Leveraging Non-uniformity in First-order Non-convex Optimization

Jincheng Mei [1 2 *] **Yue Gao** [1 *] **Bo Dai** [2] **Csaba Szepesvári** [3 1] **Dale Schuurmans** [2 1]

Abstract

Classical global convergence results for first-order methods rely on uniform smoothness and the Łojasiewicz inequality. Motivated by properties of objective functions that arise in machine learning, we propose a non-uniform refinement of these notions, leading to *Non-uniform Smoothness* (NS) and *Non-uniform Łojasiewicz inequality* (NŁ). The new definitions inspire new geometry-aware first-order methods that are able to converge to global optimality faster than the classical $\Omega(1/t^2)$ lower bounds. To illustrate the power of these geometry-aware methods and their corresponding non-uniform analysis, we consider two important problems in machine learning: policy gradient optimization in reinforcement learning (PG), and generalized linear model training in supervised learning (GLM). For PG, we find that normalizing the gradient ascent method can accelerate convergence to $O(e^{-c\cdot t})$ (where $c > 0$) while incurring less overhead than existing algorithms. For GLM, we show that geometry-aware normalized gradient descent can also achieve a linear convergence rate, which significantly improves the best known results. We additionally show that the proposed geometry-aware gradient descent methods escape landscape plateaus faster than standard gradient descent. Experimental results are used to illustrate and complement the theoretical findings.

1. Introduction

The optimization of non-convex objective functions is a topic of key interest in modern-day machine learning. Recent, intriguing results show that simple gradient-based optimization can achieve *globally* optimal solutions in certain non-convex problems arising in machine learning, such as in reinforcement learning (RL) (Agarwal et al., 2020), supervised learning (SL) (Hazan et al., 2015), and deep learning (Allen-Zhu et al., 2019). While gradient-based algorithms remain the method of choice in machine learning, the convergence of such algorithms to global minimizers has still only been established in restrictive settings where one can assert two strong assumptions about the objective function: (1) that the objective is smooth, and (2) that the objective satisfies a gradient dominance over sub-optimality such as the Łojasiewicz inequality. We will find it beneficial to recall the definitions of these properties. For the remainder of this paper let $\Theta = \mathbb{R}^d$.

Definition 1 (Smoothness)**.** *The function $f : \Theta \to \mathbb{R}$ is β-smooth ($\beta > 0$) if it is differentiable and for all $\theta, \theta' \in \Theta$,*

$$\left| f(\theta') - f(\theta) - \left\langle \frac{df(\theta)}{d\theta}, \theta' - \theta \right\rangle \right| \leq \frac{\beta}{2} \cdot \|\theta' - \theta\|_2^2. \quad (1)$$

Definition 2. *(Łojasiewicz, 1963; Polyak, 1963; Kurdyka, 1998) The differentiable function $f : \Theta \to \mathbb{R}$ satisfies the (C, ξ)-Łojasiewicz inequality if for all $\theta \in \Theta$,*

$$\left\| \frac{df(\theta)}{d\theta} \right\|_2 \geq C \cdot \Big(f(\theta) - \inf_{\theta \in \Theta} f(\theta) \Big)^{1-\xi}, \quad (2)$$

where $C > 0$ and $\xi \in [0, 1]$.

In particular, if an objective function f satisfies both assumptions, gradient-based optimization can be shown to converge to a global minimizer by noting first that uniform smoothness Eq. (1) ensures the gradient updates achieve monotonic improvement with an appropriate step size (i.e., $f(\theta_{t+1}) \leq f(\theta_t) - \frac{1}{2\beta} \cdot \|\nabla f(\theta_t)\|_2^2$, if $\theta_{t+1} \leftarrow \theta_t - \frac{1}{\beta} \cdot \nabla f(\theta_t)$), while the Łojasiewicz inequality Eq. (2) ensures the gradient does not vanish before a global minimizer is reached. Several global convergence results have recently been achieved in the machine learning literature by exploiting assumptions of this kind. For example, in reinforcement learning it has recently been shown that policy gradient (PG) methods converge to a globally optimal policy (Agarwal et al., 2020; Mei et al., 2020b); in supervised learning it has been shown that gradient descent (GD) methods converge to global minimizers of certain non-convex problems (Hazan et al., 2015); and in deep learning theory it has been shown that (stochastic) GD can converge to a global minimizer with an over-parameterized neural network (Allen-Zhu et al., 2019).

However, previous work that relies on the two *uniform* conditions in Definitions 1 and 2 assumes *universal constants*

[*]Equal contribution [1]University of Alberta [2]Google Research, Brain Team [3]DeepMind. Correspondence to: Jincheng Mei <jmei2@ualberta.ca>, Yue Gao <gao12@ualberta.ca>.

Proceedings of the 38^{th} International Conference on Machine Learning, PMLR 139, 2021.

β and C, which ignores important problem structure and limits both the applicability of the results and the strength of the results that can be obtained.

In this paper, we expand the class of problems for which gradient-based optimization is globally convergent, develop novel gradient-based methods that better exploit local structure, and improve the convergence rate analysis. We achieve these results by first defining then investigating a new set of *non-uniform* smoothness and Łojasiewicz inequalities, which generalize the classical definitions and allow a refined characterization of the space of objectives. Given these refined notions, we then tailor novel gradient-based algorithms that improve previous methods for these new problem classes, and extend the analysis to exploit these new forms of non-uniformity, achieving significantly stronger convergence rates in many cases. Importantly, these improvements are achieved in non-convex optimization problems that arise in relevant machine learning problems.

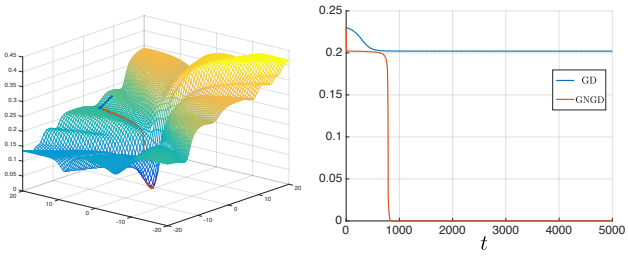
The remainder of the paper is organized as follows. First, in Section 2 we illustrate how natural optimization problems, including those in machine learning, exhibit interesting *local* structure that cannot be adequately captured by the uniform smoothness and Łojasiewicz inequalities. Then, Section 3 introduces the the Non-uniform Smoothness (NS) property and the Non-uniform Łojasiewicz (NŁ) inequality, based on which Section 4 provides non-uniform analyses. Sections 5 and 6 then present new results for policy gradient and generalized linear models respectively. Section 7 concludes the paper and discusses some future directions. Due to space limits, we relegate most of the proofs to the appendix.

2. Motivation

To illustrate the significance of non-uniformity in machine learning problems, we consider examples motivated by recent theoretical (Zhang et al., 2019; Wilson et al., 2019; Mei et al., 2020b) and empirical studies (Cohen et al., 2021).

Regarding smoothness, it is clear that a uniform smoothness constant β cannot always adequately characterize an objective over its entire domain. For example, the convex function $f : x \mapsto x^4$ cannot be informatively characterized by a uniform smoothness constant β because its Hessian $f'' : x \mapsto 12 \cdot x^2$ has the property that $f''(x) \to \infty$ as $|x| \to \infty$, and $f''(x) \to 0$ as $|x| \to 0$. Varying smoothness of this kind has motivated the study of alternative definitions to explain, for example, the effectiveness of gradient clipping in training neural networks and normalization in optimization (Zhang et al., 2019; Wilson et al., 2019). Meanwhile Cohen et al. (2021) present neural network training results that cannot be well explained using the standard smoothness condition of Definition 1.

Regarding the Łojasiewicz inequality, a recent study of policy gradient optimization in reinforcement learning has shown that, with the standard softmax parameterization, the expected return objective cannot satisfy *any* Łojasiewicz inequality with a universal constant C (Mei et al., 2020b), which removes the possibility of using Definition 2 to prove convergence. By introducing a *non-uniform* version of the Łojasiewicz inequality, the same authors were able to show a global convergence rate for the same problem.



(a) MSE landscape in GLM (b) Sub-optimality $\mathcal{L}(\theta_t)$

Figure 1. Non-uniform landscape of non-convex function.

Figure 1 illustrates another example of a non-convex objective, which arises in supervised learning. Subfigure 1(a) visualizes the mean squared error (MSE) of a generalized linear model (GLM) (Hazan et al., 2015), which is not only non-convex but also highly non-uniform. As a "teaser", subfigure 1(b) compares the convergence behavior of two algorithms: standard gradient descent (GD), which suffers from slow convergence on the plateaus due to the non-uniformity of the objective, and an alternative algorithm (GNGD), soon to be introduced. This figure previews how proper handling of non-uniformity in the optimization landscape can enable significant acceleration of optimization progress, including a quick escape from plateaus.

3. Non-uniform Properties and Algorithms

The main results in this paper depend on two core concepts, Non-uniform Smoothness (NS) and Non-uniform Łojasiewicz (NŁ) inequality. The NS property is a new, intuitive generalization of smoothness. The NŁ inequality is a recent proposal of Mei et al. (2020b) and generalizes previous Łojasiewicz inequalities. Our key contribution is to show that the *combination* of these two non-uniform concepts is particularly powerful, applicable to important non-convex objectives in machine learning, and allows the development of improved algorithms and analysis.

3.1. Non-uniform Smoothness

The first main concept we leverage is a generalized notion of smoothness that depends on the parameters non-uniformly.

Definition 3 (Non-uniform Smoothness (NS))**.** *The function $f : \Theta \to \mathbb{R}$ satisfies $\beta(\theta)$ non-uniform smoothness if f is*

differentiable and for all $\theta, \theta' \in \Theta$,

$$\left| f(\theta') - f(\theta) - \left\langle \frac{df(\theta)}{d\theta}, \theta' - \theta \right\rangle \right| \leq \frac{\beta(\theta)}{2} \cdot \|\theta' - \theta\|_2^2,$$

where β *is a positive valued function:* $\beta : \Theta \to (0, \infty)$.

We will refer to $\beta(\theta)$ in Definition 3 as the *NS coefficient*. This alternative definition generalizes and unifies several smoothness concepts from the recent literature. First, NS clearly reduces to Eq. (1) with $\beta(\theta) = \beta$. However, NS also generalizes the notion of (L_0, L_1) smoothness from Zhang et al. (2019) by using $\beta(\theta) = L_0 + L_1 \cdot \|\nabla f(\theta)\|_2$. By using $\beta(\theta) = c \cdot \|\nabla f(\theta)\|_2^{\frac{p-2}{p-1}}$, NS also reduces to the notion of strong smoothness of order p proposed in Wilson et al. (2019). Finally, with $\beta(\theta) = c/\|\theta\|_p^2$, NS reduces to a special form of non-uniform smoothness considered in Mei et al. (2020a). We will show later that NS also covers other previously unstudied smoothness variants. Below we will demonstrate the key benefits of Definition 3 in terms of its *generality*, *better convergence results*, and *practical implications* in conjunction with the NŁ inequality.

3.2. Non-uniform Łojasiewicz Inequality

The second main concept we leverage is a generalized Łojasiewicz inequality introduced by Mei et al. (2020b):

Definition 4 (Non-uniform Łojasiewicz (NŁ))**.** *The differentiable function* $f : \Theta \to \mathbb{R}$ *satisfies the* $(C(\theta), \xi)$ *non-uniform Łojasiewicz inequality if for all* $\theta \in \Theta$,

$$\left\| \frac{df(\theta)}{d\theta} \right\|_2 \geq C(\theta) \cdot |f(\theta) - f(\theta^*)|^{1-\xi}, \tag{3}$$

where $\xi \in (-\infty, 1]$, *and* $C(\theta) : \Theta \to \mathbb{R} > 0$ *holds for all* $\theta \in \Theta$. *In this definition, either* $\theta^* = \arg\min_{\theta \in \Theta} f(\theta)$, *or* $f(\theta^*)$ *is replaced with* $\inf_\theta f(\theta)$ *if the global optimum is not achieved within the domain* Θ.

Definition 4 extends the classical "uniform" Łojasiewicz inequalities in optimization literature, such as the Polyak-Łojasiewicz (PŁ) inequality with $C(\theta) = C > 0$ and $\xi = 1/2$ (Łojasiewicz, 1963; Polyak, 1963); and the Kurdyka-Łojasiewicz (KŁ) inequality[1] by setting $C(\theta) = C > 0$ (Kurdyka, 1998). Following (Mei et al., 2020b;a), we refer to ξ as the *NŁ degree* and $C(\theta)$ as the *NŁ coefficient*. Generally speaking, a larger NŁ degree ξ and NŁ coefficient $C(\theta)$ indicate faster convergence for gradient based algorithms. Appendix E provides an overview of remarkable non-convex functions that satisfy the NŁ inequality for various ξ and $C(\theta)$. As stated, our main contribution in this paper is to show how, when *combined* with NS, NŁ becomes a powerful tool for both algorithm design and analysis, which is a novel direction of investigation.

[1]The KŁ inequality is violated at bad local optima, since vanishing gradient norm cannot dominate non-zero sub-optimality gap. Therefore Definition 4 actually recovers global KŁ inequality.

3.3. Geometry-aware Gradient Descent

A key benefit of the non-uniform definitions is that we can introduce stepsize rules that make gradient descent adapt to the local "geometry" of the optimization objective. First consider the classical gradient decent update.

Definition 5 (Gradient Descent (GD))**.**

$$\theta_{t+1} \leftarrow \theta_t - \eta \cdot \nabla f(\theta_t). \tag{4}$$

The key challenge with deploying GD is choosing the step size η; if η is too large, instability ensues, if too small, progress becomes slow. Recall from the introduction that $\eta = 1/\beta$ is a canonical choice for assuring convergence in *uniformly* β smooth objectives. This suggests that in the presence of *non-uniform* smoothness $\beta(\theta)$ given in NS, the stepsize should be adapted to $1/\beta(\theta)$. This leads to a new variant of normalized gradient descent.

Definition 6 (Geometry-aware Normalized GD (GNGD))**.**

$$\theta_{t+1} \leftarrow \theta_t - \eta \cdot \frac{\nabla f(\theta_t)}{\beta(\theta_t)}. \tag{5}$$

Key to making this approach practical will be efficient ways to measure (or bound) $\beta(\theta)$. Below we will show how in the context of NS and NŁ properties, GNGD can be made both practical and extremely efficient at solving various global optimization problems in machine learning. These results also broaden our fundamental knowledge of the set of objectives that admit efficient global optimization.

4. Non-uniform Analysis

Our first main contribution is an analysis for GD and GNGD based in the presence of non-uniform properties. For minimization problems, we assume $\inf_\theta f(\theta) > -\infty$ ($\sup_\theta f(\theta) < \infty$ for maximization problems).

Theorem 1. *Suppose* $f : \Theta \to \mathbb{R}$ *satisfies NS with* $\beta(\theta)$ *and the NŁ inequality with* $(C(\theta), \xi)$. *Suppose* $C := \inf_{t \geq 1} C(\theta_t) > 0$ *for GD and GNGD. Let* $\delta(\theta) := f(\theta) - f(\theta^*)$ *be the sub-optimality gap. The following hold:*

- **(1a)** *if* $\beta(\theta) \leq c \cdot \delta(\theta)^{1-2\xi}$ *with* $\xi \in (-\infty, 1/2)$, *then the conclusions of (1b) hold;*
- **(1b)** *if* $\beta(\theta) \leq c \cdot \|\nabla f(\theta)\|_2^{\frac{1-2\xi}{1-\xi}}$ *with* $\xi \in (-\infty, 1/2)$, *then GD with* $\eta \in O(1)$ *achieves* $\delta(\theta_t) \in \Theta(1/t^{\frac{1}{1-2\xi}})$, *and GNGD achieves* $\delta(\theta_t) \in O(e^{-t})$.
- **(2a)** *if* $\beta(\theta) \leq L_0 + L_1 \cdot \|\nabla f(\theta)\|_2$, *then the conclusions of (2b) hold;*
- **(2b)** *if* $\beta(\theta) \leq L_0 \cdot \frac{\|\nabla f(\theta)\|^2}{\delta(\theta)^{2-2\xi}} + L_1 \cdot \|\nabla f(\theta)\|_2$, *then GD and GNGD both achieve* $\delta(\theta_t) \in O(1/t^{\frac{1}{1-2\xi}})$ *when* $\xi \in (-\infty, 1/2)$, *and* $O(e^{-t})$ *when* $\xi = 1/2$. *GNGD has strictly better constant than GD* ($1 > C > C^2$).
- **(3a)** *if* $\beta(\theta) \leq c \cdot \|\nabla f(\theta)\|_2^{\frac{1-2\xi}{1-\xi}}$ *with* $\xi \in (1/2, 1)$, *then the conclusions of (3b) hold;*

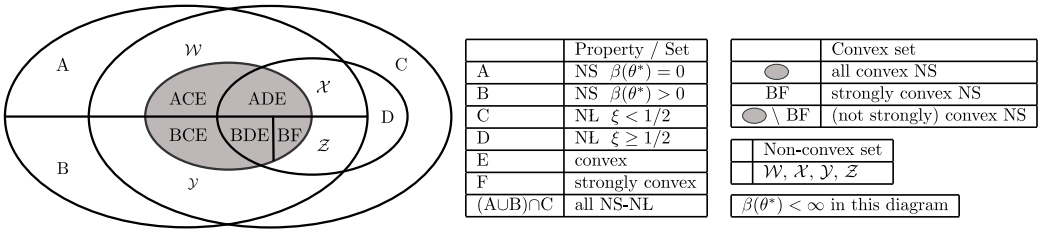


Figure 2. Different function classes for $\beta(\theta^*) < \infty$. We use a label notation where, e.g., C denotes the set of all functions that satisfy property C, and ACE $\coloneqq$ A $\cap$ C $\cap$ E. The two largest ellipsoids correspond to A $\cup$ B and C. We study the following four non-convex function classes in $(\text{A}\cup\text{B})\cap C$, i.e., $\mathcal{W} \coloneqq \text{AC} \setminus (\text{AD}\cup\text{ACE})$, $\mathcal{X} \coloneqq \text{AD}\setminus\text{ADE}$, $\mathcal{Y} \coloneqq \text{BC}\setminus(\text{BD}\cup\text{BCE})$, and $\mathcal{Z} \coloneqq \text{BD}\setminus(\text{BDE}\cup\text{BF})$.

(3b) *if $\beta(\theta) \leq c\cdot\delta(\theta)^{1-2\xi}$ with $\xi \in (1/2, 1)$, then GD with $\eta \in \Theta(1)$ does not converge, while GNGD achieves $\delta(\theta_t) \in O(e^{-t})$.*

Remark 1. *The cases (1)-(3) cover all three possibilities of $\beta(\theta^*)$. Since θ^* is the global minimum, $\nabla^2 f(\theta^*)$ is positive semi-definite (negative if θ^* is maximum) if it exists.*

(1) If $\nabla^2 f(\theta^*) = \mathbf{0}$, then $\beta(\theta) \to 0$ as $\theta, \theta' \to \theta^*$, which means the landscape around θ^* is flat;
(2) If $\nabla^2 f(\theta^*)$ has at least one strictly positive (negative) eigenvalue, then $\beta(\theta) \to \beta > 0$ as $\theta, \theta' \to \theta^*$.

The cases (1)-(2) also cover the situations where the Hessian $\nabla^2 f(\theta^*)$ does not exist but one can find a finite $\beta(\theta^*) > 0$ to upper bound the l.h.s. of Definition 3.

(3) The case (3) is for blow-up type non-existence of $\nabla^2 f(\theta^*)$, where $\beta(\theta^*)$ is unbounded.

Remark 2. *In Theorem 1, $C \coloneqq \inf_{t\geq 1} C(\theta_t)$ is related to the early optimization and plateau escaping behavior studied in Mei et al. (2020a). It remains open to study whether GNGD can be combined with different parameterizations in Mei et al. (2020a) to further improve C.*

Note that (1b) recovers the strong smoothness of order p with $p = 1/\xi$ in Wilson et al. (2019), and (2a) recovers the (L_0, L_1) smoothness of Zhang et al. (2019). The results here consider more general NŁ functions and establish faster rates of convergence. The other cases have not been studied in literature to our knowledge. In Sections 5 and 6 below we study practical machine learning examples that are covered by cases (1) and (2) in Theorem 1. Other cases of different $\beta(\theta)$ and ξ are discussed in Appendix A for completeness.

4.1. Function Classes and Existing Lower Bounds

Before applying these results to problems in machine learning, we first provide a refined characterization of function classes organized by their NS and NŁ properties. This also clarifies the relation between the non-uniform properties and standard notions of convexity and smoothness; see Figure 2.

Proposition 1. *The following hold for an objective f:*

(1) $D \subseteq C$. If f satisfies NŁ with degree ξ, it satisfies NŁ with degree $\xi' < \xi$;
(2) $F \subseteq D$. A strongly convex f satisfies NŁ with $\xi = 1/2$;
(3) $F \cap A = \emptyset$. A strongly convex f cannot satisfy NS with $\beta(\theta) \to 0$ as $\theta, \theta' \to \theta^*$;
(4) $E \subseteq C$. A (not strongly) convex f satisfies NŁ with $\xi = 0$.

The next proposition provides concrete examples for each convex function class in $(A \cup B) \cap C$ in Figure 2.

Proposition 2. *The following results hold: (1) ACE $\neq \emptyset$. (2) ADE $\neq \emptyset$. (3) BCE $\neq \emptyset$. (4) BDE $\neq \emptyset$. (5) BF $\neq \emptyset$.*

A more interesting result considers examples in the classes of non-convex functions $(A \cup B) \cap C$ in Figure 2. The non-uniform analysis above largely still applies to these problems, even when standard convex analysis cannot apply.

Proposition 3. *The following results hold: (1) $\mathcal{W} \coloneqq AC \setminus (AD \cup ACE) \neq \emptyset$. (2) $\mathcal{X} \coloneqq AD \setminus ADE \neq \emptyset$. (3) $\mathcal{Y} \coloneqq BC \setminus (BD \cup BCE) \neq \emptyset$. (4) $\mathcal{Z} \coloneqq BD \setminus (BDE \cup BF) \neq \emptyset$.*

We next apply the techniques to a class of convex functions, achieving results that cannot be explained by classical convex-smooth analysis.

Proposition 4. *The convex function $f : x \mapsto |x|^p$ with $p > 1$ satisfies the NŁ inequality with $\xi = 1/p$ and the NS property with $\beta(x) \leq c_1 \cdot \delta(x)^{1-2\xi}$.*

Consider any $p > 2$, such as $p = 4$, where it follows that f satisfies NŁ with degree $\xi = 1/4 < 1/2$. According to (1a) in Theorem 1, GD will achieve $\delta(x_t) \in \Theta(1/t^2)$, while GNGD attains $\delta(x_t) \in O(e^{-c\cdot t})$ (where $c > 0$). Note that standard convex analysis can only give a $O(1/t)$ rate on (not strongly) convex smooth functions. The $\Theta(1/t^2)$ rate for GD here follows from using NŁ degree $\xi = 1/4$, which improves on $\xi = 0$ from mere convexity ((4) in Proposition 1). The $O(e^{-c\cdot t})$ rate has also been observed for this example by exploiting strong smoothness ((1b), as noted) (Wilson et al., 2019). Figure 2 provides a more general understanding of when this happens.

$\Omega(1/t^2)$ **lower bound for convexity-smoothness.** Note that GNGD satisfies $x_{t+1} = x_1 - \sum_{i=1}^{t} \frac{\eta}{\beta(x_i)} \cdot \nabla f(x_i) \in \text{Span}\{x_1, \nabla f(x_1), \ldots, \nabla f(x_t)\}$, which is a first-order ora-

cle (Nesterov, 2003). Thus there exists a worst-case objective in the convex-smooth class that forces $\delta(x_t) \in \Omega(1/t^2)$ for $t \in O(n)$, where n is the parameter dimension (Nemirovski & Yudin, 1983; Nesterov, 2003; Bubeck, 2014). This is not a contradiction, since the lower bound is established by constructing a convex smooth function with a *constant* $\beta > 0$ (Bubeck, 2014), and $\beta(x) \to \beta > 0$ as $x, x' \to x^*$ in Definition 3. Hence, the $\Omega(1/t^2)$ result covers *some* functions in BCE in Figure 2. Meanwhile $f : x \mapsto |x|^p$ with $p > 2$ satisfies $\beta(x) \to 0$ as $x, x' \to 0$ in Definition 3 (ACE in Figure 2), which implies that the standard convex-smooth class consists of two subclasses. One subclass (BCE) admits first-order sub-linear lower bounds, while the other (ACE) allows linear convergence using first-order methods. This illustrates the *necessity* of non-uniformity in subdividing the NS class as A $\cup$ B in Figure 2. This partition also inspires geometry-aware GD.

$\Omega(1/\sqrt{t})$ **lower bound for** (L_0, L_1)**-smoothness.** For $\beta(\theta) = L_0 + L_1 \cdot \|\nabla f(\theta)\|_2$ ((2a) in Theorem 1) with $L_0, L_1 \geq 1$, standard normalized GD is subject to a $\Omega(1/\sqrt{t})$ lower bound (Zhang et al., 2019). However, in Section 5, we will show that normalized policy gradient (PG) method achieves a linear rate of $O(e^{-c \cdot t})$. Again, this is not a contradiction for similar reasons. With $L_0 \geq 1$, $\beta(\theta) \to L_0 > 0$ as $\theta, \theta' \to \theta^*$, the $\Omega(1/\sqrt{t})$ lower bound will hold for *some* functions in BCE $\cup\, \mathcal{Y}$ in Figure 2. While in Section 5 the objective satisfies $L_0 = 0$ and $L_1 > 0$, hence $\beta(\theta) \to 0$ as $\theta, \theta' \to \theta^*$ (ACE $\cup\, \mathcal{W}$ in Figure 2). This shows a similar separation of rates for first-order methods will also occur based on NS conditions. Furthermore, in Section 6, we will show that both GD and GNGD achieve a $O(e^{-c \cdot t})$ rate for GLM, but here the objective is in $\mathcal{Z}$ in Figure 2 so the lower bounds do not apply.

Unbounded Hessian. Consider any $p \in (1, 2)$, such as $p = 3/2$ where f satisfies $\xi = 2/3$. According to Theorem 1(3a), GD diverges since the Hessian is unbounded near 0. This makes it *necessary* to introduce geometry-aware normalization to ensure convergence, which is verified in Appendix F. This has practical implications for RL, for example ensuring exploration using state distribution entropy, which has unbounded Hessian near probability simplex boundary (Hazan et al., 2019) and alternative probability transforms (Mei et al., 2020a).

5. Policy Gradient

Our second main contribution is to show that the expected return objective considered in direct policy optimization in RL falls under the function class $\mathcal{W}$ in Figure 2, in particular satisfying case (1) of Theorem 1 with NŁ degree $\xi = 0$. The key point is that value functions in Markov decision processes (MDPs) satisfy NS properties with coefficient being the PG norm (Lemmas 2 and 6). This novel finding not only provides a much more precise characterization than existing standard smoothness results (Agarwal et al., 2020; Mei et al., 2020b), but also enables PG with normalization to use the NŁ inequalities (Lemmas 1 and 5) differently than for standard PG ($\|\nabla f(\theta)\|_2$ vs. $\|\nabla f(\theta)\|_2^2$), which leads to faster convergence as well as plateau escaping.

5.1. RL Settings and Notations

For a finite set $\mathcal{N}$, let $\Delta(\mathcal{N})$ denote the set of all probability distributions on $\mathcal{N}$. A finite MDP $\mathcal{M} \coloneqq (\mathcal{S}, \mathcal{A}, \mathcal{P}, r, \gamma)$ is determined by a finite state space $\mathcal{S}$, a finite action space $\mathcal{A}$, a transition function $\mathcal{P} : \mathcal{S} \times \mathcal{A} \to \Delta(\mathcal{S})$, a scalar reward function $r : \mathcal{S} \times \mathcal{A} \to \mathbb{R}$, and a discount factor $\gamma \in [0, 1)$.

In policy-based RL, an agent interacts with the environment, i.e., the MDP $\mathcal{M}$, using a policy $\pi : \mathcal{S} \to \Delta(\mathcal{A})$. Given a state s_t, the agent takes an action $a_t \sim \pi(\cdot|s_t)$, receives a one-step scalar reward $r(s_t, a_t)$ and a next-state $s_{t+1} \sim \mathcal{P}(\cdot|s_t, a_t)$. The long-term expected reward, also known as the value function of π under s, is defined as

$$V^\pi(s) \coloneqq \mathop{\mathbb{E}}_{\substack{s_0 = s, a_t \sim \pi(\cdot|s_t), \\ s_{t+1} \sim \mathcal{P}(\cdot|s_t, a_t)}} \left[\sum_{t=0}^{\infty} \gamma^t r(s_t, a_t) \right]. \tag{6}$$

The state-action value of π at $(s, a) \in \mathcal{S} \times \mathcal{A}$ is defined as $Q^\pi(s, a) \coloneqq r(s, a) + \gamma \sum_{s'} \mathcal{P}(s'|s, a) V^\pi(s')$, and $A^\pi(s, a) \coloneqq Q^\pi(s, a) - V^\pi(s)$ is the advantage function of π. The state distribution of π is defined as,

$$d^\pi_{s_0}(s) \coloneqq (1 - \gamma) \sum_{t=0}^{\infty} \gamma^t \Pr(s_t = s | s_0, \pi, \mathcal{P}). \tag{7}$$

Given an initial state distribution $\rho \in \Delta(\mathcal{S})$, we denote $V^\pi(\rho) \coloneqq \mathbb{E}_{s \sim \rho}[V^\pi(s)]$ and $d^\pi_\rho(s) \coloneqq \mathbb{E}_{s_0 \sim \rho}\left[d^\pi_{s_0}(s)\right]$. There exists an optimal policy π^* such that $V^{\pi^*}(\rho) = \sup_{\pi : \mathcal{S} \to \Delta(\mathcal{A})} V^\pi(\rho)$. For convenience, we denote $V^* \coloneqq V^{\pi^*}$. Consider a tabular representation, i.e., $\theta(s, a) \in \mathbb{R}$ for all (s, a), so that the policy π_θ can be parameterized by θ as $\pi_\theta(\cdot|s) = \mathrm{softmax}(\theta(s, \cdot))$; that is, for all (s, a),

$$\pi_\theta(a|s) = \frac{\exp\{\theta(s, a)\}}{\sum_{a' \in \mathcal{A}} \exp\{\theta(s, a')\}}. \tag{8}$$

When there is only one state the policy $\pi_\theta = \mathrm{softmax}(\theta)$ is defined as $\pi_\theta(a) = \exp\{\theta(a)\} / \sum_{a' \in \mathcal{A}} \exp\{\theta(a')$. The problem of policy-based RL is then to find a policy π_θ that maximizes the value function, i.e.,

$$\sup_{\theta : \mathcal{S} \times \mathcal{A} \to \mathbb{R}} V^{\pi_\theta}(\rho). \tag{9}$$

For convenience, and without loss of generality, we assume $r(s, a) \in [0, 1]$ for all $(s, a) \in \mathcal{S} \times \mathcal{A}$.

5.2. One-state MDPs

We first illustrate some key insights for one-state MDPs with K actions and $\gamma = 0$. The value function Eq. (6)

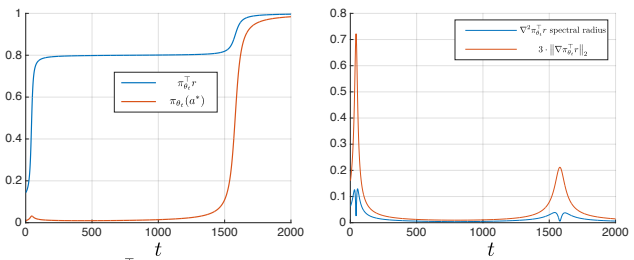
reduces to expected reward $\pi_\theta^\top r$, where $r \in [0,1]^K$, $\theta \in \mathbb{R}^K$, and $\pi_\theta = \text{softmax}(\theta)$. Mei et al. (2020b) have shown that even though $\max_\theta \pi_\theta^\top r$ is a non-concave maximization, global convergence can be achieved with a $O(1/t)$ rate using uniform smoothness and the NŁ inequality:

Lemma 1 (NŁ, Mei et al. (2020b), Lemma 3)**.** *Let a^* be the optimal action. Denote $\pi^* = \arg\max_{\pi \in \Delta} \pi^\top r$. Then,*

$$\left\| \frac{d\pi_\theta^\top r}{d\theta} \right\|_2 \geq \pi_\theta(a^*) \cdot (\pi^* - \pi_\theta)^\top r. \tag{10}$$

Note that Lemma 1 is not improvable in terms of the coefficients $C(\theta) = \pi_\theta(a^*)$ and $\xi = 0$ (Mei et al., 2020b, Remark 1 and Lemma 17). However, this result is based on only using a *uniform* smoothness coefficient $\beta = 5/2$ (Mei et al., 2020b, Lemma 2), which even empirical evidence suggests can be significantly refined. To illustrate, we run standard policy gradient (PG) on a 3-action one-state MDP. As shown in Figure 3(a), PG first goes through a long suboptimal plateau, and then eventually escapes to approach π^*. Figure 3(b) presents the spectral radius of the Hessian and the PG norm $3 \cdot \left\| \frac{d\pi_{\theta_t}^\top r}{d\theta_t} \right\|_2$ as functions of time t. It is evident that the smoothness behaves non-uniformly: it is close to zero at the suboptimal plateau and near π^*, highly aligned with the PG norm. Compared to any universal constant β, the PG norm characterizes the non-uniform landscape information far more precisely. We formalize this observation by proving the following key result:

Lemma 2 (NS)**.** *Denote $\theta_\zeta := \theta + \zeta \cdot (\theta' - \theta)$ with some $\zeta \in [0,1]$. For any $r \in [0,1]^K$, $\theta \mapsto \pi_\theta^\top r$ satisfies $\beta(\theta_\xi)$ non-uniform smoothness with $\beta(\theta_\zeta) = 3 \cdot \left\| \frac{d\pi_{\theta_\zeta}^\top r}{d\theta_\zeta} \right\|_2$.*



(a) $\pi_{\theta_t}^\top r$ and $\pi_{\theta_t}(a^*)$ (b) Hessian spectral radius and PG norm

Figure 3. PG results on $r = (1.0, 0.8, 0.1)^\top$.

Comparing Lemma 2 with (1b) in Theorem 1, we have $\xi = 0$, and GNGD requires normalizing $\beta(\theta_\zeta)$, which is the PG norm of θ_ζ rather than θ. However, ζ is unknown. Fortunately, the next lemma shows that, if we still normalize the PG norm of θ, the $\beta(\theta_\zeta)$ in Lemma 2 can be upper bounded by $\left\| \frac{d\pi_\theta^\top r}{d\theta} \right\|_2$, given the learning rate is small enough:

Lemma 3. *Let $\theta' = \theta + \eta \cdot \frac{d\pi_\theta^\top r}{d\theta} \Big/ \left\| \frac{d\pi_\theta^\top r}{d\theta} \right\|_2$. Denote $\theta_\zeta := \theta + \zeta \cdot (\theta' - \theta)$ with some $\zeta \in [0,1]$. We have, for all $\eta \in (0, 1/3)$,*

$$\left\| \frac{d\pi_{\theta_\zeta}^\top r}{d\theta_\zeta} \right\|_2 \leq \frac{1}{1 - 3\eta} \cdot \left\| \frac{d\pi_\theta^\top r}{d\theta} \right\|_2. \tag{11}$$

Next, the NŁ coefficient $\pi_\theta(a^*)$ is bounded away from 0, which provides constants in the convergence rate results.

Lemma 4 (Non-vanishing NŁ coefficient)**.** *Using normalized policy gradient method, we have $\inf_{t \geq 1} \pi_{\theta_t}(a^*) > 0$.*

To this point, we demonstrate that the non-concave function $\pi_\theta^\top r$ satisfies (1b) in Theorem 1 with $\xi = 0$ *in each iteration of normalized PG*[2]: Lemmas 2 and 3 show that the NS coefficient $\beta(\theta_{\zeta_t}) \leq c_1 \cdot \left\| \frac{d\pi_{\theta_t}^\top r}{d\theta_t} \right\|_2$, while Lemmas 1 and 4 guarantee $\left\| \frac{d\pi_{\theta_t}^\top r}{d\theta_t} \right\|_2 \geq c_2 \cdot (\pi^* - \pi_{\theta_t})^\top r$. Therefore, combining Lemmas 1 to 4, we prove the global linear convergence rate $O(e^{-c \cdot t})$ of normalized PG:

Theorem 2. *Using normalized PG $\theta_{t+1} = \theta_t + \eta \cdot \frac{d\pi_{\theta_t}^\top r}{d\theta_t} \Big/ \left\| \frac{d\pi_{\theta_t}^\top r}{d\theta_t} \right\|_2$, with $\eta = 1/6$, for all $t \geq 1$, we have,*

$$(\pi^* - \pi_{\theta_t})^\top r \leq e^{-\frac{c \cdot (t-1)}{12}} \cdot (\pi^* - \pi_{\theta_1})^\top r, \tag{12}$$

where $c = \inf_{t \geq 1} \pi_{\theta_t}(a^*) > 0$ is from Lemma 4, and c is a constant that depends on r and θ_1, but not on the time t.

Remark 3. *If π_{θ_1} is uniform, i.e., $\pi_{\theta_1}(a) = 1/K$, $\forall\, a \in [K]$, then we have $c \geq 1/K$ in Theorem 2. This can be proved by showing that $\pi_{\theta_{t+1}}(a^*) \geq \pi_{\theta_t}(a^*)$, similar to Mei et al. (2020b, Proposition 2).*

5.3. Geometry-aware Normalized PG (GNPG)

Next, we generalize from one-state to finite MDPs, using the GNPG[3] on value function, as shown in Algorithm 1.

Algorithm 1 Geometry-aware Normalized Policy Gradient

Input: Learning rate $\eta > 0$.
Initialize parameter $\theta_1(s,a)$ for all (s,a).
while $t \geq 1$ **do**
$\theta_{t+1} \leftarrow \theta_t + \eta \cdot \frac{\partial V^{\pi_{\theta_t}}(\mu)}{\partial \theta_t} \Big/ \left\| \frac{\partial V^{\pi_{\theta_t}}(\mu)}{\partial \theta_t} \right\|_2$.
end while

[2]This essentially means we prove that a uniform Łojasiewicz inequality holds for the entire sequence $\{\theta_t\}_{t \geq 1}$, but this does not imply that the NŁ condition is unnecessary. As shown in Mei et al. (2020b, Remark 1), Łojasiewicz-type inequalities with constant $C > 0$ cannot hold. It can only become uniform after specifying an initialization θ_1 and an algorithm (in this case, PG). Otherwise, uniform Łojasiewicz cannot hold since initialization can make the NŁ coefficient $\pi_\theta(a^*)$ arbitrarily close to 0.

[3]We use GNPG as the name of Algorithm 1, since NPG is usually used to refer to the natural PG algorithm in RL literature.

5.4. General MDPs

For general finite MDPs, we assume "sufficient exploration" for the initial state distribution μ, which is also adapted in literature (Agarwal et al., 2020; Mei et al., 2020b).

Assumption 1 (Sufficient exploration)**.** *The initial state distribution satisfies* $\min_s \mu(s) > 0$.

Given Assumption 1, Agarwal et al. (2020) prove asymptotic global convergence for PG on the non-concave $\max_\theta V^{\pi_\theta}(\rho)$ problem, while Mei et al. (2020b) strengthens this to a $O(1/t)$ rate using uniform smoothness and the following NŁ inequality that generalizes Lemma 1.

Lemma 5 (NŁ, Mei et al. (2020b), Lemma 8)**.** *Denote* $S := |\mathcal{S}|$ *as the total number of states. We have,* $\forall\, \theta \in \mathbb{R}^{\mathcal{S}\times\mathcal{A}}$,

$$\left\|\frac{\partial V^{\pi_\theta}(\mu)}{\partial \theta}\right\|_2 \geq \frac{\min_s \pi_\theta(a^*(s)|s)}{\sqrt{S}\cdot\left\|d_\rho^{\pi^*}/d_\mu^{\pi_\theta}\right\|_\infty}\cdot\left(V^*(\rho)-V^{\pi_\theta}(\rho)\right),$$

where $a^*(s)$ *is the action that* π^* *selects in state* s.

Here, the NŁ degree $\xi = 0$ is not improvable Mei et al. (2020b, Lemma 28). In one-state MDPs with $S = 1$, Lemma 5 recovers Lemma 1 with the same NŁ coefficient $C(\theta) = \pi_\theta(a^*)$, indicating that $C(\theta)$ in Lemma 5 might also be unimprovable. On the other hand, the uniform smoothness considered in (Agarwal et al., 2020; Mei et al., 2020b) $\beta = 8/(1-\gamma)^3$ is too conservative, particularly when γ is close to 1. Our next key result shows that the policy value also satisfies a stronger NS property, with the NS coefficient being the PG norm, generalizing Lemma 2:

Lemma 6 (NS)**.** *Let Assumption 1 hold and denote* $\theta_\zeta := \theta + \zeta\cdot(\theta'-\theta)$ *with some* $\zeta \in [0,1]$. $\theta \mapsto V^{\pi_\theta}(\mu)$ *satisfies* $\beta(\theta_\zeta)$ *non-uniform smoothness with*

$$\beta(\theta_\zeta) = \left[3 + \frac{2\cdot(C_\infty - (1-\gamma))}{(1-\gamma)\cdot\gamma}\right]\cdot\sqrt{S}\cdot\left\|\frac{\partial V^{\pi_{\theta_\zeta}}(\mu)}{\partial\theta_\zeta}\right\|_2,$$

where $C_\infty := \max_\pi \left\|\frac{d_\mu^\pi}{\mu}\right\|_\infty \leq \frac{1}{\min_s \mu(s)} < \infty$.

In one-state MDPs with $\gamma = 0$ and $S = 1$, we have $C_\infty = 1-\gamma$. Thus Lemma 6 reduces to Lemma 2 with the same NS coefficient $\beta(\theta_\zeta) = 3\cdot\left\|\frac{d\pi_{\theta_\zeta}^\top r}{d\theta_\zeta}\right\|_2$. Similar to Lemma 3, if we use Algorithm 1 with small enough learning rate, then $\beta(\theta_\zeta)$ in Lemma 6 is upper bounded by the PG norm of θ:

Lemma 7. *Let* $\eta = \frac{(1-\gamma)\cdot\gamma}{6\cdot(1-\gamma)\cdot\gamma+4\cdot(C_\infty-(1-\gamma))}\cdot\frac{1}{\sqrt{S}}$ *and* $\theta' = \theta + \eta\cdot\frac{\partial V^{\pi_\theta}(\mu)}{\partial\theta}\Big/\left\|\frac{\partial V^{\pi_\theta}(\mu)}{\partial\theta}\right\|_2$. *Denote* $\theta_\zeta := \theta+\zeta\cdot(\theta'-\theta)$ *with some* $\zeta\in[0,1]$. *We have,*

$$\left\|\frac{\partial V^{\pi_{\theta_\zeta}}(\mu)}{\partial\theta_\zeta}\right\|_2 \leq 2\cdot\left\|\frac{\partial V^{\pi_\theta}(\mu)}{\partial\theta}\right\|_2. \tag{13}$$

Next, the NŁ coefficient $\min_s \pi_\theta(a^*(s)|s)$ in Lemma 5 is lower bounded away from 0:

Lemma 8 (Non-vanishing NŁ coefficient)**.** *Let Assumption 1 hold. We have,* $c := \inf_{s\in\mathcal{S},t\geq 1}\pi_{\theta_t}(a^*(s)|s) > 0$, *where* $\{\theta_t\}_{t\geq 1}$ *is generated by Algorithm 1.*

Now we have the non-concave function $V^{\pi_\theta}(\rho)$ satisfies (1b) in Theorem 1 with $\xi = 0$ *in each iteration of Algorithm 1*. Therefore, combining Lemmas 5 to 8, we prove the global linear convergence rate $O(e^{-c\cdot t})$ of Algorithm 1:

Theorem 3. *Let Assumption 1 hold and let* $\{\theta_t\}_{t\geq 1}$ *be generated using Algorithm 1 with* $\eta = \frac{(1-\gamma)\cdot\gamma}{6\cdot(1-\gamma)\cdot\gamma+4\cdot(C_\infty-(1-\gamma))}\cdot\frac{1}{\sqrt{S}}$, *where* $C_\infty := \max_\pi\left\|\frac{d_\mu^\pi}{\mu}\right\|_\infty$. *Denote* $C'_\infty := \max_\pi\left\|\frac{d_\rho^\pi}{\mu}\right\|_\infty$. *Let* c *be the positive constant from Lemma 8. We have, for all* $t \geq 1$,

$$V^*(\rho) - V^{\pi_{\theta_t}}(\rho) \leq \frac{(V^*(\mu)-V^{\pi_{\theta_1}}(\mu))\cdot C'_\infty}{1-\gamma}\cdot e^{-C\cdot(t-1)},$$

where $C = \frac{(1-\gamma)^2\cdot\gamma\cdot c}{12\cdot(1-\gamma)\cdot\gamma+8\cdot(C_\infty-(1-\gamma))}\cdot\frac{1}{S}\cdot\left\|\frac{d_\mu^{\pi^*}}{\mu}\right\|_\infty^{-1}$.

Not only the $O(e^{-c\cdot t})$ rate in Theorem 3 is faster than $O(1/t)$ for standard PG without normalization, but also the constant is better than Mei et al. (2020b, Theorem 4). The strictly better dependence c ($\gg c^2$ in PG) is related to faster escaping plateaus as shown later (Mei et al., 2020a).

Remark 4. *The conclusion of GNPG has better constants than PG ($c \gg c^2$) arises from upper bounds (Theorem 3 and Mei et al. (2020b, Theorem 4)), which is also supported by empirical evidence. In fact, there exists a lower bound that shows c cannot be removed for PG (Mei et al., 2020a, Theorem 1) under one-state MDP settings. For finite MDPs, very recently, Li et al. (2021) show that for softmax PG (without normalization), c can be very small in terms of the number of states. It remains open to consider whether c is reasonably large for GNPG.*

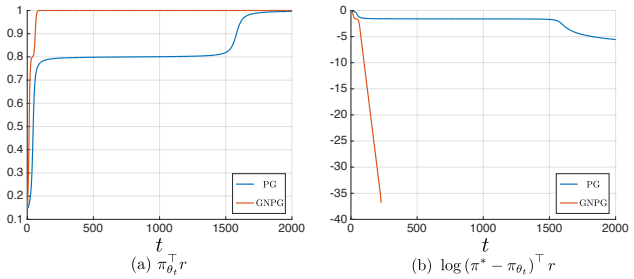
Remark 5. *To our knowledge, existing PG variants can achieve linear convergence $O(e^{-c\cdot t})$ only if using at least one of the following techniques: (a)* ***regularization****; Mei et al. (2020b) prove that entropy regularized PG enjoys $O(e^{-c\cdot t})$ convergence toward the regularized optimal policy. (b)* ***natural gradient****; Cen et al. (2020) prove that entropy regularized natural PG achieves linear convergence. (c)* ***exact line-search****; Bhandari & Russo (2020) prove that without parameterization, PG variants with exact line-search achieve linear rates by approximating policy iteration.*

Among the above techniques, regularization changes the problem to regularized MDPs, while natural PG and line-search require solving expensive optimization problems to do updates, since each update is an $\arg\max$.

On the contrary, Algorithm 1 enjoys global $O(e^{-c\cdot t})$ rate *(i)* without using regularization, since Algorithm 1 directly works on the original MDPs; *(ii)* without solving optimization problems in each iteration, and the normalized PG update is cheap. The strong results rely on the NS and NŁ

properties, and also the geometry-aware normalization that takes advantage of the non-uniform properties.

Remark 6. *Standard softmax PG with bounded learning rate follows $\Omega(1/t)$ lower bound (Mei et al., 2020b), which is consistent with the case (1) in Theorem 1. Algorithm 1 achieves faster linear convergence rates, indicating that the adaptive update stepsize $\eta/\left\|\nabla V^{\pi_{\theta_t}}(\rho)\right\|_2$ is asymptotically unbounded, since $\left\|\nabla V^{\pi_{\theta_t}}(\rho)\right\|_2 \to 0$ as $t \to \infty$.*



(a) $\pi^*_{\theta_t} r$ (b) $\log(\pi^* - \pi_{\theta_t})^\top r$

Figure 4. PG and GNPG on $r = (1.0, 0.8, 0.1)^\top$.

We compare PG and GNPG on the one-state MDP problem as shown in Figure 4. Figure 4(a) shows that GNPG escapes from the sub-optimal plateau significantly faster than PG, while Figure 4(b) shows that GNPG follows linear convergence $O(e^{-c\cdot t})$ of sub-optmality, verifying the theoretical results. Similar experimental results on synthetic tree MDPs with multiple states are presented in Appendix F.

6. Generalized Linear Models

Next, we investigate the generalized linear model (GLM) with quasi-maximum likelihood estimate (quasi-MLE), which applied widely in supervised learning. We show that the mean squared error (MSE) of GLM (Hazan et al., 2015) is in the non-convex function class $\mathcal{Z}$ in Figure 2, and it satisfies the case (2) in Theorem 1 with $\xi = 1/2$. As a result, both GD and GNGD achieve global linear convergence rates $O(e^{-c\cdot t})$, significantly improving the best existing results of $O(1/\sqrt{t})$ (Hazan et al., 2015). We also provide new understandings of using normalization in GLM based on our non-uniform analysis.

6.1. Basic Settings and Notations

Given a training data set $\mathcal{D} = \{(x_i, y_i)\}_{i\in[N]}$, which consists of N data points, there is a feature map $x_i \mapsto \phi(x_i) \in \mathbb{R}^d$ for each pair $(x_i, y_i) \in \mathcal{D}$. We denote $\phi_i := \phi(x_i)$ for conciseness. For each data point x_i, we have $y_i \in [0, 1]$ as the ground truth likelihood. Following Hazan et al. (2015), our model is parameterized by a weight vector $\theta \in \mathbb{R}^d$ as ,

$$\pi_i = \sigma(\phi_i^\top \theta) = \frac{1}{1 + \exp\{-\phi_i^\top \theta\}}, \tag{14}$$

where $\sigma : \mathbb{R} \to (0, 1)$ is the sigmoid activation. The problem is to minimize the mean squared error (MSE),

$$\min_\theta \mathcal{L}(\theta) = \min_{\theta\in\mathbb{R}^d} \frac{1}{N} \cdot \sum_{i=1}^N (\pi_i - y_i)^2. \tag{15}$$

We assume $y_i = \pi_i^* := \sigma(\phi_i^\top \theta^*)$, where $\theta^* \in \mathbb{R}^d$, and $\|\theta^*\|_2 < \infty$, which means the target y_i is realizable and non-deterministic. According to Hazan et al. (2015), the MSE in Eq. (15) is not quasi-convex (thus not convex). Fortunately, Hazan et al. (2015) manage to show that Eq. (15) satisfies a weaker Strictly-Locally-Quasi-Convex (SLQC) property, based on which they prove the following result:

Theorem 4 (Hazan et al. (2015))**.** *With diminishing learning rate $\eta_t \in \Theta(1/\sqrt{t})$, the normalized gradient descent (NGD) update $\theta_{t+1} \leftarrow \theta_t - \eta_t \cdot \frac{\partial \mathcal{L}(\theta_t)}{\partial \theta_t} \Big/ \left\|\frac{\partial \mathcal{L}(\theta_t)}{\partial \theta_t}\right\|_2$ satisfies,*

$$\delta(\theta_t) := \mathcal{L}(\theta_t) - \mathcal{L}(\theta^*) \in O(1/\sqrt{t}), \tag{16}$$

where $\theta^* := \arg\min_\theta \mathcal{L}(\theta)$ is the global optimal solution.

6.2. Fast Convergence using Non-uniform Analysis

Based on the $O(1/\sqrt{t})$ rate for NGD in Theorem 4, Hazan et al. (2015) propose to normalize gradient norm in MSE minimization. However, there is no lower bound for other methods including GD on GLM, and thus it is not clear if there exists a faster rate for GLM optimization.

Surprisingly, we prove that both GD and GNGD actually achieve much faster rates of $O(e^{-c\cdot t})$ using the non-uniform analysis. Our first key finding is to show that the MSE in GLM satisfies a new NŁ inequality with $\xi = 1/2$:

Lemma 9 (NŁ)**.** *Denote $u(\theta) := \min_{i\in[N]}\{\pi_i \cdot (1 - \pi_i)\}$, and $v := \min_{i\in[N]}\{\pi_i^* \cdot (1 - \pi_i^*)\}$. We have,*

$$\left\|\frac{\partial \mathcal{L}(\theta)}{\partial \theta}\right\|_2 \geq C(\theta, \phi) \cdot \left[\frac{1}{N} \cdot \sum_{i=1}^N (\pi_i - \pi_i^*)^2\right]^{\frac{1}{2}}, \tag{17}$$

holds for all $\theta \in \mathbb{R}^d$, where

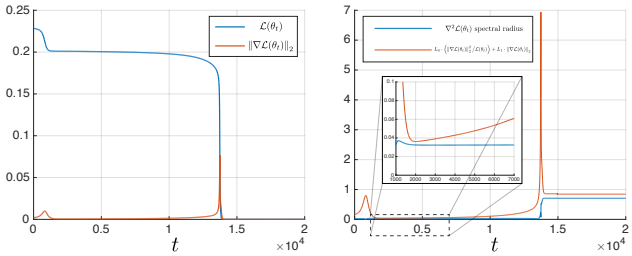
$$C(\theta, \phi) = 8 \cdot u(\theta) \cdot \min\{u(\theta), v\} \cdot \sqrt{\lambda_\phi}, \tag{18}$$

and λ_ϕ is the smallest positive eigenvalue of $\frac{1}{N} \cdot \sum_{i=1}^N \phi_i \phi_i^\top$.

Remark 7. *It is not clear if results similar to Lemma 9 hold without assuming: (i) realizable optimal prediction $y_i = \pi_i^* := \sigma(\phi_i^\top \theta^*)$; (ii) non-deterministic optimal prediction $\|\theta^*\|_2 < \infty$. We leave it as an open question to study non-uniformity of GLM without the above assumptions.*

In Lemma 9, λ_ϕ is determined by the feature ϕ, and $u(\theta)$ shows that the gradient is vanishing when π_i is near deterministic, which is consistent with the fact that the sigmoid saturates and provides uninformative gradient as the parameter magnitude becomes large.

We run GD on one example with $N = 10$ and $d = 2$. As shown in Figure 5, the gradient norm $\|\nabla \mathcal{L}(\theta_t)\|_2$ is close to



(a) $\mathcal{L}(\theta_t)$ and $\|\nabla\mathcal{L}(\theta_t)\|_2$ (b) Hessian spectral radius and NS

Figure 5. Experiments on GLM using GD.

zero at plateaus and near optimum. However, unlike the PG, the spectral radius of the Hessian $\nabla^2\mathcal{L}(\theta_t)$ is only close to zero at plateaus, while it approaches positive constant near optimum. This indicates a different NS condition other than Lemmas 2 and 6 is needed, since only gradient norm $\to 0$ cannot upper bound the spectral radius of Hessian $\to c > 0$. With some calculations, we prove the following key results:

Lemma 10 (Smoothness and NS)**.** *$\mathcal{L}(\theta)$ satisfies β smoothness with*

$$\beta = \frac{3}{8} \cdot \max_{i\in[N]} \|\phi_i\|_2^2, \tag{19}$$

and $\beta(\theta)$ NS with

$$\beta(\theta) = L_1 \cdot \left\| \frac{\partial\mathcal{L}(\theta)}{\partial\theta} \right\|_2 + L_0 \cdot \left(\left\| \frac{\partial\mathcal{L}(\theta)}{\partial\theta} \right\|_2^2 \Big/ \mathcal{L}(\theta) \right).$$

At the optimal solution θ^*, the spectral radius of the Hessian $\frac{\partial^2\mathcal{L}(\theta^*)}{\partial(\theta^*)^2}$ is strictly positive. Therefore, the MSE objective of Eq. (15) is in the non-convex function class $\mathcal{Z}$ in Figure 2, and it satisfies the case (2) in Theorem 1 with $\xi = 1/2$. Combining Lemmas 9 and 10 and applying Theorem 1, we have the following global linear convergence result:

Theorem 5. *With $\eta = 1/\beta$, GD update satisfies for all $t \geq 1$, $\mathcal{L}(\theta_t) \leq \mathcal{L}(\theta_1) \cdot e^{-C^2\cdot(t-1)}$. With $\eta \in \Theta(1)$, GNGD update satisfies for all $t \geq 1$, $\mathcal{L}(\theta_t) \leq \mathcal{L}(\theta_1) \cdot e^{-C\cdot(t-1)}$, where $C \in (0, 1)$, i.e., GNGD is strictly faster than GD.*

Theorem 5 significantly improves the $O(1/\sqrt{t})$ rate in Theorem 4. The key difference is that we discovered a new NŁ inequality of Lemma 9 that is satisfied by GLMs. The linear convergence rates are verified in Appendix F.

In Theorem 5, we have $C = \inf_{t\geq 1} C(\theta_t, \phi)$, which is very close to zero if π_i is near deterministic, and GD suffers sub-optimality plateaus as shown in Figure 1. GNGD has strictly (orders of magnitudes) better constant dependence $C \gg C^2$, and escapes plateaus significantly faster than GD. Intuitively, for the GLM in Figure 1, C in Theorem 1 is lower bounded reasonably if θ_1 is initialized within some finite distance of the central valley containing θ^*.

Combining the NŁ and NS properties (Lemmas 9 and 10), we provide new understandings of using normalization in GLM: *(i)* **First**, using standard NGD (Hazan et al., 2015) for all $t \geq 1$ is not a good choice. By examining the asymptotic behaviour as $\theta \to \theta^*$, we have $\beta(\theta) \to \beta > 0$. However, the normalization $\left\| \frac{\partial\mathcal{L}(\theta)}{\partial\theta} \right\|_2$ in standard NGD gives incremental updates with adaptive stepsize $\to \infty$. To guarantee convergence, it is necessary to use $\eta_t \to 0$, which counteracts normalization and slows down the learning, since it could be not easy to find a learning rate scheme. This is consistent with the $O(e^{-c\cdot t})$ result for GD with $\eta > 0$ and without normalization in Theorem 5. *(ii)* **Second**, using geometry-aware normalization $\beta(\theta_t)$ is a better choice than normalizing the gradient norm $\|\nabla\mathcal{L}(\theta_t)\|_2$. We elaborate this point by investigating *both the asymptotic and the early-stage behaviours* using NS-NŁ. Since $\beta(\theta_t) \to \beta > 0$ asymptotically, GNGD is approaching GD as $\theta_t \to \theta^*$, which makes GNGD enjoy the same $O(e^{-c\cdot t})$ rate. On the other hand, at early-stage optimization (e.g., close to initialization in Figure 1), when θ_t is far from θ^*, we have thus $\beta(\theta_t) \leq c \cdot \left\| \frac{\partial\mathcal{L}(\theta_t)}{\partial\theta_t} \right\|_2$. Then GNGD is close to NGD, which guarantees strictly better progresses than GD. This is because of the progress of GNGD in each iteration at this time is about $\|\nabla\mathcal{L}(\theta_t)\|_2$, while the progress of GD is $\|\nabla\mathcal{L}(\theta_t)\|_2^2$, and the gradient norm is close to 0 on plateaus. Using NŁ of Lemma 9, GNGD will have strictly better constant dependence C than C^2 in GD.

7. Conclusions and Future Work

The main contributions of this paper concern a general characterization and analysis based on non-uniform properties, which are not only sufficiently general to cover concrete examples, but also significantly improve convergence rates over previous work and even over classical lower bounds. The most exciting part is the techniques apply to important applications in machine learning that involve non-convex optimization problems. One direction is to further push the analysis to other domains with more complex function approximators, including neural networks (Allen-Zhu et al., 2019). Another valuable future work is to incorporate stochastic gradient (Karimi et al., 2016) and other adaptive gradient-based methods (Kingma & Ba, 2014) in the analysis. Finally, applying other cases of non-uniform properties beyond those mentioned in this paper would be interesting.

Acknowledgements

The authors would like to thank anonymous reviewers for their valuable comments. Jincheng Mei would like to thank Ziyi Chen for pointing that NŁ recovers global KŁ property. Jincheng Mei and Bo Dai would like to thank Nicolas Le Roux for providing feedback on a draft of this manuscript. Csaba Szepesvári and Dale Schuurmans gratefully acknowledge funding from the Canada CIFAR AI Chairs Program, Amii and NSERC.

References

Agarwal, A., Kakade, S. M., Lee, J. D., and Mahajan, G. Optimality and approximation with policy gradient methods in markov decision processes. In *Conference on Learning Theory*, pp. 64–66. PMLR, 2020.

Allen-Zhu, Z., Li, Y., and Song, Z. A convergence theory for deep learning via over-parameterization. In *International Conference on Machine Learning*, pp. 242–252. PMLR, 2019.

Bhandari, J. and Russo, D. A note on the linear convergence of policy gradient methods. *arXiv preprint arXiv:2007.11120*, 2020.

Bubeck, S. Convex optimization: Algorithms and complexity. *arXiv preprint arXiv:1405.4980*, 2014.

Cen, S., Cheng, C., Chen, Y., Wei, Y., and Chi, Y. Fast global convergence of natural policy gradient methods with entropy regularization. *arXiv preprint arXiv:2007.06558*, 2020.

Cohen, J., Kaur, S., Li, Y., Kolter, J. Z., and Talwalkar, A. Gradient descent on neural networks typically occurs at the edge of stability. In *International Conference on Learning Representations*, 2021.

Hazan, E., Levy, K., and Shalev-Shwartz, S. Beyond convexity: Stochastic quasi-convex optimization. *Advances in neural information processing systems*, 28:1594–1602, 2015.

Hazan, E., Kakade, S., Singh, K., and Van Soest, A. Provably efficient maximum entropy exploration. In *International Conference on Machine Learning*, pp. 2681–2691. PMLR, 2019.

Kakade, S. and Langford, J. Approximately optimal approximate reinforcement learning. In *ICML*, volume 2, pp. 267–274, 2002.

Karimi, H., Nutini, J., and Schmidt, M. Linear convergence of gradient and proximal-gradient methods under the polyak-łojasiewicz condition. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, pp. 795–811. Springer, 2016.

Kingma, D. P. and Ba, J. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

Kurdyka, K. On gradients of functions definable in o-minimal structures. In *Annales de l'institut Fourier*, volume 48, pp. 769–783, 1998.

Li, G., Wei, Y., Chi, Y., Gu, Y., and Chen, Y. Softmax policy gradient methods can take exponential time to converge. *arXiv preprint arXiv:2102.11270*, 2021.

Łojasiewicz, S. Une propriété topologique des sous-ensembles analytiques réels. *Les équations aux dérivées partielles*, 117:87–89, 1963.

Mei, J., Xiao, C., Dai, B., Li, L., Szepesvári, C., and Schuurmans, D. Escaping the gravitational pull of softmax. *Advances in Neural Information Processing Systems*, 33, 2020a.

Mei, J., Xiao, C., Szepesvari, C., and Schuurmans, D. On the global convergence rates of softmax policy gradient methods. In *International Conference on Machine Learning*, pp. 6820–6829. PMLR, 2020b.

Nemirovski, A. S. and Yudin, D. B. Problem complexity and method efficiency in optimization. 1983.

Nesterov, Y. *Introductory lectures on convex optimization: A basic course*, volume 87. Springer Science & Business Media, 2003.

Polyak, B. T. Gradient methods for minimizing functionals. *Zhurnal Vychislitel'noi Matematiki i Matematicheskoi Fiziki*, 3(4):643–653, 1963.

Sutton, R. S., McAllester, D. A., Singh, S. P., and Mansour, Y. Policy gradient methods for reinforcement learning with function approximation. In *Advances in neural information processing systems*, pp. 1057–1063, 2000.

Wilson, A., Mackey, L., and Wibisono, A. Accelerating rescaled gradient descent: Fast optimization of smooth functions. *arXiv preprint arXiv:1902.08825*, 2019.

Zhang, J., He, T., Sra, S., and Jadbabaie, A. Why gradient clipping accelerates training: A theoretical justification for adaptivity. In *International Conference on Learning Representations*, 2019.

Appendix

The appendix is organized as follows.

- Appendix A: proofs for general optimization results in Section 4.
  - Appendix A.1: proofs for the main Theorem 1.
  - Appendix A.2: proofs for the function classes, i.e., Propositions 1 to 4.
- Appendix B: proofs for policy gradient in Section 5.
  - Appendix B.1: proofs for one-state MDPs.
  - Appendix B.2: proofs for general MDPs.
- Appendix C: proofs for generalized linear model in Section 6.
- Appendix D: miscellaneous extra supporting results those are not mentioned in the main paper.
- Appendix E: non-convex (non-concave) examples for the NŁ inequality in literature.
- Appendix F: additional simulation results which are not in the main paper.

A. Proofs for Section 4

A.1. Main Theorem 1

Theorem 1. Suppose $f : \Theta \to \mathbb{R}$ satisfies NS with $\beta(\theta)$ and the NŁ inequality with $(C(\theta), \xi)$. Suppose $C \coloneqq \inf_{t\geq 1} C(\theta_t) > 0$ for GD and GNGD. Let $\delta(\theta) \coloneqq f(\theta) - f(\theta^*)$ be the sub-optimality gap. The following hold:

(1a) if $\beta(\theta) \leq c \cdot \delta(\theta)^{1-2\xi}$ with $\xi \in (-\infty, 1/2)$, then the conclusions of (1b) hold;

(1b) if $\beta(\theta) \leq c \cdot \|\nabla f(\theta)\|_2^{\frac{1-2\xi}{1-\xi}}$ with $\xi \in (-\infty, 1/2)$, then GD with $\eta \in O(1)$ achieves $\delta(\theta_t) \in \Theta(1/t^{\frac{1}{1-2\xi}})$, and GNGD achieves $\delta(\theta_t) \in O(e^{-t})$.

(2a) if $\beta(\theta) \leq L_0 + L_1 \cdot \|\nabla f(\theta)\|_2$, then the conclusions of (2b) hold;

(2b) if $\beta(\theta) \leq L_0 \cdot \frac{\|\nabla f(\theta)\|^2}{\delta(\theta)^{2-2\xi}} + L_1 \cdot \|\nabla f(\theta)\|_2$, then GD and GNGD both achieve $\delta(\theta_t) \in O(1/t^{\frac{1}{1-2\xi}})$ when $\xi \in (-\infty, 1/2)$, and $O(e^{-t})$ when $\xi = 1/2$. GNGD has strictly better constant than GD ($1 \geq C \geq C^2$).

(3a) if $\beta(\theta) \leq c \cdot \|\nabla f(\theta)\|_2^{\frac{1-2\xi}{1-\xi}}$ with $\xi \in (1/2, 1)$, then the conclusions of (3b) hold;

(3b) if $\beta(\theta) \leq c \cdot \delta(\theta)^{1-2\xi}$ with $\xi \in (1/2, 1)$, then GD with $\eta \in \Theta(1)$ does not converge, while GNGD achieves $\delta(\theta_t) \in O(e^{-t})$.

Proof. **(1a) First part:** *$O(1/t^{\frac{1}{1-2\xi}})$ upper bound for GD update $\theta_{t+1} \leftarrow \theta_t - \eta \cdot \nabla f(\theta_t)$ with $\eta \in O(1)$.*

We show that using GD with learning rate $\eta = \frac{1}{c \cdot \delta(\theta_1)^{1-2\xi}}$, the sub-optimality $\delta(\theta_t)$ is monotonically decreasing. And thus there exists a universal constant $\beta > 0$ such that $\beta(\theta_t) \leq \beta$, for all $t \geq 1$.

Denote $\beta \coloneqq c \cdot \delta(\theta_1)^{1-2\xi}$. We have $\beta \in (0, \infty)$, since $f(\theta^*) > -\infty$, and $f(\theta^*) < f(\theta_1) < \infty$. By assumption, we have $\beta(\theta_1) \leq \beta$. According to Lemma 11, using GD with $\eta = \frac{1}{\beta}$, we have,

$$\delta(\theta_2) - \delta(\theta_1) = f(\theta_2) - f(\theta_1) \leq 0. \tag{20}$$

Therefore, we have,

$$\begin{aligned} \beta(\theta_2) &\leq c \cdot \delta(\theta_2)^{1-2\xi} && \text{(by assumption)} && (21)\\ &\leq c \cdot \delta(\theta_1)^{1-2\xi} && (0 < \delta(\theta_2) \leq \delta(\theta_1) \text{ and } \xi < 1/2) && (22)\\ &= \beta. && && (23) \end{aligned}$$

Repeating similar arguments of Eqs. (20) and (21), we have, for all $t \geq 1$, $\beta(\theta_t) \leq \beta$ and,

$$0 < \delta(\theta_{t+1}) \leq \delta(\theta_t). \tag{24}$$

Therefore, we have, for all $t \geq 1$ (or using Lemma 11),

$$\delta(\theta_{t+1}) - \delta(\theta_t) = f(\theta_{t+1}) - f(\theta_t) \tag{25}$$

$$\leq \nabla f(\theta_t)^\top (\theta_{t+1} - \theta_t) + \frac{\beta(\theta_t)}{2} \cdot \|\theta_{t+1} - \theta_t\|_2^2 \qquad \text{(NS)} \tag{26}$$

$$\leq \nabla f(\theta_t)^\top (\theta_{t+1} - \theta_t) + \frac{\beta}{2} \cdot \|\theta_{t+1} - \theta_t\|_2^2 \qquad (\beta(\theta_t) \leq \beta) \tag{27}$$

$$= -\frac{1}{2\beta} \cdot \|\nabla f(\theta_t)\|_2^2 \qquad \left(\theta_{t+1} \leftarrow \theta_t - \frac{1}{\beta} \cdot \nabla f(\theta_t)\right) \tag{28}$$

$$\leq -\frac{1}{2\beta} \cdot C(\theta_t)^2 \cdot \delta(\theta_t)^{2-2\xi} \qquad \text{(NŁ)} \tag{29}$$

$$\leq -\frac{1}{2\beta} \cdot C^2 \cdot \delta(\theta_t)^{2-2\xi}. \qquad \left(C \coloneqq \inf_{t \geq 1} C(\theta_t) > 0\right) \tag{30}$$

According to Lemma 13, given any $\alpha > 0$, we have, for all $x \in [0, 1]$,

$$\frac{1}{\alpha} \cdot (1 - x^\alpha) \geq x^\alpha \cdot (1 - x). \tag{31}$$

Let $\alpha = 1 - 2\xi > 0$, since $\xi < 1/2$. Also let $x = \frac{\delta(\theta_{t+1})}{\delta(\theta_t)} \in (0, 1]$ due to Eq. (24). We have,

$$\frac{1}{1-2\xi} \cdot \left[1 - \frac{\delta(\theta_{t+1})^{1-2\xi}}{\delta(\theta_t)^{1-2\xi}}\right] \geq \frac{\delta(\theta_{t+1})^{1-2\xi}}{\delta(\theta_t)^{1-2\xi}} \cdot \left[1 - \frac{\delta(\theta_{t+1})}{\delta(\theta_t)}\right]. \tag{32}$$

Next, we have,

$$\frac{1}{\delta(\theta_t)^{1-2\xi}} = \frac{1}{\delta(\theta_1)^{1-2\xi}} + \frac{1}{\delta(\theta_t)^{1-2\xi}} - \frac{1}{\delta(\theta_1)^{1-2\xi}} \tag{33}$$

$$= \frac{1}{\delta(\theta_1)^{1-2\xi}} + \sum_{s=1}^{t-1} \left[\frac{1}{\delta(\theta_{s+1})^{1-2\xi}} - \frac{1}{\delta(\theta_s)^{1-2\xi}}\right] \tag{34}$$

$$= \frac{1}{\delta(\theta_1)^{1-2\xi}} + \sum_{s=1}^{t-1} \frac{1-2\xi}{\delta(\theta_{s+1})^{1-2\xi}} \cdot \frac{1}{1-2\xi} \cdot \left[1 - \frac{\delta(\theta_{s+1})^{1-2\xi}}{\delta(\theta_s)^{1-2\xi}}\right] \tag{35}$$

$$\geq \frac{1}{\delta(\theta_1)^{1-2\xi}} + \sum_{s=1}^{t-1} \cancel{\frac{1-2\xi}{\delta(\theta_{s+1})^{1-2\xi}}} \cdot \cancel{\frac{\delta(\theta_{s+1})^{1-2\xi}}{\delta(\theta_s)^{1-2\xi}}} \cdot \left[1 - \frac{\delta(\theta_{s+1})}{\delta(\theta_s)}\right] \qquad \text{(by Eq. (32))} \tag{36}$$

$$= \frac{1}{\delta(\theta_1)^{1-2\xi}} + \sum_{s=1}^{t-1} \frac{1-2\xi}{\delta(\theta_s)^{2-2\xi}} \cdot [\delta(\theta_s) - \delta(\theta_{s+1})] \tag{37}$$

$$\geq \frac{1}{\delta(\theta_1)^{1-2\xi}} + \sum_{s=1}^{t-1} \cancel{\frac{1-2\xi}{\delta(\theta_s)^{2-2\xi}}} \cdot \frac{C^2}{2\beta} \cdot \cancel{\delta(\theta_s)^{2-2\xi}} \qquad \text{(by Eq. (25))} \tag{38}$$

$$= \frac{1}{\delta(\theta_1)^{1-2\xi}} + \frac{(1-2\xi) \cdot C^2}{2\beta} \cdot (t-1), \tag{39}$$

which implies for all $t \geq 1$,

$$f(\theta_t) - f(\theta^*) = \delta(\theta_t) \leq \left[\frac{1}{(f(\theta_1) - f(\theta^*))^{1-2\xi}} + \frac{(1-2\xi) \cdot C^2}{2\beta} \cdot (t-1)\right]^{-\frac{1}{1-2\xi}} \in O\left(\frac{1}{t^{\frac{1}{1-2\xi}}}\right). \tag{40}$$

(1a) Second part: $\Omega(1/t^{\frac{1}{1-2\xi}})$ *lower bound for GD update* $\theta_{t+1} \leftarrow \theta_t - \eta_t \cdot \nabla f(\theta_t)$ *with* $\eta_t \in (0, 1]$.

According to the NS property of Definition 3, we have, for all θ and θ',

$$f(\theta') \leq f(\theta) + \nabla f(\theta)^\top (\theta' - \theta) + \frac{\beta(\theta)}{2} \cdot \|\theta' - \theta\|_2^2. \tag{41}$$

Fix θ and take minimum over θ' on both sides of the above inequality. Then we have,

$$f(\theta^*) \le f(\theta) + \min_{\theta'} \left\{ \nabla f(\theta)^\top (\theta' - \theta) + \frac{\beta(\theta)}{2} \cdot \|\theta' - \theta\|_2^2 \right\} \tag{42}$$

$$= f(\theta) - \frac{1}{\beta(\theta)} \cdot \|\nabla f(\theta)\|_2^2 + \frac{1}{2 \cdot \beta(\theta)} \cdot \|\nabla f(\theta)\|_2^2 \qquad \left(\theta' = \theta - \frac{1}{\beta(\theta)} \cdot \nabla f(\theta)\right) \tag{43}$$

$$= f(\theta) - \frac{1}{2 \cdot \beta(\theta)} \cdot \|\nabla f(\theta)\|_2^2, \tag{44}$$

which implies,

$$\|\nabla f(\theta)\|_2^2 \le 2 \cdot \beta(\theta) \cdot \delta(\theta) \tag{45}$$

$$\le 2 \cdot c \cdot \delta(\theta)^{2-2\xi}. \qquad \left(\beta(\theta) \le c \cdot \delta(\theta)^{1-2\xi}\right) \tag{46}$$

Therefore, we have,

$$\delta(\theta_t) - \delta(\theta_{t+1}) = f(\theta_t) - f(\theta_{t+1}) + \nabla f(\theta_t)^\top (\theta_{t+1} - \theta_t) - \nabla f(\theta_t)^\top (\theta_{t+1} - \theta_t) \tag{47}$$

$$\le \frac{\beta}{2} \cdot \|\theta_{t+1} - \theta_t\|_2^2 - \nabla f(\theta_t)^\top (\theta_{t+1} - \theta_t) \qquad \text{(by NS and } \beta(\theta_t) \le \beta) \tag{48}$$

$$= \left(\frac{\beta}{2} \cdot \eta_t^2 + \eta_t\right) \cdot \|\nabla f(\theta_t)\|_2^2 \qquad (\theta_{t+1} \leftarrow \theta_t - \eta_t \cdot \nabla f(\theta_t)) \tag{49}$$

$$\le \left(\frac{\beta}{2} \cdot \eta_t^2 + \eta_t\right) \cdot 2 \cdot c \cdot \delta(\theta_t)^{2-2\xi} \qquad \text{(by Eq. (45))} \tag{50}$$

$$\le (\beta + 2) \cdot c \cdot \delta(\theta_t)^{2-2\xi}. \qquad (\eta_t \in (0, 1]) \tag{51}$$

Next, we show that $\frac{\delta(\theta_{t+1})}{\delta(\theta_t)} \ge \frac{3-4\xi}{4-4\xi}$ holds for all large enough $t \ge 1$ by contradiction. According to the upper bound results in the first part, we have $\delta(\theta_t) \to 0$ as $t \to \infty$. Suppose $\frac{\delta(\theta_{t+1})}{\delta(\theta_t)} < \frac{3-4\xi}{4-4\xi}$, where $t \ge 1$ is large enough and $\delta(\theta_t)$ is small enough. We have,

$$\delta(\theta_{t+1}) \ge \delta(\theta_t) - (\beta + 2) \cdot c \cdot \delta(\theta_t)^{2-2\xi} \qquad \text{(by Eq. (47))} \tag{52}$$

$$> \frac{4-4\xi}{3-4\xi} \cdot \delta(\theta_{t+1}) - (\beta + 2) \cdot c \cdot \left(\frac{4-4\xi}{3-4\xi}\right)^{2-2\xi} \cdot \delta(\theta_{t+1})^{2-2\xi}, \tag{53}$$

where the last inequality is because of the function $f : x \mapsto x - a \cdot x^{2-2\xi}$ with $a > 0$ is monotonically increasing for all $0 < x \le \frac{1}{[(2-2\xi)a]^{1/(1-2\xi)}}$. Eq. (52) implies that,

$$\delta(\theta_{t+1})^{1-2\xi} > \frac{1}{3-4\xi} \cdot \frac{1}{(\beta+2) \cdot c} \cdot \left(\frac{3-4\xi}{4-4\xi}\right)^{2-2\xi}, \tag{54}$$

for large enough $t \ge 1$, which is a contradiction with $\delta(\theta_t) \to 0$ as $t \to \infty$. Thus we have $\frac{\delta(\theta_{t+1})}{\delta(\theta_t)} \ge \frac{3-4\xi}{4-4\xi}$ holds for all large enough $t \ge 1$. Denote

$$t_0 \coloneqq \min \left\{ t \ge 1 : \frac{\delta(\theta_{s+1})}{\delta(\theta_s)} \ge \frac{3-4\xi}{4-4\xi}, \text{ for all } s \ge t \right\}. \tag{55}$$

According to Lemma 14, given any $\alpha > 0$, we have, for all $x \in \left[\frac{2\alpha+1}{2\alpha+2}, 1\right]$,

$$\frac{1}{2\alpha} \cdot (1 - x^\alpha) \le x^\alpha \cdot (1 - x). \tag{56}$$

Let $\alpha = 1 - 2\xi > 0$, since $\xi < 1/2$. We have $\frac{2\alpha+1}{2\alpha+2} = \frac{3-4\xi}{4-4\xi}$. Also let $x = \frac{\delta(\theta_{t+1})}{\delta(\theta_t)} \in \left[\frac{3-4\xi}{4-4\xi}, 1\right]$. We have,

$$\frac{1}{2 \cdot (1-2\xi)} \cdot \left[1 - \frac{\delta(\theta_{t+1})^{1-2\xi}}{\delta(\theta_t)^{1-2\xi}}\right] \le \frac{\delta(\theta_{t+1})^{1-2\xi}}{\delta(\theta_t)^{1-2\xi}} \cdot \left[1 - \frac{\delta(\theta_{t+1})}{\delta(\theta_t)}\right], \tag{57}$$

for all $t \ge t_0$. On the other hand, since $t_0 \in O(1)$ and $1 - 2\xi > 0$, we have, for all $t < t_0$,

$$\delta(\theta_{t+1})^{1-2\xi} \ge c_0 > 0. \tag{58}$$

Next, we have, for all $t \geq t_0$,

$$\frac{1}{\delta(\theta_t)^{1-2\xi}} = \frac{1}{\delta(\theta_1)^{1-2\xi}} + \sum_{s=1}^{t-1}\left[\frac{1}{\delta(\theta_{s+1})^{1-2\xi}} - \frac{1}{\delta(\theta_s)^{1-2\xi}}\right] \tag{59}$$

$$= \frac{1}{\delta(\theta_1)^{1-2\xi}} + \sum_{s=1}^{t_0-1}\frac{1}{\delta(\theta_{s+1})^{1-2\xi}}\cdot\left[1 - \frac{\delta(\theta_{s+1})^{1-2\xi}}{\delta(\theta_s)^{1-2\xi}}\right] + \sum_{s=t_0}^{t-1}\frac{2\cdot(1-2\xi)}{\delta(\theta_{s+1})^{1-2\xi}}\cdot\frac{1}{2\cdot(1-2\xi)}\cdot\left[1 - \frac{\delta(\theta_{s+1})^{1-2\xi}}{\delta(\theta_s)^{1-2\xi}}\right] \tag{60}$$

$$\leq \frac{1}{\delta(\theta_1)^{1-2\xi}} + \sum_{s=1}^{t_0-1}\frac{1}{c_0}\cdot 1 + \sum_{s=t_0}^{t-1}\frac{2\cdot(1-2\xi)}{\cancel{\delta(\theta_{s+1})^{1-2\xi}}}\cdot\frac{\cancel{\delta(\theta_{s+1})^{1-2\xi}}}{\delta(\theta_s)^{1-2\xi}}\cdot\left[1 - \frac{\delta(\theta_{s+1})}{\delta(\theta_s)}\right] \qquad \text{(by Eq. (57))} \tag{61}$$

$$= \frac{1}{\delta(\theta_1)^{1-2\xi}} + \frac{t_0-1}{c_0} + \sum_{s=t_0}^{t-1}\frac{2\cdot(1-2\xi)}{\delta(\theta_s)^{2-2\xi}}\cdot[\delta(\theta_s) - \delta(\theta_{s+1})] \tag{62}$$

$$\leq \frac{1}{\delta(\theta_1)^{1-2\xi}} + \frac{t_0-1}{c_0} + \sum_{s=t_0}^{t-1}\frac{2\cdot(1-2\xi)}{\cancel{\delta(\theta_s)^{2-2\xi}}}\cdot(\beta+2)\cdot c\cdot\cancel{\delta(\theta_s)^{2-2\xi}} \qquad \text{(by Eq. (47))} \tag{63}$$

$$= \frac{1}{\delta(\theta_1)^{1-2\xi}} + \frac{t_0-1}{c_0} + 2\cdot(1-2\xi)\cdot(\beta+2)\cdot c\cdot(t-t_0), \tag{64}$$

which implies for all large enough $t \geq 1$,

$$f(\theta_t) - f(\theta^*) = \delta(\theta_t) \geq \left[\frac{1}{(f(\theta_1) - f(\theta^*))^{1-2\xi}} + \frac{t_0-1}{c_0} + 2\cdot(1-2\xi)\cdot(\beta+2)\cdot c\cdot(t-t_0)\right]^{-\frac{1}{1-2\xi}} \in \Omega\left(\frac{1}{t^{\frac{1}{1-2\xi}}}\right). \tag{65}$$

(1a) Third part: *$O(e^{-t})$ upper bound for GNGD update $\theta_{t+1} \leftarrow \theta_t - \frac{\nabla f(\theta_t)}{\beta(\theta_t)}$.*

We have, for all $t \geq 1$ (or using Lemma 12),

$$\delta(\theta_{t+1}) - \delta(\theta_t) = f(\theta_{t+1}) - f(\theta_t) \tag{66}$$

$$\leq \nabla f(\theta_t)^\top(\theta_{t+1} - \theta_t) + \frac{\beta(\theta_t)}{2}\cdot\|\theta_{t+1} - \theta_t\|_2^2 \qquad \text{(NS)} \tag{67}$$

$$= -\frac{1}{\beta(\theta_t)}\cdot\|\nabla f(\theta_t)\|_2^2 + \frac{1}{2}\cdot\frac{1}{\beta(\theta_t)}\cdot\|\nabla f(\theta_t)\|_2^2 \qquad \left(\theta_{t+1} \leftarrow \theta_t - \frac{\nabla f(\theta_t)}{\beta(\theta_t)}\right) \tag{68}$$

$$= -\frac{1}{2\cdot\beta(\theta_t)}\cdot\|\nabla f(\theta_t)\|_2^2 \tag{69}$$

$$\leq -\frac{1}{2\cdot\beta(\theta_t)}\cdot C(\theta_t)^2\cdot\delta(\theta_t)^{2-2\xi} \qquad \text{(NŁ)} \tag{70}$$

$$\leq -\frac{1}{2\cdot\beta(\theta_t)}\cdot C^2\cdot\delta(\theta_t)^{2-2\xi} \qquad \left(C \coloneqq \inf_{t\geq 1} C(\theta_t) > 0\right) \tag{71}$$

$$\leq -\frac{C^2}{2\cdot c}\cdot\delta(\theta_t), \qquad \left(\beta(\theta_t) \leq c\cdot\delta(\theta_t)^{1-2\xi}\right) \tag{72}$$

which implies for all $t \geq 1$,

$$f(\theta_t) - f(\theta^*) = \delta(\theta_t) \leq \left(1 - C^2/(2\cdot c)\right)\cdot\delta(\theta_{t-1}) \tag{73}$$

$$\leq \exp\left\{-C^2/(2\cdot c)\right\}\cdot\delta(\theta_{t-1}) \tag{74}$$

$$\leq \exp\left\{-(t-1)\cdot C^2/(2\cdot c)\right\}\cdot\delta(\theta_1) \tag{75}$$

$$= \exp\left\{-(t-1)\cdot C^2/(2\cdot c)\right\}\cdot(f(\theta_1) - f(\theta^*)). \tag{76}$$

(1b) First part: *$O(1/t^{\frac{1}{1-2\xi}})$ upper bound for GD update $\theta_{t+1} \leftarrow \theta_t - \eta\cdot\nabla f(\theta_t)$ with $\eta \in O(1)$.*

Denote $\beta_1 \coloneqq c\cdot\|\nabla f(\theta_1)\|_2^{\frac{1-2\xi}{1-\xi}}$. We have $\beta_1 \in (0,\infty)$, since f is differentiable (Definition 3). Using $\eta \leq \frac{1}{\beta_1}$ and according to Lemma 11, we have $\delta(\theta_2) \leq \delta(\theta_1)$. Denote $\beta_2 \coloneqq c\cdot\|\nabla f(\theta_2)\|_2^{\frac{1-2\xi}{1-\xi}}$. We also have $\beta_2 \in (0,\infty)$. Repeating the update,

we generate $\{\theta_t\}_{t\geq 1}$ such that $\delta(\theta_{t+1}) \leq \delta(\theta_t)$. Denote

$$\beta \coloneqq \sup_{t\geq 1}\{\beta_t\} = \sup_{t\geq 1}\left\{c\cdot\|\nabla f(\theta_2)\|_2^{\frac{1-2\xi}{1-\xi}}\right\}. \tag{77}$$

Now we have $0 \leq \delta(\theta_{t+1}) \leq \delta(\theta_t) \leq \cdots \leq \delta(\theta_1)$. According to the monotone convergence theorem, $\delta(\theta_t)$ converges to some finite value. And the gradient $\|\nabla f(\theta_t)\|_2 \to 0$, otherwise a small gradient update can decrease the sub-optimality, which is a contradiction with convergence. Thus we have $\beta \in (\beta_1, \infty)$, since $\beta_t \to 0$ as $t \to \infty$. Using $\eta = \frac{1}{\beta}$, we have $\eta \leq \frac{1}{\beta_t}$ holds for all $t \geq 1$, and,

$$\beta(\theta_t) \leq c\cdot\|\nabla f(\theta_t)\|_2^{\frac{1-2\xi}{1-\xi}} = \beta_t \leq \beta. \tag{78}$$

Using similar calculations in the first part of (1a), we have the $O(1/t^{\frac{1}{1-2\xi}})$ upper bound.

(1b) Second part: $\Omega(1/t^{\frac{1}{1-2\xi}})$ *lower bound for GD update* $\theta_{t+1} \leftarrow \theta_t - \eta_t\cdot\nabla f(\theta_t)$ *with* $\eta_t \in (0, 1]$.

According to Eq. (45), we have,

$$\begin{aligned}
\|\nabla f(\theta)\|_2^2 &\leq 2\cdot\beta(\theta)\cdot\delta(\theta) && (79)\\
&\leq 2\cdot c\cdot\|\nabla f(\theta)\|_2^{\frac{1-2\xi}{1-\xi}}\cdot\delta(\theta), \qquad \left(\beta(\theta) \leq c\cdot\|\nabla f(\theta)\|_2^{\frac{1-2\xi}{1-\xi}}\right) && (80)
\end{aligned}$$

which is equivalent to,

$$\|\nabla f(\theta)\|_2^2 \leq 2\cdot c_1\cdot\delta(\theta)^{2-2\xi}, \tag{81}$$

where $c_1 \coloneqq \frac{1}{2}\cdot(2\cdot c)^{2-2\xi}$. According to Eq. (47), we have,

$$\begin{aligned}
\delta(\theta_t) - \delta(\theta_{t+1}) &\leq \left(\frac{\beta}{2}\cdot\eta_t^2 + \eta_t\right)\cdot\|\nabla f(\theta_t)\|_2^2 && (82)\\
&\leq (\beta+2)\cdot c_1\cdot\delta(\theta_t)^{2-2\xi}. \qquad \text{(by Eq. (81) and } \eta_t \in (0,1]) && (83)
\end{aligned}$$

Using similar calculations in the second part of (1a), we have the $\Omega(1/t^{\frac{1}{1-2\xi}})$ lower bound.

(1b) Third part: $O(e^{-t})$ *upper bound for GNGD update* $\theta_{t+1} \leftarrow \theta_t - \frac{\nabla f(\theta_t)}{\beta(\theta_t)}$.

According to Lemma 12, we have, for all $t \geq 1$,

$$\begin{aligned}
\delta(\theta_{t+1}) - \delta(\theta_t) &\leq -\frac{1}{2\cdot\beta(\theta_t)}\cdot\|\nabla f(\theta_t)\|_2^2 && (84)\\
&\leq -\frac{1}{2\cdot c}\cdot\|\nabla f(\theta_t)\|_2^{\frac{1}{1-\xi}} \qquad \left(\beta(\theta_t) \leq c\cdot\|\nabla f(\theta_t)\|_2^{\frac{1-2\xi}{1-\xi}}\right) && (85)\\
&\leq -\frac{1}{2\cdot c}\cdot C(\theta_t)^{\frac{1}{1-\xi}}\cdot\delta(\theta_t) \qquad (\text{NŁ}) && (86)\\
&\leq -\frac{1}{2\cdot c}\cdot C^{\frac{1}{1-\xi}}\cdot\delta(\theta_t) \qquad \left(C \coloneqq \inf_{t\geq 1} C(\theta_t) > 0\right), && (87)
\end{aligned}$$

which implies (similar to Eq. (73)),

$$f(\theta_t) - f(\theta^*) = \delta(\theta_t) \leq \exp\left\{-(t-1)\cdot C^{\frac{1}{1-\xi}}/(2\cdot c)\right\}\cdot(f(\theta_1) - f(\theta^*)). \tag{88}$$

(2a) First part: $O(1/t^{\frac{1}{1-2\xi}})$ upper bound for GD when $\xi < 1/2$.

Similar to the first part of (1b), we denote $\beta_t \coloneqq L_0 + L_1\cdot\|\nabla f(\theta_t)\|_2$ and $\beta \coloneqq \sup_{t\geq 1}\{\beta_t\} \in (L_0, \infty)$ since $\|\nabla f(\theta_t)\|_2 \to 0$ as $t \to \infty$. Using $\eta = \frac{1}{\beta}$, we have $\eta \leq \frac{1}{\beta_t}$ holds for all $t \geq 1$ and $\beta(\theta_t) \leq L_0 + L_1\cdot\|\nabla f(\theta_t)\|_2 \leq \beta$. According to Eq. (25) and the first part of (1a), we have the $O(1/t^{\frac{1}{1-2\xi}})$ upper bound.

(2a) Second part: $O(e^{-t})$ upper bound for GD when $\xi = 1/2$.

According to Lemma 11, we have, for all $t \geq 1$,

$$\delta(\theta_{t+1}) - \delta(\theta_t) \leq -\frac{1}{2\beta} \cdot \|\nabla f(\theta_t)\|_2^2 \tag{89}$$

$$\leq -\frac{1}{2\beta} \cdot C(\theta_t)^2 \cdot \delta(\theta_t) \qquad (\text{NŁ with } \xi = 1/2) \tag{90}$$

$$\leq -\frac{1}{2\beta} \cdot C^2 \cdot \delta(\theta_t), \qquad \left(C \coloneqq \inf_{t\geq 1} C(\theta_t) > 0\right) \tag{91}$$

which implies (similar to Eq. (73)),

$$f(\theta_t) - f(\theta^*) = \delta(\theta_t) \leq \exp\left\{-(t-1) \cdot C^2/(2\beta)\right\} \cdot (f(\theta_1) - f(\theta^*)) . \tag{92}$$

(2a) Third part: $O(1/t^{\frac{1}{1-2\xi}})$ upper bound for GNGD when $\xi < 1/2$.

According to Lemma 12, we have, for all $t \geq 1$,

$$\delta(\theta_{t+1}) - \delta(\theta_t) \leq -\frac{1}{2 \cdot \beta(\theta_t)} \cdot \|\nabla f(\theta_t)\|_2^2 \tag{93}$$

$$\leq -\frac{1}{2} \cdot \frac{\|\nabla f(\theta_t)\|_2^2}{L_0 + L_1 \cdot \|\nabla f(\theta_t)\|_2} \qquad (\beta(\theta_t) \leq L_0 + L_1 \cdot \|\nabla f(\theta_t)\|_2) \tag{94}$$

$$\leq -\frac{1}{2} \cdot \frac{\|\nabla f(\theta_t)\|_2^2}{L_0 + L_1 \cdot \beta} \qquad \left(\beta \coloneqq \sup_{t\geq 1} \{\|\nabla f(\theta_t)\|_2\} \in (\|\nabla f(\theta_1)\|_2, \infty)\right) \tag{95}$$

$$\leq -\frac{1}{2} \cdot \frac{C^2}{L_0 + L_1 \cdot \beta} \cdot \delta(\theta_t)^{2-2\xi} \qquad \left(\text{NŁ and } C \coloneqq \inf_{t\geq 1} C(\theta_t) > 0\right), \tag{96}$$

which is similar to Eq. (25). Using similar calculations in the first part of (1a), we have the $O(1/t^{\frac{1}{1-2\xi}})$ upper bound.

(2a) Fourth part: $O(e^{-t})$ upper bound for GNGD when $\xi = 1/2$.

We have, for all $t \geq 1$,

$$\delta(\theta_{t+1}) - \delta(\theta_t) \leq -\frac{1}{2} \cdot \frac{\|\nabla f(\theta_t)\|_2^2}{L_0 + L_1 \cdot \beta} \qquad (\text{by Eq. (95)}) \tag{97}$$

$$\leq -\frac{1}{2} \cdot \frac{C^2}{L_0 + L_1 \cdot \beta} \cdot \delta(\theta_t) \qquad \left(\text{NŁ with } \xi = 1/2 \text{ and } C \coloneqq \inf_{t\geq 1} C(\theta_t) > 0\right), \tag{98}$$

which implies (similar to Eq. (73)),

$$f(\theta_t) - f(\theta^*) = \delta(\theta_t) \leq \exp\left\{-(t-1) \cdot C^2/(2 \cdot (L_0 + L_1 \cdot \beta))\right\} \cdot (f(\theta_1) - f(\theta^*)) . \tag{99}$$

(2b) First part: $O(1/t^{\frac{1}{1-2\xi}})$ upper bound for GD when $\xi < 1/2$.

Denote $\beta_t \coloneqq L_0 \cdot \frac{\|\nabla f(\theta_t)\|^2}{\delta(\theta_t)^{2-2\xi}} + L_1 \cdot \|\nabla f(\theta_t)\|_2$ and $\beta \coloneqq \sup_{t\geq 1} \{\beta_t\} \in (\beta_1, \infty)$. According to Eq. (25) and the first part of (1a), we have the $O(1/t^{\frac{1}{1-2\xi}})$ upper bound.

(2b) Second part: $O(e^{-t})$ upper bound for GD when $\xi = 1/2$.

According to Lemma 11, we have, for all $t \geq 1$ (same as the second part of (2a)),

$$\delta(\theta_{t+1}) - \delta(\theta_t) \leq -\frac{1}{2\beta} \cdot \|\nabla f(\theta_t)\|_2^2 \tag{100}$$

$$\leq -\frac{1}{2\beta} \cdot C(\theta_t)^2 \cdot \delta(\theta_t) \qquad (\text{NŁ with } \xi = 1/2) \tag{101}$$

$$\leq -\frac{1}{2\beta} \cdot C^2 \cdot \delta(\theta_t), \qquad \left(C \coloneqq \inf_{t\geq 1} C(\theta_t) > 0\right) \tag{102}$$

which implies (similar to Eq. (73)),

$$f(\theta_t) - f(\theta^*) = \delta(\theta_t) \leq \exp\left\{-(t-1) \cdot C^2/(2\beta)\right\} \cdot (f(\theta_1) - f(\theta^*)) . \tag{103}$$

(2b) Third part: $O(1/t^{\frac{1}{1-2\xi}})$ upper bound for GNGD when $\xi < 1/2$.

According to Lemma 12, we have, for all $t \geq 1$,

$$\delta(\theta_{t+1}) - \delta(\theta_t) \leq -\frac{1}{2 \cdot \beta(\theta_t)} \cdot \|\nabla f(\theta_t)\|_2^2 \tag{104}$$

$$\leq -\frac{1}{2} \cdot \frac{\|\nabla f(\theta_t)\|_2^2}{L_0 \cdot \frac{\|\nabla f(\theta_t)\|^2}{\delta(\theta_t)^{2-2\xi}} + L_1 \cdot \|\nabla f(\theta_t)\|_2} \qquad \left(\beta(\theta_t) \leq L_0 \cdot \frac{\|\nabla f(\theta_t)\|^2}{\delta(\theta_t)^{2-2\xi}} + L_1 \cdot \|\nabla f(\theta_t)\|_2\right) \tag{105}$$

$$= -\frac{1}{2} \cdot \frac{\delta(\theta_t)^{2-2\xi}}{L_0 + L_1 \cdot \frac{\delta(\theta_t)^{2-2\xi}}{\|\nabla f(\theta_t)\|_2}} \tag{106}$$

$$\leq -\frac{1}{2} \cdot \frac{\delta(\theta_t)^{2-2\xi}}{L_0 + L_1 \cdot \frac{\delta(\theta_t)^{1-\xi}}{C(\theta_t)}} \qquad \left(\text{NŁ: } \|\nabla f(\theta_t)\|_2 \geq C(\theta_t) \cdot \delta(\theta_t)^{1-\xi}\right) \tag{107}$$

$$\leq -\frac{1}{2} \cdot \frac{\delta(\theta_t)^{2-2\xi}}{L_0 + L_1 \cdot \frac{\delta(\theta_t)^{1-\xi}}{C}} \qquad \left(C \coloneqq \inf_{t \geq 1} C(\theta_t) > 0\right) \tag{108}$$

$$\leq -\frac{1}{2} \cdot \frac{\delta(\theta_t)^{2-2\xi}}{L_0 + L_1 \cdot \frac{\delta(\theta_1)^{1-\xi}}{C}}, \qquad (\delta_{t+1} \leq \delta_t, \text{ by Eq. (69)}) \tag{109}$$

which is similar to Eq. (25). Using similar calculations in the first part of (1a), we have the $O(1/t^{\frac{1}{1-2\xi}})$ upper bound.

(2b) Fourth part: $O(e^{-t})$ upper bound for GNGD when $\xi = 1/2$.

We have, for all $t \geq 1$,

$$\delta(\theta_{t+1}) - \delta(\theta_t) \leq -\frac{1}{2} \cdot \frac{\delta(\theta_t)^{2-2\xi}}{L_0 + L_1 \cdot \frac{\delta(\theta_1)^{1-\xi}}{C}} \qquad (\delta_{t+1} \leq \delta_t \text{ by Eq. (104)}) \tag{110}$$

$$= -\frac{1}{2} \cdot \frac{\delta(\theta_t)}{L_0 + L_1 \cdot \frac{\delta(\theta_1)^{1/2}}{C}}, \qquad (\xi = 1/2) \tag{111}$$

which implies (similar to Eq. (73)),

$$f(\theta_t) - f(\theta^*) = \delta(\theta_t) \leq \exp\left\{-\frac{C \cdot (t-1)}{2 \cdot (L_0 \cdot C + L_1 \cdot \delta(\theta_1)^{1/2})}\right\} \cdot (f(\theta_1) - f(\theta^*)) \tag{112}$$

$$\leq \exp\left\{-\frac{C \cdot (t-1)}{2 \cdot (L_0 + L_1 \cdot \delta(\theta_1)^{1/2})}\right\} \cdot (f(\theta_1) - f(\theta^*))\,. \qquad (\text{if } C \leq 1) \tag{113}$$

(3a) $O(e^{-t})$ *upper bound for GNGD update when* $\xi \in (1/2, 1)$.

According to Lemma 12, we have, for all $t \geq 1$ (same as the third part of (1b)),

$$\delta(\theta_{t+1}) - \delta(\theta_t) \leq -\frac{1}{2 \cdot \beta(\theta_t)} \cdot \|\nabla f(\theta_t)\|_2^2 \tag{114}$$

$$\leq -\frac{1}{2 \cdot c} \cdot \|\nabla f(\theta_t)\|_2^{\frac{1}{1-\xi}} \qquad \left(\beta(\theta_t) \leq c \cdot \|\nabla f(\theta_t)\|_2^{\frac{1-2\xi}{1-\xi}}\right) \tag{115}$$

$$\leq -\frac{1}{2 \cdot c} \cdot C(\theta_t)^{\frac{1}{1-\xi}} \cdot \delta(\theta_t) \qquad (\text{NŁ}) \tag{116}$$

$$\leq -\frac{1}{2 \cdot c} \cdot C^{\frac{1}{1-\xi}} \cdot \delta(\theta_t) \qquad \left(C \coloneqq \inf_{t \geq 1} C(\theta_t) > 0\right), \tag{117}$$

which implies (similar to Eq. (73)),

$$f(\theta_t) - f(\theta^*) = \delta(\theta_t) \leq \exp\left\{-(t-1) \cdot C^{\frac{1}{1-\xi}}/(2 \cdot c)\right\} \cdot (f(\theta_1) - f(\theta^*))\,. \tag{118}$$

(3b) $O(e^{-t})$ *upper bound for GNGD update when* $\xi \in (1/2, 1)$. □

According to Lemma 12, we have, for all $t \geq 1$ (same as the third part of (1a)),

$$\delta(\theta_{t+1}) - \delta(\theta_t) \leq -\frac{1}{2 \cdot \beta(\theta_t)} \cdot \|\nabla f(\theta_t)\|_2^2 \tag{119}$$

$$\leq -\frac{1}{2 \cdot \beta(\theta_t)} \cdot C(\theta_t)^2 \cdot \delta(\theta_t)^{2-2\xi} \qquad (\text{NŁ}) \tag{120}$$

$$\leq -\frac{1}{2 \cdot \beta(\theta_t)} \cdot C^2 \cdot \delta(\theta_t)^{2-2\xi} \qquad \left(C \coloneqq \inf_{t \geq 1} C(\theta_t) > 0\right) \tag{121}$$

$$\leq -\frac{C^2}{2 \cdot c} \cdot \delta(\theta_t), \qquad \left(\beta(\theta_t) \leq c \cdot \delta(\theta_t)^{1-2\xi}\right) \tag{122}$$

which implies (similar to Eq. (73)),

$$f(\theta_t) - f(\theta^*) = \delta(\theta_t) \leq \exp\left\{-(t-1) \cdot C^2/(2 \cdot c)\right\} \cdot (f(\theta_1) - f(\theta^*)) \,. \tag{123}$$

A.2. Function Classes in Figure 2

Proposition 1. The following results hold:

(1) D $\subseteq$ C. If a function satisfies NŁ with degree ξ, then it satisfies NŁ with degree $\xi' < \xi$.

(2) F $\subseteq$ D. A strongly convex function satisfies NŁ with $\xi \geq 1/2$.

(3) F $\cap$ A $= \emptyset$. A strongly convex function cannot satisfy NS with $\beta(\theta) \to 0$ as $\theta, \theta' \to \theta^*$.

(4) E $\subseteq$ C. A (not strongly) convex function satisfies NŁ with $\xi < 1/2$.

Proof. (1) D $\subseteq$ C. Suppose a function $f : \Theta \to \mathbb{R}$ satisfies NŁ with ξ, i.e.,

$$\left\| \frac{df(\theta)}{d\theta} \right\|_2 \geq C(\theta) \cdot |f(\theta) - f(\theta^*)|^{1-\xi} \,, \tag{124}$$

where $\xi \in (-\infty, 1]$, and $C(\theta) > 0$ holds for all $\theta \in \Theta$. Let $\xi' < \xi$. If $|f(\theta) - f(\theta^*)| > 0$, then we have,

$$|f(\theta) - f(\theta^*)|^{1-\xi} = \frac{|f(\theta) - f(\theta^*)|^{1-\xi'}}{|f(\theta) - f(\theta^*)|^{\xi - \xi'}} \tag{125}$$

$$\geq c(\theta) \cdot |f(\theta) - f(\theta^*)|^{1-\xi'} \,, \tag{126}$$

where $c(\theta) \coloneqq \frac{1}{|f(\theta) - f(\theta^*)|^{\xi - \xi'}} > 0$, and $c(\theta) \not\to 0$ as $\theta \to \theta^*$ (or $c(\theta) > c > 0$ for all θ within a finite distance of θ^*). If $|f(\theta) - f(\theta^*)| = 0$, then it trivially holds that

$$|f(\theta) - f(\theta^*)|^{1-\xi} \geq |f(\theta) - f(\theta^*)|^{1-\xi'} \,. \tag{127}$$

(2) F $\subseteq$ D. Suppose a function $f : \Theta \to \mathbb{R}$ is strongly convex. We have, there exists $\mu > 0$, for all θ, $\theta' \in \Theta$,

$$f(\theta') \geq f(\theta) + \nabla f(\theta)^\top (\theta' - \theta) + \frac{\mu}{2} \cdot \|\theta' - \theta\|_2^2 \,. \tag{128}$$

Fix θ and take minimum over θ' on both sides of the above inequality. Then we have,

$$f(\theta^*) \geq f(\theta) + \min_{\theta'} \left\{ \nabla f(\theta)^\top (\theta' - \theta) + \frac{\mu}{2} \cdot \|\theta' - \theta\|_2^2 \right\} \tag{129}$$

$$= f(\theta) - \frac{1}{\mu} \cdot \|\nabla f(\theta)\|_2^2 + \frac{1}{2\mu} \cdot \|\nabla f(\theta)\|_2^2 \qquad \left(\theta' = \theta - \frac{1}{\mu} \cdot \nabla f(\theta)\right) \tag{130}$$

$$= f(\theta) - \frac{1}{2\mu} \cdot \|\nabla f(\theta)\|_2^2 \,, \tag{131}$$

which is equivalent to,

$$\|\nabla f(\theta)\|_2 \geq \sqrt{2\mu} \cdot (f(\theta) - f(\theta^*))^{\frac{1}{2}} \,, \tag{132}$$

which means f satisfies NŁ inequality with $\xi = 1/2$.

(3) $\mathrm{F} \cap \mathrm{A} = \emptyset$. Suppose a function $f : \Theta \to \mathbb{R}$ is strongly convex. There exists $\mu > 0$, for all $\theta \in \Theta$,

$$\left| z^\top \frac{\partial^2 f(\theta)}{\partial \theta^2} z \right| \geq \mu \cdot \|z\|_2^2 , \tag{133}$$

holds for all vector z that has the same dimension as θ. Next we show $f \notin \mathrm{A}$. Suppose $f \in \mathrm{A}$. We have,

$$\beta(\theta^*) = \sup_z \left| z^\top \frac{\partial^2 f(\theta^*)}{\partial (\theta^*)^2} z \right| = 0, \tag{134}$$

which is a contradiction with Eq. (133). Therefore $f \notin \mathrm{A}$, and $\mathrm{F} \cap \mathrm{A} = \emptyset$.

(4) $\mathrm{E} \subseteq \mathrm{C}$. Suppose a function $f : \Theta \to \mathbb{R}$ is convex. We have, for all $\theta, \theta' \in \Theta$,

$$f(\theta') \geq f(\theta) + \nabla f(\theta)^\top (\theta' - \theta). \tag{135}$$

Take $\theta' = \theta^*$. We have,

$$f(\theta^*) \geq f(\theta) + \nabla f(\theta)^\top (\theta^* - \theta), \tag{136}$$

which implies,

$$\|\nabla f(\theta)\|_2 = \frac{1}{\|\theta - \theta^*\|_2} \cdot \|\nabla f(\theta)\|_2 \cdot \|\theta - \theta^*\|_2 \tag{137}$$

$$\geq \frac{1}{\|\theta - \theta^*\|_2} \cdot \nabla f(\theta)^\top (\theta^* - \theta) \qquad \text{(by Cauchy-Schwarz)} \tag{138}$$

$$\geq \frac{1}{\|\theta - \theta^*\|_2} \cdot \left(f(\theta) - f(\theta^*) \right), \qquad \text{(by Eq. (136))} \tag{139}$$

and $C(\theta) = \frac{1}{\|\theta - \theta^*\|_2} \not\to 0$ as $\theta \to \theta^*$ (or $C(\theta) > c > 0$ for all $\|\theta - \theta^*\|_2$ smaller than a finite value, e.g., within a bounded constraint). Therefore f satisfies NŁ inequality with $\xi = 0$. □

Proposition 2.

(1) $\mathrm{ACE} \neq \emptyset$. There exists at least one (not strongly) convex function which satisfies NŁ with $\xi < 1/2$ and NS with $\beta(\theta) \to 0$ as $\theta, \theta' \to \theta^*$.

(2) $\mathrm{ADE} \neq \emptyset$. There exists at least one (not strongly) convex function which satisfies NŁ with $\xi \geq 1/2$ and NS with $\beta(\theta) \to 0$ as $\theta, \theta' \to \theta^*$.

(3) $\mathrm{BCE} \neq \emptyset$. There exists at least one (not strongly) convex function which satisfies NŁ with $\xi < 1/2$ and NS with $\beta(\theta) \to \beta > 0$ as $\theta, \theta' \to \theta^*$.

(4) $\mathrm{BDE} \neq \emptyset$. There exists at least one (not strongly) convex function which satisfies NŁ with $\xi \geq 1/2$ and NS with $\beta(\theta) \to \beta > 0$ as $\theta, \theta' \to \theta^*$.

(5) $\mathrm{BF} \neq \emptyset$. There exists at least one strongly convex function which satisfies NS with $\beta(\theta) \to \beta > 0$ as $\theta, \theta' \to \theta^*$.

Proof. (1) $\mathrm{ACE} \neq \emptyset$. Consider minimizing the following function $f : \mathbb{R} \to \mathbb{R}$,

$$f(x) = x^4. \tag{140}$$

The second order derivative (Hessian) is $f''(x) = 12 \cdot x^2 \geq 0$, which means f is (not strongly) convex. According to Taylor's theorem, we have, for all $x, x' \in \mathbb{R}$,

$$\left| f(x') - f(x) - \left\langle \frac{df(x)}{dx}, x' - x \right\rangle \right| \leq \frac{|f''(x_\zeta)|}{2} \cdot \|\theta' - \theta\|_2^2 \tag{141}$$

$$= \frac{12 \cdot x_\zeta^2}{2} \cdot \|\theta' - \theta\|_2^2, \tag{142}$$

where $x_\zeta \coloneqq x + \zeta \cdot (x' - x)$ with some $\zeta \in [0, 1]$. Thus we have $\beta(x) = 12 \cdot x_\zeta^2 \to 0$ as $x, x' \to 0$. Next, we have,

$$|f'(x)| = \left| 4 \cdot x^3 \right| = 4 \cdot \left(|x|^4 \right)^{\frac{3}{4}} = 4 \cdot \left(f(x) - f(0) \right)^{1 - \frac{1}{4}}, \tag{143}$$

which means f satisfies NŁ inequality with $\xi = 1/4 < 1/2$.

(2) ADE $\neq \emptyset$. Consider minimizing the following function $f : \mathbb{R}^K \to \mathbb{R}$,

$$f(\theta) = D_{\mathrm{KL}}(y\|\pi_\theta) = D_{\mathrm{KL}}(y\|\mathrm{softmax}(\theta)), \tag{144}$$

where $y \in \{0,1\}^K$ is a one-hot vector. We show that f is a (not strongly) convex function. The gradient of f is,

$$\frac{\partial f(\theta)}{\partial \theta} = \left(\frac{d\pi_\theta}{d\theta}\right)^\top \left(\frac{d\{D_{\mathrm{KL}}(y\|\pi_\theta)\}}{d\pi_\theta}\right) \tag{145}$$

$$= \left(\mathrm{diag}(\pi_\theta) - \pi_\theta\pi_\theta^\top\right)\mathrm{diag}\left(\frac{1}{\pi_\theta}\right)(-y) \tag{146}$$

$$= \pi_\theta - y. \tag{147}$$

Therefore the Hessian is,

$$\frac{\partial^2 f(\theta)}{\partial \theta^2} = \frac{d\pi_\theta}{d\theta} = \mathrm{diag}(\pi_\theta) - \pi_\theta\pi_\theta^\top. \tag{148}$$

According to Mei et al. (2020b, Lemma 22), we have,

$$\mathrm{diag}(\pi_\theta) - \pi_\theta\pi_\theta^\top \succeq \mathbf{0}, \tag{149}$$

and the minimum eigenvalue of $\mathrm{diag}(\pi_\theta) - \pi_\theta\pi_\theta^\top$ is 0, which means f is convex but not strongly convex. Next, according to Mei et al. (2020a, Lemma 17), we have,

$$D_{\mathrm{KL}}(y\|\pi_\theta) = \sum_a y(a)\cdot\log\left(\frac{y(a)}{\pi_\theta(a)}\right) \tag{150}$$

$$\leq \sum_a y(a)\cdot\left(\frac{y(a)}{\pi_\theta(a)} - 1\right) \qquad (\log x \leq x-1) \tag{151}$$

$$= \sum_a \left(y(a) - \pi_\theta(a) + \pi_\theta(a)\right)\cdot\frac{y(a)-\pi_\theta(a)}{\pi_\theta(a)} \tag{152}$$

$$= \sum_a \frac{\left(y(a)-\pi_\theta(a)\right)^2}{\pi_\theta(a)} \tag{153}$$

$$\leq \frac{1}{\min_a \pi_\theta(a)}\cdot\sum_a \left(y(a)-\pi_\theta(a)\right)^2, \tag{154}$$

which implies,

$$\left\|\frac{\partial f(\theta)}{\partial\theta}\right\|_2 = \|\pi_\theta - y\|_2 \qquad \text{(by Eq. (145))} \tag{155}$$

$$\geq \min_a \sqrt{\pi_\theta(a)}\cdot\left[D_{\mathrm{KL}}(y\|\pi_\theta) - D_{\mathrm{KL}}(y\|y)\right]^{\frac{1}{2}}, \qquad \text{(by Eq. (150))} \tag{156}$$

which means f satisfies NŁ with $\xi = 1/2$. Denote $\theta_\zeta \coloneqq \theta + \zeta\cdot(\theta'-\theta)$ with some $\zeta \in [0,1]$. We have, as $\pi_\theta, \pi_{\theta'} \to y$,

$$\beta(\theta) = \sup_z \left|z^\top \frac{\partial^2 f(\theta_\zeta)}{\partial\theta_\zeta^2} z\right| \tag{157}$$

$$= \sup_z \left|z^\top\left(\mathrm{diag}(\pi_{\theta_\zeta}) - \pi_{\theta_\zeta}\pi_{\theta_\zeta}^\top\right)z\right| \qquad \text{(by Eq. (148))} \tag{158}$$

$$\to \sup_z \left|z^\top\left(\mathrm{diag}(y) - yy^\top\right)z\right| \tag{159}$$

$$= \sup_z \left|z^\top\mathbf{0}z\right| \qquad (y \text{ is one-hot}) \tag{160}$$

$$= 0. \tag{161}$$

(3) BCE $\neq \emptyset$. Consider the (modified) Huber loss function,

$$f(x) = \begin{cases} x^2, & \text{if } |x| \leq 1, \\ 2\cdot|x| - 1, & \text{otherwise} \end{cases} \tag{162}$$

which is a (not strongly) convex function. According to (4) in Proposition 1, f satisfies NŁ inequality with $\xi = 0$. Denote $x_\zeta \coloneqq x + \zeta\cdot(x'-x)$ with some $\zeta \in [0,1]$. We have $\beta(x) = |f''(x_\zeta)| \to 2 > 0$, as $x, x' \to 0$.

(4) BDE $\neq \emptyset$. Consider minimizing the same function as in (2),

$$f(\theta) = D_{\mathrm{KL}}(y\|\pi_\theta) = D_{\mathrm{KL}}(y\|\mathrm{softmax}(\theta)), \tag{163}$$

where $y \in (0,1)^K$ is a probability vector with $\min_a y(a) > 0$, i.e., y is bounded away from the boundary of probability simplex. As shown in (2), f is (not strongly) convex and f satisfies NŁ with $\xi = 1/2$. Next, we have,

$$\begin{aligned}
\beta(\theta) &= \sup_z \left| z^\top \left(\mathrm{diag}(\pi_{\theta_\zeta}) - \pi_{\theta_\zeta}\pi_{\theta_\zeta}^\top \right) z \right| \qquad \text{(by Eq. (148))} && (164)\\
&\to \sup_z \left| z^\top \left(\mathrm{diag}(y) - yy^\top \right) z \right| && (165)\\
&= \sup_z \left| \mathop{\mathbb{E}}_{a\sim y}[z(a)^2] - \left(\mathop{\mathbb{E}}_{a\sim y}[z(a)] \right)^2 \right| && (166)\\
&= \sup_z \left| \mathrm{Var}_{a\sim y}[z(a)] \right| > 0. && (167)
\end{aligned}$$

(5) BF $\neq \emptyset$. Consider minimizing the following function,

$$f(x) = x^2, \tag{168}$$

where $x \in \mathbb{R}$. f is strongly convex, and $\beta(x) = \beta = 2$. Thus $\beta(x) \to 2 > 0$ as $x, x' \to 0$ in Definition 3. □

Proposition 3. The following results hold:

(1) $\mathcal{W} \coloneqq \mathrm{AC} \setminus (\mathrm{AD} \cup \mathrm{ACE}) \neq \emptyset$. There exists at least one non-convex function which satisfies NŁ with $\xi < 1/2$ and NS with $\beta(\theta) \to 0$ as $\theta, \theta' \to \theta^*$.

(2) $\mathcal{X} \coloneqq \mathrm{AD} \setminus \mathrm{ADE} \neq \emptyset$. There exists at least one non-convex function which satisfies NŁ with $\xi \geq 1/2$ and NS with $\beta(\theta) \to 0$ as $\theta, \theta' \to \theta^*$.

(3) $\mathcal{Y} \coloneqq \mathrm{BC} \setminus (\mathrm{BD} \cup \mathrm{BCE}) \neq \emptyset$. There exists at least one non-convex function which satisfies NŁ with $\xi < 1/2$ and NS with $\beta(\theta) \to \beta > 0$ as $\theta, \theta' \to \theta^*$.

(4) $\mathcal{Z} \coloneqq \mathrm{BD} \setminus (\mathrm{BDE} \cup \mathrm{BF}) \neq \emptyset$. There exists at least one non-convex function which satisfies NŁ with $\xi \geq 1/2$ and NS with $\beta(\theta) \to \beta > 0$ as $\theta, \theta' \to \theta^*$.

Proof. (1) $\mathcal{W} \coloneqq \mathrm{AC} \setminus (\mathrm{AD} \cup \mathrm{ACE}) \neq \emptyset$. Consider maximizing the expected reward,

$$f(\theta) = \pi_\theta^\top r, \tag{169}$$

where $\pi_\theta = \mathrm{softmax}(\theta)$ and $\theta \in \mathbb{R}^K$. According to Mei et al. (2020b, Proposition 1), f is non-concave. According to Lemma 1, we have,

$$\left\| \frac{d\pi_\theta^\top r}{d\theta} \right\|_2 \geq \pi_\theta(a^*) \cdot (\pi^* - \pi_\theta)^\top r, \tag{170}$$

which means f satisfies NŁ inequality with $\xi = 0$. As shown in Lemma 2, we have $\beta(\theta_\zeta) = 3 \cdot \left\| \frac{d\pi_{\theta_\zeta}^\top r}{d\theta_\zeta} \right\|_2$. Therefore, $\beta(\theta_\zeta) \to 0$ as $\pi_\theta, \pi_{\theta'} \to \pi^*$.

(2) $\mathcal{X} \coloneqq \mathrm{AD} \setminus \mathrm{ADE} \neq \emptyset$. Consider minimizing the function $f : \mathbb{R}^K \to \mathbb{R}$,

$$f(\theta) = \|\pi_\theta - y\|_2^2, \tag{171}$$

where $\pi_\theta = \mathrm{softmax}(\theta)$, $\theta \in \mathbb{R}^K$, and $y \in \{0,1\}$ is a one-hot vector. We show that f is non-convex using one example. Let $y = (1,0,0)^\top$. Let $\theta_1 = (0,0,0)^\top$, $\pi_{\theta_1} = \mathrm{softmax}(\theta_1) = (1/3, 1/3, 1/3)^\top$, $\theta_2 = (\log 4, \log 36, \log 100)^\top$, and $\pi_{\theta_2} = \mathrm{softmax}(\theta_2) = (4/140, 36/140, 100/140)^\top$. We have,

$$f(\theta_1) = \|\pi_{\theta_1} - y\|_2^2 = \frac{2}{3}, \text{ and } f(\theta_2) = \|\pi_{\theta_2} - y\|_2^2 = \frac{38}{25}. \tag{172}$$

Denote $\bar{\theta} = \frac{1}{2} \cdot (\theta_1 + \theta_2) = (\log 2, \log 6, \log 10)^\top$ we have $\pi_{\bar{\theta}} = \mathrm{softmax}(\bar{\theta}) = (2/18, 6/18, 10/18)^\top$ and

$$f(\bar{\theta}) = \|\pi_{\bar{\theta}} - y\|_2^2 = \frac{98}{81}. \tag{173}$$

Therefore we have,

$$\frac{1}{2}\cdot(f(\theta_1)+f(\theta_2))=\frac{82}{75}=\frac{2214}{2025}<\frac{2450}{2025}=\frac{98}{81}=f(\bar{\theta}), \tag{174}$$

which means f is non-convex. Denote $H(\pi_\theta)\coloneqq \mathrm{diag}(\pi_\theta)-\pi_\theta\pi_\theta^\top$ as the Jacobian of $\theta\mapsto \mathrm{softmax}(\theta)$. We have,

$$\left\|\frac{\partial f(\theta)}{\partial\theta}\right\|_2=\left\|\left(\frac{d\pi_\theta}{d\theta}\right)^\top\left(\frac{df(\theta)}{d\pi_\theta}\right)\right\|_2 \tag{175}$$

$$=2\cdot\left\|H(\pi_\theta)\left(\pi_\theta-y\right)\right\|_2 \tag{176}$$

$$\geq 2\cdot\min_a\pi_\theta(a)\cdot\left\|\pi_\theta-y\right\|_2 \qquad \text{(by Mei et al. (2020b, Lemma 23))} \tag{177}$$

$$=2\cdot\min_a\pi_\theta(a)\cdot\left[f(\theta)-f(y)\right]^{\frac{1}{2}}, \tag{178}$$

which means f satisfies NŁ inequality with $\xi=1/2$. Denote $S\coloneqq S(y,\theta)\in\mathbb{R}^{K\times K}$ as the second derivative (Hessian) of f. We have,

$$S=\frac{d}{d\theta}\left\{\frac{df(\theta)}{d\theta}\right\} \tag{179}$$

$$=\frac{d}{d\theta}\left\{H(\pi_\theta)\left(\pi_\theta-y\right)\right\}. \tag{180}$$

Continuing with our calculation fix $i,j\in[K]$. Then,

$$S_{(i,j)}=\frac{d\{\pi_\theta(i)\cdot\left[\pi_\theta(i)-y(i)-\pi_\theta^\top\left(\pi_\theta-y\right)\right]\}}{d\theta(j)} \tag{181}$$

$$=\frac{d\pi_\theta(i)}{d\theta(j)}\cdot\left[\pi_\theta(i)-y(i)-\pi_\theta^\top\left(\pi_\theta-y\right)\right]+\pi_\theta(i)\cdot\frac{d\{\pi_\theta(i)-y(i)-\pi_\theta^\top\left(\pi_\theta-y\right)\}}{d\theta(j)} \tag{182}$$

$$=\left(\delta_{ij}\pi_\theta(j)-\pi_\theta(i)\pi_\theta(j)\right)\cdot\left[\pi_\theta(i)-y(i)-\pi_\theta^\top\left(\pi_\theta-y\right)\right] \tag{183}$$

$$+\pi_\theta(i)\cdot\left[\delta_{ij}\pi_\theta(j)-\pi_\theta(i)\pi_\theta(j)-\pi_\theta(j)\cdot\left(\pi_\theta(j)-y(j)-\pi_\theta^\top\left(\pi_\theta-y\right)\right)-\pi_\theta(j)\cdot\left(\pi_\theta(j)-\pi_\theta^\top\pi_\theta\right)\right] \tag{184}$$

$$=\delta_{ij}\pi_\theta(j)\cdot\left[\pi_\theta(i)-y(i)-\pi_\theta^\top\left(\pi_\theta-y\right)\right]-\pi_\theta(i)\pi_\theta(j)\cdot\left[\pi_\theta(i)-y(i)-\pi_\theta^\top\left(\pi_\theta-y\right)\right] \tag{185}$$

$$-\pi_\theta(i)\pi_\theta(j)\cdot\left[\pi_\theta(j)-y(j)-\pi_\theta^\top\left(\pi_\theta-y\right)\right]+\pi_\theta(i)\pi_\theta(j)\cdot\left[\delta_{ij}-\pi_\theta(i)-\pi_\theta(j)+\pi_\theta^\top\pi_\theta\right], \tag{186}$$

where

$$\delta_{ij}=\begin{cases}1, & \text{if } i=j,\\ 0, & \text{otherwise}\end{cases} \tag{187}$$

is Kronecker's δ-function. To show the bound on the spectral radius of S, pick $z\in\mathbb{R}^K$. Then,

$$\left|z^\top Sz\right|=\left|\sum_{i=1}^K\sum_{j=1}^K S_{(i,j)}\cdot z(i)\cdot z(j)\right| \tag{188}$$

$$=\Big|\left(H(\pi_\theta)\left(\pi_\theta-y\right)\right)^\top\left(z\odot z\right)-2\cdot\left(H(\pi_\theta)\left(\pi_\theta-y\right)\right)^\top z\cdot\left(\pi_\theta^\top z\right) \tag{189}$$

$$+\left(\pi_\theta\odot\pi_\theta\right)^\top\left(z\odot z\right)-2\cdot\left(\pi_\theta\odot\pi_\theta\right)^\top z\cdot\left(\pi_\theta^\top z\right)+\left(\pi_\theta^\top z\right)^2\cdot\left(\pi_\theta^\top\pi_\theta\right)\Big|, \tag{190}$$

where $\odot$ is Hadamard (component-wise) product. We have, as $\pi_\theta\to y$,

$$\left(H(\pi_\theta)\left(\pi_\theta-y\right)\right)^\top\left(z\odot z\right)-2\cdot\left(H(\pi_\theta)\left(\pi_\theta-y\right)\right)^\top z\cdot\left(\pi_\theta^\top z\right)\to\left(H(y)\mathbf{0}\right)^\top\left(z\odot z\right)-2\cdot\left(H(y)\mathbf{0}\right)^\top z\cdot\left(y^\top z\right) \tag{191}$$

$$=0. \tag{192}$$

Since y is one-hot vector, we have, as $\pi_\theta\to y$,

$$\left(\pi_\theta\odot\pi_\theta\right)^\top\left(z\odot z\right)-2\cdot\left(\pi_\theta\odot\pi_\theta\right)^\top z\cdot\left(\pi_\theta^\top z\right)+\left(\pi_\theta^\top z\right)^2\cdot\pi_\theta^\top\pi_\theta\to y^\top\left(z\odot z\right)-2\cdot\left(y^\top z\right)^2+\left(y^\top z\right)^2\cdot y^\top y \tag{193}$$

$$=\left(y^\top z\right)^2-2\cdot\left(y^\top z\right)^2+\left(y^\top z\right)^2=0, \tag{194}$$

which means $\beta(\theta)\to 0$ as $\theta,\theta'\to\theta^*$ in Definition 3.

(3) $\mathcal{Y} := \text{BC} \setminus (\text{BD} \cup \text{BCE}) \neq \emptyset$. Consider minimizing the function $f : \mathbb{R} \to \mathbb{R}$,

$$f(\theta) = \begin{cases} 2 \cdot (\pi_\theta - \pi_{\theta^*})^2, & \text{if } |\pi_\theta - \pi_{\theta^*}| \leq 1, \\ (\pi_\theta - \pi_{\theta^*})^4 + 1, & \text{otherwise} \end{cases} \tag{195}$$

where $\theta \in \mathbb{R}$, $\theta^* = 0$, and π_θ is defined as,

$$\pi_\theta = \sigma(\theta) = \frac{1}{1 + e^{-\theta}}, \tag{196}$$

where $\sigma : \mathbb{R} \to (0, 1)$ is the sigmoid activation. Figure 6 shows the image of f, indicating that f is a non-convex function.

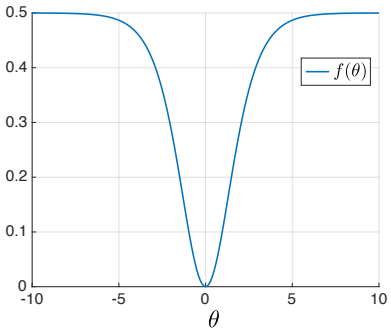


Figure 6. The image of f.

Since $\theta^* = 0$, we have $\pi_{\theta^*} = 1/2$, and for all $|\pi_\theta - \pi_{\theta^*}| > 1$,

$$\left|\frac{df(\theta)}{d\theta}\right| = \left|\frac{d\pi_\theta}{d\theta} \cdot \frac{df(\theta)}{d\pi_\theta}\right| \tag{197}$$

$$= \left|\pi_\theta \cdot (1 - \pi_\theta) \cdot 4 \cdot (\pi_\theta - \pi_{\theta^*})^3\right| \tag{198}$$

$$= 4 \cdot \pi_\theta \cdot (1 - \pi_\theta) \cdot \left[(\pi_\theta - \pi_{\theta^*})^4\right]^{\frac{3}{4}} \tag{199}$$

$$= 4 \cdot \pi_\theta \cdot (1 - \pi_\theta) \cdot [f(\theta) - f(\theta^*)]^{1 - \frac{1}{4}}, \tag{200}$$

which means f satisfies NŁ inequality with $\xi = 1/4 < 1/2$. For all $|\pi_\theta - \pi_{\theta^*}| \leq 1$, the Hessian of f is,

$$\left|\frac{d^2 f(\theta)}{d\theta^2}\right| = \left|\frac{d}{d\theta}\left\{4 \cdot \pi_\theta \cdot (1 - \pi_\theta) \cdot (\pi_\theta - \pi_{\theta^*})\right\}\right| \tag{201}$$

$$= \left|4 \cdot \pi_\theta \cdot (1 - \pi_\theta) \cdot (\pi_\theta - \pi_{\theta^*}) \cdot (1 - 2\pi_\theta) + 4 \cdot \pi_\theta^2 \cdot (1 - \pi_\theta)^2\right|. \tag{202}$$

As $\pi_\theta \to \pi_{\theta^*} = 1/2$, we have

$$4 \cdot \pi_\theta \cdot (1 - \pi_\theta) \cdot (\pi_\theta - \pi_{\theta^*}) \cdot (1 - 2\pi_\theta) \to 0, \tag{203}$$

and,

$$4 \cdot \pi_\theta^2 \cdot (1 - \pi_\theta)^2 \to 4 \cdot \frac{1}{4} \cdot \frac{1}{4} = \frac{1}{4} > 0, \tag{204}$$

which means $\beta(\theta) \to \beta > 0$ as $\theta, \theta' \to \theta^*$ in Definition 3.

(4) $\mathcal{Z} := \text{BD} \setminus (\text{BDE} \cup \text{BF}) \neq \emptyset$. Consider minimizing the same function as in (2),

$$f(\theta) = \|\pi_\theta - y\|_2^2, \tag{205}$$

where $\pi_\theta = \text{softmax}(\theta)$, $\theta \in \mathbb{R}^K$, and $y \in (0, 1)$ is a probability vector with $\min_a y(a) > 0$, i.e., y is bounded away from the boundary of probability simplex. We show that f is non-convex using one example. Let $y = (1/2, 1/4, 1/4)^\top$. Let $\theta_1 = (0, 0, 0)^\top$, $\pi_{\theta_1} = \text{softmax}(\theta_1) = (1/3, 1/3, 1/3)^\top$, $\theta_2 = (\log 4, \log 36, \log 100)^\top$, and $\pi_{\theta_2} = \text{softmax}(\theta_2) = (4/140, 36/140, 100/140)^\top$. We have,

$$f(\theta_1) = \|\pi_{\theta_1} - y\|_2^2 = \frac{1}{24}, \text{ and } f(\theta_2) = \|\pi_{\theta_2} - y\|_2^2 = \frac{613}{1400}. \tag{206}$$

Denote $\bar{\theta} = \frac{1}{2} \cdot (\theta_1 + \theta_2) = (\log 2, \log 6, \log 10)^\top$ we have $\pi_{\bar{\theta}} = \mathrm{softmax}(\bar{\theta}) = (2/18, 6/18, 10/18)^\top$ and

$$f(\bar{\theta}) = \|\pi_{\bar{\theta}} - y\|_2^2 = \frac{163}{648}. \tag{207}$$

Therefore we have,

$$\frac{1}{2} \cdot (f(\theta_1) + f(\theta_2)) = \frac{1007}{4200} = \frac{27189}{113400} < \frac{28525}{113400} = \frac{163}{648} = f(\bar{\theta}), \tag{208}$$

which means f is non-convex. Similar as (2), we have the Hessian of f,

$$S_{(i,j)} = \underbrace{\delta_{ij}\pi_\theta(j) \cdot \left[\pi_\theta(i) - y(i) - \pi_\theta^\top (\pi_\theta - y)\right]}_{\text{(a)}} - \underbrace{\pi_\theta(i)\pi_\theta(j) \cdot \left[\pi_\theta(i) - y(i) - \pi_\theta^\top (\pi_\theta - y)\right]}_{\text{(b)}} \tag{209}$$

$$- \underbrace{\pi_\theta(i)\pi_\theta(j) \cdot \left[\pi_\theta(j) - y(j) - \pi_\theta^\top (\pi_\theta - y)\right]}_{\text{(c)}} + \underbrace{\pi_\theta(i)\pi_\theta(j) \cdot \left[\delta_{ij} - \pi_\theta(i) - \pi_\theta(j) + \pi_\theta^\top \pi_\theta\right]}_{\text{(d)}}, \tag{210}$$

where $(a) = (b) = (c) = 0$ when $\pi_\theta = y$. Hence, at the optimal point θ^*, we have,

$$S = \frac{1}{128} \cdot \begin{bmatrix} 12 & -6 & -6 \\ -6 & 7 & -1 \\ -6 & -1 & 7 \end{bmatrix}, \tag{211}$$

and the eigenvalues of S are 0, $\frac{1}{16}$, and $\frac{9}{64}$. Thus as $\theta, \theta' \to \theta^*$, the Hessian spectral radius of f satisfies $\beta(\theta) \to \beta = \frac{9}{64}$. □

Proposition 4. The convex function $f : x \mapsto |x|^p$ with $p > 1$ satisfies the NŁ inequality with $\xi = 1/p$ and the NS property with $\beta(x) \le c_1 \cdot \delta(x)^{1-2\xi}$.

Proof. For $p > 1$, f is differentiable, and we have,

$$|f'(x)| = \left|p \cdot |x|^{p-1} \cdot \mathrm{sign}\{x\}\right| = p \cdot (|x|^p)^{\frac{p-1}{p}} = p \cdot (f(x) - f(0))^{1-\frac{1}{p}}, \tag{212}$$

which means f satisfies NŁ inequality with $\xi = 1/p$. On the other hand, the Hessian of f is,

$$|f''(x)| = \left|p \cdot (p-1) \cdot |x|^{p-2}\right| = p \cdot (p-1) \cdot (|x|^p)^{\frac{p-2}{p}} = p \cdot (p-1) \cdot (f(x) - f(0))^{1-\frac{2}{p}}. \qquad \square$$

B. Proofs for Section 5

B.1. One-state MDPs

Lemma 1 (NŁ) **.** Let a^* be the uniqe optimal action. Denote $\pi^* = \arg\max_{\pi \in \Delta} \pi^\top r$. Then,

$$\left\| \frac{d\pi_\theta^\top r}{d\theta} \right\|_2 \ge \pi_\theta(a^*) \cdot (\pi^* - \pi_\theta)^\top r. \tag{213}$$

Proof. See the proof in (Mei et al., 2020b, Lemma 3). We include a proof for completeness.

Using the expression of the policy gradient,

$$\begin{aligned} \left\| \frac{d\pi_\theta^\top r}{d\theta} \right\|_2 &= \left(\sum_{a \in \mathcal{A}} \left[\pi_\theta(a) \cdot (r(a) - \pi_\theta^\top r)\right]^2 \right)^{\frac{1}{2}} \\ &\ge \pi_\theta(a^*) \cdot (r(a^*) - \pi_\theta^\top r). \end{aligned} \tag{214}$$

□

Lemma 2 (NS) **.** Let $\pi_\theta = \mathrm{softmax}(\theta)$ and $\pi_{\theta'} = \mathrm{softmax}(\theta')$. Denote $\theta_\zeta \coloneqq \theta + \zeta \cdot (\theta' - \theta)$ with some $\zeta \in [0, 1]$. For any $r \in [0, 1]^K$, $\theta \mapsto \pi_\theta^\top r$ is $\beta(\theta_\zeta)$ non-uniform smooth with $\beta(\theta_\zeta) = 3 \cdot \left\| \frac{d\pi_{\theta_\zeta}^\top r}{d\theta_\zeta} \right\|_2$.

Proof. Let $S \coloneqq S(r, \theta) \in \mathbb{R}^{K \times K}$ be the second derivative of the value map $\theta \mapsto \pi_\theta^\top r$. By Taylor's theorem, it suffices to show that the spectral radius of S is upper bounded. Denote $H(\pi_\theta) \coloneqq \mathrm{diag}(\pi_\theta) - \pi_\theta \pi_\theta^\top$ as the Jacobian of $\theta \mapsto \mathrm{softmax}(\theta)$.

Now, by its definition we have

$$S = \frac{d}{d\theta}\left\{\frac{d\pi_\theta^\top r}{d\theta}\right\} \tag{215}$$

$$= \frac{d}{d\theta}\left\{H(\pi_\theta) r\right\} \tag{216}$$

$$= \frac{d}{d\theta}\left\{(\mathrm{diag}(\pi_\theta) - \pi_\theta \pi_\theta^\top) r\right\}. \tag{217}$$

Continuing with our calculation fix $i, j \in [K]$. Then,

$$S_{(i,j)} = \frac{d\{\pi_\theta(i) \cdot (r(i) - \pi_\theta^\top r)\}}{d\theta(j)} \tag{218}$$

$$= \frac{d\pi_\theta(i)}{d\theta(j)} \cdot (r(i) - \pi_\theta^\top r) + \pi_\theta(i) \cdot \frac{d\{r(i) - \pi_\theta^\top r\}}{d\theta(j)} \tag{219}$$

$$= (\delta_{ij}\pi_\theta(j) - \pi_\theta(i)\pi_\theta(j)) \cdot (r(i) - \pi_\theta^\top r) - \pi_\theta(i) \cdot (\pi_\theta(j) r(j) - \pi_\theta(j)\pi_\theta^\top r) \tag{220}$$

$$= \delta_{ij}\pi_\theta(j) \cdot (r(i) - \pi_\theta^\top r) - \pi_\theta(i)\pi_\theta(j) \cdot (r(i) - \pi_\theta^\top r) - \pi_\theta(i)\pi_\theta(j) \cdot (r(j) - \pi_\theta^\top r), \tag{221}$$

where

$$\delta_{ij} = \begin{cases} 1, & \text{if } i = j, \\ 0, & \text{otherwise} \end{cases} \tag{222}$$

is Kronecker's δ-function as defined in Eq. (187). To show the bound on the spectral radius of S, pick $y \in \mathbb{R}^K$. Then,

$$\left|y^\top S y\right| = \left|\sum_{i=1}^K \sum_{j=1}^K S_{(i,j)} \cdot y(i) \cdot y(j)\right| \tag{223}$$

$$= \left|\sum_i \pi_\theta(i)(r(i) - \pi_\theta^\top r) y(i)^2 - 2 \sum_i \pi_\theta(i)(r(i) - \pi_\theta^\top r) y(i) \sum_j \pi_\theta(j) y(j)\right| \tag{224}$$

$$= \left|\left(H(\pi_\theta) r\right)^\top (y \odot y) - 2 \cdot \left(H(\pi_\theta) r\right)^\top y \cdot \left(\pi_\theta^\top y\right)\right| \tag{225}$$

$$\leq \left\|H(\pi_\theta) r\right\|_\infty \cdot \|y \odot y\|_1 + 2 \cdot \left\|H(\pi_\theta) r\right\|_2 \cdot \|y\|_2 \cdot \|\pi_\theta\|_1 \cdot \|y\|_\infty \tag{226}$$

$$\leq 3 \cdot \left\|H(\pi_\theta) r\right\|_2 \cdot \|y\|_2^2. \tag{227}$$

According to Taylor's theorem, $\forall \theta,\ \theta'$,

$$\left|(\pi_{\theta'} - \pi_\theta)^\top r - \left\langle \frac{d\pi_\theta^\top r}{d\theta}, \theta' - \theta \right\rangle\right| = \frac{1}{2} \cdot \left|(\theta' - \theta)^\top S(r, \theta_\zeta)\,(\theta' - \theta)\right| \tag{228}$$

$$\leq \frac{3}{2} \cdot \left\|H(\pi_{\theta_\zeta}) r\right\|_2 \cdot \|\theta' - \theta\|_2^2 \qquad \text{(by Eq. (223))} \tag{229}$$

$$= \frac{3}{2} \cdot \left\|\frac{d\pi_{\theta_\zeta}^\top r}{d\theta_\zeta}\right\|_2 \cdot \|\theta' - \theta\|_2^2. \qquad \text{(by Lemma 16)} \qquad \square$$

Lemma 3. Let

$$\theta' = \theta + \eta \cdot \frac{d\pi_\theta^\top r}{d\theta} \Big/ \left\|\frac{d\pi_\theta^\top r}{d\theta}\right\|_2. \tag{230}$$

Denote $\theta_\zeta \coloneqq \theta + \zeta \cdot (\theta' - \theta)$ with some $\zeta \in [0, 1]$. We have, for all $\eta \in (0, 1/3)$,

$$\left\|\frac{d\pi_{\theta_\zeta}^\top r}{d\theta_\zeta}\right\|_2 \leq \frac{1}{1 - 3\eta} \cdot \left\|\frac{d\pi_\theta^\top r}{d\theta}\right\|_2. \tag{231}$$

Proof. Denote $\zeta_1 \coloneqq \zeta$. Also denote $\theta_{\zeta_2} \coloneqq \theta + \zeta_2 \cdot (\theta_{\zeta_1} - \theta)$ with some $\zeta_2 \in [0,1]$. We have,

$$\left\| \frac{d\pi_{\theta_{\zeta_1}}^\top r}{d\theta_{\zeta_1}} - \frac{d\pi_\theta^\top r}{d\theta} \right\|_2 = \left\| \int_0^1 \left\langle \frac{d^2\{\pi_{\theta_{\zeta_2}}^\top r\}}{d\theta_{\zeta_2}^2}, \theta_{\zeta_1} - \theta \right\rangle d\zeta_2 \right\|_2 \tag{232}$$

$$\leq \int_0^1 \left\| \frac{d^2\{\pi_{\theta_{\zeta_2}}^\top r\}}{d\theta_{\zeta_2}^2} \right\|_2 \cdot \|\theta_{\zeta_1} - \theta\|_2 \, d\zeta_2 \tag{233}$$

$$\leq \int_0^1 3 \cdot \left\| \frac{d\pi_{\theta_{\zeta_2}}^\top r}{d\theta_{\zeta_2}} \right\|_2 \cdot \zeta_1 \cdot \|\theta' - \theta\|_2 \, d\zeta_2 \qquad \text{(by Eq. (223))} \tag{234}$$

$$\leq \int_0^1 3 \cdot \left\| \frac{d\pi_{\theta_{\zeta_2}}^\top r}{d\theta_{\zeta_2}} \right\|_2 \cdot \eta \, d\zeta_2, \qquad \left(\zeta_1 \in [0,1], \text{ using } \theta' = \theta + \eta \cdot \frac{d\pi_\theta^\top r}{d\theta} \Big/ \left\| \frac{d\pi_\theta^\top r}{d\theta} \right\|_2 \right) \tag{235}$$

where the second last inequality is because of the Hessian is symmetric, and its operator norm is equal to its spectral radius. Therefore we have,

$$\left\| \frac{d\pi_{\theta_{\zeta_1}}^\top r}{d\theta_{\zeta_1}} \right\|_2 \leq \left\| \frac{d\pi_\theta^\top r}{d\theta} \right\|_2 + \left\| \frac{d\pi_{\theta_{\zeta_1}}^\top r}{d\theta_{\zeta_1}} - \frac{d\pi_\theta^\top r}{d\theta} \right\|_2 \qquad \text{(by triangle inequality)} \tag{236}$$

$$\leq \left\| \frac{d\pi_\theta^\top r}{d\theta} \right\|_2 + 3\eta \cdot \int_0^1 \left\| \frac{d\pi_{\theta_{\zeta_2}}^\top r}{d\theta_{\zeta_2}} \right\|_2 d\zeta_2. \qquad \text{(by Eq. (232))} \tag{237}$$

Denote $\theta_{\zeta_3} \coloneqq \theta + \zeta_3 \cdot (\theta_{\zeta_2} - \theta)$ with some $\zeta_3 \in [0,1]$. Using similar calculation as in Eq. (232), we have,

$$\left\| \frac{d\pi_{\theta_{\zeta_2}}^\top r}{d\theta_{\zeta_2}} \right\|_2 \leq \left\| \frac{d\pi_\theta^\top r}{d\theta} \right\|_2 + \left\| \frac{d\pi_{\theta_{\zeta_2}}^\top r}{d\theta_{\zeta_2}} - \frac{d\pi_\theta^\top r}{d\theta} \right\|_2 \tag{238}$$

$$\leq \left\| \frac{d\pi_\theta^\top r}{d\theta} \right\|_2 + 3\eta \cdot \int_0^1 \left\| \frac{d\pi_{\theta_{\zeta_3}}^\top r}{d\theta_{\zeta_3}} \right\|_2 d\zeta_3. \tag{239}$$

Combining Eqs. (236) and (238), we have,

$$\left\| \frac{d\pi_{\theta_{\zeta_1}}^\top r}{d\theta_{\zeta_1}} \right\|_2 \leq (1+3\eta) \cdot \left\| \frac{d\pi_\theta^\top r}{d\theta} \right\|_2 + (3\eta)^2 \cdot \int_0^1 \int_0^1 \left\| \frac{d\pi_{\theta_{\zeta_3}}^\top r}{d\theta_{\zeta_3}} \right\|_2 d\zeta_3 d\zeta_2, \tag{240}$$

which implies,

$$\left\| \frac{d\pi_{\theta_{\zeta_1}}^\top r}{d\theta_{\zeta_1}} \right\|_2 \leq \left[\sum_{i=0}^{\infty} (3\eta)^i \right] \cdot \left\| \frac{d\pi_\theta^\top r}{d\theta} \right\|_2 \tag{241}$$

$$= \frac{1}{1-3\eta} \cdot \left\| \frac{d\pi_\theta^\top r}{d\theta} \right\|_2. \qquad (\eta \in (0, 1/3)) \qquad \square$$

Lemma 4 (Non-vanishing NŁ coefficient) **.** Using normalized policy gradient method, we have $\inf_{t \geq 1} \pi_{\theta_t}(a^*) > 0$.

Proof. The proof is similar to Mei et al. (2020b, Lemma 5). Let

$$c = \frac{K}{2\Delta} \cdot \left(1 - \frac{\Delta}{K} \right) \tag{242}$$

and

$$\Delta = r(a^*) - \max_{a \neq a^*} r(a) > 0 \tag{243}$$

denote the reward gap of r. We will prove that $\inf_{t \geq 1} \pi_{\theta_t}(a^*) = \min_{1 \leq t \leq t_0} \pi_{\theta_t}(a^*)$, where $t_0 = \min\{t : \pi_{\theta_t}(a^*) \geq \frac{c}{c+1}\}$.

Note that t_0 depends only on θ_1 and c, and c depends only on the problem. Define the following regions,

$$\mathcal{R}_1 = \left\{\theta : \frac{d\pi_\theta^\top r}{d\theta(a^*)} \geq \frac{d\pi_\theta^\top r}{d\theta(a)},\ \forall a \neq a^*\right\}, \tag{244}$$

$$\mathcal{R}_2 = \left\{\theta : \pi_\theta(a^*) \geq \pi_\theta(a),\ \forall a \neq a^*\right\}, \tag{245}$$

$$\mathcal{N}_c = \left\{\theta : \pi_\theta(a^*) \geq \frac{c}{c+1}\right\}. \tag{246}$$

We make the following three-part claim.

Claim 6. *The following hold :*

a) *Following a NPG update* $\theta_{t+1} = \theta_t + \eta \cdot \frac{d\pi_{\theta_t}^\top r}{d\theta_t} \Big/ \left\|\frac{d\pi_{\theta_t}^\top r}{d\theta_t}\right\|_2$, *if* $\theta_t \in \mathcal{R}_1$, *then* (i) $\theta_{t+1} \in \mathcal{R}_1$ *and* (ii) $\pi_{\theta_{t+1}}(a^*) \geq \pi_{\theta_t}(a^*)$.

b) *We have* $\mathcal{R}_2 \subset \mathcal{R}_1$ *and* $\mathcal{N}_c \subset \mathcal{R}_1$.

c) *For* $\eta = 1/6$, *there exists a finite time* $t_0 \geq 1$, *such that* $\theta_{t_0} \in \mathcal{N}_c$, *and thus* $\theta_{t_0} \in \mathcal{R}_1$, *which implies that* $\inf_{t\geq 1} \pi_{\theta_t}(a^*) = \min_{1\leq t\leq t_0} \pi_{\theta_t}(a^*)$..

Claim a) Part (i): We want to show that if $\theta_t \in \mathcal{R}_1$, then $\theta_{t+1} \in \mathcal{R}_1$. Let

$$\mathcal{R}_1(a) = \left\{\theta : \frac{d\pi_\theta^\top r}{d\theta(a^*)} \geq \frac{d\pi_\theta^\top r}{d\theta(a)}\right\}. \tag{247}$$

Note that $\mathcal{R}_1 = \cap_{a\neq a^*}\mathcal{R}_1(a)$. Pick $a \neq a^*$. Clearly, it suffices to show that if $\theta_t \in \mathcal{R}_1(a)$ then $\theta_{t+1} \in \mathcal{R}_1(a)$. Hence, suppose that $\theta_t \in \mathcal{R}_1(a)$. We consider two cases.

Case (a): $\pi_{\theta_t}(a^*) \geq \pi_{\theta_t}(a)$. Since $\pi_{\theta_t}(a^*) \geq \pi_{\theta_t}(a)$, we also have $\theta_t(a^*) \geq \theta_t(a)$. After an update of the parameters,

$$\theta_{t+1}(a^*) = \theta_t(a^*) + \frac{\eta}{\left\|\frac{d\pi_{\theta_t}^\top r}{d\theta_t}\right\|_2} \cdot \frac{d\pi_{\theta_t}^\top r}{d\theta_t(a^*)} \tag{248}$$

$$\geq \theta_t(a) + \frac{\eta}{\left\|\frac{d\pi_{\theta_t}^\top r}{d\theta_t}\right\|_2} \cdot \frac{d\pi_{\theta_t}^\top r}{d\theta_t(a)} \tag{249}$$

$$= \theta_{t+1}(a), \tag{250}$$

which implies that $\pi_{\theta_{t+1}}(a^*) \geq \pi_{\theta_{t+1}}(a)$. Since $r(a^*) - \pi_{\theta_{t+1}}^\top r > 0$ and $r(a^*) > r(a)$,

$$\pi_{\theta_{t+1}}(a^*) \cdot \left(r(a^*) - \pi_{\theta_{t+1}}^\top r\right) \geq \pi_{\theta_{t+1}}(a) \cdot \left(r(a) - \pi_{\theta_{t+1}}^\top r\right), \tag{251}$$

which is equivalent to $\frac{d\pi_{\theta_{t+1}}^\top r}{d\theta_{t+1}(a^*)} \geq \frac{d\pi_{\theta_{t+1}}^\top r}{d\theta_{t+1}(a)}$, i.e., $\theta_{t+1} \in \mathcal{R}_1(a)$.

Case (b): Suppose now that $\pi_{\theta_t}(a^*) < \pi_{\theta_t}(a)$. First note that for any θ and $a \neq a^*$, $\theta \in \mathcal{R}_1(a)$ holds if and only if

$$r(a^*) - r(a) \geq \left(1 - \frac{\pi_\theta(a^*)}{\pi_\theta(a)}\right) \cdot \left(r(a^*) - \pi_\theta^\top r\right). \tag{252}$$

Indeed, from the condition $\frac{d\pi_\theta^\top r}{d\theta(a^*)} \geq \frac{d\pi_\theta^\top r}{d\theta(a)}$, we get

$$\pi_\theta(a^*) \cdot \left(r(a^*) - \pi_\theta^\top r\right) \geq \pi_\theta(a) \cdot \left(r(a) - \pi_\theta^\top r\right) \tag{253}$$

$$= \pi_\theta(a) \cdot \left(r(a^*) - \pi_\theta^\top r\right) - \pi_\theta(a) \cdot \left(r(a^*) - r(a)\right), \tag{254}$$

which, after rearranging, is equivalent to Eq. (252). Hence, it suffices to show that Eq. (252) holds for θ_{t+1} provided it holds for θ_t. From the latter condition, we get

$$r(a^*) - r(a) \geq \left(1 - \exp\left\{\theta_t(a^*) - \theta_t(a)\right\}\right) \cdot \left(r(a^*) - \pi_{\theta_t}^\top r\right). \tag{255}$$

After an update of the parameters, according to Lemma 12 (or Eq. (270) below), $\pi_{\theta_{t+1}}^\top r \geq \pi_{\theta_t}^\top r$, i.e.,

$$0 < r(a^*) - \pi_{\theta_{t+1}}^\top r \leq r(a^*) - \pi_{\theta_t}^\top r\,. \tag{256}$$

On the other hand,

$$\theta_{t+1}(a^*) - \theta_{t+1}(a) = \theta_t(a^*) + \frac{\eta}{\left\|\frac{d\pi_{\theta_t}^\top r}{d\theta_t}\right\|_2} \cdot \frac{d\pi_{\theta_t}^\top r}{d\theta_t(a^*)} - \theta_t(a) - \frac{\eta}{\left\|\frac{d\pi_{\theta_t}^\top r}{d\theta_t}\right\|_2} \cdot \frac{d\pi_{\theta_t}^\top r}{d\theta_t(a)} \tag{257}$$

$$\geq \theta_t(a^*) - \theta_t(a) \tag{258}$$

which implies that

$$1 - \exp\left\{\theta_{t+1}(a^*) - \theta_{t+1}(a)\right\} \leq 1 - \exp\left\{\theta_t(a^*) - \theta_t(a)\right\}. \tag{259}$$

Furthermore, by our assumption that $\pi_{\theta_t}(a^*) < \pi_{\theta_t}(a)$, we have $1 - \exp\left\{\theta_t(a^*) - \theta_t(a)\right\} = 1 - \frac{\pi_{\theta_t}(a^*)}{\pi_{\theta_t}(a)} > 0$. Putting things together, we get

$$\left(1 - \exp\left\{\theta_{t+1}(a^*) - \theta_{t+1}(a)\right\}\right) \cdot \left(r(a^*) - \pi_{\theta_{t+1}}^\top r\right) \leq \left(1 - \exp\left\{\theta_t(a^*) - \theta_t(a)\right\}\right) \cdot \left(r(a^*) - \pi_{\theta_t}^\top r\right) \tag{260}$$

$$\leq r(a^*) - r(a), \tag{261}$$

which is equivalent to

$$\left(1 - \frac{\pi_{\theta_{t+1}}(a^*)}{\pi_{\theta_{t+1}}(a)}\right) \cdot \left(r(a^*) - \pi_{\theta_{t+1}}^\top r\right) \leq r(a^*) - r(a), \tag{262}$$

and thus by our previous remark, $\theta_{t+1} \in \mathcal{R}_1(a)$, thus, finishing the proof of part (i).

Part (ii): Assume again that $\theta_t \in \mathcal{R}_1$. We want to show that $\pi_{\theta_{t+1}}(a^*) \geq \pi_{\theta_t}(a^*)$. Since $\theta_t \in \mathcal{R}_1$, we have $\frac{d\pi_{\theta_t}^\top r}{d\theta_t(a^*)} \geq \frac{d\pi_{\theta_t}^\top r}{d\theta_t(a)}$, $\forall a \neq a^*$. Hence,

$$\pi_{\theta_{t+1}}(a^*) = \frac{\exp\left\{\theta_{t+1}(a^*)\right\}}{\sum_a \exp\left\{\theta_{t+1}(a)\right\}} \tag{263}$$

$$= \frac{\exp\left\{\theta_t(a^*) + \eta \cdot \frac{d\pi_{\theta_t}^\top r}{d\theta_t(a^*)} \Big/ \left\|\frac{d\pi_{\theta_t}^\top r}{d\theta_t}\right\|_2\right\}}{\sum_a \exp\left\{\theta_t(a) + \eta \cdot \frac{d\pi_{\theta_t}^\top r}{d\theta_t(a)} \Big/ \left\|\frac{d\pi_{\theta_t}^\top r}{d\theta_t}\right\|_2\right\}} \tag{264}$$

$$\geq \frac{\exp\left\{\theta_t(a^*) + \eta \cdot \frac{d\pi_{\theta_t}^\top r}{d\theta_t(a^*)} \Big/ \left\|\frac{d\pi_{\theta_t}^\top r}{d\theta_t}\right\|_2\right\}}{\sum_a \exp\left\{\theta_t(a) + \eta \cdot \frac{d\pi_{\theta_t}^\top r}{d\theta_t(a^*)} \Big/ \left\|\frac{d\pi_{\theta_t}^\top r}{d\theta_t}\right\|_2\right\}} \qquad \left(\text{using } \frac{d\pi_{\theta_t}^\top r}{d\theta_t(a^*)} \geq \frac{d\pi_{\theta_t}^\top r}{d\theta_t(a)}\right) \tag{265}$$

$$= \frac{\exp\left\{\theta_t(a^*)\right\}}{\sum_a \exp\left\{\theta_t(a)\right\}} = \pi_{\theta_t}(a^*). \tag{266}$$

Claim b); Claim c) The proof of those claims are exactly the same as Mei et al. (2020b, Lemma 5), since they do not involve the update rule. □

Theorem 2. Using NPG $\theta_{t+1} = \theta_t + \eta \cdot \frac{d\pi_{\theta_t}^\top r}{d\theta_t} \Big/ \left\|\frac{d\pi_{\theta_t}^\top r}{d\theta_t}\right\|_2$, with $\eta = 1/6$, for all $t \geq 1$, we have,

$$\left(\pi^* - \pi_{\theta_t}\right)^\top r \leq e^{-\frac{c\cdot(t-1)}{12}} \cdot \left(\pi^* - \pi_{\theta_1}\right)^\top r, \tag{267}$$

where $c = \inf_{t\geq 1} \pi_{\theta_t}(a^*) > 0$ is from Lemma 4, and c is a constant that depends on r and θ_1, but not on the time t.

Proof. Denote $\theta_{\zeta_t} \coloneqq \theta_t + \zeta_t \cdot (\theta_{t+1} - \theta_t)$ with some $\zeta_t \in [0, 1]$. According to Lemma 2,

$$\left|\left(\pi_{\theta_{t+1}} - \pi_{\theta_t}\right)^\top r - \left\langle \frac{d\pi_{\theta_t}^\top r}{d\theta_t}, \theta_{t+1} - \theta_t \right\rangle\right| \leq \frac{3}{2} \cdot \left\|\frac{d\pi_{\theta_{\zeta_t}}^\top r}{d\theta_{\zeta_t}}\right\|_2 \cdot \|\theta_{t+1} - \theta_t\|_2^2 \tag{268}$$

$$\leq \frac{3}{2} \cdot \frac{1}{1 - 3\eta} \cdot \left\|\frac{d\pi_{\theta_t}^\top r}{d\theta_t}\right\|_2 \cdot \|\theta_{t+1} - \theta_t\|_2^2, \qquad (\eta = 1/6, \text{ by Lemma 3}) \tag{269}$$

which implies,

$$\pi_{\theta_t}^\top r - \pi_{\theta_{t+1}}^\top r \leq -\left\langle \frac{d\pi_{\theta_t}^\top r}{d\theta_t}, \theta_{t+1} - \theta_t \right\rangle + \frac{3}{2\cdot(1-3\eta)} \cdot \left\| \frac{d\pi_{\theta_t}^\top r}{d\theta_t} \right\|_2 \cdot \|\theta_{t+1} - \theta_t\|_2^2 \tag{270}$$

$$= -\eta \cdot \left\| \frac{d\pi_{\theta_t}^\top r}{d\theta_t} \right\|_2 + \frac{3\cdot\eta^2}{2\cdot(1-3\eta)} \cdot \left\| \frac{d\pi_{\theta_t}^\top r}{d\theta_t} \right\|_2 \qquad \left(\text{using } \theta_{t+1} = \theta_t + \eta \cdot \frac{d\pi_{\theta_t}^\top r}{d\theta_t} \Big/ \left\| \frac{d\pi_{\theta_t}^\top r}{d\theta_t} \right\|_2\right) \tag{271}$$

$$= -\frac{1}{12} \cdot \left\| \frac{d\pi_{\theta_t}^\top r}{d\theta_t} \right\|_2 \qquad (\text{using } \eta = 1/6) \tag{272}$$

$$\leq -\frac{1}{12} \cdot \pi_{\theta_t}(a^*) \cdot (\pi^* - \pi_{\theta_t})^\top r \qquad (\text{by Lemma 1}) \tag{273}$$

$$\leq -\frac{1}{12} \cdot \inf_{t\geq 1} \pi_{\theta_t}(a^*) \cdot (\pi^* - \pi_{\theta_t})^\top r. \tag{274}$$

According to Eq. (270), we have,

$$(\pi^* - \pi_{\theta_t})^\top r \leq \left(1 - \frac{c}{12}\right) \cdot \left(\pi^* - \pi_{\theta_{t-1}}\right)^\top r \qquad \left(c \coloneqq \inf_{t\geq 1} \pi_{\theta_t}(a^*) > 0\right) \tag{275}$$

$$\leq \exp\{-c/12\} \cdot \left(\pi^* - \pi_{\theta_{t-1}}\right)^\top r \tag{276}$$

$$\leq \exp\{-(t-1)\cdot c/12\} \cdot (\pi^* - \pi_{\theta_1})^\top r. \qquad \square$$

B.2. General MDPs

Lemma 5 (NŁ) **.** Denote $S \coloneqq |\mathcal{S}|$ as the total number of states. We have, for all $\theta \in \mathbb{R}^{\mathcal{S}\times\mathcal{A}}$,

$$\left\| \frac{\partial V^{\pi_\theta}(\mu)}{\partial \theta} \right\|_2 \geq \frac{\min_s \pi_\theta(a^*(s)|s)}{\sqrt{S} \cdot \left\| d_\rho^{\pi^*} / d_\mu^{\pi_\theta} \right\|_\infty} \cdot \left(V^*(\rho) - V^{\pi_\theta}(\rho)\right), \tag{277}$$

where $a^*(s)$ is the action that π^* selects in state s.

Proof. See the proof in (Mei et al., 2020b, Lemma 8). We include a proof for completeness. We have,

$$\left\| \frac{\partial V^{\pi_\theta}(\mu)}{\partial \theta} \right\|_2 = \left[\sum_{s,a} \left(\frac{\partial V^{\pi_\theta}(\mu)}{\partial \theta(s,a)} \right)^2 \right]^{\frac{1}{2}} \tag{278}$$

$$\geq \left[\sum_{s} \left(\frac{\partial V^{\pi_\theta}(\mu)}{\partial \theta(s,a^*(s))} \right)^2 \right]^{\frac{1}{2}} \tag{279}$$

$$\geq \frac{1}{\sqrt{S}} \sum_s \left| \frac{\partial V^{\pi_\theta}(\mu)}{\partial \theta(s,a^*(s))} \right| \qquad (\text{by Cauchy-Schwarz, } \|x\|_1 = |\langle \mathbf{1},\ |x|\rangle| \leq \|\mathbf{1}\|_2 \cdot \|x\|_2) \tag{280}$$

$$= \frac{1}{1-\gamma} \cdot \frac{1}{\sqrt{S}} \sum_s \left| d_\mu^{\pi_\theta}(s) \cdot \pi_\theta(a^*(s)|s) \cdot A^{\pi_\theta}(s,a^*(s)) \right| \qquad (\text{by Lemma 15}) \tag{281}$$

$$= \frac{1}{1-\gamma} \cdot \frac{1}{\sqrt{S}} \sum_s d_\mu^{\pi_\theta}(s) \cdot \pi_\theta(a^*(s)|s) \cdot \left| A^{\pi_\theta}(s,a^*(s)) \right|. \qquad \left(\text{because } d_\mu^{\pi_\theta}(s) \geq 0 \text{ and } \pi_\theta(a^*(s)|s) \geq 0\right) \tag{282}$$

Define the distribution mismatch coefficient as $\left\|\frac{d_\rho^{\pi^*}}{d_\mu^{\pi_\theta}}\right\|_\infty = \max_s \frac{d_\rho^{\pi^*}(s)}{d_\mu^{\pi_\theta}(s)}$. We have,

$$\left\|\frac{\partial V^{\pi_\theta}(\mu)}{\partial \theta}\right\|_2 \geq \frac{1}{1-\gamma}\cdot\frac{1}{\sqrt{S}}\sum_s \frac{d_\mu^{\pi_\theta}(s)}{d_\rho^{\pi^*}(s)}\cdot d_\rho^{\pi^*}(s)\cdot\pi_\theta(a^*(s)|s)\cdot|A^{\pi_\theta}(s,a^*(s))| \tag{283}$$

$$\geq \frac{1}{1-\gamma}\cdot\frac{1}{\sqrt{S}}\cdot\left\|\frac{d_\rho^{\pi^*}}{d_\mu^{\pi_\theta}}\right\|_\infty^{-1}\cdot\min_s\pi_\theta(a^*(s)|s)\cdot\sum_s d_\rho^{\pi^*}(s)\cdot|A^{\pi_\theta}(s,a^*(s))| \tag{284}$$

$$\geq \frac{1}{1-\gamma}\cdot\frac{1}{\sqrt{S}}\cdot\left\|\frac{d_\rho^{\pi^*}}{d_\mu^{\pi_\theta}}\right\|_\infty^{-1}\cdot\min_s\pi_\theta(a^*(s)|s)\cdot\sum_s d_\rho^{\pi^*}(s)\cdot A^{\pi_\theta}(s,a^*(s)) \tag{285}$$

$$= \frac{1}{\sqrt{S}}\cdot\left\|\frac{d_\rho^{\pi^*}}{d_\mu^{\pi_\theta}}\right\|_\infty^{-1}\cdot\min_s\pi_\theta(a^*(s)|s)\cdot\frac{1}{1-\gamma}\sum_s d_\rho^{\pi^*}(s)\sum_a\pi^*(a|s)\cdot A^{\pi_\theta}(s,a) \tag{286}$$

$$= \frac{1}{\sqrt{S}}\cdot\left\|\frac{d_\rho^{\pi^*}}{d_\mu^{\pi_\theta}}\right\|_\infty^{-1}\cdot\min_s\pi_\theta(a^*(s)|s)\cdot\left[V^*(\rho)-V^{\pi_\theta}(\rho)\right], \tag{287}$$

where the one but last equality used that π^* is deterministic and in state s chooses $a^*(s)$ with probability one, and the last equality uses the performance difference formula (Lemma 17). □

Lemma 6 (NS) **.** Let Assumption 1 hold and denote $\theta_\zeta \coloneqq \theta+\zeta\cdot(\theta'-\theta)$ with some $\zeta\in[0,1]$. $\theta\mapsto V^{\pi_\theta}(\mu)$ satisfies $\beta(\theta_\zeta)$ non-uniform smoothness with

$$\beta(\theta_\zeta) = \left[3+\frac{2\cdot(C_\infty-(1-\gamma))}{(1-\gamma)\cdot\gamma}\right]\cdot\sqrt{S}\cdot\left\|\frac{\partial V^{\pi_{\theta_\zeta}}(\mu)}{\partial\theta_\zeta}\right\|_2, \tag{288}$$

where $C_\infty \coloneqq \max_\pi\left\|\frac{d_\mu^\pi}{\mu}\right\|_\infty \leq \frac{1}{\min_s\mu(s)} < \infty$.

Proof. The main part is to prove that for all $y\in\mathbb{R}^{SA}$ and θ,

$$\left|y^\top\frac{\partial^2 V^{\pi_\theta}(\mu)}{\partial\theta^2}y\right| \leq \left[3+\frac{2\cdot(C_\infty-(1-\gamma))}{(1-\gamma)\cdot\gamma}\right]\cdot\sqrt{S}\cdot\left\|\frac{\partial V^{\pi_\theta}(\mu)}{\partial\theta}\right\|_2\cdot\|y\|_2^2. \tag{289}$$

We first calculate the second order derivative of $V^{\pi_\theta}(\mu)$ w.r.t. θ.

Denote $\theta_\alpha = \theta+\alpha u$, where $\alpha\in\mathbb{R}$ and $u\in\mathbb{R}^{SA}$. For any $(s,a)\in\mathcal{S}\times\mathcal{A}$,

$$\frac{\partial\pi_{\theta_\alpha}(a|s)}{\partial\alpha}\Big|_{\alpha=0} = \Big\langle\frac{\partial\pi_{\theta_\alpha}(a|s)}{\partial\theta_\alpha}\Big|_{\alpha=0},\frac{\partial\theta_\alpha}{\partial\alpha}\Big\rangle \tag{290}$$

$$= \Big\langle\frac{\partial\pi_\theta(a|s)}{\partial\theta},u\Big\rangle \tag{291}$$

$$= \Big\langle\frac{\partial\pi_\theta(a|s)}{\partial\theta(s,\cdot)},u(s,\cdot)\Big\rangle \qquad \left(\frac{\partial\pi_\theta(a|s)}{\partial\theta(s',\cdot)}=\mathbf{0},\ \forall s'\neq s\right) \tag{292}$$

Similarly, for any $(s,a)\in\mathcal{S}\times\mathcal{A}$,

$$\frac{\partial^2\pi_{\theta_\alpha}(a|s)}{\partial\alpha^2}\Big|_{\alpha=0} = \Big\langle\frac{\partial}{\partial\theta_\alpha}\left\{\frac{\partial\pi_{\theta_\alpha}(a|s)}{\partial\alpha}\right\}\Big|_{\alpha=0},\frac{\partial\theta_\alpha}{\partial\alpha}\Big\rangle \tag{293}$$

$$= \Big\langle\frac{\partial^2\pi_{\theta_\alpha}(a|s)}{\partial\theta_\alpha^2}\Big|_{\alpha=0}\frac{\partial\theta_\alpha}{\partial\alpha},\frac{\partial\theta_\alpha}{\partial\alpha}\Big\rangle \tag{294}$$

$$= \Big\langle\frac{\partial^2\pi_\theta(a|s)}{\partial\theta^2(s,\cdot)}u(s,\cdot),u(s,\cdot)\Big\rangle. \tag{295}$$

Define $\Pi(\alpha)\in\mathbb{R}^{S\times SA}$ as follows,

$$\Pi(\alpha) \coloneqq \begin{bmatrix}\pi_{\theta_\alpha}(\cdot|1)^\top & \mathbf{0}^\top & \cdots & \mathbf{0}^\top\\ \mathbf{0}^\top & \pi_{\theta_\alpha}(\cdot|2)^\top & \cdots & \mathbf{0}^\top\\ \vdots & \vdots & \ddots & \vdots\\ \mathbf{0}^\top & \mathbf{0}^\top & \cdots & \pi_{\theta_\alpha}(\cdot|S)^\top\end{bmatrix}. \tag{296}$$

Denote $\mathcal{P} \in \mathbb{R}^{SA\times S}$ such that,

$$\mathcal{P}_{(sa,s')} \coloneqq \mathcal{P}(s'|s,a). \tag{297}$$

Define $P(\alpha) \coloneqq \Pi(\alpha)\mathcal{P} \in \mathbb{R}^{S\times S}$, where $\forall(s,s')$,

$$[P(\alpha)]_{(s,s')} = \sum_a \pi_{\theta_\alpha}(a|s)\cdot\mathcal{P}(s'|s,a). \tag{298}$$

The derivative w.r.t. α is

$$\frac{\partial P(\alpha)}{\partial\alpha} = \frac{\partial\Pi(\alpha)\mathcal{P}}{\partial\alpha} = \frac{\partial\Pi(\alpha)}{\partial\alpha}\mathcal{P}. \tag{299}$$

And $\forall(s,s')$, we have,

$$\left[\frac{\partial P(\alpha)}{\partial\alpha}\Big|_{\alpha=0}\right]_{(s,s')} = \sum_a\left[\frac{\partial\pi_{\theta_\alpha}(a|s)}{\partial\alpha}\Big|_{\alpha=0}\right]\cdot\mathcal{P}(s'|s,a). \tag{300}$$

Next, consider the state value function of π_{θ_α},

$$V^{\pi_{\theta_\alpha}}(s) = \sum_a \pi_{\theta_\alpha}(a|s)\cdot r(s,a) + \gamma\sum_a\pi_{\theta_\alpha}(a|s)\sum_{s'}\mathcal{P}(s'|s,a)\cdot V^{\pi_{\theta_\alpha}}(s'), \tag{301}$$

which implies,

$$V^{\pi_{\theta_\alpha}}(s) = e_s^\top M(\alpha) r_{\theta_\alpha} \tag{302}$$

$$V^{\pi_{\theta_\alpha}}(\mu) = \mu^\top M(\alpha) r_{\theta_\alpha}, \tag{303}$$

where

$$M(\alpha) = \left(\mathbf{Id} - \gamma P(\alpha)\right)^{-1}, \tag{304}$$

and $r_{\theta_\alpha}\in\mathbb{R}^S$ is given by

$$r_{\theta_\alpha} = \Pi(\alpha) r, \tag{305}$$

where $r\in\mathbb{R}^{SA}$. Taking derivative w.r.t. α in Eq. (303),

$$\frac{\partial V^{\pi_{\theta_\alpha}}(\mu)}{\partial\alpha} = \gamma\cdot\mu^\top M(\alpha)\frac{\partial P(\alpha)}{\partial\alpha}M(\alpha)r_{\theta_\alpha} + \mu^\top M(\alpha)\frac{\partial r_{\theta_\alpha}}{\partial\alpha} \tag{306}$$

$$= \mu^\top M(\alpha)\left[\gamma\cdot\frac{\partial P(\alpha)}{\partial\alpha}M(\alpha)r_{\theta_\alpha} + \frac{\partial r_{\theta_\alpha}}{\partial\alpha}\right] \tag{307}$$

$$= \mu^\top M(\alpha)\left[\gamma\cdot\frac{\partial\Pi(\alpha)}{\partial\alpha}\mathcal{P}M(\alpha)r_{\theta_\alpha} + \frac{\partial\Pi(\alpha)}{\partial\alpha}r\right] \qquad \text{(by Eqs. (299) and (305))} \tag{308}$$

$$= \mu^\top M(\alpha)\frac{\partial\Pi(\alpha)}{\partial\alpha}Q^{\pi_{\theta_\alpha}}, \tag{309}$$

where $Q^{\pi_{\theta_\alpha}}\in\mathbb{R}^{SA}$ is the state-action value and it satisfies,

$$Q^{\pi_{\theta_\alpha}} = r + \gamma\cdot\mathcal{P}M(\alpha)r_{\theta_\alpha} \tag{310}$$

$$= r + \gamma\cdot\mathcal{P}V^{\pi_{\theta_\alpha}} \qquad \text{(by Eq. (302))} \tag{311}$$

Similarly, taking second derivative w.r.t. α,

$$\frac{\partial^2 V^{\pi_{\theta_\alpha}}(\mu)}{\partial\alpha^2} = 2\gamma^2\cdot\mu^\top M(\alpha)\frac{\partial P(\alpha)}{\partial\alpha}M(\alpha)\frac{\partial P(\alpha)}{\partial\alpha}M(\alpha)r_{\theta_\alpha} + \gamma\cdot\mu^\top M(\alpha)\frac{\partial^2 P(\alpha)}{\partial\alpha^2}M(\alpha)r_{\theta_\alpha} \tag{312}$$

$$+ 2\gamma\cdot\mu^\top M(\alpha)\frac{\partial P(\alpha)}{\partial\alpha}M(\alpha)\frac{\partial r_{\theta_\alpha}}{\partial\alpha} + \mu^\top M(\alpha)\frac{\partial^2 r_{\theta_\alpha}}{\partial\alpha^2} \tag{313}$$

$$= 2\gamma\cdot\mu^\top M(\alpha)\frac{\partial P(\alpha)}{\partial\alpha}M(\alpha)\frac{\partial\Pi(\alpha)}{\partial\alpha}\left(\gamma\cdot\mathcal{P}M(\alpha)r_{\theta_\alpha} + r\right) + \mu^\top M(\alpha)\frac{\partial^2\Pi(\alpha)}{\partial\alpha^2}\left(\gamma\cdot\mathcal{P}M(\alpha)r_{\theta_\alpha} + r\right) \tag{314}$$

$$= 2\gamma\cdot\mu^\top M(\alpha)\frac{\partial P(\alpha)}{\partial\alpha}M(\alpha)\frac{\partial\Pi(\alpha)}{\partial\alpha}Q^{\pi_{\theta_\alpha}} + \mu^\top M(\alpha)\frac{\partial^2\Pi(\alpha)}{\partial\alpha^2}Q^{\pi_{\theta_\alpha}} \tag{315}$$

For the last term, we have,

$$\left[\frac{\partial^2\Pi(\alpha)}{\partial\alpha^2}Q^{\pi_{\theta_\alpha}}\Big|_{\alpha=0}\right]_{(s)} = \sum_a \frac{\partial^2\pi_{\theta_\alpha}(a|s)}{\partial\alpha^2}\Big|_{\alpha=0}\cdot Q^{\pi_\theta}(s,a) \tag{316}$$

$$= \sum_a \left\langle \frac{\partial^2\pi_\theta(a|s)}{\partial\theta^2(s,\cdot)}u(s,\cdot), u(s,\cdot)\right\rangle \cdot Q^{\pi_\theta}(s,a) \qquad \text{(by Eq. (293))} \tag{317}$$

$$= u(s,\cdot)^\top \left[\sum_a \frac{\partial^2\pi_\theta(a|s)}{\partial\theta^2(s,\cdot)}\cdot Q^{\pi_\theta}(s,a)\right] u(s,\cdot) \tag{318}$$

Let $S(a,\theta) = \frac{\partial^2\pi_\theta(a|s)}{\partial\theta^2(s,\cdot)} \in \mathbb{R}^{A\times A}$. $\forall i,j\in[A]$, the value of $S(a,\theta)$ is,

$$S_{(i,j)} = \frac{\partial\{\delta_{ia}\pi_\theta(a|s) - \pi_\theta(a|s)\pi_\theta(i|s)\}}{\partial\theta(s,j)} \tag{319}$$

$$= \delta_{ia}\cdot[\delta_{ja}\pi_\theta(a|s) - \pi_\theta(a|s)\pi_\theta(j|s)] - \pi_\theta(a|s)\cdot[\delta_{ij}\pi_\theta(j|s) - \pi_\theta(i|s)\pi_\theta(j|s)] - \pi_\theta(i|s)\cdot[\delta_{ja}\pi_\theta(a|s) - \pi_\theta(a|s)\pi_\theta(j|s)]\,, \tag{320}$$

where the δ notation is as defined in Eq. (187). Then we have,

$$\left[\sum_a \frac{\partial^2\pi_\theta(a|s)}{\partial\theta^2(s,\cdot)}\cdot Q^{\pi_\theta}(s,a)\right]_{(i,j)} = \sum_a S_{(i,j)}\cdot Q^{\pi_\theta}(s,a) \tag{321}$$

$$= \delta_{ij}\cdot\pi_\theta(i|s)\cdot[Q^{\pi_\theta}(s,i) - V^{\pi_\theta}(s)] - \pi_\theta(i|s)\cdot\pi_\theta(j|s)\cdot[Q^{\pi_\theta}(s,i) - V^{\pi_\theta}(s)] - \pi_\theta(i|s)\cdot\pi_\theta(j|s)\cdot[Q^{\pi_\theta}(s,j) - V^{\pi_\theta}(s)]\,. \tag{322}$$

Therefore we have,

$$\left[\frac{\partial^2\Pi(\alpha)}{\partial\alpha^2}Q^{\pi_{\theta_\alpha}}\Big|_{\alpha=0}\right]_{(s)} = \sum_{i=1}^A\sum_{j=1}^A u(s,i)\cdot u(s,j)\cdot\left[\sum_a\frac{\partial^2\pi_\theta(a|s)}{\partial\theta^2(s,\cdot)}\cdot Q^{\pi_\theta}(s,a)\right]_{(i,j)} \tag{323}$$

$$= (H(\pi_\theta(\cdot|s))Q^{\pi_\theta}(s,\cdot))^\top(u(s,\cdot)\odot u(s,\cdot)) - 2\cdot\left[(H(\pi_\theta(\cdot|s))Q^{\pi_\theta}(s,\cdot))^\top u(s,\cdot)\right]\cdot\left(\pi_\theta(\cdot|s)^\top u(s,\cdot)\right), \tag{324}$$

where $H(\pi) \coloneqq \mathrm{diag}(\pi) - \pi\pi^\top$. Combining the above results with Eq. (312), we have,

$$\left|\mu^\top M(\alpha)\frac{\partial^2\Pi(\alpha)}{\partial\alpha^2}Q^{\pi_{\theta_\alpha}}\Big|_{\alpha=0}\right| \le \frac{1}{1-\gamma}\cdot\sum_s d_\mu^{\pi_\theta}(s)\cdot\left|\left[\frac{\partial^2\Pi(\alpha)}{\partial\alpha^2}Q^{\pi_{\theta_\alpha}}\Big|_{\alpha=0}\right]_{(s)}\right| \qquad \text{(by triangle inequality)} \tag{325}$$

$$\le \frac{1}{1-\gamma}\cdot\sum_s d_\mu^{\pi_\theta}(s)\cdot 3\cdot\|H(\pi_\theta(\cdot|s))Q^{\pi_\theta}(s,\cdot)\|_2\cdot\|u\|_2^2 \qquad \text{(by Hölder's inequality)} \tag{326}$$

$$\le \frac{3\cdot\sqrt{S}}{1-\gamma}\cdot\left[\sum_s d_\mu^{\pi_\theta}(s)^2\cdot\|H(\pi_\theta(\cdot|s))Q^{\pi_\theta}(s,\cdot)\|_2^2\right]^{\frac{1}{2}}\cdot\|u\|_2^2 \qquad \text{(by Cauchy-Schwarz)} \tag{327}$$

$$= 3\cdot\sqrt{S}\cdot\left\|\frac{\partial V^{\pi_\theta}(\mu)}{\partial\theta}\right\|_2\cdot\|u\|_2^2\,. \qquad \text{(by Lemma 16)} \tag{328}$$

For the first term in Eq. (312), we have,

$$\mu^\top M(\alpha)\frac{\partial P(\alpha)}{\partial\alpha}M(\alpha)\frac{\partial\Pi(\alpha)}{\partial\alpha}Q^{\pi_{\theta_\alpha}}\Big|_{\alpha=0} = \sum_{s'}\left[\mu^\top M(\alpha)\frac{\partial P(\alpha)}{\partial\alpha}\Big|_{\alpha=0}\right]_{(s')}\cdot\left[M(\alpha)\frac{\partial\Pi(\alpha)}{\partial\alpha}Q^{\pi_{\theta_\alpha}}\Big|_{\alpha=0}\right]_{(s')}, \tag{329}$$

since,

$$\left(\mu^\top M(\alpha)\frac{\partial P(\alpha)}{\partial\alpha}\right)^\top\in\mathbb{R}^S, \quad \text{and} \quad M(\alpha)\frac{\partial\Pi(\alpha)}{\partial\alpha}Q^{\pi_{\theta_\alpha}}\in\mathbb{R}^S. \tag{330}$$

Next we have,

$$\left[M(\alpha)\frac{\partial \Pi(\alpha)}{\partial \alpha}Q^{\pi_{\theta_\alpha}}\Big|_{\alpha=0}\right]_{(s')} = \frac{1}{1-\gamma}\cdot\sum_s d_{s'}^{\pi_\theta}(s)\cdot\left[\frac{\partial \Pi(\alpha)}{\partial \alpha}Q^{\pi_{\theta_\alpha}}\Big|_{\alpha=0}\right]_{(s)} \qquad \left(\frac{\partial \Pi(\alpha)}{\partial \alpha}Q^{\pi_{\theta_\alpha}} \in \mathbb{R}^S\right) \tag{331}$$

$$= \frac{1}{1-\gamma}\cdot\sum_s d_{s'}^{\pi_\theta}(s)\cdot\sum_a \frac{\partial \pi_{\theta_\alpha}(a|s)}{\partial \alpha}\Big|_{\alpha=0}\cdot Q^{\pi_\theta}(s,a) \tag{332}$$

$$= \frac{1}{1-\gamma}\cdot\sum_s d_{s'}^{\pi_\theta}(s)\cdot\sum_a \left\langle \frac{\partial \pi_\theta(a|s)}{\partial \theta(s,\cdot)}, u(s,\cdot)\right\rangle\cdot Q^{\pi_\theta}(s,a) \qquad \text{(by Eq. (290))} \tag{333}$$

$$= \frac{1}{1-\gamma}\cdot\sum_s d_{s'}^{\pi_\theta}(s)\cdot\left\langle \sum_a \frac{\partial \pi_\theta(a|s)}{\partial \theta(s,\cdot)}\cdot Q^{\pi_\theta}(s,a), u(s,\cdot)\right\rangle \tag{334}$$

$$= \frac{1}{1-\gamma}\cdot\sum_s d_{s'}^{\pi_\theta}(s)\cdot\left(H(\pi_\theta(\cdot|s))Q^{\pi_\theta}(s,\cdot)\right)^\top u(s,\cdot), \qquad (H(\pi_\theta) \text{ is the Jacobian of } \theta \mapsto \mathrm{softmax}(\theta)) \tag{335}$$

which implies,

$$\left|\left[M(\alpha)\frac{\partial \Pi(\alpha)}{\partial \alpha}Q^{\pi_{\theta_\alpha}}\Big|_{\alpha=0}\right]_{(s')}\right| \le \frac{1}{1-\gamma}\cdot\sum_s d_{s'}^{\pi_\theta}(s)\cdot\left\|H(\pi_\theta(\cdot|s))Q^{\pi_\theta}(s,\cdot)\right\|_2\cdot\left\|u(s,\cdot)\right\|_2 \tag{336}$$

$$\le \frac{\|u\|_2}{1-\gamma}\cdot\sum_s d_{s'}^{\pi_\theta}(s)\cdot\left\|H(\pi_\theta(\cdot|s))Q^{\pi_\theta}(s,\cdot)\right\|_2. \tag{337}$$

On the other hand,

$$\left[\mu^\top M(\alpha)\frac{\partial P(\alpha)}{\partial \alpha}\Big|_{\alpha=0}\right]_{(s')} = \frac{1}{1-\gamma}\cdot\sum_s d_\mu^{\pi_\theta}(s)\cdot\left[\frac{\partial P(\alpha)}{\partial \alpha}\Big|_{\alpha=0}\right]_{(s,s')} \qquad \left(\frac{\partial P(\alpha)}{\partial \alpha} \in \mathbb{R}^{S\times S}\right) \tag{338}$$

$$= \frac{1}{1-\gamma}\cdot\sum_s d_\mu^{\pi_\theta}(s)\cdot\sum_a\left[\frac{\partial \pi_{\theta_\alpha}(a|s)}{\partial \alpha}\Big|_{\alpha=0}\right]\cdot\mathcal{P}(s'|s,a) \qquad \text{(by Eq. (300))} \tag{339}$$

$$= \frac{1}{1-\gamma}\cdot\sum_s d_\mu^{\pi_\theta}(s)\cdot\sum_a\left\langle \frac{\partial \pi_\theta(a|s)}{\partial \theta(s,\cdot)}, u(s,\cdot)\right\rangle\cdot\mathcal{P}(s'|s,a) \qquad \text{(by Eq. (290))} \tag{340}$$

$$= \frac{1}{1-\gamma}\cdot\sum_s d_\mu^{\pi_\theta}(s)\cdot\sum_a \pi_\theta(a|s)\cdot\mathcal{P}(s'|s,a)\cdot\left[u(s,a)-\pi_\theta(\cdot|s)^\top u(s,\cdot)\right], \tag{341}$$

which implies,

$$\left|\left[\mu^\top M(\alpha)\frac{\partial P(\alpha)}{\partial \alpha}\Big|_{\alpha=0}\right]_{(s')}\right| \le \frac{1}{1-\gamma}\cdot\sum_s d_\mu^{\pi_\theta}(s)\cdot\sum_a \pi_\theta(a|s)\cdot\mathcal{P}(s'|s,a)\cdot 2\cdot\|u(s,\cdot)\|_\infty \tag{342}$$

$$\le \frac{2\cdot\|u\|_2}{1-\gamma}\cdot\sum_s d_\mu^{\pi_\theta}(s)\cdot\sum_a \pi_\theta(a|s)\cdot\mathcal{P}(s'|s,a). \tag{343}$$

According to

$$d_\mu^{\pi_\theta}(s') = (1-\gamma)\cdot\mu(s') + \gamma\cdot\sum_s d_\mu^{\pi_\theta}(s)\cdot\sum_a \pi_\theta(a|s)\cdot\mathcal{P}(s'|s,a), \quad \forall s'\in\mathcal{S} \tag{344}$$

we have,

$$\left|\left[\mu^\top M(\alpha)\frac{\partial P(\alpha)}{\partial \alpha}\Big|_{\alpha=0}\right]_{(s')}\right| \le \frac{2\cdot\|u\|_2}{(1-\gamma)\cdot\gamma}\cdot\left[d_\mu^{\pi_\theta}(s') - (1-\gamma)\cdot\mu(s')\right] \tag{345}$$

$$= \frac{2\cdot\|u\|_2}{(1-\gamma)\cdot\gamma}\cdot\left[\frac{d_\mu^{\pi_\theta}(s')}{\mu(s')}\cdot\mu(s') - (1-\gamma)\cdot\mu(s')\right] \tag{346}$$

$$\le \frac{2\cdot\|u\|_2}{(1-\gamma)\cdot\gamma}\cdot(C_\infty-(1-\gamma))\cdot\mu(s'). \qquad \left(C_\infty \coloneqq \max_\pi\left\|\frac{d_\mu^\pi}{\mu}\right\|_\infty < \left\|\frac{1}{\mu}\right\|_\infty < \infty\right) \tag{347}$$

Combining Eqs. (329), (336) and (345), we have,

$$\left|\mu^\top M(\alpha)\frac{\partial P(\alpha)}{\partial \alpha}M(\alpha)\frac{\partial \Pi(\alpha)}{\partial \alpha}Q^{\pi_{\theta_\alpha}}\Big|_{\alpha=0}\right| \tag{348}$$

$$\leq \sum_{s'}\frac{2\cdot\|u\|_2}{(1-\gamma)\cdot\gamma}\cdot(C_\infty-(1-\gamma))\cdot\mu(s')\cdot\frac{\|u\|_2}{1-\gamma}\cdot\sum_s d_{s'}^{\pi_\theta}(s)\cdot\|H(\pi_\theta(\cdot|s))Q^{\pi_\theta}(s,\cdot)\|_2 \tag{349}$$

$$= \frac{2\cdot(C_\infty-(1-\gamma))}{(1-\gamma)^2\cdot\gamma}\cdot\sum_s d_\mu^{\pi_\theta}(s)\cdot\|H(\pi_\theta(\cdot|s))Q^{\pi_\theta}(s,\cdot)\|_2\cdot\|u\|_2^2 \tag{350}$$

$$\leq \frac{2\cdot(C_\infty-(1-\gamma))\cdot\sqrt{S}}{(1-\gamma)^2\cdot\gamma}\cdot\left[\sum_s d_\mu^{\pi_\theta}(s)^2\cdot\|H(\pi_\theta(\cdot|s))Q^{\pi_\theta}(s,\cdot)\|_2^2\right]^{\frac{1}{2}}\cdot\|u\|_2^2 \qquad \text{(by Cauchy-Schwarz)} \tag{351}$$

$$= \frac{2\cdot(C_\infty-(1-\gamma))\cdot\sqrt{S}}{(1-\gamma)\cdot\gamma}\cdot\left\|\frac{\partial V^{\pi_\theta}(\mu)}{\partial\theta}\right\|_2\cdot\|u\|_2^2. \qquad \text{(by Lemma 16)} \tag{352}$$

Combining Eqs. (312), (325) and (348),

$$\left|\frac{\partial^2 V^{\pi_{\theta_\alpha}}(\mu)}{\partial\alpha^2}\Big|_{\alpha=0}\right| \leq \left[3+\frac{2\cdot(C_\infty-(1-\gamma))}{(1-\gamma)\cdot\gamma}\right]\cdot\sqrt{S}\cdot\left\|\frac{\partial V^{\pi_\theta}(\mu)}{\partial\theta}\right\|_2\cdot\|u\|_2^2, \tag{353}$$

which implies for all $y\in\mathbb{R}^{SA}$ and θ,

$$\left|y^\top\frac{\partial^2 V^{\pi_\theta}(\mu)}{\partial\theta^2}y\right| = \left|\left(\frac{y}{\|y\|_2}\right)^\top\frac{\partial^2 V^{\pi_\theta}(\mu)}{\partial\theta^2}\left(\frac{y}{\|y\|_2}\right)\right|\cdot\|y\|_2^2 \tag{354}$$

$$\leq \max_{\|u\|_2=1}\left|\left\langle\frac{\partial^2 V^{\pi_\theta}(\mu)}{\partial\theta^2}u,u\right\rangle\right|\cdot\|y\|_2^2 \tag{355}$$

$$= \max_{\|u\|_2=1}\left|\left\langle\frac{\partial^2 V^{\pi_{\theta_\alpha}}(\mu)}{\partial\theta_\alpha^2}\Big|_{\alpha=0}\frac{\partial\theta_\alpha}{\partial\alpha},\frac{\partial\theta_\alpha}{\partial\alpha}\right\rangle\right|\cdot\|y\|_2^2 \tag{356}$$

$$= \max_{\|u\|_2=1}\left|\left\langle\frac{\partial}{\partial\theta_\alpha}\left\{\frac{\partial V^{\pi_{\theta_\alpha}}(\mu)}{\partial\alpha}\right\}\Big|_{\alpha=0},\frac{\partial\theta_\alpha}{\partial\alpha}\right\rangle\right|\cdot\|y\|_2^2 \tag{357}$$

$$= \max_{\|u\|_2=1}\left|\frac{\partial^2 V^{\pi_{\theta_\alpha}}(\mu)}{\partial\alpha^2}\Big|_{\alpha=0}\right|\cdot\|y\|_2^2 \tag{358}$$

$$\leq \left[3+\frac{2\cdot(C_\infty-(1-\gamma))}{(1-\gamma)\cdot\gamma}\right]\cdot\sqrt{S}\cdot\left\|\frac{\partial V^{\pi_\theta}(\mu)}{\partial\theta}\right\|_2\cdot\|y\|_2^2. \qquad \text{(by Eq. (353))} \tag{359}$$

Denote $\theta_\zeta=\theta+\zeta(\theta'-\theta)$, where $\zeta\in[0,1]$. According to Taylor's theorem, $\forall s,\forall\theta,\ \theta'$,

$$\left|V^{\pi_{\theta'}}(\mu)-V^{\pi_\theta}(\mu)-\left\langle\frac{\partial V^{\pi_\theta}(\mu)}{\partial\theta},\theta'-\theta\right\rangle\right| = \frac{1}{2}\cdot\left|(\theta'-\theta)^\top\frac{\partial^2 V^{\pi_{\theta_\zeta}}(\mu)}{\partial\theta_\zeta^2}(\theta'-\theta)\right| \tag{360}$$

$$\leq \frac{3\cdot(1-\gamma)\cdot\gamma+2\cdot(C_\infty-(1-\gamma))}{2\cdot(1-\gamma)\cdot\gamma}\cdot\sqrt{S}\cdot\left\|\frac{\partial V^{\pi_{\theta_\zeta}}(\mu)}{\partial\theta_\zeta}\right\|_2\cdot\|\theta'-\theta\|_2^2. \qquad \text{(by Eq. (354))} \tag{361}$$

□

Lemma 7. Let $\eta=\frac{(1-\gamma)\cdot\gamma}{6\cdot(1-\gamma)\cdot\gamma+4\cdot(C_\infty-(1-\gamma))}\cdot\frac{1}{\sqrt{S}}$ and

$$\theta'=\theta+\eta\cdot\frac{\partial V^{\pi_\theta}(\mu)}{\partial\theta}\Big/\left\|\frac{\partial V^{\pi_\theta}(\mu)}{\partial\theta}\right\|_2. \tag{362}$$

Denote $\theta_\zeta\coloneqq\theta+\zeta\cdot(\theta'-\theta)$ with some $\zeta\in[0,1]$. We have,

$$\left\|\frac{\partial V^{\pi_{\theta_\zeta}}(\mu)}{\partial\theta_\zeta}\right\|_2\leq 2\cdot\left\|\frac{\partial V^{\pi_\theta}(\mu)}{\partial\theta}\right\|_2. \tag{363}$$

Proof. Using the similar arguments of Lemma 3 (replacing 3 in Lemma 1 with $\frac{3\cdot(1-\gamma)\cdot\gamma+2\cdot(C_\infty-(1-\gamma))}{(1-\gamma)\cdot\gamma}\cdot\sqrt{S}$ in Lemma 5), we have the results. □

Lemma 8 (Non-vanishing NŁ coefficient) **.** Let Assumption 1 hold. We have, $c:=\inf_{s\in\mathcal{S},t\geq1}\pi_{\theta_t}(a^*(s)|s)>0$, where $\{\theta_t\}_{t\geq1}$ is generated by Algorithm 1.

Proof. The proof is similar to Mei et al. (2020b, Lemma 9) and is an extension of the proof for Lemma 4. Denote $\Delta^*(s) = Q^*(s, a^*(s)) - \max_{a \neq a^*(s)} Q^*(s, a) > 0$ as the optimal value gap of state s, where $a^*(s)$ is the action that the optimal policy selects under state s, and $\Delta^* = \min_{s \in \mathcal{S}} \Delta^*(s) > 0$ as the optimal value gap of the MDP. For each state $s \in \mathcal{S}$, define the following sets:

$$\mathcal{R}_1(s) = \left\{ \theta : \frac{\partial V^{\pi_\theta}(\mu)}{\partial \theta(s, a^*(s))} \geq \frac{\partial V^{\pi_\theta}(\mu)}{\partial \theta(s, a)},\ \forall a \neq a^* \right\}, \tag{363}$$

$$\mathcal{R}_2(s) = \left\{ \theta : Q^{\pi_\theta}(s, a^*(s)) \geq Q^*(s, a^*(s)) - \Delta^*(s)/2 \right\}, \tag{364}$$

$$\mathcal{R}_3(s) = \left\{ \theta_t : V^{\pi_{\theta_t}}(s) \geq Q^{\pi_{\theta_t}}(s, a^*(s)) - \Delta^*(s)/2,\ \text{for all } t \geq 1 \text{ large enough} \right\}, \tag{365}$$

$$\mathcal{N}_c(s) = \left\{ \theta : \pi_\theta(a^*(s)|s) \geq \frac{c(s)}{c(s)+1} \right\},\ \text{where } c(s) = \frac{A}{(1-\gamma) \cdot \Delta^*(s)} - 1. \tag{366}$$

Similarly to the previous proof, we have the following claims:

Claim I. $\mathcal{R}_1(s) \cap \mathcal{R}_2(s) \cap \mathcal{R}_3(s)$ is a "nice" region, in the sense that, following a gradient update, (i) if $\theta_t \in \mathcal{R}_1(s) \cap \mathcal{R}_2(s) \cap \mathcal{R}_3(s)$, then $\theta_{t+1} \in \mathcal{R}_1(s) \cap \mathcal{R}_2(s) \cap \mathcal{R}_3(s)$; while we also have (ii) $\pi_{\theta_{t+1}}(a^*(s)|s) \geq \pi_{\theta_t}(a^*(s)|s)$.

Claim II. $\mathcal{N}_c(s) \cap \mathcal{R}_2(s) \cap \mathcal{R}_3(s) \subset \mathcal{R}_1(s) \cap \mathcal{R}_2(s) \cap \mathcal{R}_3(s)$.

Claim III. There exists a finite time $t_0(s) \geq 1$, such that $\theta_{t_0(s)} \in \mathcal{N}_c(s) \cap \mathcal{R}_2(s) \cap \mathcal{R}_3(s)$, and thus $\theta_{t_0(s)} \in \mathcal{R}_1(s) \cap \mathcal{R}_2(s) \cap \mathcal{R}_3(s)$, which implies $\inf_{t \geq 1} \pi_{\theta_t}(a^*(s)|s) = \min_{1 \leq t \leq t_0(s)} \pi_{\theta_t}(a^*(s)|s)$.

Claim IV. Define $t_0 = \max_s t_0(s)$. Then, we have $\inf_{s \in \mathcal{S}, t \geq 1} \pi_{\theta_t}(a^*(s)|s) = \min_{1 \leq t \leq t_0} \min_s \pi_{\theta_t}(a^*(s)|s)$.

Clearly, claim IV suffices to prove the lemma since for any θ, $\min_{s,a} \pi_\theta(a|s) > 0$. In what follows we provide the proofs of these four claims.

Claim I. First we prove part (i) of the claim. If $\theta_t \in \mathcal{R}_1(s) \cap \mathcal{R}_2(s) \cap \mathcal{R}_3(s)$, then $\theta_{t+1} \in \mathcal{R}_1(s) \cap \mathcal{R}_2(s) \cap \mathcal{R}_3(s)$. Suppose $\theta_t \in \mathcal{R}_1(s) \cap \mathcal{R}_2(s) \cap \mathcal{R}_3(s)$. We have $\theta_{t+1} \in \mathcal{R}_3(s)$ by the definition of $\mathcal{R}_3(s)$. We have,

$$Q^{\pi_{\theta_t}}(s, a^*(s)) \geq Q^*(s, a^*(s)) - \Delta^*(s)/2. \tag{367}$$

According to monotonic improvement of Eq. (415), we have $V^{\pi_{\theta_{t+1}}}(s') \geq V^{\pi_{\theta_t}}(s')$, and

$$\begin{aligned}
Q^{\pi_{\theta_{t+1}}}(s, a^*(s)) &= Q^{\pi_{\theta_t}}(s, a^*(s)) + Q^{\pi_{\theta_{t+1}}}(s, a^*(s)) - Q^{\pi_{\theta_t}}(s, a^*(s)) && (368)\\
&= Q^{\pi_{\theta_t}}(s, a^*(s)) + \gamma \sum_{s'} \mathcal{P}(s'|s, a^*(s)) \cdot [V^{\pi_{\theta_{t+1}}}(s') - V^{\pi_{\theta_t}}(s')] && (369)\\
&\geq Q^{\pi_{\theta_t}}(s, a^*(s)) + 0 && (370)\\
&\geq Q^*(s, a^*(s)) - \Delta^*(s)/2, && (371)
\end{aligned}$$

which means $\theta_{t+1} \in \mathcal{R}_2(s)$. Next we prove $\theta_{t+1} \in \mathcal{R}_1(s)$. Note that $\forall a \neq a^*(s)$,

$$\begin{aligned}
Q^{\pi_{\theta_t}}(s, a^*(s)) - Q^{\pi_{\theta_t}}(s, a) &= Q^{\pi_{\theta_t}}(s, a^*(s)) - Q^*(s, a^*(s)) + Q^*(s, a^*(s)) - Q^{\pi_{\theta_t}}(s, a) && (372)\\
&\geq -\Delta^*(s)/2 + Q^*(s, a^*(s)) - Q^*(s, a) + Q^*(s, a) - Q^{\pi_{\theta_t}}(s, a) && (373)\\
&\geq -\Delta^*(s)/2 + Q^*(s, a^*(s)) - \max_{a \neq a^*(s)} Q^*(s, a) + Q^*(s, a) - Q^{\pi_{\theta_t}}(s, a) && (374)\\
&= -\Delta^*(s)/2 + \Delta^*(s) + \gamma \sum_{s'} \mathcal{P}(s'|s, a) \cdot [V^*(s') - V^{\pi_{\theta_t}}(s')] && (375)\\
&\geq -\Delta^*(s)/2 + \Delta^*(s) + 0 && (376)\\
&= \Delta^*(s)/2. && (377)
\end{aligned}$$

Using similar arguments we also have $Q^{\pi_{\theta_{t+1}}}(s, a^*(s)) - Q^{\pi_{\theta_{t+1}}}(s, a) \geq \Delta^*(s)/2$. According to Lemma 15,

$$\begin{aligned}
\frac{\partial V^{\pi_{\theta_t}}(\mu)}{\partial \theta_t(s, a)} &= \frac{1}{1-\gamma} \cdot d_\mu^{\pi_{\theta_t}}(s) \cdot \pi_{\theta_t}(a|s) \cdot A^{\pi_{\theta_t}}(s, a) && (378)\\
&= \frac{1}{1-\gamma} \cdot d_\mu^{\pi_{\theta_t}}(s) \cdot \pi_{\theta_t}(a|s) \cdot [Q^{\pi_{\theta_t}}(s, a) - V^{\pi_{\theta_t}}(s)]. && (379)
\end{aligned}$$

Furthermore, since $\frac{\partial V^{\pi_{\theta_t}}(\mu)}{\partial \theta_t(s, a^*(s))} \geq \frac{\partial V^{\pi_{\theta_t}}(\mu)}{\partial \theta_t(s, a)}$, we have

$$\pi_{\theta_t}(a^*(s)|s) \cdot [Q^{\pi_{\theta_t}}(s, a^*(s)) - V^{\pi_{\theta_t}}(s)] \geq \pi_{\theta_t}(a|s) \cdot [Q^{\pi_{\theta_t}}(s, a) - V^{\pi_{\theta_t}}(s)]. \tag{380}$$

Similarly to the first part in the proof for Lemma 4. There are two cases. Case (a): If $\pi_{\theta_t}(a^*(s)|s) \geq \pi_{\theta_t}(a|s)$, then $\theta_t(s,a^*(s)) \geq \theta_t(s,a)$. After an update of the parameters,

$$\theta_{t+1}(s,a^*(s)) = \theta_t(s,a^*(s)) + \eta \cdot \frac{\partial V^{\pi_{\theta_t}}(\mu)}{\partial \theta_t(s,a^*(s))} \Big/ \left\| \frac{\partial V^{\pi_{\theta_t}}(\mu)}{\partial \theta_t} \right\|_2 \tag{381}$$

$$\geq \theta_t(s,a) + \eta \cdot \frac{\partial V^{\pi_{\theta_t}}(\mu)}{\partial \theta_t(s,a)} \Big/ \left\| \frac{\partial V^{\pi_{\theta_t}}(\mu)}{\partial \theta_t} \right\|_2 = \theta_{t+1}(s,a), \tag{382}$$

which implies $\pi_{\theta_{t+1}}(a^*(s)|s) \geq \pi_{\theta_{t+1}}(a|s)$. Since $Q^{\pi_{\theta_{t+1}}}(s,a^*(s)) - Q^{\pi_{\theta_{t+1}}}(s,a) \geq \Delta^*(s)/2 \geq 0$, $\forall a$, we have $Q^{\pi_{\theta_{t+1}}}(s,a^*(s)) - V^{\pi_{\theta_{t+1}}}(s) = Q^{\pi_{\theta_{t+1}}}(s,a^*(s)) - \sum_a \pi_{\theta_{t+1}}(a|s) \cdot Q^{\pi_{\theta_{t+1}}}(s,a) \geq 0$, and

$$\pi_{\theta_{t+1}}(a^*(s)|s) \cdot [Q^{\pi_{\theta_{t+1}}}(s,a^*(s)) - V^{\pi_{\theta_{t+1}}}(s)] \geq \pi_{\theta_{t+1}}(a|s) \cdot [Q^{\pi_{\theta_{t+1}}}(s,a) - V^{\pi_{\theta_{t+1}}}(s)], \tag{383}$$

which is equivalent to $\frac{\partial V^{\pi_{\theta_{t+1}}}(\mu)}{\partial \theta_{t+1}(s,a^*(s))} \geq \frac{\partial V^{\pi_{\theta_{t+1}}}(\mu)}{\partial \theta_{t+1}(s,a)}$, i.e., $\theta_{t+1} \in \mathcal{R}_1(s)$.

Case (b): If $\pi_{\theta_t}(a^*(s)|s) < \pi_{\theta_t}(a|s)$, then by $\frac{\partial V^{\pi_{\theta_t}}(\mu)}{\partial \theta_t(s,a^*(s))} \geq \frac{\partial V^{\pi_{\theta_t}}(\mu)}{\partial \theta_t(s,a)}$,

$$\pi_{\theta_t}(a^*(s)|s) \cdot [Q^{\pi_{\theta_t}}(s,a^*(s)) - V^{\pi_{\theta_t}}(s)] \geq \pi_{\theta_t}(a|s) \cdot [Q^{\pi_{\theta_t}}(s,a) - V^{\pi_{\theta_t}}(s)] \tag{384}$$

$$= \pi_{\theta_t}(a|s) \cdot [Q^{\pi_{\theta_t}}(s,a^*(s)) - V^{\pi_{\theta_t}}(s) + Q^{\pi_{\theta_t}}(s,a) - Q^{\pi_{\theta_t}}(s,a^*(s))], \tag{385}$$

which, after rearranging, is equivalent to

$$Q^{\pi_{\theta_t}}(s,a^*(s)) - Q^{\pi_{\theta_t}}(s,a) \geq \left(1 - \frac{\pi_{\theta_t}(a^*(s)|s)}{\pi_{\theta_t}(a|s)}\right) \cdot [Q^{\pi_{\theta_t}}(s,a^*(s)) - V^{\pi_{\theta_t}}(s)] \tag{386}$$

$$= (1 - \exp\{\theta_t(s,a^*(s)) - \theta_t(s,a)\}) \cdot [Q^{\pi_{\theta_t}}(s,a^*(s)) - V^{\pi_{\theta_t}}(s)]. \tag{387}$$

Since $\theta_{t+1} \in \mathcal{R}_3(s)$, we have,

$$Q^{\pi_{\theta_{t+1}}}(s,a^*(s)) - V^{\pi_{\theta_{t+1}}}(s) \leq \Delta^*(s)/2 \leq Q^{\pi_{\theta_{t+1}}}(s,a^*(s)) - Q^{\pi_{\theta_{t+1}}}(s,a). \tag{388}$$

On the other hand,

$$\theta_{t+1}(s,a^*(s)) - \theta_{t+1}(s,a) \tag{389}$$

$$= \theta_t(s,a^*(s)) + \eta \cdot \frac{\partial V^{\pi_{\theta_t}}(\mu)}{\partial \theta_t(s,a^*(s))} \Big/ \left\| \frac{\partial V^{\pi_{\theta_t}}(\mu)}{\partial \theta_t} \right\|_2 - \theta_t(s,a) - \eta \cdot \frac{\partial V^{\pi_{\theta_t}}(\mu)}{\partial \theta_t(s,a)} \Big/ \left\| \frac{\partial V^{\pi_{\theta_t}}(\mu)}{\partial \theta_t} \right\|_2 \tag{390}$$

$$\geq \theta_t(s,a^*(s)) - \theta_t(s,a), \tag{391}$$

which implies

$$1 - \exp\{\theta_{t+1}(s,a^*(s)) - \theta_{t+1}(s,a)\} \leq 1 - \exp\{\theta_t(s,a^*(s)) - \theta_t(s,a)\}. \tag{392}$$

Furthermore, since $1 - \exp\{\theta_t(s,a^*(s)) - \theta_t(s,a)\} = 1 - \frac{\pi_{\theta_t}(a^*(s)|s)}{\pi_{\theta_t}(a|s)} > 0$ (in this case $\pi_{\theta_t}(a^*(s)|s) < \pi_{\theta_t}(a|s)$),

$$(1 - \exp\{\theta_{t+1}(s,a^*(s)) - \theta_{t+1}(s,a)\}) \cdot [Q^{\pi_{\theta_{t+1}}}(s,a^*(s)) - V^{\pi_{\theta_{t+1}}}(s)] \leq Q^{\pi_{\theta_{t+1}}}(s,a^*(s)) - Q^{\pi_{\theta_{t+1}}}(s,a), \tag{393}$$

which after rearranging is equivalent to

$$\pi_{\theta_{t+1}}(a^*(s)|s) \cdot [Q^{\pi_{\theta_{t+1}}}(s,a^*(s)) - V^{\pi_{\theta_{t+1}}}(s)] \geq \pi_{\theta_{t+1}}(a|s) \cdot [Q^{\pi_{\theta_{t+1}}}(s,a) - V^{\pi_{\theta_{t+1}}}(s)], \tag{394}$$

which means $\frac{\partial V^{\pi_{\theta_{t+1}}}(\mu)}{\partial \theta_{t+1}(s,a^*(s))} \geq \frac{\partial V^{\pi_{\theta_{t+1}}}(\mu)}{\partial \theta_{t+1}(s,a)}$ i.e., $\theta_{t+1} \in \mathcal{R}_1(s)$. Now we have (i) if $\theta_t \in \mathcal{R}_1(s) \cap \mathcal{R}_2(s) \cap \mathcal{R}_3(s)$, then $\theta_{t+1} \in \mathcal{R}_1(s) \cap \mathcal{R}_2(s) \cap \mathcal{R}_3(s)$.

Let us now turn to proving part (ii). We have $\pi_{\theta_{t+1}}(a^*(s)|s) \geq \pi_{\theta_t}(a^*(s)|s)$. If $\theta_t \in \mathcal{R}_1(s) \cap \mathcal{R}_2(s) \cap \mathcal{R}_3(s)$, then

$\frac{\partial V^{\pi_{\theta_t}}(\mu)}{\partial \theta_t(s,a^*(s))} \geq \frac{\partial V^{\pi_{\theta_t}}(\mu)}{\partial \theta_t(s,a)}$, $\forall a \neq a^*$. After an update of the parameters,

$$\pi_{\theta_{t+1}}(a^*(s)|s) = \frac{\exp\{\theta_{t+1}(s,a^*(s))\}}{\sum_a \exp\{\theta_{t+1}(s,a)\}} \tag{395}$$

$$= \frac{\exp\left\{\theta_t(s,a^*(s)) + \eta \cdot \frac{\partial V^{\pi_{\theta_t}}(\mu)}{\partial \theta_t(s,a^*(s))} \Big/ \left\|\frac{\partial V^{\pi_{\theta_t}}(\mu)}{\partial \theta_t}\right\|_2\right\}}{\sum_a \exp\left\{\theta_t(s,a) + \eta \cdot \frac{\partial V^{\pi_{\theta_t}}(\mu)}{\partial \theta_t(s,a)} \Big/ \left\|\frac{\partial V^{\pi_{\theta_t}}(\mu)}{\partial \theta_t}\right\|_2\right\}} \tag{396}$$

$$\geq \frac{\exp\left\{\theta_t(s,a^*(s)) + \eta \cdot \frac{\partial V^{\pi_{\theta_t}}(\mu)}{\partial \theta_t(s,a^*(s))} \Big/ \left\|\frac{\partial V^{\pi_{\theta_t}}(\mu)}{\partial \theta_t}\right\|_2\right\}}{\sum_a \exp\left\{\theta_t(s,a) + \eta \cdot \frac{\partial V^{\pi_{\theta_t}}(\mu)}{\partial \theta_t(s,a^*(s))} \Big/ \left\|\frac{\partial V^{\pi_{\theta_t}}(\mu)}{\partial \theta_t}\right\|_2\right\}} \qquad \left(\text{because } \frac{\partial V^{\pi_{\theta_t}}(\mu)}{\partial \theta_t(s,a^*(s))} \geq \frac{\partial V^{\pi_{\theta_t}}(\mu)}{\partial \theta_t(s,a)}\right) \tag{397}$$

$$= \frac{\exp\{\theta_t(s,a^*(s))\}}{\sum_a \exp\{\theta_t(s,a)\}} = \pi_{\theta_t}(a^*(s)|s). \tag{398}$$

Claim II, Claim III, Claim IV. The proof of those claims are exactly the same as Mei et al. (2020b, Lemma 9), since they do not involve the update rule. □

Theorem 3. Let Assumption 1 hold and let $\{\theta_t\}_{t\geq 1}$ be generated using Algorithm 1 with

$$\eta = \frac{(1-\gamma)\cdot\gamma}{6\cdot(1-\gamma)\cdot\gamma + 4\cdot(C_\infty - (1-\gamma))} \cdot \frac{1}{\sqrt{S}}, \tag{399}$$

where $C_\infty := \max_\pi \left\|\frac{d_\mu^\pi}{\mu}\right\|_\infty < \infty$. Denote $C'_\infty := \max_\pi \left\|\frac{d_\rho^\pi}{\mu}\right\|_\infty$. Let c be the positive constant from Lemma 8. We have, for all $t \geq 1$,

$$V^*(\rho) - V^{\pi_{\theta_t}}(\rho) \leq \frac{(V^*(\mu) - V^{\pi_{\theta_1}}(\mu))\cdot C'_\infty}{1-\gamma} \cdot e^{-C\cdot(t-1)}, \tag{400}$$

where

$$C = \frac{(1-\gamma)^2\cdot\gamma\cdot c}{12\cdot(1-\gamma)\cdot\gamma + 8\cdot(C_\infty - (1-\gamma))} \cdot \frac{1}{S} \cdot \left\|\frac{d_\mu^{\pi^*}}{\mu}\right\|_\infty^{-1}. \tag{401}$$

Proof. First note that for any θ and μ,

$$d_\mu^{\pi_\theta}(s) = \mathop{\mathbb{E}}_{s_0\sim\mu}\left[d_\mu^{\pi_\theta}(s)\right] \tag{402}$$

$$= \mathop{\mathbb{E}}_{s_0\sim\mu}\left[(1-\gamma)\cdot\sum_{t=0}^{\infty}\gamma^t \Pr(s_t = s|s_0,\pi_\theta,\mathcal{P})\right] \tag{403}$$

$$\geq \mathop{\mathbb{E}}_{s_0\sim\mu}\left[(1-\gamma)\cdot\Pr(s_0 = s|s_0)\right] \tag{404}$$

$$= (1-\gamma)\cdot\mu(s). \tag{405}$$

Next, according to Lemma 18, we have,

$$V^*(\rho) - V^{\pi_\theta}(\rho) = \frac{1}{1-\gamma} \sum_s d_\rho^{\pi_\theta}(s) \sum_a (\pi^*(a|s) - \pi_\theta(a|s)) \cdot Q^*(s,a) \tag{406}$$

$$= \frac{1}{1-\gamma} \sum_s \frac{d_\rho^{\pi_\theta}(s)}{d_\mu^{\pi_\theta}(s)} \cdot d_\mu^{\pi_\theta}(s) \sum_a (\pi^*(a|s) - \pi_\theta(a|s)) \cdot Q^*(s,a) \tag{407}$$

$$\leq \frac{1}{1-\gamma} \cdot \left\| \frac{d_\rho^{\pi_\theta}}{d_\mu^{\pi_\theta}} \right\|_\infty \sum_s d_\mu^{\pi_\theta}(s) \sum_a (\pi^*(a|s) - \pi_\theta(a|s)) \cdot Q^*(s,a) \qquad \left(\sum_a (\pi^*(a|s) - \pi_\theta(a|s)) \cdot Q^*(s,a) \geq 0 \right) \tag{408}$$

$$\leq \frac{1}{(1-\gamma)^2} \cdot \left\| \frac{d_\rho^{\pi_\theta}}{\mu} \right\|_\infty \sum_s d_\mu^{\pi_\theta}(s) \sum_a (\pi^*(a|s) - \pi_\theta(a|s)) \cdot Q^*(s,a) \qquad \left(\text{by Eq. (402) and } \min_s \mu(s) > 0 \right) \tag{409}$$

$$\leq \frac{1}{(1-\gamma)^2} \cdot C'_\infty \cdot \sum_s d_\mu^{\pi_\theta}(s) \sum_a (\pi^*(a|s) - \pi_\theta(a|s)) \cdot Q^*(s,a) \tag{410}$$

$$= \frac{1}{1-\gamma} \cdot C'_\infty \cdot [V^*(\mu) - V^{\pi_\theta}(\mu)]. \qquad \text{(by Lemma 18)} \tag{411}$$

Denote $\theta_{\zeta_t} \coloneqq \theta_t + \zeta_t \cdot (\theta_{t+1} - \theta_t)$ with some $\zeta_t \in [0,1]$. And note $\eta = \frac{(1-\gamma)\cdot\gamma}{6\cdot(1-\gamma)\cdot\gamma + 4\cdot(C_\infty - (1-\gamma))} \cdot \frac{1}{\sqrt{S}}$. According to Lemma 6, we have,

$$\left| V^{\pi_{\theta_{t+1}}}(\mu) - V^{\pi_{\theta_t}}(\mu) - \left\langle \frac{\partial V^{\pi_{\theta_t}}(\mu)}{\partial \theta_t}, \theta_{t+1} - \theta_t \right\rangle \right| \tag{412}$$

$$\leq \frac{3\cdot(1-\gamma)\cdot\gamma + 2\cdot(C_\infty - (1-\gamma))}{2\cdot(1-\gamma)\cdot\gamma} \cdot \sqrt{S} \cdot \left\| \frac{\partial V^{\pi_{\theta_{\zeta_t}}}(\mu)}{\partial \theta_{\zeta_t}} \right\|_2 \cdot \|\theta_{t+1} - \theta_t\|_2^2 \tag{413}$$

$$\leq \frac{3\cdot(1-\gamma)\cdot\gamma + 2\cdot(C_\infty - (1-\gamma))}{(1-\gamma)\cdot\gamma} \cdot \sqrt{S} \cdot \left\| \frac{\partial V^{\pi_{\theta_t}}(\mu)}{\partial \theta_t} \right\|_2 \cdot \|\theta_{t+1} - \theta_t\|_2^2. \qquad \text{(by Lemma 7)} \tag{414}$$

Denote $\delta_t = V^*(\mu) - V^{\pi_{\theta_t}}(\mu)$. We have,

$$\delta_{t+1} - \delta_t = V^{\pi_{\theta_t}}(\mu) - V^{\pi_{\theta_{t+1}}}(\mu) \tag{415}$$

$$\leq -\left\langle \frac{\partial V^{\pi_{\theta_t}}(\mu)}{\partial \theta_t}, \theta_{t+1} - \theta_t \right\rangle + \frac{3\cdot(1-\gamma)\cdot\gamma + 2\cdot(C_\infty - (1-\gamma))}{(1-\gamma)\cdot\gamma} \cdot \sqrt{S} \cdot \left\| \frac{\partial V^{\pi_{\theta_t}}(\mu)}{\partial \theta_t} \right\|_2 \cdot \|\theta_{t+1} - \theta_t\|_2^2 \tag{416}$$

$$= -\frac{(1-\gamma)\cdot\gamma}{12\cdot(1-\gamma)\cdot\gamma + 8\cdot(C_\infty - (1-\gamma))} \cdot \frac{1}{\sqrt{S}} \cdot \left\| \frac{\partial V^{\pi_{\theta_t}}(\mu)}{\partial \theta_t} \right\|_2 \qquad \text{(using the value of } \eta) \tag{417}$$

$$\leq -\frac{(1-\gamma)\cdot\gamma}{12\cdot(1-\gamma)\cdot\gamma + 8\cdot(C_\infty - (1-\gamma))} \cdot \frac{1}{\sqrt{S}} \cdot \frac{\min_s \pi_{\theta_t}(a^*(s)|s)}{\sqrt{S} \cdot \left\| d_\mu^{\pi^*} / d_\mu^{\pi_{\theta_t}} \right\|_\infty} \cdot \delta_t \qquad \text{(by Lemma 5)} \tag{418}$$

$$\leq -\frac{(1-\gamma)^2\cdot\gamma}{12\cdot(1-\gamma)\cdot\gamma + 8\cdot(C_\infty - (1-\gamma))} \cdot \frac{1}{S} \cdot \left\| \frac{d_\mu^{\pi^*}}{\mu} \right\|_\infty^{-1} \cdot \inf_{s\in\mathcal{S}, t\geq 1} \pi_{\theta_t}(a^*(s)|s) \cdot \delta_t, \tag{419}$$

where the last inequality is by $d_\mu^{\pi_{\theta_t}}(s) \geq (1-\gamma)\cdot\mu(s)$ (cf. Eq. (402)). According to Lemma 8, $c = \inf_{s\in\mathcal{S}, t\geq 1} \pi_{\theta_t}(a^*(s)|s) > 0$. Therefore we have,

$$V^*(\mu) - V^{\pi_{\theta_t}}(\mu) \leq (V^*(\mu) - V^{\pi_{\theta_1}}(\mu)) \cdot \exp\left\{ -\frac{(1-\gamma)^2\cdot\gamma\cdot c\cdot(t-1)}{12\cdot(1-\gamma)\cdot\gamma + 8\cdot(C_\infty - (1-\gamma))} \cdot \frac{1}{S} \cdot \left\| \frac{d_\mu^{\pi^*}}{\mu} \right\|_\infty^{-1} \right\}, \tag{420}$$

which leads to the final result,

$$V^*(\rho) - V^{\pi_{\theta_t}}(\rho) \leq \frac{(V^*(\mu) - V^{\pi_{\theta_1}}(\mu)) \cdot C'_\infty}{1-\gamma} \cdot \exp\left\{ -\frac{(1-\gamma)^2\cdot\gamma\cdot c\cdot(t-1)}{12\cdot(1-\gamma)\cdot\gamma + 8\cdot(C_\infty - (1-\gamma))} \cdot \frac{1}{S} \cdot \left\| \frac{d_\mu^{\pi^*}}{\mu} \right\|_\infty^{-1} \right\}, \tag{421}$$

thus, finishing the proof. □

C. Proofs for Section 6

Lemma 9 (NŁ) **.** Denote $u(\theta) \coloneqq \min_i \{\pi_i \cdot (1-\pi_i)\}$, and $v \coloneqq \min_i \{\pi_i^* \cdot (1-\pi_i^*)\}$. We have, for all $i \in [N]$,

$$\left\| \frac{\partial \mathcal{L}(\theta)}{\partial \theta} \right\|_2 \geq 8 \cdot u(\theta) \cdot \min\{u(\theta), v\} \cdot \sqrt{\lambda_\phi} \cdot \left[\frac{1}{N} \cdot \sum_{i=1}^N (\pi_i - \pi_i^*)^2 \right]^{\frac{1}{2}}, \tag{422}$$

where λ_ϕ is the smallest positive eigenvalue of $\frac{1}{N} \cdot \sum_{i=1}^N \phi_i \phi_i^\top$.

Proof. Denote $\pi_i' \coloneqq \sigma(z_i')$, where $z_i' \coloneqq \phi_i^\top \theta + \zeta \cdot \left(\phi_i^\top \theta - \phi_i^\top \theta^*\right)$ for some $\zeta \in [0,1]$. We have,

$$(\pi_i - \pi_i^*)^2 = (\pi_i - \pi_i^*) \cdot \frac{d\sigma(z_i')}{dz_i'} \cdot \left(\phi_i^\top \theta - \phi_i^\top \theta^*\right) \qquad \text{(by the mean value theorem)} \tag{423}$$

$$= \pi_i' \cdot (1-\pi_i') \cdot (\pi_i - \pi_i^*) \cdot \left(\phi_i^\top \theta - \phi_i^\top \theta^*\right) \tag{424}$$

$$\leq \frac{1}{4} \cdot (\pi_i - \pi_i^*) \cdot \left(\phi_i^\top \theta - \phi_i^\top \theta^*\right). \qquad \left(x \cdot (1-x) \leq \frac{1}{4},\ \forall x \in [0,1];\ (\pi_i - \pi_i^*) \cdot \left(\phi_i^\top \theta - \phi_i^\top \theta^*\right) \geq 0\right) \tag{425}$$

Therefore we have,

$$\frac{1}{N} \cdot \sum_{i=1}^N (\pi_i - \pi_i^*)^2 \leq \frac{1}{4N} \cdot \sum_{i=1}^N (\pi_i - \pi_i^*) \cdot \left(\phi_i^\top \theta - \phi_i^\top \theta^*\right) \qquad \text{(by Eq. (423))} \tag{426}$$

$$= \frac{1}{4N} \cdot \sum_{i=1}^N \frac{1}{\pi_i \cdot (1-\pi_i)} \cdot \pi_i \cdot (1-\pi_i) \cdot (\pi_i - \pi_i^*) \cdot \left(\phi_i^\top \theta - \phi_i^\top \theta^*\right) \tag{427}$$

$$\leq \frac{1}{4N} \cdot \frac{1}{\min_i \pi_i \cdot (1-\pi_i)} \cdot \sum_{i=1}^N \pi_i \cdot (1-\pi_i) \cdot (\pi_i - \pi_i^*) \cdot \left(\phi_i^\top \theta - \phi_i^\top \theta^*\right) \qquad \left((\pi_i - \pi_i^*) \cdot \left(\phi_i^\top \theta - \phi_i^\top \theta^*\right) \geq 0\right) \tag{428}$$

$$= \frac{1}{8} \cdot \frac{1}{\min_i \pi_i \cdot (1-\pi_i)} \cdot \left(\frac{2}{N} \cdot \sum_{i=1}^N \pi_i \cdot (1-\pi_i) \cdot (\pi_i - \pi_i^*) \cdot \phi_i \right)^\top (\theta - \theta^* - c \cdot v_{\phi,\perp}) \tag{429}$$

$$= \frac{1}{8} \cdot \frac{1}{\min_i \pi_i \cdot (1-\pi_i)} \cdot \left(\frac{\partial \mathcal{L}(\theta)}{\partial \theta} \right)^\top (\theta - \theta^* - c \cdot v_{\phi,\perp}) \qquad \left(\frac{\partial \mathcal{L}(\theta)}{\partial \theta} = \frac{2}{N} \cdot \sum_{i=1}^N \pi_i \cdot (1-\pi_i) \cdot (\pi_i - \pi_i^*) \cdot \phi_i \right) \tag{430}$$

$$\leq \frac{1}{8} \cdot \frac{1}{\min_i \pi_i \cdot (1-\pi_i)} \cdot \left\| \frac{\partial \mathcal{L}(\theta)}{\partial \theta} \right\|_2 \cdot \|\theta - \theta^* - c \cdot v_{\phi,\perp}\|_2 \qquad \text{(by Cauchy-Schwarz)} \tag{431}$$

$$= \frac{1}{8} \cdot \frac{1}{u(\theta)} \cdot \left\| \frac{\partial \mathcal{L}(\theta)}{\partial \theta} \right\|_2 \cdot \|\theta - \theta^* - c \cdot v_{\phi,\perp}\|_2, \qquad \left(u(\theta) \coloneqq \min_i \{\pi_i \cdot (1-\pi_i)\}\right) \tag{432}$$

where $v_{\phi,\perp}$ is orthogonal to the space Span $\{\phi_1, \phi_2, \ldots, \phi_N\}$, and $\theta - \theta^* - c \cdot v_{\phi,\perp}$ refers to the vector after cutting off all

the components $v_{\phi,\perp}$ from $\theta-\theta^*$, such that $\theta-\theta^*-c\cdot v_{\phi,\perp}\in \text{Span}\left\{\phi_1,\phi_2,\ldots,\phi_N\right\}$. Next, we have,

$$\frac{1}{N}\cdot\sum_{i=1}^{N}\left(\pi_i-\pi_i^*\right)^2=\frac{1}{N}\cdot\sum_{i=1}^{N}\left(\frac{d\sigma(z_i')}{dz_i'}\right)^2\cdot\left(\phi_i^\top\theta-\phi_i^\top\theta^*\right)^2 \qquad \text{(by the mean value theorem)} \tag{433}$$

$$=\frac{1}{N}\cdot\sum_{i=1}^{N}\left(\pi_i'\right)^2\cdot\left(1-\pi_i'\right)^2\cdot\left(\phi_i^\top\theta-\phi_i^\top\theta^*\right)^2 \qquad \text{(by Eq. (423))} \tag{434}$$

$$\geq\min_i\left\{\left(\pi_i'\right)^2\cdot\left(1-\pi_i'\right)^2\right\}\cdot\frac{1}{N}\cdot\sum_{i=1}^{N}\left(\phi_i^\top\theta-\phi_i^\top\theta^*\right)^2 \tag{435}$$

$$=\min_i\left\{\left(\pi_i'\right)^2\cdot\left(1-\pi_i'\right)^2\right\}\cdot\left(\theta-\theta^*\right)^\top\left(\frac{1}{N}\cdot\sum_{i=1}^{N}\phi_i\phi_i^\top\right)\left(\theta-\theta^*\right) \tag{436}$$

$$=\min_i\left\{\left(\pi_i'\right)^2\cdot\left(1-\pi_i'\right)^2\right\}\cdot\left(\theta-\theta^*-c\cdot v_{\phi,\perp}\right)^\top\left(\frac{1}{N}\cdot\sum_{i=1}^{N}\phi_i\phi_i^\top\right)\left(\theta-\theta^*-c\cdot v_{\phi,\perp}\right) \tag{437}$$

$$\geq\min\left\{u(\theta)^2,v^2\right\}\cdot\left(\theta-\theta^*-c\cdot v_{\phi,\perp}\right)^\top\left(\frac{1}{N}\cdot\sum_{i=1}^{N}\phi_i\phi_i^\top\right)\left(\theta-\theta^*-c\cdot v_{\phi,\perp}\right)\qquad\left(v\coloneqq\min_i\left\{\pi_i^*\cdot\left(1-\pi_i^*\right)\right\}\right) \tag{438}$$

$$\geq\min\left\{u(\theta)^2,v^2\right\}\cdot\lambda_\phi\cdot\left\|\theta-\theta^*-c\cdot v_{\phi,\perp}\right\|_2^2, \tag{439}$$

where λ_ϕ is the smallest positive eigenvalue of $\frac{1}{N}\cdot\sum_{i=1}^{N}\phi_i\phi_i^\top$. Therefore, we have,

$$\frac{1}{N}\cdot\sum_{i=1}^{N}\left(\pi_i-\pi_i^*\right)^2\leq\frac{1}{8}\cdot\frac{1}{u(\theta)}\cdot\left\|\frac{\partial\mathcal{L}(\theta)}{\partial\theta}\right\|_2\cdot\left\|\theta-\theta^*-c\cdot v_{\phi,\perp}\right\|_2 \qquad \text{(by Eq. (426))} \tag{440}$$

$$\leq\frac{1}{8}\cdot\frac{1}{u(\theta)}\cdot\left\|\frac{\partial\mathcal{L}(\theta)}{\partial\theta}\right\|_2\cdot\frac{1}{\min\left\{u(\theta),v\right\}}\cdot\frac{1}{\sqrt{\lambda_\phi}}\cdot\left[\frac{1}{N}\cdot\sum_{i=1}^{N}\left(\pi_i-\pi_i^*\right)^2\right]^{\frac{1}{2}}, \qquad \text{(by Eq. (433))} \tag{441}$$

which implies,

$$\left\|\frac{\partial\mathcal{L}(\theta)}{\partial\theta}\right\|_2\geq 8\cdot u(\theta)\cdot\min\left\{u(\theta),v\right\}\cdot\sqrt{\lambda_\phi}\cdot\left[\frac{1}{N}\cdot\sum_{i=1}^{N}\left(\pi_i-\pi_i^*\right)^2\right]^{\frac{1}{2}}. \qquad \square$$

Lemma 10. Denote $u(\theta)\coloneqq\min_i\left\{\pi_i\cdot(1-\pi_i)\right\}$, $v\coloneqq\min_i\left\{\pi_i^*\cdot(1-\pi_i^*)\right\}$, and λ_ϕ is the smallest positive eigenvalue of $\frac{1}{N}\cdot\sum_{i=1}^{N}\phi_i\phi_i^\top$. We have, $\mathcal{L}(\theta)$ satisfies β smoothness with

$$\beta=\frac{3}{8}\cdot\max_{i\in[N]}\left\|\phi_i\right\|_2^2, \tag{442}$$

and $\beta(\theta)$ NS with

$$\beta(\theta)=L_1\cdot\left\|\frac{\partial\mathcal{L}(\theta)}{\partial\theta}\right\|_2+L_0\cdot\left(\left\|\frac{\partial\mathcal{L}(\theta)}{\partial\theta}\right\|_2^2\Big/\mathcal{L}(\theta)\right). \tag{443}$$

where

$$L_1=\frac{\max_i\left\|\phi_i\right\|_2^2}{32\cdot(\min\{u(\theta),v\}\cdot\sqrt{\lambda_\phi})^{3/2}},\ \text{and}\ L_0=\frac{17\cdot\max_i\left\|\phi_i\right\|_2^2}{512\cdot u(\theta)^2\cdot\min\left\{u(\theta)^2,v^2\right\}\cdot\lambda_\phi}. \tag{444}$$

Proof. Note that the gradient of $\mathcal{L}(\theta)$ is,

$$\frac{\partial\mathcal{L}(\theta)}{\partial\theta}=\frac{2}{N}\cdot\sum_{i=1}^{N}\pi_i\cdot(1-\pi_i)\cdot(\pi_i-\pi_i^*)\cdot\phi_i\in\mathbb{R}^d. \tag{445}$$

Denote the second order derivative (Hessian) of $\mathcal{L}(\theta)$ as,

$$S(\theta)\coloneqq\frac{\partial}{\partial\theta}\left\{\frac{\partial\mathcal{L}(\theta)}{\partial\theta}\right\}\in\mathbb{R}^{d\times d}. \tag{446}$$

For all $j, k \in [d]$, we calculate the corresponding component value of $S(\theta)$ matrix as follows,

$$S_{(j,k)} = \frac{d}{d\theta(k)} \left\{ \frac{2}{N} \cdot \sum_{i=1}^{N} \pi_i \cdot (1-\pi_i) \cdot (\pi_i - \pi_i^*) \cdot \phi_i(j) \right\} \tag{447}$$

$$= \frac{2}{N} \cdot \sum_{i=1}^{N} \frac{d\left\{\pi_i \cdot (1-\pi_i) \cdot (\pi_i - \pi_i^*)\right\}}{d\theta(k)} \cdot \phi_i(j) \tag{448}$$

$$= \frac{2}{N} \cdot \sum_{i=1}^{N} \frac{d\left\{\pi_i \cdot (1-\pi_i) \cdot (\pi_i - \pi_i^*)\right\}}{d\left\{\phi_i^\top \theta\right\}} \cdot \frac{d\left\{\phi_i^\top \theta\right\}}{d\theta(k)} \cdot \phi_i(j) \tag{449}$$

$$= \frac{2}{N} \cdot \sum_{i=1}^{N} \left[\pi_i \cdot (1-\pi_i)^2 \cdot (\pi_i - \pi_i^*) - \pi_i^2 \cdot (1-\pi_i) \cdot (\pi_i - \pi_i^*) + \pi_i^2 \cdot (1-\pi_i)^2\right] \cdot \phi_i(k) \cdot \phi_i(j) \tag{450}$$

$$= \frac{2}{N} \cdot \sum_{i=1}^{N} \left[\pi_i \cdot (1-\pi_i) \cdot (1-2\pi_i) \cdot (\pi_i - \pi_i^*) + \pi_i^2 \cdot (1-\pi_i)^2\right] \cdot \phi_i(k) \cdot \phi_i(j). \tag{451}$$

To calculate the smoothness coefficient, take a vector $z \in \mathbb{R}^d$. We have,

$$\left|z^\top S(\theta) z\right| = \left| \sum_{j=1}^{d} \sum_{k=1}^{d} S_{(j,k)} \cdot z(j) \cdot z(k) \right| \tag{452}$$

$$= \left| \frac{2}{N} \cdot \sum_{i=1}^{N} \left[\pi_i \cdot (1-\pi_i) \cdot (1-2\pi_i) \cdot (\pi_i - \pi_i^*) + \pi_i^2 \cdot (1-\pi_i)^2\right] \cdot \left(\phi_i^\top z\right)^2 \right| \qquad \text{(by Eq. (447))} \tag{453}$$

$$\leq \frac{2}{N} \cdot \max_i \left(\phi_i^\top z\right)^2 \cdot \sum_{i=1}^{N} \left| \pi_i \cdot (1-\pi_i) \cdot (1-2\pi_i) \cdot (\pi_i - \pi_i^*) + \pi_i^2 \cdot (1-\pi_i)^2 \right| \qquad \text{(by Hölder's inequality)} \tag{454}$$

$$\leq \frac{2}{N} \cdot \max_i \left(\phi_i^\top z\right)^2 \cdot \sum_{i=1}^{N} \left[\pi_i \cdot (1-\pi_i) \cdot |1-2\pi_i| \cdot |\pi_i - \pi_i^*| + \pi_i^2 \cdot (1-\pi_i)^2\right] \qquad \text{(by triangle inequality)} \tag{455}$$

$$\leq \frac{2}{N} \cdot \max_i \left(\phi_i^\top z\right)^2 \cdot \sum_{i=1}^{N} \left[\frac{1}{8} + \frac{1}{16}\right] \qquad (x \cdot (1-x) \leq 1/4, \text{ and } x \cdot (1-x) \cdot |1-2x| \leq 1/8, \ \forall x \in [0,1]) \tag{456}$$

$$= \frac{3}{8} \cdot \max_i \left[\phi_i^\top \left(\frac{z}{\|z\|_2}\right)\right]^2 \cdot \|z\|_2^2 \tag{457}$$

$$\leq \frac{3}{8} \cdot \max_i \|\phi_i\|_2^2 \cdot \|z\|_2^2 . \tag{458}$$

Therefore, $\mathcal{L}(\theta)$ satisfies β (uniform) smoothness with $\beta = \frac{3}{8} \cdot \max_i \|\phi_i\|_2^2$. Next, we calculate the NS. We have,

$$\sum_{i=1}^{N} \pi_i^2 \cdot (1-\pi_i)^2 \cdot \mathcal{L}(\theta) = \sum_{i=1}^{N} \pi_i^2 \cdot (1-\pi_i)^2 \cdot \frac{1}{N} \cdot \sum_{j=1}^{N} (\pi_j - \pi_j^*)^2 \tag{459}$$

$$\leq \frac{N}{16} \cdot \frac{1}{N} \cdot \sum_{j=1}^{N} (\pi_j - \pi_j^*)^2 \tag{460}$$

$$\leq \frac{N}{16} \cdot \frac{1}{64 \cdot u(\theta)^2 \cdot \min\left\{u(\theta)^2, v^2\right\} \cdot \lambda_\phi} \cdot \left\| \frac{\partial \mathcal{L}(\theta)}{\partial \theta} \right\|_2^2, \qquad \text{(by Lemma 9)} \tag{461}$$

which implies,

$$\sum_{i=1}^{N} \pi_i^2 \cdot (1-\pi_i)^2 \leq \frac{N}{2} \cdot \frac{1}{512 \cdot u(\theta)^2 \cdot \min\left\{u(\theta)^2, v^2\right\} \cdot \lambda_\phi} \cdot \left\| \frac{\partial \mathcal{L}(\theta)}{\partial \theta} \right\|_2^2 \Big/ \mathcal{L}(\theta). \tag{462}$$

According to Eq. (433), we have

$$\sum_{i=1}^{N}\frac{(\pi_i-\pi_i^*)^2}{\sqrt{\mathcal{L}(\theta)}\cdot\|\theta-\theta^*-c\cdot v_{\phi,\perp}\|_2^{3/2}}\geq\sum_{i=1}^{N}\frac{(\pi_i-\pi_i^*)^2}{\sqrt{\mathcal{L}(\theta)}}\cdot(\min\{u(\theta)^2,v^2\}\cdot\lambda_\phi)^{3/4}\cdot\frac{1}{\mathcal{L}(\theta)^{3/4}} \tag{463}$$

$$=(\min\{u(\theta)^2,v^2\}\cdot\lambda_\phi)^{3/4}\cdot\sum_{i=1}^{N}\frac{(\pi_i-\pi_i^*)^2}{\mathcal{L}(\theta)^{5/4}} \tag{464}$$

$$=N\cdot(\min\{u(\theta)^2,v^2\}\cdot\lambda_\phi)^{3/4}\cdot\frac{\mathcal{L}(\theta)}{\mathcal{L}(\theta)^{5/4}} \tag{465}$$

$$\geq N\cdot(\min\{u(\theta),v\}\cdot\sqrt{\lambda_\phi})^{3/2}.\qquad(\mathcal{L}(\theta)\in(0,1]) \tag{466}$$

Therefore we have,

$$\sum_{i=1}^{N}\pi_i\cdot(1-\pi_i)\cdot|1-2\pi_i|\cdot|\pi_i-\pi_i^*|\leq\sum_{i=1}^{N}\pi_i\cdot(1-\pi_i)\cdot|\pi_i-\pi_i^*| \tag{467}$$

$$\leq\left(\sum_{i=1}^{N}\pi_i\cdot(1-\pi_i)\cdot|\pi_i-\pi_i^*|\right)\cdot\left(\sum_{i=1}^{N}\frac{(\pi_i-\pi_i^*)^2}{\sqrt{\mathcal{L}(\theta)}\cdot\|\theta-\theta^*-c\cdot v_{\phi,\perp}\|_2^{3/2}}\right)\cdot\frac{1}{N\cdot(\min\{u(\theta),v\}\cdot\sqrt{\lambda_\phi})^{3/2}} \tag{468}$$

$$=\frac{1}{N\cdot(\min\{u(\theta),v\}\cdot\sqrt{\lambda_\phi})^{3/2}}\cdot\left(\sum_{i=1}^{N}\frac{\pi_i\cdot(1-\pi_i)\cdot|\pi_i-\pi_i^*|}{\sqrt{\|\theta-\theta^*-c\cdot v_{\phi,\perp}\|_2}}\right)\cdot\left(\sum_{i=1}^{N}\frac{(\pi_i-\pi_i^*)^2}{\sqrt{\mathcal{L}(\theta)}\cdot\|\theta-\theta^*-c\cdot v_{\phi,\perp}\|_2}\right) \tag{469}$$

$$\leq\frac{1}{(\min\{u(\theta),v\}\cdot\sqrt{\lambda_\phi})^{3/2}}\cdot\left(\sum_{i=1}^{N}\frac{\pi_i^2\cdot(1-\pi_i)^2\cdot(\pi_i-\pi_i^*)^2}{2\cdot\|\theta-\theta^*-c\cdot v_{\phi,\perp}\|_2}+\frac{(\pi_i-\pi_i^*)^4}{2\cdot\mathcal{L}(\theta)\cdot\|\theta-\theta^*-c\cdot v_{\phi,\perp}\|_2^2}\right) \tag{470}$$

$$\leq\frac{1}{(\min\{u(\theta),v\}\cdot\sqrt{\lambda_\phi})^{3/2}}\cdot\left(\frac{1}{32}\cdot\sum_{i=1}^{N}\frac{\pi_i\cdot(1-\pi_i)\cdot(\pi_i-\pi_i^*)\cdot(\phi_i^\top\theta-\phi_i^\top\theta^*)}{\|\theta-\theta^*-c\cdot v_{\phi,\perp}\|_2}\right) \tag{471}$$

$$+\frac{1}{(\min\{u(\theta),v\}\cdot\sqrt{\lambda_\phi})^{3/2}}\cdot\left(\frac{1}{32\cdot u(\theta)^2}\cdot\sum_{i=1}^{N}\frac{\pi_i^2\cdot(1-\pi_i)^2\cdot(\pi_i-\pi_i^*)^2\cdot(\phi_i^\top\theta-\phi_i^\top\theta^*)^2}{\mathcal{L}(\theta)\cdot\|\theta-\theta^*-c\cdot v_{\phi,\perp}\|^2}\right) \tag{472}$$

$$\leq\frac{N}{64\cdot(\min\{u(\theta),v\}\cdot\sqrt{\lambda_\phi})^{3/2}}\cdot\left\|\frac{\partial\mathcal{L}(\theta)}{\partial\theta}\right\|_2+\frac{N}{64\cdot u(\theta)^2\cdot(\min\{u(\theta),v\}\cdot\sqrt{\lambda_\phi})^{3/2}}\cdot\left\|\frac{\partial\mathcal{L}(\theta)}{\partial\theta}\right\|_2^2\Big/\mathcal{L}(\theta), \tag{473}$$

where the second inequality is according to,

$$\left(\sum_{i=1}^{N}a_i\right)\cdot\left(\sum_{i=1}^{N}b_i\right)=\sum_{i=1}^{N}\sum_{j=1}^{N}a_i\cdot b_j\leq\frac{1}{2}\cdot\sum_{i=1}^{N}\sum_{j=1}^{N}\left(a_i^2+b_j^2\right)=\frac{N}{2}\cdot\sum_{i=1}^{N}\left(a_i^2+b_i^2\right), \tag{474}$$

and the last inequality is from the intermediate results in Eq. (426). Combining Eqs. (452), (462) and (467), we have

$$\left|z^\top S(\theta)z\right|\leq\frac{2}{N}\cdot\max_i\left(\phi_i^\top z\right)^2\cdot\left[\sum_{i=1}^{N}\pi_i\cdot(1-\pi_i)\cdot|\pi_i-\pi_i^*|+\sum_{i=1}^{N}\pi_i^2\cdot(1-\pi_i)^2\right] \tag{475}$$

$$\leq\max_i\left(\phi_i^\top z\right)^2\cdot\left(\frac{1}{32\cdot(\min\{u(\theta),v\}\cdot\sqrt{\lambda_\phi})^{3/2}}\cdot\left\|\frac{\partial\mathcal{L}(\theta)}{\partial\theta}\right\|_2+\frac{17}{512\cdot u(\theta)^2\cdot\min\left\{u(\theta)^2,v^2\right\}\cdot\lambda_\phi}\cdot\left\|\frac{\partial\mathcal{L}(\theta)}{\partial\theta}\right\|_2^2\Big/\mathcal{L}(\theta)\right) \tag{476}$$

$$\leq\max_i\|\phi_i\|_2^2\cdot\|z\|_2^2\cdot\left(\frac{1}{32\cdot(\min\{u(\theta),v\}\cdot\sqrt{\lambda_\phi})^{3/2}}\cdot\left\|\frac{\partial\mathcal{L}(\theta)}{\partial\theta}\right\|_2+\frac{17}{512\cdot u(\theta)^2\cdot\min\left\{u(\theta)^2,v^2\right\}\cdot\lambda_\phi}\cdot\left\|\frac{\partial\mathcal{L}(\theta)}{\partial\theta}\right\|_2^2\Big/\mathcal{L}(\theta)\right). \tag{477}$$

Therefore, $\mathcal{L}(\theta)$ satisfies $\beta(\theta)$ NS with

$$\beta(\theta)=L_1\cdot\left\|\frac{\partial\mathcal{L}(\theta)}{\partial\theta}\right\|_2+L_0\cdot\left(\left\|\frac{\partial\mathcal{L}(\theta)}{\partial\theta}\right\|_2^2\Big/\mathcal{L}(\theta)\right), \tag{478}$$

where

$$L_1 = \frac{\max_i \|\phi_i\|_2^2}{32 \cdot (\min\{u(\theta), v\} \cdot \sqrt{\lambda_\phi})^{3/2}}, \text{ and } L_0 = \frac{17 \cdot \max_i \|\phi_i\|_2^2}{512 \cdot u(\theta)^2 \cdot \min\{u(\theta)^2, v^2\} \cdot \lambda_\phi}. \qquad \square$$

Theorem 5. With $\eta = 1/\beta$, GD update satisfies for all $t \geq 1$, $\mathcal{L}(\theta_t) \leq \mathcal{L}(\theta_1) \cdot e^{-C^2 \cdot (t-1)}$. With $\eta \in \Theta(1)$, GNGD update satisfies for all $t \geq 1$, $\mathcal{L}(\theta_t) \leq \mathcal{L}(\theta_1) \cdot e^{-C \cdot (t-1)}$, where $C \in (0, 1)$, i.e., GNGD is strictly faster than GD.

Proof. Combining Lemmas 9 and 10, and the second part of (2b) in Theorem 1, we have the results for GD. Using the fourth part of (2b) in Theorem 1, we have the results for GNGD. $\square$

D. Miscellaneous Extra Supporting Results

Lemma 11 (Descent lemma for smooth function)**.** *Let $f : \mathbb{R}^d \to \mathbb{R}$ be a β-smooth function, $\theta \in \mathbb{R}^d$ and $\theta' = \theta - \eta \cdot \frac{\partial f(\theta)}{\partial \theta}$. We have, for any $0 < \eta < 2/\beta$,*

$$f(\theta') \leq f(\theta). \tag{479}$$

In particular, for $\eta = \frac{1}{\beta}$, we have,

$$f(\theta') \leq f(\theta) - \frac{1}{2\beta} \cdot \left\| \frac{\partial f(\theta)}{\partial \theta} \right\|_2^2. \tag{480}$$

Proof. According to Definition 1, we have,

$$\left| f(\theta') - f(\theta) - \left\langle \frac{\partial f(\theta)}{\partial \theta}, \theta' - \theta \right\rangle \right| \leq \frac{\beta}{2} \cdot \|\theta' - \theta\|_2^2, \tag{481}$$

which implies,

$$f(\theta') - f(\theta) \leq \left\langle \frac{\partial f(\theta)}{\partial \theta}, \theta' - \theta \right\rangle + \frac{\beta}{2} \cdot \|\theta' - \theta\|_2^2 \tag{482}$$

$$= \eta \cdot \left(-1 + \frac{\beta}{2} \cdot \eta \right) \cdot \left\| \frac{\partial f(\theta)}{\partial \theta} \right\|_2^2 \qquad \left(\theta' = \theta - \eta \cdot \frac{\partial f(\theta)}{\partial \theta} \right) \tag{483}$$

$$\leq 0 \qquad \left(0 < \eta < \frac{2}{\beta} \right). \tag{484}$$

Let $\eta = \frac{1}{\beta}$ in Eq. (483), we have Eq. (480). $\square$

Lemma 12 (Descent lemma for NS function)**.** *Let $f : \mathbb{R}^d \to \mathbb{R}$ be a function that satisfies NS with $\beta(\theta) > 0$, for all $\theta \in \mathbb{R}^d$ and $\theta' = \theta - \frac{1}{\beta(\theta)} \cdot \frac{\partial f(\theta)}{\partial \theta}$. We have,*

$$f(\theta') \leq f(\theta) - \frac{1}{2 \cdot \beta(\theta)} \cdot \left\| \frac{\partial f(\theta)}{\partial \theta} \right\|_2^2. \tag{485}$$

Proof. According to Definition 3, we have,

$$f(\theta') - f(\theta) \leq \left\langle \frac{\partial f(\theta)}{\partial \theta}, \theta' - \theta \right\rangle + \frac{\beta(\theta)}{2} \cdot \|\theta' - \theta\|_2^2 \tag{486}$$

$$= -\frac{1}{\beta(\theta)} \cdot \left\| \frac{\partial f(\theta)}{\partial \theta} \right\|_2^2 + \frac{1}{2 \cdot \beta(\theta)} \cdot \left\| \frac{\partial f(\theta)}{\partial \theta} \right\|_2^2 \qquad \left(\theta' = \theta - \frac{1}{\beta(\theta)} \cdot \frac{\partial f(\theta)}{\partial \theta} \right) \tag{487}$$

$$= -\frac{1}{2 \cdot \beta(\theta)} \cdot \left\| \frac{\partial f(\theta)}{\partial \theta} \right\|_2^2. \qquad \square$$

Lemma 13. *Given any $\alpha > 0$, we have, for all $x \in [0, 1]$,*

$$\frac{1}{\alpha} \cdot (1 - x^\alpha) \geq x^\alpha \cdot (1 - x). \tag{488}$$

Proof. Define $f : x \mapsto \frac{1}{\alpha} \cdot (1 - x^\alpha) - x^\alpha \cdot (1 - x)$. We show that $f(x) \geq 0$ for all $x \in [0, 1]$. Note that,

$$f(0) = \frac{1}{\alpha} > 0, \text{ and } f(1) = 0. \tag{489}$$

On the other hand,

$$\begin{aligned} f'(x) &= -x^{\alpha-1} - \alpha \cdot x^{\alpha-1} \cdot (1 - x) + x^\alpha && (490)\\ &= -x^{\alpha-1} \cdot [1 + \alpha \cdot (1 - x) - x] && (491)\\ &= -x^{\alpha-1} \cdot (1 + \alpha) \cdot (1 - x) && (492)\\ &\leq 0, \qquad (\alpha > 0, \text{ and } x \in [0, 1]) && (493) \end{aligned}$$

which means f is monotonically decreasing over $[0, 1]$. Therefore $f(x) \geq 0$ for all $x \in [0, 1]$, finishing the proof. □

Lemma 14. *Given any $\alpha > 0$, we have, for all $x \in \left[\frac{2\alpha+1}{2\alpha+2}, 1\right]$,*

$$\frac{1}{2\alpha} \cdot (1 - x^\alpha) \leq x^\alpha \cdot (1 - x)\,. \tag{494}$$

Proof. Define $g : x \mapsto x^\alpha \cdot (1 - x) - \frac{1}{2\alpha} \cdot (1 - x^\alpha)$. The derivative of g is,

$$\begin{aligned} g'(x) &= \alpha \cdot x^{\alpha-1} \cdot (1 - x) - x^\alpha + (1/2) \cdot x^{\alpha-1} && (495)\\ &= x^{\alpha-1} \cdot [\alpha \cdot (1 - x) - x + 1/2] && (496)\\ &= x^{\alpha-1} \cdot [(1 + \alpha) \cdot (1 - x) - 1/2]\,. && (497) \end{aligned}$$

Then we have,

$$\begin{aligned} g'(x) > 0 \text{ for all } x \in [0, (2\alpha + 1)/(2\alpha + 2))\,, \text{ and} && (498)\\ g'(x) \leq 0 \text{ for all } x \in [(2\alpha + 1)/(2\alpha + 2), 1]\,, && (499) \end{aligned}$$

which means g is monotonically increasing over $[0, (2\alpha + 1)/(2\alpha + 2))$ and decreasing over $[(2\alpha + 1)/(2\alpha + 2), 1]$. On the other hand,

$$\begin{aligned} g((2\alpha + 1)/(2\alpha + 2)) &= \left(\frac{2\alpha+1}{2\alpha+2}\right)^\alpha \cdot \left(1 - \frac{2\alpha+1}{2\alpha+2}\right) - \frac{1}{2\alpha} \cdot \left[1 - \left(\frac{2\alpha+1}{2\alpha+2}\right)^\alpha\right] && (500)\\ &= \frac{1}{2\alpha} \cdot \left[\left(\frac{2\alpha+1}{2\alpha+2}\right)^\alpha \cdot \frac{2\alpha+1}{\alpha+1} - 1\right] && (501)\\ &= \frac{1}{2\alpha} \cdot \left[\exp\left\{\log\left(\frac{2\alpha+1}{\alpha+1}\right) - \alpha \cdot \log\left(1 + \frac{1}{2\alpha+1}\right)\right\} - 1\right] && (502)\\ &\geq \frac{1}{2\alpha} \cdot \left[\exp\left\{\log\left(\frac{2\alpha+1}{\alpha+1}\right) - \frac{\alpha}{2\alpha+1}\right\} - 1\right] \qquad (1 + x \leq e^x) && (503)\\ &\geq \frac{1}{2\alpha} \cdot \left[\exp\left\{\frac{\alpha}{2\alpha+1} - \frac{\alpha}{2\alpha+1}\right\} - 1\right] \qquad (\log(x) \geq 1 - 1/x \text{ for } x > 0) && (504)\\ &= 0. && (505) \end{aligned}$$

Also note that $g(1) = 0$. Therefore we have $g(x) \geq 0$ for all $x \in [(2\alpha + 1)/(2\alpha + 2), 1]$, finishing the proof. □

Lemma 15. *Denote $H(\pi) \coloneqq diag(\pi) - \pi\pi^\top$. Softmax policy gradient w.r.t. θ is*

$$\frac{\partial V^{\pi_\theta}(\mu)}{\partial \theta(s, \cdot)} = \frac{1}{1 - \gamma} \cdot d_\mu^{\pi_\theta}(s) \cdot H(\pi_\theta(\cdot|s)) Q^{\pi_\theta}(s, \cdot), \quad \forall s \in \mathcal{S}. \tag{506}$$

Proof. See the proof in (Mei et al., 2020b, Lemma 1). We include a proof for completeness.

According to the policy gradient theorem (Sutton et al., 2000),

$$\frac{\partial V^{\pi_\theta}(\mu)}{\partial \theta} = \frac{1}{1 - \gamma} \mathop{\mathbb{E}}_{s' \sim d_\mu^{\pi_\theta}} \left[\sum_a \frac{\partial \pi_\theta(a|s')}{\partial \theta} \cdot Q^{\pi_\theta}(s', a)\right]. \tag{507}$$

For $s' \neq s$, $\frac{\partial \pi_\theta(a|s')}{\partial \theta(s,\cdot)} = \mathbf{0}$ since $\pi_\theta(a|s')$ does not depend on $\theta(s,\cdot)$. Therefore, we have,

$$\frac{\partial V^{\pi_\theta}(\mu)}{\partial \theta(s,\cdot)} = \frac{1}{1-\gamma}\sum_{s'} d_\mu^{\pi_\theta}(s') \cdot \left[\sum_a \frac{\partial \pi_\theta(a|s')}{\partial \theta(s,\cdot)} \cdot Q^{\pi_\theta}(s',a)\right] \tag{508}$$

$$= \frac{1}{1-\gamma} \cdot d_\mu^{\pi_\theta}(s) \cdot \left[\sum_a \frac{\partial \pi_\theta(a|s)}{\partial \theta(s,\cdot)} \cdot Q^{\pi_\theta}(s,a)\right] \qquad \left(\frac{\partial \pi_\theta(a|s')}{\partial \theta(s,\cdot)} = \mathbf{0},\ \forall s' \neq s\right) \tag{509}$$

$$= \frac{1}{1-\gamma} \cdot d_\mu^{\pi_\theta}(s) \cdot \left(\frac{d\pi(\cdot|s)}{d\theta(s,\cdot)}\right)^\top Q^{\pi_\theta}(s,\cdot) \tag{510}$$

$$= \frac{1}{1-\gamma} \cdot d_\mu^{\pi_\theta}(s) \cdot H(\pi_\theta(\cdot|s)) Q^{\pi_\theta}(s,\cdot). \qquad (H(\pi_\theta) \text{ is the Jacobian of } \theta \mapsto \mathrm{softmax}(\theta)) \tag{511}$$

Note that in one-state MDPs, we have,

$$\frac{d\pi_\theta^\top r}{d\theta} = \left(\frac{d\pi_\theta}{d\theta}\right)^\top r = H(\pi_\theta) r. \qquad \square$$

Lemma 16. *Softmax policy gradient norm is*

$$\left\|\frac{\partial V^{\pi_\theta}(\mu)}{\partial \theta}\right\|_2 = \frac{1}{1-\gamma} \cdot \left[\sum_s d_\mu^{\pi_\theta}(s)^2 \cdot \|H(\pi_\theta(\cdot|s)) Q^{\pi_\theta}(s,\cdot)\|_2^2\right]^{\frac{1}{2}}. \tag{512}$$

Proof. We have,

$$\left\|\frac{\partial V^{\pi_\theta}(\mu)}{\partial \theta}\right\|_2 = \left[\sum_{s,a}\left(\frac{\partial V^{\pi_\theta}(\mu)}{\partial \theta(s,a)}\right)^2\right]^{\frac{1}{2}} \tag{513}$$

$$= \left[\sum_s \left\|\frac{\partial V^{\pi_\theta}(\mu)}{\partial \theta(s,\cdot)}\right\|_2^2\right]^{\frac{1}{2}} \tag{514}$$

$$= \frac{1}{1-\gamma} \cdot \left[\sum_s d_\mu^{\pi_\theta}(s)^2 \cdot \|H(\pi_\theta(\cdot|s)) Q^{\pi_\theta}(s,\cdot)\|_2^2\right]^{\frac{1}{2}}. \qquad \text{(by Lemma 15)} \qquad \square$$

Lemma 17 (Performance difference lemma (Kakade & Langford, 2002))**.** *For any policies π and π',*

$$V^{\pi'}(\rho) - V^{\pi}(\rho) = \frac{1}{1-\gamma}\sum_s d_\rho^{\pi'}(s) \sum_a (\pi'(a|s) - \pi(a|s)) \cdot Q^{\pi}(s,a) \tag{515}$$

$$= \frac{1}{1-\gamma}\sum_s d_\rho^{\pi'}(s) \sum_a \pi'(a|s) \cdot A^{\pi}(s,a). \tag{516}$$

Proof. According to the definition of value function,

$$V^{\pi'}(s) - V^{\pi}(s) = \sum_a \pi'(a|s) \cdot Q^{\pi'}(s,a) - \sum_a \pi(a|s) \cdot Q^{\pi}(s,a) \tag{517}$$

$$= \sum_a \pi'(a|s) \cdot \left(Q^{\pi'}(s,a) - Q^{\pi}(s,a)\right) + \sum_a (\pi'(a|s) - \pi(a|s)) \cdot Q^{\pi}(s,a) \tag{518}$$

$$= \sum_a (\pi'(a|s) - \pi(a|s)) \cdot Q^{\pi}(s,a) + \gamma \sum_a \pi'(a|s) \sum_{s'} \mathcal{P}(s'|s,a) \cdot \left[V^{\pi'}(s') - V^{\pi}(s')\right] \tag{519}$$

$$= \frac{1}{1-\gamma}\sum_{s'} d_s^{\pi'}(s') \sum_{a'} (\pi'(a'|s') - \pi(a'|s')) \cdot Q^{\pi}(s',a') \tag{520}$$

$$= \frac{1}{1-\gamma}\sum_{s'} d_s^{\pi'}(s') \sum_{a'} \pi'(a'|s') \cdot (Q^{\pi}(s',a') - V^{\pi}(s')) \tag{521}$$

$$= \frac{1}{1-\gamma}\sum_{s'} d_s^{\pi'}(s') \sum_{a'} \pi'(a'|s') \cdot A^{\pi}(s',a'). \qquad \square$$

Lemma 18 (Value sub-optimality lemma)**.** *For any policy π,*

$$V^*(\rho) - V^\pi(\rho) = \frac{1}{1-\gamma} \sum_s d_\rho^\pi(s) \sum_a (\pi^*(a|s) - \pi(a|s)) \cdot Q^*(s,a). \tag{522}$$

Proof. See the proof in (Mei et al., 2020b, Lemma 21). We include a proof for completeness.

We denote $V^*(s) \coloneqq V^{\pi^*}(s)$ and $Q^*(s,a) \coloneqq Q^{\pi^*}(s,a)$ for conciseness. We have, for any policy π,

$$\begin{aligned}
V^*(s) - V^\pi(s) &= \sum_a \pi^*(a|s) \cdot Q^*(s,a) - \sum_a \pi(a|s) \cdot Q^\pi(s,a) && (523)\\
&= \sum_a (\pi^*(a|s) - \pi(a|s)) \cdot Q^*(s,a) + \sum_a \pi(a|s) \cdot (Q^*(s,a) - Q^\pi(s,a)) && (524)\\
&= \sum_a (\pi^*(a|s) - \pi(a|s)) \cdot Q^*(s,a) + \gamma \sum_a \pi(a|s) \sum_{s'} \mathcal{P}(s'|s,a) \cdot \left[V^{\pi^*}(s') - V^\pi(s') \right] && (525)\\
&= \frac{1}{1-\gamma} \sum_{s'} d_s^\pi(s') \sum_{a'} (\pi^*(a'|s') - \pi(a'|s')) \cdot Q^*(s',a'). && \square
\end{aligned}$$

E. Non-convex (Non-concave) Examples for NŁ Inequality

We list some non-convex (or non-concave in maximization problems) functions which satisfy NŁ inequalities here from literature. See corresponding references for details.

Expected reward, softmax parameterization. As shown in Lemma 1 and Mei et al. (2020b, Lemma 3),

$$\left\| \frac{d \pi_\theta^\top r}{d\theta} \right\|_2 \geq \pi_\theta(a^*) \cdot (\pi^* - \pi_\theta)^\top r. \tag{526}$$

Value function, softmax parameterization. As shown in Lemma 5 and Mei et al. (2020b, Lemma 8),

$$\left\| \frac{\partial V^{\pi_\theta}(\mu)}{\partial \theta} \right\|_2 \geq \frac{\min_s \pi_\theta(a^*(s)|s)}{\sqrt{S} \cdot \left\| d_\rho^{\pi^*} / d_\mu^{\pi_\theta} \right\|_\infty} \cdot \left[V^*(\rho) - V^{\pi_\theta}(\rho) \right]. \tag{527}$$

Entropy regularized expected reward, softmax parameterization. As shown in Mei et al. (2020b, Proposition 5),

$$\left\| \frac{d\{\pi_\theta^\top (r - \tau \log \pi_\theta)\}}{d\theta} \right\|_2 \geq \sqrt{2\tau} \cdot \min_a \pi_\theta(a) \cdot \left[\pi_\tau^{*\top} (r - \tau \log \pi_\tau^*) - \pi_\theta^\top (r - \tau \log \pi_\theta) \right]^{\frac{1}{2}}. \tag{528}$$

Entropy regularized value function, softmax parameterization. As shown in Mei et al. (2020b, Lemma 15),

$$\left\| \frac{\partial \tilde{V}^{\pi_\theta}(\mu)}{\partial \theta} \right\|_2 \geq \frac{\sqrt{2\tau}}{\sqrt{S}} \cdot \min_s \sqrt{\mu(s)} \cdot \min_{s,a} \pi_\theta(a|s) \cdot \left\| \frac{d_\rho^{\pi_\tau^*}}{d_\mu^{\pi_\theta}} \right\|_\infty^{-\frac{1}{2}} \cdot \left[\tilde{V}^{\pi_\tau^*}(\rho) - \tilde{V}^{\pi_\theta}(\rho) \right]^{\frac{1}{2}}. \tag{529}$$

Expected reward, escort parameterization. As shown in Mei et al. (2020a, Lemma 3),

$$\left\| \frac{d \pi_\theta^\top r}{d\theta} \right\|_2 \geq \frac{p}{\|\theta\|_p} \cdot \pi_\theta(a^*)^{1-1/p} \cdot (\pi^* - \pi_\theta)^\top r. \tag{530}$$

Value function, escort parameterization. As shown in Mei et al. (2020a, Lemma 7),

$$\left\| \frac{\partial V^{\pi_\theta}(\mu)}{\partial \theta} \right\|_2 \geq \frac{p}{\sqrt{S}} \cdot \left\| \frac{d_\rho^{\pi^*}}{d_\mu^{\pi_\theta}} \right\|_\infty^{-1} \cdot \frac{\min_s \pi_\theta(a^*(s)|s)^{1-1/p}}{\max_s \|\theta(s,\cdot)\|_p} \cdot \left[V^*(\rho) - V^{\pi_\theta}(\rho) \right]. \tag{531}$$

Entropy regularized value function, escort parameterization. As shown in Mei et al. (2020a, Lemma 12),

$$\left\| \frac{\partial \tilde{V}^{\pi_\theta}(\mu)}{\partial \theta} \right\|_2 \geq \frac{p \cdot \sqrt{2\tau}}{\sqrt{S}} \cdot \min_s \sqrt{\mu(s)} \cdot \frac{\min_{s,a} \pi_\theta(a|s)^{1-1/p}}{\max_s \|\theta(s,\cdot)\|_p} \cdot \left\| \frac{d_\rho^{\pi_\tau^*}}{d_\mu^{\pi_\theta}} \right\|_\infty^{-\frac{1}{2}} \cdot \left[\tilde{V}^{\pi_\tau^*}(\rho) - \tilde{V}^{\pi_\theta}(\rho) \right]^{\frac{1}{2}}. \tag{532}$$

Cross entropy, escort parameterization. As shown in Mei et al. (2020a, Lemma 17),

$$\left\| \frac{d\{D_{\mathrm{KL}}(y\|\pi_\theta)\}}{d\theta} \right\|_2 \geq \frac{p}{\|\theta\|_p} \cdot \min_a \pi_\theta(a)^{\frac{1}{2}-\frac{1}{p}} \cdot D_{\mathrm{KL}}(y\|\pi_\theta)^{\frac{1}{2}}. \tag{533}$$

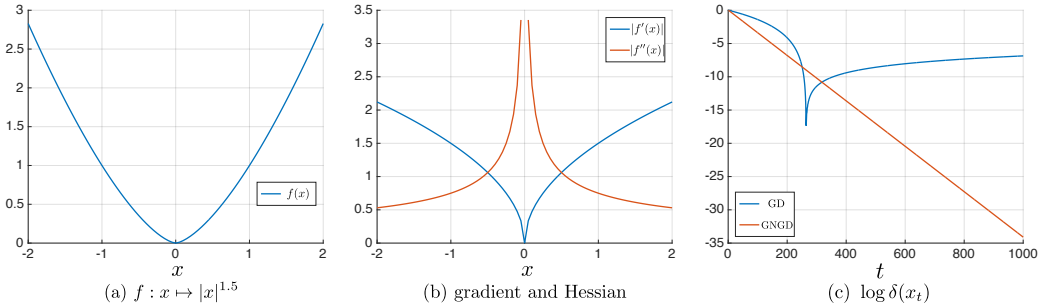
Generalized linear models, sigmoid activation, mean squared error. As shown in Lemma 9,

$$\left\| \frac{\partial \mathcal{L}(\theta)}{\partial \theta} \right\|_2 \geq 8 \cdot u(\theta) \cdot \min\{u(\theta), v\} \cdot \sqrt{\lambda_\phi} \cdot \left[\frac{1}{N} \cdot \sum_{i=1}^{N} (\pi_i - \pi_i^*)^2 \right]^{\frac{1}{2}}. \tag{534}$$

F. Additional Simulation Results

F.1. $f : x \mapsto |x|^p, p \in (1, 2)$

As shown in Proposition 4, with $p \in (1, 2)$, $f : x \mapsto |x|^p$ satisfies NŁ inequality with $\xi = 1/p \in (1/2, 1)$, which is the case (3) in Theorem 1. The function f is differentiable, and the Hessian $|f''(x)| = p \cdot (p-1) \cdot |x|^{p-2} \to \infty$, as $x \to 0$, which indicates GD with $\eta \in \Theta(1)$ does not converge.



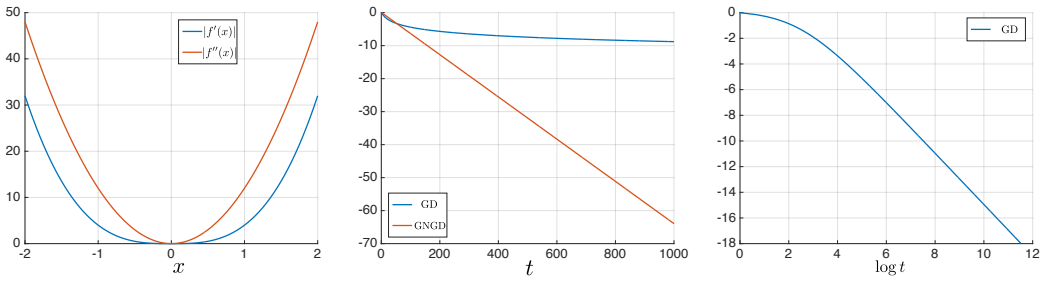
(a) $f : x \mapsto |x|^{1.5}$ (b) gradient and Hessian (c) $\log \delta(x_t)$

Figure 7. GD and GNGD on $f : x \mapsto |x|^p, p = 1.5$.

Figure 7(a) shows the image of $f : x \mapsto |x|^{1.5}$. As shown in subfigure (b), the gradient of f exists at $x = 0$, and the Hessian $|f''(x)| \to \infty$ as $x \to 0$. The results of GD with $\eta = 0.005$ and GNGD are presented in subfigure (c). The sub-optimality of GD update decreased for some time, and then it increased later. This is due to the Hessian is unbounded near $x = 0$, and thus constant learning rates cannot guarantee monotonic progresses for GD. On the other hand, GNGD with $\eta = 0.01$ enjoys $O(e^{-c \cdot t})$ convergence rate, verifying the results in the case (3) in Theorem 1.

F.2. $f : x \mapsto |x|^p, p > 2$

As shown in Proposition 4, with $p \in (1, 2)$, $f : x \mapsto |x|^p$ satisfies NŁ inequality with $\xi = 1/p \in (0, 1/2)$. As shown in Figure 8(a), the spectral radius of Hessian approaches 0 as $x \to 0$, which is the case (1) in Theorem 1.



(a) gradient and Hessian (b) $\log \delta(x_t)$ (c) $\log \delta(x_t)$, slope ≈ -1.9996

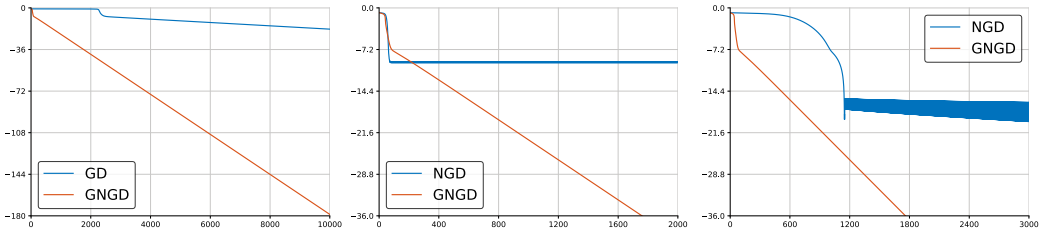
Figure 8. GD and GNGD on $f : x \mapsto |x|^p, p = 4$.

Subfigure (c) shows that the standard GD with constant learning rate $\eta = 0.01$ achieves sublinear rate about $O(1/t^2)$, while

subfigure (b) shows that GNGD with $\eta = 0.01$ enjoys linear rate $O(e^{-c\cdot t})$, verifying Theorem 1.

F.3. Convergence Rates on GLM

Theorem 5 proves linear convergence rates $O(e^{-c\cdot t})$ for both GD and GNGD on GLM. We compare GD, NGD (Hazan et al., 2015), and GNGD on GLM, as shown in Figure 9.



(a) $\log \delta(\theta_t)$ (b) $\log \delta(\theta_t)$ (c) $\log \delta(\theta_t)$

Figure 9. Convergence rates for GD, NGD, and GNGD on GLM.

Subfigure (a) presents the results of GD with $\eta = 0.09$ and GNGD with $\eta = 0.09$. Both GD and GNGD achieve linear $O(e^{-c\cdot t})$ rates, verifying Theorem 5. GD suffers from the plateaus at the early-stage optimization, which is consistent with Figure 1 and the explanations after Theorem 5. On the other hand, the slopes indicate that GNGD converges strictly faster than GD, which justifies the constant dependences ($C \geq C^2$) in Theorem 5. Subfigure (b) shows that standard NGD (Hazan et al., 2015) with constant learning rate $\eta = 0.09$ does not converge. The NGD update keeps oscillating, which verifies our argument of using standard normalization for all $t \geq 1$ is not a good idea. Subfigure (c) presents the NGD using adaptive learning rate $\eta_t = \frac{0.09}{\sqrt{t}}$, which has faster convergence than NGD with constant η. However, GNGD still significantly outperforms NGD with $\eta_t = \frac{0.09}{\sqrt{t}}$, verifying the $O(e^{-c\cdot t})$ in Theorem 5 and $O(1/\sqrt{t})$ in Theorem 4.

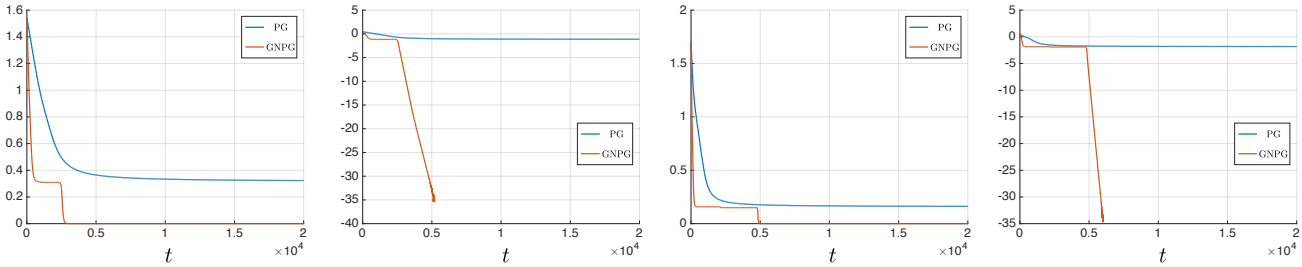
F.4. Tree MDPs

Figure 10 shows the results for PG and GNPG beyond one-state MDPs. The environment is a synthetic tree with height h and branching factor b. The total number of states is

$$S = \sum_{i=0}^{h-1} b^i. \tag{535}$$

The discount factor $\gamma = 0.99$, and we set $\mu = \rho$ (e.g., in Algorithm 1 and Theorem 3), where $\rho(s_0) = 1$ for the root state s_0. For PG, in each iteration, we calculate the policy gradient (Lemma 15) to do one update. For GNPG, Algorithm 1 is used.

Subfigures (a) and (b) show the results for $h = b = 4$, and $S = 85$. The learning rate is $\eta = 0.02$ for PG and GNPG. Subfigures (c) and (d) show the results for $h = 5$ and $b = 4$, and $S = 341$. The learning rate is $\eta = 0.05$ for PG and GNPG.



(a) $V^*(\rho) - V^{\pi_{\theta_t}}(\rho)$ (b) $\log\left(V^*(\rho) - V^{\pi_{\theta_t}}(\rho)\right)$ (c) $V^*(\rho) - V^{\pi_{\theta_t}}(\rho)$ (d) $\log\left(V^*(\rho) - V^{\pi_{\theta_t}}(\rho)\right)$

Figure 10. Results for PG and GNPG on tree MDPs. In (a) and (b), $S = 85$. In (c) and (d), $S = 341$.