \epsilon E(x,y)\big] - D_{\mathsf{KL}}(\mu|\mu_{\mathsf{ref}}) + D_{\mathsf{KL}}(\nu|\mu_{\mathsf{ref}}) \Big).$$

Let $p_\epsilon(y) = \frac{\mu_\epsilon(dy)}{\mu_{\mathsf{ref}}(dy)}$ and $q_\epsilon(x) = \frac{\nu_\epsilon(dx)}{\mu_{\mathsf{ref}}(dx)}$ be their densities. The map $\epsilon \mapsto (p_\epsilon, q_\epsilon)$ is differentiable at $\epsilon = 0$ in the sense that

$$\lim_{\epsilon\to 0} \|\epsilon^{-1}(p_\epsilon - p_0) - \dot{p}_0\|_2 = \lim_{\epsilon\to 0} \|\epsilon^{-1}(q_\epsilon - q_0) - \dot{q}_0\|_2 = 0$$

where the derivatives are given by

$$\begin{aligned}
\dot{p}_0 &= (I + J_{p_0}U^\star J_{q_0}U)^{-1}(J_{p_0}E^\star q_0 - J_{p_0}U^\star J_{q_0}Ep_0) \\
\dot{q}_0 &= -(I + J_{q_0}UJ_{p_0}U^\star)^{-1}(J_{q_0}Ep_0 + J_{q_0}UJ_{p_0}E^\star q_0)
\end{aligned}$$

where $I + J_{p_0}U^\star J_{q_0}U$ and $I + J_{q_0}UJ_{p_0}U^\star$ have bounded inverses on $L_0^2 = \{h \in L^2 : \int \mu_{\mathsf{ref}}(dz)h(z) = 0\} \subset L^2$.

Proof Let $x_\epsilon = p_\epsilon - 1$ and $y_\epsilon = q_\epsilon - 1$ so that $x_\epsilon, y_\epsilon \in L_0^2$. We then define $F : \mathbb{R} \times L_0^2 \times L_0^2 \to L_0^2 \times L_0^2$ as the residual of the fixed point equation:

$$F(\epsilon, x, y) = \begin{pmatrix} F_1(\epsilon, x, y) \\ F_2(\epsilon, x, y) \end{pmatrix} \equiv \begin{pmatrix} (x+1) - \mathcal{G}((U + \epsilon E)^\star(y+1)) \\ (y+1) - \mathcal{G}(-(U+\epsilon E)(x+1)) \end{pmatrix}$$

so that $F(\epsilon, x_\epsilon, y_\epsilon) = 0$ for all ϵ. Our goal is to show the differentiability of x_ϵ (as a map from $\mathbb{R}$ to L^2) via the Implicit function theorem on the subset L_0^2 of Hilbert spaces L^2. First, note that F indeed takes values in $L_0^2 \times L_0^2$ since $\mathcal{G}$ maps from L^∞ to the density simplex Δ^∞. Thus $F : \mathbb{R} \times L_0^2 \times L_0^2 \to L_0^2 \times L_0^2$ is well-defined.

Let $U_\epsilon = U + \epsilon E$. We next compute the derivative of F with respect to (x, y). For perturbations $(\delta x, \delta y) \in L_0^2 \times L_0^2$, using

$$\mathcal{G}(f + \delta f) = \mathcal{G}(f) + J_{\mathcal{G}(f)}\delta f + o(\|\delta f\|_\infty)$$

and the bound $\|U_\epsilon \delta x\|_\infty \le \|U_\epsilon\|_\infty \|\delta x\|_1 \le \|U_\epsilon\|_\infty \|\delta x\|_2$, we obtain

$$\begin{aligned}
F_1(\epsilon, x + \delta x, y + \delta y) &= F_1(\epsilon, x, y) + \delta x - J_{\mathcal{G}(U_\epsilon^\star(y+1))}U_\epsilon^\star \delta y + o(\|\delta x\|_2 + \|\delta y\|_2), \\
F_2(\epsilon, x + \delta x, y + \delta y) &= F_2(\epsilon, x, y) + \delta y + J_{\mathcal{G}(-U_\epsilon(x+1))}U_\epsilon \delta x + o(\|\delta x\|_2 + \|\delta y\|_2).
\end{aligned}$$

In particular, at the solution (x_ϵ, y_ϵ), where $p_\epsilon = x_\epsilon + 1$ and $q_\epsilon = y_\epsilon + 1$, we have $p_\epsilon = \mathcal{G}(U_\epsilon^\star q_\epsilon)$ and $q_\epsilon = \mathcal{G}(-U_\epsilon p_\epsilon)$, and hence

$$D_{x,y}F(\epsilon, x_\epsilon, y_\epsilon) = \begin{bmatrix} I & -J_{p_\epsilon}U_\epsilon^\star \\ J_{q_\epsilon}U_\epsilon & I \end{bmatrix}.$$

Here the derivative is understood as a bounded linear operator from $L_0^2 \times L_0^2$ to $L_0^2 \times L_0^2$ since J_{p_0}, J_{q_0} map L^2 to L_0^2.

We now show that $D_{x,y}F(0, x_0, y_0)$ is an isomorphism on $L_0^2 \times L_0^2$. Let $(r_1, r_2) \in L_0^2 \times L_0^2$. Solving

$$\begin{bmatrix} I & -J_{p_0}U^\star \\ J_{q_0}U & I \end{bmatrix} \begin{pmatrix} \delta x \\ \delta y \end{pmatrix} = \begin{pmatrix} r_1 \\ r_2 \end{pmatrix}$$

gives

$$\delta y = r_2 - J_{q_0}U\delta x.$$

Substituting this into the first equation yields the Schur complement equation

$$\big(I + J_{p_0}U^\star J_{q_0}U\big)\delta x = r_1 + J_{p_0}U^\star r_2.$$

Thus it is enough to show that $(I + J_{p_0}U^\star J_{q_0}U)$ has a bounded inverse on L_0^2. Indeed, once δx is obtained from the Schur complement equation, δy is given by $\delta y = r_2 - J_{q_0}U\delta x$, and the boundedness of the inverse follows from the boundedness of $(I + J_{p_0}U^\star J_{q_0}U)^{-1}$, $J_{p_0}U^\star$, and $J_{q_0}U$.

On L_0^2, write

$$I + J_{p_0}U^\star J_{q_0}U = I + AB, \qquad A = J_{p_0}^{1/2}, \qquad B = J_{p_0}^{1/2}U^\star J_{q_0}U.$$

The operators A and B are bounded on L_0^2. Moreover,

$$BA = J_{p_0}^{1/2}U^\star J_{q_0}U J_{p_0}^{1/2}$$

is self-adjoint and positive semidefinite on L_0^2 since J_{q_0} and J_{p_0} are self-adjoint positive semidefinite. Hence

$$\langle h, (I + BA)h\rangle_{L^2} = \|h\|_2^2 + \langle h, BAh\rangle_{L^2} \geq \|h\|_2^2.$$

By the Lax–Milgram theorem, $I + BA$ has a bounded inverse on L_0^2, with

$$\|(I + BA)^{-1}\|_{L_0^2 \to L_0^2} \leq 1.$$

Therefore $I + AB$ is also invertible on L_0^2, with inverse

$$(I + AB)^{-1} = I - A(I + BA)^{-1}B.$$

Consequently, $I + J_{p_0}U^\star J_{q_0}U$ has a bounded inverse on L_0^2, and hence

$$D_{x,y}F(0, x_0, y_0) = \begin{bmatrix} I & -J_{p_0}U^\star \\ J_{q_0}U & I \end{bmatrix}$$

has a bounded inverse on $L_0^2 \times L_0^2$.

We may therefore apply the Implicit Function Theorem on the Hilbert space $L_0^2 \times L_0^2$. Since $F(\epsilon, x_\epsilon, y_\epsilon) = 0$, the map $\epsilon \mapsto (x_\epsilon, y_\epsilon)$ is differentiable at $\epsilon = 0$ as a map into $L_0^2 \times L_0^2$. Since $p_\epsilon = 1 + x_\epsilon$ and $q_\epsilon = 1 + y_\epsilon$, this is equivalent to the differentiability of $\epsilon \mapsto (p_\epsilon, q_\epsilon)$ in L^2.

It remains to compute the derivative. Differentiating $F(\epsilon, x_\epsilon, y_\epsilon) = 0$ at $\epsilon = 0$, we obtain

$$D_{x,y}F(0, x_0, y_0)\begin{pmatrix}\dot{x}_0\\ \dot{y}_0\end{pmatrix} + \partial_\epsilon F(0, x_0, y_0) = \begin{pmatrix}0\\ 0\end{pmatrix}.$$

Since $\dot{p}_0 = \dot{x}_0$ and $\dot{q}_0 = \dot{y}_0$, it is enough to compute the system with $(\dot{x}_0, \dot{y}_0)$ replaced by $(\dot{p}_0, \dot{q}_0)$. The partial derivative with respect to ϵ is

$$\partial_\epsilon F_1(0, x_0, y_0) = -J_{p_0}E^\star q_0, \qquad \partial_\epsilon F_2(0, x_0, y_0) = J_{q_0}Ep_0.$$

Therefore

$$\begin{bmatrix} I & -J_{p_0}U^\star \\ J_{q_0}U & I\end{bmatrix}\begin{pmatrix}\dot{p}_0\\ \dot{q}_0\end{pmatrix} = \begin{pmatrix}J_{p_0}E^\star q_0\\ -J_{q_0}Ep_0\end{pmatrix}.$$

Equivalently,

$$\begin{aligned}\dot{p}_0 - J_{p_0}U^\star\dot{q}_0 &= J_{p_0}E^\star q_0,\\ J_{q_0}U\dot{p}_0 + \dot{q}_0 &= -J_{q_0}Ep_0.\end{aligned}$$

Substituting $\dot{q}_0 = -J_{q_0}U\dot{p}_0 - J_{q_0}Ep_0$ into the first equation gives

$$\big(I + J_{p_0}U^\star J_{q_0}U\big)\dot{p}_0 = J_{p_0}E^\star q_0 - J_{p_0}U^\star J_{q_0}Ep_0.$$

Hence

$$\dot{p}_0 = \big(I + J_{p_0}U^\star J_{q_0}U\big)^{-1}\big(J_{p_0}E^\star q_0 - J_{p_0}U^\star J_{q_0}Ep_0\big).$$

Similarly, eliminating $\dot{p}_0$ instead gives

$$\dot{q}_0 = -\big(I + J_{q_0}UJ_{p_0}U^\star\big)^{-1}\big(J_{q_0}Ep_0 + J_{q_0}UJ_{p_0}E^\star q_0\big).$$

Here, the bounded invertibility of $I + J_{q_0}UJ_{p_0}U^\star$ on L_0^2 follows from the same argument as $I + J_{p_0}U^\star J_{q_0}U$ above. This completes the proof. ∎

We have shown that the derivative is given by

$$\dot{p}_0 = (I + J_{p_0}U^\star J_{q_0}U)^{-1}(J_{p_0}E^\star q_0 - J_{p_0}U^\star J_{q_0}Ep_0)$$

in the sense of $\lim_{\epsilon\to 0}\|\epsilon^{-1}(p_\epsilon - p_0) - \dot{p}_0\|_2 = 0$. The rest of our goal is to bound the L_1 norm of $\dot{p}_0$.

Lemma 23 *For any $U \in L^\infty(\mu_{\mathsf{ref}} \times \mu_{\mathsf{ref}})$, $p \in \Delta^\infty$, and $f \in L^2$, it holds that*

$$\forall f \in L^2, \quad \|J_pUf\|_1 \le \|U\|_\infty\|f\|_1$$

Proof Letting $\mu(dz) = p(z)\mu_{\mathsf{ref}}(dz)$ be the tilted measure induced by $p \in \Delta^\infty$, for any $f \in L^2$,

$$\begin{aligned}
\|J_p U f\|_1 &= \int \mu_{\mathsf{ref}}(dz)\Big|p(z)Uf(z) - p(z)\int \mu_{\mathsf{ref}}(dz')p(z')Uf(z')\Big| \\
&= \int \mu(dz)\Big|Uf(z) - \int \mu(dz')Uf(z')\Big| \\
&\le \|Uf\|_{L^\infty(\mu)} && \mathbb{E}[|X - |\,\mathbb{E}[X]|] \le \|X\|_\infty \\
&\le \|Uf\|_{L^\infty(\mu_{\mathsf{ref}})} && \mu \ll \mu_{\mathsf{ref}} \\
&\le \|U\|_\infty \|f\|_1,
\end{aligned}$$

where the last inequality follows from the Cauchy–Schwarz inequality. ■

Lemma 24 *For any $U \in L^\infty(\mu_{\mathsf{ref}} \otimes \mu_{\mathsf{ref}})$, $g \in L^2$, and $p, q \in \Delta^\infty$,*

$$\|(I + J_p U^\star J_q U)^{-1} g\|_1 \le \Big(\frac{21}{8}(1 + \|U\|_\infty^2)^2\|U\|_\infty + 1 + \frac{1}{2}\|U\|_\infty\Big)\|g\|_1.$$

Proof See Section D.5.2. ■

Let us finish the proof of Lemma 21. We fix $f, g \in L^\infty$ and let $U_{g,f} = U - g \oplus f$. For any $h \in L^2_0$, using $J_{q_0}\mathbf{1} = 0$ and $\langle \mathbf{1}, h\rangle = 0$, we have

$$J_{q_0} U h = J_{q_0}(U - g \oplus f)h + J_{q_0}\mathbf{1}\langle f, h\rangle + J_{q_0} g\langle 1, h\rangle = J_{q_0} U_{g,f} h$$

and hence $J_{q_0}U$ operates on L^2_0 as

$$J_{q_0} U|_{L^2_0} = J_{q_0} U_{g,f}$$

By the same argument, we have

$$J_{p_0} U^\star|_{L^2_0} = J_{p_0} U^\star_{g,f}$$

Thus, we can rewrite $\dot{p}_0$ as

$$\begin{aligned}
\dot{p}_0 &= (I + J_{p_0}U^\star J_{q_0} U)^{-1}(J_{p_0}E^\star q_0 - J_{p_0}U^\star J_{q_0} E p_0) \\
&= \big(I + J_{p_0}U^\star_{g,f} J_{q_0} U_{g,f}\big)^{-1}\big(J_{p_0}E^\star q_0 - J_{p_0}U^\star_{g,f} J_{q_0} E p_0\big).
\end{aligned}$$

From Lemma 20, noting $\|p_0\|_1 = \|q_0\|_1 = 1$, we get

$$\begin{aligned}
\big\|J_{p_0}E^\star q_0 - J_{p_0}U^\star_{g,f} J_{q_0} E p_0\big\|_1 &\le \|J_{p_0}E^\star q_0\|_1 + \|J_{p_0}U^\star_{g,f} J_{q_0} E p_0\|_1 \\
&\le \|E\|_\infty\|q_0\|_1 + \|U_{g,f}\|_\infty\|E\|_\infty\|p_0\|_1 \\
&= (1 + \|U_{g,f}\|_\infty)\|E\|_\infty.
\end{aligned}$$

Combined with Lemma 24, we have

$$\|\dot{p}_0\|_1 \le \Big(\frac{21}{8}(1 + \|U_{g,f}\|_\infty^2)^2\|U_{g,f}\|_\infty + 1 + \frac{1}{2}\|U_{g,f}\|_\infty\Big)(1 + \|U_{g,f}\|_\infty)\|E\|_\infty.$$

Here, expanding the RHS, we can find an absolute constant C such that

$$\|\dot{p}_0\|_1 \le C(1 + \|U_{g,f}\|_\infty^6)\|E\|_\infty.$$

Since $g, f \in L^\infty$ 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 Lemma 21.

D.5.2. 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) satisfy

$$\begin{aligned} f + J_p U^\star h &= g \\ -J_q U f + h &= 0. \end{aligned}$$

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 on L^2 defined as

$$D_{\sqrt{p}} f(z) = \sqrt{p(z)} f(z), \quad pp^\star f(z) = p(z) \cdot \langle p, f \rangle.$$

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

$$\begin{aligned} f + D_{\sqrt{p}} u_h - p \cdot c_h &= g \\ -D_{\sqrt{q}} u_f + q \cdot c_f + h &= 0 \end{aligned} \tag{18}$$

where we defined the scalars c_h, c_f as

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

Here, from the first equation, taking L1 norm $\|\cdot\|_1$ and using triangle inequality, we see

$$\|f\|_1 \le \|D_{\sqrt{p}} u_h\|_1 + c_h \|p\|_1 + \|g\|_1 \le \|u_h\|_2 + |c_h| + \|g\|_1 \tag{19}$$

where the second inequality follows from $\|p\|_1 = 1$ and the Cauchy–Schwarz inequality applied to $\|D_{\sqrt{p}} u_h\|_1 = \int \mu_{\mathsf{ref}}(dz) |\sqrt{p}(x) u_h(x)| \le \|p\|_1^{1/2} \|u_h\|_2$. From this result, we see that it suffices to bound the $\|u_h\|_2$ and $|c_h| = |\langle \sqrt{p}, u_h \rangle|$.

Now, multiplying the first equation in (18) 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

$$\begin{aligned} u_f + A u_h - c_h A \sqrt{p} &= D_{\sqrt{q}} U g \\ -A^\star u_f + u_h + c_f A^\star \sqrt{q} &= 0. \end{aligned} \tag{20}$$

Here A is a bounded linear operator on L^2 with its operator norm bounded by $\|U\|_\infty$. Indeed, using the Cauchy–Schwarz inequality and $p, q \in \Delta^\infty$, for any $f \in L^\infty$ with $\|f\|_2 \le 1$,

$$\begin{aligned} \|Af\|_2^2 &= \|D_{\sqrt{q}} U D_{\sqrt{p}} f\|_2^2 \\ &= \int \mu_{\mathsf{ref}}(dx) \Big(\sqrt{q(x)} \int \mu_{\mathsf{ref}}(dy) U(x,y) f(y) \sqrt{p(y)} \Big)^2 \\ &\le \|q\|_1 \cdot \operatorname*{esssup}_x \Big(\int \mu_{\mathsf{ref}}(dy) U(x,y) f(y) \sqrt{p(y)} \Big)^2 \\ &\le \|q\|_1 \cdot \|U\|_\infty^2 \Big(\int \mu_{\mathsf{ref}}(dy) |f(y)| \sqrt{p(y)} \Big)^2 \\ &\le \|q\|_1 \cdot \|U\|_\infty^2 \cdot \|f\|_2^2 \cdot \|p\|_1 \\ &= \|U\|_\infty^2 \qquad \text{by } \|q\|_1 = \|p\|_1 = 1 \text{ and } \|f\|_2 \le 1 \end{aligned}$$

Now we solve the system for (u_f, u_h) and (c_f, c_h). Substituting the second equation $u_h = A^\star u_f - c_f A^\star \sqrt{q}$ in (20) to the first equation, rearranging it, we get

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

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

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

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

$$\forall f \in L^2, \quad (1 + \|U\|^2_\infty)^{-1} \|f\|^2_2 \le \langle f, Tf \rangle \le \|f\|^2_2. \tag{21}$$

Then, u_f can be solved as

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

where we used $I - T = TA^\star$ and $TAA^\star = 1 - T$, 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

$$\forall f \in L^2, \quad (1 + \|U\|^2_\infty)^{-1} \|f\|^2_2 \le \langle f, Sf \rangle \le \|f\|^2_2 \tag{22}$$

and u_f can be solved explicitly as

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

Putting all together, we get

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

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

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

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

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

Since $\|\sqrt{p}\|_2 = \|\sqrt{q}\|_2 = 1$, using the established bound of the spectrum of S and T (21)-(22), we know

$$\frac{1}{1 + \|U\|_\infty^2} \le \alpha \le 1, \quad \frac{1}{1 + \|U\|_\infty^2} \le \beta \le 1, \quad |\gamma| \le \frac{1}{2}. \tag{23}$$

Here $|\gamma| \le 1/2$ follows from the fact that $\|TA\|_{L^2 \to L^2} = \sqrt{\|(TA)(TA)^\star\|_{L^2 \to L^2}}$ where $(TA)(TA)^\star = AA^\star(I + AA^\star)^{-2}$, whose spectrum is given by $\sigma(1+\sigma)^{-2}$ with $\sigma \in [0, +\infty)$ where σ is the spectrum of $AA^\star$. Since $\inf_{x \ge 0} x/(1+x)^2 = 1/4$, taking square root, we get $\|TA\|_{L^2 \to L^2} \le 1/2$. By the same argument, $\|SA^\star\|_{L^2 \to L^2} \le 1/2$ holds.

Now, using the lower bound of α and β in (23), the determinant of the matrix is strictly positive:

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

Therefore, that matrix is invertible, and we obtain

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

Taking the absolute value and using the estimate of $|\alpha|, |\beta|, |\gamma|$ from (23), with $|\langle \sqrt{q}, TD_{\sqrt{p}} Ug \rangle| \le \|D_{\sqrt{p}} Ug\|_2$ and $|\langle \sqrt{p}, SA^\star D_{\sqrt{q}} Ug \rangle| \le 2^{-1} \|D_{\sqrt{q}} Ug\|_2$, the following inequality holds for each coordinate:

$$\begin{bmatrix} |c_f| \\ |c_h| \end{bmatrix} \le (1 + \|U\|_\infty^2)^2 \begin{bmatrix} 1 & 2^{-1} \\ 2^{-1} & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 2^{-1} \end{bmatrix} \cdot \|D_{\sqrt{p}} Ug\|_2 = (1 + \|U\|_\infty^2)^2 \|D_{\sqrt{p}} Ug\|_2 \begin{bmatrix} 5/4 \\ 1 \end{bmatrix}$$

With $\|p\|_1 = 1$, we have

$$\begin{aligned} \|D_{\sqrt{p}} Ug\|_2^2 &= \int \mu_{\mathsf{ref}}(dx) \big(\sqrt{p}(x) \int U(x,y) g(y) \mu_{\mathsf{ref}}(dy) \big)^2 \\ &\le \|p\|_1 \cdot \|U\|_\infty^2 \cdot \|g\|_1^2 \\ &= (\|U\|_\infty \|g\|_1)^2, \end{aligned}$$

so we get

$$\begin{bmatrix} |c_f| \\ |c_h| \end{bmatrix} \le (1 + \|U\|_\infty^2)^2 \|U\|_\infty \|g\|_1 \begin{bmatrix} 5/4 \\ 1 \end{bmatrix}. \tag{24}$$

Let us finish the proof. By the second equation of (20),

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

Taking $\|\cdot\|_2$, using $\|I - S\|_{L^2 \to L^2} \le 1$, $\|SA^\star\|_{L^2 \to L^2} \le 1/2$ and $\|D_{\sqrt{p}} U g\|_2 \le \|U\|_\infty \|g\|_1$, we obtain

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

Combined with the upper bound $\|f\|_1 \le \|u_h\|_2 + |c_h| + \|g\|_1$ from (19),

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

Finally, substituting the upper bounds of $[|c_f|, |c_h|]$ in (24), with $[2^{-1}, 2] \cdot [5/4, 1]^\top = 21/8$, we complete the proof.

Appendix E. Dynamical analysis for general utility

We provide dynamical analysis of MCHF and NLHF for general U. Note that Theorem 5 and Theorem 6 in the main document hold as corollaries of the general theorem.

For any $U \in L^\infty(\mu_{\mathsf{ref}} \otimes \mu_{\mathsf{ref}})$, we define the linear operators $U, U^\star : L^1(\mu_{\mathsf{ref}}) \to L^1(\mu_{\mathsf{ref}})$ as

$$Uf(x) = \int \mu_{\mathsf{ref}}(dy) U(x,y) f(y), \quad U^\star f(y) = \int \mu_{\mathsf{ref}}(dx) f(x) U(x,y).$$

For any $p \in L^\infty(\mu_{\mathsf{ref}})$ such that $p(\cdot) \ge 0$ and $\|p\|_{L^1(\mu_{\mathsf{ref}})} = 1$, define the linear operator $J_p : L^1(\mu_{\mathsf{ref}}) \to L^1(\mu_{\mathsf{ref}})$ as $J_p = \mathrm{Diag}(p) + pp^\star$ or more precisely

$$J_p f(x) = p(x) f(x) - p(x) \int \mu_{\mathsf{ref}}(dx') p(x') f(x').$$

For $* \in \{\mathsf{MC}, \mathsf{NL}, \mathsf{RL}\}$, let $p_* = \frac{d\mu_*}{d\mu_{\mathsf{ref}}}$ be the density with respect to μ_{ref}. The next theorem gives a Taylor expansion of p_{MC} and p_{NL} around a given additive utility $g \oplus f$.

Theorem 25 *For any $U \in L^\infty(\mu_{\mathsf{ref}} \otimes \mu_{\mathsf{ref}})$ and $f, g \in L^\infty(\mu_{\mathsf{ref}})$, it holds that*

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

where $\|\cdot\|_\infty = \|\cdot\|_{L^\infty(\mu_{\mathsf{ref}} \otimes \mu_{\mathsf{ref}})}$.

Note that the above theorem holds for any $f, g \in L^\infty(\mu_{\mathsf{ref}})$ and any $U \in L^\infty(\mu_{\mathsf{ref}} \otimes \mu_{\mathsf{ref}})$ (possibly non-antisymmetric). If we take the L2 solution $(\hat{g}, \hat{f}) \in \mathrm{argmin}_{g,f} \|U - g \oplus f\|_2$ for Theorem 25, using $\|U - \hat{f} \oplus \hat{g}\|_\infty \asymp \|U\|_\oplus$ from Proposition 12, we get

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

We see that the MHCH and the NLHF differs in their first order term by $p_{\mathsf{ref}}(\hat{f})$ and $p_{\mathsf{ref}}(-\hat{g})$.

Next, we characterize the dynamics of the algorithm for MCHF.

Theorem 26 *Consider the iteration $\mu^t_{\mathsf{MC}} = \mu^{t-1}_{\mathsf{MC}}\mathsf{P}$ initialize at $\mu^0_{\mathsf{MC}} = \mu_{\mathsf{ref}}$. Then,*

$$\begin{aligned} \|p^1_{\mathsf{MC}} - p_{\mathsf{RL}}(f) - J_{p_{\mathsf{RL}}(f)}(U - g \oplus f)^\star \mathbf{1}\|_1 &= o(\|U - g \oplus f\|_\infty) \\ \forall t \geq 2, \quad \|p^t_{\mathsf{MC}} - p_{\mathsf{RL}}(f) - J_{p_{\mathsf{RL}}(f)}(U - g \oplus f)^\star p_{\mathsf{RL}}(f)\|_1 &= o(\|U - g \oplus f\|_\infty). \end{aligned}$$

Next, we discuss iterative algorithms for NLHF.

Theorem 27 *Consider the following iteration:*

$$\begin{aligned} \mu^t_{\mathsf{NL}} &\in \operatorname*{argmax}_\mu \int \nu^{t-1}_{\mathsf{NL}}(dx)\mu(dy)U(x,y) - D_{\mathsf{KL}}(\mu \mid \mu_{\mathsf{ref}}) \\ \nu^t_{\mathsf{NL}} &\in \operatorname*{argmin}_\nu \int \nu(dx)\mu^t_{\mathsf{NL}}(dy)U(x,y) + D_{\mathsf{KL}}(\nu \mid \mu_{\mathsf{ref}}) \end{aligned} \tag{25}$$

with $\nu^0_{\mathsf{NL}} = \mu_{\mathsf{ref}}$. Then, the densities $p^t_{\mathsf{NL}} = d\mu^t_{\mathsf{NL}}/d\mu_{\mathsf{ref}}$ and $q^t_{\mathsf{NL}} = d\nu^t_{\mathsf{NL}}/d\mu_{\mathsf{ref}}$ satisfy

$$\begin{aligned} \|p^1_{\mathsf{NL}} - p_{\mathsf{RL}}(f) - J_{p_{\mathsf{RL}}(f)}(U - g \oplus f)^\star \mathbf{1}\|_1 &= o(\|U - g \oplus f\|_{L^\infty}) \\ \forall t \geq 1, \quad \|q^t_{\mathsf{NL}} - p_{\mathsf{RL}}(-g) + J_{p_{\mathsf{RL}}(-g)}(U - g \oplus f)p_{\mathsf{RL}}(f)\|_1 &= o(\|U - g \oplus f\|_{L^\infty}) \\ \forall t \geq 2, \quad \|p^t_{\mathsf{NL}} - p_{\mathsf{RL}}(f) - J_{p_{\mathsf{RL}}(f)}(U - g \oplus f)^\star p_{\mathsf{RL}}(-g)\|_1 &= o(\|U - g \oplus f\|_{L^\infty}). \end{aligned}$$

E.1. Proof of Theorem 5

If U is antisymmetric, substituting $(g, f) = (-\hat{f}, \hat{f})$ for Theorem 25, noting that $J_p X p = (\mathrm{Diag}(p) - pp^\star)Xp = \mathrm{Diag}(p)Xp$ for any antisymmetric X due to $p^\star Xp = 0$, and using $\|U - (-\hat{f}) \oplus \hat{f}\|_\infty \asymp \|U\|_\oplus$ from Proposition 13, we complete the proof.

E.2. Proof of Theorem 6

When U is antisymmetric, the update of ν from μ and the update of μ from ν in the definition of iterates (25) can be written as the same mapping. Thus, if we define the new iterate $\tilde{\mu}^t_{\mathsf{NL}}$ as $\tilde{\mu}^{2t-1}_{\mathsf{NL}} = \mu^t_{\mathsf{NL}}$ and $\tilde{\mu}^{2t}_{\mathsf{NL}} = \nu^t_{\mathsf{NL}}$, and substituting $-g = f = \hat{f}$ to Theorem 27, we complete the proof.

E.3. Proof of Theorem 25

E.3.1. MCHF

We write U as

$$U = g \oplus f + \epsilon E, \quad E = \frac{U - g \oplus f}{\|U - g \oplus f\|_\infty}, \quad \epsilon = \|U - g \oplus f\|_\infty$$

and let

$$p_\epsilon \equiv p_{\mathsf{MC}}(g \oplus f + \epsilon E).$$

Notice that $p_0 = p_{\mathsf{MC}}(g \oplus f) = p_{\mathsf{RL}}(f)$. Applying Theorem 17, we have

$$\|p_\epsilon - p_0 - \epsilon \dot{p}_0\|_1 = o(\epsilon)$$

where $\dot{p}_0$ is given by

$$\dot{p}_0 = \Big((I - P_0^\star)\,|_{L^1_0}\Big)^{-1}\left(y \mapsto \int \mu_{\mathsf{ref}}(dx)p_0(x)P_0(x,y)\Big(E(x,y) - \int \mu_{\mathsf{ref}}(dy')P_0(x,y')E(x,y')\Big)\right)$$

where

$$P_0(x,y) = \frac{\exp(g(x)+f(y))}{\int \mu_{\mathsf{ref}}(dy)\exp(g(x)+f(y'))} = \frac{\exp(f(y))}{\int \mu_{\mathsf{ref}}(dy)\exp(f(y'))} = p_{\mathsf{RL}}(f)(y) = p_0(y).$$

Note that $P^\star h = 0$ for any $h \in L_0^1$ since

$$P_0^\star h(y) = \int \mu_{\mathsf{ref}}(dx)h(x)P_0(x,y) = \Big(\int \mu_{\mathsf{ref}}(dx)h(x)\Big)p_0(y) = 0,$$

and hence

$$\Big((I - P_0^\star)\mid_{L_0^1}\Big)^{-1} = I$$

On the other hand, substituting $P_0(x,y) = p_0(y)$, we have

$$\begin{aligned}
&\int \mu_{\mathsf{ref}}(dx)p_0(x)P_0(x,y)\Big(E(x,y) - \int \mu_{\mathsf{ref}}(dy')P_0(x,y')E(x,y')\Big) \\
&= \int \mu_{\mathsf{ref}}(dx)p_0(x)p_0(y)\Big(E(x,y) - \int \mu_{\mathsf{ref}}(dy')p_0(y')E(x,y')\Big) \\
&= p_0(y)\Big(\int \mu_{\mathsf{ref}}(dx)p_0(x)E(x,y) - \iint \mu_{\mathsf{ref}}(dx)\mu_{\mathsf{ref}}(dy')p_0(x)p_0(y')E(x,y')\Big) \\
&= p_0(y)\Big(E^\star p_0(y) - \int \mu_{\mathsf{ref}}(dy')p_0(y')E^\star p_0(y)\Big) \\
&= J_{p_0}E^\star p_0(y).
\end{aligned}$$

Therefore, the derivative p_0 is given by

$$\dot{p}_0 = J_{p_0}E^\star p_0$$

and hence

$$\|p_\epsilon - p_0 - \epsilon J_{p_0}E^\star p_0\|_1 = o(\epsilon).$$

Note that $\epsilon J_{p_0}E^\star p_0 = J_{p_0}(\epsilon E)^\star p_0$ since J_{p_0} and $E^\star$ are linear operators. Therefore, substituting $p_\epsilon = p_{\mathsf{MC}}(U)$, $p_0 = p_{\mathsf{RL}}(f)$, $\epsilon E^\star = (U - g \oplus f)$, and $\epsilon = \|U - g \oplus f\|_\infty$, we complete the proof for MCHF.

E.3.2. NLHF

We write U as

$$U = g \oplus f + \epsilon E, \quad E = \frac{U - g \oplus f}{\|U - g \oplus f\|_\infty}, \quad \epsilon = \|U - g \oplus f\|_\infty$$

and let

$$p_\epsilon = p_{\mathsf{NL}}(g \oplus f + \epsilon E), \quad q_\epsilon = p_{\mathsf{NL}}(g \oplus f + \epsilon E).$$

Note that $p_0 = p_{\mathsf{RL}}(f)$ and $q_0 = p_{\mathsf{RL}}(-g)$. By Theorem 22, we have

$$\|p_\epsilon - p_0 - \epsilon\dot{p}_0\|_2 = o(\epsilon)$$

where the derivative is given by

$$\begin{aligned}\dot{p}_0 &= (I + J_{p_0}(g \oplus f)^\star J_{q_0}(g \oplus f))^{-1}(J_{p_0}E^\star q_0 - J_{p_0}(g \oplus f)^\star J_{q_0}Ep_0) \\ &= (I + J_{p_0}(f \oplus g)J_{q_0}(g \oplus f))^{-1}(J_{p_0}E^\star q_0 - J_{p_0}(f \oplus g)J_{q_0}Ep_0).\end{aligned}$$

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

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

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

$$J_{p_0}(f \oplus g) \mid_{L_0^2} = 0.$$

By the same argument, we also get

$$J_{q_0}(g \oplus f) \mid_{L_0^2} = 0.$$

Since the image space of J_{p_0} and J_{q_0} are both L_0^2, we get

$$\begin{aligned}\Big((I + J_{p_0}(f \oplus g)J_{q_0}(g \oplus f)) \mid_{L_0^2}\Big)^{-1} &= I \\ J_{p_0}E^\star q_0 - J_{p_0}(f \oplus g)J_{q_0}Ep_0 &= J_{p_0}E^\star q_0\end{aligned}$$

and hence

$$\dot{p}_0 = J_{p_0}E^\star q_0.$$

Substituting $p_0 = p_{\mathsf{RL}}(f)$, $q_0 = p_{\mathsf{RL}}(-g)$, $E = (U - g \oplus f)/\epsilon$, and $\epsilon = \|U - g \oplus f\|_\infty$, we get

$$\|p_{\mathsf{NL}}(U) - p_{\mathsf{RL}}(f) - J_{p_{\mathsf{RL}}(f)}(U - g \oplus f)^\star p_{\mathsf{RL}}(-g)\|_2 = o(\|U - g \oplus f\|_\infty).$$

Since $\|\cdot\|_{L^1} \le \|\cdot\|_2$, this completes the proof.

E.4. Proof of Theorem 26

We write U as

$$U = g \oplus f + \epsilon E, \quad E = \frac{U - g \oplus f}{\|U - g \oplus f\|_\infty}, \quad \epsilon = \|U - g \oplus f\|_\infty$$

and let

$$p_\epsilon^t = P_\epsilon^\star p_\epsilon^{t-1}, \quad p_\epsilon^0 = 1$$

where

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

Then it suffices to show

$$\lim_{\epsilon \to 0} \|\epsilon^{-1}(p_\epsilon^t - p_0^t) - \dot{p}_0^t\|_1 = 0$$

where

$$\dot{p}_0^t = \begin{cases} J_{p_{\mathsf{RL}}(f)}(U - g \oplus f)^\star 1 & t = 1 \\ J_{p_{\mathsf{RL}}(f)}(U - g \oplus f)^\star p_{\mathsf{RL}}(f) & t \geq 2 \end{cases}$$

Note that $p_0^t = p_{\mathsf{RL}}(f)$ for all $t \geq 1$ and $P_0(x, y) = p_{\mathsf{RL}}(f)(y)$. Furthermore, by the chain rule, we have

$$\partial_\epsilon P_\epsilon(x, y) = P_\epsilon(x, y)E(x, y) - P_\epsilon(x, y) \int \mu_{\mathsf{ref}}(dy') P_\epsilon(x, y')E(x, y')$$

Taking $\epsilon = 0$,

$$\dot{P}_0(x, y) = p_{\mathsf{RL}}(f)(y)E(x, y) - p_{\mathsf{RL}}(f)(y) \int \mu_{\mathsf{ref}}(dy') p_{\mathsf{RL}}(f)(y')E(x, y')$$

By the chain rule, $p_0^1 = P_0^\star \mathbf{1}$ is differentiable with derivative given by

$$\begin{aligned} \dot{p}_0^1(y) = \dot{P}_0^\star 1(y) &= \int \mu_{\mathsf{ref}}(dx) \Big(p_{\mathsf{RL}}(f)(y)E(x, y) - p_{\mathsf{RL}}(f)(y) \int \mu_{\mathsf{ref}}(dy') p_{\mathsf{RL}}(f)(y')E(x, y') \Big) \\ &= p_{\mathsf{RL}}(f)(y) \int \mu_{\mathsf{ref}}(dx)E(x, y) - p_{\mathsf{RL}}(f)(y) \iint \mu_{\mathsf{ref}}(dx)\mu_{\mathsf{ref}}(dy') p_{\mathsf{RL}}(f)(y')E(x, y') \\ &= p_{\mathsf{RL}}(f)(y)E^\star 1(y) - p_{\mathsf{RL}}(f)(y) \int \mu_{\mathsf{ref}}(dy') p_{\mathsf{RL}}(f)(y')E^\star 1(y') \\ &= J_{p_{\mathsf{RL}}(f)}(U - g \oplus f)^\star 1. \end{aligned}$$

For $t = 2$, by the chain rule, p_ϵ^2 is differentiable at $\epsilon = 0$ with derivative given by

$$\dot{p}_0^2 = \dot{P}_0^\star p_0^1(y) + P_0^\star \dot{p}_0^1$$

Since $P_0^\star \mid_{L_0^1} = 0$ and $\dot{p}_0^1 \in L_0^1$, we get $\dot{p}_0^2 = \dot{P}_0^\star p_0^1(y)$. By the same calculation for $\dot{p}_0^1$, we get

$$\dot{p}_0^2 = J_{p_{\mathsf{RL}}(f)}(U - g \oplus f)^\star p_0^1.$$

Since $p_0^1 = p_{\mathsf{RL}}(f)$, we obtain

$$\dot{p}_0^2 = J_{p_{\mathsf{RL}}(f)}(U - g \oplus f)^\star p_{\mathsf{RL}}(f).$$

Now we suppose that p_ϵ^t is differentiable at $\epsilon = 0$ with derivative given by $\dot{p}_0^t = J_{p_{\mathsf{RL}}(f)}(U - g \oplus f)^\star p_{\mathsf{RL}}(f)$. Since $\dot{p}_0^t \in L_0^1$ and $p_0^t = p_{\mathsf{RL}}(f)$, by the chain rule, p_ϵ^{t+1} is differentiable with derivative given by

$$\dot{p}_0^{t+1} = \dot{P}_0^\star p_0^t + P_0^\star \dot{p}_0^t = \dot{P}_0^\star p_0^t = J_{p_{\mathsf{RL}}(f)}(U - g \oplus f)^\star p_0^t = J_{p_{\mathsf{RL}}(f)}(U - g \oplus f)^\star p_{\mathsf{RL}}(f).$$

By induction, this completes the proof.

E.5. Proof of Theorem 27

We write U as

$$U = g \oplus f + \epsilon E, \quad E = \frac{U - g \oplus f}{\|U - g \oplus f\|_\infty}, \quad \epsilon = \|U - g \oplus f\|_\infty$$

and let $p_\epsilon = d\mu^t_{\mathsf{NL}}(U)/d\mu_{\mathsf{ref}}$ and $q_\epsilon = d\nu^t_{\mathsf{NL}}(U)/d\mu_{\mathsf{ref}}$ where

$$\mu^t_{\mathsf{NL}} \in \operatorname*{argmax}_\mu \int \nu^{t-1}_{\mathsf{NL}}(dx)\mu(dy)U(x,y) - D_{\mathsf{KL}}(\mu \mid \mu_{\mathsf{ref}})$$
$$\nu^t_{\mathsf{NL}} \in \operatorname*{argmin}_\nu \int \nu(dx)\mu^t_{\mathsf{NL}}(dy)U(x,y) + D_{\mathsf{KL}}(\nu \mid \mu_{\mathsf{ref}})$$

with $\nu^0_{\mathsf{NL}} = \mu_{\mathsf{ref}}$. Note that p^t_ϵ and q^t_ϵ satisfies

$$p^t_\epsilon = \mathcal{G}((g \oplus f + \epsilon E)^\star q^{t-1}_\epsilon), \quad q^t_\epsilon = \mathcal{G}(-(g \oplus f + \epsilon E)p^t_\epsilon)$$

with $q^0_\epsilon = 1$. Noting $p_0 = p_{\mathsf{RL}}(f)$ and $q_0 = p_{\mathsf{ref}}(-g)$, it suffices to show

$$\dot{p}^t_0 = \begin{cases} J_{p_{\mathsf{RL}}(f)}E^\star 1 & t = 1 \\ J_{p_{\mathsf{RL}}(f)}E^\star p_{\mathsf{RL}}(-g) & t \geq 2 \end{cases}$$

and

$$\forall t \geq 1, \quad \dot{q}^t_0 = -J_{q_{\mathsf{RL}}(-g)}Ep_{\mathsf{RL}}(f).$$

Now, by the chain rule, p^1_ϵ is differentiable at ϵ with derivative given by

$$p^1_\epsilon = J_{p^1_\epsilon}E^\star q^0_\epsilon.$$

Taking $\epsilon = 0$, noting $q^0_0 = 1$ and $p^1_0 = p_{\mathsf{ref}}(f)$, we get

$$\dot{p}^1_0 = J_{p_{\mathsf{RL}}(f)}E^\star 1.$$

By the same argument, applying the chain rule for $q^1_\epsilon = \mathcal{G}(-(g \oplus f + \epsilon E)p^1_\epsilon)$, we get

$$\dot{q}^1_\epsilon = J_{q^1_\epsilon}(-Ep^1_\epsilon - (g \oplus f + \epsilon E)\dot{p}^1_\epsilon).$$

Taking $\epsilon = 0$, recalling that $J_q(g \oplus f) \mid_{L^1_0} = 0$ and $p^1_\epsilon \in L^1_0$, we get

$$\dot{q}^1_0 = J_{q^1_0}(-Ep^1_0) = -J_{q^1_0}Ep^1_0 = -J_{p_{\mathsf{RL}}(-g)}Ep_{\mathsf{RL}}(f)$$

Now we assume that $\dot{q}^t_\epsilon$ is differentiable for $t \geq 1$. Then, by the chain rule for $p^{t+1}_\epsilon = \mathcal{G}((g \oplus f + \epsilon E)^\star q^t_\epsilon)$, p^{t+1}_ϵ is also differentiable with derivative given by

$$\dot{p}^{t+1}_\epsilon = J_{p^{t+1}_\epsilon}(E^\star q^t_\epsilon + (g \oplus f + \epsilon E)^\star \dot{q}^t_\epsilon).$$

Taking $\epsilon = 0$, using $J_p(g \oplus f) \mid_{L^1_0} = 0$ and $\dot{q}^t_\epsilon \in L^1_0$, we get

$$\dot{p}^{t+1}_0 = J_{p^{t+1}_0}(E^\star q^t_0 + (g \oplus f)^\star \dot{q}^t_0) = J_{p^{t+1}_0}E^\star q^t_0 = J_{p_{\mathsf{RL}}(f)}E^\star p_{\mathsf{RL}}(-g)$$

Now we further assume that p_0^{t+1} is differentiable. Then applying the chain rule for $q_\epsilon^{t+1} = \mathcal{G}(-(g \oplus f + \epsilon E)p_\epsilon^{t+1})$, we know that q_ϵ^{t+1} is also differentiable with

$$\dot{q}_\epsilon^{t+1} = -J_{q_\epsilon^{t+1}}(Ep_\epsilon^{t+1} + (g \oplus f + \epsilon)\dot{p}_\epsilon^{t+1}).$$

Taking $\epsilon = 0$, using $J_q(g \oplus f) \mid_{L_0^1} = 0$ and $\dot{p}_\epsilon^{t+1} \in L_0^1$, we have

$$\dot{q}_\epsilon^{t+1} = -J_{q_0^{t+1}} E p_0^{t+1} = -J_{p_{\mathsf{RL}}(-g)} E p_{\mathsf{RL}}(f).$$

By induction, this completes the proof.