Global Linear Convergence of Inexact TD Under Generalized Smoothness

author names withheld

Under Review for the Workshop on High-dimensional Learning Dynamics, 2026

Abstract

Recent work has analyzed temporal-difference (TD) learning with target networks through an optimization view and established linear convergence under a force-dominance condition, but these results typically rely on global smoothness, i.e., a uniform upper bound on curvature. This assumption can fail even when the inner problem is well posed, since curvature encountered during training can grow with the scale of TD-residual-induced gradients. We retain the stabilized regime in which the inner problem is strongly convex in the optimization variable, in order to isolate upper-curvature growth effects. Under generalized smoothness, where the Hessian norm may grow with gradient scale via a nondecreasing profile $\ell(\cdot)$, we analyze the inexact TD recursion with K inner gradient steps per target refresh and propose a curvature-checked constant stepsize rule that ensures global stability without requiring a global smoothness constant. Our main result proves global linear convergence under force dominance with a single trajectory-dependent admissibility requirement governed by the maximum gradient magnitude M encountered along the run. This yields an explicit scaling law: the largest admissible constant stepsize decays as $1/\ell(cM)$, for a universal constant c, and maintaining a fixed contraction requires K to grow proportionally to $\ell(cM)$. In the special case of uniformly bounded curvature, our result reduces to the classical global-smoothness regime; under curvature growth, the worst trajectory gradient scale controls both stability and attainable convergence speed, yielding a mechanism-level interpretation of why curvature-aware step control can matter in stabilized TD-style optimization.

1. Introduction

Temporal-difference learning is a central algorithmic primitive in reinforcement learning, and many modern value-based methods retain its bootstrapping structure [15, 18, 19]. At the same time, TD is known to exhibit divergence in simple examples with function approximation [2, 20]. A useful recent perspective recasts TD with target networks as a two-level optimization recursion [1, 6, 10]. At outer iteration t, a target parameter θ_t is frozen, an online parameter w is updated to reduce a target-induced objective $H(\theta_t, w)$, and the target is refreshed after a finite number of inner steps:

$$\theta_{t+1} \approx \arg\min_{w} H(\theta_t, w). \tag{1}$$

This formulation separates two effects: the optimization force in the inner variable w (curvature of the objective in the inner variable w), and the target force induced by the dependence of the next objective on θ_t (the sensitivity of the induced target mapping). The force-dominance condition requires the former to dominate the latter: the inner problem must contract strongly enough to overcome the drift caused by changing the target. Under this condition, Asadi et al. [1] obtain

fixed-point convergence for arbitrary finite K, the number of inner gradient steps per target refresh.

Algorithm 1 Inexact TD via K inner gradient steps

Input: $\theta_0, T, K, \{\gamma_{t,k}\}$
for $t = 0, \dots, T-1$ **do**
  $w_{t,0} \leftarrow \theta_t$ **for** $k = 0, \dots, K-1$ **do**
    $w_{t,k+1} \leftarrow w_{t,k} - \gamma_{t,k} \nabla_w H(\theta_t, w_{t,k})$
  end
  $\theta_{t+1} \leftarrow w_{t,K}$
end
return θ_T

The limitation addressed here is the separate global-smoothness requirement used to certify inner descent in the global-smoothness convergence theory of Asadi et al. [1]. There, $w \mapsto H(\theta, w)$ is assumed to have a globally Lipschitz gradient, uniformly over the trajectory. This can be a poor descriptor of high-dimensional learning dynamics. In deep learning, training trajectories can exhibit progressive sharpening and edge-of-stability behavior, where the largest Hessian eigenvalues evolve with the stepsize scale rather than remaining governed by a fixed global constant [4, 5, 8, 12, 13, 17]. TD adds a further nonstationarity: the frozen objective itself changes across target refreshes because bootstrapped targets depend on the current or lagged value parameters [1, 3, 15].

We therefore ask whether this force-dominance convergence mechanism survives once global upper curvature is replaced by trajectory-dependent curvature growth. We do not attempt to solve fully stochastic, nonconvex deep TD. Instead, as in the optimization-theoretic TD view, we retain a stabilized regime where, for each fixed target, the inner problem is strongly convex in the online parameter. This removes nonconvex optimization pathologies and isolates the upper-curvature effect. The frozen objective $f_t(w) := H(\theta_t, w)$ is assumed to satisfy the generalized-smoothness condition

$$\left\|\nabla^2 f_t(w)\right\|_{\mathrm{op}} \le \ell\big(\|\nabla f_t(w)\|\big), \tag{2}$$

where ℓ is positive and nondecreasing. This includes global smoothness as $\ell \equiv L$, but also permits curvature to grow with the local gradient scale [9, 11, 21, 23, 24].

Our contribution is a trajectory-dependent linear convergence theory for inexact TD under (2). The central device is a curvature-checked constant stepsize: the method proposes a baseline γ_0, but caps it by a local value computed from ℓ so that one-step descent holds without a global Hessian bound. We then prove that, on stable trajectories, the local cap is uniformly bounded below by a quantity determined by the maximum gradient magnitude M encountered during the run. Consequently, if $\gamma_0 \le [2\ell(3M/2)]^{-1}$, the curvature cap is inactive and the algorithm exhibits a uniform linear rate. The theorem recovers the classical global-smoothness regime as a special case. More importantly, it shows what changes under curvature growth: if the trajectory reaches large gradient regions, then the admissible constant stepsize becomes smaller, and more inner gradient steps may be needed to obtain the same contraction before each target refresh.

2. Problem setting

We consider the prediction setting for a discounted Markov reward process [16]. Given a value approximator $v(\cdot;\theta)$, TD with a frozen target can be written as gradient descent on

$$H(\theta, w) = \frac{1}{2} \sum_s d(s) \Big(\mathbb{E}_{r,s'}[r + \gamma v(s';\theta)] - v(s;w) \Big)^2, \tag{3}$$

where the term inside the square is the expected one-step TD residual under the frozen target θ. The analysis below applies to a general two-input objective H, with (3) as the motivating TD instance.

Algorithm 1 gives the inexact TD recursion. At outer iteration t, the target parameter θ_t is frozen, K gradient steps are taken on $w \mapsto H(\theta_t, w)$, and the target is refreshed by setting $\theta_{t+1} = w_{t,K}$. The vector $\nabla_w H(\theta, w)$ aggregates TD residuals weighted by critic Jacobians, and should therefore be read as a geometry-weighted TD residual scale.

We study Algorithm 1 for a general function $H : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ under the following assumptions. We define the frozen objective $f_t(w) = H(\theta_t, w)$.

Assumption 1 (Generalized smooth in w) *Fix t. The function $f_t : \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$ is ℓ-smooth on $\mathcal{W}$ if f_t is twice differentiable on $\mathcal{W}$, continuous on the closure of $\mathcal{W}$, and there exists a non-decreasing, positive, locally Lipschitz function $\ell : [0, \infty) \to (0, \infty)$ such that*

$$\left\|\nabla^2 f_t(w)\right\|_{\mathrm{op}} \le \ell\big(\|\nabla f_t(w)\|\big) \qquad \text{for all } w \in \mathcal{W}. \tag{4}$$

Assumption 2 (F_θ-Lipschitz partial gradient in θ) There exists $F_\theta \ge 0$ such that, for all θ_1, θ_2 and all w,

$$\left\|\nabla_w H(\theta_1, w) - \nabla_w H(\theta_2, w)\right\| \le F_\theta \left\|\theta_1 - \theta_2\right\|.$$

Assumption 3 (F_w-strong convexity in w) For every fixed θ, the function $w \mapsto H(\theta, w)$ is F_w-strongly convex. Equivalently, there exists $F_w > 0$ such that, for all w_1, w_2 and all θ,

$$\left\langle \nabla_w H(\theta, w_1) - \nabla_w H(\theta, w_2), w_1 - w_2 \right\rangle \ge F_w \left\|w_1 - w_2\right\|^2.$$

Assumption 4 (Fixed point) There exists $\theta^\star \in \mathbb{R}^d$ such that

$$\nabla_w H(\theta^\star, \theta^\star) = 0.$$

The force-dominance condition is defined as

$$F_\theta < F_w. \tag{5}$$

Thus the sensitivity of the inner gradient to target changes is smaller than the strong-convexity force of the frozen inner problem. This is the same stability lens as in the global-smoothness theory, but inner descent is certified below using trajectory-dependent curvature rather than a global Lipschitz-gradient constant.

3. Curvature-checked descent

Under generalized smoothness, the usual quadratic descent model with a global constant L is replaced by an integral modulus. Following Tyurin [21], for $a \ge 0$ define $q(s; a) = \int_0^s \frac{\mathrm{d}v}{\ell(a+v)}$. The inverse of $q(\cdot; a)$ controls gradient variation and yields the descent inequality

$$f_t(u) \le f_t(w) + \nabla f_t(w)u - w + \int_0^{\|u-w\|} q^{-1}\big(\tau; \|\nabla f_t(w)\|\big)\, \mathrm{d}\tau. \tag{6}$$

For an inner iterate, let $g_{t,k} = \|\nabla f_t(w_{t,k})\|$. If $g_{t,k} > 0$, define the local safe cap

$$\gamma_{t,k}^{\mathrm{safe}} := \frac{1}{g_{t,k}} q\left(\frac{g_{t,k}}{2}; g_{t,k}\right) = \int_0^{1/2} \frac{\mathrm{d}v}{\ell((1+v)g_{t,k})}, \tag{7}$$

and set $\gamma_{t,k} = \min\{\gamma_0, \gamma_{t,k}^{\mathrm{safe}}\}$. If $g_{t,k} = 0$, the update is stationary and we set $\gamma_{t,k} = \gamma_0$. The choice (7) is constructed so that the generalized-smoothness remainder in (6) consumes at most half of the linear decrease.

Lemma 1 (One-step descent) *For every inner update in Algorithm 1 using the stepsize rule (7),*

$$f_t(w_{t,k+1}) - f_t(w_{t,k}) \le -\frac{\gamma_{t,k}}{2} \|\nabla f_t(w_{t,k})\|^2 = -\frac{1}{2\gamma_{t,k}} \|w_{t,k+1} - w_{t,k}\|^2 . \tag{8}$$

This lemma is the point at which global smoothness is removed. It is purely local and remains valid even when the Hessian is unbounded, provided the curvature-growth profile controls the region visited by the inner update.

4. Linear convergence and scaling law

Let $\Delta_t := \|\nabla_w H(\theta_t, \theta^\star)\|$, $\kappa_{t,k} := F_w \gamma_{t,k}$. The quantity Δ_t measures the mismatch between the frozen inner objective at outer iteration t and the fixed point. The point $\theta^\star$ satisfies $\nabla_w H(\theta^\star, \theta^\star) = 0$, but the inner loop at iteration t optimizes $w \mapsto H(\theta_t, w)$. Thus, when $\theta_t \neq \theta^\star$, $\theta^\star$ need not minimize the current frozen objective, and $\nabla_w H(\theta_t, \theta^\star)$ appears as a forcing term.

Proposition 2 (Inner-step contraction with forcing) *For every inner update using the curvature-checked stepsize,*

$$\|w_{t,k+1} - \theta^\star\|^2 \le (1 - \kappa_{t,k}) \|w_{t,k} - \theta^\star\|^2 + \kappa_{t,k} \frac{\Delta_t^2}{F_w^2}. \tag{9}$$

Proposition 3 (Outer recursion with forcing) *Assume $0 < \gamma_{t,k} < 1/F_w$ for the nonstationary inner updates. Iterating (17) over K inner steps gives*

$$\|\theta_{t+1} - \theta^\star\|^2 \le \chi_{t,K} \|\theta_t - \theta^\star\|^2 + (1 - \chi_{t,K}) \frac{\Delta_t^2}{F_w^2}, \qquad \chi_{t,K} := \prod_{k=0}^{K-1} (1 - F_w \gamma_{t,k}). \tag{10}$$

Since $w \mapsto H(\theta_t, w)$ is F_w-strongly convex and satisfies the generalized-smoothness upper bound along the trajectory, $\ell(g_{t,k}) \ge F_w$. Hence, for $g_{t,k} > 0$, monotonicity of ℓ gives $\gamma_{t,k}^{\text{safe}} \le \int_0^{1/2} \frac{\mathrm{d}v}{F_w} = \frac{1}{2F_w}$. Thus the curvature-checked rule ensures $0 < F_w \gamma_{t,k} \le 1/2$ on nonstationary steps, and consequently $0 < 1 - F_w \gamma_{t,k} < 1$. Therefore $\chi_{t,K} \in (0, 1)$ whenever the inner loop makes nonzero progress.

The forcing term is controlled by the target-drift condition and the fixed-point condition: $\Delta_t = \|\nabla_w H(\theta_t, \theta^\star) - \nabla_w H(\theta^\star, \theta^\star)\| \le F_\theta \|\theta_t - \theta^\star\|$. Writing $\eta := F_\theta / F_w$, this gives $\frac{\Delta_t}{F_w} \le \eta \|\theta_t - \theta^\star\|$. Substituting this bound into (18) yields

$$\|\theta_{t+1} - \theta^\star\|^2 \le \Big(\eta^2 + (1 - \eta^2)\chi_{t,K}\Big) \|\theta_t - \theta^\star\|^2. \tag{11}$$

Under force dominance $F_\theta < F_w$, we have $\eta \in (0, 1)$, so the target-induced forcing radius is at most an η-fraction of the current fixed-point error. Since $\chi_{t,K} \in (0, 1)$, the multiplier in (11) is strictly smaller than one, and the outer TD error contracts.

To obtain a uniform constant-stepsize statement, let $R = \|\theta_0 - \theta^\star\|$ and define

$$M := \sup \{\|\nabla_w H(\theta, w)\| : \|\theta - \theta^\star\| \le R,\ \|w - \theta^\star\| \le R\}. \tag{12}$$

The contraction above gives bounded outer iterates, and (17) gives bounded inner iterates[1]. Thus, if $H \in C^1$, then $M < \infty$ and $g_{t,k} \le M$ along the run. By monotonicity of ℓ, $\gamma_{t,k}^{\text{safe}} = \int_0^{1/2} \frac{\mathrm{d}v}{\ell((1+v)g_{t,k})} \ge \frac{1}{2\ell(3M/2)}$. This gives the admissibility threshold used in the next theorem.

1. A short proof about this gradient bound is given in the Proof of Theorem 8 in the Appendix.

Theorem 4 (Constant-stepsize convergence under generalized smoothness) *Assume 1 2 3 4. Suppose $H \in C^1(\mathbb{R}^d \times \mathbb{R}^d)$, $F_\theta < F_w$, and let M be defined by (12). Define $\eta = F_\theta / F_w \in (0,1)$ and choose a constant baseline $\gamma_0 > 0$ satisfying*

$$\gamma_0 \le \frac{1}{2\ell(3M/2)}, \tag{13}$$

Run Algorithm 1 with stepsizes $\gamma_{t,k} = \min\{\gamma_0, \gamma_{t,k}^{\mathrm{safe}}\}$. Then the curvature check never clips, $\gamma_{t,k} \equiv \gamma_0$, and the outer iterates contract linearly:

$$\|\theta_{t+1} - \theta^\star\|^2 \;\le\; \rho\, \|\theta_t - \theta^\star\|^2, \qquad \rho \triangleq \eta^2 + (1-\eta^2)(1 - F_w\gamma_0)^K \in (0,1). \tag{14}$$

Consequently, $\|\theta_t - \theta^\star\|^2 \le \rho^t \|\theta_0 - \theta^\star\|^2$.

The theorem replaces the global condition number L/F_w [1] by the trajectory-dependent amplification

$$B(M) := \frac{\ell(3M/2)}{F_w}. \tag{15}$$

If γ_0 saturates (19), then $(1 - F_w\gamma_0)^K = \left(1 - \frac{1}{2B(M)}\right)^K \le \exp\left(-\frac{K}{2B(M)}\right)$. Thus a fixed-strength inner contraction requires $K \gtrsim B(M)$. For $\ell \equiv L$, this reduces to the classical smooth regime. For $\ell(s) = L_0 + L_1 s^p$, the admissible baseline decays as M^{-p} at large M, and the required inner effort grows as M^p/F_w. For exponential profiles, a globally safe constant step can become extremely conservative, which motivates the curvature check as a fail-safe during high-curvature transients.

5. Related work and conclusion

The closest point of comparison is the force-dominance analysis of TD as a two-level optimization procedure [1], together with related work on target networks and TD stabilization [6, 7, 14, 22, 25]. Our contribution is orthogonal to mechanisms such as slower target updates, projection, or regularization: we keep the stabilized inner-problem viewpoint and remove the global upper-curvature assumption used to certify descent. On the optimization side, the analysis connects to generalized smoothness and clipping under nonuniform curvature [9, 11, 21, 23, 24].

We proved that force-dominance convergence of inexact TD persists under generalized smoothness, provided the constant baseline is admissible relative to the trajectory gradient envelope. The main message is not that the envelope M is easy to know in advance; rather, it is the correct stability quantity once curvature can grow with gradient scale. Curvature-aware step control and fast linear contraction are compatible in the stabilized TD setting, but the attainable constant-step regime is governed by the worst curvature amplification encountered along the learning trajectory.

References

[1] Kavosh Asadi, Shoham Sabach, Yao Liu, Omer Gottesman, and Rasool Fakoor. Td convergence: An optimization perspective. *Advances in Neural Information Processing Systems*, 36: 49169–49186, 2023.

[2] Leemon Baird. Residual algorithms: Reinforcement learning with function approximation. In *Machine learning proceedings 1995*, pages 30–37. Elsevier, 1995.

[3] Emmanuel Bengio, Joelle Pineau, and Doina Precup. Interference and generalization in temporal difference learning. In *International Conference on Machine Learning*, pages 767–777. PMLR, 2020.

[4] Jeremy Cohen, Simran Kaur, Yuanzhi Li, J Zico Kolter, and Ameet Talwalkar. Gradient descent on neural networks typically occurs at the edge of stability. In *International Conference on Learning Representations*, 2021.

[5] Alex Damian, Eshaan Nichani, and Jason D. Lee. Self-stabilization: The implicit bias of gradient descent at the edge of stability. In *The Eleventh International Conference on Learning Representations*, 2023. URL https://openreview.net/forum?id=nhKHA59gXz.

[6] Mattie Fellows, Matthew J. A. Smith, and Shimon Whiteson. Why target networks stabilise temporal difference methods. In *Proceedings of the 40th International Conference on Machine Learning (ICML)*, pages 9886–9909. PMLR, 2023.

[7] Sina Ghiassian, Andrew Patterson, Shivam Garg, Dhawal Gupta, Adam White, and Martha White. Gradient temporal-difference learning with regularized corrections. In *Proceedings of the 37th International Conference on Machine Learning (ICML)*, pages 3524–3534. PMLR, 2020.

[8] Justin Gilmer, Behrooz Ghorbani, Ankush Garg, Sneha Kudugunta, Behnam Neyshabur, David Cardoze, George Edward Dahl, Zachary Nado, and Orhan Firat. A loss curvature perspective on training instabilities of deep learning models. In *International Conference on Learning Representations*, 2022.

[9] Eduard Gorbunov, Nazarii Tupitsa, Sayantan Choudhury, Alen Aliev, Peter Richtarik, Samuel Horváth, and Martin Takáč. Methods for convex (l_0, l_1)-smooth optimization: Clipping, acceleration, and adaptivity. In *International Conference on Learning Representations (ICLR)*, 2025.

[10] Donghwan Lee and Niao He. Target-based temporal-difference learning. In *International Conference on Machine Learning*, pages 3713–3722. PMLR, 2019.

[11] Haochuan Li, Jian Qian, Yi Tian, Alexander Rakhlin, and Ali Jadbabaie. Convex and nonconvex optimization under generalized smoothness. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2023.

[12] Philip M Long and Peter L Bartlett. Sharpness-aware minimization and the edge of stability. *Journal of Machine Learning Research*, 25(179):1–20, 2024.

[13] Clare Lyle, Zeyu Zheng, Evgenii Nikishin, Bernardo Avila Pires, Razvan Pascanu, and Will Dabney. Understanding plasticity in neural networks. In *International Conference on Machine Learning*, pages 23190–23211. PMLR, 2023.

[14] Gaurav Manek and J. Zico Kolter. The pitfalls of regularization in off-policy td learning. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2022.

[15] Volodymyr Mnih, Koray Kavukcuoglu, David Silver, Andrei A Rusu, Joel Veness, Marc G Bellemare, Alex Graves, Martin Riedmiller, Andreas K Fidjeland, Georg Ostrovski, et al. Human-level control through deep reinforcement learning. *nature*, 518(7540):529–533, 2015.

[16] Martin L Puterman. *Markov decision processes: discrete stochastic dynamic programming*. John Wiley & Sons, 2014.

[17] Vincent Roulet, Atish Agarwala, Jean-Bastien Grill, Grzegorz Swirszcz, Mathieu Blondel, and Fabian Pedregosa. Stepping on the edge: Curvature aware learning rate tuners. *Advances in Neural Information Processing Systems*, 37:47708–47740, 2024.

[18] Richard S Sutton. Learning to predict by the methods of temporal differences. *Machine learning*, 3(1):9–44, 1988.

[19] Richard S Sutton, Andrew G Barto, et al. *Reinforcement learning: An introduction*, volume 1. MIT press Cambridge, 1998.

[20] John Tsitsiklis and Benjamin Van Roy. Analysis of temporal-diffference learning with function approximation. *Advances in neural information processing systems*, 9, 1996.

[21] Alexander Tyurin. Toward a unified theory of gradient descent under generalized smoothness. *arXiv preprint arXiv:2412.11773*, 2024.

[22] Simon Weissmann, Tilman Aach, Benedikt Wille, Sebastian Kassing, and Leif Döring. The role of target update frequencies in q-learning. arXiv preprint, 2026. arXiv:2602.03911.

[23] Bohang Zhang, Jikai Jin, Cong Fang, and Liwei Wang. Improved analysis of clipping algorithms for non-convex optimization. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2020. arXiv:2010.02519.

[24] Jingzhao Zhang, Tianxing He, Suvrit Sra, and Ali Jadbabaie. Why gradient clipping accelerates training: A theoretical justification for adaptivity. *arXiv preprint arXiv:1905.11881*, 2019.

[25] Shangtong Zhang, Hengshuai Yao, and Shimon Whiteson. Breaking the deadly triad with a target network. In *Proceedings of the 38th International Conference on Machine Learning (ICML)*, pages 12621–12631. PMLR, 2021.

Appendix A. Proof details

Lemma 5 (One-step descent) *For every inner update in Algorithm 1 using the stepsize rule (7),*

$$f_t(w_{t,k+1}) - f_t(w_{t,k}) \le -\frac{\gamma_{t,k}}{2} \|\nabla f_t(w_{t,k})\|^2 = -\frac{1}{2\gamma_{t,k}} \|w_{t,k+1} - w_{t,k}\|^2 . \tag{16}$$

Proof If $g_{t,k} = 0$, then $w_{t,k+1} = w_{t,k}$ and the claim is immediate. Assume $g_{t,k} > 0$, and write $x = w_{t,k}$, $x^+ = w_{t,k+1} = x - \gamma_{t,k}\nabla f_t(x)$, and $g = \|\nabla f_t(x)\|$. By (6),

$$f_t(x^+) - f_t(x) \le -\gamma_{t,k} g^2 + \int_0^{\gamma_{t,k} g} q^{-1}(\tau; g)\, \mathrm{d}\tau.$$

Since $q^{-1}(\cdot; g)$ is increasing,

$$\int_0^{\gamma_{t,k} g} q^{-1}(\tau; g)\, \mathrm{d}\tau \le \gamma_{t,k} g\, q^{-1}(\gamma_{t,k} g; g).$$

The definition of $\gamma_{t,k}^{\mathrm{safe}}$ and the choice $\gamma_{t,k} \le \gamma_{t,k}^{\mathrm{safe}}$ imply

$$\gamma_{t,k} g \le q(g/2; g), \qquad q^{-1}(\gamma_{t,k} g; g) \le g/2.$$

Therefore

$$f_t(x^+) - f_t(x) \le -\gamma_{t,k} g^2 + \frac{\gamma_{t,k}}{2} g^2 = -\frac{\gamma_{t,k}}{2} g^2.$$

Since $x^+ - x = -\gamma_{t,k}\nabla f_t(x)$, this is equivalent to

$$f_t(w_{t,k+1}) - f_t(w_{t,k}) \le -\frac{1}{2\gamma_{t,k}} \|w_{t,k+1} - w_{t,k}\|^2 .$$

■

Proposition 6 (Inner-step contraction with forcing) *For every inner update using the curvature-checked stepsize,*

$$\|w_{t,k+1} - \theta^\star\|^2 \le (1 - \kappa_{t,k}) \|w_{t,k} - \theta^\star\|^2 + \kappa_{t,k} \frac{\Delta_t^2}{F_w^2}. \tag{17}$$

Proof Fix t, write $f_t(w) = H(\theta_t, w)$, and set $w = w_{t,k}$, $w^+ = w_{t,k+1}$, and $\gamma = \gamma_{t,k}$. Strong convexity of f_t gives

$$f_t(y) \ge f_t(\theta^\star) + \nabla f_t(\theta^\star) y - \theta^\star + \frac{F_w}{2} \|y - \theta^\star\|^2 .$$

Taking $y = w^+$ and applying Young's inequality,

$$f_t(\theta^\star) - f_t(w^+) \le \nabla f_t(\theta^\star)\theta^\star - w^+ - \frac{F_w}{2} \left\|w^+ - \theta^\star\right\|^2 \le \frac{\Delta_t^2}{2F_w}.$$

Combining this with (16) gives

$$f_t(\theta^\star) - f_t(w) \le \frac{\Delta_t^2}{2F_w} - \frac{1}{2\gamma} \left\|w^+ - w\right\|^2 .$$

Now expand the update:

$$\left\|w^+ - \theta^\star\right\|^2 = \|w - \theta^\star\|^2 + 2\gamma \nabla f_t(w)\theta^\star - w + \left\|w^+ - w\right\|^2.$$

By strong convexity,

$$\nabla f_t(w)\theta^\star - w \le f_t(\theta^\star) - f_t(w) - \frac{F_w}{2}\|w - \theta^\star\|^2.$$

Substituting the last two bounds cancels the step-length term and yields

$$\left\|w^+ - \theta^\star\right\|^2 \le (1 - F_w\gamma)\|w - \theta^\star\|^2 + \gamma\frac{\Delta_t^2}{F_w}.$$

Since $\kappa_{t,k} = F_w\gamma_{t,k}$, this is exactly (17). ■

Proposition 7 (Outer recursion with forcing) *Assume $0 < \gamma_{t,k} < 1/F_w$ for the nonstationary inner updates. Iterating (17) over K inner steps gives*

$$\|\theta_{t+1} - \theta^\star\|^2 \le \chi_{t,K}\|\theta_t - \theta^\star\|^2 + (1 - \chi_{t,K})\frac{\Delta_t^2}{F_w^2}, \qquad \chi_{t,K} := \prod_{k=0}^{K-1}(1 - F_w\gamma_{t,k}). \tag{18}$$

Proof For fixed t, define

$$x_k := \|w_{t,k} - \theta^\star\|^2, \qquad a_t := \frac{\Delta_t^2}{F_w^2}.$$

Proposition 2 gives

$$x_{k+1} \le (1 - \kappa_{t,k})x_k + \kappa_{t,k}a_t,$$

or equivalently

$$x_{k+1} - a_t \le (1 - \kappa_{t,k})(x_k - a_t).$$

Since $0 < \gamma_{t,k} < 1/F_w$, each factor $1 - \kappa_{t,k}$ is positive. Iterating over $k = 0, \dots, K-1$ gives

$$x_K - a_t \le \left(\prod_{k=0}^{K-1}(1 - \kappa_{t,k})\right)(x_0 - a_t).$$

Using $w_{t,0} = \theta_t$, $w_{t,K} = \theta_{t+1}$, and $\kappa_{t,k} = F_w\gamma_{t,k}$, we obtain (18). ■

Theorem 8 (Constant-stepsize convergence under generalized smoothness) *Assume 1 2 3 4. Suppose $H \in C^1(\mathbb{R}^d \times \mathbb{R}^d)$, $F_\theta < F_w$, and let M be defined by (12). Define $\eta = F_\theta/F_w \in (0,1)$ and choose a constant baseline $\gamma_0 > 0$ satisfying*

$$\gamma_0 \le \frac{1}{2\ell(3M/2)}, \tag{19}$$

Run Algorithm 1 with stepsizes $\gamma_{t,k} = \min\{\gamma_0, \gamma_{t,k}^{\text{safe}}\}$. Then the curvature check never clips, $\gamma_{t,k} \equiv \gamma_0$, and the outer iterates contract linearly:

$$\|\theta_{t+1} - \theta^\star\|^2 \;\le\; \rho\|\theta_t - \theta^\star\|^2, \qquad \rho \triangleq \eta^2 + (1 - \eta^2)(1 - F_w\gamma_0)^K \in (0,1). \tag{20}$$

Consequently, $\|\theta_t - \theta^\star\|^2 \le \rho^t\|\theta_0 - \theta^\star\|^2$.

Proof By the target-drift assumption and the fixed-point condition,

$$\Delta_t = \|\nabla_w H(\theta_t, \theta^\star) - \nabla_w H(\theta^\star, \theta^\star)\| \le F_\theta \|\theta_t - \theta^\star\| .$$

Let $\eta = F_\theta / F_w$. Substituting this into Proposition 3 gives

$$\|\theta_{t+1} - \theta^\star\|^2 \le \left(\eta^2 + (1-\eta^2)\chi_{t,K}\right) \|\theta_t - \theta^\star\|^2 .$$

Since $F_\theta < F_w$, $\eta \in (0,1)$. Also, the curvature-checked rule gives $0 < F_w \gamma_{t,k} \le 1/2$ on non-stationary steps, and hence $\chi_{t,K} \in (0,1)$ whenever the inner loop makes progress. Thus the outer iterates are bounded:

$$\|\theta_t - \theta^\star\| \le \|\theta_0 - \theta^\star\| = R.$$

The inner recursion (17) then implies bounded inner iterates. Indeed, since

$$\frac{\Delta_t^2}{F_w^2} \le \eta^2 \|\theta_t - \theta^\star\|^2 \le R^2$$

and $w_{t,0} = \theta_t$, induction in k gives $\|w_{t,k} - \theta^\star\| \le R$. Therefore all pairs $(\theta_t, w_{t,k})$ lie in the compact set defining M. Since $H \in C^1(\mathbb{R}^d \times \mathbb{R}^d)$, the map $(\theta, w) \mapsto \|\nabla_w H(\theta, w)\|$ is continuous, so $M < \infty$ and $g_{t,k} \le M$ along the run.

Using (7) and monotonicity of ℓ,

$$\gamma_{t,k}^{\text{safe}} = \int_0^{1/2} \frac{\mathrm{d}v}{\ell((1+v)g_{t,k})} \ge \int_0^{1/2} \frac{\mathrm{d}v}{\ell(3M/2)} = \frac{1}{2\ell(3M/2)}.$$

If $\gamma_0 \le [2\ell(3M/2)]^{-1}$, then $\gamma_0 \le \gamma_{t,k}^{\text{safe}}$ for all t, k, so the curvature check is inactive and $\gamma_{t,k} = \gamma_0$. Hence

$$\chi_{t,K} = (1 - F_w\gamma_0)^K.$$

Substituting into the outer recursion gives

$$\|\theta_{t+1} - \theta^\star\|^2 \le \rho \|\theta_t - \theta^\star\|^2 , \qquad \rho = \eta^2 + (1-\eta^2)(1 - F_w\gamma_0)^K.$$

Because $\eta \in (0,1)$ and $0 < F_w\gamma_0 \le 1/2$, we have $\rho < 1$. Iterating gives

$$\|\theta_t - \theta^\star\|^2 \le \rho^t \|\theta_0 - \theta^\star\|^2 .$$

■