MASTERARBEIT / MASTER'S THESIS

Titel der Masterarbeit / Title of the Master's Thesis

"Convergence rates of the Nesterov accelerated gradient method under geometric conditions and time scaling"

verfasst von / submitted by

Rossen Nenov, BSc

angestrebter akademischer Grad / in partial fulfilment of the requirements for the degree of

Master of Science (MSc)

Wien, 2022 / Vienna 2022

Studienkennzahl lt. Studienblatt / degree programme code as it appears on the student record sheet:	UA 066 821
Studienrichtung lt. Studienblatt / degree programme as it appears on the student record sheet:	Masterstudium Mathematik UG2002
Betreut von / Supervisor:	Univ.-Prof. Dr. Radu Ioan Boţ, Privatdoz.

Abstract

In dieser Masterarbeit untersuchen wir die Trajektorien von zwei Arten von Differentialgleichungen zweiter Ordnung in Verbindung mit der beschleunigten Gradientenmethode von Nesterov, einschließlich konstanter Dämpfung und asymptotisch verschwindender viskoser Dämpfung. Diese Differentialgleichungen werden durch eine konvexe differenzierbare Funktion F gesteuert, und es wurde gezeigt, dass jede Trajektorie x der Differentialgleichung minimierende Eigenschaften hat, das heißt, dass $F(x(t))$ gegen den Minimalwert F^* von F konvergiert, wenn t gegen unendlich geht. In früheren Arbeiten wurde untersucht, wie bestimmte geometrische Eigenschaften von F, wie die Lojasiewicz-Eigenschaft, und die Zeitskalierung der Differentialgleichung die Konvergenzraten von $F(x(t))$ zu F^* beeinflussen und verbessern. Diese wurden bisher separat untersucht, während in dieser Arbeit beide kombiniert werden, um die bekannten Schranken für $F(x(t)) - F^*$ zu verfeinern. Um die verbesserten Konvergenzraten zu veranschaulichen, werden die Ergebnisse numerischer Implementierungen vorgestellt.

Abstract

In this master thesis we study the trajectories of two types of second-order ODEs linked to the accelerated gradient method of Nesterov, including constant damping and asymptotically vanishing viscous damping. These ODEs are governed by a convex differentiable function F and it was shown that their trajectories x have minimizing properties, that means $F(x(t))$ converges to the minimal value F^* of F as t goes to infinity. In previous works it was studied, how certain geometrical properties of F, such as the Lojasiewicz property, and time scaling of the ODEs effect and improve the convergence rates of $F(x(t))$ to F^*. These were studied separately, whereas in this thesis both of them will be combined to refine the known bounds on $F(x(t)) - F^*$. To visualize the improved convergence rates the results of numerical implementations will be presented.

Contents

1 Introduction

We will start by clarifying the general mathematical setting of this thesis.

Throughout this thesis we will make the following preliminary assumptions unless specified:

the underlying space is $\mathbb{R}^n$,

$F : \mathcal{H} \to \mathbb{R}$ is a convex differentiable function with $\operatorname{argmin} F \neq \emptyset$,

$\beta : [t_0, +\infty) \to \mathbb{R}^+$ is a differentiable function with $t_0 > 0$

and $\alpha \in \mathbb{R}^+$ is a positive constant.

The minimal value of F will be denoted by F^*. We will study the convergence of the trajectories of two non-perturbed second-order ordinary differential equations as $t \to +\infty$, where we add certain geometrical assumptions on F. The first differential equation is

$$\ddot{x}(t) + \alpha\dot{x}(t) + \beta(t)\nabla F(x(t)) = 0, \tag{CD}$$

for $t \geq t_0$, where α is called the constant viscous damping coefficient and β is referred to as the time scale parameter. We define the second one as

$$\ddot{x}(t) + \frac{\alpha}{t}\dot{x}(t) + \beta(t)\nabla F(x(t)) = 0, \tag{AVD}$$

for $t \geq t_0$. Since $\frac{\alpha}{t}$ vanishes asymptotically as $t \to +\infty$ we refer to this system as the asymptotic vanishing damping case and to the first one as the constant damping case, hence the choice of the abbreviations AVD and CD respectively. Throughout the chapters $x \in \mathcal{C}^2(\mathbb{R}, \mathbb{R}^n)$ will denote the solution to the associated Cauchy problem to either (AVD) or (CD) for unspecified initial conditions $(x_0, v_0) \in \mathbb{R}^n \times \mathbb{R}^n$. The existence and uniqueness of such a solution will be assumed, but one can see, that this can be guaranteed under various circumstances, for example if ∇F is locally Lipschitz as seen in [1].

This thesis is based on the paper [2], where, under geometrical conditions on F around its set of minimizers, improved convergence rates of $F(x(t))$ to F^* were found for the system (AVD) and (CD) with no time scaling, namely $\beta \equiv 1$. The paper [3] analysed among other things the positive effects of time scaling of (AVD) on the convergence rate of $F(x(t))$ for general convex differentiable functions F. Naturally the question arises, if the rates presented in [2] can be refined by the introduction of time scaling, which will be successfully answered as the main goal of this thesis. In the end the improvements will be underscored by numerical results of implementations in `Python`.

2 The geometrical conditions

In this section we will introduce the two geometrical conditions from [2] and [4] which on the one hand ensure that F is not to sharp in the neighbourhood of any of its minimizers and on the other hand require a certain magnitude of the gradient ∇F around the minimizers of F.

In this paper $\langle \cdot, \cdot \rangle$ will be the scalar product of $\mathbb{R}^n$ and $\|\cdot\|$ the induced norm. As customary an open ball with center $x \in \mathbb{R}^n$ and radius $r > 0$ will be denoted by $B(x, r)$. Furthermore for any set $S \subset \mathbb{R}^n$ the distance of a point $x \in \mathbb{R}^n$ to the set S is defined by $d(x, X) = \inf_{y \in \mathbb{R}^n} \|x - y\|$.

Definition 2.1. Let $F : \mathbb{R}^n \to \mathbb{R}$ be a convex differentiable function with $\operatorname{argmin} F \neq \emptyset$ and $F^* = \inf F$. We introduce the following geometrical conditions:

(i) Let $\gamma \geq 1$. The function F satisfies the condition $(H_1(\gamma))$ if, for any minimizer $x^* \in \operatorname{argmin} F$, there exists $\eta > 0$ such that

$$\forall x \in B(x^*, \eta), \quad F(x) - F^* \leq \frac{1}{\gamma} \langle \nabla F(x), x - x^* \rangle . \tag{$H_1(\gamma)$}$$

(ii) Let $p \geq 1$. The function F satisfies the growth condition $(H_2(p))$ if for any minimizer $x^* \in \operatorname{argmin} F$, there exist $K_p > 0$ and $\varepsilon > 0$ such that

$$\forall x \in B(x^*, \varepsilon), \quad K_p \cdot d(x, \operatorname{argmin} F)^p \leq F(x) - F^* \tag{$H_2(p)$}$$

To get a better understanding on the implications of these conditions, we will look into a few examples as seen in [4].

For example any convex differentiable function F satisfies $H_1(\gamma)$ with $\gamma = 1$, since

$$F(x) - F^* \leq \langle \nabla F(x), x - x^* \rangle$$

holds for all $x \in \mathbb{R}^n$. Furthermore any differentiable function for which $(F - F^*)^{\frac{1}{\gamma}}$ is convex for $\gamma \geq 1$ satisfies $H_1(\gamma)$.

The condition $H_1(\gamma)$ can also be interpreted in a geometrical sense. It implies a certain type of flatness of F around its set of minimizers, as proven in the following Lemma.

Lemma 2.2. *In the same setting as in Definition 2.1 we can state that if F satisfies $H_1(\gamma)$ for some $\gamma \geq 1$, then the following properties hold:*

(i) F satisfies $H_1(\gamma')$ for all $\gamma' \in [1, \gamma]$.

(ii) For any minimizer $x^* \in \operatorname{argmin} F$, there exist $M > 0$ and $\eta > 0$ such that

$$\forall x \in B(x^*, \eta), \quad F(x) - F^* \leq M \|x - x^*\|^{\gamma} .$$

Proof. For (i) we just need to recall that $\frac{1}{\gamma} \leq \frac{1}{\gamma'}$ if $\gamma \geq \gamma'$ and that any convex differentiable function satisfies $H_1(1)$.

For (ii) we need to prove a simple result for a convex and differentiable function $g : \mathbb{R} \to \mathbb{R}$ with 0 as one of its minimizers with $g(0) = 0$. If for all $t \in [0, 1]$

$$g(t) \leq \frac{t}{\gamma} g'(t)$$

holds, then $t \mapsto t^{-\gamma} g(t)$ is monotonically increasing on $[0, 1]$, since its derivative

$$(-\gamma)t^{-\gamma-1}g(t) + t^{-\gamma}g'(t) \geq -t^{-\gamma}g'(t) + t^{-\gamma}g'(t) = 0$$

for all $t \in [0, 1]$. Therefore

$$\forall t \in [0, 1], \quad g(t)t^{-\gamma} \leq g(1) \tag{2.1}$$

holds. Now recall that $H_1(\gamma)$ means that for F there exists a $\eta > 0$ such that

$$\forall x \in B(x^*, \eta), \quad F(x) - F^* \leq \frac{1}{\gamma} \langle \nabla F(x), x - x^* \rangle .$$

Let $\eta' \in (0, \eta)$ and $x \in \overline{B}(x^*, \eta')$ with $x \neq x^*$. We define the following function on $[0, 1]$

$$g_x(t) := F\left(x^* + t\eta' \frac{x - x^*}{\|x - x^*\|}\right) - F^*,$$

which is convex and differentiable by construction with minimizer 0 and $g_x(0) = 0$. There exists an $M > 0$ s.t. for all $x \in \overline{B}(x^*, \eta')$ with $x \neq x^*$ and all $t \in [0, 1]$

$$g_x(t) \leq M$$

holds, which is implied by the continuity of F and since $x^* + t\eta' \frac{x-x^*}{\|x-x^*\|}$ stays in the compact set $\overline{B}(x^*, \eta')$. This constant only depends on x^* and η'. Using property $H_1(\gamma)$, we see that

$$\begin{aligned} g_x'(t) &= \left\langle \nabla F\left(x^* + t\eta' \frac{x - x^*}{\|x - x^*\|}\right), \eta' \frac{x - x^*}{\|x - x^*\|} \right\rangle \\ &\geq \frac{\gamma}{t}\left(F\left(x^* + t\eta' \frac{x - x^*}{\|x - x^*\|}\right) - F^*\right) = \frac{\gamma}{t} g_x(t) \end{aligned}$$

holds for any $t \in [0, 1]$. The result (2.1) and the uniform bound M yield

$$\forall x \neq x^* \in \overline{B}(x^*, \eta'), \quad g_x(t) \leq g_x(1)t^\gamma \leq Mt^\gamma$$

for all $t \in [0, 1]$. Choosing $t = \frac{1}{\eta'} \|x - x^*\|$, we obtain the desired result. □

This lemma states that F is at least as flat as $\|\cdot\|^\gamma$ in a neighbourhood of any of its minimizers.

Conversly the condition $H_2(p)$ with $p \geq 1$ ensures that F is sufficiently sharp in the neighborhood of any of its minimizers $x^* \in \operatorname{argmin} F$. It is closely tied to Lojasiewicz inequality, which is used in the mathematical analysis of dynamical systems.

Definition 2.3. A differentiable function $F : \mathbb{R}^n \to \mathbb{R}$ is said to have the Lojasiewicz property with exponent $\theta \in [0, 1)$, if for any critical point x^*, there exist $c > 0$ and $\varepsilon > 0$ such that

$$\forall x \in B(x^*, \varepsilon), \quad \|\nabla F(x)\| \geq c(F(x) - F(x^*))^\theta, \tag{2.2}$$

where $0^0 = 0$ when $\theta = 0$ by convention.

In this thesis we are focused on functions F with a unique minimizer. In fact one can prove, that in this convex setting $H_2(p)$ is equivalent to the Lojasiewicz inequality with exponent $\theta = 1 - \frac{1}{p}$ for functions F with unique minimizer x^* as seen in [2].

Lemma 2.4. *In the same setting as in Definition 2.1 we can state, that functions F with unique minimizers satisfy the condition $H_2(p)$ for some $p \geq 1$, if and only if they satisfy Lojasiewicz inequality* (2.2) *with exponent $\theta = 1 - \frac{1}{p} \in [0, 1)$.*

Proof. We denote x^* as the unique minimizer of F $\operatorname{argmin} F = \{x^*\}$, therefore $d(x, \operatorname{argmin} F) = \|x - x^*\|$. Assume that $H_2(p)$ is satisfied. Then there exist $K_p > 0$ and $\varepsilon > 0$ such that

$$\forall x \in B(x^*, \varepsilon), \quad K_p \cdot d(x, \operatorname{argmin} F)^p = K_p \|x - x^*\| \leq F(x) - F^*.$$

Since F is convex, it holds for $x \in B(x^*, \varepsilon)$

$$\begin{aligned} F(x) - F^* \leq \langle \nabla F(x), x - x^* \rangle &\leq \|\nabla F(x)\| \cdot \|x - x^*\| \\ &\leq \|\nabla F(x)\| \cdot \left(\frac{1}{K^p} (F(x) - F^*) \right)^{\frac{1}{p}}. \end{aligned}$$

Therefore for all $x \in B(x^*, \varepsilon)$ with $x \neq x^*$ and $c = K_p^{\frac{1}{p}} > 0$

$$c\,(F(x) - F^*)^{1-\frac{1}{p}} \leq \|\nabla F(x)\|.$$

If we assume that (2.2) holds for $\theta \in [0, 1)$, then the following holds for all $x \in B(x^*, \varepsilon)$:

$$\frac{1}{c} \|\nabla F(x)\| \geq (F(x) - F^*)^\theta. \tag{2.3}$$

Choose K large enough such that for $x \in \overline{B}(x^*, \varepsilon')$ with $x \neq x^*$ and $\varepsilon > \varepsilon' > 0$

$$\frac{1}{c} \|\nabla F(x)\| \leq K \cdot \|x - x^*\|^{-1} \cdot (F(x) - F^*).$$

Therefore we can further estimate (2.3)

$$\begin{aligned} (F(x) - F^*)^\theta \leq \frac{1}{c} \|\nabla F(x)\| &\leq K \|x - x^*\|^{-1} (F(x) - F^*), \\ &\Rightarrow (F(x) - F^*)^{\theta - 1} \leq K \|x - x^*\|^{-1}, \\ &\Rightarrow (F(x) - F^*)^{1-\theta} \geq \frac{1}{K} \|x - x^*\|, \\ &\Rightarrow F(x) - F^* \geq \left(\frac{1}{K} \|x - x^*\| \right)^{\frac{1}{1-\theta}}. \end{aligned}$$

This proves $H_2(p)$ for $p := \frac{1}{1-\theta}$ and $K_p := \frac{1}{K^p}$.
For $x = x^*$ both $H_2(p)$ and (2.2) are satisfied, which concludes the proof of the equivalence. $\square$

Strong convexity is another example for a sufficient condition that implies the property $H_2(p)$ with $p = 2$. For any strongly convex function F with modulus $m > 0$

$$F(x) \geq F(y) + \langle \nabla F(y), x - y \rangle + \frac{m}{2} \|x - y\|^2$$

holds. By setting $y = x^*$ we get

$$F(x) - F^* \geq \frac{m}{2} \|x - x^*\|^2 \geq \frac{m}{2} d(x, X^*)^2,$$

therefore $H_2(2)$ holds.
In the main results of this master thesis we will look at convex differential functions that satisfy both $H_1(\gamma)$ and $H_2(p)$. If for any function F both conditions hold, then this would visually mean that F has to be at least as flat as $\|x - x^*\|^\gamma$ and as sharp as $\|x - x^*\|^p$ in a neighbourhood of its minimizer x^*. This intuitively implies the following result.

Lemma 2.5. *If a convex differentiable function F with unique minimizer x^* satisfies both $H_1(\gamma)$ and $H_2(p)$ with $\gamma, p \geq 1$, then necessarily $p \geq \gamma$.*

Proof. This can easily be seen as $H_2(p)$ implies for $x \in B(x^*, \varepsilon)$

$$K_p \|x - x^*\|^p \leq F(x) - F^*$$

and Lemma 2.2 implies for $x \in B(x^*, \eta)$

$$F(x) - F^* \leq M \cdot \|x - x^*\|^\gamma .$$

$\square$

Compared to a general convex differentiable function F, both of the conditions $H_1(\gamma)$ and $H_2(p)$ on F combined allow us to prove better convergence rates for F to its minimal value along the trajectories $x(\cdot)$, which are the solutions of either differential equation (CD) or (AVD). These conditions imply estimations which only hold in a certain neighbourhood of a minimizer. Therefore, if we want to apply these in our proofs, we first need to establish that $x(t)$ converges to a minimizer of F as t goes to infinity.

3 Convergence with time scaling

In this chapter we discuss the convergence of $F(x(t))-F^*$ as $t \to +\infty$, where $x(\cdot)$ is a solution of (CD) or (AVD). This will be achieved by the use of Lyapunov functions for each system and proving that they are non-increasing or even decreasing. This will also be the fundamental proof strategy throughout this thesis.
As in [2] we define the functions $a, b_1, b_2, c : [t_0, +\infty) \to [0, +\infty)$

$$\begin{aligned} a(t) &:= F(x(t)) - F^*, \\ b_1(t) &:= \frac{1}{2} \left\| \lambda(x(t) - x^*) + \dot{x}(t) \right\|^2 \text{ with } \lambda \in [0, +\infty), \\ b_2(t) &:= \frac{1}{2} \left\| \lambda(x(t) - x^*) + t\dot{x}(t) \right\|^2 \text{ with } \lambda \in [0, +\infty), \\ \text{and } c(t) &:= \frac{1}{2} \left\| x(t) - x^* \right\|^2, \end{aligned}$$

of which the energy functions will consist.

Definition 3.1. For a solution $x : [t_0, +\infty) \to \mathbb{R}^n$ of (CD) we define the energy function $\mathcal{E}_{(\mathrm{CD}),\xi} : [t_0, +\infty) \to \mathbb{R}$ as

$$\mathcal{E}_{(\mathrm{CD}),\xi} := \beta(t)a(t) + b_1(t) + \xi c(t), \tag{3.1}$$

where $\xi \in \mathbb{R}$ is a constant.

This definition of $\mathcal{E}_{(\mathrm{CD}),\xi}$ is inspired by the energy functions in [3] and adapted for the proofs in this paper. For the differential equation (AVD) the energy function can be chosen as in [3].

Definition 3.2. For a solution $x : [t_0, +\infty) \to \mathbb{R}^n$ of (AVD) we define the energy function $\mathcal{E}_{(\mathrm{AVD}),\xi} : [t_0, +\infty) \to \mathbb{R}$ as

$$\mathcal{E}_{(\mathrm{AVD}),\xi} := t^2\beta(t)a(t) + b_2(t) + \xi c(t), \tag{3.2}$$

where $\xi \in \mathbb{R}$ is a constant.

The choice of the constants λ and ξ will vary in the different proofs and will be chosen and limited accordingly to achieve the desired results.

Since looking at the derivatives of both $\mathcal{E}_{(\mathrm{CD}),\xi}$ and $\mathcal{E}_{(\mathrm{AVD}),\xi}$ will be crucial in proving all of the theorems in this paper, we will calculate and simplify both $\mathcal{E}'_{(\mathrm{CD}),\xi}$ and $\mathcal{E}'_{(\mathrm{AVD}),\xi}$ as done in [2] and [3].

Lemma 3.3. *For any solution $x : [t_0, +\infty) \to \mathbb{R}^n$ of (CD) the following equality holds for any $t \geq t_0$, $\lambda \in [0, +\infty)$ and $\xi \in \mathbb{R}$*

$$\begin{aligned} \mathcal{E}'_{(CD),\xi}(t) = \beta'(t)a(t) + 2(\lambda-\alpha)\left(b_1(t) - \lambda^2 c(t)\right) - \lambda\beta(t)\left\langle \nabla F(x(t)), x(t) - x^* \right\rangle \\ + (\xi - \lambda(\lambda - \alpha))\left\langle \dot{x}(t), x(t) - x^* \right\rangle. \end{aligned}$$

Proof. All statements throughout this proof will hold for all $t \geq t_0$. We will calculate the derivative of $\mathcal{E}_{(\mathrm{CD}),\xi}$ by looking at the derivatives of the summands:

$$\frac{d}{dt}(\beta(t)a(t)) = \beta'(t)(F(x(t)) - F^*) + \beta(t)\left\langle \nabla F(x(t)), \dot{x}(t) \right\rangle, \tag{3.3}$$

$$\frac{d}{dt}b_1(t) = \left\langle \lambda\dot{x}(t) + \ddot{x}(t), \lambda(x(t) - x^*) + \dot{x}(t) \right\rangle,$$

$$\text{and } \frac{d}{dt}c(t) = \left\langle \dot{x}(t), x(t) - x^* \right\rangle. \tag{3.4}$$

Since $x(\cdot)$ is a solution to our differential equation (CD), we know that

$$\lambda\dot{x} + \ddot{x}(t) = (\lambda - \alpha)\dot{x}(t) - \beta(t)\nabla F(x(t))$$

holds. This yields

$$\frac{d}{dt}b_1(t) = (\lambda - \alpha)\left\langle \dot{x}(t), \lambda(x(t) - x^*) + \dot{x}(t) \right\rangle + \beta(t)\left\langle \nabla F(x(t)), -\lambda(x(t) - x^*) - \dot{x}(t) \right\rangle.$$

Furthermore we can rewrite

$$\left\langle \dot{x}(t), \lambda(x(t) - x^*) + \dot{x}(t) \right\rangle = \left\langle \dot{x}(t), \lambda(x(t) - x^*) \right\rangle + \|\dot{x}(t)\|^2.$$

Therefore we can write the derivative as

$$\mathcal{E}'_{(\mathrm{CD}),\xi}(t) = \beta'(t)(F(x(t)) - F^*) - \beta(t)\left\langle \nabla F(x(t)), \lambda(x(t) - x^*) \right\rangle + (\xi + \lambda(\lambda - \alpha))\left\langle \dot{x}(t), x(t) - x^* \right\rangle + (\lambda - \alpha)\|\dot{x}(t)\|^2.$$

Now notice that

$$2b_1(t) = \|\lambda(x(t) - x^*) + \dot{x}(t)\|^2 = 2\lambda^2 c(t) + 2\lambda\left\langle x(t) - x^*, \dot{x}(t) \right\rangle + \|\dot{x}(t)\|^2,$$

which yields

$$\begin{aligned}\mathcal{E}'_{(\mathrm{CD}),\xi}(t) = \beta'(t)(F(x(t)) - F^*) - \beta(t)\left\langle \nabla F(x(t)), \lambda(x(t) - x^*) \right\rangle \\ +(\xi - \lambda(\lambda - \alpha))\left\langle \dot{x}(t), x(t) - x^* \right\rangle \\ -2\lambda^2(\lambda - \alpha)c(t) + 2(\lambda - \alpha)b_1(t)\end{aligned}$$

or, equivalently,

$$\mathcal{E}'_{(\mathrm{CD}),\xi}(t) = \beta'(t)a(t) + 2(\lambda - \alpha)\left(b_1(t) - \lambda^2 c(t)\right) - \lambda\beta(t)\left\langle \nabla F(x(t)), x(t) - x^* \right\rangle + (\xi - \lambda(\lambda - \alpha))\left\langle \dot{x}(t), x(t) - x^* \right\rangle.$$

□

Lemma 3.4. *For any solution $x : [t_0, +\infty) \to \mathbb{R}^n$ of* (AVD) *the following equality holds for any $t \geq t_0$, $\lambda \in [0, +\infty)$ and $\xi \in \mathbb{R}$*

$$\mathcal{E}'_{(AVD),\xi} = (2t\beta(t) + t^2\beta'(t))a(t) + \frac{2(\lambda - \alpha + 1)}{t}\left(b_2(t) - \lambda^2 c(t)\right) - t\lambda\beta(t)\left\langle \nabla F(x(t)), x(t) - x^* \right\rangle + (\xi - \lambda(\lambda - \alpha + 1))\left\langle \dot{x}(t), x(t) - x^* \right\rangle.$$

Proof. This proof is structured as in the lemma before and we take $t \geq t_0$. We only need to calculate the derivatives

$$\frac{d}{dt}(t^2\beta(t)a(t)) = (2t\beta(t) + t^2\beta'(t))a(t) + t^2\beta(t)\langle\nabla F(x(t)), \dot{x}(t)\rangle \tag{3.5}$$
$$\frac{d}{dt}b_2(t) = \langle\lambda\dot{x}(t) + t\ddot{x}(t) + \dot{x}(t), \lambda(x(t) - x^*) + t\dot{x}(t)\rangle$$

and the derivative of c can be taken as in (3.4). Since $x(\cdot)$ is here a solution of (AVD), we see that

$$t\ddot{x}(t) = -\alpha\dot{x}(t) - t\beta(t)\nabla F(x(t)).$$

Hence

$$\begin{aligned}\frac{d}{dt}b_2(t) = t(\lambda - \alpha + 1)\|\dot{x}(t)\|^2 + \lambda(\lambda - \alpha + 1)\langle\dot{x}(t), x(t) - x^*\rangle \\ - t^2\beta(t)\langle\nabla F(x(t)), \dot{x}(t)\rangle - t\lambda\beta(t)\langle\nabla F(x(t)), x(t) - x^*\rangle.\end{aligned} \tag{3.6}$$

We notice that

$$\begin{aligned}t\|\dot{x}(t)\|^2 + \lambda\langle\dot{x}(t), x(t) - x^*\rangle &= \\ &= \frac{1}{t}\left(\|t\dot{x}(t)\|^2 + \langle t\dot{x}(t), \lambda(x(t) - x^*)\rangle\right) \\ &= \frac{1}{t}\left(\|\lambda(x(t) - x^*) + t\dot{x}(t)\|^2 - \langle t\dot{x}(t), \lambda(x(t) - x^*)\rangle - \lambda^2\|x(t) - x^*\|^2\right) \\ &= \frac{1}{t}\left(2b_2(t) - \langle t\dot{x}(t), \lambda(x(t) - x^*)\rangle - 2\lambda^2 c(t)\right).\end{aligned}$$

Therefore

$$\begin{aligned}\frac{d}{dt}b_2(t) = \frac{\lambda - \alpha + 1}{t}\left(2b_2(t) - \langle t\dot{x}(t), \lambda(x(t) - x^*)\rangle - 2\lambda^2 c(t)\right) \\ - t^2\beta(t)\langle\nabla F(x(t)), \dot{x}(t)\rangle - t\lambda\beta(t)\langle\nabla F(x(t)), x(t) - x^*\rangle.\end{aligned} \tag{3.7}$$

By adding up (3.4), (3.5) and (3.7) we conclude

$$\begin{aligned}\mathcal{E}'_{(\mathrm{AVD}),\xi} = (2t\beta(t) + t^2\beta'(t))a(t) + \frac{2(\lambda - \alpha + 1)}{t}\left(b_2(t) - \lambda^2 c(t)\right) \\ - t\lambda\beta(t)\langle\nabla F(x(t)), x(t) - x^*\rangle + (\xi - \lambda(\lambda - \alpha + 1)\langle\dot{x}(t), x(t) - x^*\rangle.\end{aligned}$$

□

As part of the proof strategies, a variation of the Grönwall's inequality as stated in [5] will play an important role in converting estimations the derivative of $\mathcal{E}'_{(\mathrm{CD}),\xi}$ into estimations of $\mathcal{E}_{(\mathrm{CD}),\xi}$.

Lemma 3.5 (Grönwall's Inequality)**.** *Let $u : [a, +\infty) \to \mathbb{R}$ be a differentiable function. If*

$$u'(t) \leq g(t)u(t)$$

holds for all $t \in [a, +\infty)$ and a continuous function $g : [a, +\infty) \to \mathbb{R}$, then

$$u(t) \leq u(a)e^{\int_a^b g(t)dt}$$

for all $t \in [a, +\infty)$.

Simple integration proves this lemma.

Now we will look at the convergence of the trajectories of the two systems (CD) and (AVD) separately.

3.1 Constant damping

The following proof is an adapted version of the proof presented in [6], where we added the time scaling parameter β.

Theorem 3.6. *Let $\alpha \geq 0$, $t_0 > 0$ and $x : [t_0, \infty) \to \mathbb{R}^n$ be a solution of* (CD). *If for all $t \geq t_0$*

$$\beta'(t) \leq \frac{\alpha}{2}\beta(t) \tag{3.8}$$

holds, then $F(x(t)) - F^* = \mathcal{O}\left(\frac{e^{-\frac{\alpha}{2}t}}{\beta(t)}\right)$ as $t \to +\infty$.

Proof. We choose $\lambda = \alpha$ and $\xi = 0$ reduces the result of Lemma 3.3 for all $t \geq t_0$ to

$$\mathcal{E}'_{(\mathrm{CD}),0}(t) = \beta'(t)a(t) - \alpha\beta(t)\left\langle \nabla F(x(t)), x(t) - x^* \right\rangle.$$

By the definition of convexity $a(t) = F(x) - F^* \leq \langle \nabla F(x), x - x^* \rangle$ holds for all $x \in \mathbb{R}^n$ and we can deduce

$$\mathcal{E}'_{(\mathrm{CD}),0}(t) \leq (\beta'(t) - \alpha\beta(t))a(t) \quad \forall t \geq t_0.$$

The assumption (3.8) on β implies $\beta'(t) - \alpha\beta(t) \leq -\frac{\alpha}{2}\beta(t)$ and combined with $\mathcal{E}_{(\mathrm{CD}),0} = \beta(t)a(t) + b_2(t) \geq \beta(t)a(t)$, we get

$$\mathcal{E}'_{(\mathrm{CD}),0}(t) \leq -\frac{\alpha}{2}\mathcal{E}_{(\mathrm{CD}),0}(t)$$

for all $t \geq t_0$. Grönwall's inequality yields $\mathcal{E}_{(\mathrm{CD}),0}(t) = \mathcal{O}(e^{-\frac{\alpha}{2}t})$. Therefore by Definition 3.1 we conclude $F(x(t)) - F^* = a(t) \leq \frac{\mathcal{E}_{(\mathrm{CD}),0}(t)}{\beta(t)}$. □

3.2 Asymptotically vanishing damping

The proof of convergence for the system with asymptotically vanishing damping is structured as in [3], where we leave out the Hessian damping parameter.

Theorem 3.7. *Let $\alpha \geq 1$, $t_0 > 0$ and $x : [t_0, \infty) \to \mathbb{R}^n$ be a solution of* (AVD). *If for all $t \geq t_0$*

$$t\beta'(t) \leq (\alpha - 3)\beta(t) \tag{3.9}$$

holds, then $F(x(t)) - F^* = \mathcal{O}\left(\frac{1}{t^2\beta(t)}\right)$ as $t \to +\infty$.

Proof. We choose $\lambda = \alpha - 1$, $\xi = 0$ and $t \geq t_0$. By Lemma 3.4 we obtain for the derivative of the energy function the following:

$$\mathcal{E}'_{(\mathrm{AVD}),0}(t) \;=\; (2t\beta(t) + t^2\beta'(t))a(t) - t(\alpha - 1)\beta(t)\left\langle \nabla F(x(t)), x(t) - x^* \right\rangle . \tag{3.10}$$

Again by convexity $a(t) \geq \langle \nabla F(x), x^* - x(t) \rangle$ holds. By using this property (3.10) becomes

$$\mathcal{E}'_{(\mathrm{AVD}),0}(t) \leq ((3-\alpha)t\beta(t) + t^2\beta'(t))a(t).$$

For $\beta(\cdot)$ the inequality (3.9) holds, hence $\mathcal{E}'_{(\mathrm{AVD}),0}(t) \leq 0$. Consequently $\mathcal{E}_{(\mathrm{AVD}),0}$ is non-increasing and $\mathcal{E}_{(\mathrm{AVD}),0}(t) \leq \mathcal{E}_{(\mathrm{AVD}),0}(t_0)$ for all $t \geq t_0$. Therefore

$$\begin{aligned} t^2\beta(t)(F(x(t)) - F^*) + \frac{1}{2}\left\|(\alpha-1)(x(t)-x^*) + t\dot{x}(t)\right\|^2 &\leq \mathcal{E}_{(\mathrm{AVD}),0}(t_0), \\ t^2\beta(t)(F(x(t)) - F^*) &\leq \mathcal{E}_{(\mathrm{AVD}),0}(t_0), \end{aligned}$$

which concludes that $F(x(t)) - F^* = \mathcal{O}\left(\frac{1}{t^2\beta(t)}\right)$.

□

We have established that $F(x(t)) \to F^*$ and we will see how these initial convergence rates can be refined by adding the geometrical conditions on F. This will be done in the following two chapters, where we first leave out time scaling, in other words choose $\beta \equiv 1$.

4 Convergence under geometrical conditions

In this chapter we will prove the main results of [2], where time scaling is not considered.

First of all we choose $\beta \equiv 1$. We will look at functions that satisfy $H_1(\gamma)$ with $\gamma \in [1, 2]$ and $H_2(p)$ with $p = 2$ and assume that F has a unique minimizer which we denote by x^*. The latter has certain practical implications. The theorems in the previous chapter imply that $F(x(t))$ converges to F^* as $t \to +\infty$ for solutions of either (CD) or (AVD), because $\beta \equiv 1$ fulfills the required differential inequalities (3.8) and (3.9). Since F is convex and the minimizer is unique, we can deduce that $x(t)$ converges to x^* as $t \to +\infty$. This means that t can be chosen large enough for $x(t)$ to be in any neighbourhood of x^*, so that the definitions of $H_1(\gamma)$ and $H_2(2)$ apply. Lastly we notice that $d(x, \operatorname{argmin} F)$ reduces to $\|x - x^*\|$. These observations prove the following lemma:

Lemma 4.1. *Let $x : [t_0, +\infty) \to \mathbb{R}^n$ be a solution of either* (CD) *or* (AVD) *and $\beta \equiv 1$. If F admits a unique minimizer x^*, then $x(t)$ converges to x^* as $t \to +\infty$. Furthermore if there exists a $\gamma \leq 2$ such that F fulfills the conditions $H_1(\gamma)$ and $H_2(2)$, then there exists a $t_1 \geq t_0$ such that*

$$F(x(t)) - F^* \leq \frac{1}{\gamma} \langle \nabla F(x(t)), x(t) - x^* \rangle$$
$$K_2 \cdot \|x(t) - x^*\|^2 \leq F(x(t)) - F^*$$

hold for all $t \geq t_1$.

This lemma makes it possible to use the geometrical conditions on F for estimating. We can even rewrite the inequalities of Lemma 4.1 with the definitions of a and c as follows:

$$\gamma \cdot a(t) \leq \langle \nabla F(x(t)), x(t) - x^* \rangle \tag{4.1}$$
$$2 \cdot K_2 \cdot c(t) \leq a(t) \tag{4.2}$$

for $t \geq t_1$. These estimates play a central role in proving the convergence rates of $F(x(t)) - F^*$ in the following proofs in this thesis. First we look at the trajectories of the system (CD) and derive the convergence rate as done in [2].

4.1 Constant damping

Theorem 4.2. *Let $\gamma \in [1, 2]$ and let $x : [t_0, +\infty) \to \mathbb{R}^n$ be a solution of* (CD) *with $\beta \equiv 1$. If F satisfies $H_1(\gamma)$ and $H_2(2)$ with a real constant $K_2 > 0$ and admits a unique minimizer x^*, then for any*

$$m \in \left(0, \min\left(1, \frac{2\alpha^2}{\gamma(\gamma+2)K_2}\right)\right)$$

we have

$$F(x(t)) - F^* = \mathcal{O}\left(e^{-m(1-m)\frac{\gamma^2 K_2}{\alpha} t}\right)$$

as $t \to +\infty$.

Proof. Under our assumptions on F and β, Lemma 4.1 implies the existence of a $t_1 \geq t_0$ such that (4.1) and (4.2) hold for all $t \geq t_1$. In this proof we take $t \geq t_1$.

We restrict $\lambda \leq \alpha$ for now and set $\xi = \lambda(\lambda - \alpha)$, where we notice that $\xi \leq 0$ holds. With the result of Lemma 3.3 and the facts that $\beta \equiv 1$ and $\beta' \equiv 0$ we get

$$\mathcal{E}'_{(\mathrm{CD}),\xi}(t) = 2(\lambda - \alpha)b_1(t) - 2\xi\lambda c(t) - \lambda \left\langle \nabla F(x(t)), x(t) - x^* \right\rangle.$$

By (4.1) we see that

$$\mathcal{E}'_{(\mathrm{CD}),\xi}(t) \leq 2(\lambda - \alpha)b_1(t) - 2\xi\lambda c(t) - \lambda\gamma a(t).$$

Furthermore applying (4.2) yields

$$\mathcal{E}'_{(\mathrm{CD}),\xi}(t) \leq 2(\lambda - \alpha)b_1(t) - \lambda\gamma \left(1 + \frac{\xi\lambda}{K_2\gamma}\right) a(t), \tag{4.3}$$

since ξ is negative. From now on we choose $\lambda = m\frac{\gamma K_2}{\alpha}$, where $m \in (0,1)$ has to be chosen small enough for

$$1 + \frac{\xi}{K_2\gamma} > 1 - m > 0, \tag{4.4}$$

$$\text{and } 2(\lambda - \alpha) \leq -\lambda\gamma \tag{4.5}$$

to hold. The second inequality is equivalent to $\lambda \leq \frac{2}{\gamma+2}\alpha$, which ensures $\lambda < \alpha$, and furthermore implies that m has to satisfy

$$0 < m < \min\left(1, \frac{2\alpha^2}{(\gamma+2)\gamma K_2}\right).$$

Inserting (4.4) and (4.5) in (4.3) we get

$$\begin{aligned}\mathcal{E}'_{(\mathrm{CD}),\xi}(t) &\leq -\lambda\gamma b_1(t) - \lambda\gamma\,(1-m)\,a(t), \\ &= -\lambda\gamma(1-m)\left(a(t) + \frac{1}{1-m}b_1(t)\right).\end{aligned}$$

Since ξ is negative we know that $a(t) + \frac{1}{1-m}b_1(t) \geq a(t) + b_1(t) \geq \mathcal{E}_{(\mathrm{CD}),\xi}(t)$ holds. Hence we obtain

$$\mathcal{E}'_{(\mathrm{CD}),\xi}(t) \leq -(1-m)\lambda\gamma\mathcal{E}_{(\mathrm{CD}),\xi}(t).$$

By Grönwall's inequality we conclude

$$\mathcal{E}_{(\mathrm{CD}),\xi}(t) = \mathcal{O}\left(e^{-(1-m)\lambda\gamma t}\right).$$

Since by the definition of m and (4.2)

$$\mathcal{E}_{(\mathrm{CD}),\xi}(t) = a(t) + b_1(t) + c(t) \geq (1-m)a(t) + b_1(t) \geq (1-m)a(t)$$

holds, $F(x(t)) - F^* = a(t) \leq \frac{1}{1-m}\mathcal{E}_{(\mathrm{CD}),\xi}(t)$ yields the desired result. $\square$

The question arises for which $m \in \left(0, \min\left(1, \frac{2\alpha^2}{\gamma(\gamma+2)K_2}\right)\right)$ the presented convergence rate is optimal.

Remark 4.3. Consider the case, where

$$\frac{1}{2} < \frac{2\alpha^2}{\gamma(\gamma+2)K_2} \tag{4.6}$$

holds. This can occur if K_2 is rather small or α is large. Then the optimal rate is achieved for $m = \frac{1}{2}$, because then $m(1-m) = \frac{1}{4}$ is maximal at this point. Furthermore (4.6) implies

$$\frac{1}{2} < \frac{2\alpha^2}{\gamma^2 K_2} \Rightarrow \frac{\gamma^2 K_2}{4\alpha} < \alpha.$$

Therefore we can notice that in this case the lower bound of the presented convergence rate is

$$e^{-m(1-m)\frac{\gamma^2 K_2}{\alpha}t} > e^{-\alpha t}.$$

For $\frac{1}{2} > \frac{2\alpha^2}{\gamma(\gamma+2)K_2} > m$ one can notice that $m(1-m)\frac{\gamma^2 K_2}{\alpha}t$ gets larger if m increases. This means that m has to be chosen in the interval as large as possible to achieve the best convergence rate.

4.2 Asymptotically vanishing damping

In the case with vanishing damping we need a slightly different ansatz to prove the convergence rate. As presented in [2] we will show that $F(x(t)) - F^*$ is bounded by a polynomial multiplied with a function $\mathcal{G} : [t_0, +\infty) \to \mathbb{R}$, which will turn out to be bounded.

Theorem 4.4. *Let $\gamma \in [1, 2]$ and let $x : [t_0, +\infty) \to \mathbb{R}^n$ be a solution of* (AVD) *with $\beta \equiv 1$. If F satisfies $H_1(\gamma)$ and $H_2(2)$ with a real constant $K_2 > 0$ and admits a unique minimizer x^*, and if $\alpha > 1 + \frac{2}{\gamma}$ we have*

$$F(x(t)) - F^* = \mathcal{O}\left(\frac{1}{t^{\frac{2\gamma\alpha}{\gamma+2}}}\right)$$

as $t \to +\infty$.

Proof. Under our assumptions on F and β, Lemma 4.1 implies the existence of a $t_1 \geq t_0$ such that (4.1) and (4.2) hold for all $t \geq t_1$. In this proof we take $t \geq t_1$.

We define $\mathcal{G}(t) = t^p \mathcal{E}(t)$ for $p > 0$. Recall that by Lemma 3.4 with $\beta(t) \equiv 1$

$$\mathcal{E}'_{(\mathrm{AVD}),\xi} = 2ta(t) + \frac{2(\lambda - \alpha + 1)}{t} b_2(t) - \frac{2\lambda\xi}{t} c(t) - t\lambda \langle \nabla F(x(t)), x(t) - x^* \rangle$$

holds, where we limit $\lambda \leq \alpha - 1$ and chose $\xi = \lambda(\lambda - \alpha + 1) < 0$. By applying (4.1) we get

$$\mathcal{E}'_{(\mathrm{AVD}),\xi} \leq (2 - \lambda\gamma) ta(t) + \frac{2(\lambda - \alpha + 1)}{t} b_2(t) - \frac{2\lambda\xi}{t} c(t).$$

This implies for the derivative $\mathcal{G}'(t) = p \cdot t^{p-1}\mathcal{E}(t) + t^p\mathcal{E}'(t)$:

$$\mathcal{G}'(t) \leq t^{p+1}(p+2-\lambda\gamma)a(t) + t^{p-1}(p+2(\lambda-\alpha+1))b_2(t) + t^{p-1}(p-2\lambda)\xi c(t).$$

Choosing $p = \frac{2\gamma\alpha}{\gamma+2} - 2$ and $\lambda = \frac{2\alpha}{\gamma+2}$ leads to the coefficients of $a(t)$ and $b_2(t)$ being zero. This leaves

$$\mathcal{G}'(t) \leq \xi(p-2\lambda)t^{p-1}c(t).$$

Now we apply the second geometrical property (4.2)

$$\mathcal{G}'(t) \leq \frac{\xi(p-2\lambda)}{2K_2}t^{p-1}a(t), \tag{4.7}$$

since $\xi(p-2\lambda)$ is positive by definition. On the other hand $\xi < 0$ holds and we get

$$\mathcal{G}(t) \geq t^p(t^2a(t) + \xi c(t)) \geq t^{p+2}\left(1 + \frac{\xi}{2K_2t^2}\right)a(t),$$

by applying (4.2) again. Choosing t large enough yields

$$\mathcal{G}(t) \geq \frac{1}{2}t^{p+2}a(t) \geq 0. \tag{4.8}$$

Combining (4.7) and (4.8) results in

$$\mathcal{G}'(t) \leq C \cdot \frac{\mathcal{G}(t)}{t^3}$$

for t and $C > 0$ large enough. Grönwall's inequality implies

$$\mathcal{G}(t) \leq \mathcal{O}\left(e^{-\frac{1}{t^2}}\right)$$

and therefore $\mathcal{G}$ is bounded for t large enough. We denote the bound by M. With (4.8) we conclude

$$a(t) \leq \frac{M}{2t^{p+2}},$$

which proves the desired rate, since $p + 2 = \frac{2\gamma\alpha}{\gamma+2}$.

□

We notice that since $\gamma \in [1, 2]$ the exponent of the convergence rate $\frac{2\gamma\alpha}{\gamma+2}$ lies in the interval $[\frac{2}{3}\alpha, \alpha]$. Therefore the best possible rate can be achieved for $\gamma = 2$.

5 Convergence under geometric conditions and time scaling

We have looked at the convergence rates of $F(x(t)) - F^*$ for solutions to both systems either with time scaling or under the geometrical conditions as shown in previous works. In this chapter we analyse the combined effect of time scaling and the geometrical properties of F on the convergence rates. As seen in Chapter 3 there are certain assumptions we need to make for β in order to be able to make the right estimations.

5.1 Constant damping

For the system (CD) with constant damping the time scale parameter β needs to be increasing and unbounded in order to have an effect on the convergence rate. This assumption will also make the convergence rate independent of the factor K_2 in $H_2(2)$ compared to Theorem 4.2. Additionally we want the differential inequality

$$\beta'(t) \leq C \cdot \beta(t)$$

to be satisfied for t large enough and for a certain $C > 0$. By Grönwall's inequality one can see that this means that β can be chosen as large as an exponential function.

Theorem 5.1. *Let $\gamma \in [1, 2]$ and $t_0 > 0$. Let $x(\cdot)$ be the solution of* (CD) *with $\alpha > 0$. Assume that F satisfies $H_1(\gamma)$ and $H_2(2)$ with a real constant $K_2 > 0$ and admits a unique minimizer x^*. If $\beta : [t_0, \infty) \to (0, +\infty)$ is an increasing function with $\lim_{t\to+\infty} \beta(t) = +\infty$ and there exist a constant $C_1 \in (0, \alpha\gamma]$ and a $t_2 \geq t_0$ such that*

$$\beta'(t) \leq C_1 \beta(t) \tag{5.1}$$

for $t \geq t_2$, then for any

$$\lambda \in \left(\frac{C_1 + 2\alpha}{2 + \gamma}, \alpha \right]$$

we have

$$F(x(t)) - F^* = \mathcal{O}\left(\frac{e^{-2(\alpha-\lambda)t}}{\beta(t)} \right)$$

as $t \to +\infty$.

Proof. Under our assumptions on F and β, Lemma 4.1 implies the existence of a $t_1 \geq t_0$ such that (4.1) and (4.2) hold for all $t \geq t_1$. In this proof we take $t \geq \max\{t_1, t_2\}$ so that (5.1) holds as well.

We recall the derivative of the energy function in Lemma 3.3:

$$\begin{aligned}\mathcal{E}'_{(CD),\xi}(t) = \beta'(t)a(t) + 2(\lambda - \alpha)\left(b_1(t) - \lambda^2 c(t)\right) - \lambda\beta(t)\left\langle \nabla F(x(t)), x(t) - x^* \right\rangle \\ + (\xi - \lambda(\lambda - \alpha))\left\langle \dot{x}(t), x(t) - x^* \right\rangle.\end{aligned}$$

We limit $\lambda \leq \alpha$ and choose $\xi = \lambda(\lambda - \alpha)$:

$$\mathcal{E}'_{(\mathrm{CD}),\xi}(t) = \beta'(t)a(t) + 2(\lambda - \alpha)b_1(t) - 2\lambda\xi c(t) - \lambda\beta(t) \langle \nabla F(x(t)), x(t) - x^* \rangle .$$

Using the implication (4.1) of $H_2(2)$ yields

$$\mathcal{E}'_{(\mathrm{CD}),\xi}(t) \leq \beta'(t)a(t) - \gamma\lambda\beta(t)a(t) + 2(\lambda - \alpha)b_1(t) - 2\lambda\xi c(t)$$

and with (4.2) we deduce

$$\mathcal{E}'_{(\mathrm{CD}),\xi}(t) \leq (\beta'(t) - \gamma\lambda\beta(t))a(t) + 2(\lambda - \alpha)b_1(t) - \frac{\lambda\xi}{K_2}a(t).$$

By (5.1) and the fact that ξ is non-positive we obtain

$$\mathcal{E}'_{(\mathrm{CD}),\xi}(t) \leq (C_1 - \gamma\lambda)\beta(t)a(t) + 2(\lambda - \alpha)b_1(t) + \frac{\lambda\,|\xi|}{K_2}a(t). \tag{5.2}$$

As seen in the proofs before, we want the energy function $\mathcal{E}_{(\mathrm{CD}),\xi}$ to be non-increasing, therefore the derivative needs to be non-positive. To achieve this it will be important to choose λ appropriately, which requires to distinguish between two cases.

First case: $C_1 < \alpha\gamma$:

We restrict λ to the interval $(\frac{C_1}{\gamma}, \alpha]$, which is non-empty in this case. By this choice $C_1 - \lambda\gamma$ is negative. Rewriting (5.2) yields

$$\mathcal{E}'_{(\mathrm{CD}),\xi}(t) \leq \left(1 + \frac{\lambda\,|\xi|}{K_2(C_1 - \gamma\lambda)\beta(t)}\right)(C_1 - \gamma\lambda)\beta(t)a(t) + 2(\lambda - \alpha)b_1(t).$$

Since $(C_1 - \lambda\gamma)\beta(t) \to -\infty$ as $t \to \infty$, we deduce that

$$\left(1 + \frac{\lambda\,|\xi|}{K_2(C_1 - \gamma\lambda)\beta(t)}\right) \nearrow 1 \text{ as } t \to \infty.$$

Choose $\lambda > \frac{2\alpha + C_1}{2+\gamma} = \frac{2}{2+\gamma}\alpha + \frac{\gamma}{2+\gamma}\frac{C_1}{\gamma}$, which lies in $(\frac{C_1}{\gamma}, \alpha]$. This is equivalent to $0 \geq 2(\lambda - \alpha) > C_1 - \gamma\lambda$. Therefore we can choose t large enough such that

$$1 > \left(1 + \frac{\lambda\,|\xi|}{K_2(C_1 - \gamma\lambda)\beta(t)}\right) \geq \frac{2(\lambda - \alpha)}{C_1 - \gamma\lambda}.$$

This implies for the derivative

$$\begin{aligned}\mathcal{E}'_{(\mathrm{CD}),\xi}(t) &\leq \frac{2(\lambda - \alpha)}{C_1 - \gamma\lambda}(C_1 - \gamma\lambda)\beta(t)a(t) + 2(\lambda - \alpha)b_1(t) \\ &= 2(\lambda - \alpha)(\beta(t)a(t) + b_1(t)).\end{aligned}$$

Since by definition $\mathcal{E}_{(\mathrm{CD}),\xi}(t) = \beta(t)a(t) + b_1(t) + \xi c(t) \leq \beta(t)a(t) + b_1(t)$ this reduces to

$$\mathcal{E}'_{(\mathrm{CD}),\xi}(t) \leq 2(\lambda - \alpha)\mathcal{E}_{(\mathrm{CD}),\xi}(t).$$

Grönwall's inequality yields

$$\mathcal{E}_{(\mathrm{CD}),\xi}(t) \leq \mathcal{O}\left(e^{2(\lambda - \alpha)t}\right).$$

Since $b_1(t) \geq 0$ and the geometrical property (4.2) hold, we can estimate

$$\mathcal{E}_{(\mathrm{CD}),\xi}(t) = \beta(t)a(t) + b_1(t) + \xi \cdot c(t) \geq \left(\beta(t) - \frac{\lambda\,|\xi|}{K_2}\right) a(t).$$

Therefore from the last two equations we can conclude

$$F(x(t)) - F^* \leq \mathcal{O}\left(\frac{e^{2(\lambda-\alpha)t}}{\beta(t)}\right)$$

as $t \to +\infty$.

Second case: $C_1 = \alpha\gamma$.

We choose $\lambda = \alpha$ which implies that $\xi = \lambda(\lambda - \alpha) = 0$ and therefore by (5.2) $\mathcal{E}'_{(\mathrm{CD}),\xi}(t) \leq 0$ holds. This means that $\mathcal{E}_{(\mathrm{CD}),\xi}$ is decreasing and we get

$$\mathcal{E}_{(\mathrm{CD}),\xi}(t) \leq \mathcal{E}_{(\mathrm{CD}),\xi}(t_1)$$

for $t \geq t_1$. This concludes

$$F(x(t)) - F^* \leq \frac{\mathcal{E}_{(\mathrm{CD}),\xi}(t_1)}{\beta(t)}$$

for $t \geq \max\{t_1, t_2\}$ □

Now the question arises if this rate is truly better than in the case without time scaling. For small $\alpha > 0$ the rates are difficult to compare, since the constants K_2 and C_1 can vary for different choices of F and β. But if α is large enough for (4.6) to be fulfilled, the best possible convergence rate of $F(x(t)) - F^*$ for solutions x of (CD) with no time scaling is $e^{-(\gamma^2 K_2/4a)t}$ as $t \to +\infty$ as seen in Remark 4.3. Increasing α any further makes the rate worse. However by the addition of time scaling the exponent of $e^{-2(\alpha-\lambda)t}$ decreases linearly and β can be a faster growing function, if we choose α to be larger. Both boost the convergence rate of $F(x(t)) - F^*$ as $t \to +\infty$.

Especially for the choice $\beta(t) = c \cdot e^{\alpha\gamma t}$ for all $t \geq t_0$ one can see the improvement of the rate. The inequality (5.1) is satisfied with equality for $C_1 = \alpha\gamma$. This implies that λ has to be equal to α and the convergence rate is

$$F(x(t)) - F^* \leq \mathcal{O}\left(\frac{1}{\beta(t)}\right) = \mathcal{O}\left(e^{-\alpha\gamma t}\right) \leq \mathcal{O}\left(e^{-\alpha t}\right), \tag{5.3}$$

as $t \to +\infty$, since $\gamma \geq 1$. In Remark 4.3 we saw that for solutions of (CD) the rate proven in Theorem 4.2 cannot be better than $e^{-\alpha t}$ for small K_2 or large α.

Without the geometrical conditions the convergence rate in the constant damping case with time scaling was in Theorem 3.6 shown to be

$$F(x(t)) - F^* \leq \mathcal{O}\left(\frac{e^{-\frac{\alpha}{2}t}}{\beta(t)}\right) \tag{5.4}$$

for β satisfying $\beta'(t) \leq \frac{\alpha}{2}\beta(t)$ for t large enough. One can see that the best possible rate, which is achieved for $\beta(t) = e^{\frac{\alpha}{2}t}$, is still not better than (5.3) and only equal for $\gamma = 1$. If we choose β to be a polynomial then the differential inequality (5.1) is fulfilled for arbitrarily small $C_1 > 0$ if t is large enough. By choosing $\lambda = \frac{C_1+2\alpha}{2+\gamma} + \varepsilon$ for small $\varepsilon > 0$, we can rewrite the rate of Theorem 5.1 as

$$F(x(t)) - F^* \leq \mathcal{O}\left(\frac{e^{\left(-\frac{2\gamma\alpha}{2+\gamma}+\frac{C_2+2\alpha}{2+\gamma}+\varepsilon\right)t}}{\beta(t)}\right)$$

as $t \to +\infty$. Since $\frac{2\gamma}{2+\gamma} \in \left[\frac{2}{3}, 1\right]$ and both C_1 and ε can be chosen arbitrarily small, we see that even for polynomial time scaling the convergence rate improves compared to (5.4) by adding the geometrical assumptions on F.

5.2 Asymptotically vanishing damping

For the system (AVD) the time scaling parameter β will be required to satisfy

$$t \cdot \beta'(t) \leq C \cdot \beta(t)$$

for certain $C > 0$ and t large enough in order to prove the convergence rate, which can be attained under the geometrical conditions. This means the function cannot grow faster than a polynomial of a certain degree. For asymptotically vanishing damping β does not need to be increasing nor unbounded for the convergence rate to hold.

Theorem 5.2. *Let $\gamma \in [1, 2]$, $\alpha > 1 + \frac{2}{\gamma}$ and $t_0 > 0$. Let $x(\cdot)$ be the solution of the* (AVD)*. Assume that for $\beta : [t_0, \infty) \to (0, +\infty)$ there exist a constant $C_2 \in (0, \gamma\alpha - (\gamma + 2)]$ s.t.*

$$t \cdot \beta'(t) \leq C_2\beta(t) \tag{5.5}$$

for t large enough. If F satisfies $H_1(\gamma)$ and $H_2(2)$ with a real constant $K_2 > 0$ and admits a unique minimizer x^*, then

$$F(x(t)) - F^* = \mathcal{O}\left(\frac{1}{t^{\frac{2\gamma\alpha-2C_2}{\gamma+2}}\beta(t)}\right)$$

as $t \to +\infty$.

Proof. Under our assumptions on F and β, Lemma 4.1 implies the existence of a $t_1 \geq t_0$ such that (4.1) and (4.2) hold for all $t \geq t_1$. In this proof we take $t \geq \max\{t_1, t_2\}$ so that (5.1) holds as well.
By the result of Lemma 3.4, we know that

$$\begin{aligned}\mathcal{E}'_{(AVD),\xi} &= (2t\beta(t) + t^2\beta'(t))a(t) + \frac{2(\lambda - \alpha + 1)}{t}\left(b_2(t) - \lambda^2 c(t)\right)\\ &\quad - t\lambda\beta(t)\left\langle \nabla F(x(t)), x(t) - x^*\right\rangle + (\xi - \lambda(\lambda - \alpha + 1)\left\langle \dot{x}(t), x(t) - x^*\right\rangle\end{aligned}$$

holds. By choosing $\xi = \lambda(\lambda + 1 - \alpha)$, which is non-positive for $\lambda \leq \alpha - 1$, and by applying (4.1) we see that

$$\begin{aligned}\mathcal{E}'_{(\mathrm{AVD}),\xi}(t) \leq \left((2 - \lambda\gamma)t\beta(t) + t^2\beta'(t)\right)a(t) + \left(\frac{2(\lambda - \alpha + 1)}{t^2}\right)b_2(t)&\\ - \left(\frac{2\lambda}{t}\right)\xi c(t).&\end{aligned} \tag{5.6}$$

We define $\mathcal{G}(t) = t^p E(t)$. For the derivative

$$\mathcal{G}'(t) = pt^{p-1}\mathcal{E}_{(\mathrm{AVD}),\xi}(t) + t^p\mathcal{E}'_{(\mathrm{AVD}),\xi}(t)$$

holds. By summing up the coefficients of $a(t)$, $b_2(t)$ and $c(t)$ in (3.2) and (5.6) we can see that

$$\begin{aligned}
\mathcal{G}'(t) \leq& \left(pt^{p+1}\beta(t) + (2-\lambda\gamma)t^{p+1}\beta(t) + t^{p+2}\beta'(t)\right)a(t)\\
&+ \left(pt^{p-1} + 2(\lambda-\alpha+1)t^{p-1}\right)b_2(t)\\
&+ \left(pt^{p-1} - 2\lambda t^{p-1}\right)\xi c(t),\\
\mathcal{G}'(t) \leq& \left((p+2-\lambda\gamma)\beta(t) + t\beta'(t)\right)t^{p+1}a(t)\\
&+ (p+2(\lambda-\alpha+1))\,t^{p-1}b_2(t)\\
&+ (p-2\lambda)\,\xi t^{p-1}c(t).
\end{aligned} \tag{5.7}$$

We notice that $\frac{2\alpha+C}{\gamma+2} \leq \frac{2\alpha+\gamma\alpha-(\gamma+2)}{\gamma+2} = \alpha - 1$ holds, since $C \leq \gamma\alpha - (\gamma+2)$. Defining $p = \frac{2\gamma\alpha-2C}{\gamma+2} - 2$ and $\lambda = \frac{2\alpha+C}{\gamma+2}$ yields that the coefficient of $t^{p-1}b_2(t)$ in (5.7) becomes zero, since

$$p + 2(\lambda-\alpha+1) = \frac{2\gamma\alpha-2C}{\gamma+2} - 2 + 2\left(\frac{2\alpha-C}{\gamma+2} - \frac{\alpha(\gamma+2)}{\gamma+2} + 1\right) = 0,$$

and also that the coefficient of $c(t)$ is $(p-2\lambda)\,\xi \geq -2\xi \geq 0$, since $p - 2\lambda \leq -2$. Furthermore notice that $p + 2 - \lambda\gamma = -C$. Therefore we obtain

$$\mathcal{G}'(t) \leq \left(-C_2\beta(t) + t\beta'(t)\right)t^{p+1}a(t) + (p-2\lambda)\xi t^{p-1}c(t).$$

By the differential inequality (5.5) $-C_2\beta(t) + t\beta'(t) \leq 0$ and thus

$$\mathcal{G}'(t) \leq (p-2\lambda)\xi t^{p-1}c(t)$$

holds. The condition $H_2(2)$ implies (4.2) which yields

$$\mathcal{G}'(t) \leq \frac{(p-2\lambda)\xi}{2K_2}t^{p-1}a(t). \tag{5.8}$$

By the definition of $\mathcal{G}(t)$ and since $b_2(t) \geq 0$ we can see that

$$\mathcal{G}(t) \geq t^p(t^2\beta(t)a(t) + \xi c(t)).$$

Applying condition $H_2(2)$ yields

$$\mathcal{G}(t) \geq t^p\left(t^2\beta(t) + \frac{\xi}{2K_2}\right)a(t). \tag{5.9}$$

Combining (5.8) and (5.9) results in

$$\mathcal{G}'(t) \leq \frac{(p-2\lambda)\xi}{2K_2}\cdot\frac{1}{t}\cdot\frac{1}{t^2\beta(t)+\frac{\xi}{2K_2}}\mathcal{G}(t).$$

We can find a constant $C > 0$ for t large enough such that $\mathcal{G}'(t) \leq Ct^{-3}\mathcal{G}(t)$. By Grönwall's inequality $\mathcal{G}(t) \leq \mathcal{O}\left(e^{-t^{-2}}\right)$ holds. Therefore H is bounded and we can find a positive constant $M > 0$ such that $\mathcal{G}(t) \leq M$. This leads with (5.9) to

$$a(t) \leq \frac{M}{t^{p+2}\beta(t) + \frac{\xi}{2K_2}t^p} \leq \mathcal{O}\left(\frac{1}{t^{p+2}\beta(t)}\right)$$

for t large enough. Since $a(t) = F(x(t)) - F^*$ we conclude

$$F(x(t)) - F^* = \mathcal{O}\left(\frac{1}{t^{p+2}\beta(t)}\right),$$

which proves the result by our choice of $p = \frac{2\gamma\alpha - 2C}{\gamma+2} - 2$. $\square$

In the paper [2] as seen in Theorem 4.4 the rate

$$F(x(t)) - F^* = \mathcal{O}\left(\frac{1}{t^{\frac{2\gamma\alpha}{\gamma+2}}}\right) \tag{5.10}$$

as $t \to +\infty$ was proven with the geometrical assumptions on F but without time scaling under the same conditions on γ and α. Our condition on the function β implies by Grönwall's inequality that

$$\beta(t) \leq \mathcal{O}\left(t^{C_2}\right)$$

as $t \to +\infty$. If we choose $\beta(t) = t^{C_2}$ with $C_2 \leq \gamma\alpha - (\gamma + 2)$, then (5.5) holds with equality. Inserting this β in the proven convergence rate of 5.2 we get

$$F(x(t)) - F^* = \mathcal{O}\left(\frac{1}{t^{\frac{2\gamma\alpha + C_2\gamma}{\gamma+2}}}\right) \tag{5.11}$$

as $t \to +\infty$, which is cleary an improvement of (5.10). We see that the best possible convergence rate occurs in the case of $C_2 = \gamma\alpha - (\gamma+2)$ and $\beta(t) = t^{C_2}$:

$$F(x(t)) - F^* = \mathcal{O}\left(\frac{1}{t^{\gamma\alpha - \gamma}}\right)$$

as $t \to +\infty$. If we only apply time scaling the convergence rate was proven in Theorem 3.7 to be $\mathcal{O}\left(\frac{1}{t^2\beta(t)}\right)$ for β satisfying $t\beta'(t) \leq (\alpha - 3)\beta(t)$ for t large enough. Since $\gamma\alpha - (\gamma + 2) \in [\alpha - 3, 2\alpha - 4]$ the additions of the geometrical conditions allow us to choose polynomials of larger degree for β. Furthermore one can easily see that the convergence rate for the latter is better.

6 Numerical comparison

The aim of this chapter is to visualize the effects of different choices for time scaling on the convergence rates compared to no time scaling. Since a function either satisfies or does not satisfy the geometrical conditions, and because the comparison of the trajectories of two completely different functions would not be conclusive, we will only focus on functions F that fulfill both conditions $H_1(\gamma)$ and $H_2(2)$. The Cauchy-problems were solved in `Python` using the solver `odeint` of the package `scipy.integrate`. All figures in this chapter were generated in `Python` by the files presented in the appendix.

6.1 First example

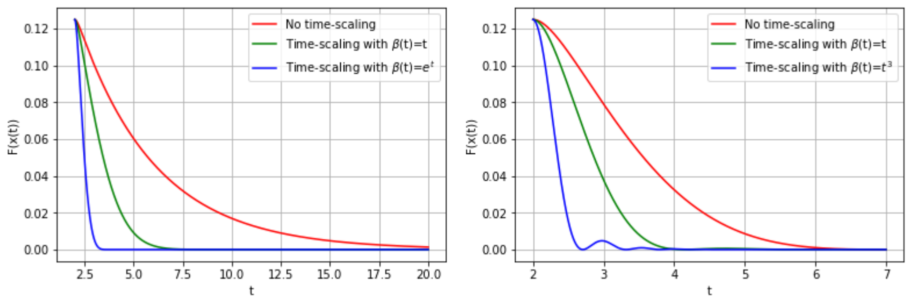
Let $F : \mathbb{R} \to \mathbb{R}$, $F(x) := \frac{x^2}{2}$. Clearly F is convex and has the mimizer $x^* = 0$ with $F(x^*) = 0$. On the one hand one can easily see that

$$F(x) - F(x^*) = \frac{x^2}{2} = \frac{1}{2} \cdot 2x \cdot x = \frac{1}{2} \langle F'(x), x - x^* \rangle$$

and therefore F fulfills $H_1(2)$. On the other hand

$$\frac{1}{2} \|x - x^*\|^2 = \frac{x^2}{2} = F(x) - F(x^*)$$

and therefore $H_2(2)$ holds with $K_2 = \frac{1}{2}$. For each system we choose $\alpha = 8$ and we will look at three different options for β. The first one is the case $\beta \equiv 1$ with no time scaling and the second one is $\beta(t) = t$ for all $t \geq t_0$. For (CD) we choose $\beta(t) = e^t$ for all $t \geq t_0$ and for (AVD) $\beta(t) = t^3$ for all $t \geq t_0$ as a third, since they fulfill the differential inequalities (5.1) and (5.5) respectively. The starting value is set as $(x_0, v_0) = (0.5, 0)$. The plots show the function value $F(x(t))$, the left one if x is a solution of (CD) and the right one for a solution of (AVD).



Accordingly to the results in the previous chapter we see improvements of the convergence rates by adding time scaling, especially for the constant damping case, where we saw in Chapter 5 that the rate significantly improves if $\frac{2\alpha^2}{\gamma(\gamma+2)K_2} > \frac{1}{2}$ holds. We also are able to notice the positive effect of choosing faster growing time scaling parameter β.

6.2 Second example

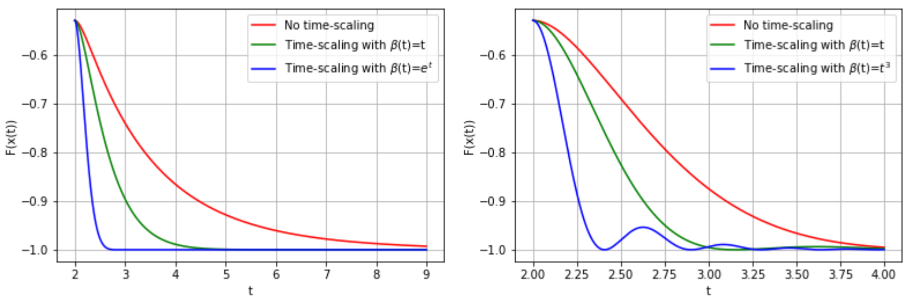
Let $F : \mathbb{R} \to \mathbb{R}$, $F(x) := -e^{-x^2} + x^2$. Again F is convex and has the minimizer $x^* = 0$ with $F(x^*) = -1$. We can check that

$$F(x) - F(x^*) = -e^{-x^2} + x^2 + 1 \leq \frac{1}{2} \cdot \left(2x^2 e^{-x^2} + 2x^2\right) = \frac{1}{2} \langle F'(x), x - x^* \rangle$$

holds for x large enough and therefore F fulfills $H_1(2)$. Furthermore

$$\frac{1}{2} \|x - x^*\|^2 = \frac{x^2}{2} \leq -e^{-x^2} + x^2 + 1 = F(x) - F(x^*)$$

holds for x large enough and therefore $H_2(2)$ holds with $K_2 = \frac{1}{2}$. The choices for α, $\beta(t)$ and the starting value will remain the same as in the previous example. The plots show the function value $F(x(t))$, the left one if x is a solution of (CD) and the right one for a solution of (AVD).



Again the same improvements of the convergence rates for both systems are visible. We notice that $\beta(t) = t^3$ has some slight oscillations.

7 Conclusion and outlook

As expected the combination of the geometrical conditions on F and the introduction of time scaling to both systems (CD) and (AVD) improved the convergence rates of $F(x(t)) - F^*$ as $t \to +\infty$ significantly. We showed in the proofs what the largest possible choices for β are, but we also saw that it is sufficient for β to be increasing and unbounded to achieve faster results compared to no time scaling. The numerical results of the implementation in `Python` were able support our established improvements.

It is further possible to study the introduction of Hessian driven damping to the systems (CD) and (AVD) with time scaling under the geometrical conditions of F. In the paper [3] was shown that adding Hessian driven damping attenuates possible oscillations of $F(x(t))$ in time, which we noticed in our numerical example, and enjoys further favorable convergence properties. It would be of interest to study if those can be enhanced if the geometrical conditions on F are added. During the writing of this thesis the question arose, if further geometrical assumptions on F can be found, that would make it possible to establish better convergence rates and allow for faster growing time scale parameters β.

In the literature the differential equations (CD) and (AVD) are commonly discretized into iterative schemes. The converges rates of the iterates generated by these algorithms are similar to those of the continuous systems. It would be interesting to analyse how the results of this thesis translate to the discretized schemes.

References

[1] Mohamed Jendoubi and Ramzi May. "Asymptotics for a second-order differential equation with nonautonomous damping and an integrable source term". In: *Applicable Analysis* 94 (Feb. 2015), pp. 435–445.

[2] Othmane Sebbouh, Charles H Dossal, and Aude Rondepierre. "Convergence rates of damped inertial dynamics under geometric conditions". In: *SIAM Journal on Optimization* 30.3 (Aug. 2020), pp. 1850–1877. URL: `https://hal.archives-ouvertes.fr/hal-02173978`.

[3] Hédy Attouch, Zaki Chbani, Jalal M. Fadili, and Hassan Riahi. "Convergence of iterates for first-order optimization algorithms with inertia and Hessian driven damping". In: *Optimization* (2021). URL: `https://hal.archives-ouvertes.fr/hal-03285272`.

[4] Jean-Francois Aujol, Charles Dossal, and Aude Rondepierre. "Optimal Convergence Rates for Nesterov Acceleration". In: *SIAM Journal on Optimization* 29.4 (2019), pp. 3131–3153. eprint: `https://doi.org/10.1137/18M1186757`. URL: `https://doi.org/10.1137/18M1186757`.

[5] Terence Tao. *Nonlinear Dispersive Equations: Local and Global Analysis*. 106. American Mathematical Soc., 2006.

[6] Felipe Alvarez. "On the Minimizing Property of a Second Order Dissipative System in Hilbert Spaces". In: *SIAM Journal on Control and Optimization* 38.4 (2000), pp. 1102–1119.

8 Appendix

The .py - file for the first example in chapter 6

```
from scipy.integrate import odeint
import numpy as np
import matplotlib.pyplot as plt


""""Define F """
def F(x): return x**2/2

"""Define the differential equations as first order systems"""

""" eq1 denotes CD with different time scaling """
def eq1(y,t,alpha):
    u,v=y
    dydt= [v,-alpha*v-u]
    return dydt

def eq1tsc(y,t,alpha):
    u,v=y
    dydt= [v,-alpha*v-t*u]
    return dydt

def eq1tsc2(y,t,alpha):
    u,v=y
    dydt= [v,-alpha*v-np.exp(t)*u]
    return dydt

""" eq2 denotes AVD with different time scaling """

def eq2(y,t,alpha):
    u,v=y
    dydt= [v,-alpha/t*v-u]
    return dydt

def eq2tsc(y,t,alpha):
    u,v=y
    dydt= [v,-alpha/t*v-t*u]
    return dydt

def eq2tsc2(y,t,alpha):
    u,v=y
    dydt= [v,-alpha/t*v-t**3*u]
    return dydt




"""The values of F(u(t)) as t gets bigger"""

alpha=8
```

```
y0=[0.5,0]


""" Plot for different alphas """
tmax=20
n=10000

t=np.linspace(2,tmax,n)
sol=odeint(eq1,y0,t, args=(alpha,))
sol2=odeint(eq1tsc,y0,t, args=(alpha,))
sol3=odeint(eq1tsc2,y0,t, args=(alpha,))

plt.plot(t, F(sol[:, 0]), 'r', label='No time-scaling')
plt.plot(t, F(sol2[:, 0]), 'g', label=r'Time-scaling with $\beta$(t)=t')
plt.plot(t, F(sol3[:, 0]), 'b', label=r'Time-scaling with $\beta$(t)=$e^{t}$')
plt.legend(loc='best')
plt.xlabel('t')
plt.ylabel('F(x(t))')
plt.grid()
plt.show()


""" Plot for different alphas """
tmax=7
n=10000

t=np.linspace(2,tmax,n)
sol=odeint(eq2,y0,t, args=(alpha,))
sol2=odeint(eq2tsc,y0,t, args=(alpha,))
sol3=odeint(eq2tsc2,y0,t, args=(alpha,))

plt.plot(t, F(sol[:, 0]), 'r', label='No time-scaling')
plt.plot(t, F(sol2[:, 0]), 'g', label=r'Time-scaling with $\beta$(t)=t')
plt.plot(t, F(sol3[:, 0]), 'b', label=r'Time-scaling with $\beta$(t)=$t^3$')
plt.legend(loc='best')
plt.xlabel('t')
plt.ylabel('F(x(t))')
plt.grid()
plt.show()
```

The .py - file for the second example in chapter 6

```
from scipy.integrate import odeint
import numpy as np
import matplotlib.pyplot as plt

""""Define F """
def F(x): return -np.exp(-x**2)+x**2

def grad(x): return 2*x*np.exp(-x**2)*abs(x)+2*x

"""Define the differential equations as first order systems"""
```

```python
""" eq1 denotes CD with different time scaling """
def eq1(y,t,alpha):
    u,v=y
    dydt= [v,-alpha*v-grad(u)]
    return dydt

def eq1tsc(y,t,alpha):
    u,v=y
    dydt= [v,-alpha*v-t*grad(u)]
    return dydt

def eq1tsc2(y,t,alpha):
    u,v=y
    dydt= [v,-alpha*v-np.exp(t)*grad(u)]
    return dydt

""" eq2 denotes AVD with different time scaling """

def eq2(y,t,alpha):
    u,v=y
    dydt= [v,-alpha/t*v-grad(u)]
    return dydt

def eq2tsc(y,t,alpha):
    u,v=y
    dydt= [v,-alpha/t*v-t*grad(u)]
    return dydt

def eq2tsc2(y,t,alpha):
    u,v=y
    dydt= [v,-alpha/t*v-t**3*grad(u)]
    return dydt



"""The values of F(u(t)) as t gets bigger"""

alpha=8

y0=[0.5,0]


""" Plot for different alphas """
tmax=9
n=10000

t=np.linspace(2,tmax,n)
sol=odeint(eq1,y0,t, args=(alpha,))
sol2=odeint(eq1tsc,y0,t, args=(alpha,))
sol3=odeint(eq1tsc2,y0,t, args=(alpha,))

plt.plot(t, F(sol[:, 0]), 'r', label='No time-scaling')
plt.plot(t, F(sol2[:, 0]), 'g', label=r'Time-scaling with $\beta$(t)=t')
plt.plot(t, F(sol3[:, 0]), 'b', label=r'Time-scaling with $\beta$(t)=$e^{t}$')
```

```python
plt.legend(loc='best')
plt.xlabel('t')
plt.ylabel('F(x(t))')
plt.grid()
plt.show()


""" Plot for different alphas """
tmax=4
n=10000

t=np.linspace(2,tmax,n)
sol=odeint(eq2,y0,t, args=(alpha,))
sol2=odeint(eq2tsc,y0,t, args=(alpha,))
sol3=odeint(eq2tsc2,y0,t, args=(alpha,))

plt.plot(t, F(sol[:, 0]), 'r', label='No time-scaling')
plt.plot(t, F(sol2[:, 0]), 'g', label=r'Time-scaling with $\beta$(t)=t')
plt.plot(t, F(sol3[:, 0]), 'b', label=r'Time-scaling with $\beta$(t)=$t^3$')
plt.legend(loc='best')
plt.xlabel('t')
plt.ylabel('F(x(t))')
plt.grid()
plt.show()
```