Nesterov's Accelerated Gradient Descent as a Nash-Equilibrium Seeking Algorithm for Quadratic Games

Jay Paek

December 19, 2025

Introduction

This research paper delves into the intricate interplay between Nesterov's Accelerated Gradient Descent (NAGD) and the ordinary differential equation (ODE) that encapsulates its dynamic behavior, offering a profound insight into the optimization landscape. NAGD has firmly established itself as a potent optimization algorithm renowned for its rapid convergence in solving convex optimization problems. One of its defining characteristics is the presence of a specific ODE, known as the Su-Boyd-Candès equation, which governs the evolution of its iterates. Furthermore, this paper delves into the transformative role of NAGD as a Nash Equilibrium Seeking Algorithm (NESA), unraveling its applications in various domains where it seeks stable equilibria among rational agents. However, in the pursuit of equilibrium, we encounter scenarios where the ODE associated with NAGD defies convergence.

This paper introduces Chetaev's Instability Theorem, shedding light on the situations where the Nesterov equation fails to reach a stable state. This theorem serves as a critical cautionary note, highlighting the limitations of NAGD in certain optimization landscapes and its implications for equilibrium-seeking applications. Through this exploration, we aim to provide a comprehensive understanding of both the promising capabilities and potential challenges posed by NAGD and its associated ODE, enriching the discourse on optimization theory and multi-agent systems.

Notations and Preliminaries

Let $\mathbb{R}_{++}$ denote the set of real numbers strictly greater than 0. For a symmetric matrix A, the notation $A \succ 0$ means that A is positive definite. The scalar $t \in \mathbb{R}_{++}$ denotes the (continuous) time variable of the trajectory. The first and second time derivatives of a vector-valued trajectory $X(t)$ are denoted by $\dot{X}(t)$ and $\ddot{X}(t)$, respectively.

For an arbitrary matrix $A \in \mathbb{R}^{m\times n}$, the range space of A is

$$\mathcal{R}(A) \triangleq \{x \in \mathbb{R}^m : \exists\, y \in \mathbb{R}^n \text{ such that } x = Ay\},$$

and the null space of A is

$$\mathcal{N}(A) \triangleq \{x \in \mathbb{R}^n : Ax = 0\}.$$

For a set of vectors $S \subseteq \mathbb{R}^n$, its orthogonal complement is denoted

$$S^{\perp} \triangleq \{x \in \mathbb{R}^n : x^{\top} y = 0,\ \forall\, y \in S\}.$$

The Euclidean norm of a vector is denoted by $\|\cdot\|_2$.

Nesterov Accelerated Gradient Dynamics

Consider the second-order differential equation that models Nesterov accelerated gradient dynamics (NAGD) [?]:

$$\ddot{X}(t) + \frac{3}{t}\dot{X}(t) + \nabla f\big(X(t)\big) = 0, \tag{1}$$

where $f : \mathbb{R}^n \to \mathbb{R}$ is an objective function (e.g., a cost function in a game or a loss function in a learning model) and $\nabla f(\cdot)$ denotes its gradient. Introducing the state variables

$$x_1(t) \triangleq X(t), \qquad x_2(t) \triangleq \dot{X}(t),$$

the dynamics (1) can be written as the first-order system

$$\begin{cases} \dot{x}_1(t) = x_2(t), \\ \dot{x}_2(t) = -\dfrac{3}{t}x_2(t) - \nabla f\big(x_1(t)\big), \end{cases} \tag{2}$$

where $x_1(t), x_2(t) \in \mathbb{R}^n$ represent the position and momentum states, respectively.

Quadratic Games and the Pseudo-Gradient

Consider a game with N players, where each player $i \in \{1, \ldots, N\}$ is equipped with a quadratic cost function

$$J_i(\cdot) : \mathbb{R}^N \to \mathbb{R}.$$

Let the joint action profile be

$$p \triangleq \begin{bmatrix} p_1 & \cdots & p_N \end{bmatrix}^\top \in \mathbb{R}^N,$$

where $p_i \in \mathbb{R}$ is the action of player i. In what follows, we use $x \in \mathbb{R}^N$ (and, when needed, an auxiliary state such as y) to denote the joint decision/state variable associated with the Nash equilibrium seeking dynamics for the N-player game.

Each player cost is assumed quadratic of the form

$$J_i(x) = x^\top Q_i x,$$

where $Q_i \in \mathbb{R}^{N\times N}$ is a given matrix (typically taken symmetric in quadratic-form representations). Define the stacked cost vector map $J : \mathbb{R}^N \to \mathbb{R}^N$ by

$$J(x) \triangleq \begin{bmatrix} J_1(x) \\ \vdots \\ J_N(x) \end{bmatrix} = \begin{bmatrix} x^\top Q_1 x \\ \vdots \\ x^\top Q_N x \end{bmatrix}.$$

Instead of the standard gradient of a scalar objective, we use the pseudo-gradient mapping (also called the game gradient) as the analogue of $\nabla f(x)$:

$$\nabla f(x) \triangleq \begin{bmatrix} \dfrac{\partial J_1}{\partial x_1}(x) \\ \vdots \\ \dfrac{\partial J_N}{\partial x_N}(x) \end{bmatrix}, \tag{3}$$

i.e., the ith component of $\nabla f(x)$ is obtained by differentiating player i's objective J_i with respect to player i's own action variable x_i.

For quadratic games of the form $J_i(x) = x^\top Q_i x$, each component of (3) is a linear combination of the entries of x. Consequently, the pseudo-gradient can be written in the compact linear form

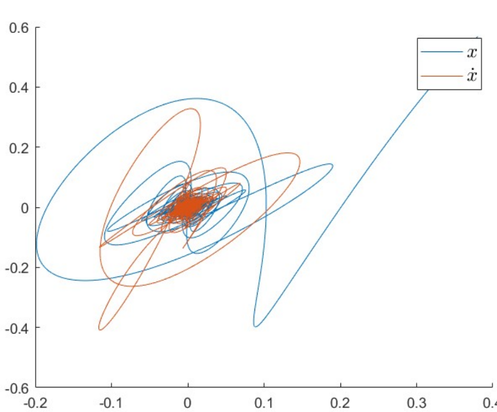
$$\nabla f(x) = \mathcal{G}x, \tag{4}$$

for some matrix $\mathcal{G} \in \mathbb{R}^{N\times N}$ (the pseudo-gradient matrix).

For such games, a Nash equilibrium corresponds to a point in the null space of the pseudo-gradient matrix. In particular, if $\mathcal{G}$ is full rank, then $\mathcal{N}(\mathcal{G}) = \{0\}$ and the equilibrium is at $x^\star = 0$.

Illustrative NAGD Trajectories

Below we illustrate trajectories x of the NAGD dynamics, where the horizontal and vertical axes correspond to the first and second components of x, respectively. All simulations are generated over 10^5 iterations with $t \in [1, 10^3]$ and time increments of 0.01.



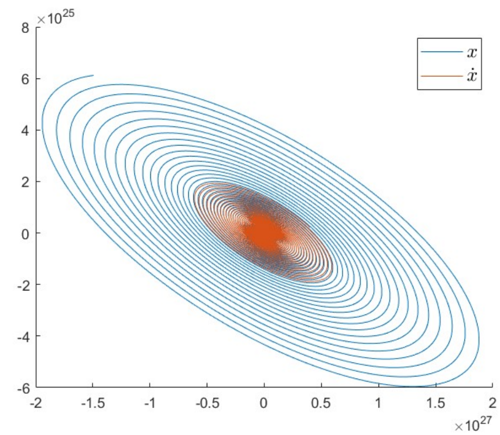


Figure 1: Nesterov dynamics with two different full-rank pseudo-gradient matrices. $\mathcal{G} = \begin{bmatrix} 0.4 & 0.2 \\ 0.2 & 0.8 \end{bmatrix}$ (left) and $\mathcal{G} = \begin{bmatrix} 0.1 & -0.1 \\ 0 & 0.1 \end{bmatrix}$ (right).

Figure 1 displays two trajectories of the NAGD dynamics initialized at a random point satisfying $\|x\|_2 < 1$, where neither trajectory converges. More broadly, in these experiments the NAGD dynamics do not converge for any full-rank $N \times N$ pseudo-gradient matrix $\mathcal{G}$.

Figure 2 shows two additional trajectories of the NAGD dynamics initialized at a random point satisfying $\|x\|_2 < 1$. In both cases, the state converges to the null space of the low-rank pseudo-gradient matrix $\mathcal{G}$. The trajectory on the left trails along $\alpha \begin{bmatrix} 1 \\ -1 \end{bmatrix}$, and the trajectory on the right trails along $\alpha \begin{bmatrix} 2 \\ -1 \end{bmatrix}$, where $\alpha \in \mathbb{R}$.

Scope of This Paper

In this paper, we prove stability and instability properties of NAGD dynamics for N-player quadratic games under various classes of pseudo-gradient matrices.

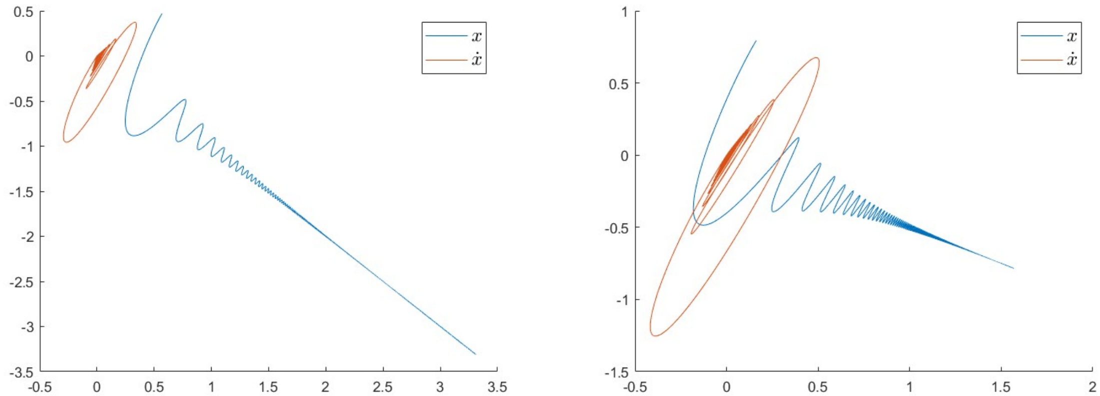


Figure 2: Nesterov dynamics with two different low-rank pseudo-gradient matrices.
$\mathcal{G} = \begin{bmatrix} 1 & 1 \\ 2 & 2 \end{bmatrix}$ (left) and $\mathcal{G} = \begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix}$ (right).

Projected Dynamics

This section introduces a projection-based reduction that can simplify Lyapunov stability analysis by decoupling the dynamics along a prescribed orthonormal basis.

Proposition 1 (Stability via orthogonal projections)**.** *Consider the autonomous ODE*

$$\dot{x}(t) = f(x(t)), \qquad x(t) \in \mathbb{R}^n, \tag{5}$$

where $f : \mathbb{R}^n \to \mathbb{R}^n$ is linear*, i.e., $f(x) = Ax$ for some $A \in \mathbb{R}^{n\times n}$. Let $\{w_i\}_{i=1}^n \subset \mathbb{R}^n$ be an orthonormal basis, and define the scalar coordinates*

$$y_i(t) \triangleq w_i^\top x(t), \qquad i = 1, \dots, n.$$

Assume that for each i there exists a function $g_i : \mathbb{R} \to \mathbb{R}$ such that, along solutions of (5)*,*

$$w_i^\top f(x(t)) = g_i\big(w_i^\top x(t)\big) = g_i\big(y_i(t)\big), \qquad \forall\, t \geq 0. \tag{6}$$

Then the following statements are equivalent:

1. *The equilibrium $x = 0$ of* (5) *is Lyapunov stable.*
2. *For every $i \in \{1, \dots, n\}$, the scalar system*

$$\dot{y}_i(t) = g_i\big(y_i(t)\big) \tag{7}$$

   has a Lyapunov-stable equilibrium at $y_i = 0$.

Proof. **Step 1 (Projected dynamics are well-defined).** By definition, $y_i(t) = w_i^\top x(t)$, hence $\dot{y}_i(t) = w_i^\top \dot{x}(t) = w_i^\top f(x(t))$. Using (6) yields $\dot{y}_i(t) = g_i(y_i(t))$, which is exactly (7).

(1 ⇒ 2). Assume $x = 0$ is Lyapunov stable for (5). Fix any i and any $\varepsilon > 0$. By Lyapunov stability, there exists $\delta > 0$ such that $\|x(0)\|_2 < \delta$ implies $\|x(t)\|_2 < \varepsilon$ for all $t \geq 0$. Since $\{w_i\}$ is orthonormal, $|y_i(t)| = |w_i^\top x(t)| \leq \|x(t)\|_2 < \varepsilon$ for all $t \geq 0$. Moreover, $|y_i(0)| \leq \|x(0)\|_2 < \delta$. Therefore $y_i = 0$ is Lyapunov stable for (7). As i was arbitrary, all scalar equilibria are stable.

(2 ⇒ 1). Assume that for each i, the equilibrium $y_i = 0$ of (7) is Lyapunov stable. Let $\varepsilon > 0$ be given and set $\varepsilon_i \triangleq \varepsilon/\sqrt{n}$ for all i. For each i, there exists $\delta_i > 0$ such that $|y_i(0)| < \delta_i$ implies $|y_i(t)| < \varepsilon_i$ for all $t \geq 0$. Define

$$\delta \triangleq \min_{1 \leq i \leq n} \delta_i.$$

If $\|x(0)\|_2 < \delta$, then $|y_i(0)| = |w_i^\top x(0)| \leq \|x(0)\|_2 < \delta \leq \delta_i$ for every i, hence $|y_i(t)| < \varepsilon/\sqrt{n}$ for all $t \geq 0$ and all i. Using Parseval's identity for an orthonormal basis,

$$\|x(t)\|_2^2 = \sum_{i=1}^{n} \big(w_i^\top x(t)\big)^2 = \sum_{i=1}^{n} y_i(t)^2 < \sum_{i=1}^{n} \left(\frac{\varepsilon}{\sqrt{n}}\right)^2 = \varepsilon^2,$$

so $\|x(t)\|_2 < \varepsilon$ for all $t \geq 0$. Therefore $x = 0$ is Lyapunov stable for (5). □

Remark 1 (Interpretation and consistency)**.** *The assumption* (6) *asserts that, in the orthonormal coordinates $y_i = w_i^\top x$, the ith projected component of the vector field depends only on the corresponding scalar coordinate. Consequently, the dynamics decouple along $\{w_i\}$ and Lyapunov stability of the origin is equivalent to Lyapunov stability of each one-dimensional subsystem* (7)*. In particular, since f is linear, each g_i is necessarily linear in y_i (possibly with a coefficient depending on i).*

Projected Analysis for Low-Rank and Full-Rank Pseudo-Gradients

We analyze the NAGD game dynamics

$$\ddot{x}(t) + \frac{3}{t}\dot{x}(t) + \mathcal{G}x(t) = 0, \qquad t \in [t_0, \infty),\ t_0 > 0, \tag{8}$$

equivalently, in first-order form with $x_1(t) = x(t)$ and $x_2(t) = \dot{x}(t)$,

$$\frac{d}{dt}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = A(t)\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}, \qquad A(t) \triangleq \begin{bmatrix} \mathbf{0} & I \\ -\mathcal{G} & -\frac{3}{t}I \end{bmatrix}. \tag{9}$$

Throughout, we work on $t \geq t_0$ to avoid the singularity at $t = 0$.

Low-rank pseudo-gradient: convergence of null-space projections

To connect Lyapunov stability to the long-run behavior, it is useful to isolate the components of the trajectory that lie in directions annihilated by the pseudo-gradient. Since $\mathcal{G}$ may be nonsymmetric, the appropriate projection uses the *left* null space.

Lemma 1 (Left-nullspace projection for rank-deficient $\mathcal{G}$)**.** *Assume rank($\mathcal{G}$) $< n$. Then there exists a nonzero vector $w \in \mathcal{N}(\mathcal{G}^\top)$, i.e., $w \neq \mathbf{0}$ and $w^\top \mathcal{G} = \mathbf{0}^\top$. For any such w, define the scalar projected states*

$$y_1(t) \triangleq w^\top x_1(t), \qquad y_2(t) \triangleq w^\top x_2(t).$$

Then (y_1, y_2) satisfies the decoupled linear time-varying system

$$\dot{y}_1(t) = y_2(t), \qquad \dot{y}_2(t) = -\frac{3}{t}y_2(t). \tag{10}$$

Proof. Since $\operatorname{rank}(\mathcal{G}) < n$, $\dim \mathcal{N}(\mathcal{G}^\top) = n - \operatorname{rank}(\mathcal{G}) > 0$, so there exists $w \neq \mathbf{0}$ with $w^\top \mathcal{G} = \mathbf{0}^\top$. Left-multiplying (9) by $w^\top$ gives

$$\frac{d}{dt}(w^\top x_1) = w^\top x_2, \qquad \frac{d}{dt}(w^\top x_2) = -\frac{3}{t}w^\top x_2 - w^\top \mathcal{G} x_1,$$

and $w^\top \mathcal{G} = \mathbf{0}^\top$ eliminates the coupling term, yielding (10). □

Explicit solution and uniform boundedness. Solving (10) yields

$$y_2(t) = \left(\frac{t_0}{t}\right)^3 y_2(t_0), \tag{11}$$

and therefore

$$y_1(t) = y_1(t_0) + \int_{t_0}^{t} y_2(s)\,ds = y_1(t_0) + \frac{t_0}{2}\Big(1 - \frac{t_0^2}{t^2}\Big) y_2(t_0). \tag{12}$$

Equivalently,

$$\begin{bmatrix} y_1(t) \\ y_2(t) \end{bmatrix} = \Phi_w(t,t_0) \begin{bmatrix} y_1(t_0) \\ y_2(t_0) \end{bmatrix}, \qquad \Phi_w(t,t_0) \triangleq \begin{bmatrix} 1 & \frac{t_0}{2}\Big(1 - \frac{t_0^2}{t^2}\Big) \\ 0 & \left(\frac{t_0}{t}\right)^3 \end{bmatrix}.$$

Fixing $t_0 > 0$, $\Phi_w(t,t_0)$ is bounded uniformly for all $t \geq t_0$; e.g.,

$$\sup_{t \geq t_0} \|\Phi_w(t,t_0)\| < \infty,$$

so the projected subsystem is uniformly stable (in the standard linear time-varying sense on $[t_0, \infty)$). Moreover, (11) implies $y_2(t) \to 0$ and (12) implies $y_1(t)$ converges to a finite limit:

$$\lim_{t\to\infty} y_1(t) = y_1(t_0) + \frac{t_0}{2} y_2(t_0). \tag{13}$$

Interpretation. For every $w \in \mathcal{N}(\mathcal{G}^\top)$, the scalar projection $w^\top x_1(t)$ converges to a constant and $w^\top x_2(t)$ decays to 0. Hence, the component of $x_1(t)$ along the left-nullspace directions (in the sense of these linear functionals) converges, rather than necessarily converging to 0. In particular, if $\mathcal{G}$ is symmetric (so that $\mathcal{N}(\mathcal{G}^\top) = \mathcal{N}(\mathcal{G})$), this establishes Lyapunov stability of the null-space component and is consistent with the observed behavior that trajectories converge to (or drift along) the null space rather than the origin.

Full-rank pseudo-gradient: scalar modal reductions and long-run behavior

Assume now that $\mathcal{G}$ is full rank. Then $\mathcal{N}(\mathcal{G}) = \{0\}$, and the only equilibrium of (8) is $x = 0$. We study stability through modal projections.

Limit matrix and a first obstruction. From (9),

$$\lim_{t\to\infty} A(t) = A^* \triangleq \begin{bmatrix} \mathbf{0} & I \\ -\mathcal{G} & \mathbf{0} \end{bmatrix}.$$

The eigenvalues of A^* satisfy

$$\det(-\mathcal{G} - \lambda^2 I) = 0,$$

so λ^2 must be an eigenvalue of $-\mathcal{G}$. For $N \geq 2$, this generally does not yield eigenvalues with strictly negative real parts, so A^* is not Hurwitz in general; consequently, exponential stability for (9) cannot be inferred from the limit system alone.

Projection onto left eigenvectors of $\mathcal{G}$. Let $w \neq 0$ be a *left* eigenvector of $\mathcal{G}$ with eigenvalue $\lambda \in \mathbb{C}$, i.e.,

$$w^\top \mathcal{G} = \lambda w^\top.$$

Projecting (9) along $w^\top$ and defining $y_1(t) = w^\top x_1(t)$, $y_2(t) = w^\top x_2(t)$ gives the scalar system

$$\dot{y}_1(t) = y_2(t), \qquad \dot{y}_2(t) = -\lambda y_1(t) - \frac{3}{t} y_2(t), \tag{14}$$

which is equivalent to the second-order ODE

$$\ddot{y}(t) + \frac{3}{t}\dot{y}(t) + \lambda y(t) = 0, \qquad y(t) \triangleq y_1(t). \tag{15}$$

Closed-form solution and asymptotics. A standard reduction yields the (Bessel) representation

$$y(t) = \frac{c_1}{t} J_1\big(\sqrt{\lambda}\, t\big) + \frac{c_2}{t} Y_1\big(\sqrt{\lambda}\, t\big), \tag{16}$$

with constants c_1, c_2 determined by $(y(t_0), \dot{y}(t_0))$, and where $J_1(\cdot)$ and $Y_1(\cdot)$ are Bessel functions of the first and second kind (order 1). If λ is real and negative, then $\sqrt{\lambda}$ is imaginary and the same expression can be interpreted via the analytic continuation of Bessel functions; we keep the form (16) to emphasize the dependence on $\sqrt{\lambda}$.

For $|\sqrt{\lambda}\, t|$ large, the asymptotic expansions [1] take the form

$$J_1(\sqrt{\lambda} t) \sim \sqrt{\frac{2}{\pi\sqrt{\lambda} t}} \left(\cos\left(\sqrt{\lambda} t - \frac{3\pi}{4}\right) + e^{|\Im(\sqrt{\lambda} t)|}\, \mathcal{O}\Big(|\sqrt{\lambda} t|^{-1}\Big)\right),$$

$$Y_1(\sqrt{\lambda} t) \sim \sqrt{\frac{2}{\pi\sqrt{\lambda} t}} \left(\sin\left(\sqrt{\lambda} t - \frac{3\pi}{4}\right) + e^{|\Im(\sqrt{\lambda} t)|}\, \mathcal{O}\Big(|\sqrt{\lambda} t|^{-1}\Big)\right), \qquad |\arg \sqrt{\lambda}| < \pi,$$

where $\Im(\cdot)$ denotes the imaginary part. The factor $e^{|\Im(\sqrt{\lambda} t)|}$ becomes active when $\sqrt{\lambda}$ has nonzero imaginary part, converting the oscillatory behavior into exponential growth in t. Thus, we obtain the following theorem:

Lemma 2 (Asymptotic behavior of projected NAGD modes)**.** *Let $\mathcal{G} \in \mathbb{R}^{n\times n}$ and consider the NAGD game dynamics* (8) *on $[t_0, \infty)$ with $t_0 > 0$. Let $\lambda \in \mathbb{C}$ be a left eigenvalue of $\mathcal{G}$, i.e., there exists $w \neq 0$ such that $w^\top \mathcal{G} = \lambda w^\top$. Define the projected scalar trajectory $y(t) \triangleq w^\top x(t)$. Then y satisfies*

$$\ddot{y}(t) + \frac{3}{t}\dot{y}(t) + \lambda y(t) = 0, \tag{17}$$

and admits the Bessel-type representation (16)*. Moreover:*

1. ***Complex eigenvalue (generic exponential growth).*** *If $\Im(\sqrt{\lambda}) \neq 0$ (in particular, if $\lambda \notin \mathbb{R}_{\geq 0}$), then the large-$t$ asymptotics of the Bessel terms include the factor $e^{|\Im(\sqrt{\lambda})| t}$, and for generic initial conditions $(y(t_0), \dot{y}(t_0))$ the projected trajectory satisfies*

   $$|y(t)| \to \infty \quad \textit{exponentially as } t \to \infty,$$

   implying instability of the corresponding projected mode.

2. ***Positive real eigenvalue (polynomial decay).*** *If $\lambda > 0$ is real (so that $\sqrt{\lambda} \in \mathbb{R}$), then the oscillatory parts of the Bessel functions remain bounded and the prefactor in* (16) *yields decay. In particular,*

   $$\lim_{t\to\infty} y(t) = 0, \qquad \textit{and} \qquad y(t) = \mathcal{O}(t^{-3/2}) \quad \textit{as } t \to \infty.$$

Consequently, the full NAGD system inherits stable or unstable behavior based on the spectrum of $\mathcal{G}$ through its left eigenvalues: complex eigenvalues generate exponentially diverging projected trajectories, whereas strictly positive real eigenvalues yield decaying projected trajectories. We have the following theorem:

Theorem 1 (Spectral characterization via projected modes)**.** *Consider the NAGD game dynamics on $[t_0, \infty)$ with $t_0 > 0$,*

$$\ddot{x}(t) + \frac{3}{t}\dot{x}(t) + \mathcal{G}x(t) = 0, \tag{18}$$

equivalently, with $x_1 = x$ and $x_2 = \dot{x}$,

$$\frac{d}{dt}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} \mathbf{0} & I \\ -\mathcal{G} & -\frac{3}{t}I \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}. \tag{19}$$

Assume that $\mathcal{G}$ is normal *(in particular, this holds if $\mathcal{G} = \mathcal{G}^\top$), so that there exists an orthonormal basis $\{w_i\}_{i=1}^n$ of $\mathbb{R}^n$ consisting of (right/left) eigenvectors of $\mathcal{G}$ with eigenvalues $\{\lambda_i\}_{i=1}^n$, i.e., $\mathcal{G}w_i = \lambda_i w_i$ and $w_i^\top \mathcal{G} = \lambda_i w_i^\top$. Define the scalar modal projections*

$$y_i(t) \triangleq w_i^\top x_1(t), \qquad \dot{y}_i(t) = w_i^\top x_2(t), \qquad i = 1, \ldots, n.$$

Then each mode satisfies

$$\ddot{y}_i(t) + \frac{3}{t}\dot{y}_i(t) + \lambda_i y_i(t) = 0. \tag{20}$$

Moreover, the following hold.

1. ***Instability from non-real or negative spectrum.*** *If there exists i such that $\Im(\sqrt{\lambda_i}) \neq 0$ (in particular, if $\lambda_i \notin \mathbb{R}_{\geq 0}$), then the origin of* (19) *is not Lyapunov stable. Equivalently, there exist arbitrarily small initial conditions for which the corresponding solution has $|y_i(t)|$ (and hence $\|x_1(t)\|_2$) growing exponentially as $t \to \infty$.*

2. ***Lyapunov stability and convergence to the null space.*** *If $\lambda_i \in \mathbb{R}_{\geq 0}$ for all i, then the origin of* (19) *is Lyapunov stable. Furthermore:*

   (a) For each $\lambda_i > 0$, the associated mode satisfies

   $$y_i(t) \to 0 \quad \text{and} \quad y_i(t) = \mathcal{O}(t^{-3/2}) \qquad (t \to \infty).$$

   (b) For each $\lambda_i = 0$, the mode satisfies $\dot{y}_i(t) = \left(\frac{t_0}{t}\right)^3 \dot{y}_i(t_0)$ and hence

   $$\lim_{t\to\infty} y_i(t) = y_i(t_0) + \frac{t_0}{2}\dot{y}_i(t_0).$$

   Consequently, letting $\Pi_{\mathcal{N}(\mathcal{G})}$ denote the orthogonal projector onto $\mathcal{N}(\mathcal{G})$,

   $$x_2(t) \to 0, \qquad x_1(t) \to \Pi_{\mathcal{N}(\mathcal{G})}\left(x_1(t_0) + \frac{t_0}{2}x_2(t_0)\right), \qquad (t \to \infty).$$

   In particular, if $\mathcal{G} \succ 0$ (so $\mathcal{N}(\mathcal{G}) = \{0\}$), then $x_1(t) \to 0$ and $x_2(t) \to 0$.

Proof. Since $\mathcal{G}$ is normal, choose an orthonormal eigenbasis $\{w_i\}$ and decompose $x_1(t) = \sum_i y_i(t)w_i$. Projecting (18) onto each w_i yields the decoupled scalar dynamics (20). By Theorem 2, each mode either contains an exponentially growing component when $\Im(\sqrt{\lambda_i}) \neq 0$, or decays to 0 with rate $\mathcal{O}(t^{-3/2})$ when $\lambda_i > 0$, while the $\lambda_i = 0$ modes converge to constants given explicitly above. Because the basis is orthonormal and the dynamics decouple along these coordinates (as in Proposition 1 applied to the projected coordinates), Lyapunov stability/instability of the full system follows from the stability/instability of the scalar modes. The limit expression for $x_1(t)$ is obtained by reconstructing the trajectory from its $\lambda_i = 0$ components, which is precisely the orthogonal projection onto $\mathcal{N}(\mathcal{G})$. $\square$

Remarks on slowly-varying structure and Khalil's Lemma 5.15

The system (9) can be viewed as a linear time-varying (LTV) system with a time-varying damping coefficient $u(t) = 3/t$. Since $u(t)$ is bounded and continuously differentiable on $[t_0, \infty)$, one may be tempted to apply "frozen-time" intuition: analyze $\dot{x} = A(\alpha)x$ with α fixed, and then transfer conclusions to the time-varying case. However, the present dynamics exhibit a key obstruction: although $A(t) \to A^*$, the limiting matrix A^* is generally not Hurwitz, and even when $A(t)$ becomes Hurwitz for large t in certain projected coordinates, its spectral gap typically shrinks to 0 as $t \to \infty$ (real parts approach the imaginary axis). This prevents direct invocation of uniform exponential stability results.

To make this connection precise, we quote Lemma 5.15 from *Nonlinear Systems* by Khalil [2].

Lemma 3 (Khalil, Lemma 5.15)**.** *Consider the system $\dot{z} = A(\alpha)z$, where $\alpha \in \Gamma \subset \mathbb{R}^m$ and $A(\alpha)$ is continuously differentiable. Suppose the elements of A and their first partial derivatives with respect to α are uniformly bounded; that is,*

$$\|A(\alpha)\|_2 < c, \qquad \left\| \frac{\partial}{\partial \alpha_i} A(\alpha) \right\|_2 < b_i, \quad \forall \alpha \in \Gamma,\ \forall i = 1, \dots, m.$$

Suppose further that $A(\alpha)$ is Hurwitz uniformly in α; that is,

$$\Re[\lambda(A(\alpha))] < -\sigma < 0, \qquad \forall \alpha \in \Gamma.$$

Then the Lyapunov equation $P(\alpha)A(\alpha) + A^\top(\alpha)P(\alpha) = -I$ has a unique positive definite solution $P(\alpha)$ for every $\alpha \in \Gamma$ such that $V(z, \alpha) = z^\top P(\alpha)z$ ensures that the origin $z = 0$ is exponentially stable.

Remark 2 (Why Lemma 3 does not yield exponential stability here)**.** *In our setting, $A(t)$ in* (9) *is bounded for $t \geq t_0$, and $A(t) \to A^*$ as $t \to \infty$. One could set $\alpha = u(t) = 3/t \in (0, 3/t_0]$ and view A as a smooth map of α:*

$$A(\alpha) = \begin{bmatrix} \mathbf{0} & I \\ -\mathcal{G} & -\alpha I \end{bmatrix}.$$

The uniform Hurwitz condition required by Lemma 3 fails in general: even when $A(\alpha)$ is Hurwitz for each fixed $\alpha > 0$ in certain cases, the spectral margin typically collapses as $\alpha \downarrow 0$ (equivalently, as $t \to \infty$), so one cannot obtain a uniform $\sigma > 0$ on the set $\Gamma = (0, 3/t_0]$. This aligns with the explicit projected analysis (15)–(16)*: modes may be asymptotically decaying without being uniformly exponentially stable, and complex eigenvalues of $\mathcal{G}$ can induce exponential growth despite the presence of vanishing damping.*

Remark 3 (On alternative damping profiles)**.** *The special choice $u(t) = 3/t$ provides persistent (though vanishing) damping that can still drive decay in stable modes. By contrast, if one replaces $3/t$ with a function $u(t)$ such that $\int_{t_0}^{\infty} u(t)\,dt < \infty$, then the damping becomes "too short-lived" in an integral sense, and the long-run behavior approaches that of the undamped equation $\ddot{y} + \lambda y = 0$ in the projected dynamics, because the $\dot{y}$-term loses influence asymptotically.*

Simulated Examples

The previous section summarized the behavior for the dynamics of Nesterov's accelerated gradient descent as an ordinary differential equation when applied for 2 player quadratic games. Below are simulated trajectories on MATLAB via the Runge-Kutta method with step sizes of 0.01 and from $t \in [1, 1000]$. The blue and orange curves represent the position and momentum vectors, respectively.

Figure 3 and Figure 4 are trajectories for low rank $\mathcal{G}$. Clearly, the position vectors converge towards the null space of $\mathcal{G}$. This establishes Lyapunov stability for the momentum.

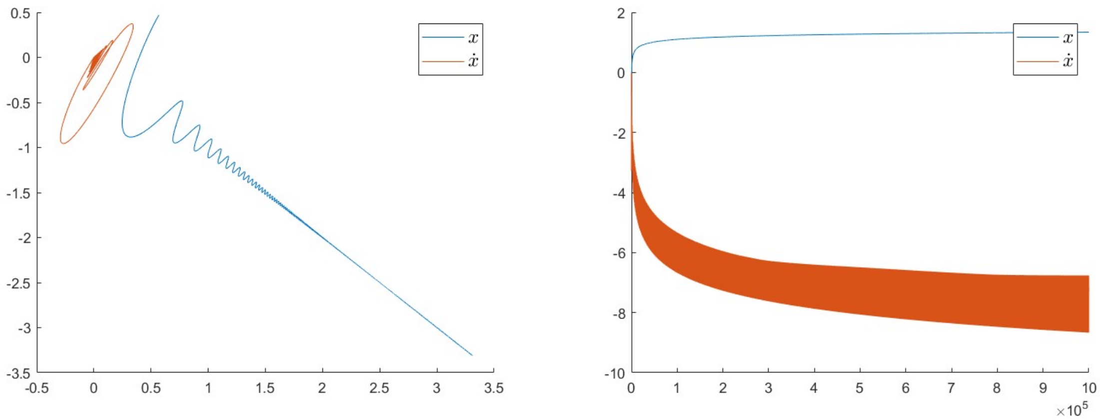


Figure 3: Nesterov dynamics with $\mathcal{G} = \begin{bmatrix} 1 & 1 \\ 2 & 2 \end{bmatrix}$ and log10-norm of trajectory

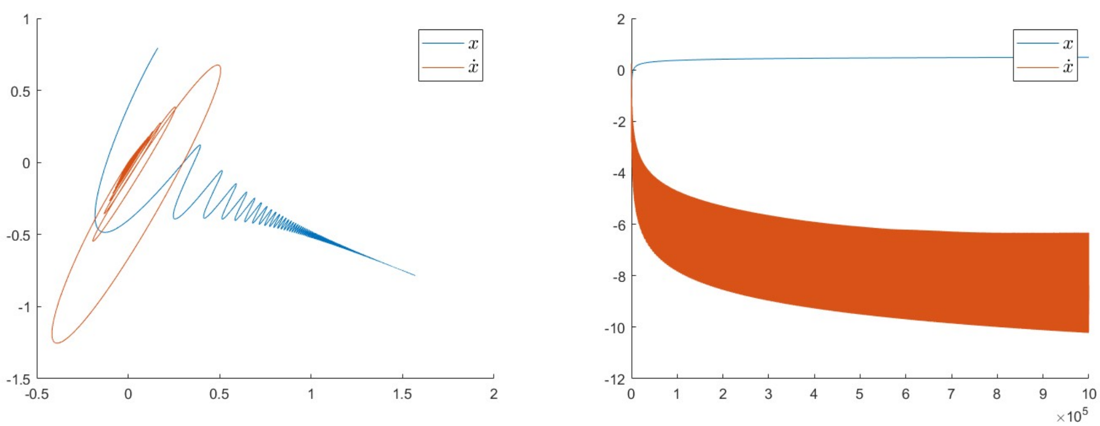


Figure 4: Nesterov dynamics with $\mathcal{G} = \begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix}$ and log10-norm of trajectory

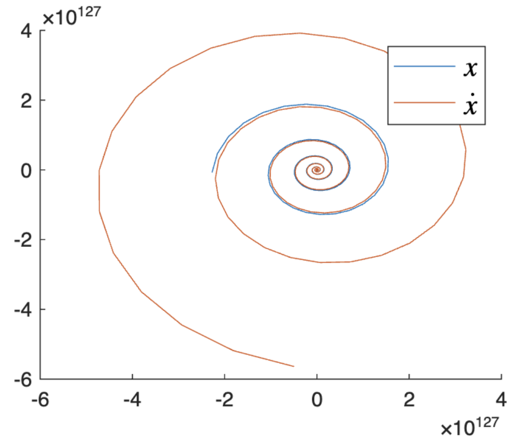


Figure 5: $\mathcal{G} = \begin{bmatrix} 6 & 1.5 \\ -1.5 & 6 \end{bmatrix}$ and $\lambda = 6 + 1.5i, 6 - 1.5i$

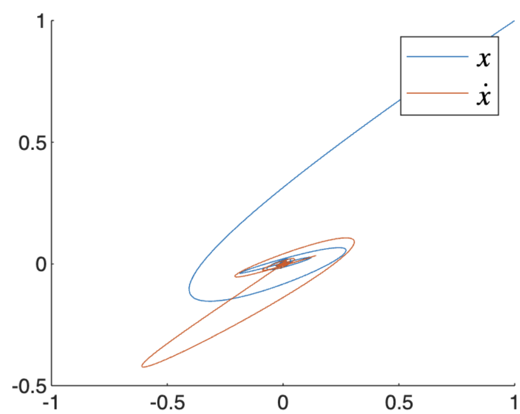


Figure 6: $\mathcal{G} = \begin{bmatrix} 0.6557 & 0.8491 \\ 0.0357 & 0.9340 \end{bmatrix}$ and $\lambda = 0.5720, 1.0178$

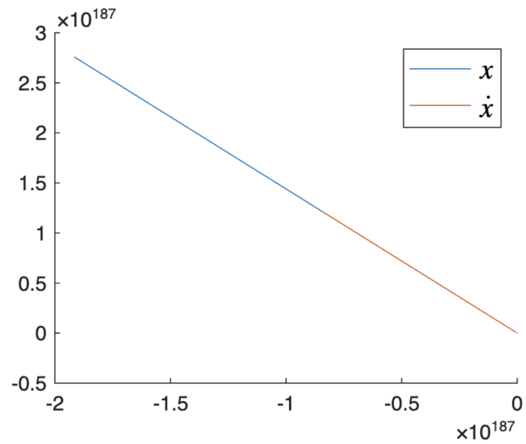
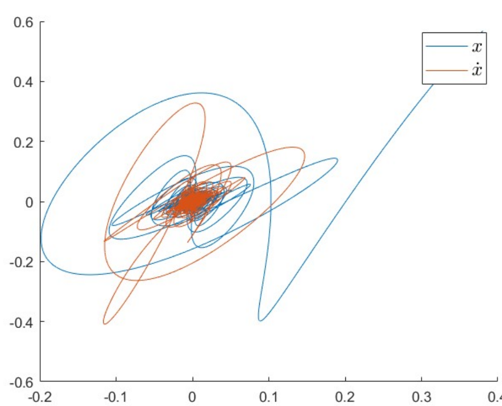


Figure 7: $\mathcal{G} = \begin{bmatrix} 0.9572 & 0.8003 \\ 0.4854 & 0.1419 \end{bmatrix}$ and $\lambda = 1.2942, -0.1952$



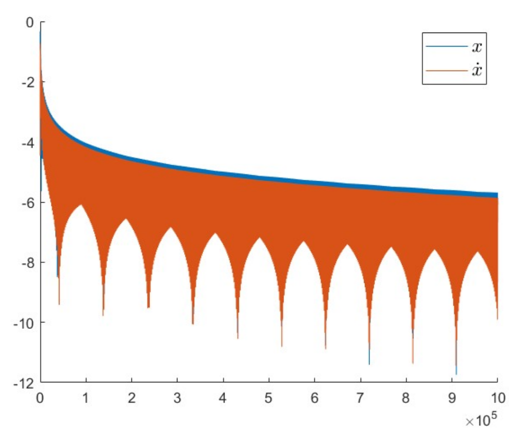


Figure 8: Nesterov dynamics with $\mathcal{G} = \begin{bmatrix} 0.4 & 0.2 \\ 0.2 & 0.8 \end{bmatrix}$ and log10-norm of trajectory

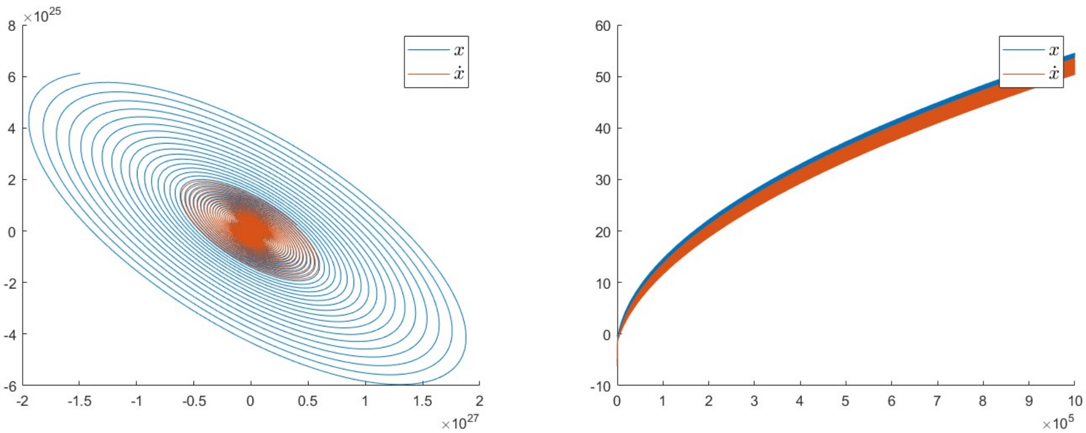


Figure 9: Nesterov dynamics with $\mathcal{G} = \begin{bmatrix} 0.1 & -0.1 \\ 0 & 0.1 \end{bmatrix}$ and log10-norm of trajectory

References

[1] Milton Abramowitz and Irene A. Stegun. *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables.* Dover, New York, ninth dover printing, tenth gpo printing edition, 1964.

[2] Hassan K Khalil. *Nonlinear systems; 3rd ed.* Prentice-Hall, Upper Saddle River, NJ, 2002. The book can be consulted by contacting: PH-AID: Wallet, Lionel.