Analysis of the Heavy-ball Algorithm using Integral Quadratic Constraints

Apurva Badithela[1] and Peter Seiler[2]

Abstract— **In this paper, we analyze the convergence rate of the Heavy-ball algorithm applied to optimize a class of continuously differentiable functions. The analysis is performed with the Heavy-ball tuned to achieve the best convergence rate on the sub-class of quadratic functions. We review recent work to characterize convergence rate upper bounds for optimization algorithms using integral quadratic constraints (IQC). This yields a linear matrix inequality (LMI) condition which is typically solved numerically to obtain convergence rate bounds. We construct an analytical solution for this LMI condition using a specific "weighted off-by-one" IQC. We also construct a specific objective function such that the Heavy-ball algorithm enters a limit cycle. These results demonstrate that IQC condition is tight for the analysis of the tuned Heavy-ball, i.e. it yields the exact condition ratio that separates global convergence from non-global convergence for the algorithm.**

I. INTRODUCTION

This paper focuses on convergence rate analysis of first-order algorithms for solving convex optimizations. The objective function is assumed to be continuously differentiable and m-strongly convex with L-Lipschitz gradients. There are many first-order algorithms in the literature including gradient descent [1], [2], Nesterov's method [3], Heavy-ball method [4] and the recently developed triple momentum method [5]. Convergence rate analysis determines asymptotic convergence rate bounds for optimization algorithms. This paper discusses convergence rate analysis for the Heavy-ball algorithm tuned for quadratics.

Our approach is motivated by recent work in [6] on analysis of optimization algorithms using integral quadratic constraints (IQCs). IQCs were originally introduced by Yakubovich [7] for control law analysis. Megretski and Rantzer gave a unified framework to incorporate IQCs for various types of nonlinearities and uncertaintities [8]. The framework, reviewed in Section II, was adapted in [6] to compute convergence rate upper bounds. The approach represents the optimization algorithm as a linear dynamical system in feedback with the gradient of the objective function. The conditions for convergence rate upper bounds are specified as linear matrix inequalities (LMIs). Numerical solutions of this LMI condition indicate that Heavy-ball tuned for quadratic functions is not globally convergent for condition ratios $\kappa := \frac{L}{m}$ greater than ≈ 18. A specific function with $\kappa = 25$ is provided in [6] for which the tuned Heavy-ball enters a limit cycle.

The main contribution of our work is to precisely characterize the condition ratio κ_S that separates global convergence from non-global convergence for the tuned Heavy-ball algorithm. This precise characterization consists of two aspects. First, we provide an analytical solution to the LMI condition in [6] using the weighted off-by-1 IQC (See Section III-A). This analytical solution is constructed using a related Riccati equation form of the LMI condition. This approach might be useful for other, related analyses. The analytical LMI solution provides an upper bound on the convergence rate of $\rho = 1$ for the condition ratio $\kappa_S = 9+5\sqrt{14}$. Second, we construct a specific objective function with condition ratio k_S that causes the tuned Heavy-ball to limit cycle, i.e. it does not globally converge (See Section III-B).

The analysis in our paper draws from related work in the controls literature [9]. This related work uses Zames-Falb IQCs to demonstrate global stability of a discrete-time system in feedback with a static, nonlinearity. They also construct counter-example nonlinearities that cause the feedback system to limit cycle and hence prove instability. The construction of the nonlinearity in Section III-B draws inspiration from the prior work in [9].

II. BACKGROUND

A. Heavy-ball Algorithm

In this work, we consider the following convex optimization involving $f : \mathbb{R} \to \mathbb{R}$,

$$\min_x f(x) \tag{1}$$

This paper focuses on functions of a single variable. However, multivariable functions $f : \mathbb{R}^n \to \mathbb{R}$ can be handled with minor modifications using the symmetry / dimension reduction argument in [6]. The function f is assumed to be continuously differentiable, m-strongly convex with gradients that are L-Lipschitz. We denote the set of such functions as $S(m, L)$ [2]. The condition ratio of any $f \in S(m, L)$ is defined as $\kappa := \frac{L}{m}$. The optimization in (1) has a unique global minimum x^* if f is strongly convex with $m > 0$. Assume, without loss of generality by a coordinate shift, that the minimum occurs at $x^* = 0$. This assumption simplifies the notation in the remainder of the paper.

There are many algorithms to solve such optimizations including gradient descent [1], Heavy-ball [4], Nesterov's method [3], and the recently developed triple-momentum

This work was supported by the National Science Foundation under Grant No. NSF-CMMI-1254129 entitled CAREER: Probabilistic Tools for High Reliability and Monitoring and Control of Wind Farms

[1]A. Badithela is a graduate student at California Institute of Technology, Pasadena, USA apurva@caltech.edu

[2]P. Seiler is with Faculty of Aerospace Engineering and Mechanics, University of Minnesota, Twin-Cities, USA seile017@umn.edu

method [5]. This paper focuses on the convergence properties of the Heavy-ball algorithm. The iterates for this algorithm are computed as:

$$x_{k+1} = x_k - \alpha \nabla f(x_k) + \beta(x_k - x_{k-1}), \tag{2}$$

where α and β are parameters that are held constant as the iterates are updated. The parameters are chosen to the values that provide the optimal convergence rate for the class of convex quadratic functions in $S(m, L)$ [10], [6]. These tuned values for parameters α and β are given in Section II-D.

Most first-order optimization algorithms can be equivalently written in the form of a discrete-time, state-space system in feedback with the gradient of f as noted in [6]. The Heavy-ball method can be written as:

$$\begin{aligned} \eta_{k+1} &= A\eta_k + Bu_k, \\ y_k &= C\eta_k, \end{aligned} \tag{3}$$

$$u_k = \nabla f(y_k), \tag{4}$$

where $\eta_k = [x_k, x_{k-1}]^T$ is the state of the linear system and the state matrices (A,B,C) are given as follows:

$$A := \begin{bmatrix} 1+\beta & -\beta \\ 1 & 0 \end{bmatrix}, B := \begin{bmatrix} -\alpha \\ 0 \end{bmatrix}, C := \begin{bmatrix} 1 \\ 0 \end{bmatrix}^T. \tag{5}$$

Let G denote the linear system in (3). This representation for the first-order optimization is a feedback interconnection of G and the gradient as shown in Figure 1. Figure 1 also shows an additional system Ψ. This filter Ψ is used in the analysis as described further in the next section.

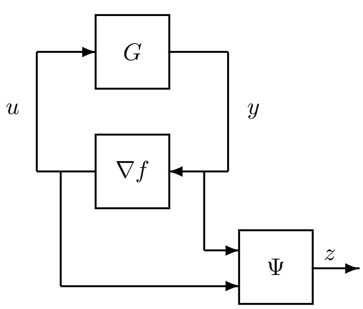

Fig. 1. Block diagram representation of a first-order algorithm given by (3). The algorithm consists of the linear system G in feedback with the gradient function ∇f. The auxiliary system Ψ, defined in Section II-B, is appended for the analysis.

B. Integral Quadratic Constraints

Figure 1 is a feedback system involving an LTI system and a nonlinear function. This section briefly reviews a class of input/output constraints satisfied by the nonlinear function. Specifically, the class of Integral Quadratic Constraints (IQCs), introduced in [7], [8], can be used to analyze feedback interconnections as shown in Figure 1. These constraints have roots tracing back to the classical absolute stability problem [11]. This section reviews the variation of ρ-hard IQCs introduced in [6] for assessing convergence rates of optimization algorithms.

Consider a static, memoryless function $\phi : \mathbb{R} \to \mathbb{R}$. This function defines a mapping from a sequence $\{y_0, y_1, \dots,\}$ to $\{u_0, u_1, \dots,\}$ by:

$$u_k = \phi(y_k) \tag{6}$$

The following discrete-time linear system Ψ can be used to construct an auxiliary sequence, z, from the signals y and u:

$$\begin{aligned} \zeta_{k+1} &= A_\Psi \zeta_k + B_\Psi^y y_k + B_\Psi^u u_k, \\ z_k &= C_\Psi \zeta_k + D_\Psi^y y_k + D_\Psi^u u_k. \\ \zeta_0 &= 0 \end{aligned} \tag{7}$$

The dimension of z is n_z so that Ψ has dimension has dimension $n_z \times 2$. The class of constraints on ϕ is given by the auxiliary system Ψ and a matrix $M \in \mathbb{R}^{n_z \times n_z}$ as formally defined below.

Definition 1: Let $\rho > 0$ be given. The function $\phi : \mathbb{R} \to \mathbb{R}$ satisfies the *ρ-hard IQC* defined by (Ψ, M) if the following constraint holds for all $N \geq 0$ and for all $y \in \ell_{2e}$ and $u = \phi(y)$:

$$\sum_{k=0}^{N} \rho^{-2k} z_k^T M z_k \geq 0, \tag{8}$$

We use the notation $\phi \in IQC(\Psi, M, \rho)$ if ϕ satisfies this constraint.

IQCs can also be defined for more general operators. However, the definition for static, memoryless functions will be sufficient here. In particular, the paper uses the following two ρ-hard IQCs derived in [6], [12].

Lemma 1 (Sector): Assume $\phi = \nabla f$ where $f \in S(m, L)$ is given. Then ϕ satisfies the ρ-hard IQC defined by:

$$\Psi := \begin{bmatrix} L & -1 \\ -m & 1 \end{bmatrix} \text{ and } M := \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \tag{9}$$

This is known as the sector ρ-hard IQC and Ψ is static, i.e. a matrix, for this constraint. This simply represents that the function ϕ lies between the sector given by lines of slope m and L. The next constraint involves a dynamic auxiliary system Ψ.

Lemma 2 (Weighted Off-by-1): Assume $\phi = \nabla f$ where $f \in S(m, L)$ is given. Then for any $h_1 \in [0, \rho^2]$, ϕ satisfies the ρ-hard IQC defined by

$$(A_\Psi, B_\Psi, C_\Psi, D_\Psi) = \left(0, \begin{bmatrix} -L & 1 \end{bmatrix}, \begin{bmatrix} h_1 \\ 0 \end{bmatrix}, \begin{bmatrix} L & -1 \\ -m & 1 \end{bmatrix}\right)$$
$$\text{and } M := \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

This is known as the weighted off-by-1 ρ-hard IQC. It is actually a class of IQCs parameterized by the free variable h_1. The dynamics Ψ represent a finite impulse response (FIR) filter with one time-step. This constraint captures the slope conditions that must hold between data at one timestep (u_k, y_k) and the previous timestep (u_{k-1}, y_{k-1}). If $h_1 = 0$ then the dynamics of Ψ are unobservable and the constraint reduces to the sector IQC, i.e. this form of the weighted off-by-1 includes the sector constraint. This ρ-hard IQC is part of the more general class of discrete-time Zames-Falb multipliers [11]. Additional details can be found in [12]. This related work includes ρ-hard constraints where Ψ has a higher order FIR filter. This captures slope conditions across multiple time steps. There has also been recent work on the use of anti-causal filters for ρ-hard IQCs [13].

C. LMI Condition for Convergence

IQCs can be used to estimate an upper bound on the asymptotic convergence rate of first-order optimization algorithms [6]. The analysis is based on the augmented feedback system shown in Figure 1. Combine the dynamics of G (3) and Ψ (7) to construct an LTI system with input u, output z, and state $\hat{x} := \left[\eta^T, \zeta^T\right]^T$. The state matrices for the combined dynamics of G and Ψ are:

$$\hat{A} = \begin{bmatrix} A & 0 \\ B_\Psi^y C & A_\Psi \end{bmatrix} \qquad \hat{B} = \begin{bmatrix} B \\ B_\Psi^u \end{bmatrix}$$
$$\hat{C} = \begin{bmatrix} D_\Psi^y C & C_\Psi \end{bmatrix} \qquad \hat{D} = \begin{bmatrix} D_\Psi^u \end{bmatrix}$$

The convergence rate condition in the next theorem is stated with these combined dynamics and an IQC satisfied by ∇f.

Theorem 1 ([6]): Assume $\nabla f > 0$ satisfies the ρ-hard IQC defined by (Ψ, M, ρ). Assume there exists $P > 0$ such that the following LMI is feasible:

$$\begin{bmatrix} \hat{A}^T P \hat{A} & \hat{A}^T P \hat{B} \\ \hat{B}^T P \hat{A} & \hat{B}^T P \hat{B} \end{bmatrix} - \begin{bmatrix} \rho^2 P & 0 \\ 0 & 0 \end{bmatrix} + \begin{bmatrix} \hat{C}^T \\ \hat{D}^T \end{bmatrix} M \begin{bmatrix} \hat{C} & \hat{D} \end{bmatrix} \leq 0, \tag{10}$$

Then the dynamics of the optimization algorithm (3) and (4) initialized from any η_o satisfy

$$||\eta_k|| \leq \sqrt{cond(P)}\rho^k ||\eta_o|| \forall k \tag{11}$$

where $cond(P)$ is the condition ratio of P.

Proof: Define the Lyapunov-like function of the combined system as $V(\hat{x}) := \hat{x}^T P \hat{x}$. Multiply the LMI condition in (10) on the left and right by $\left[\hat{x}^T, u\right]$ and its transpose. This yields:

$$V(\hat{x}_{k+1}) - \rho^2 V(\hat{x}_k) + z_k^T M z_k \leq 0 \tag{12}$$

This Lyapunov-type condition along with the ρ-hard IQC can be used to show that $V(\hat{x}_k) \leq \rho^{2k} V(\hat{x}_0)$. This is used to bound the convergence rate of η. Details are given in [6]. ■

Note that the analysis condition only depends on the combined LTI dynamics and the ρ-hard IQC. In other words, it does not explicitly depend on the nonlinearity ∇f. The ρ-hard IQC constrains the signals (u, y) and hence implicitly constrains the input/output behavior of ∇f. Also note that (10) is an LMI in P for fixed $\rho > 0$. In general, the ρ-hard IQC itself depends on ρ. For example, the weighted off-by-1 IQC has another decision variable h_1 subject to the constraint $0 \leq h_1 \leq \rho^2$. The best (smallest) upperbound ρ on the convergence rate is obtained via bisection with LMI feasibility problems for each fixed ρ.

If $\rho \geq 1$ then the upper bound on the Heavy-ball iterates grows geometrically. In this case, the iterates need not converge to the optimal value at $x^* = 0$. We will still use the term "convergence rate" for such cases with $\rho \geq 1$.

D. Heavy-ball Convergence Rate

This section summarizes the Heavy-ball analysis given in [6] using the convergence rate condition in Theorem 1. First, recall Nesterov's theoretical convergence rate lower bound [3]:

$$\rho_{LB} := \frac{\sqrt{\kappa} - 1}{\sqrt{\kappa} + 1} \tag{13}$$

This is a lower bound on the convergence rate achieved by any first-order method on the class of functions in $S(m, L)$. The Heavy-ball parameters (α, β) can be tuned to achieve this rate on the sub-class of quadratic functions in $S(m, L)$. These optimal parameters are given by:

$$\begin{aligned} \alpha_0 &:= \frac{4}{(\sqrt{L} + \sqrt{m})^2} \\ \beta_0 &:= \left(\frac{\sqrt{\kappa} - 1}{\sqrt{\kappa} + 1} \right)^2, \end{aligned} \tag{14}$$

where κ is the condition ratio of function f used in (2). Heavy-ball with (α_0, β_0), achieves Nesterov's lower bound (13) on the class of quadratic functions in $S(m, L)$. However, it is not globally convergent for all functions in $S(m, L)$. For example, Lessard, et al. [6] provide a (non-quadratic) function $f \in S(1, 25)$ for which Heavy-ball with (α_0, β_0) produces a (non-decaying) limit cycle from appropriate initial conditions. This provides a specific lower bound on the rate ρ: Heavy Ball with (α_0, β_0) has rate $\rho \geq 1$ for the class of functions $S(1, 25)$.

Lessard, et. al also use the IQC framework to compute upper bounds on the convergence rate for Heavy-ball with (α_0, β_0). Figure 2 shows the upper bounds on the convergence rate computed with both the sector IQC (Lemma 1) and weighted off-by-1 IQC (Lemma 2). This figure also shows Nesterov's lower bound. All bounds are shown as a function of the condition ratio $\kappa = \frac{L}{m}$. Note that the upper bound curve for the weighted off-by-1 IQC crosses the bound $\rho = 1$ at a condition ratio $\kappa \approx 18$. This condition ratio serves as a stability boundary for Heavy Ball with (α_0, β_0). Specifically, the algorithm is globally convergent for κ below ≈ 18. The next section precisely characterizes this stability boundary.

III. RESULTS

A. Analytical Bound on Convergence Rate

This section provides an analytical solution to the convergence rate LMI (Theorem 1) using the weighted off-by-1 IQC. The weighted off-by-1 IQC in Lemma 2 has a single state and the Heavy-ball algorithm has two states. Thus the augmented linear system $(\hat{A}, \hat{B}, \hat{C}, \hat{D})$ in Section II-C has three states. The convergence rate LMI was solved numerically with bisection to obtain the best (smallest) upper bound on the rate ρ. These numerical solutions indicate

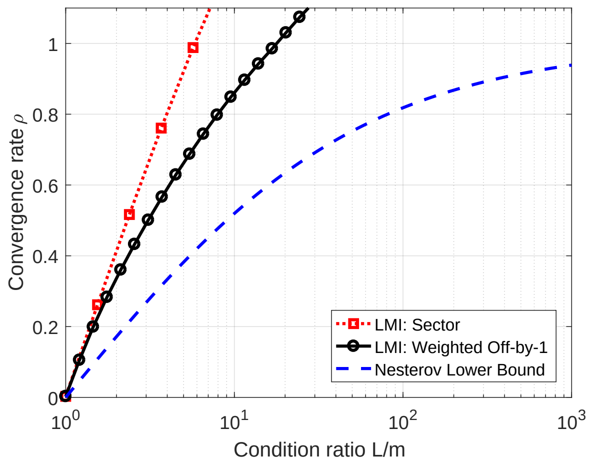

Fig. 2. Rate bounds for various algorithms. Two curves represent IQC-derived rate upper bounds for the quadratically optimized Heavy-ball algorithm. The blue dashed curve represents Nesterov's theoretical lower bound for any algorithm given by (3) and any $f \in S(m, L)$.

that the optimal convergence rate and storage matrix have following dependencies for large condition ratios:

$$\rho(\kappa) := \frac{\sqrt{\kappa}-1}{\sqrt{\kappa}+1}\rho_\infty\,, \tag{15}$$

$$P(\kappa) = \begin{bmatrix} \frac{c_{11}}{\alpha_0^2} & \frac{c_{12}\sqrt{\beta_0}}{\alpha_0^2} & \frac{c_{13}}{\alpha_0} \\ (\cdot) & \frac{c_{22}\beta_0}{\alpha_0^2} & \frac{c_{23}\sqrt{\beta_0}}{\alpha_0} \\ (\cdot) & (\cdot) & c_{33} \end{bmatrix}, \tag{16}$$

where ρ_∞ and c_{ij} are constants (independent of the condition ratio) to be determined. The dependence of (α_0, β_0) on κ has not been denoted for simplicity. Entries of P that can be inferred from symmetry are also omitted. The numerical solutions also indicate that the optimal weighted off-by-1 IQC is given with $h_1 = \rho$ for large condition ratios. The dependences given in Equations (15) and (16) were obtained by examining the numerical LMI solutions for large condition ratios. For example, it was noted that the (3,3) entry of $P(\kappa)$ was constant (to within numerical errors). It was also noted that the (1,3) entry of $P(\kappa)$ was inversely proportional to α_0. There was some trial-and-error involved to obtain these particular dependencies.

The next step is to determine analytical expressions for the unknown constants. For this step, first define the following matrices related to the weighted off-by-1 IQC:

$$\begin{bmatrix} Q & S \\ S^T & R \end{bmatrix} := \begin{bmatrix} \hat{C}^T \\ \hat{D}^T \end{bmatrix} M \begin{bmatrix} \hat{C} & \hat{D} \end{bmatrix} \tag{17}$$

Next note that the LMI condition has the following related Riccati equation form:

$$\begin{aligned} 0 = &\hat{A}^T P \hat{A} - \rho^2 P + Q \\ &- (\hat{A}^T P \hat{B} + S)(\hat{B}^T P \hat{B} + R)^{-1}(\hat{A}^T P \hat{B} + S)^T \end{aligned}$$

Symbolic solvers can be used to obtain the expressions for the remaining unknowns c_{ij} and ρ_∞ from this Riccati equation. This yields the following expressions:

$$\rho_\infty := \frac{\sqrt{\kappa_S}+1}{\sqrt{\kappa_S}-1} \text{ where } \kappa_S := 9 + 5\sqrt{14} \tag{18}$$

$$\begin{bmatrix} c_{11} & c_{12} & c_{13} \\ (\cdot) & c_{22} & c_{23} \\ (\cdot) & (\cdot) & c_{33} \end{bmatrix} = \begin{bmatrix} 3\sqrt{5} & 1+2\sqrt{5} & \frac{1+3\sqrt{5}}{2} \\ (\cdot) & 4+3\sqrt{5} & \frac{7+\sqrt{5}}{2} \\ (\cdot) & (\cdot) & \rho_\infty \end{bmatrix} \tag{19}$$

This provides a complete analytical solution (P, h_1, ρ) to the convergence rate LMI with the weighted off-by-1 IQC for large condition ratios. It can be verified that this is a solution to the Riccati equation. Moreover, the analytical expressions for (P, h_1, ρ) satisfy the convergence rate LMI in Theorem 1 using the weighted off-by-1 IQC.

This solution is valid only for condition ratios greater than or equal to $\kappa_S \approx 17.94$. In addition, $\rho(\kappa_S) = 1$. Thus κ_S is the condition ratio that defines the boundary for global convergence for the Heavy-ball with (α_0, β_0).

B. Construction of Limit Cycle

This section constructs a function f with condition ratio $\kappa_S = 9+5\sqrt{14}$ for which Heavy Ball (α_0, β_0) enters a non-trivial limit cycle. Our construction is similar to that used by Carrasco et al. in [9]. Specifically, we seek iterates for the Heavy-ball algorithm (x_1, x_2, x_3) such that Heavy-ball sustains a limit cycle. Carrasco et al. [9] use a saturation function as the non-linearity in their system. For the Heavy-ball algorithm the non-linearity is ∇f. Gradient functions of the following type have been used by other researchers [10], [6] to find limit cycles:

$$\nabla f(x) = \begin{cases} Lx + c_1 & x < a \\ mx + a(L-m) + c_1 & a \le x \le b \\ Lx + (a-b)(L-m) + c_1 & x \ge b \end{cases}, \tag{20}$$

where $\kappa := L/m$ is the condition ratio and (a, b, c_1) are constants to be determined.

If the weighted off-by-1 IQC provides a tight rate upper bound, we should be able to find a f with condition ratio $\kappa_S = 9 + 5\sqrt{14}$ whose Heavy-ball iterates result in a limit-cycle. We first simplify the gradient function ∇f in (20). By appropriate scaling, we can assume $m = 1$ so that $\kappa = L$. We further assume $a = 1$ so that the gradient is linear for $x < a$. This choice implies that $c_1 = 0$ because $\nabla f(0) = 0$ by our standing assumption that the global minima occurs at $x^* = 0$. These choices yield the following partially defined gradient function:

$$\nabla f(x) = \begin{cases} \kappa x & x < 1 \\ x + (L-1) & 1 \le x < b \\ Lx + (1-b)(L-1) & x \ge b \end{cases}, \tag{21}$$

This function has slope of L for $x < 1$ and $x \ge b$ but (smaller) slope of m in the middle interval. We assume the first two iterates "hop" across the middle interval: $x_1 := b$ and $x_2 = 1$. This leaves three unknowns for the condition ratio κ, constant $b > 1$ and iterate x_3. The limit cycle (if it exists) must satisfy the Heavy-ball update relation (2).

This yields the following three simultaneous equations for the candidate limit cycle:

$$\begin{aligned} x_3 &= x_2 - \alpha_0 L x_2 + \beta_0(x_2 - x_1) \\ x_1 &= x_3 - \alpha_0 L x_3 + \beta_0(x_3 - x_3) \\ x_2 &= x_1 - \alpha_0[L x_1 + (1-b)(L-1)] + \beta_0(x_1 - x_3), \end{aligned} \quad (22)$$

where (α_0, β_0) are given in (14). Symbolically solving these equations yields the following:

$$\begin{aligned} L = \kappa_S &= 4\sqrt{5} + 9 \approx 17.94 \\ b = x_1 &= 3\sqrt{5}/10 + 3/2 \approx 2.17 \\ x_3 &= -7\sqrt{5}/10 - 1/2 \approx -2.07 \end{aligned}$$

Figure 3 illustrates the function f that gives a 3-limit cycle for the Heavy-ball algorithm. Note that an even simpler function with $\nabla f(x) = 1$ (instead of L) $\forall x \geq 1$ will have the same limit cycle. Figure 4 gives the iteration history for the Heavy-ball algorithm.

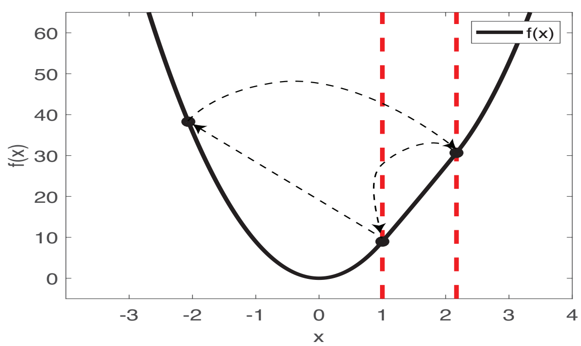

Fig. 3. Graph of a function whose gradient is given by (21). The Heavy-ball iterates cycle between 1, 2.17 and -2.07. The slope of the gradient is κ_S in the intervals (-inf, 1) and [b, inf), and $m = 1$) in the interval [1, b), where $b := 2.17$. The red dashed lines separate these intervals.

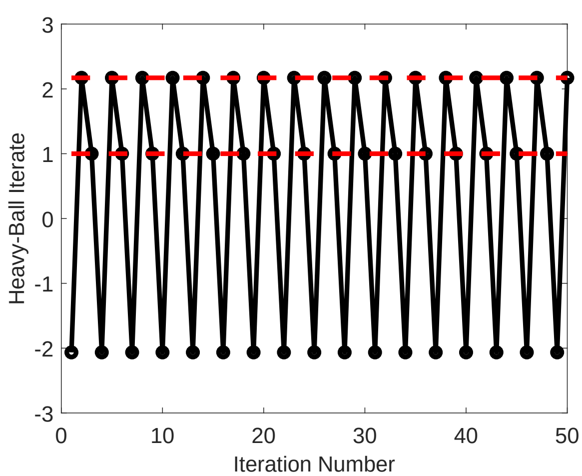

Fig. 4. Heavy-ball iterates for $f(x)$ forming a 3-limit cycle.

IV. CONCLUSIONS

In this work, we analyzed the Heavy-ball algorithm with parameters tuned for the sub-class of quadratic functions. The analysis is performed using LMI conditions and ρ-hard IQCs as introduced by Lessard, et al. Numerical studies of these LMI conditions indicate that the tuned Heavy-ball algorithm is globally convergent for functions with "small" condition ratios. An analytical solution to the LMI condition using a weighted off-by-1 IQC yields the condition ratio $\kappa_S := 9+5\sqrt{14}$ as the boundary between global convergence and non-convergence. We have also constructed a specific function with condition ratio equal to κ_S for which the tuned Heavy-ball enters into a limit cycle. This indicates that the IQC condition provides a tight characterization of the stability boundary for the tuned Heavy-ball. For future work, it would be interesting to study if there exists a general form for functions $\in S(m, L)$ with condition ratios larger than κ_S for which the weighted-off-by-1 provides a tight convergence rate upperbound.

REFERENCES

[1] D. P. Bertsekas, *Nonlinear Programming*. Athena Scientific, second ed., 1999.

[2] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.

[3] Y. Nesterov, *Introductory lectures on convex optimization: A basic course*. Springer, 2013.

[4] B. T. Polyak, "Some methods of speeding up the convergence of iteration methods," *USSR Computational Mathematics and Mathematical Physics*, vol. 4, no. 5, pp. 1–17, 1964.

[5] B. Van Scoy, R. A. Freeman, and K. M. Lynch, "The fastest known globally convergent first-order method for minimizing strongly convex functions," *IEEE Control Systems Letters*, vol. 2, no. 1, pp. 49–54, 2018.

[6] L. Lessard, B. Recht, and A. Packard, "Analysis and design of optimization algorithms via integral quadratic constraints," *SIAM Journal on Optimization*, vol. 26, no. 1, pp. 57–95, 2016.

[7] V. A. Yakubovich, "S–procedure in nonlinear control theory," *Vestnik Leningrad Univ.*, pp. 62–77, 1971, (English translation in Vestnik Leningrad Univ. Math., vol. 4, pp. 73-93, 1977).

[8] A. Megretski and A. Rantzer, "System analysis via integral quadratic constraints," *IEEE Transactions on Automatic Control*, vol. 42, no. 6, pp. 819–830, 1997.

[9] J. Carrasco, W. Heath, and M. de la Sen, "Second-order counterexample to the discrete-time Kalman conjecture," in *European Control Conference*, 2015, pp. 981–985.

[10] E. Ghadimi, H. Feyzmahdavian, and M. Johansson, "Global convergence of the Heavy-ball method for convex optimization," in *European Control Conference*, 2015, pp. 310–315.

[11] J. Carrasco, M. C. Turner, and W. P. Heath, "Zames–Falb multipliers for absolute stability: From O'Shea's contribution to convex searches," *European Journal of Control*, vol. 28, pp. 1–19, 2016.

[12] R. Boczar, L. Lessard, A. Packard, and B. Recht, "Exponential stability analysis via integral quadratic constraints," *submitted to IEEE Transactions on Automatic Control, arXiv preprint arXiv:1706.01337*, 2017.

[13] R. Freeman, "Noncausal Zames–Falb multipliers for tighter estimates of exponential convergence rates," in *American Control Conference*, 2018, pp. 2984–2989.