ANALYSIS OF CLOSED-LOOP INERTIAL GRADIENT DYNAMICS

Subhransu S. Bhattacharjee*
School of Engineering
The Australian National University
Canberra, Australia 2601

Ian R. Petersen
Professor, School of Engineering
The Australian National University
Canberra, Australia 2601

Abstract—In this paper, we analyse the performance of the closed-loop Whiplash gradient descent algorithm [1] for $L$-smooth convex cost functions. Using numerical experiments, we study the algorithm’s performance for convex cost functions, for different condition numbers. We analyse the convergence of the momentum sequence using symplectic integration and introduce the concept of relaxation sequences which analyses the non-classical character of the whiplash method. Under the additional assumption of invexity, we establish a momentum-driven adaptive convergence rate. Furthermore, we introduce an energy method for predicting the convergence rate with convex cost functions for closed-loop inertial gradient dynamics, using an integral anchored energy function and a novel lower bound asymptotic notation, by exploiting the bounded nature of the solutions. Using this, we establish a polynomial convergence rate for the whiplash inertial gradient system, for a family of scalar quadratic cost functions and an exponential rate for a quadratic scalar cost function.

Index Terms—Optimization; Non-linear dynamics and control

1. INTRODUCTION

In the field of continuous optimization, we study unconstrained minimisation in a finite-dimensional setting. We revisit classical problems in optimization theory, which form the heart of popular deep learning algorithms. Though Nemirovsky [2] and Nesterov [3] introduced decades ago an oracle-machine perspective into the field of optimization, it did not receive due attention. It is only with the recent treatment of optimization methods as ordinary differential equations (ODEs), that oracle machine perspective of solving black box problems have become popular [4], in the field of systems theory. While studying such algorithms in continuous-time, we consider finite-dimensional global minimisation problems with optima $x^* \in \mathcal{X}$ and the optimal cost

$$f^* = \min_{x \in \mathcal{X}} f(x),$$

(1)

where $f : \mathbb{R}^d \to \mathbb{R}$. A central aspect of analysing gradient-based learning methods is to make necessary assumptions to solve a class of problems. One such assumption is the Lipschitz continuity of the gradient of the cost function, where there exists a constant $L$, termed as the Lipschitz constant, such that:

$$\|\nabla f(y) - \nabla f(x)\| \leq L \|y - x\| \quad \forall x, y \in \mathbb{R}^d,$$

(2)

where $\|\cdot\|$ indicates the $\ell_2$ norm. We make this assumption for our cost functions throughout our analysis, unless specified otherwise. When we discuss iterative gradient schemes in optimization, a sufficient condition for learning the gradient of such a cost function is to take a small enough step-size $s$ such that:

$$0 < s \leq \frac{1}{L}.$$

(3)

We consider computational models that make queries to an oracle which contains a finite-dimensional vector map of the entire linear span of gradients of a cost function [2]. Unlike other optimization methods, we use a black-box model of optimization which yields the gradient of the cost function only at the point of the query. In this paper, we are specifically investigating optimisation methods in the accelerated framework [3]. The question: “Why acceleration speeds up rates of convergences, and how it ties to the exact design structure of the algorithms?”, is still an active area of research. We gain some intuition from the method of estimate sequences, introduced by Nesterov [3]. However, continuous-time methods to analyze optimization algorithms provide us with more intuition. A variational approach developed by Wilson et al. [5], for continuous-time systems, described a generalised Lyapunov analysis framework. Using this framework, they showed that “there is an equivalence between the technique of estimate sequences and a family of Lyapunov functions in both continuous and discrete-time”. Hence, carefully designed time-scaled energy functions have become a staple to prove convergence rates for optimization algorithms which can be studied as continuous-time inertial gradient flow dynamical systems. Shi et al. [6], demonstrate the impracticality of using discrete schemes to find discrete Lyapunov functions. Though in continuous time, this is much more intuitive, the task of finding such energy functions is complicated and can only be developed vaguely by using physical energy-based analogies. For closed-loop non-autonomous ODEs, which use damping coefficients, this task often becomes very difficult. This is because the damping terms themselves depend on the dynamics of the system. In their recent paper, Attouch et al. [7] uses a parametric manufactured solution for a particular cost function, to consider a specific family of dynamical systems. However, this method is limited from a control-theoretic perspective as it does not provide any energy method to prove the convergence rate for the cost function. The aim of our study is to realise a generalised framework to solve unconstrained global optimization problems, using a control-theoretic acceleration framework. The objectives of our immediate study are as follows:

1. To understand the phenomenon of oscillation attenuation of the objective value for the whiplash method [1].

2. A Lyapunov analysis for closed-loop inertial gradient systems.

*Mr. Subhransu Bhattacharjee is the corresponding author for this paper. Please direct all queries to him at his official mail u7143478@anu.edu.au.
**Organisation:** To address our first objective, we study the structure of the whiplash gradient system and assess its performance for a variety of cost functions. We apply the whiplash gradient descent method to smooth and convex cost functions and study the performance under various starting velocities. We further show that the whiplash method allows relaxations that stabilise the non-linear iterations, leading to attenuation of oscillations. For the second objective, we introduce a method that verifies the lower-bound rate of convergences for particular cost functions. Using a novel energy method, we analyse the stability of time-scaled inertial gradient systems, using an integral anchor constraint. By considering an asymptotic lower bound on the constraint, we predict the convergence rate of the objective cost. Finally, we verify the technique on the whiplash method.

## 2. Background

### A. Preliminaries

We assume the cost function $f(x)$ is such that $f : \mathbb{R}^d \rightarrow \mathbb{R}$ is a twice continuously differentiable convex function such that for all $\epsilon \in [0, 1]$ [8]:

$$f(\epsilon x + (1 - \epsilon)y) \leq \epsilon f(x) + (1 - \epsilon)f(y).$$

(4)

This implies that the Jensen’s inequality holds

$$f(y) \geq f(x) + \langle \nabla f(x), y - x \rangle.$$

(5)

Rearranging (5), we have:

$$0 \leq f(x) - f(y) \leq \langle \nabla f(x), x - y \rangle.$$

(6)

For $L$-smooth convex functions (which have Lipschitz continuous gradient), where we choose the step-size as $s = \frac{1}{\lambda}$, $\forall x, y \in \mathbb{R}$, for $\lambda > 0$ we have

$$f(y) \leq f(x) + \langle \nabla f(x), y - x \rangle + \frac{L}{2} \|y - x\|^2.$$

(7)

We will additionally consider the Polyak-Łojasiewicz inequality, which implies invexity [9], which is a weaker criterion than strong convexity. If we prove convergence for the P-L condition, it would naturally hold for stronger convexity criterion. The P-L inequality is said to be satisfied, if the following holds for some $\lambda > 0$ and for every $x \in \mathbb{R}$:

$$f(x) - f^* \leq \frac{1}{2\lambda} \|\nabla f(x)\|^2.$$

(8)

We use the asymptotic Landau notations $O$ and $o$, which denote the loose and strict asymptotic upper bounds, defined for $t_0 \in [0, \infty)$ respectively for the rate $\nu(t)$ as $\nu(t)$ as:

$$\lim_{t \rightarrow t_0} [\nu(t)F(t)] < \infty \implies F(t) = O\left(\frac{1}{\nu(t)}\right),$$

(9)

$$\lim_{t \rightarrow t_0} [\nu(t)F(t)] = 0 \implies F(t) = o\left(\frac{1}{\nu(t)}\right).$$

(10)

**Notation:** $\mathbb{R}^+ = \{x \in \mathbb{R} \mid x \geq 0\}$; $\mathbb{R}^*_+ = \{x \in \mathbb{R} \mid x > 0\}$.

### B. The whiplash method

The whiplash inertial gradient method, as introduced in [1], is given as

$$\dot{x} + (1 + \epsilon\|\dot{x}\|^2) \dot{x} = \nabla f(x) = 0.$$

(11)

The symplectic Euler discretisation of (11) leads to the discrete method or algorithm as given below:

$$\dot{x}_k = x_{k+1} = x_k - \frac{\epsilon}{s} \nabla f(x_k),$$

$$z_{k+1} = z_k - k s \delta k \|z_k\|^2,$$

$$\alpha_k = 1 - \sqrt{s} - k s \|z_k\|^2,$$

(12)

where condition (3) holds. The whiplash method generates the momentum of the cost function, to scale its next move in the discrete iteration. Figure: 1 provides a representation of the algorithm of (12). We can model the whiplash gradient descent as a physical system, consisting of mass $m = \frac{1}{\sqrt{s}}$ [1], where $s > 0$ is the step-size of the algorithm, in a fluid of damping constant $c = 1$ as shown in Figure: 2. We can model the oracle as a spring of varying restoring force $-\nabla f(x)$ and the non-linear damping as an adaptive damper, connected by rigid inextensible links.

### C. Numerical Results

1. We analyse the performance of our algorithm for convex cost functions. The standard method of determining the effectiveness the algorithm is by analysing its performance for various condition numbers. Note that, for constrained minimisation, this is a benchmark test, because it allows us to analyse the tractability of the algorithm over a wide range of geometries. We perform a few experiments to test the algorithmic performance, as shown in Figure 3 using the function $f(x, y) = \frac{1}{2}(x^2 + ky^2)$ for the starting point $(1, -1)$, where the condition number $\kappa$ is varied.

---

1. All numerical experiments can be found here. Unless mentioned otherwise, the x-axis of all graphs denote time in seconds.
We consider the following lemmas which bridge from the proof. The global asymptotic stability of the whiplash gradient sequence \( \{\|x_k - x^*\|\}_{k=0}^{\infty} \) is convergent.

\[ \lim_{k \to \infty} \|z_k\| = 0 \]  

Applying the symplectic discretisation of \( \dot{x}(t) = \frac{z_{k+1} - z_k}{\sqrt{s}} \), we obtain for a sufficiently small \( s > 0 \),

\[ \lim_{k \to \infty} \|z_k\| = 0 \]  

which implies \( \{\|z_k\|\}_{k=0}^{\infty} \) is convergent as required.

**Theorem 1.** The momentum sequence \( \{\|z_k\|\}_{k=0}^{\infty} \) converges at the rate \( O(\frac{1}{\sqrt{k}}) \) for an L-smooth convex cost function.

**Proof.** From (12), using \( s = \frac{1}{\sqrt{k}} \), we obtain

\[ k\|z_k\|^2 = - (Lz_{k+1} - (L - \sqrt{L})z_k) - \nabla f(x_k) \]

Taking the norm on both sides and applying the triangle inequality, we obtain

\[ k\|z_k\|^2 \leq L\|z_{k+1}\| + L'\|z_k\| + \|\nabla f(x_k)\| \]

where \( L' = (L - \sqrt{L}) \). Now, we take limits on both sides of the inequality to obtain

\[ \lim_{k \to \infty} k\|z_k\|^2 \leq \lim_{k \to \infty} (L\|z_{k+1}\| + L'\|z_k\| + \|\nabla f(x_k)\|) \]
Using (17), we simplify the above inequality to
\[
0 \leq \lim_{k \to \infty} k \|z_k\|^3 \leq \lim_{k \to \infty} \|\nabla f(x_k)\|.
\]
Now, using (14) for a Lipschitz continuous function \( f(x) \), we obtain
\[
0 \leq \lim_{k \to \infty} k \|z_k\|^3 \leq \|\nabla f(x^*)\|.
\]
Since \( \|\nabla f(x^*)\| = 0 \), we obtain
\[
\lim_{k \to \infty} k \|z_k\|^3 = 0,
\]
which implies that the sequence \( \{\|z_k\|\}_{k=0}^\infty \) converges at the rate of \( o\left(\frac{1}{k^3}\right) \) as required.

\[\square\]

**B. Relaxation sequence for L-smooth convex cost functions**

Using numerical experiments, we observe that unlike other accelerated methods, the first step taken by the whiplash method is the largest, followed by which it increases momentum to escape the valley, after which it attenuates rapidly to the optima. Essentially, such a phenomenon is caused due to the relaxation of the convergent sequence. closed-loop methods evolve with the momentum, which assist in escaping low curvature geometries, by increasing the amplitude of transient oscillations. For higher curvature, this leads to rapid convergence towards the nearest minima. These characteristics set apart the closed-loop inertial gradient descent algorithms from classical methods. A comparative study as shown in Figure: 6, illustrates this effect. The following theorems introduce relaxation sequences for L-smooth cost functions, a set of discrete energy terms which are strictly positive, bounded and convergent and allow us to study momentum-based scaling for the whiplash method.

**Theorem 2.** The whiplash gradient scheme admits a convergent relaxation sequence \( \{\delta_k\}_{k=0}^\infty \) for convex L-smooth cost functions, so that
\[
\|x_{k+1} - x^*\|^2 \leq \|x_k - x^*\|^2 + \|\delta_k\|, \quad \forall \ k > 0, \quad \text{where} \quad \delta_k = \|z_k\|^2 + 2\|z_k\|\|x_k - x^*\| + \left(\frac{1}{\sqrt{2L}} + \frac{k}{L}\right)\|z_k\|^2.
\]

**Proof.** From (12) and (7), we know that
\[
x_{k+1} = x_k - \frac{1}{L} \nabla f(x_k) + \alpha_k z_k.
\]
Subtracting the optimiser \( x^* \) on both sides, taking its norm and squaring, we obtain
\[
\|x_{k+1} - x^*\|^2 = \|x_k - x^* + \alpha_k z_k\|^2 - \frac{2}{L} \langle x_k - x^*, \nabla f(x_k) \rangle - 2\alpha_k \frac{1}{L} \langle \nabla f(x_k), x_k - x_{k-1} \rangle + \frac{1}{L^2} \|\nabla f(x_k)\|^2.
\]
Using (6), we obtain
\[
\|x_{k+1} - x^*\|^2 = \|x_k - x^* + \alpha_k z_k\|^2 - \frac{1}{L^2} \|\nabla f(x_k)\|^2 - 2\alpha_k \frac{1}{L} \langle \nabla f(x_k), x_k - x_{k-1} \rangle.
\]
Replacing \( \alpha_k \) and rearranging, we obtain
\[
\|x_{k+1} - x^*\|^2 \leq \|x_k - x^*\| + \left(1 - \frac{1}{\sqrt{L}}\right) \|z_k\|^2 + \frac{k}{L} \|z_k\|^2.
\]
Using the triangle inequality, we obtain
\[
\|x_{k+1} - x^*\|^2 \leq \|x_k - x^*\| + \|z_k\|^2 + \frac{k}{L} \|z_k\|^2.
\]
By rearranging the terms further and using the following inequality,
\[
a^2 \leq b^2 + c^2 \implies a \leq b + c \quad \forall \ a, b, c \geq 0,
\]
we obtain the following relaxation sequence as
\[
\|x_{k+1} - x^*\| \leq \|x_k - x^*\| + \sqrt{\delta_k}, \quad \forall \ k > 0
\]
where \( \delta_k = \|z_k\|^2 + 2\|z_k\|\|x_k - x^*\| + \left(\frac{1}{\sqrt{2L}} + \frac{k}{L}\right)\|z_k\|^2 \)
as required. Using Lemma 1, 2 and Theorem 1, it follows that the relaxation sequence \( \{\|z_k\|\}_{k=0}^\infty \) converges.

\[\square\]

**Theorem 3.** Let \( f \) be an L-smooth function, which satisfies the Polyak-Łojasiewicz inequality (8) for some \( \lambda > 0 \). Then the whiplash gradient descent method converges at a rate, given by the relaxation sequence:
\[
f(x_{k+1}) - f^* \leq \left(1 - \frac{\lambda}{L}\right)^k (f(x_0) - f^*) + \frac{L}{2} \sum_{i=1}^k \beta_i,
\]
for every \( k > 0 \) where
\[
\beta_i = \left(1 - \frac{\lambda}{L}\right)^i \left(1 - \frac{1}{\sqrt{L}} - \frac{k - i}{L} \|z_{k-i}\|^2\right) \|z_{k-i}\|^2.
\]

**Proof.** Using (12) and (7), we have
\[
f(x_{k+1}) \leq f(x_k) + \langle \nabla f(x_k), z_{k+1} \rangle + \frac{L}{2} \|z_{k+1}\|^2.
\]
Upon expanding \( z_{k+1} \) and rearranging, we obtain
\[
f(x_{k+1}) \leq f(x_k) + \langle \nabla f(x_k), \alpha_k z_k - \frac{1}{L} \nabla f(x_k) \rangle + \alpha_k^2 \frac{L}{2} \|z_k\|^2 + \frac{1}{2L} \|\nabla f(x_k)\|^2 - \langle \alpha_k z_k, \nabla f(x_k) \rangle.
\]
Upon cancellation of common terms and subtracting from both sides the optimal value \( f^* \) and rearrangement, we obtain
\[
f(x_{k+1}) - f^* \leq f(x_k) - f^* - \frac{1}{2L} \|\nabla f(x_k)\|^2 + \beta_k,
\]
where \( \beta_k = \alpha_k^2 \|z_k\|^2 \). Now, applying the inequality (8), we obtain
\[
f(x_{k+1}) - f^* \leq f(x_k) - f^* - \frac{\lambda}{L} (f(x_k) - f^*) + \frac{L}{2} \beta_k.
\]
Upon further rearrangement, we obtain
\[ f(x_{k+1}) - f^* \leq \left(1 - \frac{\lambda}{L}\right) (f(x_k) - f^*) + \frac{L}{2} \beta_k. \]
Applying this recursively, we obtain the telescoping sum,
\[ f(x_{k+1}) - f^* \leq \left(1 - \frac{\lambda}{L}\right)^k (f(x_0) - f^*) + \frac{L}{2} \sum_{i=1}^k \beta_i, \]
where \( \beta_i = (1 - \frac{\lambda}{L})^i \left(1 - \frac{1}{\sqrt{E}} - \frac{k}{E} \||\|_{k-1}^2\right)^2 \||\|_{k-1}^2 \) for all \( k > 0 \) as required. Using Lemma 2 and Theorem 3, we know that the relaxation sequence \( \{|\beta_i|\}_{i=1}^\infty \) must converge and thus \( f(x_{k+1}) - f^* \) must converge at the given rate. \( \square \)

4. ENVELOPE CONVERGENCE
In the study of the field of inertial gradient dynamics, we know that such systems are globally asymptotically stable for all damping laws which are strictly positive \([1]\). We further know that if the gradient of the cost function is Lipschitz continuous, there exists a global solution which is convergent, but such a solution is non-analytical in most cases. Our analysis starts, therefore, with consideration of the asymptotic nature of globally asymptotically stable solutions of (29). To understand the rate of convergence of the system, we need to perform Lyapunov analysis on the Lipschitz continuous scaled system \( g(t) = P_{t}x(t) \). As discussed earlier, finding such Lyapunov functions is tedious and often lacks any kind of motivational rigour. Lyapunov’s fundamental argument lead to a family of energy functions which are non-increasing along the system’s dynamics, which is typically the only constraint. Therefore, instead of finding a state dependent energy function of the scaled dynamical system, we introduce using the knowledge of implicit time dependence and the system states’ bounded nature, an integral anchor constraint function to predict their rates of convergence by stability analysis.

A. Asymptotic lower bound
We will consider the asymptotic behaviour of the system and correspondingly the associated Lyapunov functions. It is therefore necessary to introduce the concept of asymptotic lower bound of functions, in this context. We have modified it reasonably to suit our requirement as \( \nu(\cdot) \), while borrowing some of its character.

**Definition.** (Asymptotic lower bound) A function \( F(t) \) which is asymptotically lower bounded is said to be convergent at the rate \( \nu(t) > 0 \), if
\[ 0 \leq \lim_{t \to \infty} |\nu(t)F(t)| < \infty \implies F(t) = \Theta\left(\frac{1}{\nu(t)}\right). \quad (20) \]
We define the lower bound notation of two real and bounded functions \( A_t, B_t : \mathbb{R}^+ \to \mathbb{R}, \) such that \( A_t = \mathcal{O}(a(t)) \) and \( B_t = \mathcal{O}(b(t)), \) where \( \mathcal{O}(a(t)) \leq \mathcal{O}(b(t)) \). We further have the following relations:
\[ \mathcal{O}(\lambda a(t)) = \text{sgn}(\lambda) \mathcal{O}(a(t)) \text{ where } \lambda \in \mathbb{R}, \]
\[ \mathcal{O}(a(t)) \leq \lambda \mathcal{O}(b(t)) \Rightarrow A_t = \text{sgn}(\lambda) \mathcal{O}(b(t)), \]
\[ \mathcal{O}(a(t)b(t)) = \mathcal{O}(a(t)) \mathcal{O}(b(t)), \]
\[ \mathcal{O}(a(t) \pm b(t)) = \pm \mathcal{O}(b(t)). \quad (21) \]
This implies that, whichever function is relatively asymptotically slower is the dominating function amongst the two and shall be represented by the “~” symbol i.e. \( A_t \sim B_t = \mathcal{O}(b(t)) \) but \( B_t \sim A_t \).

**Lemma 3.** Suppose an \( L \)-smooth function \( F(t) \) is asymptotically bounded, and convergent at the rate \( \beta(t) \) such that \( \lim_{t \to \infty} |\beta(t)F(t)| = 0. \) If the rate \( \beta(t) \) is such that
\[ \beta(t) < o\left(\frac{1}{t}\right) \forall t \geq 0, \]
then there exists \( c \geq 0 \) such that
\[ \int_{t_0}^{t} |F(s)|ds < c, \quad \forall t > t_0 \geq 0. \quad (23) \]

**Proof.** Since \( |F(t)| \) is everywhere continuous and defined for all \( t > 0 \), it follows that for any \( m > t_0 \geq 0 \) and any \( m \leq t \), we obtain
\[ \int_{t_0}^{m} |F(s)|ds = \int_{t_0}^{m} |F(s)|ds + \int_{m}^{t} |F(s)|ds. \]
Now, since \( |F(t)| \) is everywhere bounded, it follows that there exists a \( c_1 > 0 \), such that \( \int_{t_0}^{m} |F(s)|ds = c_1 \) and there exists \( c_2 > 0 \) such that for a sufficiently large \( m \), we obtain using (20) and (21) that
\[ \int_{m}^{t} |F(s)|ds \sim \int_{m}^{t} |\mathcal{O}\left(\frac{1}{\beta(s)}\right)|ds \leq c_2 \quad \forall m \leq t. \]
The result then follows with \( c_1 + c_2 = c, \) \( \square \)

**Lemma 4.** In addition to the conditions on \( F \) in Lemma 3, if we have the conditions \( \lim_{t \to \infty} |\hat{F}(t)| = 0, \) then \( \hat{F}(t) \sim -\mathcal{O}\left(\frac{1}{\nu(t)}\right). \)

**Proof.** Consider the function \( \mu(t) = \nu(t)\hat{F}(t) \). From the previous lemma, we know that \( \mu(t) \) is bounded and must converge as \( t \to \infty. \) Upon differentiation of \( \mu(t) \), we obtain \( \dot{\mu}(t) = \nu(t)\hat{F}(t) + \nu(t)\dot{F}(t). \) Now, since we know that \( \mu(t) \) is Lipschitz continuous as well, it implies that
\[ \lim_{t \to \infty} \dot{\mu}(t) = 0. \]
It follows upon rearrangement that
\[ \lim_{t \to \infty} \hat{F} = \lim_{t \to \infty} -\mathcal{O}\left(\frac{\nu(t)}{\nu^{2}(t)}\right). \]
This implies using (20) that
\[ \hat{F} \sim -\mathcal{O}\left(\frac{\nu(t)}{\nu^{2}(t)}\right), \quad (24) \]
as required. \( \square \)

**Lemma 5.** Let us consider a finite-dimensional vector-valued \( L \)-smooth function \( q(t) : \mathbb{R}^+ \to \mathbb{R}^d, \) which is asymptotically bounded, and converges to the vector \( q^* \) such that \( \|Q(t)\| = \mathcal{O}\left(\frac{1}{\nu(t)}\right), \) where \( Q(t) = q(t) - q^* \), it follows that \( Q \sim \mathcal{O}\left(\frac{1}{\nu(t)}\right). \)

**Proof.** As \( Q(t) \) is a finite-dimensional vector-valued function, there exist scalar functions \( f_i : \mathbb{R}^+ \to \mathbb{R}^d \) which are everywhere bounded, \( \lim_{t \to \infty} f_i(t) = 0 \) for every \( i \geq 1 \) and for which
\[ Q(t) = \sum_{i=1}^{d} f_i(t) \dot{e}_i, \quad (25) \]
where \( |\dot{e}_i| = 1 \) denote the unit basis vectors. We also know that
\[ \|Q(t)\| = \sqrt{\sum_{i=1}^{d} f_i^2(t)}. \quad (26) \]
Using Cauchy-Schwarz inequality on (25), and (26), it follows that
\[ Q(t) \leq \sqrt{\sum_{i=1}^{d} f_i^2(t)} \sqrt{\sum_{i=1}^{d} \epsilon_i^2} = \sqrt{\mathcal{Q}} = \sqrt{\mathcal{Q}(\frac{1}{\nu(t)})}, \]
which implies that as \( t \) grows, for a finite \( d \geq 1 \), we have
\[ Q(t) \leq \sqrt{\mathcal{Q}(\frac{1}{\nu(t)})}. \]
Therefore, it follows using (21) that \( Q(t) \sim \mathcal{O}(\frac{1}{\nu(t)}) \), as required.

\[ \tag{27} \]

\[ \tag{28} \]

\[ \tag{29} \]

\[ \tag{30} \]

\[ \tag{31} \]

\[ \tag{32} \]

\[ \tag{33} \]

\[ \tag{34} \]

**B. Method of envelope convergence**

In [1], we defined the converger \( \mathcal{P}_t > 0 \), such that for a convex cost function \( f \), we have a convergence rate of \( \mathcal{O}(\frac{1}{\nu(t)}) \). This converger must be defined \( \forall t > 0 \), such that \( \mathcal{P}_t \to 0 \) as \( t \to \infty \). A simple way of numerically verifying these results is by *enveloping* the output. We show a convergence rate for \( x(t) \); i.e. we define \( y(t) = \mathcal{P}_t(x(t) - x^*) \), and demonstrate its numerical convergence. For the sake of simplicity in our analysis, we shall consider \( x^* \) to be the zero vector. The intuition is that if \( y(t) \) is convergent, then \( x(t) \) must converge at least at the rate of \( \mathcal{O}(\frac{1}{\nu(t)}) \). Essentially, we are extending the concept of using a solution to investigate a convergence rate for its inertial gradient system, as presented in [7].

We consider an inertial gradient dynamical system of the form:
\[ \ddot{x}(t) + (1 + \alpha(t))\dot{x}(t) + \nabla f(x(t)) = 0, \]
where the scalar function \( \alpha(t) \) is a closed-loop control law. We describe the envelope convergence method as the following:

1. **Exponential rate**: We start with \( \eta = 0.1 \) and keep running simulations for increasing values of \( \eta \) by steps of size 0.05. If a transition to the stability exists in this region, it is sharp and moves from bounded to divergent solutions, within a range of 0.01. Hence, once the transition is observed, we vary \( \eta \) by steps of size 0.001. If stability is observed for a critical range, we check different starting conditions.

2. **Polynomial rate**: We start with \( \eta = 0.5 \) and keep running simulations for increasing values of \( \eta \) by steps of size 0.1. If a transition to the stability exists in this region, it is sharp and moves from bounded oscillations to divergent within a range of 0.05. Hence, once the transition is observed, we vary \( \eta \) by steps of size 0.01. If stability is observed for a critical range, we check different starting conditions.

3. Once we have verified this, we can define a system such that there exists a energy function \( \mathcal{E}(t) \), which for a particular integral anchor constraint, proves the convergence rate of the objective value.

The envelope convergence method has been summarised in Figure: 7. The following theorems describe this.

**Theorem 4.** For a twice differentiable convex cost function \( f \), the system (29) converges with a rate of:
\[ f(x) - f^* = \mathcal{O}(\frac{1}{\mathcal{P}_t}), \]
for a converger \( \mathcal{P}_t > 0 \) and \( \mathcal{P}_t > 0 \), such that \( \mathcal{P}_t x(t) \) is bounded and convergent, if there exists a constant \( c \geq 0 \), for all \( t > 0 \) and for every solution to (29), \( \mathcal{J} + c \geq 0 \) where
\[ \mathcal{J} = \int_0^t \left( \alpha \mathcal{P}_t \dot{x}, x - x^* \right) - \left( \mathcal{P}_t \nabla f(x), \dot{x} \right) - \frac{\mathcal{P}_t}{2} \left\| x - x^* + \dot{x} \right\|^2 ds. \]

**Proof.** From (5), we know that
\[ f(x) - f(x^*) \leq \left\langle \nabla f(x), x - x^* \right\rangle. \]
Now, for the system (11), we define the energy function as:
\[ \mathcal{E}_t = \mathcal{P}_t(f(x) - f^*) + \frac{\mathcal{P}_t}{2} \left\| x - x^* + \dot{x} \right\|^2 + \mathcal{J}(t) + c \geq 0. \]
Taking its time derivative we obtain
\[ \dot{\mathcal{E}}_t = \dot{\mathcal{P}}_t(f(x) - f^*) + \mathcal{P}_t \left( \nabla f(x), \dot{x} \right) + \frac{\mathcal{P}_t}{2} \left\| x - x^* + \dot{x} \right\|^2 \]
Using (29) and rearranging terms, we obtain
\[ \dot{\mathcal{E}}_t = \mathcal{P}_t \dot{x}, x - x^* \right\rangle - \mathcal{P}_t \nabla f(x), \dot{x} \right\rangle - \frac{\mathcal{P}_t}{2} \left\| x - x^* + \dot{x} \right\|^2 \]
\[ \forall t > 0 \] and hence \( \kappa = 0 \). Therefore,
\[ \dot{\mathcal{E}}_t = \mathcal{P}_t \left( f(x) - f^* - \left\langle \nabla f(x), x - x^* \right\rangle \right) \leq 0, \]
\[ \forall t > 0, \] as \( \mathcal{P}_t > 0 \). Hence, upon integration we obtain
\[ f(x) - f^* \leq \frac{\mathcal{E}_0}{\mathcal{P}_t} = \mathcal{O}(\frac{1}{\mathcal{P}_t}), \]
as required. Essentially, if the functional \( F \) has a lower bound, then there must exist a constant \( c \) for which \( F + c \geq 0 \). It follows that if \( F(x, \dot{x}, t) + c \geq 0 \), for every solution \( x(t) \) to the dynamics (29), then \( \tilde{F}_1 \geq 0 \) and \( \tilde{F}_2 \leq 0 \).

C. Integral anchor constraint

To consider a lower bound on the functional \( F \), we use the asymptotic lower bound of the states \( x(t) \) and \( \dot{x}(t) \). Now, we use the asymptotic lower bound estimates of each term in the integrand, to check if all the generalised manufactured solutions, for a particular choice of converger, satisfy the integral anchor constraint. We consider, as before, for such an analysis the two cases: polynomial and exponential convergence settings. The rationale is that all bounded and convergent solutions, must exhibit a behaviour between weak polynomial convergence and strong exponential convergence, asymptotically. Since the gradient is Lipschitz continuous (2), we know that there exists global solutions to the globally asymptotically stable system (29) and they are at least \( C^3 \)-smooth. This implies that they are continuously integrable in \( t \in [0, \infty) \). It follows that the nature of solution is implicitly time dependent, which implies that all solutions satisfy \( \|x(t) - x^*\| = O\left(\frac{1}{\theta t^p}\right) \) for an arbitrary increasing function \( \lambda(t) > 0 \) for all \( t \geq 0 \). By Lemma 5, we therefore have \( x(t) - x^* \sim O\left(\frac{1}{\lambda(t)}\right) \), which implies from the convexity of the cost function that \( f(x) - f^* \leq O\left(\frac{1}{\lambda(t)}\right) \). The following theorem considers the lower bound of the integrand in these two cases, using necessary assumptions.

**Theorem 5.** The integral \( F \), which is defined for the dynamical system (29) \( \forall t > 0 \), for a \( C^2 \) convex cost function \( f(x) \), given by (31), is asymptotically lower bounded for every possible solution to (29), i.e.

\[
F \geq c > 0,
\]

if the following assumptions hold.

**Case I: Polynomial convergence setting:** \( (t \geq t_0 > 0) \)

**Assumption 1.** \( \lambda(t) = t^p \) where \( \theta > 0 \).

**Assumption 2.** \( \alpha(t) = O\left(\frac{1}{t^p}\right) \) where \( p \geq 0 \).

**Assumption 3.** Consider the polynomial converger \( P_t = t^q \).

The critical condition is:

\[
x - x^* \sim O\left(\frac{1}{t^q}\right) : \theta > \frac{q}{2} > 0.
\]

**Proof.** Using Lemma 4 and the assumptions above, we have

\[
\dot{x}(t) \sim -O\left(\frac{1}{t^{q+1}}\right).
\]

We know that as \( t \to \infty, x \to x^* \). Following the hypothesis of Theorem 5 and using Lemma 4, we have

\[
\nabla f(x) \sim -O\left(\frac{1}{t^{q+1}}\right),
\]

since \( f(x) - f^* \) is at least \( O\left(\frac{1}{t^p}\right) \). Using (21), we have \( \|x - x^* + \dot{x}\|^2 \sim O\left(\frac{1}{t^{2q+2p}}\right) \). Using the assumptions above, we rewrite (31) as:

\[
F \sim -\int_{t_0}^{t} O\left(s^{-2q+\theta-n-1} + O\left(s^{-2q+\theta-n+1}\right) + O\left(s^{-2q+\theta-n-3}\right)\right) ds.
\]

Let \( \tau_1 = 2\theta - n + p, \tau_2 = 2\theta - n + 1 \) and \( \tau_3 = 2\theta - n + 2 \). Using (36) and Assumption 2, we have all \( \tau_1, \tau_2, \tau_3 > 0 \). Upon integration and using (21), we obtain

\[
F \sim O\left(\frac{1}{t^q}\right),
\]

where \( \theta = 2\theta - n + \min\{p, 1\} \). It follows that for any \( \phi \), the functional \( F \) is asymptotically lower bounded, as required.

**Case II: Exponential convergence setting:** \( (t \geq t_0 > 0) \)

**Assumption 4.** \( \lambda(t) = e^{-\theta t} \) where \( \theta > 0 \).

**Assumption 5.** \( \alpha(t) = O\left(e^{-p\theta t}\right) \) where \( p \geq 0 \).

**Assumption 6.** Consider the exponential converger \( P_t = e^{\theta t} \) where \( 0 < \eta < 2\theta \). The critical condition is:

\[
x - x^* \sim O\left(e^{-\theta t}\right) : \theta > \frac{\eta}{2} > 0.
\]

**Proof.** Using assumptions above, and using Lemma 4, we have

\[
\dot{x}(t) \sim -e^{-\theta t}.
\]

This implies that \( |x - x^* + \dot{x}|^2 \sim O\left(e^{-2\theta t}\right) \), and given that \( f(x) - f^* \) is at least \( O\left(e^{-\theta t}\right) \), we have using Lemma 4,

\[
\nabla f(x) \sim -e^{-\theta t},
\]

Now, we rewrite (31) as

\[
F \sim \int_{t_0}^{t} \left(O\left(e^{-\theta t-2\eta t}\right) + O\left(e^{(\eta-2\theta) t}\right) + O\left(e^{(\eta-2\theta) t}\right)\right) ds.
\]

Upon integration, we obtain using (21)

\[
F \sim O\left(e^{-\sigma t}\right).
\]

where \( \sigma = 2\theta - n \). From (40) and Assumption 5, we know that \( \sigma > 0 \). Hence, it follows that \( F \) is asymptotically lower bounded, as required.

Thus, we complete the proof of Theorem 5 by proving the integral anchor constraint assumption, to include every possible solution to the dynamics (29), such that we may consider the rate of convergence for the objective \( f(x) - f^* \). It is important to note that Theorem 5 does not assume any specific rate but a specific form for the convergence analysis, and provides a condition to guarantee convergence rates.

D. Verification of the envelope convergence method for the whiplash inertial gradient system for scalar quadratic costs

We now apply the envelope convergence method to the whiplash inertial gradient system for a particular set of convex functions. We define this set of functions as \( f(x) = \frac{1}{2}x^2 \) for every \( \lambda \in \mathbb{R}_{++}^\lambda \), where \( x \in \mathbb{R} \), to verify our envelope method, for which the system (11) reduces to the scalar ODE:

\[
\ddot{x} + (1 + t\dot{x}^2)\dot{x} + \lambda x = 0 \quad \text{where} \quad x \in \mathbb{R}.
\]

Using our envelope convergence methodology, we choose a polynomial converger and start with a small value of \( \eta \). To demonstrate the envelope convergence method for the system (44), we choose \( \eta = 2 \), using the polynomial convergence criterion as shown in the Figure: 8, which holds for every \( \lambda \in \mathbb{R}_{++}^\lambda \). For this case, we observe the transition to instability for the cost \( \hat{f} \) at \( \eta = 0.495 \) using the exponential rate as
This means, using Theorems 4 and 5, it is sufficient to show \( \tilde{\eta} \) whiplash gradient descent for the cost function, \( r > O(\hat{\eta}) \), environment, using the methods shown in Figure 8. Since, we have found using computational methods, a suitable \( P_\theta x(t) \), which converges, it is sufficient to confirm the convergence rate of \( O(\frac{1}{t}) \) for the cost function \( f = \lambda x^2 \), or a rate of \( O(e^{-0.495t}) \) for the cost function \( \tilde{f} = \frac{1}{2} x^2 \), by showing that for the dynamics (44), there exists an asymptotic lower bound solution such that it satisfies the design conditions (36) and (40) respectively. For our case, we thus have:

\[
\begin{align*}
    f(x) - f^* &= \frac{1}{2} x^2 : x(t) = \Theta(\frac{1}{\theta}) \quad \forall \theta > 1.5, \\
    \tilde{f}(x) - \tilde{f}^* &= \frac{1}{2} \tilde{x}^2 : \tilde{x}(t) = \Theta(e^{-0.04t}) \quad \forall \tilde{\theta} > 0.2475. 
\end{align*}
\]

This means, using Theorems 4 and 5, it is sufficient to show that \( \lim_{t \to \infty} e^{\theta} x(t) = 0 \) and \( \lim_{t \to \infty} e^{\tilde{\theta}} \tilde{x}(t) = 0 \). Now, since \( P_\theta x(t) \) is experimentally convergent, in both cases, it follows from (21), that both design criterion hold true for \( \theta \) and \( \tilde{\theta} \). This implies that \( \mathcal{F} \) is asymptotically lower bounded in both cases, and thus the required rates of convergence are established.

\[2\] All numerical experiments have been performed in the SIMULINK™ environment, using the Euler ode solver with a fixed step-size of 0.001.

\[3\] Note that as per the requirement of Theorem 5, in (44), for the cost function \( \tilde{f} \), the control law \( \alpha(t) = \Theta(\frac{\tilde{\theta}}{\tilde{\theta}}) \), where \( p > 0 \). Though this does not agree explicitly with Assumption 5, the \( \Theta \) lower bound enables that there exists some \( r > 0 \), such that \( \Theta(\frac{\tilde{\theta}t}{\tilde{\theta}t}) = \Theta(e^{-rt}) \).

5. Discussion & Further Work

In this paper, we have presented a numerical approach that provides a practical methodology for analysing convergence rates. Essentially, what we have here is a generalised Lyapunov method which allows for rate prediction of inertial gradient flow systems. This method establishes using an integral anchored Lyapunov function, rates of convergences by using lower bound asymptotic analyses. The cornerstone of this approach lies in numerical computation of a converger, which separates it from classical methods of establishing stability. In other words, we have reduced the analytical task of proving convergence rates by finding Lyapunov functions into the prediction of rates of convergence by finding appropriate convergers, which satisfy the design criterion (36) or (40). This means that if \( v(\theta, t)(x(t) - x^*) \) is convergent, then \( P_\theta(\eta)(f(x) - f^*) \) is convergent as well. Therefore, as long as we choose a target \( \eta \) with respect to a suitable \( \theta \), which does not violate the design criterion, our analysis holds true. This simplification provides a structured methodology for using experimental methods to check for the system’s stability in a feasible time, without the necessity of resorting to Lyapunov functions — specific to convergence rates, as elucidated. The whiplash method performs well for various non-convex cost functions and outperforms many classical methods. However, global optimization is a much bigger challenge to undertake, as local minimas are difficult to escape using deterministic methods. On the other hand, stochastic methods are much faster and are able to escape local minimas and saddle points. A similar hybrid algorithm, in a closed-loop control framework, is a promising direction for future work.

Acknowledgement

We are grateful to the School of Engineering, ANU for their generous support. Subhansu is grateful to Prof. Emeritus Hedy Attouch, Université de Montpellier for productive discussions.

References

[1] S. S. Bhattacharjee and I. R. Petersen, “A closed loop gradient descent algorithm applied to Rosenbrock’s function,” 2021 Australian & New Zealand Control Conference (ANZCC), Nov 2021.

[2] A. Nemirovsky and D. Yudin, “Problem complexity and method efficiency in optimization,” SIAM Review, vol. 27, no. 2, pp. 264–265, 1985.

[3] Y. Nesterov, Lectures on Convex Optimization, 2nd ed. Springer Publishing Company, Incorporated, 2018.

[4] W. Su, S. Boyd, and E. J. Candès, “A differential equation for modeling Nesterov’s accelerated gradient method: Theory and insights,” Journal of Machine Learning Research, vol. 17, no. 1, 2016.

[5] A. C. Wilson, B. Recht, and M. I. Jordan, “A Lyapunov analysis of accelerated methods in optimization,” Journal of Machine Learning Research, vol. 22, no. 113, pp. 1–34, 2021.

[6] B. Shi, S. S. Du, W. Su, and M. I. Jordan, “Acceleration via symplectic discretization of high-resolution differential equations,” in Advances in Neural Information Processing Systems, vol. 32. Curran Associates, Inc., 2019.

[7] H. Attouch, R. Boţ, and E. R. Csetnek, “Fast optimization via Inertial dynamics with closed-loop damping,” Journal of European Mathematical Society, 2021.

[8] D. Bertsekas, Convex Optimization Theory. Athena Scientific, Belmont, MA, 2009.

[9] H. Karimi, J. Nutini, and M. Schmidt, “Linear convergence of gradient and proximal-gradient methods under the Polyak-Lojasiewicz condition,” Joint European Conference on Machine Learning and Knowledge Discovery in Databases, pp. 795–811, 2016.

[10] E. Hairet, C. Lubich, and G. Wanner, Geometric numerical integration, 2nd ed., ser. Springer Series in Computational Mathematics. Springer-Verlag, Berlin, 2006.