Optimization Algorithms as Uncertain Graded Dynamical Systems

Bryan Van Scoy

***Abstract*— The interpretation of iterative optimization algorithms as dynamical systems has led to a variety of advances in their analysis and design using tools from control. In this paper, we identify a structure of dynamical systems that arises naturally in a variety of optimization algorithms, and we show how to take advantage of this structure in system analysis. In particular, first-order optimization algorithms consist of the gradient of the objective in feedback with *graded* dynamical systems, which are systems whose signal spaces decompose as direct sums that are not mixed by the system dynamics.**

I. Introduction

Optimization problems are prevelant throughout science and engineering [1], with iterative algorithms used to construct numerical solutions [2]. Recent work in the controls community has interpreted such algorithms as dynamical systems in order to exploit the variety of tools from control theory [3] to study optimization algorithms [4]–[6].

The interpretation of optimization algorithms as dynamical systems has led to the application of tools from control to systematically analyze their convergence properties [4], [5]. These analysis techniques make use of integral quadratic constraints [7] and dissipativity theory [8]–[11] from robust control, and have provided novel insights into algorithm behavior, such as robustness to noisy gradient evaluations [12]. Beyond analysis, tools from control can also be used to synthesize novel optimization algorithms [13], [14].

In addition to analysis and synthesis results, the interpretation of optimization algorithms has provided insight into the structure of algorithms. A first-order method, for instance, can be modeled as a dynamical system in feedback with the gradient of the objective function, where the dynamical system must contain an integrator so that fixed points of the system correspond to first-order stationary points of the optimization problem [5]. Other examples of structure include the particular plant structures in algorithm synthesis [13], and separation of optimization and consensus components in distributed optimization [15].

In this paper, we describe a general system structure posessed by various iterative optimization algorithms. As we illustrate, a wide variety of optimization algorithms has the form of a dynamical system in feedback with an uncertainty (such as the objective function, its gradient, a projection onto the constraint set, etc.). Moreover, the iterates of the system separate into distinct subspaces that are not mixed by the system dynamics, but only by the uncertainty. Examples of optimization algorithms with such structure include first-order methods that explicitly model the evolution of the function values, primal-dual algorithms for linearly-constrained optimization, and distributed optimization algorithms.

*This material is based upon work supported by the National Science Foundation under Grant Number 2347121.

Bryan Van Scoy is with Department of Electrical and Computer Engineering, Miami University, Oxford, OH 45056, USA. bvanscoy@miamioh.edu

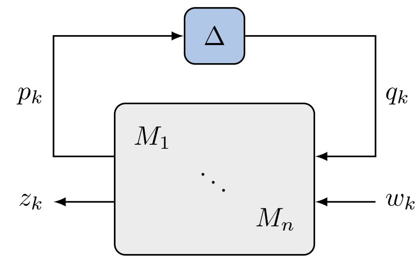

Fig. 1. Block diagram of an uncertain dynamical system consisting of a block-diagonal plant M with exogneous input w and exogenous output z in feedback with an uncertainty Δ.

Main contributions. Our main contributions are as follows:

1) We define the notion of a *graded dynamical system*, which is a system whose state space separates as a direct sum of subspaces and whose dynamics do not mix iterates between these subspaces. This notion generalizes the block-diagonal system structure in Fig. 1.
2) We motivate this system structure by illustrating how it naturally arises in the modeling of optimization algorithms as dynamical systems; optimization algorithms are graded dynamical systems in feedback with uncertain and/or nonlinear components.
3) We then show how to leverage the graded structure in the analysis of such systems. We consider Lyapunov analysis of graded autonomous systems, dissipativity analysis for graded systems with inputs and outputs, and integral quadratic constraints for graded systems with uncertainty.

The block diagram in Figure 1 illustrates the structure of optimization algorithms, where a dynamical system M is in feedback with an uncertainty Δ. Here, the plant is block diagonal, which is a special case of being graded. In general, graded systems need not be diagonal, but do not mix subspaces of their inputs and outputs.

Notation. In block diagrams, shaded gray blocks indicate dynamical systems, while shaded blue blocks represent uncertainties. The column space, row space, and null space of a matrix are denoted $\mathrm{col}(\cdot)$, $\mathrm{row}(\cdot)$, and $\mathrm{null}(\cdot)$, respectively. The symbol $\oplus$ represents the direct sum of two vector spaces. The sets of real and natural numbers are denoted $\mathbb{R}$ and $\mathbb{N}$. We denote the transfer function of a discrete-time linear time-invariant (LTI) dynamical system G by $\hat{G}(z)$.

II. Graded dynamical systems

We first introduce the novel concept of a graded dynamical system. Informally, a graded dynamical system is one in which the state, input, and output spaces each decompose into disjoint subspaces that the system dynamics do not mix. The block-diagonal system in Fig. 1 is graded, for instance. However, graded systems need not be block diagonal.

A. Preliminaries

Before describing graded systems, we first describe the notions of a graded vector space and a graded map; see [19, Section 2.1.1] and [20, Section 1.5.1] for reference.

A vector space X is *graded* if it has a decomposition as a direct sum of vector spaces. In particular, let $\mathcal{I}$ be an index set, and let X^i for $i \in \mathcal{I}$ be a set of subspaces whose direct sum is the whole space:

$$X = \bigoplus_{i\in\mathcal{I}} X^i. \tag{1}$$

This implies that each element $x \in X$ has a unique decomposition over these subspaces:

$$x = \sum_{i\in\mathcal{I}} x^i \quad \text{where} \quad x^i \in X^i. \tag{2}$$

Since there is a bijection (invertible mapping) between an element x and its subspace decomposition $\{x^i\}$ for $i \in \mathcal{I}$, we use the two representations interchangeably. Elements of a subspace X^i are called *homogeneous* of grade i.

A linear map $A : X \to X$ is *graded* with respect to the grading on X if it preserves the grading of homogeneous elements, meaning that

$$A(X^i) \subseteq X^i \quad \text{for all } i \in \mathcal{I}, \tag{3}$$

where $A(X^i)$ is the image of X^i under the linear map A.

B. Autonomous systems

Consider the autonomous linear time-invariant discrete-time dynamical system described by the difference equation

$$x_{k+1} = Ax_k \tag{4}$$

for each iteration k in the natural numbers $\mathbb{N} = \{0, 1, 2, \ldots\}$. At each iteration $k \in \mathbb{N}$, the state x_k is in the state space X, and the linear map $A : X \to X$ is the state transition map.

Definition 1: A *graded autonomous dynamical system* is a system (4) in which the state space X is a graded vector space and the state transition map A is a graded map with respect to this grading.

We now formalize the concept that graded dynamical systems do not mix subspaces. Let x_k denote a trajectory of the system, and let the homogeneous elements $x_k^i \in X^i$ denote its corresponding subspace decomposition. By linearity of the state transition map, the next state is

$$\sum_{i\in\mathcal{I}} x_{k+1}^i = x_{k+1} = Ax_k = A\sum_{i\in\mathcal{I}} x_k^i = \sum_{i\in\mathcal{I}} Ax_k^i. \tag{5}$$

Since the subspace decomposition is unique, this implies that each homogeneous element of the state also satisfies the system dynamics,

$$x_{k+1}^i = Ax_k^i \quad \text{for all } i \in \mathcal{I}.$$

Moreover, this implies that each subspace is invariant along the system dynamics, so any state that starts in a subspace X^i remains in the subspace forward in time.

C. Systems with inputs and outputs

We now extend the notion of a graded dynamical system to systems with inputs and outputs.

Consider the discrete-time linear time-invariant dynamical system that is described by the iterations

$$x_{k+1} = Ax_k + Bu_k, \tag{6a}$$
$$y_k = Cx_k + Du_k. \tag{6b}$$

For each iteration $k \in \mathbb{N}$, the state x_k is in the state space X, the input u_k is in the input space U, and the output y_k is in the output space Y. The system dynamics are then described by the linear maps

$$A : X \to X, \qquad B : U \to X,$$
$$C : X \to Y, \qquad D : U \to Y.$$

Definition 2: A *graded input-output dynamical system* is a system (6) in which the state space X, input space U, and output space Y are all graded with respect to the same index set $\mathcal{I}$, and the state-space maps are graded in that

$$A(X^i) + B(U^i) \subseteq X^i,$$
$$C(X^i) + D(U^i) \subseteq Y^i.$$

For graded dynamical systems, there exists an index set $\mathcal{I}$ and subspaces X^i, U^i, and Y^i for $i \in \mathcal{I}$ whose direct sum is the whole state, input, and output space, respectively,

$$X = \bigoplus_{i\in\mathcal{I}} X^i, \qquad U = \bigoplus_{i\in\mathcal{I}} U^i, \qquad Y = \bigoplus_{i\in\mathcal{I}} Y^i.$$

By definition of the direct sum, each state $x \in X$, input $u \in U$, and output $y \in Y$ has a unique decomposition over these subspaces,

$$x = \sum_{i\in\mathcal{I}} x^i, \qquad u = \sum_{i\in\mathcal{I}} u^i, \qquad y = \sum_{i\in\mathcal{I}} y^i,$$

where $x^i \in X^i$, $u^i \in U^i$, and $y^i \in Y^i$ for each subspace index $i \in \mathcal{I}$. Since there is a bijection between a state x and its subspace decomposition $\{x_i\}_{i\in\mathcal{I}}$ (and likewise for inputs and ouputs), we use the two representations interchangeably.

With this notion of a graded dynamical system, we are now able to state our first main result.

Theorem 1 (Graded dynamical system): The iterates of a graded dynamical system satisfy the system dynamics on each subspace:

$$x_{k+1}^i = Ax_k^i + Bu_k^i$$
$$y_k^i = Cx_k^i + Du_k^i$$

for all iterations $k \in \mathbb{N}$ and all subspace indices $i \in \mathcal{I}$, where the iterates

$$x_k^i \in X^i, \qquad u_k^i \in U^i, \qquad y_k^i \in Y^i$$

are the homogeneous elements in the subspace decompositions of the state $x_k \in X$, input $u_k \in U$, and output $y_k \in Y$, respectively.

This result implies that each subspace is invariant along the system dynamics, meaning that any state that starts in a subspace X^i remains in the subspace forward in time when an input in the corresponding subspace U^i is applied.

Proof: Let (x_k, u_k, y_k) for $k \in \mathbb{N}$ denote a trajectory of the system, and let the homogeneous elements $(x_k^i, u_k^i, y_k^i) \in X^i \times U^i \times Y^i$ denote its subspace decomposition. By linearity of the state-space maps (and similar to (5)), the next state is

$$\sum_{i \in \mathcal{I}} x_{k+1}^i = \sum_{i \in \mathcal{I}} (A x_k^i + B u_k^i)$$

and the output is

$$\sum_{i \in \mathcal{I}} y_k^i = \sum_{i \in \mathcal{I}} (C x_k^i + D u_k^i).$$

Since the subspace decomposition is unique, this implies the homogeneous elements also satisfy the system dynamics. ■

III. Motivating examples

We now motivate the notion of a graded dynamical system by illustrating how this structure naturally arises in a variety of optimization algorithms.

A. Function values

The iterates of an optimization algorithm are often in the domain of the objective function. Consider minimizing a real-valued function f, and let $x_k \in \operatorname{dom} f$ for $k \in \mathbb{N}$ denote the iterates of the algorithm. Associated with these iterates are the corresponding function values, $f_k = f(x_k) \in \mathbb{R}$. While the iterates are in the domain of the objective function, the function values are real numbers. We now show how including both the iterates and the function values leads to a graded dynamical system.

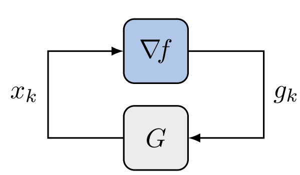

Fig. 2. Block diagram of a gradient-based optimization algorithm.

Consider the gradient-based iterative optimization algorithm described by the block diagram in Figure 2. When the dynamical system G is linear and time invariant with transfer function $\hat{G}(z) = -\alpha/(z-1)I$, for instance, this describes the gradient descent algorithm,

$$x_{k+1} = x_k - \alpha g_k \qquad \text{where} \qquad g_k = \nabla f(x_k),$$

with stepsize $\alpha \in \mathbb{R}$. This intepretation of the system enables the use of tools from robust control, such as dissipativity theory and integral quadratic constraints, to analyze its convergence properties. However, this representation does not take into account the function values of the iterates. One may want to include the function values as a measure of performance (such as the optimality gap), or the function values may enter the analysis through the interpolation conditions for the function class [16]. The combined dynamics of the iterates x_k and function values f_k are described by the block diagram in Figure 3.

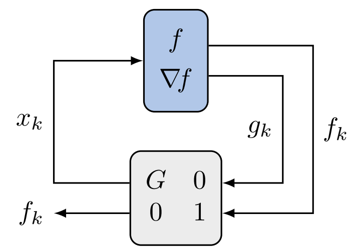

Fig. 3. Block diagram of a gradient-based optimization algorithm that includes the function values of iterates.

This system has the structure of a block-diagonal system in feedback with the operator $\Delta = \begin{bmatrix} f \\ \nabla f \end{bmatrix}$, which is the same as that in Figure 1. In particular, suppose the system G has a state-space realization with state space X, input space U, and output space Y. Then the diagonal system $\left[\begin{smallmatrix} G & 0 \\ 0 & 1 \end{smallmatrix}\right]$ is graded with respect to the spaces $X \oplus \mathbb{R}$, $U \oplus \mathbb{R}$, and $Y \oplus \mathbb{R}$.

B. Primal-dual algorithms

While the dynamical system in the previous example was block diagonal as in Fig. 1, graded dynamical systems generalize this structure as illustrated by our next two examples.

Consider the linearly-constrained optimization problem

$$\text{minimize} \quad f(x) \tag{7a}$$
$$\text{subject to} \quad Ax = b \tag{7b}$$

with decision variable $x \in \mathbb{R}^n$, objective $f : \mathbb{R}^n \to \mathbb{R}$, constraint matrix $A \in \mathbb{R}^{m \times n}$, and constraint vector $b \in \mathbb{R}^m$. One method to solve this problem is the primal-dual algorithm represented by the block diagram in Figure 4.

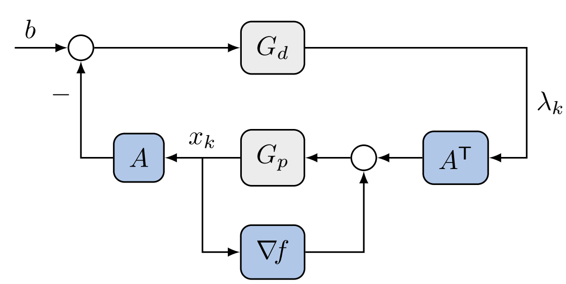

Fig. 4. Block diagram of a primal-dual algorithm, where G_p and G_d are dynamical systems that represent the primal and dual dynamics, respectively.

As a concrete example, suppose the primal and dual dynamical systems are LTI with transfer matrices

$$\hat{G}_p(z) = \frac{-\alpha}{z-1} I \qquad \text{and} \qquad \hat{G}_d(z) = \frac{\beta}{z-1} I,$$

where α and β are the primal and dual stepsizes, respectively. A state-space representation of this algorithm is then

$$x_{k+1} = x_k - \alpha\left(\nabla f(x_k) + A^\mathsf{T}\lambda_k\right), \tag{8a}$$
$$\lambda_{k+1} = \lambda_k + \beta\left(b - Ax_k\right), \tag{8b}$$

where $x_k \in \mathbb{R}^n$ and $\lambda_k \in \mathbb{R}^m$ are the primal and dual states. The structure of the algorithm is such that all fixed points of the system (8) satisfy the first-order optimality conditions for the optimization problem (7).

The primal-dual algorithm in Figure 4 has the form of a dynamical system in feedback with an operator Δ as in Figure 1, except that the plant is not block diagonal. In particular, the plant and uncertainty are

$$M = \begin{bmatrix} 0 & G_p & G_p & 0 \\ -G_d & 0 & 0 & G_d \end{bmatrix} \quad \text{and} \quad \Delta = \begin{bmatrix} A & 0 \\ 0 & A^\mathsf{T} \\ \nabla f & 0 \end{bmatrix}.$$

Here, we take the constraint vector as the exogeneous input, $w = b$, and the performance channel is omitted. The exogeneous signal spaces are then $W = \mathbb{R}^m$ and $Z = \emptyset$. The uncertainty channel is

$$p = \begin{bmatrix} x \\ \lambda \end{bmatrix} \quad \text{and} \quad q = \begin{bmatrix} Ax \\ A^\mathsf{T}\lambda \\ \nabla f(x) \end{bmatrix},$$

where the uncertainty signal spaces are

$$P = \mathbb{R}^n \times \mathbb{R}^m \qquad \text{and} \qquad Q = \mathbb{R}^m \times \mathbb{R}^n \times \mathbb{R}^n.$$

A fundamental result in linear algebra is that, for any matrix $A \in \mathbb{R}^{m\times n}$, the vector spaces $\mathbb{R}^n$ and $\mathbb{R}^m$ have the following decompositions as direct sums:

$$\mathbb{R}^n = \text{row}(A) \oplus \text{null}(A),$$
$$\mathbb{R}^m = \text{col}(A) \oplus \text{null}(A^\mathsf{T}).$$

Therefore, each of the uncertainty signal spaces has the structure of a direct sum,

$$P = \underbrace{\text{row}(A) \times \text{col}(A)}_{P^1} \oplus \underbrace{\text{null}(A) \times \text{null}(A^\mathsf{T})}_{P^2}$$

and

$$Q = \underbrace{\text{col}(A) \times \text{row}(A) \times \text{row}(A)}_{Q^1} \oplus \underbrace{\text{null}(A^\mathsf{T}) \times \text{null}(A) \times \text{null}(A)}_{Q^2}.$$

The performance channel spaces have similarly structure. Moreover, if the primal and dual dynamics are diagonal so that $G_p = g_p I$ and $G_d = g_d I$ (as in the example primal-dual algorithm in (8)), then the system dynamics do not mix these subspaces in that, if the input is in Q^i, the output will be in P^i for $i \in \{1, 2\}$. In other words, the system M is graded with respect to the input grading $P = P^1 \oplus P^2$ and output grading $Q = Q^1 \oplus Q^2$.

C. Distributed optimization algorithms

Consider a network of n agents, where each agent $j \in \{1, \ldots, n\}$ has a local objective function $f_j : \mathbb{R}^d \to \mathbb{R}$ and local decision variable $x_j \in \mathbb{R}^d$, and the agents cooperate to solve the distributed optimization problem

$$\text{minimize} \quad \sum_{j=1}^{n} f_j(x_j), \tag{9a}$$
$$\text{subject to} \quad x_1 = x_2 = \ldots = x_n. \tag{9b}$$

The goal is to minimize the objective (optimization) subject to agreement on the solution (consensus). It has been shown that every distributed optimization algorithm (from within a broad class of algorithms) has the form in Figure 5, where the consensus and optimization dynamics are separated [15]. Here, L denotes the graph Laplacian, and ∇f denotes the gradient of the global objective function.

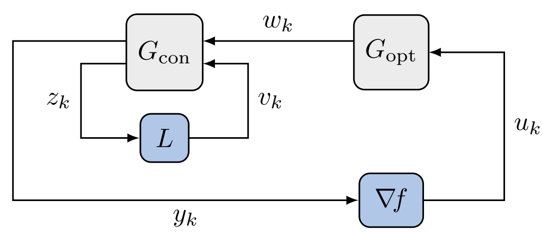

Fig. 5. Block diagram representation of a distributed optimization algorithm, where G_{opt} and G_{con} are dynamical systems that represent the optimization and consensus dynamics, respectively.

As in the previous subsection, the signal spaces decompose as direct sums. For instance, the signal spaces for the input and output of the Laplacian are

$$Z = \mathbb{R}^n = \text{row}(L) \oplus \text{null}(L),$$
$$V = \mathbb{R}^n = \text{col}(L) \oplus \text{null}(L^\mathsf{T}),$$

and these subspaces are not mixed by the system when the consensus dynamics have the form $G_{\text{con}} = g_{\text{con}} I$.

D. Discussion

The previous examples illustrate that a variety of optimization algorithms can be modeled as graded dynamical systems in feedback with other components (such as the gradient of the objective function, the constraint matrix, or the graph Laplacian). Similar results apply to a variety of optimization algorithms, such as the alternating direction method of multipliers (ADMM) [17] and proximal methods [18], among others.

IV. System analysis

We now turn to the analysis of dynamical systems with the graded structure. We first use Lyapunov analysis to study graded autonomous systems and show that the search for a Lyapunov function separates along each grade, which is expected since the dynamics do not mix subspaces. We then apply dissipativity theory to graded systems with exogeneous inputs and outputs. And finally, since optimization algorithms also have uncertainties, we use integral quadratic constraints to study uncertain graded systems.

A. Preliminaries

For each iteration $k \in \mathbb{N}$, let P_k and Q_k be functions that map the state space X to the real numbers $\mathbb{R}$. For the analysis, we will characterize the convergence properties of these two performance measures.

Given a functional V, we use the inequality $V \geq 0$ to denote that $V(x) \geq 0$ for all x in the domain of V. Similarly, $V \leq 0$ denotes that $-V \geq 0$. Given a set of functionals $V^i : X^i \to \mathbb{R}$ for each grade $i \in \mathcal{I}$, we let $V : X \to \mathbb{R}$ denote the functional

$$V(x) = \sum_{i \in \mathcal{I}} V^i(x^i),$$

where x^i is the homogeneous element in X^i for each grade $i \in \mathcal{I}$ in the decomposition (2).

B. Lyapunov analysis

Consider the autonomous system in (4), let P_k and Q_k be two performance measures that map X to $\mathbb{R}$, and suppose that the system is graded. We then have the following convergence result which states that the search for a Lyapunov function separates along each grade.

Theorem 2 (Lyapunov analysis): If, for each iteration $k \in \mathbb{N}$ and grade $i \in \mathcal{I}$, there exists a functional $V_k^i : X \to \mathbb{R}$ such that, for all states $x \in X^i$,

$$\begin{aligned} P_k^i(x) - V_k^i(x) &\leq 0, \quad \text{(positivity condition)} \\ V_{k+1}^i(Ax) - V_k^i(x) + Q_k^i(x) &\leq 0, \quad \text{(decrease condition)} \end{aligned}$$

then the state trajectory satisfies the performance bound

$$P_k(x_k) + \sum_{\ell=0}^{k-1} Q_\ell(x_\ell) \leq V_0(x_0). \tag{10}$$

Proof: Let (x_k) for $k \in \mathbb{N}$ be a trajectory of the system, and let x_k^i for $i \in \mathcal{I}$ denote its corresponding decomposition as in (2). Evaluating the positivity condition at x_k^i yields the inequality

$$P_k^i(x_k^i) - V_k^i(x_k^i) \leq 0.$$

Since the iterates satisfy the system dynamics on each subspace by Theorem 1, we have that $x_{k+1}^i = Ax_k^i$. Therefore, evaluating the decrease condition at x_k^i yields the inequality

$$V_{k+1}^i(x_{k+1}^i) - V_k^i(x_k^i) + Q_k^i(x_k^i) \leq 0.$$

Summing each of these inequalities over the grade $i \in \mathcal{I}$, we obtain the cumulative positivity and decrease conditions,

$$\begin{aligned} P_k(x_k) - V_k(x_k) &\leq 0, \\ V_{k+1}(x_{k+1}) - V_k(x_k) + Q_k(x_k) &\leq 0. \end{aligned}$$

Summing the second inequality over $\ell \in \{0, \ldots, k-1\}$ yields the telescoping sum

$$V_k(x_k) - V_0(x_0) + \sum_{\ell=0}^{k-1} Q_\ell(x_\ell) \leq 0,$$

then applying the first inequality yields the bound. ■

For the performance bound to be meaningful, $V_0(x_0)$ must be finite. This is the case if there exists a constant $c > 0$ such that $V_0(x) \leq c\,\|x\|$ for all $x \in X$. We now describe several particular cases of this result.

- Setting $Q \equiv 0$ yields an upper bound on the performance P of the last iterate,

$$P_k(x_k) \leq V_0(x_0).$$

- Setting $P \equiv 0$ yields an upper bound on the cumulative performance Q,

$$\sum_{\ell=0}^{k-1} Q_\ell(x_\ell) \leq V_0(x_0).$$

  Moreover, bounding each term in the summation by its minimum value yields the bound on the performance Q of the best iterate,

$$\underset{0 \leq \ell \leq k-1}{\text{minimum}} \; Q_\ell(x_\ell) \leq \frac{V_0(x_0)}{k}.$$

  If the performance measure is monotonically decreasing along system trajectories, then this also implies the sublinear bound on the performance of the last iterate,

$$Q_k(x_k) \leq \frac{V_0(x_0)}{k+1}.$$

In the special case when the functional V and the performance measures P and Q are the time-invariant quadratics

$$V(x) = x^\mathsf{T} \bar{V} x, \qquad P(x) = x^\mathsf{T} \bar{P} x, \qquad Q = x^\mathsf{T} \bar{Q} x,$$

we obtain the following corollary.

Corollary 1: If, for each grade $i \in \mathcal{I}$,

$$\begin{aligned} S_i^\mathsf{T}(\bar{P} - \bar{V})S_i &\preceq 0, \\ S_i^\mathsf{T}(A^\mathsf{T}\bar{V}A - \bar{V} + \bar{Q})S_i &\leq 0, \end{aligned}$$

where S_i is a matrix whose columns span subspace X_i, then the state trajectory of (4) satisfies the performance bound

$$x_k^\mathsf{T} \bar{P} x_k + \sum_{\ell=0}^{k-1} x_\ell^\mathsf{T} \bar{Q} x_\ell \leq x_0^\mathsf{T} \bar{V} x_0. \tag{11}$$

C. Dissipativity analysis

We now extend the Lyapunov convergence results from the previous section to systems with inputs and outputs.

Consider the dynamical system (6). Suppose that the system is graded and the input and output satisfy a known set of constraints

$$M(y, u) \geq 0 \qquad \text{for all } y \in Y \text{ and } u \in U \tag{12}$$

for all functions $M : Y \times U \to \mathbb{R}$ in some set $\mathcal{M}$. For instance, the system is dissipative if $Y = U$ is an inner product space and $\langle y_k, u_k \rangle \geq 0$ for all iterations $k \in \mathbb{N}$. Moreover, suppose the constraint decomposes over the subspaces as

$$M(y, u) = \sum_{i \in \mathcal{I}} M^i(y^i, u^i),$$

where $y^i \in Y^i$ and $u^i \in U^i$ are the homogeneous elements associated with the input u and output y, respectively.

We can then use this constraint to refine the search for a Lyapunov function. Doing so leads to the following result.

Theorem 3: If, for each iteration $k \in \mathbb{N}$ and grade $i \in \mathcal{I}$, there exists a functional $V_k^i : X \to \mathbb{R}$ and $M_k \in \mathcal{M}$ such that, for all $(x, u, y) \in X^i \times U^i \times Y^i$,

$$P_k^i(x) - V_k^i(x) + M_k^i(y, u) \leq 0,$$
$$V_{k+1}^i(Ax + Bu) - V_k^i(x) + Q_k^i(x) + M_k^i(y, u) \leq 0,$$

then the state trajectory satisfies the bound (10).

Proof: The proof is similar to that of Theorem 2, where the constraint (12) is used to obtain the positivity and decrease conditions. ■

Unlike Theorem 2, the conditions on the functionals V_k^i in Theorem 3 may be coupled due to the condition that $M_k = \sum_{i \in \mathcal{I}} M_k^i \in \mathcal{M}$.

D. Integral quadratic constraints

As we have seen in Section III, many optimization algorithms are graded dynamical systems in feedback with an operator. We now illustrate how to analyze such systems using integral quadratic constraints (IQCs) from robust control.

Consider the dynamical system described by the block diagram in Figure 6, in which the plant M is in feedback with an uncertainty Δ.

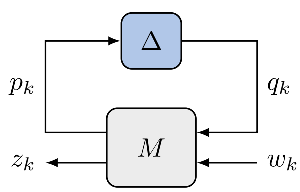

Fig. 6. Uncertain dynamical system consisting of a plant M with exogneous input w and exogenous output z in feedback with an uncertainty Δ.

The main idea behind IQCs is to replace the uncertainty with constraints on its filtered input and output, as illustrated in Figure 7.

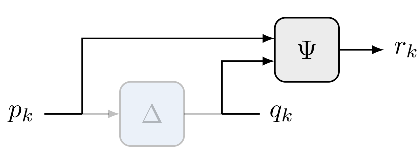

Fig. 7. Filtering the input p and output q of the uncertainty Δ through a filter Ψ to produce the signal r. The uncertainty is opaque, as it is replaced in the analysis by constraints on the output of the filter.

Filtering the input and output of the uncertainty Δ in the feedback loop in Figure 6 by a filter Ψ as in Figure 7 results in the closed-loop system from q to r.

If the plant G and filter Ψ are compatibly graded, then this closed-loop system is also graded. We can then use the results from the previous subsection to search for a Lyapunov function using the constraint on the output of the filter.

V. Conclusion

The interpretation of optimization algorithms as dynamical systems enables the breadth of tools from controls to be applied for their analysis and design. These tools can be strengthened by taking into account all available structure. In this paper, we introduced the notion of a graded dynamical system and showed that a variety of optimization algorithms can be modeled as such systems in feedback with other components. Moreover, we described how to leverage this structure in algorithm analysis. An interesting avenue for future work is to specialize techniques for algorithm synthesis (for example, [13]) to graded dynamical systems.

References

[1] S. Boyd and L. Vandenberghe, *Convex Optimization*. New York, NY, USA: Cambridge University Press, 2004.
[2] D. P. Bertsekas, *Convex Optimization Algorithms*. Athena Scientific, 2015.
[3] K. Zhou, J. Doyle, and K. Glover, *Robust and Optimal Control*. Prentice Hall, 1996.
[4] L. Lessard, B. Recht, and A. Packard, "Analysis and design of optimization algorithms via integral quadratic constraints," *SIAM J. Optim.*, vol. 26, no. 1, pp. 57–95, 2016.
[5] S. Michalowsky, C. Scherer, and C. Ebenbauer, "Robust and structure exploiting optimisation algorithms: an integral quadratic constraint approach," *Int. J. Contr.*, vol. 94, no. 11, pp. 2956–2979, 2020.
[6] B. Van Scoy and L. Lessard, "A tutorial on a Lyapunov-based approach to the analysis of iterative optimization algorithms," in *IEEE Conference on Decision and Control*, 2023. [Online]. Available: https://arxiv.org/abs/2309.11377
[7] A. Megretski and A. Rantzer, "System analysis via integral quadratic constraints," *IEEE Trans. Automat. Contr.*, vol. 42, no. 6, pp. 819–830, 1997.
[8] J. C. Willems, "Dissipative dynamical systems—Part I: General theory; Part II: Linear systems with quadratic supply rates," *Archive for Rational Mechanics and Analysis*, vol. 45, no. 5, pp. 321–393, 1972.
[9] M. Arcak, C. Meissen, and A. Packard, *Networks of Dissipative Systems: Compositional Certification of Stability, Performance, and Safety*, 1st ed. Springer Publishing Company, Incorporated, 2016.
[10] P. Seiler, "Stability analysis with dissipation inequalities and integral quadratic constraints," *IEEE Trans. Automat. Contr.*, vol. 60, no. 6, pp. 1704–1709, 2015.
[11] B. Hu and L. Lessard, "Dissipativity theory for Nesterov's accelerated method," in *Int. Conf. Mach. Learn.*, vol. 70, 2017, pp. 1549–1557.
[12] H. Mohammadi, M. Razaviyayn, and M. R. Jovanović, "Robustness of accelerated first-order algorithms for strongly convex optimization problems," *IEEE Trans. Automat. Contr.*, vol. 66, no. 6, pp. 2480–2495, 2021.
[13] C. W. Scherer, C. Ebenbauer, and T. Holicki, "Optimization algorithm synthesis based on integral quadratic constraints: A tutorial," in *IEEE Conference on Decision and Control*, 2023, pp. 2995–3002.
[14] B. Van Scoy, R. A. Freeman, and K. M. Lynch, "The fastest known globally convergent first-order method for minimizing strongly convex functions," *IEEE Contr. Syst. Lett.*, vol. 2, no. 1, pp. 49–54, 2018.
[15] B. Van Scoy and L. Lessard, "A universal decomposition for distributed optimization algorithms," *IEEE Control Systems Letters*, vol. 6, pp. 3044–3049, 2022.
[16] A. B. Taylor, J. M. Hendrickx, and F. Glineur, "Smooth strongly convex interpolation and exact worst-case performance of first-order methods," *Math. Program.*, vol. 161, no. 1, pp. 307–345, 1 2017.
[17] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, *Distributed Optimization and Statistical Learning via the Alternating Direction Method of Multipliers*. Foundations and Trends in Machine Learning, 2010, vol. 3.
[18] N. Parikh and S. Boyd, *Proximal Algorithms*. Foundations and Trends in Optimization, 2013, vol. 1, no. 3.
[19] M. Fairon, "Introduction to graded geometry," *European Journal of Mathematics*, vol. 3, no. 2, pp. 208–222, June 2017.
[20] J. Loday and B. Vallette, *Algebraic Operads*, ser. Grundlehren der mathematischen Wissenschaften. Springer Berlin Heidelberg, 2012.