Information and Inference: A Journal of the IMA (2022) Page 1 of 70
doi:10.1093/imaiai/drn000

Exit Time Analysis for Approximations of Gradient Descent Trajectories Around Saddle Points

RISHABH DIXIT
Department of Electrical and Computer Engineering
Rutgers University–New Brunswick, NJ 08854 USA
Corresponding author: rishabh.dixit@rutgers.edu

MERT GÜRBÜZBALABAN
Department of Management Science and Information Systems
Department of Electrical and Computer Engineering
Department of Statistics
Rutgers University–New Brunswick, NJ 08854 USA
mg1366@rutgers.edu

WAHEED U. BAJWA
Department of Electrical and Computer Engineering
Department of Statistics
Rutgers University–New Brunswick, NJ 08854 USA
waheed.bajwa@rutgers.edu

[Received on 10 September 2022]

This paper considers the problem of understanding the exit time for trajectories of gradient-related first-order methods from saddle neighborhoods under some initial boundary conditions. Given the 'flat' geometry around saddle points, first-order methods can struggle to escape these regions in a fast manner due to the small magnitudes of gradients encountered. In particular, while it is known that gradient-related first-order methods escape strict-saddle neighborhoods, existing analytic techniques do not explicitly leverage the local geometry around saddle points in order to control behavior of gradient trajectories. It is in this context that this paper puts forth a rigorous geometric analysis of the gradient-descent method around strict-saddle neighborhoods using matrix perturbation theory. In doing so, it provides a key result that can be used to generate an approximate gradient trajectory for any given initial conditions. In addition, the analysis leads to a linear exit-time solution for gradient-descent method under certain necessary initial conditions, which explicitly bring out the dependence on problem dimension, conditioning of the saddle neighborhood, and more, for a class of strict-saddle functions.

Keywords: Exit-time analysis; Gradient descent; Non-convex optimization; Strict-saddle property.

2010 Math Subject Classification: 90C26 ; 15Axx ; 41A58 ; 65Hxx

1. Introduction

The problem of finding the convergence rate/time of gradient-related methods to a stationary point of a convex function has been studied extensively. Moreover, it has been well established that stronger conditions on function geometry yield better convergence guarantees for the class of gradient-related first-order methods. For instance, conditions like strong convexity and quadratic growth result in the

so-called 'linear convergence rate' to a stationary point for gradient-related first-order methods. Though there is also a class of second-order (Hessian-related) methods like the Newton method that yield super-linear convergence to stationary points of strongly convex functions, that comes at the cost of very high iteration complexity.

More recently much of the focus has shifted towards obtaining rates of convergence for gradient-related methods to stationary points of non-convex functions. To this end, there are some local geometric conditions like the Kurdyka-Łojasiewicz property [21, 26] that guarantee linear convergence rates provided the iterate is in some bounded neighborhood of the function's second-order stationary point [25]. Such guarantees, however, are hard to obtain for non-convex functions in a global sense and the linear convergence rates are often eventual, i.e., these methods usually exhibit such linear convergence only asymptotically. The main reason that restricts this speedup behavior to the asymptotic setting is the non-convex geometry that can impede fast traversal of these methods across the geometric landscape of the function. This is due to the fact that trajectories of gradient-related methods can encounter extremely flat curvature regions very near to first-order saddle points. Such regions are characterized by gradients that have very small magnitudes and it can take exponential time for the trajectory of an algorithm to traverse this extremely flat region. A natural question to ask then is whether there exist gradient-related first-order methods for which a subset of non-zero measure trajectories escape first-order saddle points of a class of non-convex functions in 'linear' time.[1] The non-zero measure of such fast escaping trajectories is important since studying fast escape is only useful when the initialization set is dense in such trajectories. Section 3.2 (see Remark 3.1) in particular establishes that indeed fast saddle escape is possible from an initialization set of positive measure.

We address this question in this work by deriving an upper bound on the exit time for a certain class of gradient-descent trajectories escaping some bounded neighborhood of the first-order saddle point of a class of smooth, non-convex functions. Specifically, let $\mathbf{x}^*$ be a saddle point of a smooth, non-convex function $f : \mathbb{R}^n \to \mathbb{R}$ and, without loss of generality, define the bounded neighborhood around the saddle point to be an open ball of radius ε around $\mathbf{x}^*$, denoted by $\mathscr{B}_\varepsilon(\mathbf{x}^*)$. Recall that the gradient at saddle point $\mathbf{x}^*$ is a zero vector, i.e., it is necessarily a first-order stationary point. In addition, the saddle neighborhood $\mathscr{B}_\varepsilon(\mathbf{x}^*)$ exhibits certain properties that depend on Lipschitz boundedness of the function and its derivatives as well as eigenvalues of the Hessian at $\mathbf{x}^*$. The class of trajectories we focus on in here is assumed to have the current iterate sitting on the boundary of $\mathscr{B}_\varepsilon(\mathbf{x}^*)$ and it comprises of all those trajectories of gradient descent that escape this saddle neighborhood with *at least* linear rate. Note that the current iterate could have reached the boundary of $\mathscr{B}_\varepsilon(\mathbf{x}^*)$ using *any* gradient-related method, but that problem is not our concern. Rather, our focus here is whether there exists any gradient-descent trajectory from the current iterate that can escape $\mathscr{B}_\varepsilon(\mathbf{x}^*)$ in almost linear time of order $\mathscr{O}\left(\log(\varepsilon^{-1})\right)$ or better. And if such a trajectory exists, then an immediate subsequent question asks for the necessary conditions required for the existence of such gradient-descent trajectories. To answer both these questions effectively, we present a rigorous analysis of gradient-descent trajectories $\{\mathbf{x}_k\}$ starting at time $k = 0$, when the initial iterate $\mathbf{x}_0$ sits on the boundary of the ball $\mathscr{B}_\varepsilon(\mathbf{x}^*)$, till the time they exit $\mathscr{B}_\varepsilon(\mathbf{x}^*)$, which we term the *exit time* and denote by K_{exit}. It should be noted that we analyze in this work the first-order approximations of the exact trajectories, instead of the exact trajectories themselves, where the approximation error is sufficiently small. Specifically, the presence of higher-order terms ($\mathscr{O}(\varepsilon^2)$ terms) in the forthcoming analysis accounts for the approximation in our

[1] We are slightly abusing terminology here and, in keeping with the convention of linear convergence rates in optimization literature, we are defining 'linear exit time' for the trajectory of a discrete method to be one in which the trajectory escapes an $\mathscr{O}(\varepsilon)$ saddle neighborhood in $\mathscr{O}(\log(\varepsilon^{-1}))$ number of iterations.

analysis, while things are proved about trajectories and perturbations up to the first order in ε.

We conclude by noting the relevance of the exit-time results derived in this paper to the broader field of non-convex optimization. First, to the best of our knowledge, there are no works other than the ones listed in Table 1 that *explicitly* investigate the exit times from saddle neighborhoods of the trajectories of discrete first-order methods. Rather, the focus in much of the related works discussed in Section 1.1 is on providing rates of convergence to second-order stationary points. While such analysis necessarily implies saddle escape, this is typically accomplished through the use of noisy perturbations that allow the trajectories to move along a negative curvature direction; in particular, such approaches do not yield an explicit expression for the exit time of a trajectory from a saddle neighborhood. Second, and most importantly, the rate of convergence to a second-order stationary point is trivially a function of the time a trajectory spends near a saddle point. It therefore stands to reason that the existing convergence rates for some of the recent first-order methods can possibly be improved by identifying trajectories with linear exit time, which is the focus of this paper.

1.1 *Relation to prior work*

Convergence rates of optimization methods to the minima of convex functions have been studied for quite some time. For instance, the seminal work dealing with convergence rate analysis of gradient-related methods has been well summarized in [36], while a recent work by [33] summarizes convergence rates of Newton-type methods. These prior works rely heavily on the Lipschitz boundedness of the function along with some other form of curvature property. The works [2] and [3] utilize the local Kurdyka–Łojasiewicz property [21, 26] of a function to develop convergence guarantees and the ergodic rates using monotonicity of gradient sequences in a bounded neighborhood of the function's stationary point. However, for non-convex functions these seminal works do not analyze the exit time from a bounded neighborhood of a first-order saddle point. With the focus shifting towards characterizing the efficacy of gradient-related methods on non-convex geometries in recent years, it becomes imperative to conduct such an analysis. To the best of our knowledge, currently no work exists that analyzes (discrete) gradient-descent trajectories in the saddle neighborhood using eigenvector perturbations. Therefore, this is the first work that incorporates matrix perturbation theory to extract the local geometric information around a saddle point necessary for analyzing gradient trajectories at such small scales. As a result of the perturbation analysis, the hidden dependence of exit time on the trajectory's initialization point, conditioning of the saddle neighborhood, problem dimension, and more, is also revealed in this work (cf. Table 2 in Section 3.5).

There is a plethora of existing methods in the literature that deal with non-convex optimization problems. Within the context of this paper, we broadly classify these methods into *continuous-time* Ordinary Differential Equations (ODE)-type methods/analysis and *discrete-time* gradient-descent related algorithms/analysis. The latter class of methods can be further categorized into first-order and higher-order methods. Starting with the continuous-time ODE-type algorithms, we first refer to [3] that has developed upon the gradient flow curve analysis of non-smooth convex functions. Although this work focuses on convex problems, yet it is important in the sense that it motivates us in drawing some parallels between the discrete gradient trajectories and the continuous flow curves in our analysis of non-convex functions.

Another recent work [15] within the continuous-time setting analyzes the saddle escape problem using a stochastic ODE to characterize the rates of escape in terms of a multiplicative noise factor. Remarkably, the results in [15] give a linear rate of escape in expectation for very small stochastic noise. This work also extends these results to cascaded saddle geometries. Note that the analysis in [15] relies on an earlier important work by [20], which characterizes the probability distribution of the exit

time of gradient curves from saddle point vicinities. The hyberbolic flow curves discussed in [3, 15, 20] are the building blocks of our intuition towards analyzing discrete gradient trajectories in this work.

A Stochastic Differential Equation (SDE) approach has also been utilized in a recent work [39] to study gradient-based (stochastic) methods for non-convex functions in the continuous-time setting. While this work also guarantees linear rates of global convergence for non-convex problems under certain assumptions, a few of which are more restrictive than our work, it does not lend itself to understanding the behavior of discrete gradient trajectories around first-order saddle points. Similarly the analysis done in [32] shows that fast evasion is possible for trajectories generated by normalized gradient flow from strict saddle neighborhoods of Morse functions but such an analysis is not sufficient to explain the behavior of discrete trajectories around saddle points.

Next, there exists a large collection of work analyzing discrete gradient-related methods in non-convex settings. The very basic yet most often investigated approach in these works is the Stochastic Gradient Descent (SGD) method and its variants. Such methods have been extensively studied in the literature for the purpose of escaping saddles, specifically first-order saddle points. For instance, [12, 17] provide the rates of convergence to a second order stationary point with very high probability using perturbed gradient descent, where the perturbation vector is an isotropic noise. In contrast, the work in [12] shows that—in the worst case—the time to escape cascaded saddles scales exponentially with the problem dimension, thereby making the method impractical for highly pathological problems like optimization over jagged functions.

The work in [23] provides new insights into the efficacy of gradient-descent method around strict saddle points. The authors in this work present a measure-theoretic analysis of the gradient-descent trajectories escaping strict saddle points almost surely. Their analysis uses the stable center manifold theorem in [19] to prove that random initializations of gradient-descent trajectories in the vicinity of a strict saddle point almost never terminate into this saddle point. Note that while this is an intuitive inference, it is somewhat hard to prove for gradient flow curves around saddle points. The work [9] also provides rates and escape guarantees under certain strong assumptions of high correlation between the negative curvature direction and a random perturbation vector. Interestingly, the convergence rate put forth in this work does not depend on the problem dimension. However due to the nature of the somewhat restrictive assumptions in [9], the resulting method is not suited to work over a general class of non-convex problems. We also note two related recent works [13, 37] that analyze global convergence behavior of Langevin dynamics-based variants of the SGD (and simulated annealing) for non-convex functions. Neither of these works, however, focuses on the escape behavior of trajectories around saddle neighborhoods.

There also is a sub-category of first-order methods leveraging acceleration and momentum techniques to escape saddle points. For instance, [34] uses the stable center manifold theorem to show that the heavy-ball method almost surely escapes a strict saddle neighborhood. But the rate of escape derived in this work is limited to quadratic functions; further, the ensuing analysis does not bring out the dependence on problem dimension, conditioning of the saddle neighborhood, etc. The work in [38] provides extensions of SGD methods like the Stochastic Variance Reduced Gradient (SVRG) algorithm for escaping saddles. Recently, in works like [18] and [41], methods approximating the second-order information of the function (i.e., Hessian) have been employed to escape the saddles and at the same time preserve the first-order nature of the algorithm. Specifically, [18] shows that the acceleration step in gradient descent guarantees escape from saddle points with provably better rates; yet the rate is still worse than the linear rate. Along similar lines, the method in [41] utilizes the second-order nature of the acceleration step combined with a stochastic perturbation to guarantee escape and provide escape rates.

Finally, higher-order methods are discussed in [30, 35], which utilize the Hessian of the function or

its combinations with first-order algorithms to escape saddle neighborhoods with an impressive super linear rate while trading-off heavily with per-iteration complexity. Going even a step further, the work in [1] poses the problem with second-order saddles, thereby making higher-order methods an absolute necessity. Though these techniques optimize well over certain pathological functions like degenerate saddles or very ill-conditioned geometries, yet they suffer heavily in terms of complexity. In addition, none of these methods leverage the initial boundary condition of their methods around saddle points, which could not only influence the future trajectory but also control its exit time from some bounded neighborhood of the saddle point. This further motivates us to conduct a rigorous analysis of (approximations of) gradient-descent trajectories around saddle points for some fixed initial boundary conditions.

We conclude by noting that the use of careful initial boundary conditions in order to avoid saddle points in non-convex optimization is not a fundamentally new idea. Consider, for instance, the non-convex formulation of the phase retrieval problem in [5]. A variant of the gradient descent method, termed the Wirtinger flow algorithm, can be used to solve this problem as long as the algorithm is carefully initialized along the direction of the negative curvature by means of a spectral method [5]. However, one of the implications of the results in this paper are that spectral initializations such as the one in [5], which require costly computation of the dominant eigenvector of a matrix, are not always required for saddle escape. Rather, one might be able to escape the saddle neighborhoods in approximately linear time provided the projection of the initial gradient descent iterate along the negative curvature direction is lower bounded by a small quantity.

1.2 *Our contributions*

Having discussed the relevant works pertaining to the problem of characterizing the exit time of first-order methods from saddle neighborhoods, we now elaborate upon the contributions of our work.

First, none of the earlier discussed works exploit the dependence of the function gradient in saddle neighborhood on the eigenvectors of the Hessian at the saddle point. This dependence results from the eigenvector perturbations of the Hessian in the saddle neighborhood. Therefore, to our knowledge, this is the first work that utilizes the Rayleigh–Schrödinger perturbation theory to approximate the Hessian $\nabla^2 f(\mathbf{x})$ at any point $\mathbf{x} \in \mathscr{B}_\varepsilon(\mathbf{x}^*)$. This approximate Hessian is then used to obtain the function gradient $\nabla f(\mathbf{x})$ for any point $\mathbf{x} \in \mathscr{B}_\varepsilon(\mathbf{x}^*)$.

Second, using the value of the function gradient, for any given initialization $\mathbf{x}_0$ and some fixed step size, we generate an approximate trajectory for the gradient-descent method inside the ball $\mathscr{B}_\varepsilon(\mathbf{x}^*)$. As a consequence, we obtain the distance between the saddle point $\mathbf{x}^*$ and any point on the approximate trajectory inside the ball $\mathscr{B}_\varepsilon(\mathbf{x}^*)$ as a function of (discrete) time. Once this distance function is known, we can estimate the exit time of the approximate trajectory from the ball $\mathscr{B}_\varepsilon(\mathbf{x}^*)$. In this vein, we develop an analytical framework in this work that approximates the trajectory for gradient descent within the saddle neighborhood and establish the fact that a linear escape rate from the saddle neighborhood is possible for some approximate trajectories generated by the gradient-descent method.

Third, we utilize the initial conditions on our iterate by projecting it onto a stable and an unstable subspace of the eigenvectors of the Hessian at the saddle point. This is extremely important since the escape rate and the associated necessary conditions are heavily dependent on where the iterate or gradient trajectory started. To this end, we simply make use of the strict saddle property to split the eigenspace of the Hessian at the saddle point into orthogonal subspaces of which two are of interest, namely, the stable subspace and the unstable subspace.[2] Taking the inner product of the iterate with these

[2]There can be one more orthogonal subspace corresponding to the zero eigenvalues of the Hessian at a strict saddle point.

subspaces yields the respective projections. (Note that this analysis of ours can be readily adapted to obtain these projections for *any* gradient-related method.) As a consequence, for any given initialization of our iterate within the saddle neighborhood, we provide the approximate iterate expression for the entire trajectory as long as it stays within this saddle neighborhood.

Finally, and most importantly, this work provides an upper bound on the exit time K_{exit} for approximations of (discrete) gradient-descent trajectories that is of the order $\mathcal{O}(\log(\varepsilon^{-1}))$, where the constants inside the $\mathcal{O}(\cdot)$ term explicitly depend on the condition number, dimension, and eigenvalue gap, as detailed in Section 3.5. Also, we develop a necessary condition on the initial iterate that is required for the existence of this exit time. It is worth noting that though the trajectory analysis developed in this work for the gradient-descent method is only approximate, we show in a follow-up work [11] that this approximation can only have a maximum relative error of order $\mathcal{O}(\log^2(\varepsilon^{-1})\varepsilon^{3/2})$, provided the exit time K_{exit} is at most of the order $\mathcal{O}(\log(\varepsilon^{-1}))$. Therefore our approximate analysis of the gradient-descent trajectories and their time of exit from the saddle neighborhood can be readily adapted to develop efficient algorithms for escaping first-order saddle points at a linear rate. One such algorithm has already been developed in [11], which extends the boundary conditions developed in this work for the linear exit time gradient trajectories and escapes saddle neighborhoods in linear time. The algorithm is designed to check the initial boundary conditions, after which it decides to either keep traversing along the same gradient trajectory or switch to a higher-order method for one iteration. To get a detailed understanding of this extension of our current work, we refer the reader to [11].

We conclude with Table 1, which highlights the similarities and differences between this work and other prior works that explicitly investigate the problem of characterizing the exit time from saddle neighborhoods. The *asymptotic* analyses in this table refer to works that provide measure-theoretic results in terms of the non-convergence of trajectories to a strict saddle point, whereas the *non-asymptotic* works deal with the analysis of trajectories exiting local saddle neighborhoods. The function classes $\mathscr{C}^2$ and $\mathscr{C}^\omega$ in the table represent twice continuously differentiable functions and analytic functions, respectively, while the class of quadratics ($\langle \mathbf{x}, \mathbf{A}\mathbf{x} \rangle$) represents functions with constant Hessian. The class of Morse functions is defined in Assumption **A4** in the next section. The map $\pi_{\mathscr{E}_{US}}(.)$ is the projection map onto the unstable subspace $\mathscr{E}_{US}$ of the Hessian $\nabla^2 f(\mathbf{x}^*)$, where this subspace will be formally defined in Lemma 3.2. Notice that the references [15, 32] provide exit times from a strict saddle neighborhood for the class of $\mathscr{C}^2$ functions but analyze continuous time dynamical systems, whereas this work provides the exit time analysis for the gradient descent method, which is a discrete dynamical system. Similarly the work [34] develops escape rates for discrete dynamical systems like gradient descent and the heavy ball method but restricts itself to the class of quadratic functions. The only work that develops escape rates for a discrete dynamical system on the class of $\mathscr{C}^2$ functions is [35] but that analysis is for a second-order Newton based method whereas we provide an exit time bound for a first-order method.

1.3 *Notation*

All vectors are in bold lower-case letters, all matrices are in bold upper-case letters, $\mathbf{0}$ is the n-dimensional null vector, $\mathbf{I}$ represents the $n \times n$ identity matrix, and $\langle \cdot, \cdot \rangle$ represents the inner product of two vectors. In addition, unless otherwise stated, all vector norms $\|\cdot\|$ are ℓ_2 norms, while the matrix norm $\|\cdot\|_2$ denotes the operator norm. Also, for any matrix expressed as $\mathbf{Z} + \mathcal{O}(c)$ where c is some scalar, the matrix-valued perturbation term $\mathcal{O}(c)$ is with respect to the Frobenius norm. Further, the symbol $(\cdot)^T$ is the transpose operator, the symbols $\mathcal{O}$, Ω, and Θ represent the Big-O, Big-Omega, and Big-Theta

Under the assumption of the function being a *Morse function*, however, this subspace vanishes.

Reference	Nature of analysis	Dynamical system analyzed	Function class	Exit time bound	Necessary initial conditions
[24]	Asymptotic	Gradient descent method	$\mathscr{C}^2$ functions	✗	✗
[34]	Asymptotic	Heavy ball method	$\mathscr{C}^2$ functions	✗	✗
[34]	Non-asymptotic	General accelerated methods, Gradient descent method	Quadratics ($\langle \mathbf{x}, \mathbf{A}\mathbf{x} \rangle$)	$\mathscr{O}(\log(\frac{1}{\Delta}))$ iterations from the unit ball $\mathscr{B}_1(\mathbf{x}^*)$	$\left\|\pi_{\mathscr{E}_{US}}(\mathbf{x}_0 - \mathbf{x}^*)\right\| \geqslant \Delta$
[32]	Non-asymptotic	Normalized gradient flow	$\mathscr{C}^2$ Morse functions	$\mathscr{O}(\varepsilon)$ exit time from a small neighborhood $\mathscr{B}_\varepsilon(\mathbf{x}^*)$	$\mathbf{x}_0 \neq \mathbf{x}^*$
[15]	Non-asymptotic	SDE-based gradient flow	$\mathscr{C}^2$ Morse functions	$\mathscr{O}(\log(\frac{1}{\tau}))$ mean exit time from some open neighborhood; τ is the scale of random perturbation	✗
[35]	Non-asymptotic	Newton-based method	$\mathscr{C}^2$ functions	$\mathscr{O}(\log(\frac{1}{\Delta}))$ iterations from some open neighborhood	$\left\|\pi_{\mathscr{E}_{US}}(\nabla f(\mathbf{x}_0))\right\| \geqslant \Delta$
This work	Non-asymptotic	Gradient descent method	Locally $\mathscr{C}^\omega$ Morse functions, $\mathscr{C}^2$ Morse functions	$\mathscr{O}(\log(\frac{1}{\varepsilon}))$ iterations from the ball $\mathscr{B}_\varepsilon(\mathbf{x}^*)$	$\left\|\pi_{\mathscr{E}_{US}}(\mathbf{x}_0 - \mathbf{x}^*)\right\|^2 \geqslant \Delta > \Omega(\varepsilon)$

Table 1: Summary of the similarities and differences between this work and some related prior works.

notation, respectively, and $W(\cdot)$ is the Lambert W function [8]. Throughout the paper, t represents the continuous-time index, while k, K are used for the discrete time. Next, $\gtrapprox$ and $\lessapprox$ mean 'approximately greater than' and 'approximately less than', respectively. Finally, the operator $\mathbf{dist}(\cdot,\cdot)$ returns the distance between two sets, $\mathbf{diam}(\cdot)$ returns the diameter of a set, and all the eigenvectors in this work are normalized to be unit vectors.

2. Problem formulation

Consider a non-convex smooth function $f(\cdot)$ that has strict first-order saddle points in its geometry. By strict first-order saddle points, we mean that the Hessian of function $f(\cdot)$ at these points has at least one negative eigenvalue, i.e., the function has negative curvature. Next, consider some neighborhood around a given saddle point. Formally, let $\mathbf{x}^*$ be some first-order strict saddle point of $f(\cdot)$ and let $\mathscr{B}_\varepsilon(\mathbf{x}^*)$ be an open ball around $\mathbf{x}^*$, where ε is sufficiently small. We then generate a sequence of iterates $\mathbf{x}_k$ from a gradient-related method on the function $f(\cdot)$, where we call the vector $\mathbf{u}_k = \mathbf{x}_k - \mathbf{x}^*$ inside the ball $\mathscr{B}_\varepsilon(\mathbf{x}^*)$ the **radial vector** (see Figure 1). Also, it is assumed that the initial iterate $\mathbf{x}_0 \in \bar{\mathscr{B}}_\varepsilon(\mathbf{x}^*) \backslash \mathscr{B}_\varepsilon(\mathbf{x}^*)$, where $\bar{\mathscr{B}}_\varepsilon(\mathbf{x}^*)$ is the closure of set $\mathscr{B}_\varepsilon(\mathbf{x}^*)$. With this initial boundary condition, we are interested in analyzing the behavior of our gradient-related sequence $\mathbf{x}_k$ in the vicinity of saddle point $\mathbf{x}^*$. More importantly, we are interested in finding some K_{exit} for which the subsequence $\{\mathbf{x}_k\}_{k>K_{exit}}$ lies outside $\mathscr{B}_\varepsilon(\mathbf{x}^*)$ and establishing that $K_{exit} \leqslant \mathscr{O}(\log(\varepsilon^{-1}))$. Finally, we have to obtain any necessary conditions on $\mathbf{x}_0$ that are required for the existence of this 'linear' exit time K_{exit}.

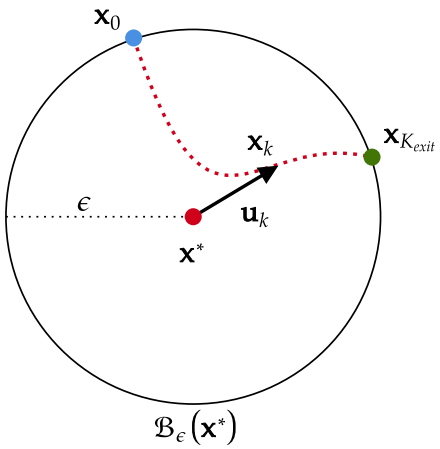


FIG. 1: The radial vector evolution in a saddle neighborhood for a function defined on $\mathbb{R}^2$.

2.1 *Assumptions*

Having briefly stated the problem, we formally state the set of assumptions that are required for this problem to be addressed in this work.

A1. *The function $f:\mathbb{R}^n\to\mathbb{R}$ is globally $\mathscr{C}^2$, i.e., twice continuously differentiable, and locally $\mathscr{C}^\omega$ in sufficiently large neighbodhoods of its saddle points, i.e., all the derivatives of this function are continuous around saddle points and the function $f(\cdot)$ also admits Taylor series expansion in these neighborhoods.*

A2. *The gradient of the function $f(\cdot)$ is $L-$Lipschitz continuous: $\|\nabla f(\mathbf{x})-\nabla f(\mathbf{y})\|\leqslant L\|\mathbf{x}-\mathbf{y}\|$.*

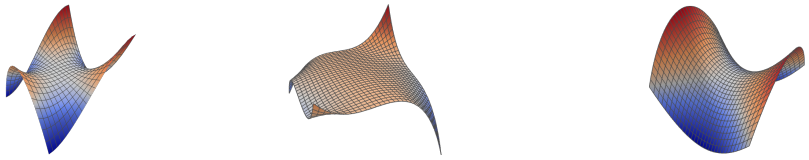
A3. *The Hessian of the function $f(\cdot)$ is $M-$Lipschitz continuous: $\left\|\nabla^2 f(\mathbf{x})-\nabla^2 f(\mathbf{y})\right\|_2\leqslant M\|\mathbf{x}-\mathbf{y}\|$.*

A4. *The function $f(\cdot)$ has only well-conditioned first-order stationary points, i.e., no eigenvalue of the function's Hessian is close to zero at these points (see Figure 2). Formally, if $\mathbf{x}^*$ is the first-order stationary point for $f(\cdot)$, then we have*

$$\nabla f(\mathbf{x}^*)=\mathbf{0},\text{ and}$$
$$\min_i|\lambda_i(\nabla^2 f(\mathbf{x}^*))|>\beta,$$

where $\lambda_i(\nabla^2 f(\mathbf{x}^*))$ denotes the i^{th} eigenvalue of the matrix $\nabla^2 f(\mathbf{x}^*)$ and $\beta>0$. Note that such a function is termed a Morse function.

We now make a few remarks concerning these assumptions as well as their implications. Notice that Assumption **A1** requires $f(\cdot)$ to be locally real analytic, which may seem too restrictive to some readers since the theory of non-convex optimization is often developed around only the assumption that $f\in\mathscr{C}^2$



Non-strict saddle Degenerate strict saddle **Morse function strict saddle**

FIG. 2: Possible cases of saddle points where the first figure corresponds to a monkey saddle, the second figure is a strict saddle with non-invertible Hessian at the saddle point and the third figure is strict saddle with invertible Hessian at the saddle point.

with Lipschitz-continuous Hessian. It is worth reminding the reader, however, that many practical non-convex problems such as quadratic programs, low-rank matrix completion, phase retrieval, etc., with appropriate smooth regularizers satisfy this assumption of real analyticity around the saddle neighborhoods; see, e.g., the formulations discussed in [7, 27]. Similarly, the loss functions in deep neural networks with analytic activation functions also satisfy Assumption **A1** under certain mild conditions [22]. It is also worth noting here that Assumption **A1** enables highly precise estimates of the exit time and the initial boundary condition, which is something that does not happen when dealing with purely $\mathscr{C}^2$ functions; see Section 3.2 for further discussion on this topic. Next, Assumptions **A2** and **A3** are satisfied locally around any saddle point since any locally analytic function is locally $\mathscr{C}^\infty$ smooth and therefore is gradient and Hessian Lipschitz continuous in some compact neighborhood of the saddle point.

Lastly, the problem formulation in this work assumes the class of Morse functions (Assumption **A4**), i.e., functions whose Hessians are invertible at their critical points. Since Morse functions can only have isolated critical points [28], the insights from this work are not *directly* applicable to non-convex optimization problems with connected saddle points. While this may appear to be a limitation of this work, Morse functions are an important tool in the study of general non-convex optimization problems since they are dense in the class of $\mathscr{C}^2$ functions [28]. It is therefore no surprise that they are routinely invoked in the non-convex optimization literature (see, e.g., [29, 31, 42]), while neural networks with smooth activation functions are also known to be Morse functions under certain mild assumptions [22]. Additionally, since connected saddle points for smooth functions generally arise only when their Hessian at the critical points has one or more zero eigenvalues, one could always add a quadratic regularization term with a sufficiently small constant to any smooth function so as to make the Hessian of the function invertible at its critical points and thus transform the function into a Morse function. As an example, we have circumvented the problem of connected saddle points within the low-rank matrix factorization problem in our follow-up work [11] by adding a regularization term that makes the objective function a Morse function.

Assumption **A4** also implies the following two propositions, both of which will be routinely invoked as part of the forthcoming analysis.

PROPOSITION 2.1 Under Assumption **A4**, the function $f(\cdot)$ has only first-order saddle points in its geometry. Moreover, these first-order saddle points are strict saddle, i.e., for any first-order saddle point $\mathbf{x}^*$, there exists at least one eigenvalue λ_i of $\nabla^2 f(\mathbf{x}^*)$ that satisfies $\lambda_i(\nabla^2 f(\mathbf{x}^*)) < -\beta$.

Proof. For any $\mathscr{C}^m$-smooth function $f(\cdot)$ with $m \geqslant 2$, if $\mathbf{x}^*$ is its second- or higher-order saddle point then it must necessarily satisfy $\nabla f(\mathbf{x}^*) = \mathbf{0}$ and $\nabla^2 f(\mathbf{x}^*) \succeq \mathbf{0}$, where at least one of the eigenvalues of $\nabla^2 f(\mathbf{x}^*)$ is 0. But this is not possible in our case because of Assumption **A4**. The fact that an eigenvalue

λ_i exists such that $\lambda_i(\nabla^2 f(\mathbf{x}^*)) < -\beta$ is also a direct consequence of Assumption **A4**. □

PROPOSITION 2.2 Under Assumption **A4**, for any sufficiently small ε where $\varepsilon \ll \beta$, we can group the eigenvalues of the Hessian $\nabla^2 f(\mathbf{x}^*)$ at any strict saddle point $\mathbf{x}^*$ into m disjoint sets $\{\mathscr{G}_1, \mathscr{G}_2, \ldots, \mathscr{G}_m\}$ with $2 \leqslant m \leqslant n$ based on the level of degeneracy of eigenvalues (closeness to one another) such that for some $\delta = \Omega(\varepsilon^{1-a})$ where $a \in (0,1]$, we have the following conditions:

$$\textbf{dist}(\mathscr{G}_p, \mathscr{G}_q) \geqslant \delta \ \ \forall \ \ \mathscr{G}_p, \mathscr{G}_q \ \text{ s.t. } \ p \neq q, \text{ and} \tag{2.1}$$

$$\max_p \{\textbf{diam}(\mathscr{G}_p)\} = \mathscr{O}(\varepsilon^{1-a}). \tag{2.2}$$

Proof. From Assumption **A4**, the eigenvalues of the Hessian $\nabla^2 f(\mathbf{x}^*)$ at any strict saddle point $\mathbf{x}^*$ can always be separated into two distinct groups, one consisting of positive eigenvalues and the other comprising negative eigenvalues. By this construction, the distance between these groups will be at least 2β. Since $\varepsilon \ll \beta$, we get a $\delta = 2\beta$ for this construction which satisfies the constraint $\delta = \Omega(1)$. Next, we check whether the diameter of these two groups is larger than $\Theta(\varepsilon^{1-a})$; if yes then we split that particular group into two more groups at the first eigenvalue where the consecutive eigenvalue gap within that group exceeds $\Theta(\varepsilon^{1-a})$. This eigenvalue gap becomes our new δ and by construction it will satisfy the constraint $\delta = \Omega(\varepsilon^{1-a})$ for some $a > 0$ since $\delta > \Theta(\varepsilon^{1-a})$. Repeating this process recursively, we would have constructed the disjoint sets $\{\mathscr{G}_1, \mathscr{G}_2, \ldots, \mathscr{G}_m\}$ with $2 \leqslant m \leqslant n$. Since n is finite, this process will terminate in finite steps (maximum $n-1$ steps) and therefore after the final splitting, we will obtain $\delta = \Omega(\varepsilon^{1-a})$ for some $a \in (0,1]$ such that $\max_p \{\textbf{diam}(\mathscr{G}_p)\} = \mathscr{O}(\varepsilon^{1-a})$. □

Proposition 2.2 describes a fundamental property of any $\mathscr{C}^2$ function that arises due to the algebraic multiplicity / (approximate) degeneracy of the eigenvalues of its Hessian at the saddle points. Note that, as a consequence of the strict-saddle property (Assumption **A4** / Proposition 2.1) and Proposition 2.2, we get the following necessary condition:

$$\beta \geqslant \frac{\delta}{2}. \tag{2.3}$$

3. Gradient trajectories and their approximations around strict saddle point

In this section, we analyze the behavior of the gradient descent algorithm in the vicinity of our strict saddle point, i.e., the region given by the set of points contained in $\mathscr{B}_\varepsilon(\mathbf{x}^*)$. It has been already established that gradient descent converges to minimizers and almost never ends up terminating into a strict saddle point [24]. However, the geometric structure of the region $\mathscr{B}_\varepsilon(\mathbf{x}^*)$ has not been utilized completely in prior works when it comes to developing rates of escape (possibly linear). Although linear rates of divergence from a strict saddle point are provided in [34] for the Nesterov accelerated gradient method, their analysis is reserved only for quadratic functions. Intuitively, for saddle neighborhoods with sufficient curvature magnitude β (Assumption **A4**, Proposition 2.1), there should exist some gradient trajectories that escape the saddle neighborhood $\mathscr{B}_\varepsilon(\mathbf{x}^*)$ with linear rate every time. Moreover, these trajectories should have some dependence on their initialization $\mathbf{x}_0$. To support this intuition of a linear escape rate, we first need an understanding of the behavior of gradient flow curves in the saddle point neighborhood, following which parallels can be drawn between flow curves and gradient trajectories.

We start by formally defining the gradient descent update and the corresponding flow curve equation. For a constant step size, the gradient descent method is given by

$$\mathbf{x}_{k+1} = \mathbf{x}_k - \alpha \nabla f(\mathbf{x}_k), \tag{3.1}$$

where α is the step size and we require that $\alpha \leqslant \frac{1}{L}$.

Next, the corresponding gradient flow curve is defined. If the step size α in (3.1) is taken to 0, the discrete iterate equation in index k of gradient descent can be transformed into a continuous-time ODE in t given by

$$\frac{d\mathbf{x}(t)}{dt} = -\nabla f(\mathbf{x}(t)), \tag{3.2}$$

which is the gradient flow equation in the limit of $\alpha \to 0$ [4]. Note that although $\|\mathbf{x}_{k+1} - \mathbf{x}_k\|$ is $\mathscr{O}(\varepsilon)$ here since both $\mathbf{x}_k$ and $\mathbf{x}_{k+1}$ lie inside $\mathscr{B}_\varepsilon(\mathbf{x}^*)$, we still require that $\alpha \to 0$ to transform the discrete iterate update into a continuous-time ODE.

We now state the following lemma about the gradient norm $\|\nabla f(\mathbf{x})\|$ when $\mathbf{x} \in \mathscr{B}_\varepsilon(\mathbf{x}^*)$.

LEMMA 3.1 For every point $\mathbf{x} \in \mathscr{B}_\varepsilon(\mathbf{x}^*)$, the gradient $\nabla f(\mathbf{x})$ will have $\mathscr{O}(\varepsilon)$ magnitude.

Proof. This can be verified using Assumption **A2**:

$$\|\nabla f(\mathbf{x}) - \nabla f(\mathbf{x}^*)\| \leqslant L\|\mathbf{x} - \mathbf{x}^*\| \leqslant L\varepsilon. \tag{3.3}$$

□

This lemma is of importance since it will help us in characterizing the gradients in the ball $\mathscr{B}_\varepsilon(\mathbf{x}^*)$ in terms of the Hessian $\nabla^2 f(\mathbf{x}^*)$ at the saddle point, from which we will develop approximations of gradient trajectories around the saddle point.

3.1 *Intuition behind the linear time of escape*

From the ODE analysis of flow curves for gradient-related methods such as those in [15, 20], it can be readily inferred that the gradient flow curves show hyperbolic behavior in the vicinity of saddle points. Since the discrete gradient method (3.1) is the Euler discretization of the gradient flow curve ODE (3.2), the geometric behavior of these two equations should be similar to one another with a deviation between them not more than of order $\mathscr{O}(\alpha)$ when the step size α is sufficiently small.[3] Therefore a crude analysis of flow curves should be sufficient to make approximate deductions for the discrete gradient method.

Concretely, we first define a time-varying vector $\mathbf{u}(t)$ that points to our iterate $\mathbf{x}(t)$ from the first-order strict saddle point $\mathbf{x}^*$. By this definition, we have that

$$\mathbf{u}(t) = \mathbf{x}(t) - \mathbf{x}^* \implies \frac{d\mathbf{u}(t)}{dt} = \frac{d\mathbf{x}(t)}{dt}. \tag{3.4}$$

Now, computing the norm squared of $\mathbf{u}(t)$, differentiating it with respect to t and using (3.2), we get

$$\|\mathbf{u}(t)\|^2 = \|\mathbf{x}(t) - \mathbf{x}^*\|^2 \tag{3.5}$$

$$\implies \frac{d\,\|\mathbf{u}(t)\|^2}{dt} = 2\langle(\mathbf{x}(t) - \mathbf{x}^*), -\nabla f(\mathbf{x}(t))\rangle. \tag{3.6}$$

Next, let the gradient flow curve enter $\mathscr{B}_\varepsilon(\mathbf{x}^*)$ ball at time $t = 0$ and exit this ball at time $t = T$. Geometrically, the inner product defined in (3.6) is negative at the entry point of the ball $\mathscr{B}_\varepsilon(\mathbf{x}^*)$ (i.e., vectors

[3]The actual deviation between the gradient flow curve and the gradient descent method after k iterations tends to be on the order of $\mathscr{O}(k\alpha)$ for a fixed ε. However, the factor k can be suppressed provided the trajectories generated by the two methods do not have large exit times.

$(\mathbf{x}(0)-\mathbf{x}^*)$ and $-\nabla f(\mathbf{x}(0))$ form an obtuse angle), becomes equal to 0 at some point $\mathbf{x}_{critical}$ inside this ball and is positive at the exit point (i.e., vectors $(\mathbf{x}(T)-\mathbf{x}^*)$ and $-\nabla f(\mathbf{x}(T))$ form an acute angle).

Using Taylor's expansion around $\mathbf{x}^*$ along the direction $\mathbf{x}(t)-\mathbf{x}^*$, we can write $\nabla f(\mathbf{x}(t))$ in the following manner:

$$\nabla f(\mathbf{x}(t)) = \nabla f(\mathbf{x}^*) + \int_{p=0}^{p=1} \nabla^2 f(\mathbf{x}^* + p\mathbf{u}(t))\mathbf{u}(t)dp. \tag{3.7}$$

If $\|\mathbf{u}(t)\|$ is sufficiently small or is of order $\mathscr{O}(\varepsilon)$, we can approximate $\nabla^2 f(\mathbf{x}^* + p\mathbf{u}(t)) \approx \nabla^2 f(\mathbf{x}^*)$. After substituting this approximation in (3.7) we obtain

$$\nabla f(\mathbf{x}(t)) \approx \nabla^2 f(\mathbf{x}^*)\mathbf{u}(t). \tag{3.8}$$

Using this result in (3.6) yields

$$\frac{d\|\mathbf{u}(t)\|^2}{dt} = 2\langle(\mathbf{x}(t)-\mathbf{x}^*), -\nabla f(\mathbf{x}(t))\rangle \approx -2\langle\mathbf{u}(t), \nabla^2 f(\mathbf{x}^*)\mathbf{u}(t)\rangle. \tag{3.9}$$

Also using (3.2) and (3.8) we get that

$$\frac{d\mathbf{u}(t)}{dt} = \frac{d\mathbf{x}(t)}{dt} \approx -\nabla^2 f(\mathbf{x}^*)\mathbf{u}(t). \tag{3.10}$$

Now consider the case where Assumptions **A1** to **A4** are satisfied. Since the eigenvalues of $\nabla^2 f(\mathbf{x}^*)$ are both positive and negative, the approximate ODE (3.10) will have the following solution:

$$\mathbf{u}(t) = \sum_{i=1}^{n} c_i \mathbf{v}_i(0) e^{-\lambda_i(0)t}, \tag{3.11}$$

where $(\lambda_i(0), \mathbf{v}_i(0))$ represents the i^{th} eigenvalue–eigenvector pair for the Hessian $\nabla^2 f(\mathbf{x}^*)$ and c_i are non-negative constants that depend on the initialization $\mathbf{u}(0)$. (Here, the non-negativity of c_i's can be assumed without loss of generality because the sign of the eigenvectors can be chosen arbitrarily.)

From this equation it is clear that we have a solution that is exponential in t. Moreover from the approximate ODE (3.9), it is evident that a hyperbolic curve is generated with an exponential rate of change. Therefore, for any initialization, i.e., for any choice of constants c_i, $\|\mathbf{u}(t)\|^2$ eventually increases at an exponential rate, thereby giving a linear escape rate for $\mathbf{x}(t)$ from the region $\mathscr{B}_\varepsilon(\mathbf{x}^*)$ provided $c_i \neq 0$ corresponding to at least one of the negative eigenvalues.

However, the approximation $\nabla^2 f(\mathbf{x}^* + p\mathbf{u}(t)) \approx \nabla^2 f(\mathbf{x}^*)$ fails to capture the first-order perturbation terms in the Hessian $\nabla^2 f(\mathbf{x}^* + p\mathbf{u}(t))$. Given a sufficiently small saddle neighborhood $\mathscr{B}_\varepsilon(\mathbf{x}^*)$, for any $\mathbf{x} \in \mathscr{B}_\varepsilon(\mathbf{x}^*)$, the eigenvalues and eigenvectors of the Hessian $\nabla^2 f(\mathbf{x})$ can have $\mathscr{O}(\varepsilon)$ variations with respect to those of the Hessian $\nabla^2 f(\mathbf{x}^*)$. Taking this $\mathscr{O}(\varepsilon)$ perturbation into account complicates the gradient flow curve analysis inside the $\mathscr{B}_\varepsilon(\mathbf{x}^*)$ ball,[4] which otherwise is straightforward from (3.10). Moreover, for all practical purposes, we cannot take our step size $\alpha \to 0$ for the sake of using ODE analysis. Choosing arbitrarily small step sizes causes the number of iterations needed to escape from the ball $\mathscr{B}_\varepsilon(\mathbf{x}^*)$ increase to infinity. Therefore a discrete gradient trajectory analysis using matrix perturbation theory becomes an absolute necessity to obtain trajectories (or approximate trajectories) with linear exit time from $\mathscr{B}_\varepsilon(\mathbf{x}^*)$.

[4]This is formally taken into account in subsequent sections using matrix perturbation theory.

3.2 *Warm-up: Rudimentary analysis of the exit time for discrete gradient trajectories*

The intuition developed as part of the ODE-based analysis *suggests* linear time escape of discrete gradient trajectories from strict-saddle neighborhoods. We now present a rudimentary analysis of the gradient descent method that uses elementary facts about first-order methods, as opposed to matrix perturbation theory, to derive a bound on the exit time of the gradient descent method from the saddle neighborhood $\mathscr{B}_\varepsilon(\mathbf{x}^*)$. The purpose of this analysis is twofold. First, it shows that (discrete) gradient-descent trajectories can indeed escape strict-saddle neighborhoods in linear time. Second, it highlights the limitations of existing analytical techniques in deriving linear escape rates for discrete gradient trajectories, thereby motivating the need for the matrix perturbation-based analysis of gradient trajectories in the next section for derivation of a linear escape rate from $\mathscr{B}_\varepsilon(\mathbf{x}^*)$.

Note that the analysis in this section requires only a relaxed version of Assumption **A1** on the function $f(\cdot)$, namely, it is twice continuously differentiable: $f \in \mathscr{C}^2$. But the remaining assumptions (Assumptions **A2**–**A4**) stay the same. Now consider the following that follows from the gradient descent iteration:

$$\begin{aligned}
\mathbf{x}_{k+1} - \mathbf{x}^* &= \mathbf{x}_k - \mathbf{x}^* - \alpha \nabla f(\mathbf{x}_k) && (3.12)\\
&= \mathbf{x}_k - \mathbf{x}^* - \alpha \nabla^2 f(\mathbf{x}^*)(\mathbf{x}_k - \mathbf{x}^*) - \alpha(\nabla f(\mathbf{x}_k) - \nabla^2 f(\mathbf{x}^*)(\mathbf{x}_k - \mathbf{x}^*)) && (3.13)\\
&= (\mathbf{I} - \alpha \nabla^2 f(\mathbf{x}^*))(\mathbf{x}_k - \mathbf{x}^*) - \alpha r(\mathbf{x}_k), && (3.14)
\end{aligned}$$

where $r(\mathbf{x}_k) = (\nabla f(\mathbf{x}_k) - \nabla^2 f(\mathbf{x}^*)(\mathbf{x}_k - \mathbf{x}^*))$. Using the Hessian Lipschitz continuity of $f(\cdot)$ and the fact that $\nabla f(\mathbf{x}_k) = \nabla f(\mathbf{x}^*) + \int_{p=0}^{p=1} \nabla^2 f(\mathbf{x}^* + p(\mathbf{x}_k - \mathbf{x}^*))(\mathbf{x}_k - \mathbf{x}^*) dp$ since $f \in \mathscr{C}^2$, we get that

$$\begin{aligned}
\|r(\mathbf{x}_k)\| &= \left\| \int_{p=0}^{p=1} \nabla^2 f(\mathbf{x}^* + p(\mathbf{x}_k - \mathbf{x}^*))(\mathbf{x}_k - \mathbf{x}^*) dp - \nabla^2 f(\mathbf{x}^*)(\mathbf{x}_k - \mathbf{x}^*) \right\| && (3.15)\\
&\leqslant \left(\int_{p=0}^{p=1} \left\| \nabla^2 f(\mathbf{x}^* + p(\mathbf{x}_k - \mathbf{x}^*)) - \nabla^2 f(\mathbf{x}^*) \right\| dp \right) \|(\mathbf{x}_k - \mathbf{x}^*)\| && (3.16)\\
&\leqslant \frac{M \|(\mathbf{x}_k - \mathbf{x}^*)\|^2}{2}. && (3.17)
\end{aligned}$$

Thus, $\|r(\mathbf{x}_k)\| \leqslant \frac{M\varepsilon^2}{2}$ whenever $\mathbf{x}_k \in \mathscr{B}_\varepsilon(\mathbf{x}^*)$. Inducting (3.14) up to $k = 0$ yields:

$$\mathbf{x}_{k+1} - \mathbf{x}^* = (\mathbf{I} - \alpha \nabla^2 f(\mathbf{x}^*))^{k+1}(\mathbf{x}_0 - \mathbf{x}^*) - \alpha \sum_{i=0}^{k} (\mathbf{I} - \alpha \nabla^2 f(\mathbf{x}^*))^{k-i} r(\mathbf{x}_i). \tag{3.18}$$

Next, in order to analyze the worst case bounds on the exit time, assume that the unstable subspace of $\nabla^2 f(\mathbf{x}^*)$ has dimension 1, i.e., $\lambda_j > 0$ for all $j \in \{1, 2, \ldots, n-1\}$ and $\lambda_n < 0$, where λ_j is the j^{th} eigenvalue of $\nabla^2 f(\mathbf{x}^*)$. Also let $\mathbf{v}_n$ be an eigenvector of $\nabla^2 f(\mathbf{x}^*)$ of unit norm corresponding to the eigenvalue λ_n, where $\lambda_n < -\beta$ from Assumption **A4**. Since divergence can happen only from the unstable subspace, our assumption on $\nabla^2 f(\mathbf{x}^*)$ will leave only a single direction of escape, i.e. along $\mathbf{v}_n$, for the gradient trajectories. Moreover since both $\mathbf{v}_n$ and $-\mathbf{v}_n$ will be the eigenvectors of $\nabla^2 f(\mathbf{x}^*)$, hence without loss of generality let us assume that $\langle \mathbf{v}_n, (\mathbf{x}_0 - \mathbf{x}^*) \rangle \geqslant 0$, where $\mathbf{x}_0 \in \bar{\mathscr{B}}_\varepsilon(\mathbf{x}^*) \backslash \mathscr{B}_\varepsilon(\mathbf{x}^*)$, and we are required to find the exit time K_{exit} that satisfies

$$K_{exit} = \inf_{k>0} \{k \,|\, \|\mathbf{x}_k - \mathbf{x}^*\| > \varepsilon\}. \tag{3.19}$$

As we show in Lemma A.1 in Appendix A, this is equivalent to the following condition:

$$K_{exit} = \inf_{k>0}\{k|\langle \mathbf{v}_n, (\mathbf{x}_k - \mathbf{x}^*)\rangle > \gamma_k \varepsilon\}, \tag{3.20}$$

where $\gamma_k = \frac{\langle \mathbf{v}_n, (\mathbf{x}_k - \mathbf{x}^*)\rangle}{\|\mathbf{x}_k - \mathbf{x}^*\|}$ and we have assumed for the sake of the crude analysis that $\gamma_k \in (0,1]$ for every k. Now, taking the inner product of $\mathbf{v}_n$ with $(\mathbf{x}_{k+1} - \mathbf{x}^*)$ in (3.14), and using the Hessian Lipschitz continuity and $\|r(\mathbf{x}_k)\| \leqslant \frac{M\varepsilon^2}{2}$, we get:

$$\langle \mathbf{v}_n, \mathbf{x}_{k+1} - \mathbf{x}^*\rangle = \langle \mathbf{v}_n, (\mathbf{I} - \alpha\nabla^2 f(\mathbf{x}^*))(\mathbf{x}_k - \mathbf{x}^*)\rangle - \alpha\langle \mathbf{v}_n, r(\mathbf{x}_k)\rangle \geqslant (1+\alpha\beta)\langle \mathbf{v}_n, \mathbf{x}_k - \mathbf{x}^*\rangle - \frac{\alpha M\varepsilon^2}{2}, \tag{3.21}$$

where we have used the substitution $(\mathbf{I} - \alpha\nabla^2 f(\mathbf{x}^*)) = \sum_{j=1}^{n}(1-\alpha\lambda_j)\mathbf{v}_j\mathbf{v}_j^T$. To show divergence from $\mathbf{x}^*$, it then suffices to show that for some $\rho \in (0,1)$ we have

$$(1+\alpha\beta)\langle \mathbf{v}_n, \mathbf{x}_k - \mathbf{x}^*\rangle - \frac{\alpha M\varepsilon^2}{2} \geqslant (1+\rho\alpha\beta)\langle \mathbf{v}_n, \mathbf{x}_k - \mathbf{x}^*\rangle \tag{3.22}$$

hold for all k, which will then imply that $\langle \mathbf{v}_n, \mathbf{x}_k - \mathbf{x}^*\rangle$ is strictly monotonically increasing with k.[5] Further simplifying (3.22) we get the condition

$$\beta(1-\rho)\langle \mathbf{v}_n, \mathbf{x}_k - \mathbf{x}^*\rangle \geqslant \frac{M\varepsilon^2}{2}, \tag{3.23}$$

which should hold for all k. A sufficient boundary condition for this inequality to hold is:

$$\langle \mathbf{v}_n, \mathbf{x}_0 - \mathbf{x}^*\rangle \geqslant \frac{M\varepsilon^2}{2\beta(1-\rho)}. \tag{3.24}$$

Now if the boundary condition (3.24) holds then from (3.21) and (3.22) we have:

$$\langle \mathbf{v}_n, \mathbf{x}_k - \mathbf{x}^*\rangle \geqslant (1+\rho\alpha\beta)^k\langle \mathbf{v}_n, \mathbf{x}_0 - \mathbf{x}^*\rangle \geqslant (1+\rho\alpha\beta)^k\frac{M\varepsilon^2}{2\beta(1-\rho)}. \tag{3.25}$$

Then using (3.20), exit from the ball $\mathscr{B}_\varepsilon(\mathbf{x}^*)$ can be guaranteed by setting the following condition:

$$\langle \mathbf{v}_n, \mathbf{x}_k - \mathbf{x}^*\rangle \geqslant (1+\rho\alpha\beta)^k\frac{M\varepsilon^2}{2\beta(1-\rho)} > \gamma_k\varepsilon \tag{3.26}$$

$$\iff \quad k \geqslant \frac{\log(\frac{2\gamma_k\beta(1-\rho)}{M\varepsilon})}{\log(1+\rho\alpha\beta)}, \tag{3.27}$$

which implies $K_{exit} \leqslant \frac{\log\left(\frac{2\gamma_{K_{exit}}\beta(1-\rho)}{M\varepsilon}\right)}{\log(1+\rho\alpha\beta)}$ as long as the sufficient condition (3.24) is satisfied.

[5] In general, we do not need the monotonicity condition for all k but only after a sufficiently large k that is smaller than the exit time. Such trajectories will also have linear exit times as proved in a subsequent counterexample.

The preceding rudimentary analysis guarantees a linear exit time bound for the gradient descent method under the sufficient boundary condition (3.24). But the resulting exit time bound is loose due to its dependency on the unknown factors $\gamma_{K_{exit}}$ and ρ, where $\gamma_{K_{exit}}$ could be arbitrarily small and the presence of ρ in the boundary condition makes this analysis more restrictive than the matrix perturbation-based analysis presented in Section 3.5. Also, the exit time analysis in this section does not bring out the dependence of boundary conditions and exit time bound on the problem dimension, conditioning of the neighborhood, and spectral gap, etc. Such dependencies are captured in the analysis of Section 3.5 and Table 2 in Section 3.6 summarizes the corresponding differences between the two analytical approaches. More importantly this analysis guarantees a linear exit time bound only for those trajectories starting at $\mathbf{x}_0$ that satisfy the monotonicity property implied by (3.21) and (3.22). That is, it does not capture the trajectories for which $\langle \mathbf{v}_n, \mathbf{x}_k - \mathbf{x}^* \rangle$ does not increase monotonically with k. A simple counterexample to the need for this monotonicity property for derivation of a linear exit time bound can be easily constructed. We refer the reader to Appendix F for one such counterexample. This implies there exist gradient trajectories that can exit in a linear time while violating the monotonicity condition, thereby illustrating that the rudimentary exit time analysis does not capture all the trajectories with linear exit times.

REMARK 3.1 Note that the sufficient condition of $\langle \mathbf{v}_n, \mathbf{x}_0 - \mathbf{x}^* \rangle \geqslant \frac{M\varepsilon^2}{2\beta(1-\rho)}$ from (3.24) guarantees linear exit time gradient trajectories. Moreover this condition makes sure that such trajectories do not have zero measure since the set of initialization given by $\{\mathbf{x}_0 \mid \langle \mathbf{v}_n, \mathbf{x}_0 - \mathbf{x}^* \rangle \geqslant \frac{M\varepsilon^2}{2\beta(1-\rho)}\}$ has positive measure for sufficiently small ε.

In summary, to analyze the complete set of gradient trajectories around the saddle point that escape in linear time and develop a precise exit time bound we need more than the class of twice-differentiable functions; hence the need to work with analytic functions.[6] Note that many optimization and learning problems, such as quadratic functions and deep neural networks with smooth activation functions, satisfy real analyticity in some neighborhoods of stationary points, if not over the entire domain.

3.3 *An informal statement of the main result*

In this section, we provide an informal statement of the main result of this paper as well as a brief discussion of the implications of this result.

THEOREM 3.1 (Informal Main Result) Under Assumptions **A1**–**A4**, the approximate trajectories of the gradient descent method with step size $\alpha = \frac{1}{L}$, when initialized on the boundary of some ε neighborhood of a strict saddle point $\mathbf{x}^*$ of $f(\cdot)$, where $\varepsilon < \min\left\{\frac{2\beta}{M}, \Omega\left(\frac{\delta}{n^2}\right)\right\}$ and $\varepsilon \ll 1$, can exit this neighborhood in approximately linear time, i.e., $K_{exit} \lessapprox \mathcal{O}\left(\log\left(\frac{\delta}{\varepsilon n}\right)\right)$, where K_{exit} is the exit time for the approximate trajectory, n is the problem dimension and δ is the eigen gap from Proposition 2.2. However, this linear exit time bound holds only if the initial radial vector $\mathbf{u}_0 = \mathbf{x}_0 - \mathbf{x}^*$ is not orthogonal to the unstable subspace of $\nabla^2 f(\mathbf{x}^*)$ and subtends some non-zero angle with the stable subspace of $\nabla^2 f(\mathbf{x}^*)$. In particular, the cosine square of the angle between the initial radial vector and the unstable subspace

[6] In order to get a highly precise bound on exit time, we need the best possible first-order approximations of gradient trajectories, which can only be obtained for analytic functions. Therefore even the class of $\mathscr{C}^\infty$ functions is not sufficient for our analysis; see also the discussion in Remark 3.5 in Section 3.5.1 in this regard.

of $\nabla^2 f(\mathbf{x}^*)$ must be at least of the order $\Omega\left(\frac{\varepsilon n}{\delta}\right)$, where this cosine square is referred to as the unstable subspace projection value.

A formal statement of this result, which includes precise characterizations of the approximate trajectory, exit time, and the bounds on ε as well as the necessary initial unstable subspace projections, is provided in Theorem 3.3. We also refer the reader to Figure 3 for a concrete intuition of the angle between the initial radial vector and the unstable subspace of $\nabla^2 f(\mathbf{x}^*)$ as well as its relation to the unstable subspace projection.

We now briefly summarize the implications of this main result, while additional discussion is provided after Theorem 3.3. For a function $f(\cdot)$ satisfying Assumptions **A1**–**A4**, let the gradient descent method with step size $\alpha = \frac{1}{L}$ be initialized on the boundary of some ε neighborhood of a strict saddle point $\mathbf{x}^*$ of $f(\cdot)$ such that the initial radial vector $\mathbf{u}_0 = \mathbf{x}_0 - \mathbf{x}^*$ subtends some angle with the unstable subspace of $\nabla^2 f(\mathbf{x}^*)$ that is not equal to $\frac{\pi}{2}$. Then we have the following statements:

S1. There exists some lower bound on the cosine square of this angle (termed as the 'sufficient condition') for which the approximate trajectories of the gradient descent method will exit the saddle neighborhood in linear time.

S2. Also, there exists a strict lower bound on the cosine square of this angle (termed as the 'necessary condition') that is of the order $\Omega\left(\frac{\varepsilon n}{\delta}\right)$. If the cosine square of the angle between the initial radial vector and the unstable subspace of $\nabla^2 f(\mathbf{x}^*)$ is smaller than $\Theta\left(\frac{\varepsilon n}{\delta}\right)$, the approximate trajectories of the gradient descent method can never exit the saddle neighborhood in linear time.

This work rigorously establishes Statement **S2** and also shows that Statement **S1** is not vacuous (cf. Section E.0.3 in Appendix E). Note that a rigorous characterization of the lower bound in Statement **S1** requires a more sophisticated proof machinery, which has been pursued in our follow-up work [11].

REMARK 3.2 A fast exit time in terms of the scaling with $\frac{1}{\varepsilon}$ in and of itself might not preclude the gradient descent method from converging super slowly in the worst case. The carefully constructed function with cascaded saddles in [12], in particular, is a prime example of this behavior, as the gradient descent method takes an exponentially—in dimension n—large time in the worst case to escape the cascaded saddles and converge to a local minimum for this function. However, the particular class of functions within the family of Morse functions being considered in this work excludes the construction in [12]. Going further, we have established in a follow-up work [11] that the time to escape cascaded saddles and reach a second-order stationary point for functions in this class does not scale exponentially in the dimension for a simple variant of the gradient descent method.

3.4 *Brief overview of results and proof sketch for the linear exit time bound*

Our matrix perturbation-based analysis utilizes the standard gradient-descent method (3.1) in the saddle neighborhood $\mathscr{B}_\varepsilon(\mathbf{x}^*)$. Since we are interested in developing analysis suited only for the region $\mathscr{B}_\varepsilon(\mathbf{x}^*)$, we assume that initially our iterate $\mathbf{x}_0$ sits on the boundary of $\bar{\mathscr{B}}_\varepsilon(\mathbf{x}^*)$. We then follow the given sequence of steps in order to obtain linear exit time bound for approximations of gradient descent trajectories around a saddle point.

1. Starting with Lemma 3.2 we show that the region $\mathscr{B}_\varepsilon(\mathbf{x}^*)$ around the strict saddle point $\mathbf{x}^*$ is comprised of a stable and an unstable subspace, which are orthogonal to one another.

2. Next, for any $\mathbf{x} \in \mathscr{B}_\varepsilon(\mathbf{x}^*)$ we write $\nabla f(\mathbf{x})$ in terms of the radial vector $\mathbf{u} = \mathbf{x} - \mathbf{x}^*$ as $\nabla f(\mathbf{x}) = \left(\int_{p=0}^{p=1} \nabla^2 f(\mathbf{x}^* + p\mathbf{u}) dp\right)\mathbf{u}$.

3. Then in Lemma 3.3 using matrix perturbation theory we express the Hessian $\nabla^2 f(\mathbf{x})$ at $\mathbf{x} = \mathbf{x}^* + p\mathbf{u}$, where $\mathbf{x} \in \mathscr{B}_\varepsilon(\mathbf{x}^*)$, $p \in [0,1]$, and $\|\mathbf{u}\| \leqslant \varepsilon$ in terms of a perturbation of $\nabla^2 f(\mathbf{x}^*)$, as

$$\nabla^2 f(\mathbf{x}^* + p\mathbf{u}) = \nabla^2 f(\mathbf{x}^*) + \mathbf{D}(\mathbf{x}),$$

with the perturbation matrix $\mathbf{D}(\mathbf{x})$ bounded as

$$\|\mathbf{D}(\mathbf{x})\| \leqslant Mp\varepsilon.$$

4. We iterate the Gradient descent method in terms of the radial vector $\mathbf{u}_k$ as follows:

$$\begin{aligned} \mathbf{u}_{k+1} = \mathbf{x}_k - \mathbf{x}^* - \alpha\nabla f(\mathbf{x}_k) &= \left(\mathbf{I} - \alpha\int_0^1 \nabla^2 f(\mathbf{x}^* + p\mathbf{u}_k)dp\right)\mathbf{u}_k \\ \implies \mathbf{u}_{k+1} &= \Bigg(\mathbf{I} - \alpha\nabla^2 f(\mathbf{x}^*) - \underbrace{\alpha\int_0^1 \mathbf{D}(\mathbf{x}^* + p\mathbf{u}_k)dp}_{\mathbf{R}(\mathbf{u}_k) = \mathscr{O}(\varepsilon)}\Bigg)\mathbf{u}_k \end{aligned}$$

where $\|\mathbf{R}(\mathbf{u}_k)\| = \left\|\alpha\int_0^1 \mathbf{D}(\mathbf{x}^* + p\mathbf{u}_k)dp\right\| = \mathscr{O}(\varepsilon)$ from the last step. Using this radial vector update in Lemma 3.4, we induct the above recursion up to initialization $\mathbf{u}_0$ and obtain the exact trajectory expression:

$$\mathbf{u}_{K+1} = \Pi_{k=0}^K \left(\mathbf{I} - \alpha\nabla^2 f(\mathbf{x}^*) - \mathbf{R}(\mathbf{u}_k)\right)\mathbf{u}_0.$$

5. In Lemma 3.5 we expand the product of the $K+1$ non-commuting matrices from the last step up to first order as follows:

$$\tilde{\mathbf{u}}_{K+1} := \Pi_{k=0}^K \mathbf{A}_k\mathbf{u}_0 - \sum_{r=0}^K (\Pi_{k=r+1}^K \mathbf{A}_r\mathbf{R}(\mathbf{u}_r)\Pi_{k=0}^{r-1}\mathbf{A}_r)\mathbf{u}_0,$$

where $\tilde{\mathbf{u}}_{K+1} \approx \mathbf{u}_{K+1}$ and $\mathbf{A}_k := \mathbf{I} - \alpha\nabla^2 f(\mathbf{x}^*)$ for all k in the case of gradient descent. This is the most crucial step in the analysis since we obtain the approximate trajectory $\{\tilde{\mathbf{u}}_K\}$ in this step.[7]

6. The approximate trajectory $\{\tilde{\mathbf{u}}_K\}$ obtained above cannot be uniquely determined since it is a function of the eigenvalues of the Hessian $\nabla^2 f(\mathbf{x}^*)$, which are known only up to an interval. Therefore in Lemma 3.6 we obtain a parametrized family of approximate trajectories for a fixed $\mathbf{u}_0$, denoted by $\{\tilde{\mathbf{u}}_K^\tau\}$, where the parameter $\tau \in \mathbb{R}$ varies with variations in the eigenvalues of the Hessian $\nabla^2 f(\mathbf{x}^*)$. Next, we construct the *minimal* approximate trajectory from this family, defined as one that stays closest to $\mathbf{x}^*$ for each K and show that this *minimal* approximate trajectory has the maximum exit time among all approximate trajectories.

[7]Even though $\mathbf{A}_k$ is constant for the gradient-descent iteration (3.1), we have purposefully not removed its subscript k since it may not be constant for a general dynamical system. Consider, for instance, gradient descent with variable step size α_k instead of constant step size α and we then have $\mathbf{A}_k = \mathbf{I} - \alpha_k\nabla^2 f(\mathbf{x}^*)$. Hence, with the subscript k intact, the expression for the approximate trajectory $\{\tilde{\mathbf{u}}_K\}$ can be easily adapted to a general class of first-order methods.

7. In Theorem 3.2 we obtain the closed form expression of the normalized radial distance for the *minimal* approximate trajectory given by $\Psi(K)$ where $\varepsilon^2\Psi(K) \leqslant \inf_\tau \|\tilde{\mathbf{u}}_K^\tau\|^2 < \varepsilon^2$.
8. Finally in Theorem 3.3 we obtain the smallest upper bound on K of the order $\mathscr{O}(\log(\varepsilon^{-1}))$ that satisfies the condition $\Psi(K) > 1$ which will imply $\varepsilon^2 < \varepsilon^2\Psi(K) \leqslant \inf_\tau \|\tilde{\mathbf{u}}_K^\tau\|^2$. This condition gives the linear exit time bound from the saddle neighborhood. We then derive any necessary conditions on $\mathbf{x}_0$ for guaranteeing this linear exit time.

Before formally beginning our analysis of discrete gradient trajectories, we state the following lemma that will be utilized frequently in our analysis.

LEMMA 3.2 For any point $\mathbf{x} \in \mathscr{B}_\varepsilon(\mathbf{x}^*)$, the vector $\mathbf{u}$ given by $\mathbf{u} = \mathbf{x} - \mathbf{x}^*$ belongs to a vector space $\mathscr{E}$ that is comprised of a stable subspace $\mathscr{E}_S$ (subspace corresponding to contraction dynamics) and an unstable subspace $\mathscr{E}_{US}$ (subspace corresponding to expansive dynamics). Formally, this can be written as

$$\mathscr{E} = \mathscr{E}_S \bigoplus \mathscr{E}_{US},$$

where $\bigoplus$ denotes the direct sum of two spaces.

Proof. The eigenvalues of the Hessian $\nabla^2 f(\mathbf{x}^*)$ are both positive and negative. Without loss of generality, these can be classified into two sets of stable and unstable eigenvalues with the stable set comprising positive eigenvalues and the unstable set having negative eigenvalues. Then the corresponding subspaces can be written as

$$\mathscr{E}_S = span\{\mathbf{v}_i | \lambda_i(\nabla^2 f(\mathbf{x}^*)) > 0\}, \text{ and} \tag{3.28}$$

$$\mathscr{E}_{US} = span\{\mathbf{v}_j | \lambda_j(\nabla^2 f(\mathbf{x}^*)) < 0\}, \tag{3.29}$$

where $\lambda_i(\nabla^2 f(\mathbf{x}^*)), \mathbf{v}_i$ represent the i^{th} eigenvalue-eigenvector pair. Since these subspaces are orthogonal and span the complete space $\mathscr{E} = \mathbb{R}^n$, any vector $\mathbf{u} = \mathbf{x} - \mathbf{x}^*$ is spanned by these subspaces. Next, we define the two index sets $\mathscr{N}_S = \{i | \lambda_i(\nabla^2 f(\mathbf{x}^*)) > 0\}$ and $\mathscr{N}_{US} = \{j | \lambda_j(\nabla^2 f(\mathbf{x}^*)) < 0\}$ for the two subspaces. Since these subspaces are orthogonal, their index sets are disjoint. □

3.5 *Analysis of discrete gradient trajectories using matrix perturbation theory*

Now that we have established all the necessary preliminaries, we can move on to develop approximate bounds on the escape time from the region $\mathscr{B}_\varepsilon(\mathbf{x}^*)$ for gradient descent. From here onward we restrict ourselves to discrete time iterates denoted by subscripts k and the entire analysis is carried out in discrete time. Also, we assume that Assumptions **A1** to **A4** hold along with the additional condition of $m = n$ in Proposition 2.2, i.e., there are no degenerate eigenvalues. Section 3.5.1 after Lemma 3.3 discusses the analysis for the degenerate eigenvalues, i.e., the case when $m \neq n$ in Proposition 2.2. In there, we show that the analysis for the degenerate case is very straightforward and easy to extend from the non-degenerate analysis. It should also be noted that instead of analyzing exact trajectories, we analyze from here onward the first-order approximations of the exact trajectories, where the approximation error is sufficiently small. The presence of the higher-order terms ($\mathscr{O}(\varepsilon^2)$ terms) in the forthcoming analysis accounts for the approximation in our analysis, and things are proved about trajectories and perturbations up to the first order in ε. To summarize our next set of steps, we begin with a lemma that characterizes the approximate Hessian behavior in the region $\mathscr{B}_\varepsilon(\mathbf{x}^*)$, followed by a lemma that expresses $\mathbf{x}_k$ for any $k \geqslant 0$ approximately in terms of $\mathbf{x}_0$ and a theorem that characterizes an approximate lower bound on the distance of $\mathbf{x}_k$ from $\mathbf{x}^*$.

LEMMA 3.3 Let $r_j(\mathbf{u})$ be a function of the vector $\mathbf{u}$ defined as $r_j(\mathbf{u}) = \left\| \left(\frac{d^j}{dw^j} \nabla^2 f(\mathbf{x}^* + w\mathbf{u}) \Big|_{w=0} \right) \right\|_2$ and $\varepsilon > 0$ be a constant that satisfies the necessary condition of $\varepsilon < \inf_{\|\mathbf{u}\|=1} \left(\limsup_{j\to\infty} \sqrt[j]{\frac{r_j(\mathbf{u})}{j!}} \right)^{-1}$. Then for any $\mathbf{x}_k \in \mathscr{B}_\varepsilon(\mathbf{x}^*)$ such that $\mathbf{x}_k = \mathbf{x}^* + p\mathbf{u}_k$ with $0 < p \leqslant 1$, the Hessian $\nabla^2 f(\mathbf{x}_k)$ is given by

$$\nabla^2 f(\mathbf{x}_k) = \nabla^2 f(\mathbf{x}^*) + p\,\|\mathbf{u}_k\|\,\mathbf{H}(\hat{\mathbf{u}}_k) + \mathscr{O}(\varepsilon^2), \tag{3.30}$$

where $\hat{\mathbf{u}}_k = \frac{\mathbf{u}_k}{\|\mathbf{u}_k\|}$ and we have that

$$\begin{aligned}\mathbf{H}(\hat{\mathbf{u}}_k) = \sum_{i=1}^{n} \Bigg(& \langle \mathbf{v}_i(0), \mathbf{H}(\hat{\mathbf{u}}_k)\mathbf{v}_i(0)\rangle \mathbf{v}_i(0)\mathbf{v}_i(0)^T + \lambda_i(0)\Big(\sum_{l\neq i} \frac{\langle \mathbf{v}_l(0), \mathbf{H}(\hat{\mathbf{u}}_k)\mathbf{v}_i(0)\rangle}{\lambda_i(0) - \lambda_l(0)} \mathbf{v}_l(0)\Big)\mathbf{v}_i(0)^T \\ & + \lambda_i(0)\mathbf{v}_i(0)\Big(\sum_{l\neq i} \frac{\langle \mathbf{v}_l(0), \mathbf{H}(\hat{\mathbf{u}}_k)\mathbf{v}_i(0)\rangle}{\lambda_i(0) - \lambda_l(0)} \mathbf{v}_l(0)\Big)^T \Bigg)\end{aligned} \tag{3.31}$$

with $\lambda_i(0), \mathbf{v}_i(0)$ being the i^{th} eigenvalue–eigenvector pair of the Hessian $\nabla^2 f(\mathbf{x}^*)$.

The proof of this lemma is given in Appendix B. Note that the expression for $\mathbf{H}(\hat{\mathbf{u}}_k)$ in the lemma statement is more of a property rather than a definition, where $\|\mathbf{H}(\hat{\mathbf{u}}_k)\|_2$ is bounded. However, it may not be the case that $\mathbf{H}(\hat{\mathbf{u}}_k) = \mathscr{O}(\varepsilon)$. In particular, we have the following bound from inequality (C.16) in Appendix C:

$$\|\mathbf{H}(\hat{\mathbf{u}}_k)\|_2 \leqslant M + \mathscr{O}(\varepsilon), \tag{3.32}$$

which suggests that $\|\mathbf{H}(\hat{\mathbf{u}}_k)\|_2$ could even be a constant-order term; see Appendix C for further details.

REMARK 3.3 The condition $\varepsilon < \inf_{\|\mathbf{u}\|=1} \left(\limsup_{j\to\infty} \sqrt[j]{\frac{r_j(\mathbf{u})}{j!}} \right)^{-1}$ is necessary but may not be sufficient to guarantee this lemma's result. Since evaluating the radius of convergence for an expansion generated by the Rayleigh–Schrödinger perturbation analysis is beyond the scope of this work, we only put forth this necessary condition here.

REMARK 3.4 Note that the quantity $\inf_{\|\mathbf{u}\|=1} \left(\limsup_{j\to\infty} \sqrt[j]{\frac{r_j(\mathbf{u})}{j!}} \right)^{-1}$ is exactly equal to the radius of convergence for the Taylor series expansion of the matrix $\nabla^2 f(\mathbf{x}^* + w\mathbf{u})$ about $w > 0$ for all $\{\mathbf{u} : \|\mathbf{u}\|_2 = 1\}$, which is strictly positive due to the analytic nature of $f(\cdot)$. A proof of this claim is given in Appendix B.

3.5.1 *Statement about the generality of Lemma 3.3.* It should be noted that while obtaining (3.31), we assumed a minimum gap of δ between any two eigenvalues of the Hessian $\nabla^2 f(\mathbf{x}^*)$. However, we can have many groups of equal or almost similar eigenvalues from Proposition 2.2; this creates singular terms in the coefficient denominators of first-order eigenvector corrections in (3.31). This can be solved easily from the degenerate matrix perturbation theory, which extends the results of Rayleigh–Schrödinger theory. From that we obtain the following new first-order correction term in place of (B.16) in the proof of the lemma for the i^{th} eigenvector $\tilde{\mathbf{v}}_i(w)$:

$$\frac{d}{dw}(\tilde{\mathbf{v}}_i(w))\Big|_{w=0} = \sum_{l\notin\mathscr{G}_p} \frac{\langle \tilde{\mathbf{v}}_l(0), \mathbf{H}(\hat{\mathbf{u}}_k)\tilde{\mathbf{v}}_i(0)\rangle}{\lambda_i(0) - \lambda_l(0)} \tilde{\mathbf{v}}_l(0), \tag{3.33}$$

where the corresponding i^{th} unperturbed eigenvalue $\lambda_i(0)$ belongs to the set $\mathscr{G}_p$. Also note that we have a new basis of eigenvectors $\tilde{\mathbf{v}}_i$ instead of $\mathbf{v}_i$, which resolves the degeneracy issue within the groups of similar eigenvalues. This change of basis can always be done since there are infinitely many solutions to the eigenvectors belonging to the degenerate subspaces. More importantly, we are never required to compute these eigenvectors explicitly in our analysis. To get a detailed understanding of the degenerate matrix perturbation theory, the reader can refer to [6, 14].

Therefore for the case with degenerate eigenvalue sets, the analysis will remain the same, but with fewer first-order perturbation terms ((3.33) has $n-|\mathscr{G}_p|$ orthogonal terms in the summation instead of the $n-1$ orthogonal terms that appear in (B.16)). Now, these fewer $\mathscr{O}(\varepsilon)$ terms in (3.33) will result in weaker first-order perturbations on the distance $\|\mathbf{x}_k-\mathbf{x}^*\|$ when compared to that from (B.16). In a subsequent lemma (Lemma 3.6), it will be established that the worst-case trajectory is obtained when the first-order perturbation terms are used to minimize $\|\mathbf{x}_k-\mathbf{x}^*\|$ for every k. This worst-case trajectory stays inside the ball $\mathscr{B}_\varepsilon(\mathbf{x}^*)$ for the maximum number of iterations. For the case of degenerate eigenvalues, fewer first-order terms from (3.33) means a weaker perturbation effect over $\|\mathbf{x}_k-\mathbf{x}^*\|$, which implies that $\|\mathbf{x}_k-\mathbf{x}^*\|$ cannot be minimized completely. This is in contrast to the case of (B.16) which has more first-order terms $(n-1)$ and hence a stronger perturbation effect over $\|\mathbf{x}_k-\mathbf{x}^*\|$. Now, a stronger perturbation can be used to contain the worst-case trajectory inside $\mathscr{B}_\varepsilon(\mathbf{x}^*)$ for a longer duration (part of the proof for Lemma 3.6). As a consequence, the worst-case trajectory from the non-degenerate case will have a larger exit time compared to that of the degenerate case. Therefore, we are not required to perform the analysis for the degenerate case because the worst-case performance in terms of exit time is captured in the current analysis for the non-degenerate case.

REMARK 3.5 It is worth noting here that the exit time analysis in this work could have been carried out using the Davis–Kahan theorem [10]. Such an analysis would have required the function $f(\cdot)$ to only be $\mathscr{C}^2$, as opposed to analytic, but it would have necessitated the eigensubspaces of the Hessian $\nabla^2 f(\mathbf{x}^*)$ to be non-degenerate. However, non-degeneracy of the eigensubspaces is a much stronger assumption in many real-world problems than the analyticity assumption of the function $f(\cdot)$, which is needed for use of the degenerate matrix perturbation theory in our analysis.

We now move on to the lemmas that express $\mathbf{x}_K \in \mathscr{B}_\varepsilon(\mathbf{x}^*)$ for any $K \geqslant 0$ approximately in terms of $\mathbf{x}_0$ provided K and ε satisfy certain necessary conditions.

LEMMA 3.4 Given an initialization of the radial vector $\mathbf{u}_0$ and $\varepsilon < \inf_{\|\mathbf{u}\|=1}\left(\limsup_{j\to\infty}\sqrt[j]{\frac{r_j(\mathbf{u})}{j!}}\right)^{-1}$, at any iteration K the radial vector $\mathbf{u}_K$ is given by the product

$$\mathbf{u}_K = \prod_{k=0}^{K-1}\Big[\mathbf{A}_k + \varepsilon\mathbf{P}_k\Big]\mathbf{u}_0, \tag{3.34}$$

where $\varepsilon\mathbf{P}_k = \mathbf{B}_k + \mathscr{O}(\varepsilon^2)$, $\mathbf{B}_k = \mathscr{O}(\varepsilon)$ for $\mathbf{x}_k \in \mathscr{B}_\varepsilon(\mathbf{x}^*)$ and $\mathbf{A}_k, \mathbf{B}_k$ are given by the following equations:

$$\mathbf{A}_k = \sum_{i\in\mathscr{N}_S} c_i^s(k)\mathbf{v}_i(0)\mathbf{v}_i(0)^T + \sum_{j\in\mathscr{N}_{US}} c_j^{us}(k)\mathbf{v}_j(0)\mathbf{v}_j(0)^T \tag{3.35}$$

$$\mathbf{B}_k = \sum_{i=1}^{n}\sum_{l\neq i}\Big(d_{l,i}(k)\mathbf{v}_l(0)\mathbf{v}_i(0)^T + d_{i,l}(k)\mathbf{v}_i(0)\mathbf{v}_l(0)^T\Big). \tag{3.36}$$

The coefficient terms $c_i^s(k)$, $c_j^{us}(k)$, $d_{i,l}(k)$ and $d_{l,i}(k)$ are as follows:

$$c_i^s(k) = \left(1 - \alpha\lambda_i(0) - \alpha\frac{\|\mathbf{u}_k\|}{2}\langle \mathbf{v}_i(0), \mathbf{H}(\hat{\mathbf{u}}_k)\mathbf{v}_i(0)\rangle\right), \tag{3.37}$$

$$c_j^{us}(k) = \left(1 - \alpha\lambda_j(0) - \alpha\frac{\|\mathbf{u}_k\|}{2}\langle \mathbf{v}_j(0), \mathbf{H}(\hat{\mathbf{u}}_k)\mathbf{v}_j(0)\rangle\right), \text{ and} \tag{3.38}$$

$$d_{i,l}(k) = d_{l,i}(k) = \frac{\langle \mathbf{v}_l(0), \mathbf{H}(\hat{\mathbf{u}}_k)\mathbf{v}_i(0)\rangle\lambda_i(0)\alpha\|\mathbf{u}_k\|}{2(\lambda_l(0) - \lambda_i(0))}. \tag{3.39}$$

Further, suppose $\nu_n \leqslant \cdots \leqslant \nu_1$ are the absolute values of the eigenvalues of the matrix $\prod_{k=0}^{K-1}\left[\mathbf{A}_k + \varepsilon\mathbf{P}_k\right]$ and we have that $\sup_{0\leqslant k\leqslant K-1}\|\mathbf{A}_k\|_2 = \|\mathbf{A}\|_2$, $\sup_{0\leqslant k\leqslant K-1}\left\|\mathbf{A}_k^{-1}\right\|_2 = \left\|\mathbf{A}^{-1}\right\|_2$ and $\sup_{0\leqslant k\leqslant K-1}\|\mathbf{P}_k\|_2 = \|\mathbf{P}\|_2$ for some matrices $\mathbf{A}$ and $\mathbf{P}$. Then for $\varepsilon < \min\left\{\inf_{\|\mathbf{u}\|=1}\left(\limsup_{j\to\infty}\sqrt[j]{\frac{r_j(\mathbf{u})}{j!}}\right)^{-1}, \frac{\|\mathbf{A}^{-1}\|_2^{-1}}{\|\mathbf{P}\|_2}\right\}$ and $K\varepsilon \ll 1$, the following condition holds provided $\mathbf{A}_k$ is non-singular for all k:

$$\left\|\mathbf{A}^{-1}\right\|_2^{-K}\left(1 - K\varepsilon\frac{\|\mathbf{P}\|_2}{\|\mathbf{A}^{-1}\|_2^{-1}} - \mathscr{O}\left((K\varepsilon)^2\right)\right) \leqslant \nu_n \leqslant \cdots \leqslant \nu_1 \leqslant \|\mathbf{A}\|_2^K\left(1 + K\varepsilon\frac{\|\mathbf{P}\|_2}{\|\mathbf{A}\|_2} + \mathscr{O}\left((K\varepsilon)^2\right)\right). \tag{3.40}$$

The proof of this lemma is given in Appendix C. This lemma states that the radial vector $\mathbf{u}_K$ evolves linearly at every iteration K, where the transition matrix from the initial state $\mathbf{u}_0$ to the state $\mathbf{u}_K$ is given by $\prod_{k=0}^{K-1}\left[\mathbf{A}_k + \varepsilon\mathbf{P}_k\right]$. This lemma also states that the absolute value of the eigenvalues of this transition matrix are bounded between terms that are expressed up to $K\varepsilon$ precision if $K\varepsilon \ll 1$ and ε is upper bounded by the value provided in the lemma. This result is extremely useful in establishing that the matrix product given by $\prod_{k=0}^{K-1}\left[\mathbf{A}_k + \varepsilon\mathbf{P}_k\right]$ can be computed explicitly up to $K\varepsilon$ precision without trading off much on the accuracy of the radial vector $\mathbf{u}_K$.

REMARK 3.6 Notice that the matrix product $\prod_{k=0}^{K-1}\left[\mathbf{A}_k + \varepsilon\mathbf{P}_k\right]$ in this lemma is hard to compute where expansion of this product will generate K terms. The hardness lies in the fact that the higher order terms in ε appearing in the expansion do not simplify due to the fact that matrices $\mathbf{P}_k$ do not commute. Beyond first order the expansion of this matrix product cannot be simplified with ease. Therefore Lemma 3.4 is of utmost importance in the sense that it provides the conditions under which the the tail error generated by the first order approximation $\prod_{k=0}^{K-1}\left[\mathbf{A}_k + \varepsilon\mathbf{P}_k\right] \approx \prod_{k=0}^{K-1}\mathbf{A}_k + \varepsilon\sum_{r=0}^{K}(\Pi_{k=r+1}^{K}\mathbf{A}_r\mathbf{P}_r\Pi_{k=0}^{r-1}\mathbf{A}_r)$ remains bounded.

LEMMA 3.5 Given an initialization of the radial vector $\mathbf{u}_0$, at any iteration K such that $K = \mathscr{O}\left(\frac{1}{\varepsilon}\right)$ and $\varepsilon < \min\left\{\inf_{\|\mathbf{u}\|=1}\left(\limsup_{j\to\infty}\sqrt[j]{\frac{r_j(\mathbf{u})}{j!}}\right)^{-1}, \frac{2\delta(1-\alpha L)}{\alpha M(2Ln^2+\delta)} + \mathscr{O}(\varepsilon^2)\right\}$ when $\alpha \in \left(0, \frac{1}{L} - \mathscr{O}(\varepsilon)\right]$ or $\varepsilon < \min\left\{\inf_{\|\mathbf{u}\|=1}\left(\limsup_{j\to\infty}\sqrt[j]{\frac{r_j(\mathbf{u})}{j!}}\right)^{-1}, \frac{2L\delta}{M(2Ln^2-\delta)} + \mathscr{O}(\varepsilon^2)\right\}$ when $\alpha \in \left(\frac{1}{L} - \mathscr{O}(\varepsilon), \frac{1}{L}\right]$, the radial

vector $\mathbf{u}_K$ can be approximately given as

$$\begin{aligned}\mathbf{u}_K \approx \tilde{\mathbf{u}}_K = \varepsilon \sum_{i\in\mathscr{N}_S}\Bigg(&\prod_{k=0}^{K-1} c_i^s(k)\theta_i^s + \sum_{l\in\mathscr{N}_S}\sum_{r=0}^{K-1}\prod_{k=0}^{r-1} c_i^s(k)d_{i,l}(r)\prod_{k=r+1}^{K-1} c_l^s(k)\theta_l^s \\ &+ \sum_{l\in\mathscr{N}_{US}}\sum_{r=0}^{K-1}\prod_{k=0}^{r-1} c_i^s(k)d_{i,l}(r)\prod_{k=r+1}^{K-1} c_l^{us}(k)\theta_l^{us}\Bigg)\mathbf{v}_i(0) + \\ \varepsilon \sum_{j\in\mathscr{N}_{US}}\Bigg(&\prod_{k=0}^{K-1} c_j^{us}(k)\theta_j^{us} + \sum_{l\in\mathscr{N}_S}\sum_{r=0}^{K-1}\prod_{k=0}^{r-1} c_j^{us}(k)d_{j,l}(r)\prod_{k=r+1}^{K-1} c_l^s(k)\theta_l^s \\ &+ \sum_{l\in\mathscr{N}_{US}}\sum_{r=0}^{K-1}\prod_{k=0}^{r-1} c_j^{us}(k)d_{j,l}(r)\prod_{k=r+1}^{K-1} c_l^{us}(k)\theta_l^{us}\Bigg)\mathbf{v}_j(0), \end{aligned} \tag{3.41}$$

where $\varepsilon\theta_i^s = \langle \mathbf{u}_0, \mathbf{v}_i(0)\rangle$, $\varepsilon\theta_j^{us} = \langle \mathbf{u}_0, \mathbf{v}_j(0)\rangle$ and we have that

$$\mathbf{u}_0 = \varepsilon \sum_{i\in\mathscr{N}_S} \theta_i^s \mathbf{v}_i(0) + \varepsilon \sum_{j\in\mathscr{N}_{US}} \theta_j^{us}\mathbf{v}_j(0) \tag{3.42}$$

with $\theta_i^s \geqslant 0$, $\theta_j^{us} \geqslant 0$ for all i, j. The coefficient terms $c_i^s(k)$, $c_j^{us}(k)$, $d_{i,l}(k)$, $d_{l,i}(k)$ are the same as in Lemma 3.4.

The proof of this lemma is given in Appendix C. The approximation $\tilde{\mathbf{u}}_K$ in this lemma for the radial vector $\mathbf{u}_K$ is generated by explicitly computing the matrix product $\prod_{k=0}^{K-1}\left[\mathbf{A}_k + \varepsilon\mathbf{P}_k\right]$ from Lemma 3.4 up to first order in ε. Also note that the non-negativity of θ_i^s and θ_j^{us} here can be assumed without loss of generality.

REMARK 3.7 The conditions $\varepsilon < \min\left\{\inf_{\|\mathbf{u}\|=1}\left(\limsup_{j\to\infty}\sqrt[j]{\frac{r_j(\mathbf{u})}{j!}}\right)^{-1}, \frac{2\delta(1-\alpha L)}{\alpha M(2Ln^2+\delta)} + \mathscr{O}(\varepsilon^2)\right\}$ when $\alpha \in \left(0, \frac{1}{L} - \mathscr{O}(\varepsilon)\right]$ or $\varepsilon < \min\left\{\inf_{\|\mathbf{u}\|=1}\left(\limsup_{j\to\infty}\sqrt[j]{\frac{r_j(\mathbf{u})}{j!}}\right)^{-1}, \frac{2L\delta}{M(2Ln^2-\delta)} + \mathscr{O}(\varepsilon^2)\right\}$ when we have $\alpha \in \left(\frac{1}{L} - \mathscr{O}(\varepsilon), \frac{1}{L}\right]$ are necessary but may not be sufficient due to unavailability of the radius of convergence from the Rayleigh–Schrödinger perturbation analysis. Also note that here $r_j(\mathbf{u})$ has the same definition as in Lemma 3.3.

In words, this lemma states that the radial vector $\mathbf{u}_K$ can be expressed by explicitly computing the matrix product $\prod_{k=0}^{K-1}\left[\mathbf{A}_k + \varepsilon\mathbf{P}_k\right]$ from Lemma 3.4 to $K\varepsilon$ precision provided $K\varepsilon \ll 1$ and ε is bounded above. This approximate solution represented by $\tilde{\mathbf{u}}_K$ generates the trajectory $\{\tilde{\mathbf{u}}_K\}_{K=1}^{K_{exit}}$, which we refer to as the ε-precision trajectory.

REMARK 3.8 Notice that from (3.41) we obtain a closed form expression for the ε precision trajectory inside $\mathscr{B}_\varepsilon(\mathbf{x}^*)$ for some initialization $\mathbf{u}_0$. However the solution is not unique due to the fact that the coefficients $c_i^s(k), c_j^{us}(k), d_{l,i}(k)$ from Lemma 3.4 are known only up to an interval. This is due to the fact that the eigenvalues $\lambda_i(0), \lambda_j(0)$ are known up to an interval. Hence we will obtain a **family of ε precision trajectories** from the expression of $\tilde{\mathbf{u}}_K$. The next lemma provides a handle on the exit times for such a family of approximate trajectories.

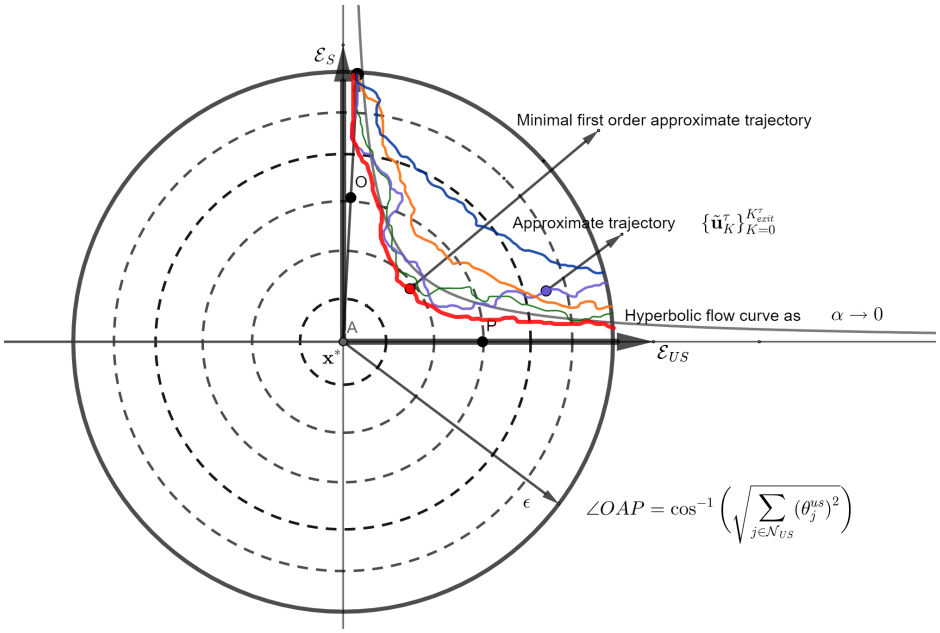


FIG. 3: A 2-D representation of the approximate trajectories, where every approximate trajectory has its own exit time and the *minimal* approximate trajectory is the one which has the largest exit time. The initial radial vector subtends a very large angle of $\angle OAP$ ($0 \ll \angle OAP < \frac{\pi}{2}$) from the unstable subspace, where the initial unstable projection is given by $\sum_{j\in\mathcal{N}_{US}} (\theta_j^{us})^2$.

LEMMA 3.6 Let $S_\varepsilon = \left\{ \{\tilde{\mathbf{u}}_K^\tau\}_{K=1}^{K_{exit}^\tau} \middle| \mathbf{u}_0 \right\}$ be the set of all possible τ-parameterized ε-precision trajectories generated by the approximate equation (3.41) in Lemma 3.5, where the parameter $\tau \in \mathbb{R}$ varies with variations in the sequence $\left\{ \{c_i^s(k), c_j^{us}(k), d_{l,i}(k)\}_{k=0}^{K-1} \right\}_{K=1}^{K_{exit}}$. Let K_{exit}^τ be the exit time of the τ-parameterized trajectory $\{\tilde{\mathbf{u}}_K^\tau\}_{K=1}^{K_{exit}^\tau}$ from the ball $\mathscr{B}_\varepsilon(\mathbf{x}^*)$ where we have that

$$K_{exit}^\tau = \inf_{K\geqslant 1} \left\{ K \;\middle|\; \|\tilde{\mathbf{u}}_K^\tau\|^2 > \varepsilon^2 \right\}. \tag{3.43}$$

Formally, $\tilde{\mathbf{u}}_K^\tau$ is a possible solution to the equation (3.41) in $\tilde{\mathbf{u}}_K$, where $1 \leqslant K \leqslant K_{exit}^\tau$ and $\tilde{\mathbf{u}}_K$ varies with variations in the sequence $\{c_i^s(k), c_j^{us}(k), d_{l,i}(k)\}_{k=0}^{K-1}$.

Let K^ι be the exit time of the infimum over all possible τ-parameterized trajectories, where infimum is taken with respect to the squared radial distance $\|\tilde{\mathbf{u}}_K^\tau\|^2$. This K^ι can be defined as

$$K^\iota = \inf_{K\geqslant 1} \left\{ K \;\middle|\; \inf_\tau \left\{ \|\tilde{\mathbf{u}}_K^\tau\|^2 \right\} > \varepsilon^2 \right\}. \tag{3.44}$$

Then we have the following condition:

$$K^{\iota} \geqslant \sup_{\tau}\left\{K^{\tau}_{exit}\right\} = \sup_{\tau}\inf_{K\geqslant 1}\left\{K \;\middle|\; \|\tilde{\mathbf{u}}^{\tau}_{K}\|^2 > \varepsilon^2\right\}. \tag{3.45}$$

The proof of this lemma is given in Appendix D. This particular lemma states an important result about the exit time K^{ι} of the trajectory generated by selecting that approximate vector $\tilde{\mathbf{u}}^{\tau}_{K}$ from all possible τ that has the minimum radial distance from $\mathbf{x}^*$ at each K. It claims that this minimal trajectory has the maximum exit time from $\mathscr{B}_{\varepsilon}(\mathbf{x}^*)$. Though seemingly trivial, this result is extremely important in proving the worst-case exit time for trajectories with linear escape rates. A representation of the family of approximate trajectories along with the constructed minimal approximate trajectory is provided in Figure 3.

THEOREM 3.2 For every value of the parameter τ, there exists a lower bound on the squared radial distance $\|\tilde{\mathbf{u}}^{\tau}_{K}\|^2$ for all K in the range $1 \leqslant K \leqslant \sup_{\tau}\left\{K^{\tau}_{exit}\right\}$ provided $K\varepsilon \ll 1$. Moreover, this lower bound can be expressed using a function of K called the trajectory function $\Psi(K)$. Formally, for $1 \leqslant K < \sup_{\tau}\left\{K^{\tau}_{exit}\right\}$ we have that

$$\varepsilon^2 \geqslant \inf_{\tau}\|\tilde{\mathbf{u}}^{\tau}_{K}\|^2 > \varepsilon^2\Psi(K), \tag{3.46}$$

where the trajectory function $\Psi(K)$ is defined as follows:

$$\Psi(K) = \left(c_1^{2K} - 2Kc_2^{2K-1}b_1 - b_2c_3^Kc_2^K - b_2c_3^{2K}\right)\sum_{i\in\mathscr{N}_S}(\theta_i^s)^2 + \left(c_4^{2K} - 2Kc_3^{2K-1}b_1 - b_2c_3^Kc_2^K - b_2c_3^{2K}\right)\sum_{j\in\mathscr{N}_{US}}(\theta_j^{us})^2 \tag{3.47}$$

with $c_1 = \left(1-\alpha L - \frac{\alpha\varepsilon M}{2} - \mathscr{O}(\varepsilon^2)\right)$, $c_2 = \left(1-\alpha\beta + \frac{\alpha\varepsilon M}{2} + \mathscr{O}(\varepsilon^2)\right)$, $c_3 = \left(1+\alpha L + \frac{\alpha\varepsilon M}{2} + \mathscr{O}(\varepsilon^2)\right)$, $c_4 = \left(1+\alpha\beta - \frac{\alpha\varepsilon M}{2} - \mathscr{O}(\varepsilon^2)\right)$, $b_1 = \left(\frac{\alpha\varepsilon MLn}{2\delta} + \mathscr{O}(\varepsilon^2)\right)$, $b_2 = \frac{\left(\frac{\alpha\varepsilon MLn}{2\delta}+\mathscr{O}(\varepsilon^2)\right)\left(1+\mathscr{O}(K\varepsilon)\right)}{\left(\alpha L+\alpha\beta+\mathscr{O}(\varepsilon^2)\right)}$ and δ is defined in Proposition 2.2.

We also require that $\varepsilon < \min\left\{\inf_{\|\mathbf{u}\|=1}\left(\limsup_{j\to\infty}\sqrt[j]{\frac{r_j(\mathbf{u})}{j!}}\right)^{-1}, \frac{2\delta(1-\alpha L)}{\alpha M(2Ln^2+\delta)} + \mathscr{O}(\varepsilon^2)\right\}$ when $\alpha \in \left(0, \frac{1}{L} - \mathscr{O}(\varepsilon)\right]$, while $\varepsilon < \min\left\{\inf_{\|\mathbf{u}\|=1}\left(\limsup_{j\to\infty}\sqrt[j]{\frac{r_j(\mathbf{u})}{j!}}\right)^{-1}, \frac{2L\delta}{M(2Ln^2-\delta)} + \mathscr{O}(\varepsilon^2)\right\}$ when we have $\alpha \in \left(\frac{1}{L} - \mathscr{O}(\varepsilon), \frac{1}{L}\right]$.

The proof of this theorem is given in Appendix D. Theorem 3.2 states that for a given initialization $\mathbf{u}_0$, all the possible ε-precision trajectories generated have their radial distance from $\mathbf{x}^*$ lower bounded using some function $\Psi(K)$. Now this $\Psi(K)$ can be used to determine K^{ι} and hence K_{exit} for any choice of the step size α.

REMARK 3.9 Notice that the trajectory function $\Psi(K)$ corresponds to the minimal approximate trajectory $\inf_\tau \|\tilde{\mathbf{u}}_K^\tau\|$. Now K^ι can be obtained by solving the condition $K^\iota = \inf_{\substack{K\geqslant 1 \\ K=\mathscr{O}(\log(\varepsilon^{-1}))}} \{K \mid \Psi(K) > 1\}$. The condition $K = \mathscr{O}(\log(\varepsilon^{-1}))$ ensures linear time solutions, which are the only solutions of interest to the problem. Then from Lemma 3.6 we will have $K_{exit} < K^\iota = \mathscr{O}(\log(\varepsilon^{-1}))$. It is worth reminding the reader here that the $\mathscr{O}(\varepsilon^2)$ terms in the theorem statement account for the approximation in analysis and things are proved about trajectories and perturbations up to first order in ε.

Observe that in the expression for the trajectory function $\Psi(K)$, the term accompanying $\sum_{i\in\mathscr{N}_S}(\theta_i^s)^2$ is $\left(c_1^{2K} - 2Kc_2^{2K-1}b_1 - b_2c_3^Kc_2^K - b_2c_3^{2K}\right)$, which is a decreasing function of K since $c_1 < 1$. Moreover, the rate of decrease of the term $\left(c_1^{2K} - 2Kc_2^{2K-1}b_1 - b_2c_3^Kc_2^K - b_2c_3^{2K}\right)$ for small values of K is governed by c_1 and not by c_3, where $c_3 > 1$ due to the fact that b_1, b_2 are of order $\mathscr{O}(\varepsilon)$ and so $-2Kc_2^{2K-1}b_1$, $-b_2c_3^{2K}$ will not decrease as fast as c_1^{2K} for small enough K since we assumed $K\varepsilon \ll 1$. Next, by a similar argument the term accompanying $\sum_{j\in\mathscr{N}_{US}}(\theta_j^{us})^2$ given by $\left(c_4^{2K} - 2Kc_3^{2K-1}b_1 - b_2c_3^Kc_2^K - b_2c_3^{2K}\right)$ is an increasing function of K for $K\varepsilon \ll 1$ since $c_4 > 1$ and so c_4^{2K} dominates the term $2Kc_3^{2K-1}b_1 + b_2c_3^Kc_2^K + b_2c_3^{2K}$. Also notice that $\Psi(K) < 1$ at $K = 0$ since $\sum_{i\in\mathscr{N}_S}(\theta_i^s)^2 + \sum_{j\in\mathscr{N}_{US}}(\theta_j^{us})^2 = 1$. Therefore, provided the initial unstable subspace projection $\sum_{j\in\mathscr{N}_{US}}(\theta_j^{us})^2$ is not too small, the trajectory function $\Psi(K)$ first increases for small K, where $K\varepsilon \ll 1$, and then decreases to $-\infty$. Then for some small K if $\Psi(K) > 1$, we are guaranteed that the minimal approximate trajectory $\inf_\tau \|\tilde{\mathbf{u}}_K^\tau\|$ escapes $\mathscr{B}_\varepsilon(\mathbf{x}^*)$. Section 4.1 simulates the evolution of the trajectory function $\Psi(K)$ on the phase retrieval problem, which corroborates this theoretical understanding.

Before moving on to the next theorem, we introduce the notion of conditioning of a function. The condition number at the stationary point of a non-convex function is given by the ratio of the largest absolute eigenvalue to the smallest absolute eigenvalue of the Hessian of the function at that point. Also, a function is called perfectly conditioned if the condition number is equal to 1. In the current problem setting, the condition number of the function $f(\cdot)$ at the saddle point $\mathbf{x}^*$ is given by $\frac{L}{\beta}$. Now, the function $f(\cdot)$ is well-conditioned if the condition number $\frac{L}{\beta}$ is not arbitrarily large or equivalently $\frac{\beta}{L}$ is bounded away from 0.

THEOREM 3.3 For the gradient update equation with the step size $\alpha = \frac{1}{L}$, there exists a minimum projection Δ of the radial vector initialization $\mathbf{u}_0$ on the unstable subspace $\mathscr{E}_{US}$ such that whenever $\sum_{j\in\mathscr{N}_{US}}(\theta_j^{us})^2 \geqslant \Delta$, where $\mathbf{u}_0 + \mathbf{x}^* \in \bar{\mathscr{B}}_\varepsilon(\mathbf{x}^*)\backslash\mathscr{B}_\varepsilon(\mathbf{x}^*)$, $\mathbf{u}_0 = \varepsilon\sum_{i\in\mathscr{N}_S}\theta_i^s\mathbf{v}_i(0) + \varepsilon\sum_{j\in\mathscr{N}_{US}}\theta_j^{us}\mathbf{v}_j(0)$, the ε-precision trajectories $\{\tilde{\mathbf{u}}_K\}_{K=1}^{K_{exit}}$ can exit $\mathscr{B}_\varepsilon(\mathbf{x}^*)$ in linear time. Moreover their exit time K_{exit} from the ball $\mathscr{B}_\varepsilon(\mathbf{x}^*)$ is approximately upper bounded as follows:

$$K_{exit} < K^\iota \lessapprox \frac{\log\left(\left(2+\frac{\varepsilon M}{2L}\right)\log\left(\frac{2+\frac{\varepsilon M}{2L}}{1+\frac{\beta}{L}-\frac{\varepsilon M}{2L}}\right)\frac{2\delta}{\varepsilon Mn}\right)}{2\log\left(\frac{2+\frac{\varepsilon M}{2L}}{1+\frac{\beta}{L}-\frac{\varepsilon M}{2L}}\right)}, \tag{3.48}$$

where $\varepsilon < \min\left\{\inf_{\|\mathbf{u}\|=1}\left(\limsup_{j\to\infty}\sqrt[j]{\frac{r_j(\mathbf{u})}{j!}}\right)^{-1}, \frac{2L\delta}{M(2Ln^2-\delta)} + \mathscr{O}(\varepsilon^2), \frac{2\beta}{M}\right\}$ and we must necessarily have that $\Delta > \varepsilon\frac{MLn}{\delta(L+\beta)}$ with δ defined in Proposition 2.2.

The proof of this theorem is given in Appendix E. In terms of the order notation, we have $K_{exit} \lessapprox \mathscr{O}\left(\log\left(\frac{\delta}{\varepsilon n}\right)\right)$ and the initial unstable subspace projection satisfies $\sum_{j\in\mathscr{N}_{US}}(\theta_j^{us})^2 \geqslant \Delta > \Omega\left(\frac{\varepsilon n}{\delta}\right)$.

REMARK 3.10 This theorem guarantees the existence of ε-precision trajectories with linear exit time and gives an upper bound on their exit time K_{exit} from the ball $\mathscr{B}_\varepsilon(\mathbf{x}^*)$. However, the sufficient conditions that guarantee the existence of this exit time $K_{exit} \lessapprox \mathscr{O}\left(\log\left(\frac{1}{\varepsilon}\right)\right)$ depend on the quantity Δ. Note that the condition $\Delta > \varepsilon\frac{MLn}{\delta(L+\beta)}$ is necessary for the existence of order $\mathscr{O}\left(\log\left(\frac{1}{\varepsilon}\right)\right)$ solution of K^t but not sufficient. Since this work only deals with the existence of linear exit time solutions, we refrain from developing tighter lower bounds on Δ. Obtaining such sufficient conditions requires a more rigorous analysis of the trajectory function $\Psi(K)$, which is beyond the scope of current work. In particular, our followup work [11] derives one such sufficient condition.

REMARK 3.11 Observe that the bound on the exit time from Theorem 3.3 depends on quantities like the Lipschitz parameters, condition number, problem dimension and the eigen gap. However for structured problems such as those in [7], one can leverage the specialized function geometry and obtain rates of convergence independent of these parameters. But in the absence of any other assumption on the function, and since we are dealing with a much larger function class, i.e., the class of Morse functions, these parameters become necessary to evaluate the escape rates. In order to better understand the utility of these local Lipschitz parameters in the derivation of our results for the general (as opposed to the specialized) non-convex functions, observe that the local Hessian Lipschitz parameter M is required to bound $\|\mathbf{H}(\hat{\mathbf{u}}_k)\|_2$, where $\mathbf{H}(\hat{\mathbf{u}}_k)$ is used to determine the Hessian at any point $\mathbf{x}_k \in \mathscr{B}_\varepsilon(\mathbf{x}^*)$ from Lemma 3.3. Next, the local gradient Lipschitz parameter L controls the coefficient terms $c_i^s(k)$, $c_j^{us}(k)$, $d_{i,l}(k)$ from Lemmas 3.4 and 3.5, where these terms depend on the eigenvalues of $\nabla^2 f(\mathbf{x}^*)$, the difference between these eigenvalues, and the matrix $\mathbf{H}(\hat{\mathbf{u}}_k)$, which comes from Lemma 3.3. Since these coefficient terms determine the expression for the approximate gradient trajectory in Lemma 3.5, one cannot generate a closed-form expression of the approximate gradient trajectory in the absence of the gradient Lipschitz parameter. Finally, the minimal approximate trajectory function from Theorem 3.2 relies on the precise bounds for these coefficients. Without the gradient Lipschitz parameter, the eigenvalues of $\nabla^2 f(\mathbf{x}^*)$ cannot be bounded and similarly without the Hessian Lipschitz parameter one cannot obtain an upper bound on $\|\mathbf{H}(\hat{\mathbf{u}}_k)\|_2$.

Theorem 3.3 guarantees a linear exit time bound from the ball $\mathscr{B}_\varepsilon(\mathbf{x}^*)$ for ε-precision trajectories under some necessary initial conditions on $\mathbf{x}_0$. The necessary condition of $\Delta > \varepsilon\frac{MLn}{\delta(L+\beta)}$ requires that the initial radial vector is not aligned too much with the stable subspace of the Hessian $\nabla^2 f(\mathbf{x}^*)$ and has some order $\Omega(\varepsilon)$ alignment with the unstable subspace so as to facilitate the linear time escape. It should be noted that this necessary condition of $\Delta > \varepsilon\frac{MLn}{\delta(L+\beta)}$ is sufficient to claim that these gradient descent trajectories for $\alpha < \frac{1}{L}$ will almost surely not terminate into the strict saddle point $\mathbf{x}^*$ from the following Lemma 3.7.

LEMMA 3.7 The discrete gradient trajectories for $\alpha < \frac{1}{L}$ ending into the first-order strict saddle point $\mathbf{x}^*$ have zero Lebesgue measure with respect to the space $\mathscr{E}$ and are referred to as trivial trajectories. This result can be established using the stable center manifold theorem from [40].

We refer the reader to [24] for a proof of this lemma. Note that the assumption on the step size $\alpha < \frac{1}{L}$

in Lemma 3.7 is necessary since the zero measure result can only be developed when the map $G : \mathbf{x}_k \mapsto \mathbf{x}_{k+1}$, where $\mathbf{x}_{k+1} = \mathbf{x}_k - \alpha\nabla f(\mathbf{x}_k) =: G(\mathbf{x}_k)$, is a diffeomorphism (or is at least locally bi-Lipschitz). A crucial step in [24] where this diffeomorphism property is utilized involves pulling back measure zero sets under the diffeomorphism G to again get measure zero sets. However for the case of $\alpha = \frac{1}{L}$, the map $G : \mathbf{x}_k \mapsto \mathbf{x}_{k+1}$ fails to be a diffeomorphism (or even locally bi-Lipschitz); see details in [24].

We also note that the condition of minimal non-zero projection of the initial point on the unstable subspace of $\nabla^2 f(\mathbf{x}^*)$ from Theorem 3.3, given by the bound $\Delta > \varepsilon\frac{MLn}{\delta(L+\beta)}$, is tight. Moreover, this necessary condition does not contradict any existing results regarding the almost sure non-convergence of randomly initialized gradient descent to strict saddle points. Further, recall that the gradient descent method may get stuck at the saddle point for a particular set of initializations. In Theorem 3.3, however, we provide a condition on the initialization that ensures its exclusion from such a set. This condition, which is one of the major contributions of this work, requires the projection of the initial point on the unstable subspace of the Hessian $\nabla^2 f(\mathbf{x}^*)$ at the saddle point $\mathbf{x}^*$ to be at least on the order of $\Omega(\varepsilon)$. Take, for instance, a specific example of the strict saddle Morse function $f(x,y) = x^2 - y^2$ with the initialization scheme of $(x_0, 0)$ for any $x_0 \in \mathbb{R}$. Under this given initialization scheme, the gradient descent method will eventually get stuck at the origin, which is a strict saddle point. However, since the initialization point completely lies in the stable subspace of $\nabla^2 f(0,0)$, which is $\mathsf{span}\{(1,0)\}$, it has a null projection on the unstable subspace of $\nabla^2 f(0,0)$, which is $\mathsf{span}\{(0,1)\}$. Therefore, this example violates the minimal projection condition of Theorem 3.3 and does not affect the validity of our claims.

3.6 *Comparison with the exit time bound from Section 3.2*

It can be seen from Theorem 3.3 that the exit time bound for the approximate trajectory and the necessary initial condition using the matrix perturbation-based analysis depend on quantities like the inverse of the condition number $\frac{\beta}{L}$, minimum eigenvalue gap δ, function's dimension n and the size of the saddle neighborhood ε. In contrast, the rudimentary analysis in Section 3.2 does not bring out the dependence of the exit time bound and the initial boundary condition on these key problem parameters. Moreover, the analysis developed in Section 3.2 leaves more open questions by introducing unknown parameters like ρ and $\gamma_{K_{exit}}$, where $\gamma_{K_{exit}}$ could be arbitrarily small and the presence of ρ in the boundary condition makes the analysis from Section 3.2 more restrictive than the analysis presented in Section 3.5 where matrix perturbation theory is used. The main reason for this difference between the results of Section 3.2 and those of Theorem 3.3 is that, by restricting the class of functions from $\mathscr{C}^2$ to real analytic, we are able to develop tight approximations to discrete trajectories using the matrix perturbation theory that lead to precise expressions for the exit time bound and the initial boundary condition that depend on the key problem parameters. These differences between the rudimentary analytical approach of Section 3.2 and the matrix perturbation-based approach of Section 3.5 are also summarized in Table 2. Notice that there is a cross (✗) marked against the 'Closed form expression for the trajectory' in Table 2 in the column corresponding to the analysis of Section 3.2. This is because although (3.18) provides an expression for the trajectory inside the ball $\mathscr{B}_\varepsilon(\mathbf{x}^*)$, its exact closed form cannot be determined due to the fact that we only have information on $\|r(\mathbf{x}_k)\|$ in Section 3.2. In contrast, the same $r(\mathbf{x}_k)$ is known up to first-order precision in Section 3.5 and therefore a closed-form expression for the ε-precision trajectory is available from Lemma 3.5.

Assumptions / Techniques / Metrics	**Exit Time Analysis from Section 3.2**	**Exit Time Analysis from Section 3.5**
Function class	$\mathscr{C}^2$ Morse functions	locally $\mathscr{C}^\omega$ Morse functions
Proof techniques	Sequential monotonicity of the unstable subspace projection	Matrix perturbation theory and approximation theory
Key metrics	Saddle neighborhood's radius ε, unknown factors $\gamma_{\mathcal{K}_{exit}}, \rho$	Saddle neighborhood's radius ε, dimension n and eigenvalue gap δ
Closed-form expression for the trajectory / approximate trajectory inside $\mathscr{B}_\varepsilon(\mathbf{x}^*)$	✗	✓
Constraints on the set of trajectories / approximate trajectories analyzed	Gradient trajectories for which $\langle \mathbf{v}_n, \mathbf{x}_k - \mathbf{x}^* \rangle$ increases monotonically with k	No constraints
Linear exit time bound	$\mathscr{O}\left(\log\left(\frac{\gamma_{\mathcal{K}_{exit}}(1-\rho)}{\varepsilon}\right)\right)$	$\mathscr{O}\left(\log\left(\frac{\delta}{\varepsilon n}\right)\right)$
Nature of the exit time bound	Exact	Approximate
Initial boundary conditions	$\langle \mathbf{v}_0, \mathbf{x}_0 - \mathbf{x}^* \rangle \geqslant \Omega\left(\frac{\varepsilon^2}{1-\rho}\right)$	$\sum_{j \in \mathcal{N}_{US}} (\theta_j^{us})^2 \geqslant \Delta > \Omega\left(\frac{\varepsilon n}{\delta}\right)$
Bounds on ε	✗	✓

Table 2: Comparison of the exit time analyses that follow from existing analytical techniques (Section 3.2) and the novel matrix perturbation-based analytical approach of Section 3.5.

4. Numerical results

To support the theoretical framework developed in this work and showcase the effectiveness of gradient trajectories with large initial unstable projections in escaping from strict saddle neighborhoods, we evaluate the performance of the gradient descent method on the phase retrieval problem [5]. Briefly, the phase retrieval problem formulation is given by

$$\min_{\mathbf{x} \in \mathbb{R}^n} f(\mathbf{x}) = \frac{1}{4m} \sum_{j=1}^{m} \left[\langle \mathbf{a}_j, \mathbf{x} \rangle^2 - y_j \right]^2, \tag{4.1}$$

where the y_j's are known observations and the $\mathbf{a}_j$'s are independent and identically distributed (i.i.d.) random vectors whose entries are generated from a normal distribution. Note that the variable 'm' here in (4.1) should not be confused with the number of eigenvalue groups 'm' defined in proposition 2.2. The formulation in (4.1) is the least-squares problem reformulation for the Short-Time Fourier Transform (STFT) of the actual phase retrieval problem (see [16]). Moreover, the above least-squares reformulation of the original phase retrieval problem can also be found in recent works like [27], which highlight the efficacy of simple gradient descent method on structured non-convex functions. Clearly, the function in (4.1) satisfies Assumption **A1** and also Assumptions **A2** and **A3** locally in every compact set.

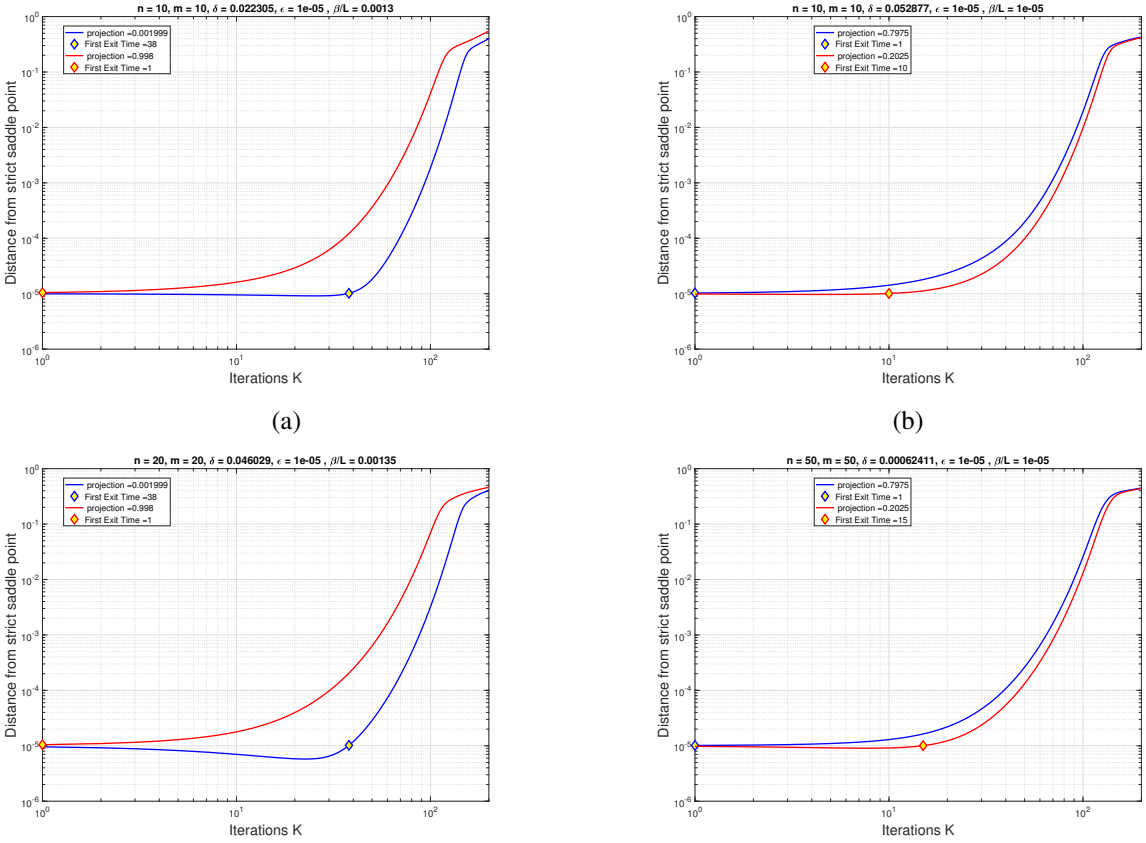


FIG. 4: Simulating gradient trajectories on the phase retrieval problem with $\alpha = 0.1/L$ under certain initial unstable projections for various values of m, n and ϵ.

In the simulations, we set $y_j = 1$ for $1 \leqslant j \leqslant \lfloor \frac{m}{2} \rfloor$ and $y_j = -1$ otherwise. Also, for the sake of simplicity we always set $m = n$ so that the system of equations $y_j = \langle \mathbf{a}_j, \mathbf{x} \rangle^2$ is neither under determined nor over determined and the Hessian of the function $f(\cdot)$ is full rank. The i.i.d. nature of the $\mathbf{a}_j$'s thus implies that the parameter $\frac{\beta}{L}$ is not too small and therefore Assumption **A4** gets satisfied. The closed-form expressions for the gradient and the Hessian of the function in (4.1) are, respectively, as follows:

$$\nabla f(\mathbf{x}) = \frac{1}{m} \sum_{j=1}^{m} \Big(\langle \mathbf{a}_j, \mathbf{x} \rangle^2 - y_j \Big) \langle \mathbf{a}_j, \mathbf{x} \rangle \mathbf{a}_j, \quad \text{and} \tag{4.2}$$

$$\nabla^2 f(\mathbf{x}) = \frac{1}{m} \sum_{j=1}^{m} \Big(3\langle \mathbf{a}_j, \mathbf{x} \rangle^2 - y_j \Big) \mathbf{a}_j \mathbf{a}_j^T. \tag{4.3}$$

For the particular choice of y_j's it is observed that $\mathbf{x}^* = \mathbf{0}$ is a strict saddle point. We now initialize the

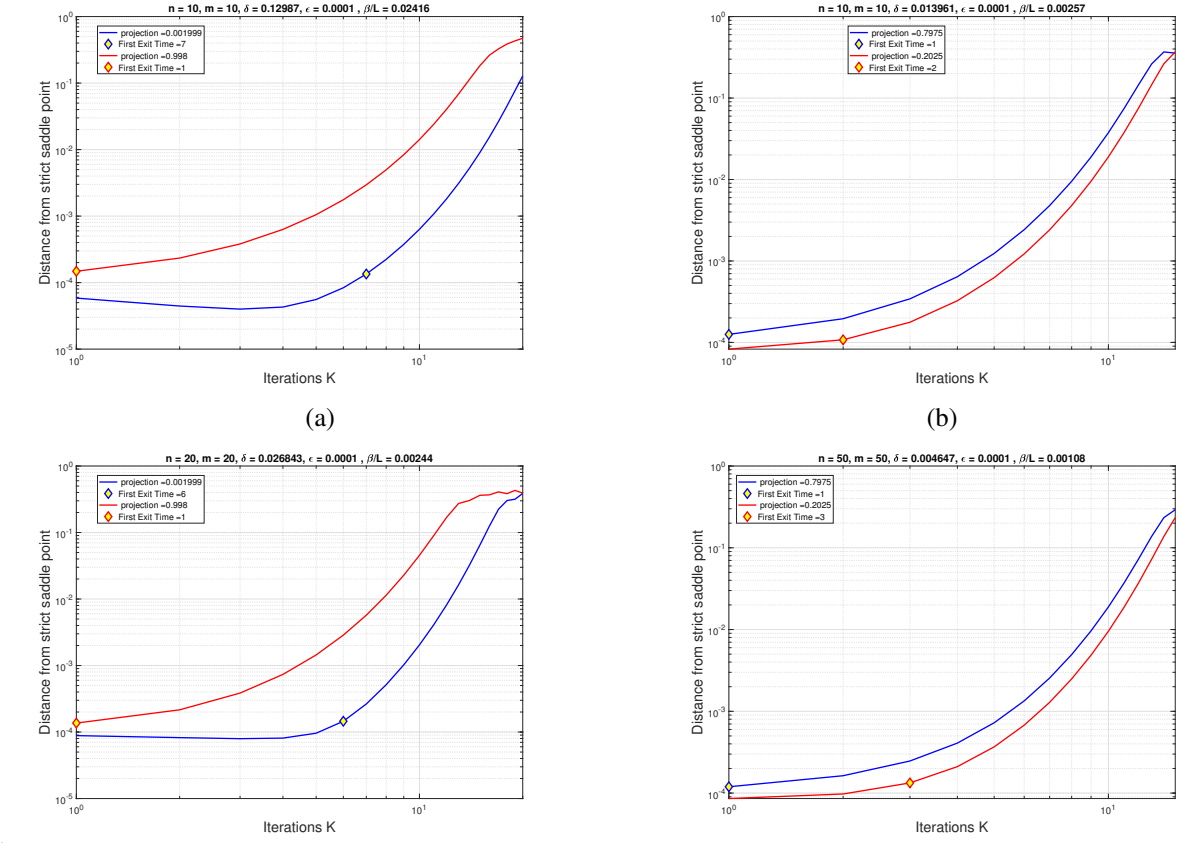


FIG. 5: Simulating gradient trajectories on the phase retrieval problem with $\alpha = 1/L$ under certain initial unstable projections for various values of m, n and ε.

gradient descent method in the ε-neighborhood of $\mathbf{x}^*$ and examine the exit-time behavior of its trajectories for different values of n, m, ε, and the 'projection' of the initial iterate on the unstable subspace, which corresponds to the quantity $\sum_{j \in \mathcal{N}_{US}} (\theta_j^{us})^2$. The results are reported in Figure 4 for the step size of $\alpha = 0.1/L$ and in Figure 5 for the step size of $\alpha = 1/L$, with L being the largest eigenvalue of $\nabla^2 f(\mathbf{x}^*)$. Note that each subplot in both of the figures corresponds to different random $\mathbf{a}_j$'s. In order to highlight the dependence of the exit time on the unstable projection, we compare two different initializations of the gradient descent method for the same set of problem parameters in terms of the radial distance of the respective generated trajectories from the saddle point. Also the "first exit time" (the iteration when the gradient trajectory exits $\mathscr{B}_\varepsilon(\mathbf{x}^*)$ for the first time) from the saddle neighborhood for the two trajectories are marked on each of the curves in colors matching with their respective radial distance curves.

It is evident from the two figures that, as suggested by the theoretical developments in this paper, a larger initial unstable subspace projection results in a faster exit time. More importantly, Figure 5 corroborates our findings from Theorem 3.3 that for the step size of $\frac{1}{L}$, even with very small initial

unstable subspace projections, i.e., $\sum_{j\in\mathcal{N}_{US}}(\theta_j^{us})^2 = \mathscr{O}(\varepsilon)$ such as those in Figure 5(a) and Figure 5(b), faster exit times are possible. Such conclusion does not necessarily hold for small step size, as in Figure 4(a) and Figure 4(b), where small initial unstable subspace projections yield relatively larger exit times.

We next illustrate the dependence of the exit time estimate on the dimension n and eigen gap δ. We first develop a numerical setup to showcase the dependence on δ. To give an idea of our experimental setup, below is a step-by-step methodology used to perform simulations:

1. Suppose the Hessian of function $f(\cdot)$ for the phase retrieval problem (4.1) has three distinct groups of eigenvalues,[8] where the eigenvalues within any group are identical such that one group has eigenvalues equal to the gradient Lipschitz constant L (as before, L is the largest eigenvalue of $\nabla^2 f(\mathbf{x}^*)$, where $\mathbf{x}^*$ is the strict saddle point), the other group has eigenvalues equal to $-\beta$, and the third group is placed on the eigenvalue spectrum so that it is at a δ distance from one of these groups. Further, suppose the third eigenvalue group has eigenvalues $-\beta+\delta$ where $(L+\beta)/2 > \delta > 2\beta$. This construction preserves the parameters L,β from Assumptions **A2**, **A4** for the function $f(\cdot)$ as the eigen gap δ is varied. Though the Hessian Lipschitz parameter M for the function $f(\cdot)$ may not be preserved by this construction,[9] yet this setup is still able to control a given maximum number of parameters, i.e., L,β and the problem dimension n.

2. Next, we set $m = n = 100$ in the phase retrieval problem (4.1), where $\mathbf{x}\in\mathbb{R}^n$, $\mathbf{a}_j$'s are taken to be the canonical basis of $\mathbb{R}^n$, and the eigen gap δ varies in the range $[0.15, 2.13]$. Using the setup described in the previous bullet point, we then set the y_j's as follows:

$$y_j = \begin{cases} \frac{m}{20} \ ; & 1 \leqslant j \leqslant \lfloor \frac{m}{3} \rfloor \\ \frac{m}{20} - m\delta \ ; & \lfloor \frac{m}{3} \rfloor + 1 \leqslant j \leqslant 2\lfloor \frac{m}{3} \rfloor \\ -5m \ ; & \textit{otherwise}. \end{cases} \tag{4.4}$$

3. Since the $\mathbf{a}_j$'s are orthonormal, it can be readily checked using (4.3) that the eigenvalues of $\nabla^2 f(\mathbf{x})$ at $\mathbf{x} = \mathbf{0}$ are $-y_j/m$ and we have $L = 5$, $\beta = 1/20$ from the given choice of y_j's. By the above choice of y_j's, the eigenvalues belong to three distinct groups and $\mathbf{x} = \mathbf{0}$ is a strict saddle point. In particular, the choice $y_j = \frac{m}{20} - m\delta$ from above corresponds to the case where the free eigenvalue group has eigenvalues equal to $(\delta - \frac{1}{20})$.

4. Finally, for the eigen gap δ in the range $[0.15, 2.13]$, we compute the exit time from ε-neighborhood of the origin for different values of the initial unstable subspace projections.

The results for this numerical setup are plotted in Figure 6 for two values of the initial unstable subspace projections for $\alpha = 0.1/L$, where we have displayed the exit time versus δ on the logarithmic scale. We observe from the figure that the exit time increases with increasing eigen gap δ at least initially, which agrees with Theorem 3.3 where we have $K_{exit} \lessapprox \mathscr{O}(\log\delta)$.

Next, we illustrate the dependence of the exit time on the problem dimension. Note that in general as the dimension n increases, the gradient as well as the Hessian Lipschitz parameters (L,M) increase. In particular, we have $L = \Theta(n)$, $M = \Theta(n)$ (see the discussion within Section 3 of [7]). However, we

[8] We can introduce more groups of eigenvalues but refrain from doing so for the sake of simplicity.

[9] The Hessian Lipschitz parameter M may change but will remain bounded in every compact set and therefore will be upper bounded by a constant term in the ball $\mathscr{B}_\varepsilon(\mathbf{x}^*)$. Also, M will remain constant with respect to the dimension n since n is fixed here.

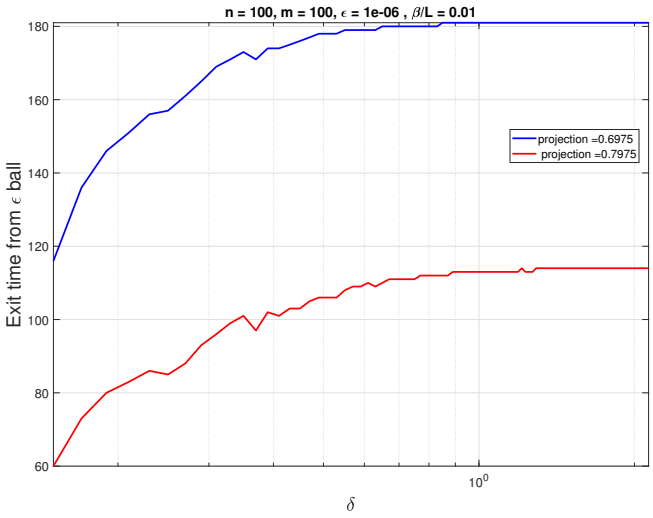


FIG. 6: Exit time versus the eigen gap δ (logarithmic scale) under certain initial unstable subspace projections for given values of n, L, β, and ε.

can showcase the dependence of the exit time on the problem dimension for very particular choice of functions by keeping the gradient and Hessian Lipschitz parameters fixed with respect to the order of dimension. To this end, we modify the cost function in the phase retrieval problem (4.1) by normalizing it with dimension n and rewriting (4.1) as:

$$\min_{\mathbf{x}\in\mathbb{R}^n} f(\mathbf{x}) = \frac{1}{4mn}\sum_{j=1}^{m}\left[\langle \mathbf{a}_j, \mathbf{x}\rangle^2 - y_j\right]^2, \tag{4.5}$$

where the normalization factor of $1/n$ helps in keeping the Hessian Lipschitz parameter independent of the dimension n. Note that in the earlier formulation (4.1) if we had $M = \Theta(n)$ then in the new formulation (4.5) we will have $M = \Theta(1)$.

Next, we once again set $m = n$ in (4.5), where $\mathbf{x} \in \mathbb{R}^n$, and vary n in the interval $[20, 800]$. As before the $\mathbf{a}_j$'s are the canonical basis of $\mathbb{R}^n$, while the eigen gap δ is fixed at 0.1. We then set the y_j's as follows:

$$y_j = \begin{cases} \frac{mn}{20} \ ; & 1 \leqslant j \leqslant \left\lfloor \frac{m}{2} \right\rfloor \\ -\frac{mn}{20} \ ; & \left\lfloor \frac{m}{2} \right\rfloor + 1 \leqslant j \leqslant 2\left\lfloor \frac{m}{2} \right\rfloor - 1 \\ -5mn \ ; & \textit{otherwise}. \end{cases} \tag{4.6}$$

Since the $\mathbf{a}_j$'s are orthonormal, it can be readily checked after adapting (4.3) for the modified formulation (4.5) that the eigenvalues of $\nabla^2 f(\mathbf{x})$ at $\mathbf{x} = \mathbf{0}$ are $\frac{-y_j}{mn}$ and we have $L = 5$, $\beta = 1/20$, and $\delta = 0.1$ from the given choice of y_j's. This construction preserves the parameters L, β from Assumptions **A2**, **A4** and the eigen gap δ from Proposition 2.2 for the function $f(\cdot)$ as the problem dimension n is varied (the parameter M from Assumption **A3** also gets independent of the dimension n). Finally, for $n \in [20, 800]$

we compute the exit time from the ε-neighborhood of origin for different values of initial unstable subspace projections. The results are plotted in Figure 7 for two values of initial unstable subspace projections for $\alpha = 0.1/L$, where we have displayed the exit time versus dimension n on the logarithmic scale. We observe that the exit time decreases with increasing dimension n, which agrees with Theorem 3.3 where we have $K_{exit} \lessapprox \mathcal{O}(\log n^{-1})$.

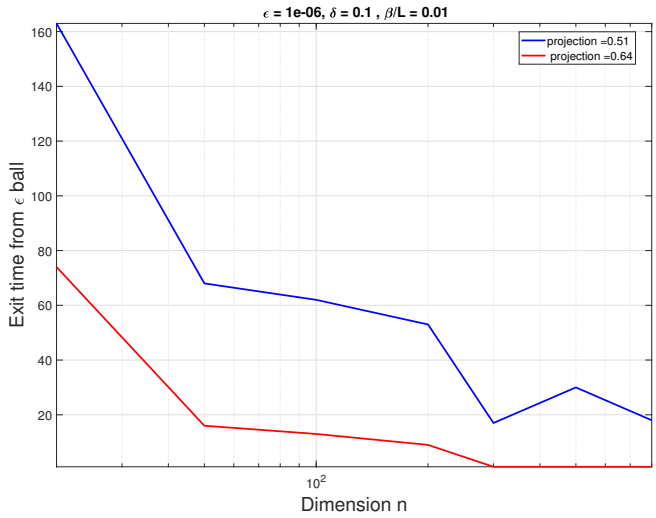


FIG. 7: Exit time vs dimension n under certain initial unstable subspace projections for given values of δ, L, β and ε.

4.1 *Evolution of the trajectory function $\Psi(K)$ from Theorem 3.2 on phase retrieval problem*

We now illustrate that the trajectory function $\Psi(K)$ first increases to a maximum and then continuously decreases to $-\infty$ from the example of the phase retrieval problem. In particular, if the initial unstable subspace projection is not too small then there exists a non-trivial finite K where $\Psi(K) > 1$, which is the upper bound on the exit time. In the phase retrieval problem (4.1) we set $m = n = 20$, where $\mathbf{x} \in \mathbb{R}^n$, the $\mathbf{a}_j$'s are taken to be the canonical basis of $\mathbb{R}^n$, the eigen gap $\delta = 0.5$, $L = 20$, and $\beta = 2$. We then set the y_j's as follows:

$$y_j = \begin{cases} m(\beta+\delta) \ ; & j = 1 \\ m\beta \ ; & j = 2 \\ -m\beta \ ; & 3 \leqslant j \leqslant m-1 \\ -mL \ ; & otherwise. \end{cases} \tag{4.7}$$

The results are plotted in Figure 8 for two values of initial unstable subspace projections for $\alpha = 1/L$. Clearly, the trajectory function $\Psi(K)$ first increases to a maximum and then continuously decreases to $-\infty$, which agrees with Theorem 3.2 (see also the discussion following Remark 3.9).

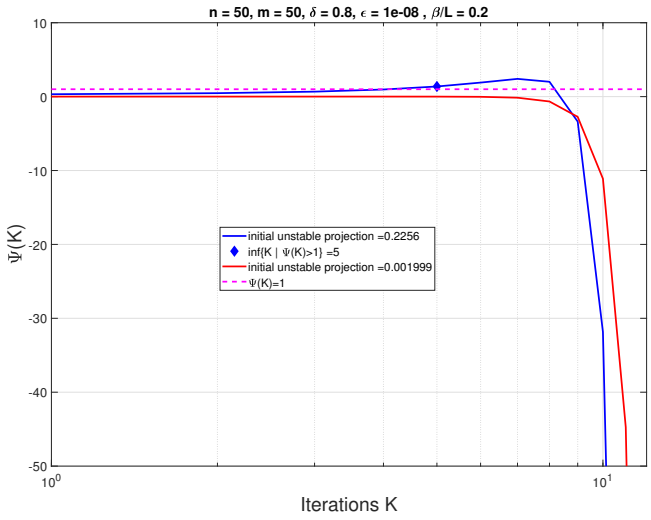


FIG. 8: $\Psi(K)$ vs K under certain initial unstable subspace projections for given values of n, L, β, δ and ε. The blue curve has a sufficient initial unstable subspace projection that allows it to first increase, become greater than 1 and then decrease whereas the red curve always remains below $\Psi(K) = 1$ and keeps on decreasing since it has a very small initial unstable subspace projection.

5. Conclusion

This work has focused on the analysis of gradient-descent trajectories in some small neighborhoods of a strict saddle point. Using tools from matrix perturbation theory and first-order eigenvector perturbations, a proof technique has been developed that describes the behavior of gradient-descent method as a function of the local geometry around a strict saddle point. Two novel lemmas have been presented in this work that quantify the radius of a saddle neighborhood within which an approximate analysis for the gradient-descent trajectory can be developed, provided the trajectory stays inside this neighborhood for a bounded interval. Next, this work has also presented two novel theorems that quantify this approximate trajectory distance from the saddle point at every iteration and provide an exit time from the saddle neighborhood based on the initial unstable projection of the radial vector. Developing a robust algorithm that can leverage this analysis so as to efficiently escape saddle neighborhood and a rigorous analysis of the trajectory function are some of the directions that have been pursued in a follow-up paper [11] to this work.

6. Data Availability Statement

The data underlying this paper are available in the paper and in its online supplementary material.

Acknowledgments

This work was supported in part by the National Science Foundation under grants CCF-1453073, CCF-1907658, CCF-1910110, OAC-1940209, CNS-2148104, CCF-1814888, and DMS-2053485, by

the Army Research Office under grants W911NF-17-1-0546 and W911NF-21-1-0301, by the Office of Naval Research Award Number N00014-21-1-2244 and by the DARPA Lagrange Program under ONR/SPAWAR contract N660011824020. The authors would also like to thank H. Vincent Poor, an anonymous reader, and the reviewers for their careful reading and many helpful suggestions that have helped improve the paper.

Appendices

A. On the equivalence of (3.19) and (3.20)

LEMMA A.1 In the setting of Section 3.2, the exit time (3.19) is equivalent to (3.20).

Proof. First, we show that the condition $\|\mathbf{x}_k-\mathbf{x}^*\|=0$ does not hold for any finite $k\geqslant 0$. For $k=0$, this is a trivial statement as our initialization is such that $\|\mathbf{x}_k-\mathbf{x}^*\|=\varepsilon>0$. We then proceed by induction. Suppose that $\|\mathbf{x}_{k-1}-\mathbf{x}^*\|>0$ and $\|\mathbf{x}_k-\mathbf{x}^*\|=0$ for some finite $k\geqslant 1$. Since $\mathbf{x}_k-\mathbf{x}^*=\mathbf{x}_{k-1}-\mathbf{x}^*-\alpha\nabla f(\mathbf{x}_{k-1})$, we can write $\|\mathbf{x}_k-\mathbf{x}^*\|=\|(\mathbf{I}-\alpha\mathbf{M})(\mathbf{x}_{k-1}-\mathbf{x}^*)\|$ with $\mathbf{M}=\int_{p=0}^{1}\nabla^2 f(\mathbf{x}^*+p(\mathbf{x}_{k-1}-\mathbf{x}^*))dp$, where we have used Taylor's formula (with an integral form) to represent the gradient of f as an integral over the Hessian of f. By assumption, we have $\alpha\leqslant\frac{1}{L}$ and we first consider the case of $\alpha<\frac{1}{L}$ so that $\alpha\|\mathbf{M}\|_2<1$, which implies $\left\|(\mathbf{I}-\alpha\mathbf{M})^{-1}\right\|_2^{-1}>0$ and we can therefore write $(\mathbf{I}-\alpha\mathbf{M})^{-1}(\mathbf{x}_k-\mathbf{x}^*)=\mathbf{x}_{k-1}-\mathbf{x}^*$. Then, $\|\mathbf{x}_k-\mathbf{x}^*\|\geqslant\left\|(\mathbf{I}-\alpha\mathbf{M})^{-1}\right\|_2^{-1}\|(\mathbf{x}_{k-1}-\mathbf{x}^*)\|>0$, which leads to a contradiction. Therefore, we conclude that $\|\mathbf{x}_k-\mathbf{x}^*\|>0$ for every k. Correspondingly, the quantity $\gamma_k=\frac{\langle\mathbf{v}_n,(\mathbf{x}_k-\mathbf{x}^*)\rangle}{\|\mathbf{x}_k-\mathbf{x}^*\|}$ is well-defined in the sense that its denominator cannot vanish. Here, $\gamma_k\in[0,1]$ because the vectors $\mathbf{v}_n$ and $\frac{\mathbf{x}_k-\mathbf{x}^*}{\|\mathbf{x}_k-\mathbf{x}^*\|}$ are both unit vectors and if the dot product is negative, we can always flip the sign of the eigenvector $\mathbf{v}_n$. Note that throughout this crude analysis section, for the sake of simplicity, we assume the dot product does not vanish, i.e., $\gamma_k\neq 0$ for any k, because otherwise the set $\{k|\langle\mathbf{v}_n,(\mathbf{x}_k-\mathbf{x}^*)\rangle>\gamma_k\varepsilon\}$ can be empty.[10]

Next, notice that by the definition of γ_k, we have $\langle\mathbf{v}_n,(\mathbf{x}_k-\mathbf{x}^*)\rangle>\gamma_k\varepsilon\iff\|\mathbf{x}_k-\mathbf{x}^*\|>\varepsilon$; this is because by multiplying the latter inequality with the positive scalar γ_k, we can simply obtain the former inequality. Therefore, we conclude that the sets $\{k|\,\|\mathbf{x}_k-\mathbf{x}^*\|>\varepsilon\}$ and $\{k|\langle\mathbf{v}_n,(\mathbf{x}_k-\mathbf{x}^*)\rangle>\gamma_k\varepsilon\}$ (defined in (3.19) and (3.20) respectively) are identical for $\gamma_k\in(0,1]$ and $\alpha<\frac{1}{L}$. When $\alpha=\frac{1}{L}$, we can have $\|\mathbf{x}_k-\mathbf{x}^*\|=0$ for some finite $k=K$, but since $\mathbf{x}^*$ is a fixed point of the gradient descent iteration, we will get $\|\mathbf{x}_k-\mathbf{x}^*\|=0$ for all $k>K$, which implies $\inf_{k>0}\{k|\,\|\mathbf{x}_k-\mathbf{x}^*\|>\varepsilon\}=\infty$. Since we are looking for finite exit times in the crude analysis, we can disregard the case of $\|\mathbf{x}_k-\mathbf{x}^*\|=0$ for some finite k when $\alpha=\frac{1}{L}$, and then for $\gamma_k\in(0,1]$, we again conclude that the sets $\{k|\,\|\mathbf{x}_k-\mathbf{x}^*\|>\varepsilon\}$ and $\{k|\langle\mathbf{v}_n,(\mathbf{x}_k-\mathbf{x}^*)\rangle>\gamma_k\varepsilon\}$ are identical. Therefore, we conclude that (3.19) and (3.20) are equivalent. This completes the proof. □

[10]This assumption would be satisfied for instance for quadratic objectives if the initialization has a non-zero component in the stable subspace of the Hessian at the saddle point; this can be verified as the solutions admit an explicit formula for every k in the quadratic case.

B. Proof of Lemma 3.3 (Hessian perturbation)

Proof. From the Taylor expansion around the strict saddle point $\mathbf{x}^*$ along the direction $\mathbf{x}_k - \mathbf{x}^*$ we have the following:

$$\nabla f(\mathbf{x}_k) = \nabla f(\mathbf{x}^*) + \int_{p=0}^{p=1} \nabla^2 f(\mathbf{x}^* + p\mathbf{u}_k)\mathbf{u}_k dp \tag{B.1}$$

$$\implies \nabla f(\mathbf{x}_k) = \nabla f(\mathbf{x}^*) + \int_{p=0}^{p=1} \nabla^2 f(\mathbf{x}^* + p\|\mathbf{u}_k\|\hat{\mathbf{u}}_k)\mathbf{u}_k dp, \tag{B.2}$$

where $\mathbf{u}_k = \mathbf{x}_k - \mathbf{x}^*$ and $\{\mathbf{x}_k\}$ is the sequence of iterates generated from the gradient descent method (3.1).

Note that here in the last step, we have made the substitution of $\mathbf{u}_k = \|\mathbf{u}_k\|\hat{\mathbf{u}}_k$ and we have that $\|\mathbf{u}_k\| \leqslant \varepsilon$ since our iterate $\mathbf{x}_k$ lies inside the ball $\mathscr{B}_\varepsilon(\mathbf{x}^*)$. $\hat{\mathbf{u}}_k$ represents the unit vector in the direction of $\mathbf{u}_k$.

Next, we start developing the term $\nabla^2 f(\mathbf{x}^* + p\|\mathbf{u}_k\|\hat{\mathbf{u}}_k)$ using matrix perturbation theory and variational calculus. We start with introducing a matrix function $\mathbf{G}(\cdot) : \mathbb{R} \to \mathbb{R}^{n\times n}$ which is given by

$$\mathbf{G}(w) = \nabla^2 f(\mathbf{x}^* + w\hat{\mathbf{u}}_k), \tag{B.3}$$

where $w = p\|\mathbf{u}_k\|$, p being the variable of previous integration and therefore $w = \mathscr{O}(\varepsilon)$. For sufficiently small ε we can utilize the Taylor series expansion of $\mathbf{G}(w)$ around $w = 0$:

$$\mathbf{G}(w) = \mathbf{G}(0) + w\frac{d\mathbf{G}}{dw}\bigg|_{w=0} + \frac{w^2}{2}\frac{d^2\mathbf{G}}{dw^2}\bigg|_{w=0} + \quad + \ldots \tag{B.4}$$

$$\implies \nabla^2 f(\mathbf{x}^* + w\hat{\mathbf{u}}_k) = \underbrace{\nabla^2 f(\mathbf{x}^*) + w\frac{d}{dw}(\nabla^2 f(\mathbf{x}^* + w\hat{\mathbf{u}}_k))\bigg|_{w=0}}_{S_1} + \underbrace{\frac{w^2}{2}\frac{d^2}{dw^2}(\nabla^2 f(\mathbf{x}^* + w\hat{\mathbf{u}}_k))\bigg|_{w=0} + \quad + \ldots}_{R_1}. \tag{B.5}$$

With $w = \mathscr{O}(\varepsilon)$ and the eigenvalues of $\nabla^2 f(\mathbf{x}^*)$ separated by δ or more, we can get rid of all the higher-order terms in the Taylor sequence from w^2 onwards. It is a reasonable approximation from the Rayleigh–Schrödinger perturbation theory ([6, 14, 43]) as long as we have Proposition 2.2, i.e., there are at least two eigenvalue groups of $\nabla^2 f(\mathbf{x}^*)$ that are not degenerate or too close to one another. This leaves us with the following first order approximation:

$$\nabla^2 f(\mathbf{x}^* + w\hat{\mathbf{u}}_k) = \nabla^2 f(\mathbf{x}^*) + w\frac{d}{dw}(\nabla^2 f(\mathbf{x}^* + w\hat{\mathbf{u}}_k))\bigg|_{w=0} + \mathscr{O}(\varepsilon^2), \tag{B.6}$$

where we have that $S_1 = \nabla^2 f(\mathbf{x}^*) + w\frac{d}{dw}(\nabla^2 f(\mathbf{x}^* + w\hat{\mathbf{u}}_k))\Big|_{w=0}$ and the order of the remainder term R_1 is $\mathscr{O}(\varepsilon^2)$. This remainder term R_1 is easy to obtain from Taylor's Remainder theorem. Applying this theorem to (B.5) with the substitution $\nabla^2 f(\mathbf{x}^* + u\hat{\mathbf{u}}_k) = \mathbf{G}(u)$ yields

$$R_1 = \int_0^w u\frac{d^2\mathbf{G}}{du^2}du \tag{B.7}$$

$$\implies \|R_1\|_2 = \left\|\int_0^w u\frac{d^2\mathbf{G}}{du^2}du\right\|_2 < \left(\int_0^w \left\|\frac{d^2\mathbf{G}}{du^2}\right\|_2^2 du\right)^{\frac{1}{2}}\left(\int_0^w u^2 du\right)^{\frac{1}{2}} \leqslant \frac{B_2 w^2}{\sqrt{3}} < \frac{B_2\varepsilon^2}{\sqrt{3}}. \tag{B.8}$$

Here in the last step we applied the Cauchy-Schwarz inequality followed by an extra assumption on the spectral radius of $\frac{d^2\mathbf{G}}{du^2}$ which is $\left\|\frac{d^2\mathbf{G}}{du^2}\right\|_2 \leqslant B_2$ for some finite positive value B_2. The final inequality follows from the fact that $w = p\|\mathbf{u}_k\| < \varepsilon$ where $0 < p \leqslant 1$. Hence the remainder term R_1 is of order $\mathscr{O}(\varepsilon^2)$. Note that the condition of $\left\|\frac{d^2\mathbf{G}}{du^2}\right\|_2 \leqslant B_2 < \infty$ is valid for any analytic function $f(\cdot)$. Moreover, it bounds the variations of the Hessian inside the ball $\mathscr{B}_\varepsilon(\mathbf{x}^*)$.

Next, using a matrix substitution of $\mathbf{H}(\hat{\mathbf{u}}_k) = \frac{d}{dw}(\nabla^2 f(\mathbf{x}^* + w\hat{\mathbf{u}}_k))|_{w=0}$, our first order Hessian approximation becomes

$$\nabla^2 f(\mathbf{x}^* + w\hat{\mathbf{u}}_k) = \nabla^2 f(\mathbf{x}^*) + w\mathbf{H}(\hat{\mathbf{u}}_k) + \mathscr{O}(\varepsilon^2) \tag{B.9}$$

$$\implies \nabla^2 f(\mathbf{x}^* + p\mathbf{u}_k) = \nabla^2 f(\mathbf{x}^*) + p\|\mathbf{u}_k\|\,\mathbf{H}(\hat{\mathbf{u}}_k) + \mathscr{O}(\varepsilon^2). \tag{B.10}$$

B.1 *Rayleigh–Schrödinger perturbation analysis*

We can now find the matrix $\mathbf{H}(\hat{\mathbf{u}}_k)$ using the spectral theorem and the Rayleigh–Schrödinger perturbation theory. Note that this matrix $\mathbf{H}(\hat{\mathbf{u}}_k)$ depends on the unit vector $\hat{\mathbf{u}}_k$.

We first apply the spectral theorem on the real symmetric matrix $\nabla^2 f(\mathbf{x}^* + w\hat{\mathbf{u}}_k)$ to get the following decomposition in terms of its eigenvalues $\lambda_i(w)$ and the eigenvectors $\mathbf{v}_i(w)$:

$$\nabla^2 f(\mathbf{x}^* + w\hat{\mathbf{u}}_k) = \sum_{i=1}^{n} \lambda_i(w)\mathbf{v}_i(w)\mathbf{v}_i(w)^T. \tag{B.11}$$

Now, differentiating this decomposition w.r.t. the variable w and obtaining its value at the point $w = 0$ we get

$$\frac{d}{dw}(\nabla^2 f(\mathbf{x}^* + w\hat{\mathbf{u}}_k)) = \sum_{i=1}^{n} \frac{d}{dw}(\lambda_i(w)\mathbf{v}_i(w)\mathbf{v}_i(w)^T) \tag{B.12}$$

$$\begin{aligned}\implies \frac{d}{dw}(\nabla^2 f(\mathbf{x}^* + w\hat{\mathbf{u}}_k))\bigg|_{w=0} &= \sum_{i=1}^{n} \left(\frac{d}{dw}(\lambda_i(w))\bigg|_{w=0} \mathbf{v}_i(0)\mathbf{v}_i(0)^T + \lambda_i(0)\frac{d}{dw}(\mathbf{v}_i(w))\bigg|_{w=0} \mathbf{v}_i(0)^T \right. \\ &\quad \left. + \lambda_i(0)\mathbf{v}_i(0)\frac{d}{dw}(\mathbf{v}_i(w)^T)\bigg|_{w=0} \right) \end{aligned} \tag{B.13}$$

$$\begin{aligned}\implies \mathbf{H}(\hat{\mathbf{u}}_k) &= \sum_{i=1}^{n} \left(\frac{d}{dw}(\lambda_i(w))\bigg|_{w=0} \mathbf{v}_i(0)\mathbf{v}_i(0)^T + \lambda_i(0)\frac{d}{dw}(\mathbf{v}_i(w))\bigg|_{w=0} \mathbf{v}_i(0)^T \right. \\ &\quad \left. + \lambda_i(0)\mathbf{v}_i(0)\frac{d}{dw}(\mathbf{v}_i(w)^T)\bigg|_{w=0} \right). \end{aligned} \tag{B.14}$$

Note that the pair $(\lambda_i(0), \mathbf{v}_i(0))$ represents the i^{th} eigenvalue-eigenvector pair of the unperturbed matrix $\nabla^2 f(\mathbf{x}^*)$. From the Rayleigh–Schrödinger perturbation theory ([43]), for a given first order perturbation matrix $\mathbf{H}(\hat{\mathbf{u}}_k)$ in (B.9), we have the following first order correction terms:

$$\frac{d}{dw}(\lambda_i(w))\bigg|_{w=0} = \langle \mathbf{v}_i(0), \mathbf{H}(\hat{\mathbf{u}}_k)\mathbf{v}_i(0) \rangle \tag{B.15}$$

$$\frac{d}{dw}(\mathbf{v}_i(w))\bigg|_{w=0} = \sum_{l \neq i} \frac{\langle \mathbf{v}_l(0), \mathbf{H}(\hat{\mathbf{u}}_k)\mathbf{v}_i(0) \rangle}{\lambda_i(0) - \lambda_l(0)} \mathbf{v}_l(0). \tag{B.16}$$

Observe that under Proposition 2.2, we are considering the case of $m = n$, i.e., no degenerate eigenvalues in our analysis. However, we have a subsection after Lemma 3.3 (generality of Lemma 3.3) that discusses the degenerate case as well.

Substituting these first-order correction terms in (B.14), we get the following result:

$$\mathbf{H}(\hat{\mathbf{u}}_k) = \sum_{i=1}^{n} \Bigg(\langle \mathbf{v}_i(0), \mathbf{H}(\hat{\mathbf{u}}_k)\mathbf{v}_i(0)\rangle \mathbf{v}_i(0)\mathbf{v}_i(0)^T + \lambda_i(0)\Big(\sum_{l\neq i} \frac{\langle \mathbf{v}_l(0), \mathbf{H}(\hat{\mathbf{u}}_k)\mathbf{v}_i(0)\rangle}{\lambda_i(0)-\lambda_l(0)}\mathbf{v}_l(0)\Big)\mathbf{v}_i(0)^T + \lambda_i(0)\mathbf{v}_i(0)\Big(\sum_{l\neq i} \frac{\langle \mathbf{v}_l(0), \mathbf{H}(\hat{\mathbf{u}}_k)\mathbf{v}_i(0)\rangle}{\lambda_i(0)-\lambda_l(0)}\mathbf{v}_l(0)\Big)^T \Bigg). \tag{B.17}$$

Now, combining this result with (B.14) and substituting the subsequent matrix approximation in (B.2) leads to the following result:

$$\nabla f(\mathbf{x}_k) = \nabla f(\mathbf{x}^*) + \int_{p=0}^{p=1} (\nabla^2 f(\mathbf{x}^*) + p\,\|\mathbf{u}_k\|\,\mathbf{H}(\hat{\mathbf{u}}_k) + \mathscr{O}(\varepsilon^2))\mathbf{u}_k dp \tag{B.18}$$

$$= \left(\nabla^2 f(\mathbf{x}^*) + \frac{\|\mathbf{u}_k\|}{2}\mathbf{H}(\hat{\mathbf{u}}_k) + \mathscr{O}(\varepsilon^2)\right)\mathbf{u}_k. \tag{B.19}$$

Note that $\|\mathbf{u}_k\|\,\mathbf{H}(\hat{\mathbf{u}}_k)$ and $\mathbf{u}_k$ do not depend on p and hence can be pulled out of the integral.

B.2 *Validity of the Taylor expansion in Rayleigh–Schrödinger analysis*

Recall that we used the Taylor expansion in (B.5) for the matrix $\mathbf{G}(w)$ around $w = 0$. Next, we evaluated the first-order perturbation term $\mathbf{H}(\hat{\mathbf{u}}_k)$ in this expansion using the Rayleigh–Schrödinger perturbation analysis, which is dependent on this Taylor expansion (see derivations in [6, 14]). In other words, the perturbation analysis is only valid for those values of w where the Taylor expansion for the matrix $\mathbf{G}(w)$ around $w = 0$ converges. This directly reduces to the problem of finding the radius of convergence for the expansion (B.5).

Although evaluating the radius of convergence in the Rayleigh–Schrödinger perturbation analysis remains an open problem in general, we can still find the radius of convergence for the expansion (B.5) using matrix power series.

For the Taylor expansion in (B.5), consider the sequence $\{r_j(\hat{\mathbf{u}}_k)\}$ for all $j \in \{1, 2, ...\}$ such that

$$r_j(\hat{\mathbf{u}}_k) = \left\| \left(\frac{d^j \mathbf{G}}{dw^j}\bigg|_{w=0} \right) \right\|_2, \tag{B.20}$$

where $\mathbf{G}(w) = \nabla^2 f(\mathbf{x}^* + w\hat{\mathbf{u}}_k)$ and $w = p\,\|\mathbf{u}_k\|$ with $0 < p \leqslant 1$.

Next by the Cauchy–Hadamard theorem, for any power series defined by

$$h(z) = \sum_{j=0}^{\infty} c_j (z-a)^j \tag{B.21}$$

where $z \in \mathbb{C}$, the radius of convergence for the series is given by

$$r = \left(\limsup_{j\to\infty} \sqrt[j]{|c_j|} \right)^{-1}. \tag{B.22}$$

For the case of matrix power series, the spectral radius of a matrix is used to determine the radius of convergence. From the expression of the $r_j(\hat{\mathbf{u}}_k)$ in (B.20), it is clear that the matrix $\frac{d^j\mathbf{G}}{dw^j}\Big|_{w=0}$ is real-symmetric since $\mathbf{G}$ is real-symmetric. Hence, the spectral radius of this matrix is equal to its l_2 norm.

Using the Cauchy–Hadamard theorem on our expansion (B.5) for $|c_j| = \frac{r_j(\hat{\mathbf{u}}_k)}{j!}$, we get the following radius of convergence:

$$r(\hat{\mathbf{u}}_k) = \left(\limsup_{j\to\infty} \sqrt[j]{\frac{r_j(\hat{\mathbf{u}}_k)}{j!}} \right)^{-1}. \tag{B.23}$$

Therefore, if $\sqrt[j]{\frac{r_j(\hat{\mathbf{u}}_k)}{j!}}$ is upper bounded for all j, then a non-zero radius of convergence is guaranteed. This implies that

$$w = p\,\|\mathbf{u}_k\| < \left(\limsup_{j\to\infty} \sqrt[j]{\frac{r_j(\hat{\mathbf{u}}_k)}{j!}} \right)^{-1}. \tag{B.24}$$

Since $w < \varepsilon$ for any $\mathbf{x}_k \in \mathscr{B}_\varepsilon(\mathbf{x}^*)$, where $\mathbf{x}_k = \mathbf{x}^* + w\hat{\mathbf{u}}_k$, by setting a condition on ε such that $\varepsilon < \left(\limsup_{j\to\infty} \sqrt[j]{\frac{r_j(\hat{\mathbf{u}}_k)}{j!}} \right)^{-1}$, we can guarantee the inequality (B.24). However this result should hold for any possible unit directional vector $\hat{\mathbf{u}}_k$. Hence we must have

$$\varepsilon < \inf_{\hat{\mathbf{u}}_k} \left(\limsup_{j\to\infty} \sqrt[j]{\frac{r_j(\hat{\mathbf{u}}_k)}{j!}} \right)^{-1} \tag{B.25}$$

$$\implies \varepsilon < \inf_{\|\mathbf{u}\|=1} \left(\limsup_{j\to\infty} \sqrt[j]{\frac{r_j(\mathbf{u})}{j!}} \right)^{-1}, \tag{B.26}$$

where

$$r_j(\mathbf{u}) = \left\| \left(\frac{d^j}{dw^j} \nabla^2 f(\mathbf{x}^* + w\mathbf{u}) \Big|_{w=0} \right) \right\|_2. \tag{B.27}$$

It is to be noted that this bound on ε only guarantees convergence of the expansion (B.5) and not the convergence of terms generated by the Rayleigh–Schrödinger perturbation analysis. Evaluating the convergence radius from the Rayleigh–Schrödinger perturbation theory is beyond the scope of the current work. Hence this condition on ε is necessary but may not be sufficient.

B.3 *Note on the existence of a positive upper bound on ε*

For the condition (B.26) to make sense, we must have $\inf_{\|\mathbf{u}\|=1} \left(\limsup_{j\to\infty} \sqrt[j]{\frac{r_j(\mathbf{u})}{j!}} \right)^{-1} > 0$. To this end, consider the following Taylor expansion with respect to the variable $w \geqslant 0$:

$$\nabla^2 f(\mathbf{x}^* + w\mathbf{u}) = \sum_{j=0}^{\infty} \frac{d^j}{dw^j} \nabla^2 f(\mathbf{x}^* + w\mathbf{u}) \Big|_{w=0} \frac{w^j}{j!}, \tag{B.28}$$

where the above matrix-valued series converges with some strictly positive *radius of convergence* (ROC) R (i.e., $w \leqslant R$) for all $\{\mathbf{u} : \|\mathbf{u}\|_2 = 1\}$ due to the analytic nature of $f(\cdot)$. Here, we focus on convergence of the series with respect to the operator (spectral) norm and note that for any n-dimensional symmetric matrix $\mathbf{A}$ we have the inequality $\frac{1}{n}\max_{i,l}\{|[\mathbf{A}]_{i,l}|\} \leqslant \frac{1}{n}\|\mathbf{A}\|_F \leqslant \|\mathbf{A}\|_2 \leqslant \|\mathbf{A}\|_F$. Thus, if the matrix-valued series (B.28) converges for $w \leqslant R$ in the spectral norm then the matrix sum on the right-hand side of (B.28) must also element-wise converge for the same ROC R. For the $(i,l)^{th}$ element of $\nabla^2 f(\mathbf{x}^* + w\mathbf{u})$ to converge in (B.28), we must have $w \leqslant R \leqslant \left(\limsup_{j\to\infty} \sqrt[j]{\frac{1}{j!}\left|\frac{d^j}{dw^j}[\nabla^2 f(\mathbf{x}^* + w\mathbf{u})]_{i,l}\right|_{w=0}}\right)^{-1}$ for any unit vector $\mathbf{u}$. Precisely, the ROC for (B.28) is given by

$$R = \min_{i,l} \inf_{\|\mathbf{u}\|=1} \left(\limsup_{j\to\infty} \sqrt[j]{\frac{1}{j!}\left|\frac{d^j}{dw^j}[\nabla^2 f(\mathbf{x}^* + w\mathbf{u})]_{i,l}\right|_{w=0}}\right)^{-1}$$

which is strictly positive. Next, due to $\frac{1}{n}\max_{i,l}\{|[\mathbf{A}]_{i,l}|\} \leqslant \|\mathbf{A}\|_2$, we will have the following for any $(i,l)^{th}$ element of $\mathbf{A} = \frac{d^j}{dw^j}\nabla^2 f(\mathbf{x}^* + w\mathbf{u})\Big|_{w=0}$:

$$\frac{1}{j!}\left|\frac{d^j}{dw^j}[\nabla^2 f(\mathbf{x}^* + w\mathbf{u})]_{i,l}\Big|_{w=0}\right| \leqslant \frac{n}{j!}\left\|\frac{d^j}{dw^j}\nabla^2 f(\mathbf{x}^* + w\mathbf{u})\Big|_{w=0}\right\|_2 \tag{B.29}$$

$$\implies \sqrt[j]{\frac{1}{j!}\left|\frac{d^j}{dw^j}[\nabla^2 f(\mathbf{x}^* + w\mathbf{u})]_{i,l}\Big|_{w=0}\right|} \leqslant \sqrt[j]{\frac{n}{j!}\left\|\frac{d^j}{dw^j}\nabla^2 f(\mathbf{x}^* + w\mathbf{u})\Big|_{w=0}\right\|_2} \tag{B.30}$$

$$\implies \limsup_{j\to\infty} \sqrt[j]{\frac{1}{j!}\left|\frac{d^j}{dw^j}[\nabla^2 f(\mathbf{x}^* + w\mathbf{u})]_{i,l}\Big|_{w=0}\right|} \leqslant \limsup_{j\to\infty} n^{1/j}\sqrt[j]{\frac{1}{j!}\left\|\frac{d^j}{dw^j}\nabla^2 f(\mathbf{x}^* + w\mathbf{u})\Big|_{w=0}\right\|_2} \tag{B.31}$$

$$\implies \left(\limsup_{j\to\infty} \sqrt[j]{\frac{1}{j!}\left|\frac{d^j}{dw^j}[\nabla^2 f(\mathbf{x}^* + w\mathbf{u})]_{i,l}\Big|_{w=0}\right|}\right)^{-1} \geqslant \left(\limsup_{j\to\infty} \sqrt[j]{\frac{1}{j!}\left\|\frac{d^j}{dw^j}\nabla^2 f(\mathbf{x}^* + w\mathbf{u})\Big|_{w=0}\right\|_2}\right)^{-1}$$

$$= \left(\limsup_{j\to\infty} \sqrt[j]{\frac{r_j(\mathbf{u})}{j!}}\right)^{-1} \tag{B.32}$$

$$\implies R = \min_{i,l} \inf_{\|\mathbf{u}\|=1} \left(\limsup_{j\to\infty} \sqrt[j]{\frac{1}{j!}\left|\frac{d^j}{dw^j}[\nabla^2 f(\mathbf{x}^* + w\mathbf{u})]_{i,l}\Big|_{w=0}\right|}\right)^{-1} \geqslant \inf_{\|\mathbf{u}\|=1} \left(\limsup_{j\to\infty} \sqrt[j]{\frac{r_j(\mathbf{u})}{j!}}\right)^{-1}, \tag{B.33}$$

where we used the lim sup product rule in the second last step. Now (B.33) implies that the quantity $\inf_{\|\mathbf{u}\|=1}\left(\limsup_{j\to\infty}\sqrt[j]{\frac{r_j(\mathbf{u})}{j!}}\right)^{-1}$ is upper bounded by the radius of convergence R of the series in (B.28). Next, due to the inequality $\frac{1}{n^2}\|\mathbf{A}\|_2 \leqslant \frac{1}{n^2}\|\mathbf{A}\|_F \leqslant \frac{1}{n}\max_{i,l}\{|[\mathbf{A}]_{i,l}|\}$, for the maximum absolute element of $\mathbf{A} = \frac{d^j}{dw^j}\nabla^2 f(\mathbf{x}^* + w\mathbf{u})\Big|_{w=0}$ denoted by $\left|\frac{d^j}{dw^j}[\nabla^2 f(\mathbf{x}^* + w\mathbf{u})]_{m(j),q(j)}\Big|_{w=0}\right|$ we have the

following:[11]

$$\frac{n}{j!}\left|\frac{d^j}{dw^j}[\nabla^2 f(\mathbf{x}^*+w\mathbf{u})]_{m(j),q(j)}\Big|_{w=0}\right| \geqslant \frac{1}{j!}\left\|\frac{d^j}{dw^j}\nabla^2 f(\mathbf{x}^*+w\mathbf{u})\Big|_{w=0}\right\|_2 \tag{B.34}$$

$$\implies \sqrt[j]{\frac{n}{j!}\left|\frac{d^j}{dw^j}[\nabla^2 f(\mathbf{x}^*+w\mathbf{u})]_{m(j),q(j)}\Big|_{w=0}\right|} \geqslant \sqrt[j]{\frac{1}{j!}\left\|\frac{d^j}{dw^j}\nabla^2 f(\mathbf{x}^*+w\mathbf{u})\Big|_{w=0}\right\|_2} \tag{B.35}$$

$$\implies \limsup_{j\to\infty} n^{1/j}\sqrt[j]{\frac{1}{j!}\left|\frac{d^j}{dw^j}[\nabla^2 f(\mathbf{x}^*+w\mathbf{u})]_{m(j),q(j)}\Big|_{w=0}\right|} \geqslant \limsup_{j\to\infty}\sqrt[j]{\frac{1}{j!}\left\|\frac{d^j}{dw^j}\nabla^2 f(\mathbf{x}^*+w\mathbf{u})\Big|_{w=0}\right\|_2} \tag{B.36}$$

$$\implies \left(\limsup_{j\to\infty}\sqrt[j]{\frac{1}{j!}\left|\frac{d^j}{dw^j}[\nabla^2 f(\mathbf{x}^*+w\mathbf{u})]_{m(j),q(j)}\Big|_{w=0}\right|}\right)^{-1} \leqslant \left(\limsup_{j\to\infty}\sqrt[j]{\frac{1}{j!}\left\|\frac{d^j}{dw^j}\nabla^2 f(\mathbf{x}^*+w\mathbf{u})\Big|_{w=0}\right\|_2}\right)^{-1}$$
$$= \left(\limsup_{j\to\infty}\sqrt[j]{\frac{r_j(\mathbf{u})}{j!}}\right)^{-1} \tag{B.37}$$

$$\implies R \leqslant \inf_{\|\mathbf{u}\|=1}\left(\limsup_{j\to\infty}\sqrt[j]{\frac{1}{j!}\left|\frac{d^j}{dw^j}[\nabla^2 f(\mathbf{x}^*+w\mathbf{u})]_{m(j),q(j)}\Big|_{w=0}\right|}\right)^{-1} \leqslant \inf_{\|\mathbf{u}\|=1}\left(\limsup_{j\to\infty}\sqrt[j]{\frac{r_j(\mathbf{u})}{j!}}\right)^{-1}, \tag{B.38}$$

where the L.H.S. of the last inequality holds by $\min_{i,l}\inf_{\|\mathbf{u}\|=1}\left(\limsup_{j\to\infty}\sqrt[j]{\frac{1}{j!}\left|\frac{d^j}{dw^j}[\nabla^2 f(\mathbf{x}^*+w\mathbf{u})]_{i,l}\Big|_{w=0}\right|}\right)^{-1} \leqslant \inf_{\|\mathbf{u}\|=1}\left(\limsup_{j\to\infty}\sqrt[j]{\frac{1}{j!}\left|\frac{d^j}{dw^j}[\nabla^2 f(\mathbf{x}^*+w\mathbf{u})]_{m(j),q(j)}\Big|_{w=0}\right|}\right)^{-1}$. Finally, combining (B.33) and (B.38) we get:

$$R \leqslant \inf_{\|\mathbf{u}\|=1}\left(\limsup_{j\to\infty}\sqrt[j]{\frac{r_j(\mathbf{u})}{j!}}\right)^{-1} \leqslant R \tag{B.39}$$

$$\implies \inf_{\|\mathbf{u}\|=1}\left(\limsup_{j\to\infty}\sqrt[j]{\frac{r_j(\mathbf{u})}{j!}}\right)^{-1} = R. \tag{B.40}$$

□

C. Radial vector $\mathbf{u}_k$ in terms of the initialization $\mathbf{u}_0$

C.1 *Proof of Lemma 3.4*

Proof. Combining the equation $\mathbf{u}_k = \mathbf{x}_k - \mathbf{x}^*$ this with gradient descent update yields

$$\mathbf{u}_{k+1} - \mathbf{u}_k = -\alpha\nabla f(\mathbf{x}_k). \tag{C.1}$$

[11] Notice that the position $(m(j), q(j))$ of the maximum absolute element depends on j.

Next, substituting (B.19) here, we get the following recursion:

$$\mathbf{u}_{k+1}-\mathbf{u}_k=-\alpha\left(\nabla^2 f(\mathbf{x}^*)+\frac{\|\mathbf{u}_k\|}{2}\mathbf{H}(\hat{\mathbf{u}}_k)+\mathscr{O}(\varepsilon^2)\right)\mathbf{u}_k \tag{C.2}$$

$$\mathbf{u}_{k+1}=\left(\mathbf{I}-\alpha\left(\nabla^2 f(\mathbf{x}^*)+\frac{\|\mathbf{u}_k\|}{2}\mathbf{H}(\hat{\mathbf{u}}_k)+\mathscr{O}(\varepsilon^2)\right)\right)\mathbf{u}_k. \tag{C.3}$$

Finally substituting (B.17) here and applying the spectral theorem to the matrices $\mathbf{I}$ and $\nabla^2 f(\mathbf{x}^*)$ yields

$$\begin{aligned}\mathbf{u}_{k+1}=\Bigg(\sum_{i=1}^n \mathbf{v}_i(0)\mathbf{v}_i(0)^T-\alpha\Bigg(\sum_{i=1}^n \lambda_i(0)\mathbf{v}_i(0)\mathbf{v}_i(0)^T+\frac{\|\mathbf{u}_k\|}{2}\Bigg(\sum_{i=1}^n\Bigg(\langle\mathbf{v}_i(0),\mathbf{H}(\hat{\mathbf{u}}_k)\mathbf{v}_i(0)\rangle\mathbf{v}_i(0)\mathbf{v}_i(0)^T\\ +\lambda_i(0)(\sum_{l\neq i}\frac{\langle\mathbf{v}_l(0),\mathbf{H}(\hat{\mathbf{u}}_k)\mathbf{v}_i(0)\rangle}{\lambda_i(0)-\lambda_l(0)}\mathbf{v}_l(0))\mathbf{v}_i(0)^T+\lambda_i(0)\mathbf{v}_i(0)(\sum_{l\neq i}\frac{\langle\mathbf{v}_l(0),\mathbf{H}(\hat{\mathbf{u}}_k)\mathbf{v}_i(0)\rangle}{\lambda_i(0)-\lambda_l(0)}\mathbf{v}_l(0))^T\Bigg)\Bigg)\Bigg)+\mathscr{O}(\varepsilon^2)\Bigg)\mathbf{u}_k\end{aligned} \tag{C.4}$$

$$\begin{aligned}\mathbf{u}_{k+1}=\Bigg[\sum_{i=1}^n\left(1-\alpha\lambda_i(0)-\alpha\frac{\|\mathbf{u}_k\|}{2}\langle\mathbf{v}_i(0),\mathbf{H}(\hat{\mathbf{u}}_k)\mathbf{v}_i(0)\rangle\right)\mathbf{v}_i(0)\mathbf{v}_i(0)^T\\ -\alpha\frac{\|\mathbf{u}_k\|}{2}\sum_{i=1}^n\sum_{l\neq i}\frac{\langle\mathbf{v}_l(0),\mathbf{H}(\hat{\mathbf{u}}_k)\mathbf{v}_i(0)\rangle\lambda_i(0)}{\lambda_i(0)-\lambda_l(0)}\left(\mathbf{v}_l(0)\mathbf{v}_i(0)^T+\mathbf{v}_i(0)\mathbf{v}_l(0)^T\right)\Bigg]\mathbf{u}_k+\mathscr{O}(\varepsilon^2)\mathbf{u}_k.\end{aligned} \tag{C.5}$$

Next, we start analyzing the coefficients of spectral components $\mathbf{v}_i(0)\mathbf{v}_l(0)^T$ for any (i,l) pair.

C.1.1 *Coefficient bounds:.* We start with (C.5) and analyze it in terms of the stable subspace $\mathscr{E}_S$ and unstable subspace $\mathscr{E}_{US}$ of the Hessian $\nabla^2 f(\mathbf{x}^*)$. To this end we rewrite (C.5) and split its first term into the stable and unstable spectral components. The stable spectral components result from the positive eigenvalues of $\nabla^2 f(\mathbf{x}^*)$ whereas the unstable spectral components result from its negative eigenvalues.

$$\begin{aligned}\mathbf{u}_{k+1}=&\Bigg[\sum_{i=1}^n\left(1-\alpha\lambda_i(0)-\alpha\frac{\|\mathbf{u}_k\|}{2}\langle\mathbf{v}_i(0),\mathbf{H}(\hat{\mathbf{u}}_k)\mathbf{v}_i(0)\rangle\right)\mathbf{v}_i(0)\mathbf{v}_i(0)^T\\ &-\alpha\frac{\|\mathbf{u}_k\|}{2}\sum_{i=1}^n\sum_{l\neq i}\frac{\langle\mathbf{v}_l(0),\mathbf{H}(\hat{\mathbf{u}}_k)\mathbf{v}_i(0)\rangle\lambda_i(0)}{\lambda_i(0)-\lambda_l(0)}\left(\mathbf{v}_l(0)\mathbf{v}_i(0)^T+\mathbf{v}_i(0)\mathbf{v}_l(0)^T\right)\Bigg]\mathbf{u}_k+\mathscr{O}(\varepsilon^2)\mathbf{u}_k \qquad \text{(C.6)}\\ =&\Bigg[\sum_{i\in\mathscr{N}_S}\left(1-\alpha\lambda_i(0)-\alpha\frac{\|\mathbf{u}_k\|}{2}\langle\mathbf{v}_i(0),\mathbf{H}(\hat{\mathbf{u}}_k)\mathbf{v}_i(0)\rangle\right)\mathbf{v}_i(0)\mathbf{v}_i(0)^T\\ &+\sum_{j\in\mathscr{N}_{US}}\left(1-\alpha\lambda_j(0)-\alpha\frac{\|\mathbf{u}_k\|}{2}\langle\mathbf{v}_j(0),\mathbf{H}(\hat{\mathbf{u}}_k)\mathbf{v}_j(0)\rangle\right)\mathbf{v}_j(0)\mathbf{v}_j(0)^T\\ &-\alpha\frac{\|\mathbf{u}_k\|}{2}\sum_{i=1}^n\sum_{l\neq i}\frac{\langle\mathbf{v}_l(0),\mathbf{H}(\hat{\mathbf{u}}_k)\mathbf{v}_i(0)\rangle\lambda_i(0)}{\lambda_i(0)-\lambda_l(0)}\left(\mathbf{v}_l(0)\mathbf{v}_i(0)^T+\mathbf{v}_i(0)\mathbf{v}_l(0)^T\right)\Bigg]\mathbf{u}_k+\mathscr{O}(\varepsilon^2)\mathbf{u}_k \qquad \text{(C.7)}\\ =&\Bigg[\sum_{i\in\mathscr{N}_S}c_i^s(k)\mathbf{v}_i(0)\mathbf{v}_i(0)^T+\sum_{j\in\mathscr{N}_{US}}c_j^{us}(k)\mathbf{v}_j(0)\mathbf{v}_j(0)^T\\ &+\sum_{i=1}^n\sum_{l\neq i}\left(d_{l,i}(k)\mathbf{v}_l(0)\mathbf{v}_i(0)^T+d_{i,l}(k)\mathbf{v}_i(0)\mathbf{v}_l(0)^T\right)\Bigg]\mathbf{u}_k+\mathscr{O}(\varepsilon^2)\mathbf{u}_k, \qquad \text{(C.8)}\end{aligned}$$

where the coefficient terms $c_i^s(k)$, $c_j^{us}(k)$ and $d_{l,i}(k)$ in (C.8) are as follows:

$$c_i^s(k) = \left(1 - \alpha\lambda_i(0) - \alpha\frac{\|\mathbf{u}_k\|}{2}\langle \mathbf{v}_i(0), \mathbf{H}(\hat{\mathbf{u}}_k)\mathbf{v}_i(0)\rangle\right) \tag{C.9}$$

$$c_j^{us}(k) = \left(1 - \alpha\lambda_j(0) - \alpha\frac{\|\mathbf{u}_k\|}{2}\langle \mathbf{v}_j(0), \mathbf{H}(\hat{\mathbf{u}}_k)\mathbf{v}_j(0)\rangle\right) \tag{C.10}$$

$$d_{i,l}(k) = d_{l,i}(k) = \frac{\langle \mathbf{v}_l(0), \mathbf{H}(\hat{\mathbf{u}}_k)\mathbf{v}_i(0)\rangle\lambda_i(0)\alpha\|\mathbf{u}_k\|}{2(\lambda_l(0) - \lambda_i(0))}. \tag{C.11}$$

Now, from (B.10) and the Lipschitz continuity of the Hessian (Assumption **A3**), we get the following bound:

$$\nabla^2 f(\mathbf{x}^* + p\mathbf{u}_k) = \nabla^2 f(\mathbf{x}^*) + p\|\mathbf{u}_k\|\mathbf{H}(\hat{\mathbf{u}}_k) + \mathscr{O}(\varepsilon^2). \tag{C.12}$$

Recall that the term $\mathscr{O}(\varepsilon^2)$ comes from (B.7). Therefore, to further simplify the above equation, we replace $\mathscr{O}(\varepsilon^2)$ with $\int_0^w u\frac{d^2\mathbf{G}}{du^2}du$ from (B.7) where $w = p\|\mathbf{u}_k\|$. Then taking the norm of both sides, followed by triangle inequality and using Assumption **A3** yields

$$\nabla^2 f(\mathbf{x}^* + p\mathbf{u}_k) = \nabla^2 f(\mathbf{x}^*) + p\|\mathbf{u}_k\|\mathbf{H}(\hat{\mathbf{u}}_k) + \int_0^w u\frac{d^2\mathbf{G}}{du^2}du \tag{C.13}$$

$$\|\mathbf{H}(\hat{\mathbf{u}}_k)\|_2 = \frac{1}{p\|\mathbf{u}_k\|}\left\|\nabla^2 f(\mathbf{x}^* + p\mathbf{u}_k) - \nabla^2 f(\mathbf{x}^*) - \int_0^w u\frac{d^2\mathbf{G}}{du^2}du\right\|_2 \tag{C.14}$$

$$\leqslant \frac{M}{p\|\mathbf{u}_k\|}\|\mathbf{x}^* + p\mathbf{u}_k - \mathbf{x}^*\| + \frac{\left\|\int_0^w u\frac{d^2\mathbf{G}}{du^2}du\right\|_2}{p\|\mathbf{u}_k\|} \tag{C.15}$$

$$\leqslant M + \frac{\left(\int_0^w \left\|\frac{d^2\mathbf{G}}{du^2}\right\|_2^2 du\right)^{\frac{1}{2}}\left(\int_0^w u^2 du\right)^{\frac{1}{2}}}{w} \leqslant M + \frac{B_2 w}{\sqrt{3}} \leqslant M + \mathscr{O}(\varepsilon). \tag{C.16}$$

Note that in the last step, we used the Cauchy Schwarz inequality followed by the same bound $\left\|\frac{d^2\mathbf{G}}{du^2}\right\|_2 \leqslant B_2$ as in the steps following (B.7). For the case when $p\|\mathbf{u}_k\| \to 0$, the bound on $\|\mathbf{H}(\hat{\mathbf{u}}_k)\|_2$ can be evaluated by using the substitution $w = p\|\mathbf{u}_k\|$:

$$\|\mathbf{H}(\hat{\mathbf{u}}_k)\|_2 \leqslant \lim_{p\|\mathbf{u}_k\|\to 0}\frac{M}{p\|\mathbf{u}_k\|}\|\mathbf{x}^* + p\mathbf{u}_k - \mathbf{x}^*\| + \lim_{p\|\mathbf{u}_k\|\to 0}\frac{\left(\int_0^w \left\|\frac{d^2\mathbf{G}}{du^2}\right\|_2^2 du\right)^{\frac{1}{2}}\left(\int_0^w u^2 du\right)^{\frac{1}{2}}}{p\|\mathbf{u}_k\|} \tag{C.17}$$

$$\leqslant \lim_{w\to 0}\frac{M}{w}w + \lim_{w\to 0}\frac{\left(\int_0^w \left\|\frac{d^2\mathbf{G}}{du^2}\right\|_2^2 du\right)^{\frac{1}{2}}\left(\int_0^w u^2 du\right)^{\frac{1}{2}}}{w} \tag{C.18}$$

$$\leqslant M + \lim_{w\to 0}\frac{\left(\int_0^w \left\|\frac{d^2\mathbf{G}}{du^2}\right\|_2^2 du\right)^{\frac{1}{2}} w^{1/2}}{\sqrt{3}} = M + \lim_{w\to 0}\left(\int_0^w \left\|\frac{d^2\mathbf{G}}{du^2}\right\|_2^2 du\right)^{\frac{1}{2}}\lim_{w\to 0}\frac{w^{1/2}}{\sqrt{3}} = M. \tag{C.19}$$

Note that in the last step, we used $\lim_{w\to 0}\left(\int_0^w \left\|\frac{d^2\mathbf{G}}{du^2}\right\|_2^2 du\right)^{\frac{1}{2}} = \left(\int_0^1 \lim_{w\to 0} \mathbb{1}_{[0,w]} \left\|\frac{d^2\mathbf{G}}{du^2}\right\|_2^2 du\right)^{\frac{1}{2}} = 0$ by the dominated convergence theorem where $\mathbb{1}_{[0,w]}$ is the indicator function on $[0,w]$.

Hence for any eigenvectors $\mathbf{v}_i(0), \mathbf{v}_j(0)$ we have that

$$-M - \mathscr{O}(\varepsilon) \leqslant \langle \mathbf{v}_i(0), \mathbf{H}(\hat{\mathbf{u}}_k)\mathbf{v}_j(0)\rangle \leqslant M + \mathscr{O}(\varepsilon). \tag{C.20}$$

Using Assumptions **A2** and **A4**, for the stable subspace $\mathscr{E}_S$, we have the following bound on $\lambda_i(0)$:

$$\beta \leqslant \lambda_i(0) \leqslant L. \tag{C.21}$$

Similarly for the unstable subspace $\mathscr{E}_{US}$, we have the following bound on $\lambda_j(0)$ from Assumptions **A2** and **A4**:

$$-L \leqslant \lambda_j(0) \leqslant -\beta. \tag{C.22}$$

Now substituting these bounds into (C.9), (C.10), (C.11) and using the fact that $\|\mathbf{u}_k\| < \varepsilon$, we get the following bounds on the coefficients:

$$\left(1 - \alpha L - \frac{\alpha\varepsilon M}{2} - \mathscr{O}(\varepsilon^2)\right) \leqslant c_i^s(k) \leqslant \left(1 - \alpha\beta + \frac{\alpha\varepsilon M}{2} + \mathscr{O}(\varepsilon^2)\right) \tag{C.23}$$

$$\left(1 + \alpha\beta - \frac{\alpha\varepsilon M}{2} - \mathscr{O}(\varepsilon^2)\right) \leqslant c_j^{us}(k) \leqslant \left(1 + \alpha L + \frac{\alpha\varepsilon M}{2} + \mathscr{O}(\varepsilon^2)\right) \tag{C.24}$$

$$-\frac{\alpha\varepsilon ML}{2\delta} - \mathscr{O}(\varepsilon^2) \leqslant d_{i,l}(k) \leqslant \frac{\alpha\varepsilon ML}{2\delta} + \mathscr{O}(\varepsilon^2). \tag{C.25}$$

After establishing the bounds on the coefficients $c_i^s(k), c_j^{us}(k), d_{i,l}(k)$, we further analyze the recursive vector update equation (C.8) and induct it from $k=0$ to $k=K-1$ so as to obtain $\mathbf{u}_K$ in terms of $\mathbf{u}_0$:

$$\begin{aligned}\mathbf{u}_{k+1} = &\Bigg[\sum_{i\in\mathscr{N}_S} c_i^s(k)\mathbf{v}_i(0)\mathbf{v}_i(0)^T + \sum_{j\in\mathscr{N}_{US}} c_j^{us}(k)\mathbf{v}_j(0)\mathbf{v}_j(0)^T \\ &+ \sum_{i=1}^n \sum_{l\neq i}\left(d_{l,i}(k)\mathbf{v}_l(0)\mathbf{v}_i(0)^T + d_{i,l}(k)\mathbf{v}_i(0)\mathbf{v}_l(0)^T\right)\Bigg]\mathbf{u}_k + \mathscr{O}(\varepsilon^2)\mathbf{u}_k \end{aligned} \tag{C.26}$$

$$\begin{aligned}\implies \mathbf{u}_K = \prod_{k=0}^{K-1}\Bigg[&\mathscr{O}(\varepsilon^2) + \sum_{i\in\mathscr{N}_S} c_i^s(k)\mathbf{v}_i(0)\mathbf{v}_i(0)^T + \sum_{j\in\mathscr{N}_{US}} c_j^{us}(k)\mathbf{v}_j(0)\mathbf{v}_j(0)^T \\ &+ \sum_{i=1}^n \sum_{l\neq i}\left(d_{l,i}(k)\mathbf{v}_l(0)\mathbf{v}_i(0)^T + d_{i,l}(k)\mathbf{v}_i(0)\mathbf{v}_l(0)^T\right)\Bigg]\mathbf{u}_0. \end{aligned} \tag{C.27}$$

Observe that in the above expression, the vector $\mathbf{u}_K$ results from a product of K matrices. Each of these matrices comes from a linear combination of n^2 different matrices given by $\mathbf{v}_i(0)\mathbf{v}_i(0)^T$ for $i \in \mathscr{N}_S$, $\mathbf{v}_j(0)\mathbf{v}_j(0)^T$ for $j \in \mathscr{N}_{US}$, the cross terms $\mathbf{v}_l(0)\mathbf{v}_i(0)^T, \mathbf{v}_i(0)\mathbf{v}_l(0)^T$ with $i \neq l$ and in addition to this a matrix term of order $\mathscr{O}(\varepsilon^2)$.

Next, using the orthogonality of eigenvectors we obtain $\mathbf{v}_i(0)^T\mathbf{v}_j(0) = 0$ for $i \neq j$ and $\mathbf{v}_i(0)^T\mathbf{v}_j(0) = 1$ for $i = j$. Therefore by induction it can be readily inferred that the K matrix product is a linear

combination of the same n^2 matrices plus all the matrix error terms of the order $\mathscr{O}(\varepsilon^2)$ and above. Hence we rewrite (C.27) as follows:

$$\mathbf{u}_K = \prod_{k=0}^{K-1} \Big[\mathbf{A}_k + \mathbf{B}_k + \mathscr{O}(\varepsilon^2) \Big] \mathbf{u}_0, \tag{C.28}$$

where $\mathbf{A}_k = \sum_{i \in \mathscr{N}_S} c_i^s(k) \mathbf{v}_i(0) \mathbf{v}_i(0)^T + \sum_{j \in \mathscr{N}_{US}} c_j^{us}(k) \mathbf{v}_j(0) \mathbf{v}_j(0)^T$ and $\mathbf{B}_k = \sum_{i=1}^n \sum_{l \neq i} \Big(d_{l,i}(k) \mathbf{v}_l(0) \mathbf{v}_i(0)^T + d_{i,l}(k) \mathbf{v}_i(0) \mathbf{v}_l(0)^T \Big)$. From (C.25) the term $\mathbf{B}_k$ is of order $\mathscr{O}(\varepsilon)$. Therefore, this equation can be written more compactly as

$$\mathbf{u}_K = \prod_{k=0}^{K-1} \Big[\mathbf{A}_k + \varepsilon \mathbf{P}_k \Big] \mathbf{u}_0, \tag{C.29}$$

where $\varepsilon \mathbf{P}_k = \mathbf{B}_k + \mathscr{O}(\varepsilon^2)$.

Next we analyze the matrix product $\prod_{k=0}^{K-1} \Big[\mathbf{A}_k + \varepsilon \mathbf{P}_k \Big]$. Taking the norm of this product, followed by the supremum over k and using the triangle inequality yields

$$\begin{aligned}
\left\| \prod_{k=0}^{K-1} \Big[\mathbf{A}_k + \varepsilon \mathbf{P}_k \Big] \right\|_2 &\leqslant \prod_{k=0}^{K-1} \left\| \Big[\mathbf{A}_k + \varepsilon \mathbf{P}_k \Big] \right\|_2 && \text{(C.30)} \\
&\leqslant \prod_{k=0}^{K-1} \sup_{0 \leqslant k \leqslant K-1} \left\| \Big[\mathbf{A}_k + \varepsilon \mathbf{P}_k \Big] \right\|_2 && \text{(C.31)} \\
&\leqslant \prod_{k=0}^{K-1} \Big[\sup_{0 \leqslant k \leqslant K-1} \|\mathbf{A}_k\|_2 + \varepsilon \sup_{0 \leqslant k \leqslant K-1} \|\mathbf{P}_k\|_2 \Big] && \text{(C.32)} \\
&\leqslant \prod_{k=0}^{K-1} \Big[\|\mathbf{A}\|_2 + \varepsilon \|\mathbf{P}\|_2 \Big] = \Big(\|\mathbf{A}\|_2 + \varepsilon \|\mathbf{P}\|_2 \Big)^K, && \text{(C.33)}
\end{aligned}$$

where in the last step we have used the substitutions $\sup_{0 \leqslant k \leqslant K-1} \|\mathbf{A}_k\|_2 = \|\mathbf{A}\|_2$ and $\sup_{0 \leqslant k \leqslant K-1} \|\mathbf{P}_k\|_2 = \|\mathbf{P}\|_2$ for some arbitrary matrices $\mathbf{A}$ and $\mathbf{P}$.

Now observe that the product term on the right-hand side of (C.33) has a binomial expansion which can be written compactly as

$$\left\| \prod_{k=0}^{K-1} \Big[\mathbf{A}_k + \varepsilon \mathbf{P}_k \Big] \right\|_2 \leqslant \sum_{r=0}^{K} \binom{K}{r} (\varepsilon \|\mathbf{P}\|_2)^r \|\mathbf{A}\|_2^{K-r} = \|\mathbf{A}\|_2^K \left(1 + \varepsilon \frac{\|\mathbf{P}\|_2}{\|\mathbf{A}\|_2} \right)^K. \tag{C.34}$$

Next, consider the term $\left(1 + \varepsilon \frac{\|\mathbf{P}\|_2}{\|\mathbf{A}\|_2} \right)^K$ on the right-hand side of above bound. For the function $g_\omega(x) = (1+x)^\omega$ such that $\omega \in \mathbb{R}$, its expansion and the remainder term are given by

$$\begin{aligned}
(1+x)^\omega &= \sum_{k=0}^{\infty} \binom{\omega}{k} x^k && \text{(C.35)} \\
R_j(x) &= \int_0^x \frac{(x-z)^j}{j!} (j+1)! \binom{\omega}{j+1} (1+z)^{\omega-j-1} dz, && \text{(C.36)}
\end{aligned}$$

where we have that $\limsup_{j\to\infty} R_j(x) = 0$ for $|x| < 1$.

Therefore using this remainder expression for the term $\left(1+\varepsilon\frac{\|\mathbf{P}\|_2}{\|\mathbf{A}\|_2}\right)^K$ with $x = \varepsilon\frac{\|\mathbf{P}\|_2}{\|\mathbf{A}\|_2}$ we will have

$$R_1(x) = \int_0^x \frac{(x-z)}{1!} 2! \binom{K}{2} (1+z)^{K-2} dz \tag{C.37}$$

$$= K(K-1)\left(\frac{(1+x)^K}{K(K-1)} - \frac{1+x}{K-1} + \frac{1}{K}\right). \tag{C.38}$$

For $|x| < 1$ and $|Kx| \ll 1$, we can use the approximation $(1+x)^K \approx 1 + Kx + \binom{K}{2}x^2$. Substituting this approximation in (C.38), we get $R_1(x)$ as

$$R_1(x) \approx K(K-1)\left(\frac{1+Kx+\binom{K}{2}x^2}{K(K-1)} - \frac{1+x}{K-1} + \frac{1}{K}\right) = \frac{K(K-1)}{2}x^2 \tag{C.39}$$

$$R_1\left(\varepsilon\frac{\|\mathbf{P}\|_2}{\|\mathbf{A}\|_2}\right) \approx \frac{K(K-1)}{2}\left(\varepsilon\frac{\|\mathbf{P}\|_2}{\|\mathbf{A}\|_2}\right)^2 = \mathscr{O}\left((K\varepsilon)^2\right). \tag{C.40}$$

Hence for $\varepsilon\frac{\|\mathbf{P}\|_2}{\|\mathbf{A}\|_2} < 1$ and $K\varepsilon \ll 1$, we can substitute this bound in (C.34) as follows:

$$\left\|\prod_{k=0}^{K-1}\left[\mathbf{A}_k + \varepsilon\mathbf{P}_k\right]\right\|_2 \leqslant \|\mathbf{A}\|_2^K \left(1+\varepsilon\frac{\|\mathbf{P}\|_2}{\|\mathbf{A}\|_2}\right)^K \tag{C.41}$$

$$= \|\mathbf{A}\|_2^K \left(1 + K\varepsilon\frac{\|\mathbf{P}\|_2}{\|\mathbf{A}\|_2} + R_1\left(\varepsilon\frac{\|\mathbf{P}\|_2}{\|\mathbf{A}\|_2}\right)\right) \tag{C.42}$$

$$\approx \|\mathbf{A}\|_2^K \left(1 + K\varepsilon\frac{\|\mathbf{P}\|_2}{\|\mathbf{A}\|_2} + \mathscr{O}\left((K\varepsilon)^2\right)\right). \tag{C.43}$$

This approximate upper bound implies that the upper bound on the norm of matrix product $\prod_{k=0}^{K-1}\left[\mathbf{A}_k + \varepsilon\mathbf{P}_k\right]$ can be approximately expanded up to an ε precision term accompanied with a remainder term of $\mathscr{O}\left(\|\mathbf{A}\|_2^K (K\varepsilon)^2\right)$ as long as $K\varepsilon \ll 1$.

Next we obtain a lower bound on the inverse of the norm of matrix product $\prod_{k=0}^{K-1}\left[\mathbf{A}_k + \varepsilon\mathbf{P}_k\right]^{-1}$. Taking the inverse of the norm of this product, using the identities $\|\mathbf{Z}\|_2 \geqslant \left\|\mathbf{Z}^{-1}\right\|_2^{-1}$, $\left\|(\mathbf{I}+\mathbf{Z})^{-1}\right\|_2^{-1} \geqslant (1 - \|\mathbf{Z}\|_2)$, followed by taking the infimum over k yields

$$\left\|\prod_{k=0}^{K-1}\left[\mathbf{A}_k + \varepsilon\mathbf{P}_k\right]^{-1}\right\|_2^{-1} \geqslant \prod_{k=0}^{K-1}\left\|\mathbf{A}_k^{-1}\right\|_2^{-1}\left(1 - \varepsilon\left\|\mathbf{A}_k^{-1}\mathbf{P}_k\right\|_2\right) \tag{C.44}$$

$$\geqslant \prod_{k=0}^{K-1}\inf_{0\leqslant k\leqslant K-1}\left\|\mathbf{A}_k^{-1}\right\|_2^{-1}\left(1 - \varepsilon\sup_{0\leqslant k\leqslant K-1}\left\|\mathbf{A}_k^{-1}\mathbf{P}_k\right\|_2\right) \tag{C.45}$$

$$\geqslant \left(\inf_{0\leqslant k\leqslant K-1}\left\|\mathbf{A}_k^{-1}\right\|_2^{-1}\right)^K \left(1 - \varepsilon\sup_{0\leqslant k\leqslant K-1}\left\|\mathbf{A}_k^{-1}\mathbf{P}_k\right\|_2\right)^K. \tag{C.46}$$

Now repeating the previous analysis of the upper bound here will give the conclusion that the lower bound on inverse of the norm of matrix product $\prod_{k=0}^{K-1}\left[\mathbf{A}_k+\varepsilon\mathbf{P}_k\right]^{-1}$ can be approximately computed up to $K\varepsilon$ precision provided $\varepsilon\sup_{0\leqslant k\leqslant K-1}\left\|\mathbf{A}_k^{-1}\mathbf{P}_k\right\|_2<1$ and $K\varepsilon\ll 1$ if the step size $\alpha<\frac{1}{L}$. The reasoning for having $\alpha<\frac{1}{L}$ will be discussed in the subsequent section when we derive some feasible range for ε as well as the case where $\alpha\approx\frac{1}{L}$. In particular, the inequality (C.46) can be simplified further as

$$\left\|\prod_{k=0}^{K-1}\left[\mathbf{A}_k+\varepsilon\mathbf{P}_k\right]^{-1}\right\|_2^{-1}\geqslant\left\|\mathbf{A}^{-1}\right\|_2^{-K}\left(1-K\varepsilon\frac{\|\mathbf{P}\|_2}{\left\|\mathbf{A}^{-1}\right\|_2^{-1}}-\mathscr{O}\left((K\varepsilon)^2\right)\right) \tag{C.47}$$

for $K\varepsilon\ll 1$ and $\varepsilon\frac{\|\mathbf{P}\|_2}{\left\|\mathbf{A}^{-1}\right\|_2^{-1}}<1$ where we have that $\sup_{0\leqslant k\leqslant K-1}\left\|\mathbf{A}_k^{-1}\right\|_2=\left\|\mathbf{A}^{-1}\right\|_2$ and $\sup_{0\leqslant k\leqslant K-1}\|\mathbf{P}_k\|_2=\|\mathbf{P}\|_2$ for the matrices $\mathbf{A}$ and $\mathbf{P}$ used previously.

Now, if $\nu_n\leqslant\cdots\leqslant\nu_1$ are the absolute value of the eigenvalues of the matrix product $\prod_{k=0}^{K-1}\Big[\mathbf{A}_k+\varepsilon\mathbf{P}_k\Big]$, then using (C.43) and (C.47), we have the condition

$$\left\|\mathbf{A}^{-1}\right\|_2^{-K}\left(1-K\varepsilon\frac{\|\mathbf{P}\|_2}{\left\|\mathbf{A}^{-1}\right\|_2^{-1}}-\mathscr{O}\left((K\varepsilon)^2\right)\right)\leqslant\nu_n\leqslant\cdots\leqslant\nu_1\leqslant\|\mathbf{A}\|_2^K\left(1+K\varepsilon\frac{\|\mathbf{P}\|_2}{\|\mathbf{A}\|_2}+\mathscr{O}\left((K\varepsilon)^2\right)\right). \tag{C.48}$$

Therefore we can conclude that the matrix product (C.27) can be approximately computed up to $K\varepsilon$ precision provided $K\varepsilon\ll 1$, $\varepsilon\frac{\|\mathbf{P}\|_2}{\|\mathbf{A}\|_2}<1$ and $\varepsilon\frac{\|\mathbf{P}\|_2}{\left\|\mathbf{A}^{-1}\right\|_2^{-1}}<1$. At this point, we are interested in analyzing the matrix product in (C.27) only for iterations $K=\mathscr{O}(\frac{1}{\varepsilon})$. This is done so as to derive exit times and initial conditions for trajectories that can escape a strict saddle point in linear time. It is also remarked that we could have retained the higher-order terms $\mathscr{O}\left(\|\mathbf{A}\|_2^K(K\varepsilon)^r\right)$ in the above matrix product (C.46) if we wanted to analyze trajectories with polynomial or even exponential rates of escape. □

C.2 *Proof of Lemma 3.5*

Proof. For values of $K=\mathscr{O}(\frac{1}{\varepsilon})$ we explicitly compute the matrix product in (C.27) up to $K\varepsilon$ precision and drop all the higher order terms (ε^2 and above) that collectively act as a single remainder term of an approximate order $\mathscr{O}\left(\|\mathbf{A}\|_2^K(K\varepsilon)^2\right)$. From (C.25) we know that only the coefficients $d_{i,l}(k)$ are of order $\mathscr{O}(\varepsilon)$, hence we now expand (C.27) only up to first order in $d_{i,l}(k)$ to obtain the following

approximation:

$$\mathbf{u}_K \approx \tilde{\mathbf{u}}_K = \Bigg[\sum_{i\in\mathscr{N}_S} \Bigg(\prod_{k=0}^{K-1} c_i^s(k) \Bigg) \mathbf{v}_i(0)\mathbf{v}_i(0)^T + \sum_{j\in\mathscr{N}_{US}} \prod_{k=0}^{K-1} \Big(c_j^{us}(k) \Big) \mathbf{v}_j(0)\mathbf{v}_j(0)^T$$
$$+ \sum_{i\in\mathscr{N}_S} \sum_{l\in\mathscr{N}_S} \sum_{r=0}^{K-1} \Bigg(\prod_{k=0}^{r-1} c_i^s(k) \Bigg) d_{i,l}(r) \Bigg(\prod_{k=r+1}^{K-1} c_l^s(k) \Bigg) \Big(\mathbf{v}_l(0)\mathbf{v}_i(0)^T + \mathbf{v}_i(0)\mathbf{v}_l(0)^T \Big)$$
$$+ \sum_{i\in\mathscr{N}_S} \sum_{l\in\mathscr{N}_{US}} \sum_{r=0}^{K-1} \Bigg(\prod_{k=0}^{r-1} c_i^s(k) \Bigg) d_{i,l}(r) \Bigg(\prod_{k=r+1}^{K-1} c_l^{us}(k) \Bigg) \Big(\mathbf{v}_l(0)\mathbf{v}_i(0)^T + \mathbf{v}_i(0)\mathbf{v}_l(0)^T \Big)$$
$$+ \sum_{i\in\mathscr{N}_{US}} \sum_{l\in\mathscr{N}_S} \sum_{r=0}^{K-1} \Bigg(\prod_{k=0}^{r-1} c_i^{us}(k) \Bigg) d_{i,l}(r) \Bigg(\prod_{k=r+1}^{K-1} c_l^s(k) \Bigg) \Big(\mathbf{v}_l(0)\mathbf{v}_i(0)^T + \mathbf{v}_i(0)\mathbf{v}_l(0)^T \Big)$$
$$+ \sum_{i\in\mathscr{N}_{US}} \sum_{l\in\mathscr{N}_{US}} \sum_{r=0}^{K-1} \Bigg(\prod_{k=0}^{r-1} c_i^{us}(k) \Bigg) d_{i,l}(r) \Bigg(\prod_{k=r+1}^{K-1} c_l^{us}(k) \Bigg) \Big(\mathbf{v}_l(0)\mathbf{v}_i(0)^T + \mathbf{v}_i(0)\mathbf{v}_l(0)^T \Big) \Bigg] \mathbf{u}_0, \quad \text{(C.49)}$$

where we have that $\tilde{\mathbf{u}}_K$ as the ε approximate trajectory.

Next we express $\mathbf{u}_0$ as the sum of projections onto the stable subspace and unstable subspace of $\nabla^2 f(\mathbf{x}^*)$ as follows:

$$\mathbf{u}_0 = \varepsilon \sum_{i\in\mathscr{N}_S} \theta_i^s \mathbf{v}_i(0) + \varepsilon \sum_{j\in\mathscr{N}_{US}} \theta_j^{us} \mathbf{v}_j(0) \quad \text{(C.50)}$$

$$\sum_{i\in\mathscr{N}_S} (\theta_i^s)^2 + \sum_{j\in\mathscr{N}_{US}} (\theta_j^{us})^2 = 1, \quad \text{(C.51)}$$

where $\varepsilon\theta_i^s = \langle \mathbf{u}_0, \mathbf{v}_i(0) \rangle$, $\varepsilon\theta_j^{us} = \langle \mathbf{u}_0, \mathbf{v}_j(0) \rangle$ with $\mathbf{v}_i(0) \in \mathscr{E}_S$ and $\mathbf{v}_j(0) \in \mathscr{E}_{US}$ respectively. Observe that (C.50) has an ε multiplier because $\|\mathbf{u}_0\| = \varepsilon$. This is due to the fact that $\mathbf{u}_0 + \mathbf{x}^* = \mathbf{x}_0$ and $\mathbf{x}_0 \in \bar{\mathscr{B}}_\varepsilon(\mathbf{x}^*) \backslash \mathscr{B}_\varepsilon(\mathbf{x}^*)$.

Now for all i and j, the Hessian $\nabla^2 f(\mathbf{x}^*)$ can have eigenvectors $\mathbf{v}_i(0)$ and $\mathbf{v}_j(0)$ as well as $-\mathbf{v}_i(0)$ and $-\mathbf{v}_j(0)$. Therefore for the sake of analysis, the signs with these eigenvectors are chosen such that the respective coefficients θ_i^s and θ_j^{us} are positive for all i and j. It is easy to show that such a choice always exists for all i and j because if $\langle \mathbf{u}_0, \mathbf{v}_i(0) \rangle > 0$ then $\langle \mathbf{u}_0, -\mathbf{v}_i(0) \rangle < 0$ and vice versa for any i (and analogously for the index j).

Finally substituting $\mathbf{u}_0$ in (C.49), we get the following result for $\mathbf{u}_K$:

$$\mathbf{u}_K \approx \tilde{\mathbf{u}}_K = \varepsilon \sum_{i\in\mathscr{N}_S} \Bigg(\prod_{k=0}^{K-1} c_i^s(k)\theta_i^s + \sum_{l\in\mathscr{N}_S} \sum_{r=0}^{K-1} \prod_{k=0}^{r-1} c_i^s(k) d_{i,l}(r) \prod_{k=r+1}^{K-1} c_l^s(k)\theta_l^s$$
$$+ \sum_{l\in\mathscr{N}_{US}} \sum_{r=0}^{K-1} \prod_{k=0}^{r-1} c_i^s(k) d_{i,l}(r) \prod_{k=r+1}^{K-1} c_l^{us}(k)\theta_l^{us} \Bigg) \mathbf{v}_i(0) +$$
$$\varepsilon \sum_{j\in\mathscr{N}_{US}} \Bigg(\prod_{k=0}^{K-1} c_j^{us}(k)\theta_j^{us} + \sum_{l\in\mathscr{N}_S} \sum_{r=0}^{K-1} \prod_{k=0}^{r-1} c_j^{us}(k) d_{j,l}(r) \prod_{k=r+1}^{K-1} c_l^s(k)\theta_l^s$$
$$+ \sum_{l\in\mathscr{N}_{US}} \sum_{r=0}^{K-1} \prod_{k=0}^{r-1} c_j^{us}(k) d_{j,l}(r) \prod_{k=r+1}^{K-1} c_l^{us}(k)\theta_l^{us} \Bigg) \mathbf{v}_j(0). \quad \text{(C.52)}$$

C.2.1 *Bounds on ε:.* Recall that from (C.43) we established that the first-order approximation of the matrix product (C.27) is only valid for $\varepsilon \frac{\|\mathbf{P}\|_2}{\|\mathbf{A}\|_2} < 1$ and $K\varepsilon \ll 1$. Next, from (C.33) we have that $\sup_{0\leqslant k\leqslant K-1} \|\mathbf{A}_k\|_2 = \|\mathbf{A}\|_2$ and $\sup_{0\leqslant k\leqslant K-1} \|\mathbf{P}_k\|_2 = \|\mathbf{P}\|_2$.

From (C.28) we have the following:

$$\mathbf{A}_k = \sum_{i\in\mathcal{N}_S} c_i^s(k)\mathbf{v}_i(0)\mathbf{v}_i(0)^T + \sum_{j\in\mathcal{N}_{US}} c_j^{us}(k)\mathbf{v}_j(0)\mathbf{v}_j(0)^T, \text{ and} \tag{C.53}$$

$$\mathbf{B}_k = \sum_{i=1}^{n}\sum_{l\neq i}\Big(d_{l,i}(k)\mathbf{v}_l(0)\mathbf{v}_i(0)^T + d_{i,l}(k)\mathbf{v}_i(0)\mathbf{v}_l(0)^T\Big), \tag{C.54}$$

with $\varepsilon\mathbf{P}_k = \mathbf{B}_k + \mathscr{O}(\varepsilon^2)$.

Observe that $\mathbf{A}_k$ is a matrix in its spectral decomposed form where the coefficients $c_i^s(k)$ and $c_i^{us}(k)$ correspond to the eigenvalues of $\mathbf{A}_k$. Therefore applying the bounds (C.23) and (C.24) we have the following result:

$$\|\mathbf{A}\|_2 = \sup_{0\leqslant k\leqslant K-1} \|\mathbf{A}_k\|_2 \tag{C.55}$$

$$= \sup_{0\leqslant k\leqslant K-1}\left\{\max_{i\in\mathcal{N}_S, j\in\mathcal{N}_{US}}\{c_i^s(k), c_j^{us}(k)\}\right\} \tag{C.56}$$

$$= \left(1 + \alpha L + \frac{\alpha\varepsilon M}{2} + \mathscr{O}(\varepsilon^2)\right). \tag{C.57}$$

Next, taking the norm of both sides of $\varepsilon\mathbf{P}_k = \mathbf{B}_k + \mathscr{O}(\varepsilon^2)$, taking supremum over k followed by the triangle inequality and then using (C.25) we get the following upper bound:

$$\sup_{0\leqslant k\leqslant K-1} \|\varepsilon\mathbf{P}_k\|_2 = \sup_{0\leqslant k\leqslant K-1} \big\|\mathbf{B}_k + \mathscr{O}(\varepsilon^2)\big\|_2 \tag{C.58}$$

$$\leqslant \sup_{0\leqslant k\leqslant K-1} \|\mathbf{B}_k\|_2 + \mathscr{O}(\varepsilon^2) \tag{C.59}$$

$$\leqslant \sum_{i=1}^{n}\sum_{l\neq i}\left(\sup_{0\leqslant k\leqslant K-1}\big\|d_{l,i}(k)\mathbf{v}_l(0)\mathbf{v}_i(0)^T\big\|_2 + \sup_{0\leqslant k\leqslant K-1}\big\|d_{i,l}(k)\mathbf{v}_i(0)\mathbf{v}_l(0)^T\big\|_2\right) + \mathscr{O}(\varepsilon^2) \tag{C.60}$$

$$\leqslant \sum_{i=1}^{n}\sum_{l\neq i}\left(\sup_{0\leqslant k\leqslant K-1}\big\|d_{l,i}(k)\mathbf{v}_l(0)\mathbf{v}_i(0)^T\big\|_F + \sup_{0\leqslant k\leqslant K-1}\big\|d_{i,l}(k)\mathbf{v}_i(0)\mathbf{v}_l(0)^T\big\|_F\right) + \mathscr{O}(\varepsilon^2) \tag{C.61}$$

$$= \sum_{i=1}^{n}\sum_{l\neq i}\left(\sup_{0\leqslant k\leqslant K-1}|d_{l,i}(k)| + \sup_{0\leqslant k\leqslant K-1}|d_{i,l}(k)|\right) + \mathscr{O}(\varepsilon^2) \tag{C.62}$$

$$\leqslant \frac{\alpha\varepsilon M L n^2}{\delta} + \mathscr{O}(\varepsilon^2), \tag{C.63}$$

where in the last couple of steps we used the following properties of any matrix $\mathbf{Z}$: $\|\mathbf{Z}\|_2 \leqslant \|\mathbf{Z}\|_F$, and $\|\mathbf{Z}\|_F = \sqrt{\mathrm{tr}(\mathbf{Z}\mathbf{Z}^T)}$.

Now we require that $\varepsilon \frac{\|\mathbf{P}\|_2}{\|\mathbf{A}\|_2} < 1$. Using (C.57), this condition becomes

$$\sup_{0\leqslant k\leqslant K-1} \|\varepsilon \mathbf{P}_k\|_2 = \|\varepsilon \mathbf{P}\|_2 < \|\mathbf{A}\|_2 = \left(1+\alpha L+\frac{\alpha\varepsilon M}{2}+\mathscr{O}(\varepsilon^2)\right). \tag{C.64}$$

Therefore to obtain a bound on ε we can utilize (C.63) and set this condition as follows:

$$\sup_{0\leqslant k\leqslant K-1} \|\varepsilon \mathbf{P}_k\|_2 = \|\varepsilon \mathbf{P}\|_2 \leqslant \frac{\alpha\varepsilon MLn^2}{\delta}+\mathscr{O}(\varepsilon^2) < \|\mathbf{A}\|_2 = \left(1+\alpha L+\frac{\alpha\varepsilon M}{2}+\mathscr{O}(\varepsilon^2)\right) \tag{C.65}$$

$$\frac{\alpha\varepsilon MLn^2}{\delta}-\frac{\alpha\varepsilon M}{2} < 1+\alpha L+\mathscr{O}(\varepsilon^2) \tag{C.66}$$

$$\varepsilon < \frac{2\delta(1+\alpha L)}{\alpha M(2Ln^2-\delta)}+\mathscr{O}(\varepsilon^2). \tag{C.67}$$

Note that this condition on ε is sufficient but may not be necessary since we are using an upper bound on $\sup_{0\leqslant k\leqslant K-1}\|\varepsilon\mathbf{P}_k\|_2$ from (C.63) as a lower bound for $\|\mathbf{A}\|_2$. Hence, the inequality may shrink the feasible set for ε making it a sufficient condition but not necessary.

Having established a range for ε from the upper bound (C.43), we utilize the lower bound (C.46) to get the complete feasible range for ε. From the bound (C.46) we need that $\varepsilon \sup_{0\leqslant k\leqslant K-1}\left\|\mathbf{A}_k^{-1}\mathbf{P}_k\right\|_2 < 1$. Now for this particular condition to work, $\mathbf{A}_k$ should not have eigenvalues close to 0 or of order $\mathscr{O}(\varepsilon)$. Recall that from (C.53), $\mathbf{A}_k$ has its eigenvalues as $c_i^s(k)$ and $c_j^{us}(k)$ which are bounded by the inequalities in (C.23), (C.24). For $\alpha \approx \frac{1}{L}$, the lower bound in (C.23) becomes $\mathscr{O}(\varepsilon)$. Hence we analyze the two cases corresponding to different ranges of α separately.

C.2.2 *Case 1—*$\alpha \in \left(0, \frac{1}{L}-\mathscr{O}(\varepsilon)\right]$*:.* For this case, we can use the condition $\varepsilon \sup_{0\leqslant k\leqslant K-1}\left\|\mathbf{A}_k^{-1}\mathbf{P}_k\right\|_2 < 1$ in (C.46). To obtain a certain feasible range on ε, this condition can be set as follows:

$$\varepsilon \sup_{0\leqslant k\leqslant K-1}\left\|\mathbf{A}_k^{-1}\mathbf{P}_k\right\|_2 < \sup_{0\leqslant k\leqslant K-1}\left\|\mathbf{A}_k^{-1}\right\|_2 \sup_{0\leqslant k\leqslant K-1}\|\varepsilon\mathbf{P}_k\|_2 < 1 \tag{C.68}$$

$$\sup_{0\leqslant k\leqslant K-1}\left\{\max_{i\in\mathcal{N}_S, j\in\mathcal{N}_{US}}\left\{\frac{1}{c_i^s(k)},\frac{1}{c_j^{us}(k)}\right\}\right\}\sup_{0\leqslant k\leqslant K-1}\|\varepsilon\mathbf{P}_k\|_2 < 1 \tag{C.69}$$

$$\left(1-\alpha L-\frac{\alpha\varepsilon M}{2}-\mathscr{O}(\varepsilon^2)\right)^{-1}\left(\frac{\alpha\varepsilon MLn^2}{\delta}+\mathscr{O}(\varepsilon^2)\right) < 1 \tag{C.70}$$

$$\frac{2\delta(1-\alpha L)}{\alpha M(2Ln^2+\delta)}+\mathscr{O}(\varepsilon^2) > \varepsilon. \tag{C.71}$$

Note that this condition on ε is sufficient but may not be necessary.
Moreover, combining the conditions (C.67) and (C.71) with (B.26) we get the following necessary bound:

$$\varepsilon < \min\left\{\inf_{\|\mathbf{u}\|=1}\left(\limsup_{j\to\infty}\sqrt[j]{\frac{r_j(\mathbf{u})}{j!}}\right)^{-1}, \frac{2\delta(1-\alpha L)}{\alpha M(2Ln^2+\delta)}+\mathscr{O}(\varepsilon^2)\right\}. \tag{C.72}$$

Finally it is also required to have $K\varepsilon \ll 1$ or $K \ll \frac{1}{\varepsilon}$. Therefore this condition implies

$$K = \mathscr{O}\left(\frac{1}{\varepsilon}\right). \tag{C.73}$$

C.2.3 *Case 2—*$\alpha \in \left(\frac{1}{L} - \mathscr{O}(\varepsilon), \frac{1}{L}\right]$: . For this case, observe that the lower bound in (C.46) is of order $\mathscr{O}(\varepsilon^K)$. Further simplifying this lower bound and taking the infimum term inside, we obtain the following:

$$\left\|\prod_{k=0}^{K-1}\left[\mathbf{A}_k + \varepsilon\mathbf{P}_k\right]\right\|_2 \geqslant \left(\inf_{0\leqslant k\leqslant K-1}\left\|\mathbf{A}_k^{-1}\right\|_2^{-1}\right)^K \left(1 - \varepsilon \sup_{0\leqslant k\leqslant K-1}\left\|\mathbf{A}_k^{-1}\mathbf{P}_k\right\|_2\right)^K \tag{C.74}$$

$$\geqslant \left(\inf_{0\leqslant k\leqslant K-1}\left\|\mathbf{A}_k^{-1}\right\|_2^{-1}\right)^K \left(1 - \sup_{0\leqslant k\leqslant K-1}\left\|\mathbf{A}_k^{-1}\right\|_2 \sup_{0\leqslant k\leqslant K-1}\|\varepsilon\mathbf{P}_k\|_2\right)^K \tag{C.75}$$

$$\geqslant \left(\inf_{0\leqslant k\leqslant K-1}\left\|\mathbf{A}_k^{-1}\right\|_2^{-1} - \frac{\sup_{0\leqslant k\leqslant K-1}\left\|\mathbf{A}_k^{-1}\right\|_2}{\sup_{0\leqslant k\leqslant K-1}\left\|\mathbf{A}_k^{-1}\right\|_2} \sup_{0\leqslant k\leqslant K-1}\|\varepsilon\mathbf{P}_k\|_2\right)^K \tag{C.76}$$

$$\geqslant \left(\left|\left(1 - \alpha L - \frac{\alpha\varepsilon M}{2} - \mathscr{O}(\varepsilon^2)\right)\right| - \left(\frac{\alpha\varepsilon M L n^2}{\delta} + \mathscr{O}(\varepsilon^2)\right)\right)^K. \tag{C.77}$$

Now for $\alpha = \frac{1}{L}$, the above lower bound will be $(C\varepsilon)^K$ where C is some constant. Therefore, for this lower bound to converge to 0 for large K we must necessarily have $C\varepsilon < 1$ which implies

$$\left|\frac{\varepsilon M}{2L} - \frac{\varepsilon M n^2}{\delta} + \mathscr{O}(\varepsilon^2)\right| < 1 \tag{C.78}$$

$$\frac{1}{\frac{Mn^2}{\delta} - \frac{M}{2L}} + \mathscr{O}(\varepsilon^2) > \varepsilon \tag{C.79}$$

$$\frac{2L\delta}{M(2Ln^2 - \delta)} + \mathscr{O}(\varepsilon^2) > \varepsilon. \tag{C.80}$$

Finally, combining this condition on ε with (C.67) and (B.26) for $\alpha = \frac{1}{L}$, we get that

$$\varepsilon < \min\left\{\inf_{\|\mathbf{u}\|=1}\left(\limsup_{j\to\infty}\sqrt[j]{\frac{r_j(\mathbf{u})}{j!}}\right)^{-1}, \frac{4L\delta}{M(2Ln^2-\delta)} + \mathscr{O}(\varepsilon^2), \frac{2L\delta}{M(2Ln^2-\delta)} + \mathscr{O}(\varepsilon^2)\right\} \tag{C.81}$$

$$\varepsilon < \min\left\{\inf_{\|\mathbf{u}\|=1}\left(\limsup_{j\to\infty}\sqrt[j]{\frac{r_j(\mathbf{u})}{j!}}\right)^{-1}, \frac{2L\delta}{M(2Ln^2-\delta)} + \mathscr{O}(\varepsilon^2)\right\}. \tag{C.82}$$

The condition $K = \mathscr{O}\left(\frac{1}{\varepsilon}\right)$ is still required to hold. □

D. Lower bounds on the distance between $\mathbf{x}_K$ and $\mathbf{x}^*$

D.1 *Proof of Lemma 3.6*

Proof. An approximate equation for $\mathbf{u}_K$ in terms of $\mathbf{u}_0$ is given by (C.52). This approximation holds for all values of K from 1 to K_{exit}, where K_{exit} denotes the iteration number of escape from $\mathscr{B}_\varepsilon(\mathbf{x}^*)$. Formally K_{exit} can be expressed as

$$K_{exit} = \inf_{K\geqslant 1}\left\{K \;\middle|\; \|\tilde{\mathbf{u}}_K\|^2 > \varepsilon^2\right\}, \tag{D.1}$$

where the squared norm is used for the sake of simplifying subsequent analysis involving lower bounds. However, the sequence $\{\tilde{\mathbf{u}}_K\}_{K=0}^{K_{exit}}$ cannot be determined solely from the initialization $\mathbf{u}_0$. To uniquely determine any $\tilde{\mathbf{u}}_K$, we still need to know the coefficient terms $c_i^s(k)$, $c_j^{us}(k)$ and $d_{l,i}(k)$ for all values of k from 0 to $K-1$. The only information available in this regard is the bound on these coefficients from (C.23), (C.24) and (C.25). Therefore it becomes impossible to predetermine the entire sequence $\{\tilde{\mathbf{u}}_K\}_{K=0}^{K_{exit}}$ just based on the knowledge of $\mathbf{u}_0$.

To circumvent this problem, we introduce a set S_ε which is the set of all possible ε precision trajectories generated by the approximate equation (C.52). Recall that while deriving the approximation (C.52), we expanded terms appearing in the product (C.27) only up to order $\mathscr{O}(\varepsilon)$; hence we can call these approximate sequences as ε precision trajectories with respect to $\mathbf{x}^*$. For a fixed initialization of $\mathbf{u}_0$, the set S_ε is given by

$$S_\varepsilon = \left\{ \{\tilde{\mathbf{u}}_K^\tau\}_{K=1}^{K_{exit}^\tau} \;\middle|\; \mathbf{u}_0 \right\}, \tag{D.2}$$

where each possible ε precision trajectory is parameterized by some $\tau \in \mathbb{R}$, K_{exit}^τ is the escape iteration for the τ-parameterized ε-precision trajectory and $\tilde{\mathbf{u}}_K^\tau$ satisfies (C.52) for every τ. Note that τ varies with variations in the sequence $\left\{ \{c_i^s(k), c_j^{us}(k), d_{l,i}(k)\}_{k=0}^{K-1} \right\}_{K=1}^{K_{exit}}$ which are in turn controlled by variations in the coefficient terms from the bounds (C.23), (C.24) and (C.25). Since the set S_ε contains all possible ε precision trajectories, the actual ε-precise trajectory that the radial vector $\mathbf{u}_K$ takes inside the ball $\mathscr{B}_\varepsilon(\mathbf{x}^*)$ will also belong to the set S_ε. Let this actual ε precision trajectory be parameterized by some $\tau = \omega$. Therefore we have that

$$\{\tilde{\mathbf{u}}_K^\omega\}_{K=1}^{K_{exit}^\omega} \in S_\varepsilon. \tag{D.3}$$

Moreover, $\tilde{\mathbf{u}}_K^\omega$ satisfies the approximate equation (C.52). Next using (D.1), we can write the escape iteration for the τ-parameterized ε precision trajectory as

$$K_{exit}^\tau = \inf_{K \geqslant 1} \left\{ K \;\middle|\; \|\tilde{\mathbf{u}}_K^\tau\|^2 > \varepsilon^2 \right\}. \tag{D.4}$$

We now define a quantity K^ι such that

$$K^\iota = \inf_{K \geqslant 1} \left\{ K \;\middle|\; \inf_\tau \left\{ \|\tilde{\mathbf{u}}_K^\tau\|^2 \right\} > \varepsilon^2 \right\}. \tag{D.5}$$

D.1.1 *Claim of the lemma:*.

$$K^\iota \geqslant \sup_\tau \left\{ K_{exit}^\tau \right\} = \sup_\tau \inf_{K \geqslant 1} \left\{ K \;\middle|\; \|\tilde{\mathbf{u}}_K^\tau\|^2 > \varepsilon^2 \right\}. \tag{D.6}$$

Proof by contradiction: Let us assume that for some $\tau = a$ the escape iteration K_{exit}^a is such that $K_{exit}^a > K^\iota$. From the definition of K^ι in (D.5), K^ι is the smallest iteration such that $\inf_\tau \left\{ \left\| \tilde{\mathbf{u}}_{K^\iota}^\tau \right\|^2 \right\} > \varepsilon^2$. This implies $\left\| \tilde{\mathbf{u}}_{K^\iota}^a \right\|^2 > \varepsilon^2$. However, this is not possible since it contradicts the definition of infimum from (D.4) for $\tau = a$. Therefore we must have $K_{exit}^a \leqslant K^\iota$ and this should hold for any a. Hence, we must have $K^\iota \geqslant \sup_\tau \left\{ K_{exit}^\tau \right\}$.

Since the actual ε-precise trajectory given by $\{\tilde{\mathbf{u}}_K^{\omega}\}_{K=1}^{K_{exit}^{\omega}}$ belongs to the τ-parameterized set S_ε, hence $K_{exit}^{\omega} \leqslant K^{\iota}$. Therefore it is sufficient to develop an upper bound on K^{ι} in order to draw conclusions about K_{exit}^{ω}. In the subsequent section, we analyze the lower bound on $\|\tilde{\mathbf{u}}_K\|^2$ to obtain this K^{ι}. □

D.2 *Proof of Theorem 3.2*

Proof.

Taking the norm squared on both sides of (C.52) we get the following:

$$\begin{aligned}\|\tilde{\mathbf{u}}_K\|^2 = \varepsilon^2 \sum_{i\in\mathscr{N}_S}\Bigg(&\underbrace{\prod_{k=0}^{K-1} c_i^s(k)\theta_i^s}_{T_1} + \underbrace{\sum_{l\in\mathscr{N}_S}\sum_{r=0}^{K-1}\prod_{k=0}^{r-1} c_i^s(k)d_{i,l}(r)\prod_{k=r+1}^{K-1} c_l^s(k)\theta_l^s}_{T_2} \\ &+ \underbrace{\sum_{l\in\mathscr{N}_{US}}\sum_{r=0}^{K-1}\prod_{k=0}^{r-1} c_i^s(k)d_{i,l}(r)\prod_{k=r+1}^{K-1} c_l^{us}(k)\theta_l^{us}}_{T_3}\Bigg)^2 + \\ \varepsilon^2 \sum_{j\in\mathscr{N}_{US}}\Bigg(&\underbrace{\prod_{k=0}^{K-1} c_j^{us}(k)\theta_j^{us}}_{T_4} + \underbrace{\sum_{l\in\mathscr{N}_S}\sum_{r=0}^{K-1}\prod_{k=0}^{r-1} c_j^{us}(k)d_{j,l}(r)\prod_{k=r+1}^{K-1} c_l^s(k)\theta_l^s}_{T_5} \\ &+ \underbrace{\sum_{l\in\mathscr{N}_{US}}\sum_{r=0}^{K-1}\prod_{k=0}^{r-1} c_j^{us}(k)d_{j,l}(r)\prod_{k=r+1}^{K-1} c_l^{us}(k)\theta_l^{us}}_{T_6}\Bigg)^2 \qquad \text{(D.7)}\end{aligned}$$

$$= \varepsilon^2\Bigg(\sum_{i\in\mathscr{N}_S}(T_1+T_2+T_3)^2 + \sum_{j\in\mathscr{N}_{US}}(T_4+T_5+T_6)^2\Bigg). \qquad \text{(D.8)}$$

Now this equation is satisfied by $\tilde{\mathbf{u}}_K^{\tau}$ for every τ. Hence for any given τ we can write

$$\|\tilde{\mathbf{u}}_K^{\tau}\|^2 = \varepsilon^2\Bigg(\sum_{i\in\mathscr{N}_S}(T_1(\tau)+T_2(\tau)+T_3(\tau))^2 + \sum_{j\in\mathscr{N}_{US}}(T_4(\tau)+T_5(\tau)+T_6(\tau))^2\Bigg), \qquad \text{(D.9)}$$

where τ varies with variations in the sequence $\left\{\{c_i^s(k), c_j^{us}(k), d_{l,i}(k)\}_{k=0}^{K-1}\right\}_{K=1}^{K_{exit}}$.

Using (C.23), (C.24) and (C.25) we get the bounds on these coefficient product terms from $T_1(\tau)$ to $T_6(\tau)$. Starting with the term $T_1(\tau)$ we have that

$$\inf_{\tau} T_1(\tau) = \prod_{k=0}^{K-1}\inf_{\tau}\left\{c_i^s(k)\right\}\theta_i^s = \left(1-\alpha L - \frac{\alpha\varepsilon M}{2} - \mathscr{O}(\varepsilon^2)\right)^K \theta_i^s, \text{ and} \qquad \text{(D.10)}$$

$$\sup_{\tau} T_1(\tau) = \prod_{k=0}^{K-1}\sup_{\tau}\left\{c_i^s(k)\right\}\theta_i^s = \left(1-\alpha\beta + \frac{\alpha\varepsilon M}{2} + \mathscr{O}(\varepsilon^2)\right)^K \theta_i^s \qquad \text{(D.11)}$$

for positive $c_i^s(k)$. Next for the term $T_2(\tau)$, first consider the lower bound

$$\inf_\tau T_2(\tau) \geqslant \sum_{l\in\mathcal{N}_S}\sum_{r=0}^{K-1} \inf_\tau \left\{ d_{i,l}(r)\prod_{k=0}^{r-1} c_i^s(k) \prod_{k=r+1}^{K-1} c_l^s(k) \right\} \theta_l^s \tag{D.12}$$

$$\geqslant \sum_{l\in\mathcal{N}_S}\sum_{r=0}^{K-1} -\sup_\tau \left\{ |d_{i,l}(r)| \right\} \sup_\tau \left\{ \prod_{k=0}^{r-1} c_i^s(k) \prod_{k=r+1}^{K-1} c_l^s(k) \right\} \theta_l^s \tag{D.13}$$

$$\geqslant \sum_{l\in\mathcal{N}_S}\sum_{r=0}^{K-1} -\left(\frac{\alpha\varepsilon ML}{2\delta} + \mathcal{O}(\varepsilon^2) \right) \sup_\tau \left\{ \prod_{k=0}^{r-1} c_i^s(k) \prod_{k=r+1}^{K-1} c_l^s(k) \right\} \theta_l^s \tag{D.14}$$

$$= \sum_{l\in\mathcal{N}_S}\sum_{r=0}^{K-1} -\left(\frac{\alpha\varepsilon ML}{2\delta} + \mathcal{O}(\varepsilon^2) \right) \prod_{k=0}^{r-1} \sup_\tau \left\{ c_i^s(k) \right\} \prod_{k=r+1}^{K-1} \sup_\tau \left\{ c_l^s(k) \right\} \theta_l^s \tag{D.15}$$

$$= -K\left(1 - \alpha\beta + \frac{\alpha\varepsilon M}{2} + \mathcal{O}(\varepsilon^2) \right)^{K-1} \left(\frac{\alpha\varepsilon ML}{2\delta} + \mathcal{O}(\varepsilon^2) \right) \sum_{l\in\mathcal{N}_S} \theta_l^s, \tag{D.16}$$

where we have $c_i^s(k) \geqslant 0$ for all i and k. The upper bound on $T_2(\tau)$ is as follows:

$$\sup_\tau T_2(\tau) \leqslant \sum_{l\in\mathcal{N}_S}\sum_{r=0}^{K-1} \sup_\tau \left\{ d_{i,l}(r)\prod_{k=0}^{r-1} c_i^s(k) \prod_{k=r+1}^{K-1} c_l^s(k) \right\} \theta_l^s \tag{D.17}$$

$$\leqslant \sum_{l\in\mathcal{N}_S}\sum_{r=0}^{K-1} \sup_\tau \left\{ |d_{i,l}(r)| \right\} \sup_\tau \left\{ \prod_{k=0}^{r-1} c_i^s(k) \prod_{k=r+1}^{K-1} c_l^s(k) \right\} \theta_l^s \tag{D.18}$$

$$= \sum_{l\in\mathcal{N}_S}\sum_{r=0}^{K-1} \left(\frac{\alpha\varepsilon ML}{2\delta} + \mathcal{O}(\varepsilon^2) \right) \sup_\tau \left\{ \prod_{k=0}^{r-1} c_i^s(k) \prod_{k=r+1}^{K-1} c_l^s(k) \right\} \theta_l^s \tag{D.19}$$

$$= K\left(1 - \alpha\beta + \frac{\alpha\varepsilon M}{2} + \mathcal{O}(\varepsilon^2) \right)^{K-1} \left(\frac{\alpha\varepsilon ML}{2\delta} + \mathcal{O}(\varepsilon^2) \right) \sum_{l\in\mathcal{N}_S} \theta_l^s. \tag{D.20}$$

For the term $T_3(\tau)$, first consider the lower bound

$$\inf_{\tau} T_3(\tau) \geqslant \sum_{l\in\mathcal{N}_{US}} \sum_{r=0}^{K-1} \inf_{\tau} \left\{ \prod_{k=0}^{r-1} c_i^s(k) d_{i,l}(r) \prod_{k=r+1}^{K-1} c_l^{us}(k) \right\} \theta_l^{us} \tag{D.21}$$

$$\geqslant \sum_{l\in\mathcal{N}_{US}} \sum_{r=0}^{K-1} -\sup_{\tau} \left\{ |d_{i,l}(r)| \right\} \sup \left\{ \prod_{k=0}^{r-1} c_i^s(k) \prod_{k=r+1}^{K-1} c_l^{us}(k) \right\} \theta_l^{us} \tag{D.22}$$

$$= \sum_{l\in\mathcal{N}_{US}} \sum_{r=0}^{K-1} -\left(\frac{\alpha\varepsilon ML}{2\delta} + \mathcal{O}(\varepsilon^2) \right) \prod_{k=0}^{r-1} \sup_{\tau} \left\{ c_i^s(k) \right\} \prod_{k=r+1}^{K-1} \sup_{\tau} \left\{ c_l^{us}(k) \right\} \theta_l^{us} \tag{D.23}$$

$$= \sum_{l\in\mathcal{N}_{US}} \sum_{r=0}^{K-1} -\left(\frac{\alpha\varepsilon ML}{2\delta} + \mathcal{O}(\varepsilon^2) \right) \left(1 - \alpha\beta + \frac{\alpha\varepsilon M}{2} + \mathcal{O}(\varepsilon^2) \right)^r \left(1 + \alpha L + \frac{\alpha\varepsilon M}{2} + \mathcal{O}(\varepsilon^2) \right)^{K-r-1} \theta_l^{us} \tag{D.24}$$

$$= -\left(\frac{\alpha\varepsilon ML}{2\delta} + \mathcal{O}(\varepsilon^2) \right) \frac{\left(1 + \alpha L + \frac{\alpha\varepsilon M}{2} + \mathcal{O}(\varepsilon^2) \right)^K - \left(1 - \alpha\beta - \frac{\alpha\varepsilon M}{2} + \mathcal{O}(\varepsilon^2) \right)^K}{(\alpha L + \alpha\beta + \mathcal{O}(\varepsilon^2))} \sum_{l\in\mathcal{N}_{US}} \theta_l^{us} \tag{D.25}$$

$$> -\left(\frac{\alpha\varepsilon ML}{2\delta} + \mathcal{O}(\varepsilon^2) \right) \frac{\left(1 + \alpha L + \frac{\alpha\varepsilon M}{2} + \mathcal{O}(\varepsilon^2) \right)^K}{(\alpha L + \alpha\beta + \mathcal{O}(\varepsilon^2))} \sum_{l\in\mathcal{N}_{US}} \theta_l^{us}. \tag{D.26}$$

Note that here in the last step we used a loose lower bound by dropping the negative term from the numerator for the sake of simplifying the subsequent analysis. Similarly, an upper bound on $T_3(\tau)$ can be obtained, which is as follows:

$$\sup_{\tau} T_3(\tau) < \left(\frac{\alpha\varepsilon ML}{2\delta} + \mathcal{O}(\varepsilon^2) \right) \frac{\left(1 + \alpha L + \frac{\alpha\varepsilon M}{2} + \mathcal{O}(\varepsilon^2) \right)^K}{(\alpha L + \alpha\beta + \mathcal{O}(\varepsilon^2))} \sum_{l\in\mathcal{N}_{US}} \theta_l^{us}. \tag{D.27}$$

Now that we have derived the bounds for the terms $T_1(\tau), T_2(\tau), T_3(\tau)$, the bounds for remaining terms $T_4(\tau), T_5(\tau), T_6(\tau)$ can be derived along similar lines. Since the algebra is somewhat tedious, we leave these derivations to the reader and directly present the bounds.

The term $T_4(\tau)$ is bounded as

$$\inf_{\tau} T_4(\tau) = \left(1 + \alpha\beta - \frac{\alpha\varepsilon M}{2} - \mathcal{O}(\varepsilon^2) \right)^K \theta_j^{us}, \text{ and} \tag{D.28}$$

$$\sup_{\tau} T_4(\tau) = \left(1 + \alpha L + \frac{\alpha\varepsilon M}{2} + \mathcal{O}(\varepsilon^2) \right)^K \theta_j^{us}. \tag{D.29}$$

The lower and upper bound on term $T_5(\tau)$ are as follows:

$$\inf_{\tau} T_5(\tau) > -\left(\frac{\alpha\varepsilon ML}{2\delta}+\mathscr{O}(\varepsilon^2)\right)\frac{\left(1+\alpha L+\frac{\alpha\varepsilon M}{2}+\mathscr{O}(\varepsilon^2)\right)^K}{(\alpha L+\alpha\beta+\mathscr{O}(\varepsilon^2))}\sum_{l\in\mathscr{N}_S}\theta_l^s, \text{ and} \tag{D.30}$$

$$\sup_{\tau} T_5(\tau) < \left(\frac{\alpha\varepsilon ML}{2\delta}+\mathscr{O}(\varepsilon^2)\right)\frac{\left(1+\alpha L+\frac{\alpha\varepsilon M}{2}+\mathscr{O}(\varepsilon^2)\right)^K}{(\alpha L+\alpha\beta+\mathscr{O}(\varepsilon^2))}\sum_{l\in\mathscr{N}_S}\theta_l^s. \tag{D.31}$$

The lower and upper bound on term $T_6(\tau)$ are as follows:

$$\inf_{\tau} T_6(\tau) \geqslant -K\left(1+\alpha L+\frac{\alpha\varepsilon M}{2}+\mathscr{O}(\varepsilon^2)\right)^{K-1}\left(\frac{\alpha\varepsilon ML}{2\delta}+\mathscr{O}(\varepsilon^2)\right)\sum_{l\in\mathscr{N}_{US}}\theta_l^{us}, \text{ and} \tag{D.32}$$

$$\sup_{\tau} T_6(\tau) \leqslant K\left(1+\alpha L+\frac{\alpha\varepsilon M}{2}+\mathscr{O}(\varepsilon^2)\right)^{K-1}\left(\frac{\alpha\varepsilon ML}{2\delta}+\mathscr{O}(\varepsilon^2)\right)\sum_{l\in\mathscr{N}_{US}}\theta_l^{us}. \tag{D.33}$$

Using these results and dropping higher order terms ($\mathscr{O}(\varepsilon^2)$ and above), we can get the lower bound on $\|\tilde{\mathbf{u}}_K^{\tau}\|^2$. From (D.9), observe that

$$\|\tilde{\mathbf{u}}_K\|^2 = \varepsilon^2\left(\sum_{i\in\mathscr{N}_S}(T_1+T_2+T_3)^2+\sum_{j\in\mathscr{N}_{US}}(T_4+T_5+T_6)^2\right). \tag{D.34}$$

Let $Y_1(\tau)=\sum_{i\in\mathscr{N}_S}(T_1(\tau)+T_2(\tau)+T_3(\tau))^2$ and $Y_2(\tau)=\sum_{j\in\mathscr{N}_{US}}(T_4(\tau)+T_5(\tau)+T_6(\tau))^2$. Using (D.9), we can see that

$$\|\tilde{\mathbf{u}}_K^{\tau}\|^2 = \varepsilon^2\Big(Y_1(\tau)+Y_2(\tau)\Big). \tag{D.35}$$

Now using the bounds for $T_1(\tau), T_2(\tau), T_3(\tau)$ we have the following lower bound on $Y_1(\tau)$:

$$\inf_\tau Y_1(\tau) \geqslant \sum_{i\in\mathcal{N}_S}\left(\inf_\tau\left\{T_1^2(\tau)+T_2^2(\tau)+T_3^2(\tau)+2T_1(\tau)T_2(\tau)+2T_2(\tau)T_3(\tau)+2T_3(\tau)T_1(\tau)\right\}\right) \tag{D.36}$$

$$\geqslant \sum_{i\in\mathcal{N}_S}\left(\inf_\tau \underbrace{T_1^2(\tau)}_{>0}+\inf_\tau \underbrace{T_2^2(\tau)}_{\geqslant 0}+\inf_\tau \underbrace{T_3^2(\tau)}_{\geqslant 0}+2\underbrace{\inf_\tau\left\{T_1(\tau)T_2(\tau)\right\}}_{<0}+2\underbrace{\inf_\tau\left\{T_2(\tau)T_3(\tau)\right\}}_{<0}+2\underbrace{\inf_\tau\left\{T_3(\tau)T_1(\tau)\right\}}_{<0}\right) \tag{D.37}$$

$$> \left(\sum_{i\in\mathcal{N}_S}\inf_\tau \underbrace{T_1^2(\tau)}_{>0}+0+0-2\underbrace{\sup_\tau|T_1(\tau)|}_{>0}\underbrace{\sup_\tau|T_2(\tau)|}_{>0}-2\underbrace{\sup_\tau|T_2(\tau)|}_{>0}\underbrace{\sup_\tau|T_3(\tau)|}_{>0}-2\underbrace{\sup_\tau|T_3(\tau)|}_{>0}\underbrace{\sup_\tau|T_1(\tau)|}_{>0}\right) \tag{D.38}$$

$$\begin{aligned}
&= \sum_{i\in\mathcal{N}_S}\left(\left(1-\alpha L-\frac{\alpha\varepsilon M}{2}-\mathcal{O}(\varepsilon^2)\right)^{2K}(\theta_i^s)^2-2K\theta_i^s\left(1-\alpha\beta+\frac{\alpha\varepsilon M}{2}+\mathcal{O}(\varepsilon^2)\right)^{2K-1}\left(\frac{\alpha\varepsilon ML}{2\delta}+\mathcal{O}(\varepsilon^2)\right)\sum_{l\in\mathcal{N}_S}\theta_l^s\right.\\
&-2K\left(1-\alpha\beta+\frac{\alpha\varepsilon M}{2}+\mathcal{O}(\varepsilon^2)\right)^{K-1}\left(\frac{\alpha\varepsilon ML}{2\delta}+\mathcal{O}(\varepsilon^2)\right)^2\frac{\left(1+\alpha L+\frac{\alpha\varepsilon M}{2}+\mathcal{O}(\varepsilon^2)\right)^K}{(\alpha L+\alpha\beta+\mathcal{O}(\varepsilon^2))}\sum_{l\in\mathcal{N}_{US}}\theta_l^{us}\sum_{i\in\mathcal{N}_S}\theta_i^s\\
&\left.-2\left(1-\alpha\beta+\frac{\alpha\varepsilon M}{2}+\mathcal{O}(\varepsilon^2)\right)^{K}\left(\frac{\alpha\varepsilon ML}{2\delta}+\mathcal{O}(\varepsilon^2)\right)\frac{\left(1+\alpha L+\frac{\alpha\varepsilon M}{2}+\mathcal{O}(\varepsilon^2)\right)^K}{(\alpha L+\alpha\beta+\mathcal{O}(\varepsilon^2))}\sum_{l\in\mathcal{N}_{US}}\theta_l^{us}\sum_{i\in\mathcal{N}_S}\theta_i^s\right)
\end{aligned} \tag{D.39}$$

$$\begin{aligned}
&= \sum_{i\in\mathcal{N}_S}\left(\left(1-\alpha L-\frac{\alpha\varepsilon M}{2}-\mathcal{O}(\varepsilon^2)\right)^{2K}(\theta_i^s)^2-2K\theta_i^s\left(1-\alpha\beta+\frac{\alpha\varepsilon M}{2}+\mathcal{O}(\varepsilon^2)\right)^{2K-1}\left(\frac{\alpha\varepsilon ML}{2\delta}+\mathcal{O}(\varepsilon^2)\right)\sum_{l\in\mathcal{N}_S}\theta_l^s\right)\\
&-2\left(1+\frac{K\sum_{i\in\mathcal{N}_S}\left(\frac{\alpha\varepsilon ML}{2\delta}+\mathcal{O}(\varepsilon^2)\right)}{\left(1-\alpha\beta+\frac{\alpha\varepsilon M}{2}+\mathcal{O}(\varepsilon^2)\right)}\right)\left(1-\alpha\beta+\frac{\alpha\varepsilon M}{2}+\mathcal{O}(\varepsilon^2)\right)^K\left(\frac{\alpha\varepsilon ML}{2\delta}+\mathcal{O}(\varepsilon^2)\right)\frac{\left(1+\alpha L+\frac{\alpha\varepsilon M}{2}+\mathcal{O}(\varepsilon^2)\right)^K}{(\alpha L+\alpha\beta+\mathcal{O}(\varepsilon^2))}\sum_{l\in\mathcal{N}_{US}}\theta_l^{us}\sum_{i\in\mathcal{N}_S}\theta_i^s
\end{aligned} \tag{D.40}$$

$$\begin{aligned}
&= \sum_{i\in\mathcal{N}_S}\left(\left(1-\alpha L-\frac{\alpha\varepsilon M}{2}-\mathcal{O}(\varepsilon^2)\right)^{2K}(\theta_i^s)^2-2K\theta_i^s\left(1-\alpha\beta+\frac{\alpha\varepsilon M}{2}+\mathcal{O}(\varepsilon^2)\right)^{2K-1}\left(\frac{\alpha\varepsilon ML}{2\delta}+\mathcal{O}(\varepsilon^2)\right)\sum_{l\in\mathcal{N}_S}\theta_l^s\right)\\
&-2\left(1+\mathcal{O}(K\varepsilon)\right)\left(1-\alpha\beta+\frac{\alpha\varepsilon M}{2}+\mathcal{O}(\varepsilon^2)\right)^K\left(\frac{\alpha\varepsilon ML}{2\delta}+\mathcal{O}(\varepsilon^2)\right)\frac{\left(1+\alpha L+\frac{\alpha\varepsilon M}{2}+\mathcal{O}(\varepsilon^2)\right)^K}{(\alpha L+\alpha\beta+\mathcal{O}(\varepsilon^2))}\sum_{l\in\mathcal{N}_{US}}\theta_l^{us}\sum_{i\in\mathcal{N}_S}\theta_i^s,
\end{aligned} \tag{D.41}$$

where in the last step we replaced the term $K\sum_{i\in\mathcal{N}_S}\frac{\left(\frac{\alpha\varepsilon ML}{2\delta}+\mathcal{O}(\varepsilon^2)\right)}{\left(1-\alpha\beta+\frac{\alpha\varepsilon M}{2}+\mathcal{O}(\varepsilon^2)\right)}$ with $\mathcal{O}(K\varepsilon)$ for $K\varepsilon \ll 1$ and

$\left(1-\alpha\beta+\frac{\alpha\varepsilon M}{2}+\mathcal{O}(\varepsilon^2)\right) \gg \varepsilon$. This is because the numerator $\left(\frac{\alpha\varepsilon ML}{2\delta}+\mathcal{O}(\varepsilon^2)\right)$ is of $\mathcal{O}(\varepsilon)$; hence, we require the denominator $\left(1-\alpha\beta+\frac{\alpha\varepsilon M}{2}+\mathcal{O}(\varepsilon^2)\right)$ to be of constant order, i.e., independent of ε. Similarly, using the bounds for $T_4(\tau), T_5(\tau), T_6(\tau)$ we have the following lower bound for $Y_2(\tau)$:

$$\inf_\tau Y_2(\tau) \geqslant \sum_{j\in\mathcal{N}_{US}} \left(\inf_\tau \left\{ T_4^2(\tau)+T_5^2(\tau)+T_6^2(\tau)+2T_4(\tau)T_6(\tau)+2T_6(\tau)T_5(\tau)+2T_5(\tau)T_4(\tau) \right\} \right) \tag{D.42}$$

$$\geqslant \sum_{j\in\mathcal{N}_{US}} \Bigg(\inf_\tau \underbrace{T_4^2(\tau)}_{>0} + \inf_\tau \underbrace{T_5^2(\tau)}_{\geqslant 0} + \inf_\tau \underbrace{T_6^2(\tau)}_{\geqslant 0} + 2\underbrace{\inf_\tau \left\{ T_4(\tau)T_6(\tau) \right\}}_{<0} + 2\underbrace{\inf_\tau \left\{ T_6(\tau)T_5(\tau) \right\}}_{<0} + 2\underbrace{\inf_\tau \left\{ T_5(\tau)T_4(\tau) \right\}}_{<0} \Bigg) \tag{D.43}$$

$$> \Bigg(\sum_{j\in\mathcal{N}_{US}} \inf_\tau \underbrace{T_4^2(\tau)}_{>0} + 0 + 0 - 2\underbrace{\sup_\tau |T_4(\tau)|}_{>0}\underbrace{\sup_\tau |T_6(\tau)|}_{>0} - 2\underbrace{\sup_\tau |T_6(\tau)|}_{>0}\underbrace{\sup_\tau |T_5(\tau)|}_{>0} - 2\underbrace{\sup_\tau |T_5(\tau)|}_{>0}\underbrace{\sup_\tau |T_4(\tau)|}_{>0} \Bigg) \tag{D.44}$$

$$= \sum_{j\in\mathcal{N}_{US}} \Bigg(\left(1+\alpha\beta-\frac{\alpha\varepsilon M}{2}-\mathcal{O}(\varepsilon^2)\right)^{2K} (\theta_j^{us})^2 - 2K\theta_j^{us}\left(1+\alpha L+\frac{\alpha\varepsilon M}{2}+\mathcal{O}(\varepsilon^2)\right)^{2K-1}\left(\frac{\alpha\varepsilon ML}{2\delta}+\mathcal{O}(\varepsilon^2)\right)\sum_{l\in\mathcal{N}_{US}}\theta_l^{us}\Bigg)$$
$$-2\left(1+\mathcal{O}(K\varepsilon)\right)\left(\frac{\alpha\varepsilon ML}{2\delta}+\mathcal{O}(\varepsilon^2)\right)\frac{\left(1+\alpha L+\frac{\alpha\varepsilon M}{2}+\mathcal{O}(\varepsilon^2)\right)^{2K}}{(\alpha L+\alpha\beta+\mathcal{O}(\varepsilon^2))}\sum_{l\in\mathcal{N}_S}\theta_l^s\sum_{j\in\mathcal{N}_{US}}\theta_j^{us}. \tag{D.45}$$

Finally combining these two bounds yields the following lower bound on $\inf_\tau \|\tilde{\mathbf{u}}_K^\tau\|^2$:

$$\inf_\tau \|\tilde{\mathbf{u}}_K^\tau\|^2 = \varepsilon^2\left(\inf_\tau Y_1(\tau)+\inf_\tau Y_2(\tau)\right) \tag{D.46}$$

$$> \varepsilon^2\Bigg[\left(1-\alpha L-\frac{\alpha\varepsilon M}{2}-\mathcal{O}(\varepsilon^2)\right)^{2K}\sum_{i\in\mathcal{N}_S}(\theta_i^s)^2 - 2K\left(1-\alpha\beta+\frac{\alpha\varepsilon M}{2}+\mathcal{O}(\varepsilon^2)\right)^{2K-1}\left(\frac{\alpha\varepsilon ML}{2\delta}+\mathcal{O}(\varepsilon^2)\right)\Big(\sum_{i\in\mathcal{N}_S}\theta_i^s\Big)^2 +$$
$$\left(1+\alpha\beta-\frac{\alpha\varepsilon M}{2}-\mathcal{O}(\varepsilon^2)\right)^{2K}\sum_{j\in\mathcal{N}_{US}}(\theta_j^{us})^2 - 2K\left(1+\alpha L+\frac{\alpha\varepsilon M}{2}+\mathcal{O}(\varepsilon^2)\right)^{2K-1}\left(\frac{\alpha\varepsilon ML}{2\delta}+\mathcal{O}(\varepsilon^2)\right)\Big(\sum_{j\in\mathcal{N}_{US}}\theta_j^{us}\Big)^2 -$$
$$2\left(1+\mathcal{O}(K\varepsilon)\right)\frac{\left(\frac{\alpha\varepsilon ML}{2\delta}+\mathcal{O}(\varepsilon^2)\right)}{(\alpha L+\alpha\beta+\mathcal{O}(\varepsilon^2))}\left(1+\alpha L+\frac{\alpha\varepsilon M}{2}+\mathcal{O}(\varepsilon^2)\right)^K\left(1-\alpha\beta+\frac{\alpha\varepsilon M}{2}+\mathcal{O}(\varepsilon^2)\right)^K\sum_{j\in\mathcal{N}_{US}}\theta_j^{us}\sum_{i\in\mathcal{N}_S}\theta_i^s -$$
$$2\left(1+\mathcal{O}(K\varepsilon)\right)\frac{\left(\frac{\alpha\varepsilon ML}{2\delta}+\mathcal{O}(\varepsilon^2)\right)}{(\alpha L+\alpha\beta+\mathcal{O}(\varepsilon^2))}\left(1+\alpha L+\frac{\alpha\varepsilon M}{2}+\mathcal{O}(\varepsilon^2)\right)^{2K}\sum_{j\in\mathcal{N}_{US}}\theta_j^{us}\sum_{i\in\mathcal{N}_S}\theta_i^s\Bigg] \tag{D.47}$$

$$> \varepsilon^2\Bigg[\bigg(1-\alpha L-\frac{\alpha\varepsilon M}{2}-\mathscr{O}(\varepsilon^2)\bigg)^{2K}\sum_{i\in\mathscr{N}_S}(\theta_i^s)^2-2nK\bigg(1-\alpha\beta+\frac{\alpha\varepsilon M}{2}+\mathscr{O}(\varepsilon^2)\bigg)^{2K-1}\bigg(\frac{\alpha\varepsilon ML}{2\delta}+\mathscr{O}(\varepsilon^2)\bigg)\sum_{i\in\mathscr{N}_S}(\theta_i^s)^2+$$
$$\bigg(1+\alpha\beta-\frac{\alpha\varepsilon M}{2}-\mathscr{O}(\varepsilon^2)\bigg)^{2K}\sum_{j\in\mathscr{N}_{US}}(\theta_j^{us})^2-2nK\bigg(1+\alpha L+\frac{\alpha\varepsilon M}{2}+\mathscr{O}(\varepsilon^2)\bigg)^{2K-1}\bigg(\frac{\alpha\varepsilon ML}{2\delta}+\mathscr{O}(\varepsilon^2)\bigg)\sum_{j\in\mathscr{N}_{US}}(\theta_j^{us})^2-$$
$$n\bigg(1+\mathscr{O}(K\varepsilon)\bigg)\frac{\bigg(\frac{\alpha\varepsilon ML}{2\delta}+\mathscr{O}(\varepsilon^2)\bigg)}{(\alpha L+\alpha\beta+\mathscr{O}(\varepsilon^2))}\bigg(1+\alpha L+\frac{\alpha\varepsilon M}{2}+\mathscr{O}(\varepsilon^2)\bigg)^{K}\bigg(1-\alpha\beta+\frac{\alpha\varepsilon M}{2}+\mathscr{O}(\varepsilon^2)\bigg)^{K}\bigg(\sum_{i\in\mathscr{N}_S}(\theta_i^s)^2+\sum_{j\in\mathscr{N}_{US}}(\theta_j^{us})^2\bigg)-$$
$$n\bigg(1+\mathscr{O}(K\varepsilon)\bigg)\frac{\bigg(\frac{\alpha\varepsilon ML}{2\delta}+\mathscr{O}(\varepsilon^2)\bigg)}{(\alpha L+\alpha\beta+\mathscr{O}(\varepsilon^2))}\bigg(1+\alpha L+\frac{\alpha\varepsilon M}{2}+\mathscr{O}(\varepsilon^2)\bigg)^{2K}\bigg(\sum_{i\in\mathscr{N}_S}(\theta_i^s)^2+\sum_{j\in\mathscr{N}_{US}}(\theta_j^{us})^2\bigg)\Bigg]. \tag{D.48}$$

Note that in the last step we have used the following inequalities:

$$n\sum_{i\in\mathscr{N}_S}(\theta_i^s)^2\geqslant(\sum_{i\in\mathscr{N}_S}\theta_i^s)^2,$$
$$n\sum_{j\in\mathscr{N}_{US}}(\theta_j^{us})^2\geqslant(\sum_{j\in\mathscr{N}_{US}}\theta_j^{us})^2,\text{ and}$$
$$2\sum_{j\in\mathscr{N}_{US}}\theta_j^{us}\sum_{i\in\mathscr{N}_S}\theta_i^s\leqslant n\sum_{j\in\mathscr{N}_{US}}(\theta_j^{us})^2+n\sum_{i\in\mathscr{N}_S}(\theta_i^s)^2,$$

where n is the dimension of the domain of the function $f(\cdot)$. The above condition can be more compactly written as

$$\varepsilon^2\geqslant\inf_{\tau}\|\tilde{\mathbf{u}}_K^{\tau}\|^2>\varepsilon^2\Psi(K), \tag{D.49}$$

where we have that

$$\Psi(K)=\bigg(c_1^{2K}-2Kc_2^{2K-1}b_1-b_2c_3^Kc_2^K-b_2c_3^{2K}\bigg)\sum_{i\in\mathscr{N}_S}(\theta_i^s)^2+\bigg(c_4^{2K}-2Kc_3^{2K-1}b_1-b_2c_3^Kc_2^K-b_2c_3^{2K}\bigg)\sum_{j\in\mathscr{N}_{US}}(\theta_j^{us})^2, \tag{D.50}$$

and $c_1=\bigg(1-\alpha L-\frac{\alpha\varepsilon M}{2}-\mathscr{O}(\varepsilon^2)\bigg)$, $c_2=\bigg(1-\alpha\beta+\frac{\alpha\varepsilon M}{2}+\mathscr{O}(\varepsilon^2)\bigg)$, $c_3=\bigg(1+\alpha L+\frac{\alpha\varepsilon M}{2}+\mathscr{O}(\varepsilon^2)\bigg)$,
$c_4=\bigg(1+\alpha\beta-\frac{\alpha\varepsilon M}{2}-\mathscr{O}(\varepsilon^2)\bigg)$, $b_1=\bigg(\frac{\alpha\varepsilon MLn}{2\delta}+\mathscr{O}(\varepsilon^2)\bigg)$ and $b_2=\dfrac{\bigg(\frac{\alpha\varepsilon MLn}{2\delta}+\mathscr{O}(\varepsilon^2)\bigg)\bigg(1+\mathscr{O}(K\varepsilon)\bigg)}{\bigg(\alpha L+\alpha\beta+\mathscr{O}(\varepsilon^2)\bigg)}$.

The condition in (D.49) holds for all such K where $\inf_\tau\|\tilde{\mathbf{u}}_K^{\tau}\|^2\leqslant\varepsilon^2$. Therefore to obtain K^ι defined in (D.5), we need to solve for K where $\varepsilon^2\leqslant\varepsilon^2\Psi(K)$ or equivalently $1\leqslant\Psi(K)$ where the condition $\inf_\tau\|\tilde{\mathbf{u}}_K^{\tau}\|^2\leqslant\varepsilon^2$ gets inverted using inequality (D.49).

D.2.1 *Claim for the value of K in Theorem 3.2:.* Since the infimum in (D.49) is taken over all τ, the condition in (D.49) holds true for all K in the range $1\leqslant K<\sup_\tau\bigg\{K_{exit}^{\tau}\bigg\}$.

Proof of the claim: Recall that from the definition of K^ι from (D.5), K^ι satisfies the following condition:

$$\inf_\tau \left\|\tilde{\mathbf{u}}^\tau_{K^\iota-1}\right\|^2 \leqslant \varepsilon^2 < \inf_\tau \|\tilde{\mathbf{u}}^\tau_{K^\iota}\|^2, \tag{D.51}$$

where the lower bound implies that the infimum over all τ-parameterized approximate trajectories has not yet escaped the ball $\mathscr{B}_\varepsilon(\mathbf{x}^*)$. Let there exist some $\bar{K}$ where $\bar{K} \in \left\{1,2,...,\sup_\tau \left\{K^\tau_{exit}\right\}\right\}$ such that the condition in (D.49) holds for all $K \in \left\{1,...,\bar{K}-1\right\}$ and fails to hold for all $K \in \left\{\bar{K},...,\sup_\tau \left\{K^\tau_{exit}\right\}\right\}$, i.e. we have the condition $\varepsilon^2 \leqslant \varepsilon^2\Psi(K) < \inf_\tau \|\tilde{\mathbf{u}}^\tau_K\|^2$ for $K \geqslant \bar{K}$. This implies that

$$\varepsilon^2\Psi(\bar{K}-1) < \inf_\tau \left\|\tilde{\mathbf{u}}^\tau_{\bar{K}-1}\right\|^2 \leqslant \varepsilon^2 \leqslant \varepsilon^2\Psi(\bar{K}) < \inf_\tau \left\|\tilde{\mathbf{u}}^\tau_{\bar{K}}\right\|^2. \tag{D.52}$$

From conditions (D.51) and (D.52) we get that $\bar{K} = K^\iota$. Since $\bar{K} \in \left\{1,2,...,\sup_\tau \left\{K^\tau_{exit}\right\}\right\}$ we have that $K^\iota = \bar{K} \leqslant \sup_\tau \left\{K^\tau_{exit}\right\} \leqslant K^\iota$. Hence we must have that $\bar{K} = \sup_\tau \left\{K^\tau_{exit}\right\}$. □

E. Proof of Theorem 3.3 (Exit time for the infimum of ε-precision trajectories)

Proof.

Further simplifying the inequality in (D.48) by dropping order $\mathscr{O}(\varepsilon^2)$ and $\mathscr{O}(K\varepsilon)$ terms (for $K\varepsilon \ll 1$) appearing on its right hand side and using (D.49), we get the following approximate lower bound:

$$\begin{aligned}
1 \gtrapprox & \left(\left[\left(1-\alpha L-\frac{\alpha\varepsilon M}{2}\right)^{2K} - 2K\left(1-\alpha\beta+\frac{\alpha\varepsilon M}{2}\right)^{2K-1}\frac{\alpha\varepsilon MLn}{2\delta}\right]\sum_{i\in\mathscr{N}_S}(\theta^s_i)^2 + \right. \\
& \left[\left(1+\alpha\beta-\frac{\alpha\varepsilon M}{2}\right)^{2K} - 2K\left(1+\alpha L+\frac{\alpha\varepsilon M}{2}\right)^{2K-1}\frac{\alpha\varepsilon MLn}{2\delta}\right]\sum_{j\in\mathscr{N}_{US}}(\theta^{us}_j)^2 \\
& - \frac{\alpha\varepsilon MLn}{2\delta(\alpha L+\alpha\beta)}\left(1+\alpha L+\frac{\alpha\varepsilon M}{2}\right)^K\left(1-\alpha\beta+\frac{\alpha\varepsilon M}{2}\right)^K\left(\sum_{i\in\mathscr{N}_S}(\theta^s_i)^2+\sum_{j\in\mathscr{N}_{US}}(\theta^{us}_j)^2\right) \\
& \left. - \frac{\alpha\varepsilon MLn}{2\delta(\alpha L+\alpha\beta)}\left(1+\alpha L+\frac{\alpha\varepsilon M}{2}\right)^{2K}\right)\left(\sum_{i\in\mathscr{N}_S}(\theta^s_i)^2+\sum_{j\in\mathscr{N}_{US}}(\theta^{us}_j)^2\right)
\end{aligned} \tag{E.1}$$

$$\begin{aligned}
1 \gtrapprox & \left(\left[\left(1-\alpha L-\frac{\alpha\varepsilon M}{2}\right)^{2K} - 2K\left(1-\alpha\beta+\frac{\alpha\varepsilon M}{2}\right)^{2K-1}\frac{\alpha\varepsilon MLn}{2\delta}\right]\sum_{i\in\mathscr{N}_S}(\theta^s_i)^2 + \right. \\
& \left.\left[\left(1+\alpha\beta-\frac{\alpha\varepsilon M}{2}\right)^{2K} - 2K\left(1+\alpha L+\frac{\alpha\varepsilon M}{2}\right)^{2K-1}\frac{\alpha\varepsilon MLn}{2\delta}\right]\sum_{j\in\mathscr{N}_{US}}(\theta^{us}_j)^2 - \varepsilon MLn\frac{\left(1+\alpha L+\frac{\alpha\varepsilon M}{2}\right)^{2K}}{\delta(L+\beta)}\right),
\end{aligned} \tag{E.2}$$

where in the last step we used the relation $\left(\sum_{i\in\mathscr{N}_S}(\theta^s_i)^2+\sum_{j\in\mathscr{N}_{US}}(\theta^{us}_j)^2\right)=1$ and the inequality $\left(1-\alpha\beta+\frac{\alpha\varepsilon M}{2}\right) < \left(1+\alpha L+\frac{\alpha\varepsilon M}{2}\right)$. Now, if we substitute the step size $\alpha=\frac{1}{L}$, we get the following

approximate inequality:

$$1 \gtrapprox \Bigg(\Bigg[\Bigg(-\frac{\varepsilon M}{2L} \Bigg)^{2K} - 2K \Bigg(1 - \frac{\beta}{L} + \frac{\varepsilon M}{2L} \Bigg)^{2K-1} \frac{\varepsilon Mn}{2\delta} \Bigg] \sum_{i \in \mathcal{N}_S} (\theta_i^s)^2 +$$

$$\Bigg[\Bigg(1 + \frac{\beta}{L} - \frac{\varepsilon M}{2L} \Bigg)^{2K} - 2K \Bigg(2 + \frac{\varepsilon M}{2L} \Bigg)^{2K-1} \frac{\varepsilon Mn}{2\delta} \Bigg] \sum_{j \in \mathcal{N}_{US}} (\theta_j^{us})^2 - \varepsilon MLn \frac{\Big(2 + \frac{\varepsilon M}{2L} \Big)^{2K}}{\delta(L+\beta)} \Bigg) \tag{E.3}$$

$$1 \gtrapprox \Bigg(\Bigg[-2K \Bigg(1 - \frac{\beta}{L} + \frac{\varepsilon M}{2L} \Bigg)^{2K-1} \frac{\varepsilon Mn}{2\delta} \Bigg] \sum_{i \in \mathcal{N}_S} (\theta_i^s)^2 +$$

$$\Bigg[\Bigg(1 + \frac{\beta}{L} - \frac{\varepsilon M}{2L} \Bigg)^{2K} - 2K \Bigg(2 + \frac{\varepsilon M}{2L} \Bigg)^{2K-1} \frac{\varepsilon Mn}{2\delta} \Bigg] \sum_{j \in \mathcal{N}_{US}} (\theta_j^{us})^2 - \varepsilon MLn \frac{\Big(2 + \frac{\varepsilon M}{2L} \Big)^{2K}}{\delta(L+\beta)} \Bigg), \tag{E.4}$$

where in the last step we dropped the $\Big(-\frac{\varepsilon M}{2L} \Big)^{2K}$ term from right hand side.

In order to obtain K^ι and hence the exit time K_{exit}, we need to solve for values of K where the approximate inequality in (E.4) becomes an equality. Hence, we look into the two possible cases for this value K, i.e., large K and small K. Note that in the next subsections we only consider those cases where our unstable projection $\sum_{j \in \mathcal{N}_{US}} (\theta_j^{us})^2$ is not too close to 0. We now obtain the exit time K_{exit} for the two cases.

E.0.1 *Case 1—Large K:.* If K is large with $K = \mathcal{O}\Big(\frac{1}{\varepsilon} \Big)$ then we can use the Lambert W function [8] to solve the above transcendental inequality (E.4). Specifically for obtaining linear escape rates i.e., $K = \mathcal{O}\Big(\log\Big(\frac{1}{\varepsilon} \Big) \Big)$, we set $\frac{1}{\Big(2 + \frac{\varepsilon M}{2L} \Big)^{2K}} = \rho \varepsilon^c$ for some $\rho > 0$, $c > 0$, $\Big(1 - \frac{\beta}{L} + \frac{\varepsilon M}{2L} \Big)^{2K} = \eta \varepsilon^d$ for some $\eta > 0$, $d > 0$ where $\Big(1 - \frac{\beta}{L} + \frac{\varepsilon M}{2L} \Big) < 1$ and divide both sides of (E.4) by the term $\Big(2 + \frac{\varepsilon M}{2L} \Big)^{2K}$ to get the following approximate inequality:

$$\frac{1}{\Big(2 + \frac{\varepsilon M}{2L} \Big)^{2K}} \gtrapprox \Bigg(\underbrace{\Bigg[-2K \frac{\Big(1 - \frac{\beta}{L} + \frac{\varepsilon M}{2L} \Big)^{2K-1}}{\Big(2 + \frac{\varepsilon M}{2L} \Big)^{2K}} \frac{\varepsilon Mn}{2\delta} \Bigg] \sum_{i \in \mathcal{N}_S} (\theta_i^s)^2}_{F_1} +$$

$$\Bigg[\frac{\Big(1 + \frac{\beta}{L} - \frac{\varepsilon M}{2L} \Big)^{2K}}{\Big(2 + \frac{\varepsilon M}{2L} \Big)^{2K}} - 2K \Big(2 + \frac{\varepsilon M}{2L} \Big)^{-1} \frac{\varepsilon Mn}{2\delta} \Bigg] \sum_{j \in \mathcal{N}_{US}} (\theta_j^{us})^2 - \frac{\varepsilon MLn}{\delta(L+\beta)} \Bigg). \tag{E.5}$$

Dropping the first term F_1 on right hand side for large K (this term has order $\mathscr{O}\left(\varepsilon^{(1+c+d)}\log\left(\frac{1}{\varepsilon}\right)\right)$ with $c>0$, $d>0$) and making the substitution of $\rho\varepsilon^c$ on the left hand side, we get the following bound:

$$\rho\varepsilon^c \gtrapprox \left(\frac{1+\frac{\beta}{L}-\frac{\varepsilon M}{2L}}{2+\frac{\varepsilon M}{2L}}\right)^{2K}\sum_{j\in\mathscr{N}_{US}}(\theta_j^{us})^2-2K\left(2+\frac{\varepsilon M}{2L}\right)^{-1}\frac{\varepsilon Mn}{2\delta}\sum_{j\in\mathscr{N}_{US}}(\theta_j^{us})^2-\frac{\varepsilon MLn}{\delta(L+\beta)} \tag{E.6}$$

$$\left(\frac{1+\frac{\beta}{L}-\frac{\varepsilon M}{2L}}{2+\frac{\varepsilon M}{2L}}\right)^{2K} \lessapprox 2K\left(2+\frac{\varepsilon M}{2L}\right)^{-1}\frac{\varepsilon Mn}{2\delta}+\frac{\varepsilon\left(\rho\varepsilon^{(c-1)}+\frac{MLn}{\delta(L+\beta)}\right)}{\sum_{j\in\mathscr{N}_{US}}(\theta_j^{us})^2}. \tag{E.7}$$

When the problem is well conditioned, i.e., $\left(1-\frac{\beta}{L}+\frac{\varepsilon M}{2L}\right)<1$ or equivalently $\frac{\beta}{L}>\frac{\varepsilon M}{2L}$, then we are guaranteed fast escape under good initial unstable projections. Now, solving for the values of K where the inequality (E.7) becomes equality, we make use of the general transcendental equation $q^x=ax+b$ whose solution is given by

$$x=-\frac{W(-\frac{\log q}{a}q^{-\frac{b}{a}})}{\log q}-\frac{b}{a}, \tag{E.8}$$

where $W(\cdot)$ is the Lambert W function. On comparing the coefficients, we have $x=2K$ and the constants as follows:

$$a=\left(2+\frac{\varepsilon M}{2L}\right)^{-1}\frac{\varepsilon Mn}{2\delta}, b=\frac{\varepsilon\left(\rho\varepsilon^{(c-1)}+\frac{MLn}{\delta(L+\beta)}\right)}{\sum_{j\in\mathscr{N}_{US}}(\theta_j^{us})^2}, q=\left(\frac{1+\frac{\beta}{L}-\frac{\varepsilon M}{2L}}{2+\frac{\varepsilon M}{2L}}\right). \tag{E.9}$$

For large values of any argument y, the Lambert W function is bounded by $W(y)\leqslant\log(y)$. If the quantity $\sum_{j\in\mathscr{N}_{US}}(\theta_j^{us})^2$ is not too close to 0 and is lower bounded, i.e., $\sum_{j\in\mathscr{N}_{US}}(\theta_j^{us})^2\geqslant\Delta$ then we have an initial projection onto the unstable subspace of the saddle point. Using the Lambert W function

bound and substituting the coefficients, we have following bound on K:

$$2K = \frac{1}{\log\left(\frac{2+\frac{\varepsilon M}{2L}}{1+\frac{\beta}{L}-\frac{\varepsilon M}{2L}}\right)} W\left(\left(2+\frac{\varepsilon M}{2L}\right)\frac{2\delta}{\varepsilon Mn}\log\left(\frac{2+\frac{\varepsilon M}{2L}}{1+\frac{\beta}{L}-\frac{\varepsilon M}{2L}}\right)\left(\frac{2+\frac{\varepsilon M}{2L}}{1+\frac{\beta}{L}-\frac{\varepsilon M}{2L}}\right)^{\frac{2\delta\left(2+\frac{\varepsilon M}{2L}\right)\left(\rho\varepsilon^{(c-1)}+\frac{MLn}{\delta(L+\beta)}\right)}{Mn\sum_{j\in\mathcal{N}_{US}}(\theta_j^{us})^2}}\right) - \frac{2\delta\left(2+\frac{\varepsilon M}{2L}\right)\left(\rho\varepsilon^{(c-1)}+\frac{MLn}{\delta(L+\beta)}\right)}{Mn\sum_{j\in\mathcal{N}_{US}}(\theta_j^{us})^2} \tag{E.10}$$

$$2K \leqslant \frac{1}{\log\left(\frac{2+\frac{\varepsilon M}{2L}}{1+\frac{\beta}{L}-\frac{\varepsilon M}{2L}}\right)} \log\left(\left(2+\frac{\varepsilon M}{2L}\right)\frac{2\delta}{\varepsilon Mn}\log\left(\frac{2+\frac{\varepsilon M}{2L}}{1+\frac{\beta}{L}-\frac{\varepsilon M}{2L}}\right)\left(\frac{2+\frac{\varepsilon M}{2L}}{1+\frac{\beta}{L}-\frac{\varepsilon M}{2L}}\right)^{\frac{2\delta\left(2+\frac{\varepsilon M}{2L}\right)\left(\rho\varepsilon^{(c-1)}+\frac{MLn}{\delta(L+\beta)}\right)}{Mn\sum_{j\in\mathcal{N}_{US}}(\theta_j^{us})^2}}\right) - \frac{2\delta\left(2+\frac{\varepsilon M}{2L}\right)\left(\rho\varepsilon^{(c-1)}+\frac{MLn}{\delta(L+\beta)}\right)}{Mn\sum_{j\in\mathcal{N}_{US}}(\theta_j^{us})^2} \tag{E.11}$$

$$2K \leqslant \frac{\log\left(\left(2+\frac{\varepsilon M}{2L}\right)\log\left(\frac{2+\frac{\varepsilon M}{2L}}{1+\frac{\beta}{L}-\frac{\varepsilon M}{2L}}\right)\frac{2\delta}{\varepsilon Mn}\right)}{\log\left(\frac{2+\frac{\varepsilon M}{2L}}{1+\frac{\beta}{L}-\frac{\varepsilon M}{2L}}\right)} + \frac{2\delta\left(2+\frac{\varepsilon M}{2L}\right)\left(\rho\varepsilon^{(c-1)}+\frac{MLn}{\delta(L+\beta)}\right)}{Mn\sum_{j\in\mathcal{N}_{US}}(\theta_j^{us})^2} - \frac{2\delta\left(2+\frac{\varepsilon M}{2L}\right)\left(\rho\varepsilon^{(c-1)}+\frac{MLn}{\delta(L+\beta)}\right)}{Mn\sum_{j\in\mathcal{N}_{US}}(\theta_j^{us})^2} \tag{E.12}$$

$$K \leqslant \frac{\log\left(\left(2+\frac{\varepsilon M}{2L}\right)\log\left(\frac{2+\frac{\varepsilon M}{2L}}{1+\frac{\beta}{L}-\frac{\varepsilon M}{2L}}\right)\frac{2\delta}{\varepsilon Mn}\right)}{2\log\left(\frac{2+\frac{\varepsilon M}{2L}}{1+\frac{\beta}{L}-\frac{\varepsilon M}{2L}}\right)} = \mathcal{O}\left(\log\left(\frac{1}{\varepsilon}\right)\right). \tag{E.13}$$

Notice that the K solved here is an approximate solution to (E.7) where the inequality in (E.7) gets inverted. Since the condition (E.4) gets reversed at $K = K^\iota$, we therefore get the condition $K^\iota \lessapprox \frac{\log\left(\left(2+\frac{\varepsilon M}{2L}\right)\log\left(\frac{2+\frac{\varepsilon M}{2L}}{1+\frac{\beta}{L}-\frac{\varepsilon M}{2L}}\right)\frac{2\delta}{\varepsilon Mn}\right)}{2\log\left(\frac{2+\frac{\varepsilon M}{2L}}{1+\frac{\beta}{L}-\frac{\varepsilon M}{2L}}\right)}$ and using the fact that $K_{exit} < K^\iota$ gives the desired conclusion of $K_{exit} \leqslant K^\iota = \mathcal{O}\left(\log\left(\frac{1}{\varepsilon}\right)\right)$. The bound $\varepsilon < \frac{2\beta}{M}$ follows from the fact that $\frac{\beta}{L} > \frac{\varepsilon M}{2L}$.

Hence, we have escape rates of order $\mathcal{O}\left(\log\left(\frac{1}{\varepsilon}\right)\right)$ for the case when our problem is well conditioned and does not have a very small unstable projection. It is remarked that this is only an upper bound on K and the iterate is likely to escape way before this time. Also, this result supports our analysis of the trajectory function for values of $K = \mathcal{O}\left(\frac{1}{\varepsilon}\right)$.

It is worth mentioning that dropping of the first term F_1 with order $\mathcal{O}\left(\varepsilon^{(1+c+d)}\log\left(\frac{1}{\varepsilon}\right)\right)$ from the

right hand side of inequality (E.5) is justified since from the particular upper bound of K^t from (E.13) it can be inferred that $c > 1$.

From the substitution $\frac{1}{\left(2+\frac{\varepsilon M}{2L}\right)^{2K}} = \rho\varepsilon^c$ where $2K = \frac{\log\left(\left(2+\frac{\varepsilon M}{2L}\right)\log\left(\frac{2+\frac{\varepsilon M}{2L}}{1+\frac{\beta}{L}-\frac{\varepsilon M}{2L}}\right)\frac{2\delta}{\varepsilon Mn}\right)}{\log\left(\frac{2+\frac{\varepsilon M}{2L}}{1+\frac{\beta}{L}-\frac{\varepsilon M}{2L}}\right)}$ we have that

$$\log\left(\frac{1}{\rho\varepsilon^c}\right) = 2K\log\left(2+\frac{\varepsilon M}{2L}\right) \tag{E.14}$$

$$c\log\left(\frac{1}{\sqrt[c]{\rho}\varepsilon}\right) = \frac{\log\left(\left(2+\frac{\varepsilon M}{2L}\right)\log\left(\frac{2+\frac{\varepsilon M}{2L}}{1+\frac{\beta}{L}-\frac{\varepsilon M}{2L}}\right)\frac{2\delta}{\varepsilon Mn}\right)}{\log\left(\frac{2+\frac{\varepsilon M}{2L}}{1+\frac{\beta}{L}-\frac{\varepsilon M}{2L}}\right)}\log\left(2+\frac{\varepsilon M}{2L}\right) \tag{E.15}$$

$$c = \frac{\log\left(2+\frac{\varepsilon M}{2L}\right)}{\log\left(2+\frac{\varepsilon M}{2L}\right)-\log\left(1+\frac{\beta}{L}-\frac{\varepsilon M}{2L}\right)} > 1, \tag{E.16}$$

where we have $\log\left(\frac{1}{\sqrt[c]{\rho}\varepsilon}\right) = \log\left(\left(2+\frac{\varepsilon M}{2L}\right)\log\left(\frac{2+\frac{\varepsilon M}{2L}}{1+\frac{\beta}{L}-\frac{\varepsilon M}{2L}}\right)\frac{2\delta}{\varepsilon Mn}\right)$. Now with $c > 1$, we will have the following condition for any $d > 0$:

$$\lim_{\varepsilon\to 0^+}\frac{\varepsilon^{(1+c+d)}\log\left(\frac{1}{\varepsilon}\right)}{\varepsilon^2} = 0. \tag{E.17}$$

Hence, for sufficiently small ε, the term F_1 can be of at most order $\mathscr{O}(\varepsilon^2)$.

Comments on the projection $\sum_{j\in\mathscr{N}_{US}}(\theta_j^{us})^2$: Recall that from (E.7), we solved for values of K where this inequality becomes an equality. However, this solution for such K may not necessarily exist. For instance, the left hand side of (E.7) given by $\left(\frac{1+\frac{\beta}{L}-\frac{\varepsilon M}{2L}}{2+\frac{\varepsilon M}{2L}}\right)^{2K}$ is a decreasing function of K whereas the right hand side of this inequality given by $2K\left(2+\frac{\varepsilon M}{2L}\right)^{-1}\frac{\varepsilon Mn}{2\delta}+\frac{\varepsilon\left(\rho\varepsilon^{(c-1)}+\frac{MLn}{\delta(L+\beta)}\right)}{\sum_{j\in\mathscr{N}_{US}}(\theta_j^{us})^2}$ is an increasing function of K. Hence for a solution K to exist where these two quantities become equal, we must necessarily have that

$$\left(\frac{1+\frac{\beta}{L}-\frac{\varepsilon M}{2L}}{2+\frac{\varepsilon M}{2L}}\right)^{2K}\Bigg|_{K=0} > 2K\left(2+\frac{\varepsilon M}{2L}\right)^{-1}\frac{\varepsilon Mn}{2\delta}\Bigg|_{K=0}+\frac{\varepsilon\left(\rho\varepsilon^{(c-1)}+\frac{MLn}{\delta(L+\beta)}\right)}{\sum_{j\in\mathscr{N}_{US}}(\theta_j^{us})^2} \tag{E.18}$$

$$\sum_{j\in\mathscr{N}_{US}}(\theta_j^{us})^2 > \varepsilon\left(\rho\varepsilon^{(c-1)}+\frac{MLn}{\delta(L+\beta)}\right) > \varepsilon\frac{MLn}{\delta(L+\beta)}, \tag{E.19}$$

where we can set $\Delta > \varepsilon\frac{MLn}{\delta(L+\beta)}$ and therefore require the condition $\sum_{j\in\mathscr{N}_{US}}(\theta_j^{us})^2 \geqslant \Delta$. Note that this is only a necessary condition for the existence of K from (E.13) but is not sufficient.

E.0.2 *Case 2—Small K:.* Recall that while developing the inequality (E.5) from (E.4), we used the fact that K is sufficiently large. However, for very small values of K, i.e., $K < \mathscr{O}\left(\log\left(\frac{1}{\varepsilon}\right)\right)$, the transformation of the inequality (E.4) into (E.5) may not necessarily hold true. In that case, a different approach is required to solve for K. Since the new solutions for K will be very small values, we can skip the analysis for small K case and extrapolate it to the previous result of $K \leqslant K^{\iota} = \mathscr{O}\left(\log\left(\frac{1}{\varepsilon}\right)\right)$ which is a linear exit time solution. We now complete the proof of Theorem 3.3 by establishing one last result.

E.0.3 *Claim: The set of ε-precision trajectories with linear exit times from the ball $\mathscr{B}_{\varepsilon}(\mathbf{x}^*)$ is non-empty.*
Proof. Observe that from (E.4), we need to find K where this approximate inequality becomes an equality. Let the initial condition be such that $\sum_{j\in\mathscr{N}_{US}}(\theta_j^{us})^2 = 1$; then (E.4) can be given by

$$1 \gtrapprox \left(1+\frac{\beta}{L}-\frac{\varepsilon M}{2L}\right)^{2K} - \left[2K\left(2+\frac{\varepsilon M}{2L}\right)^{2K-1}\frac{\varepsilon Mn}{2\delta} + \varepsilon MLn\frac{\left(2+\frac{\varepsilon M}{2L}\right)^{2K}}{\delta(L+\beta)}\right] \tag{E.20}$$

$$\left(2+\frac{\varepsilon M}{2L}\right)^{-2K} \gtrapprox \left(\frac{1+\frac{\beta}{L}-\frac{\varepsilon M}{2L}}{2+\frac{\varepsilon M}{2L}}\right)^{2K} - \underbrace{\left[2K\left(2+\frac{\varepsilon M}{2L}\right)^{-1}\frac{\varepsilon Mn}{2\delta} + \frac{\varepsilon MLn}{\delta(L+\beta)}\right]}_{L_1}. \tag{E.21}$$

It is easy to infer that the right-hand side of (E.21) is negative for $K = \dfrac{\log\left(\left(2+\frac{\varepsilon M}{2L}\right)\log\left(\frac{2+\frac{\varepsilon M}{2L}}{1+\frac{\beta}{L}-\frac{\varepsilon M}{2L}}\right)\frac{2\delta}{\varepsilon Mn}\right)}{2\log\left(\frac{2+\frac{\varepsilon M}{2L}}{1+\frac{\beta}{L}-\frac{\varepsilon M}{2L}}\right)}$ where this value of K comes from (E.13). Hence, the approximate inequality in (E.21) holds for this value of K. However, for small positive values of K, one can check that the right-hand side of (E.21) is greater than its left-hand side, provided ε is sufficiently small and the problem is well-conditioned. This is because the term L_1 on the right-hand side of (E.21) is of order $\mathscr{O}(\varepsilon)$ for small positive values of K whereas we have that $\left(2+\frac{\varepsilon M}{2L}\right)^{-2K} < \left(\frac{1+\frac{\beta}{L}-\frac{\varepsilon M}{2L}}{2+\frac{\varepsilon M}{2L}}\right)^{2K}$ for any positive K.

Therefore, the approximate inequality in (E.21) becomes an equality for some $K = \mathscr{O}(\log(\varepsilon^{-1}))$ and we have that $K^{\iota} = \mathscr{O}(\log(\varepsilon^{-1}))$. As a result, the exit time K_{exit} is linear for the initial condition $\sum_{j\in\mathscr{N}_{US}}(\theta_j^{us})^2 = 1$ since $K_{exit} < K^{\iota}$. It should be noted that the proof of a linear exit time for the general initial condition $\Delta \leqslant \sum_{j\in\mathscr{N}_{US}}(\theta_j^{us})^2 < 1$ can be developed along similar lines though it may require more effort. □

F. Counterexample to the monotonicity property

Consider a trajectory of the gradient descent method that satisfies the following boundary condition for some $\rho \in (0,1)$:

$$\frac{M\varepsilon^2}{2\beta(1-\rho)} = \frac{M\|\mathbf{x}_0-\mathbf{x}^*\|^2}{2\beta(1-\rho)} > \langle \mathbf{v}_n, \mathbf{x}_0-\mathbf{x}^*\rangle \geqslant \underbrace{\langle \mathbf{v}_n, \mathbf{x}_1-\mathbf{x}^*\rangle}_{I_1} \geqslant \frac{M\|\mathbf{x}_1-\mathbf{x}^*\|^2}{2\beta(1-\rho)}, \tag{F.1}$$

where $\|\mathbf{x}_0-\mathbf{x}^*\|=\varepsilon$ by definition. This trajectory violates the strict monotonicity of $\langle \mathbf{v}_n, \mathbf{x}_k-\mathbf{x}^*\rangle$ for $k=0$. But it is straightforward to see that the condition I_1 along with (3.21)–(3.23), in which ε is replaced by $\|\mathbf{x}_1-\mathbf{x}^*\|$, ensures geometric growth of $\langle \mathbf{v}_n, \mathbf{x}_k-\mathbf{x}^*\rangle$ for all $k\geqslant 1$. This in turn guarantees linear exit time from $\mathscr{B}_\varepsilon(\mathbf{x}^*)$ for the trajectory starting at $\mathbf{x}_0$, even though the strict monotonicity property is violated. Now our goal is to prove the existence of at-least one such trajectory that satisfies (F.1) so as to construct the counterexample. In order to have the condition $\langle \mathbf{v}_n, \mathbf{x}_0-\mathbf{x}^*\rangle \geqslant \langle \mathbf{v}_n, \mathbf{x}_1-\mathbf{x}^*\rangle$ from (F.1) with $\|\mathbf{x}_1-\mathbf{x}^*\|\leqslant\varepsilon$, we require

$$\langle \mathbf{v}_n, \mathbf{x}_0-\mathbf{x}^*\rangle \geqslant \langle \mathbf{v}_n, \mathbf{x}_1-\mathbf{x}^*\rangle \tag{F.2}$$

$$\iff \langle \mathbf{v}_n, \mathbf{x}_0-\mathbf{x}^*\rangle \geqslant \langle \mathbf{v}_n, \mathbf{x}_0-\mathbf{x}^*\rangle - \alpha\langle \mathbf{v}_n, \nabla f(\mathbf{x}_0)\rangle \tag{F.3}$$

$$\iff \alpha\langle \mathbf{v}_n, \nabla f(\mathbf{x}_0)\rangle \geqslant 0 \tag{F.4}$$

$$\iff \left\langle \mathbf{v}_n, \int_{p=0}^{p=1} \nabla^2 f(\mathbf{x}^*+p(\mathbf{x}_0-\mathbf{x}^*))(\mathbf{x}_0-\mathbf{x}^*)dp \right\rangle \geqslant 0 \tag{F.5}$$

$$\iff \langle \mathbf{v}_n, \nabla^2 f(\mathbf{x}^*)(\mathbf{x}_0-\mathbf{x}^*)\rangle + \langle \mathbf{v}_n, P(\mathbf{x}_0)(\mathbf{x}_0-\mathbf{x}^*)\rangle \geqslant 0, \tag{F.6}$$

where $P(\mathbf{x}_0)=\int_{p=0}^{p=1}\nabla^2 f(\mathbf{x}^*+p(\mathbf{x}_0-\mathbf{x}^*))dp-\nabla^2 f(\mathbf{x}^*)$ and $\|P(\mathbf{x}_0)\|_2\leqslant\frac{M\varepsilon}{2}$. Next, without loss of generality, write $\mathbf{x}_0-\mathbf{x}^*=\varepsilon\sum_{j=1}^{n}a_j\mathbf{v}_j$ with $a_j\in[0,1]$ for all j, $\sum_j a_j^2=1$, and $a_n=\frac{\langle \mathbf{v}_n,\mathbf{x}_0-\mathbf{x}^*\rangle}{\varepsilon}=\frac{M\varepsilon\sigma}{2\beta(1-\rho)}$ for some positive σ (note that σ cannot be 1 since we require the condition $\langle \mathbf{v}_n, \mathbf{x}_0-\mathbf{x}^*\rangle < \frac{M\varepsilon^2}{2\beta(1-\rho)}$ from the left-hand-side of (F.1)). Substituting the expression for $\mathbf{x}_0-\mathbf{x}^*$ in (F.6) followed by substituting a_n and using the fact that $\lambda_n\geqslant -L$ from Assumption **A2** yields

$$\langle \mathbf{v}_n, \nabla f(\mathbf{x}_0)\rangle = \langle \mathbf{v}_n, (\nabla^2 f(\mathbf{x}^*)+P(\mathbf{x}_0))(\mathbf{x}_0-\mathbf{x}^*)\rangle \tag{F.7}$$

$$= \varepsilon a_n\lambda_n + \langle \mathbf{v}_n, P(\mathbf{x}_0)(\mathbf{x}_0-\mathbf{x}^*)\rangle \geqslant -L\frac{M\varepsilon^2\sigma}{2\beta(1-\rho)} + \varepsilon\langle \mathbf{v}_n, P(\mathbf{x}_0)\sum_{j=1}^{n}a_j\mathbf{v}_j\rangle \geqslant 0. \tag{F.8}$$

Now there will exist some twice continuously differentiable function $f(\cdot)$ for which $\|P(\mathbf{x}_0)\|_2=\frac{M\varepsilon}{2}$ for a given $\mathbf{x}_0$. Writing $P(\mathbf{x}_0)$ in terms of the $\mathbf{v}_j$'s using the rank-one decomposition we get $P(\mathbf{x}_0)=\sum_{i=1}^{n}\sum_{j=1}^{n}c_{ij}\mathbf{v}_i\mathbf{v}_j^T$ where $c_{ij}=c_{ji}$ since $P(\mathbf{x}_0)$ is symmetric and we have the constraint $\frac{M\varepsilon}{2}\leqslant\sqrt{\sum_{i=1}^{n}\sum_{j=1}^{n}c_{ij}^2}\leqslant\frac{Mn\varepsilon}{2}$. Hence one can fix $c_{in}=\frac{M\varepsilon}{2}a_i$ for some twice continuously differentiable $f(\cdot)$ and substitute the resulting $P(\mathbf{x}_0)$ into (F.8) to get

$$-L\frac{M\varepsilon^2\sigma}{2\beta(1-\rho)} + \varepsilon\langle \mathbf{v}_n, P(\mathbf{x}_0)\sum_{j=1}^{n}a_j\mathbf{v}_j\rangle \geqslant 0 \tag{F.9}$$

$$\iff \varepsilon\left\langle \sum_{i=1}^{n}\sum_{j=1}^{n}c_{ij}\mathbf{v}_i\mathbf{v}_j^T\mathbf{v}_n, \sum_{j=1}^{n}a_j\mathbf{v}_j\right\rangle \geqslant L\frac{M\varepsilon^2\sigma}{2\beta(1-\rho)} \tag{F.10}$$

$$\iff \varepsilon\left\langle \sum_{i=1}^{n}c_{in}\mathbf{v}_i, \sum_{j=1}^{n}a_j\mathbf{v}_j\right\rangle \geqslant L\frac{M\varepsilon^2\sigma}{2\beta(1-\rho)} \tag{F.11}$$

$$\iff \frac{M\varepsilon^2}{2} \geqslant L\frac{M\varepsilon^2\sigma}{2\beta(1-\rho)} \tag{F.12}$$

$$\iff \frac{\beta(1-\rho)}{L}\geqslant\sigma. \tag{F.13}$$

Also, from (3.14) we have that $\mathbf{x}_1 - \mathbf{x}^* = (\mathbf{I} - \alpha\nabla^2 f(\mathbf{x}^*))(\mathbf{x}_0 - \mathbf{x}^*) - \alpha r(\mathbf{x}_0)$ where $\|r(\mathbf{x}_0)\| \leqslant \frac{M\varepsilon^2}{2}$. Hence, $\|\mathbf{x}_1 - \mathbf{x}^*\| \leqslant \varepsilon\sqrt{\sum_{j=1}^{n}(1-\alpha\lambda_j)^2 a_j^2} + \frac{M\varepsilon^2}{2} < \varepsilon(\sqrt{(1-\alpha\beta)^2 + 4a_n^2} + \frac{M\varepsilon}{2}) = \varepsilon(\sqrt{(1-\alpha\beta)^2 + \mathscr{O}(\varepsilon^2)} + \frac{M\varepsilon}{2})$. Next, using (F.8) and (F.13) we get that $\langle \mathbf{v}_n, \mathbf{x}_1 - \mathbf{x}^* \rangle = \langle \mathbf{v}_n, \mathbf{x}_0 - \mathbf{x}^* \rangle - \alpha\langle \mathbf{v}_n, \nabla f(\mathbf{x}_0)\rangle = \frac{M\varepsilon^2\sigma}{2\beta(1-\rho)} - \mathscr{O}(\varepsilon^3) = \frac{M\varepsilon^2}{2L} - \mathscr{O}(\varepsilon^3)$ by choosing $\sigma = \frac{\beta(1-\rho)}{L} - \mathscr{O}(\varepsilon)$ when evaluating $\langle \mathbf{v}_n, \nabla f(\mathbf{x}_0)\rangle$. For the inequality I_1 to hold, we require $\langle \mathbf{v}_n, \mathbf{x}_1 - \mathbf{x}^* \rangle \geqslant \frac{M\|\mathbf{x}_1 - \mathbf{x}^*\|^2}{2\beta(1-\rho)}$. This can be achieved by requiring the condition

$$\langle \mathbf{v}_n, \mathbf{x}_1 - \mathbf{x}^* \rangle = \frac{M\varepsilon^2\sigma}{2\beta(1-\rho)} - \mathscr{O}(\varepsilon^3) \geqslant \frac{M\varepsilon^2}{2\beta(1-\rho)}\left(\sqrt{(1-\alpha\beta)^2 + \mathscr{O}(\varepsilon^2)} + \frac{M\varepsilon}{2}\right)^2 \geqslant \frac{M\|\mathbf{x}_1 - \mathbf{x}^*\|^2}{2\beta(1-\rho)} \tag{F.14}$$

$$\implies \frac{\beta(1-\rho)}{L} - \mathscr{O}(\varepsilon) = \sigma \geqslant \left(\sqrt{(1-\alpha\beta)^2 + \mathscr{O}(\varepsilon^2)} + \frac{M\varepsilon}{2}\right)^2 + \mathscr{O}(\varepsilon). \tag{F.15}$$

Now both (F.13) and (F.15) will be satisfied for $\alpha = \frac{1}{L}$ provided $\frac{\beta}{L}$ is close to 1, ε is sufficiently small and ρ is not too large. Hence we have obtained a value of σ and in turn $a_n = \frac{\langle \mathbf{v}_n, \mathbf{x}_0 - \mathbf{x}^* \rangle}{\varepsilon} = \frac{M\varepsilon\sigma}{2\beta(1-\rho)}$, i.e., the initial boundary condition, for which (F.1) is satisfied on some twice continuously differentiable function $f(\cdot)$.

References

[1] ANANDKUMAR, A. & GE, R. (2016) Efficient approaches for escaping higher order saddle points in non-convex optimization. in *Proc. Conf. Learning Theory*, pp. 81–102.

[2] ATTOUCH, H., BOLTE, J. & SVAITER, B. F. (2013) Convergence of descent methods for semi-algebraic and tame problems: Proximal algorithms, forward–backward splitting, and regularized Gauss–Seidel methods. *Mathematical Programming*, **137**(1-2), 91–129.

[3] BOLTE, J., DANIILIDIS, A. & LEWIS, A. (2007) The Łojasiewicz inequality for nonsmooth subanalytic functions with applications to subgradient dynamical systems. *SIAM Journal on Optimization*, **17**(4), 1205–1223.

[4] BOLTE, J., DANIILIDIS, A., LEY, O. & MAZET, L. (2010) Characterizations of Łojasiewicz inequalities: Subgradient flows, talweg, convexity. *Transactions of the American Mathematical Society*, **362**(6), 3319–3363.

[5] CANDES, E. J., LI, X. & SOLTANOLKOTABI, M. (2015) Phase retrieval via Wirtinger flow: Theory and algorithms. *IEEE Transactions on Information Theory*, **61**(4), 1985–2007.

[6] CAPPELLARO, P. (2012) Matrix perturbation theory. Course Notes on *Quantum Theory of Radiation Interactions*, Massachusetts Institute of Technology, Accessed: 2019-08-19.

[7] CHEN, Y., CHI, Y., FAN, J. & MA, C. (2019) Gradient descent with random initialization: Fast global convergence for nonconvex phase retrieval. *Mathematical Programming*, **176**(1), 5–37.

[8] CORLESS, R. M., GONNET, G. H., HARE, D. E. G., JEFFREY, D. J. & KNUTH, D. E. (1996) On the Lambert W function. *Advances in Computational Mathematics*, **5**(1), 329–359.

[9] DANESHMAND, H., KOHLER, J., LUCCHI, A. & HOFMANN, T. (2018) Escaping saddles with stochastic gradients. in *Proc. 35th International Conference on Machine Learning*, pp. 1155–1164.

[10] DAVIS, C. & KAHAN, W. M. (1970) The rotation of eigenvectors by a perturbation. III. *SIAM Journal on Numerical Analysis*, **7**(1), 1–46.

[11] DIXIT, R., GÜRBÜZBALABAN, M. & BAJWA, W. U. (2021) Boundary Conditions for Linear Exit Time Gradient Trajectories Around Saddle Points: Analysis and Algorithm. *arXiv preprint arXiv:2101.02625*.

[12] DU, S. S., JIN, C., LEE, J. D., JORDAN, M. I., SINGH, A. & POCZOS, B. (2017) Gradient descent can take exponential time to escape saddle points. in *Proc. Advances in Neural Information Processing Systems*, pp. 1067–1077.

[13] ERDOGDU, M. A., MACKEY, L. & SHAMIR, O. (2018) Global non-convex optimization with discretized diffusions. in *Proc. Advances in Neural Information Processing Systems (NeurIPS'18)*, pp. 9671–9680.

[14] FITZPATRICK, R. (2010) Degenerate perturbation theory. Course notes on *Quantum Mechanics*, The University of Texas at Austin, Accessed: 2019-08-19.

[15] HU, W. & LI, C. J. (2021) On the fast convergence of random perturbations of the gradient flow. *Asymptotic Analysis*, **122**(3-4), 371–393.

[16] JAGANATHAN, K., ELDAR, Y. C. & HASSIBI, B. (2016) STFT phase retrieval: Uniqueness guarantees and recovery algorithms. *IEEE Journal of selected topics in signal processing*, **10**(4), 770–781.

[17] JIN, C., GE, R., NETRAPALLI, P., KAKADE, S. M. & JORDAN, M. I. (2017) How to escape saddle points efficiently. in *Proc. 34th International Conference on Machine Learning*, pp. 1724–1732. JMLR. org.

[18] JIN, C., NETRAPALLI, P. & JORDAN, M. I. (2018) Accelerated gradient descent escapes saddle points faster than gradient descent. in *Proc. 31st Conference on Learning Theory*, pp. 1042–1085.

[19] KELLEY, A. (1966) The stable, center-stable, center, center-unstable, unstable manifolds. *Journal of Differential Equations*.

[20] KIFER, Y. (1981) The exit problem for small random perturbations of dynamical systems with a hyperbolic fixed point. *Israel Journal of Mathematics*, **40**(1), 74–96.

[21] KURDYKA, K. (1998) On gradients of functions definable in o-minimal structures. *Annales de l'Institut Fourier*, **48**(3), 769–783.

[22] KUROCHKIN, S. V. (2021) Neural network with smooth activation functions and without bottlenecks is almost surely a Morse function. *Computational Mathematics and Mathematical Physics*, **61**(7), 1162–1168.

[23] LEE, J. D., PANAGEAS, I., PILIOURAS, G., SIMCHOWITZ, M., JORDAN, M. I. & RECHT, B. (2017) First-order methods almost always avoid saddle points. *arXiv preprint arXiv:1710.07406*.

[24] LEE, J. D., SIMCHOWITZ, M., JORDAN, M. I. & RECHT, B. (2016) Gradient descent converges to minimizers. *arXiv preprint arXiv:1602.04915*.

[25] LI, G. & PONG, T. K. (2018) Calculus of the exponent of Kurdyka–Łojasiewicz inequality and its applications to linear convergence of first-order methods. *Foundations Computational Mathematics*, **18**, 1199–1232.

[26] ŁOJASIEWICZ, S. (1961) *Sur le probleme de la division*. Instytut Matematyczny Polskiej Akademi Nauk (Warszawa).

[27] MA, C., WANG, K., CHI, Y. & CHEN, Y. (2020) Implicit Regularization in Nonconvex Statistical Estimation: Gradient Descent Converges Linearly for Phase Retrieval, Matrix Completion, and Blind Deconvolution.. *Foundations of Computational Mathematics*, **20**(3).

[28] MATSUMOTO, Y. (2002) *An introduction to Morse theory*, vol. 208. American Mathematical Soc.

[29] MEI, S., BAI, Y. & MONTANARI, A. (2018) The landscape of empirical risk for nonconvex losses. *The Annals of Statistics*, **46**(6A), 2747–2774.

[30] MOKHTARI, A., OZDAGLAR, A. & JADBABAIE, A. (2018) Escaping saddle points in constrained optimization. in *Proc. Advances in Neural Information Processing Systems*, pp. 3629–3639.

[31] ——— (2019) Efficient nonconvex empirical risk minimization via adaptive sample size methods. in *Proc. 22nd International Conference on Artificial Intelligence and Statistics*, pp. 2485–2494.

[32] MURRAY, R., SWENSON, B. & KAR, S. (2019) Revisiting normalized gradient descent: Fast evasion of saddle points. *IEEE Transactions on Automatic Control*, **64**(11), 4818–4824.

[33] NESTEROV, Y. & POLYAK, B. T. (2006) Cubic regularization of Newton method and its global performance. *Mathematical Programming*, **108**(1), 177–205.

[34] O'NEILL, M. & WRIGHT, S. J. (2019) Behavior of accelerated gradient methods near critical points of nonconvex functions. *Mathematical Programming*, **176**(1-2), 403–427.

[35] PATERNAIN, S., MOKHTARI, A. & RIBEIRO, A. (2019) A Newton-based method for nonconvex optimization with fast evasion of saddle points. *SIAM Journal on Optimization*, **29**(1), 343–368.

[36] POLYAK, B. T. (1964) Some methods of speeding up the convergence of iteration methods. *USSR Computational Mathematics and Mathematical Physics*, **4**(5), 1–17.

[37] RAGINSKY, M., RAKHLIN, A. & TELGARSKY, M. (2017) Non-convex learning via stochastic gradient Langevin dynamics: A nonasymptotic analysis. in *Proc. Conf. Learning Theory (COLT'17)*, pp. 1674–1703, Amsterdam, Netherlands.

[38] REDDI, S. J., ZAHEER, M., SRA, S., POCZOS, B., BACH, F., SALAKHUTDINOV, R. & SMOLA, A. J. (2018) A generic approach for escaping saddle points. in *Proc. 21st Intl. Conf. Artificial Intelligence and Statistics (AISTATS'18)*, pp. 1233–1242.

[39] SHI, B., SU, W. J. & JORDAN, M. I. (2020) On learning rates and Schrödinger operators. *arXiv preprint arXiv:2004.06977*.

[40] SHUB, M. (2013) *Global stability of dynamical systems*. Springer Science & Business Media.

[41] XU, Y., RONG, J. & YANG, T. (2018) First-order stochastic algorithms for escaping from saddle points in almost linear time. in *Proc. Advances in Neural Information Processing Systems*, pp. 5530–5540.

[42] YANG, J., HU, W. & LI, C. J. (2021) On the fast convergence of random perturbations of the gradient flow. *Asymptotic Analysis*, **122**(3-4), 371–393.

[43] ZHONG, Y. (2017) Eigenvector under random perturbation: A nonasymptotic Rayleigh-Schrödinger theory. *arXiv preprint arXiv:1702.00139*.