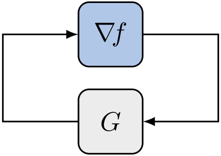
Optimization Algorithms as Uncertain Graded Dynamical Systems

Bryan Van Scoy
Miami University

Background

- Iterative first-order methods can be interpreted as dynamical systems in feedback with the gradient.



- Can use tools from robust control to analyze convergence properties.

(Bhaya, Kaszkurewicz, 2006), (Lessard, Recht, Packard, 2016), (Taylor, Van Scoy, Lessard, 2018), (Michalowsky, Scherer, Ebenbauer, 2021), (Van Scoy, Lessard, 2023), (Wu, Petersen, Ugrinovskii, Shames, 2024), (Mahammadi, Razaviyayn, Jovanović, 2025), . . .

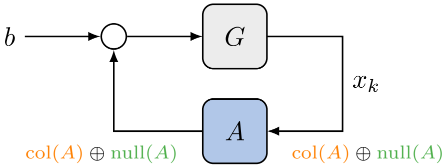
Motivation

Consider minimizing an n-dimensional quadratic function

$$f(x) = \frac{1}{2}x^\mathsf{T}Ax + b^\mathsf{T}x$$

using gradient descent

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



Observations:

- each signal space decomposes as $\mathbb{R}^n = \text{col}(A) \oplus \text{null}(A)$
- both G and A preserve each subspace

Contributions:

- identify a simple structure that arises when modeling optimization algorithms as dynamical systems
- provide some examples
- show how to exploit the structure for analysis

Preliminaries

- The sum of two vector spaces is

$$U + V = \{u + v \mid u \in U \text{ and } v \in V\}$$

- Called the *direct sum*, denoted $U \oplus V$, if the decomposition is unique.

- A vector space X is *graded* if it has a decomposition as a direct sum:

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

Each vector in a graded vector space has a unique decomposition

$$x = \sum_{i \in \mathcal{I}} x^i \qquad \text{where} \qquad x^i \in X^i$$

Example: $X = \mathbb{R}^2 = \text{span} \begin{bmatrix} 1 \\ 0 \end{bmatrix} \oplus \text{span} \begin{bmatrix} 0 \\ 1 \end{bmatrix}$

- The image of a vector space X under a linear map $A : X \to Y$ is

$$A(X) = \{Ax \mid x \in X\} \subseteq Y$$

- A linear map $A : X \to X$ is *graded* if

$$A(X^i) \subseteq X^i \qquad \text{for all } i \in \mathcal{I}$$

Example: $X = \mathbb{R}^2 = \text{span} \begin{bmatrix} 1 \\ 0 \end{bmatrix} \oplus \text{span} \begin{bmatrix} 0 \\ 1 \end{bmatrix}$ and $A = \begin{bmatrix} a_1 & 0 \\ 0 & a_2 \end{bmatrix}$

Graded systems

The system $x_{k+1} = Ax_k$ with $x_k \in X$ is *graded* if X is a graded vector space and A is a graded linear map with respect to this grading.

Example: $X = \mathbb{R}^2 = \text{span} \begin{bmatrix} 1 \\ 0 \end{bmatrix} \oplus \text{span} \begin{bmatrix} 0 \\ 1 \end{bmatrix}$ and $A = \begin{bmatrix} a_1 & 0 \\ 0 & a_2 \end{bmatrix}$

Graded systems to not “mix” subspaces:

$$x^i_{k+1} = Ax^i_k \quad \text{for all } i \in \mathcal{I}$$

Proof: The subspace decomposition is unique, and

$$\sum_{i \in \mathcal{I}} x^i_{k+1} = x_{k+1} = Ax_k = A \sum_{i \in \mathcal{I}} x^i_k = \sum_{i \in \mathcal{I}} Ax^i_k$$

A *graded input-output dynamical system* is a system

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

with $x_k \in X$, $u_k \in U$, and $y_k \in Y$ in which

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

and the state-space maps are graded in that

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

Example: $X = U = Y = \mathbb{R}^n$ and A, B, C, D diagonal

Main result

The iterates of a graded system satisfy the dynamics on each subspace:

$$\begin{aligned} x^i_{k+1} &= Ax^i_k + Bu^i_k \\ y^i_k &= Cx^i_k + Du^i_k \end{aligned}$$

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

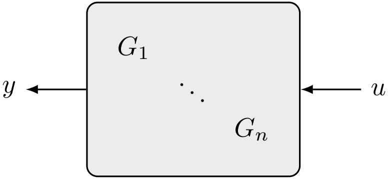
$$x^i_k \in X^i \qquad u^i_k \in U^i \qquad y^i_k \in Y^i$$

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.

Graded systems do not “mix” subspaces.

Examples

Diagonal systems are graded.

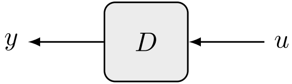


Gradings are the canonical subspaces:

$$U = Y = \mathsf{span}\, e_1 \oplus \cdots \oplus \mathsf{span}\, e_n$$

But graded systems do not have to be diagonal...

Static systems have a canonical grading.



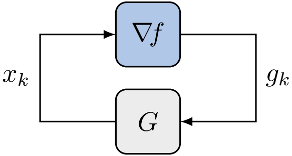
Any matrix $D \in \mathbb{R}^{m \times n}$ is graded with respect to

$$U = \mathbb{R}^n = \mathrm{row}(D) \oplus \mathrm{null}(D)$$
$$Y = \mathbb{R}^m = \mathrm{col}(D) \oplus \mathrm{null}(D^\mathsf{T})$$

D^T is also graded with input and output spaces swapped.

Applications to optimization algorithms

First-order methods are dynamical systems in feedback with the gradient.



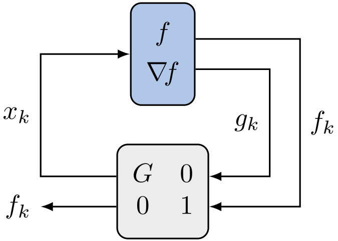
Example: gradient descent

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

has transfer function $\hat{G}(z) = -\frac{\alpha}{z-1} I$

But this does not model the function values $f_k = f(x_k)$.

Including function values results in a graded system.



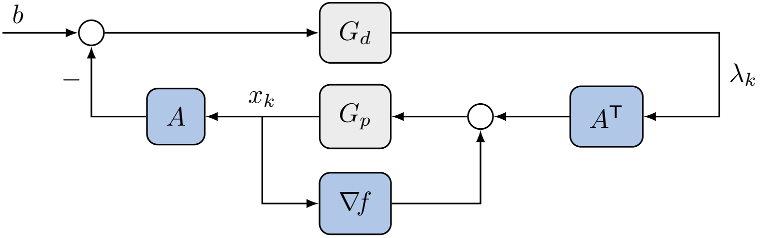
The system is graded with respect to $\mathbb{R}^n \oplus \mathbb{R}$.

Primal-dual algorithms: To solve the linearly-constrained problem

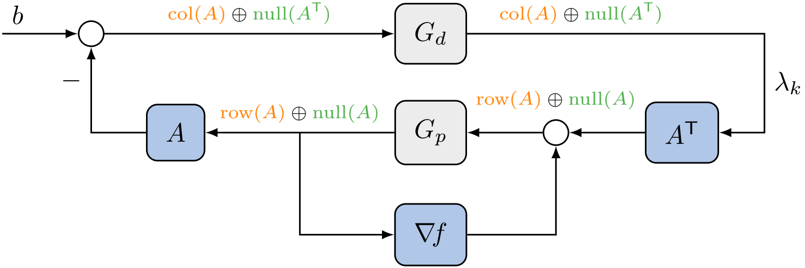
$$\begin{aligned} &\text{minimize} && f(x) \\ &\text{subject to} && Ax = b \end{aligned}$$

we can use the primal-dual method

$$\begin{aligned} x_{k+1} &= x_k - \alpha \left(\nabla f(x_k) + A^\mathsf{T} \lambda_k \right) \\ \lambda_{k+1} &= \lambda_k + \beta \left(b - A x_k \right) \end{aligned}$$



Example: previous algorithm has $\hat{G}_p(z) = \frac{-\alpha}{z-1} I$ and $\hat{G}_d(z) = \frac{\beta}{z-1} I$



- each signal space is graded (i.e., decomposes as a direct sum)
- G_p, G_d, A, and A^T do not "mix" subspaces (only mixed by ∇f)

Lyapunov analysis

To analyze Lyapunov stability of a graded system $x_{k+1} = Ax_k$, search for Lyapunov function that decomposes over the subspaces:

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

If $V^i(x) = x^\mathsf{T} P^i x$ quadratic, then V is a Lyapunov function iff $P^i \succ 0$ and

$$U^i \big(A^\mathsf{T} P^i A - P^i \big) U^i \prec 0 \quad \text{for all } i \in \mathcal{I}$$

where U^i is a matrix whose columns form a basis for X^i.

Lyapunov analysis of graded systems separates along each subspace.

Similar ideas extend to dissipativity analysis, IQCs, ...

Summary

- introduced new notion of *graded* dynamical systems
- showed that this structure arises in modeling optimization algorithms
- can exploit this structure for analysis

Extensions:

- can also define graded systems in terms of their transfer function
- interconnections of graded systems
- exploit for algorithm design?