Quadratic optimization from a control theory perspective

Jean-Jacques Godeme*

In honor, of the 90th Birthday of Professor emiritus R. Tyrrell Rockafellar for all his contributions to mathematics and his role as an inspiration to younger generations.

Abstract. In this short notes, we introduce and study the controllability of the trajectories of a linear dynamical system which is usually used to solve the minimization of a quadratic function in finite dimension. We coin this dynamical system the *controlled gradient flow*. Finally, we introduce what we call the *controlled gradient descent* and the *controlled proximity operator* which are respectively the Euler explicit and implicit discretization of the controlled gradient flow.

Key words. Quadratic optimization, controlled gradient flow, *etc*.

Problem statement

In this draft, we consider the following quadratic optimization problem

$$\min_{x\in\mathbb{R}^n} f(x) = \frac{1}{2}\langle x, Ax\rangle + \langle b, x\rangle + c, \tag{$\mathcal{P}$}$$

where ($\mathcal{P}$) obey to the following hypothesis

Premise A.

- $\mathbb{R}$ is the set of real vectors, n, m are positive integers,
- $A \in \mathbb{R}^{n\times n}$ is a symmetric, positive semidefinite, $b \in \mathbb{R}^n$ a vector.

($\mathcal{P}$) is the one of the most famous approach to solve problems arising in many fields of research like operational research, inverse problems, machine learning and automatic differentiation with applications like X-ray crystallography and light scattering. For instance, the least-square formulation of the compressed sensing can be written as:

$$\min_{x\in\mathbb{R}^n} \frac{1}{2}\|Dx - y\|^2 = \frac{1}{2}\left\langle x, D^\top Dx\right\rangle + \left\langle -2D^\top y, x\right\rangle + \|y\|^2. \tag{0.1}$$

where we want to recover the signal $\bar{x} \in \mathbb{R}^n$ from the measurements vector $y = D\bar{x}$. In such case, one can identify that $A = D^\top D, b = -2D^\top y$ and $c = \|y\|^2$. The set of critical points of ($\mathcal{P}$) is given by

$$\mathrm{crit}(f) = \left\{x \in \mathbb{R}^n \quad \text{s.t.} \quad Ax + b = 0\right\},$$

which is the set of linear equations. From an analytical point of view, if A is positive definite then the solution is unique $x^\star = -A^{-1}b$.

To solve the ($\mathcal{P}$), the common approach is to follow the gradient-flow introduced by the pioneered work of Cauchy [1] and have been intensively studied by [9, 8, 5, 3, 2].

Let us now turn our attention to the gradient flow of this problem, we have

$$\dot{x}(t) = -Ax(t) - b. \tag{GF}$$

*INRIA Sophia & Université Côte d'Azur, CNRS, LJAD, France. e-mail:jean-jacques.godeme@inria.fr

Controlability

At this step, our main inspiration comes from control theory [4, 6, 10]. Our goal is to control the trajectories of this dynamical system by acting on it with a another "function" u commonly called *control*. Why control theory in optimization? Indeed, roughly speaking, control theory answers the question of whether it is possible or not to control the trajectories of a dynamical system between two known states. To the best of our knowledge the most close approach to ours is this paper [7] where the author proposed a model of reinforcement learning where we have an objective function subject to a constraint which is a controlled dynamical system. The two approaches are completely different because we impose the control directly on the gradient flow dynamics.

We consider in this work (GF) on which we act with an additive control. We can rewrite (GF) as

$$\dot{x}(t) = -Ax(t) - b + Bu(t), \tag{0.2}$$

where $B \in \mathbb{R}^{n\times m}$ with $m \leq n$ and u is the control taking values in $\mathbb{R}^m$. Let us simplify the notations, we get

$$\dot{x} = -Ax + Bu - b. \tag{0.3}$$

(0.3) is also known in the control theory literature as the *time invariant linear control system*. Let T_0 and T_1 be two real numbers such that $T_0 < T_1$. We want to study the following controlled system

$$\begin{cases} \dot{x} = -Ax + Bu - b \\ x(T_0) = x_0 \in \mathbb{R}^n, u \in L^\infty((T_0, T_1); \mathbb{R}^m). \end{cases} \tag{c-GF}$$

Let us properly define when a system is called "controllable" in finite dimension.

Definition 0.1 (Controllability). Let us consider the system (c-GF), we say that (c-GF) is controllable from $x_0 \in \mathbb{R}^n$ in time $T > 0$ if their exist $\tilde{T} \in [T_0, T_1]$ such that $T = \tilde{T} - T_0$ and one can reach from the initial point $x(T_0) = x_0$ any point of the space by acting on the system with a control $u \in L^\infty((T_0, T_1); \mathbb{R}^m)$.

The following result is fundamental in linear control theory. It gives a necessary and sufficient condition for controllability without any direct integration procedure. It is also known as the rank condition.

Proposition 0.2. *The time invariant linear control system* (c-GF) *is controllable on* $[T_0, T_1]$ *if and only if we have*

$$\operatorname{span}\left\{(-A)^i Bv; v \in \mathbb{R}^m, i = 0, \cdots, n-1\right\} = \mathbb{R}^n. \tag{0.4}$$

Futhermore, whatever $T_0 < T_1$ *and* $\widetilde{T}_0 < \widetilde{T}_1$, (c-GF) *is controllable on* $[T_0, T_1]$ *if and only if it is controllable on* $[\widetilde{T}_0, \widetilde{T}_1]$.

Proof. The proof can be found in monograph such as [4, 10]. □

Remark 0.3.

- Proposition 0.2 states that under the rank condition (0.4), the system (c-GF) is controllable. Roughly speaking, the gradient flow for minimizing the function $f(x) = \frac{1}{2}\langle x, Ax\rangle + \langle b, x\rangle + c$ is controllable. Therefore, given an initial point $x(T_0) = x_0$ and a desired target $x(T_1) = x_d^\star \in \operatorname{crit}(f)$ one can control gradient flow from x_0 to $x_d^\star$. This result is quite important since it means that instead of running the gradient flow to find any element in $\operatorname{crit}(f)$, one can control the gradient flow in such a way that we can reach our target state $x_d^\star$.
- The condition (0.4) appears first in a paper by Lev Pontryagin in 1959 and later in a joint paper by Rudolph Kalman *et al.* in 1963, it also known as the Kalman condition in control theory.

- The *Kalman condition* can be equivalently stated as a rank condition *i.e.*

$$\mathrm{rank}(K) = n \quad \text{where} \quad K \overset{\text{def}}{=} \left[B|(-A)B|A^2B\cdots|(-A)^{n-1}B\right]. \tag{0.5}$$

- Although controllability of (c-GF) is really interesting, it is not realistic since we can not control a real-world system with any desired control typically with a control which has «high energy».

Explicit discretization The forward Euler method or the explicit discretization of (c-GF) yield to

$$\forall k \in \mathbb{N}, \quad \frac{x_{k+1} - x_k}{\gamma_k} = -Ax_k - b + Bu_k \Leftrightarrow x_{k+1} = x_k - \gamma_k\,(Ax_k + b) + Bu_k, \tag{0.6}$$

where $x_k = x(t_k), u_k = u(t_k)$ with $t_k = k\gamma_k$ and γ_k is the stepsize of the discretization of the time interval $[T_0, T_1]$.

Finally, we get the following system

$$\forall k \in \mathbb{N}, \quad \begin{cases} x_{k+1} = x_k - \gamma\,(Ax_k + b) + Bu_k \\ x_0 \in \mathbb{R}^n, u_k \in \mathbb{R}^m, B \in \mathbb{R}^{n\times m}, \end{cases} \tag{c-GD}$$

We call the system (c-GD) the controlled gradient descent applied on the quadratic program ($\mathcal{P}$).

Implicit discretization The backward Euler method or the implicit discretization of (c-GF) yield to

$$\forall k \in \mathbb{N}, \quad \frac{x_{k+1} - x_k}{\gamma_k} = -Ax_{k+1} - b + Bu_{k+1} \Leftrightarrow x_{k+1} = x_k - \gamma_k Ax_{k+1} - b + \gamma_k Bu_{k+1}, \tag{0.7}$$

where $x_k = x(t_k), u_k = u(t_k)$ with $t_k = k\gamma_k$ and γ_k is the stepsize of the discretization of the time interval $[T_0, T_1]$. We recall that $\nabla f(x) = Ax + b$.

$$\forall k \in \mathbb{N}, \quad \begin{cases} x_{k+1} = \left(\mathrm{Id} + \gamma_k\{\nabla f\}\right)^{-1} \circ \psi_{u_{k+1}}(x_k) \\ x_0 \in \mathbb{R}^n, u_{k+1} \in \mathbb{R}^m, B \in \mathbb{R}^{n\times m}, \end{cases} \quad \text{where} \quad \psi_{u_{k+1}}(x) \overset{\text{def}}{=} x_k + \gamma_k Bu_{k+1} \tag{c-prox}$$

$\forall k \in \mathbb{N}, \psi_{u_{k+1}}$ can be seen as a control over the resolvent.

Remark 0.4.
- This expression (c-prox) is closely related to a controlled version of the Moreau's envelop and proximity operator.
- This *controlled resolvent* has probably a link with *the Bregman D-prox*.

Conclusion & Perspectives

To close this really short notes, our approach to optimization is new to the best of our knowledge and we have introduced a control which can cost computationally however it has a rich literature and toolboxes such as optimization. It is also possible to have an explicit expression of the control u. *We want to investigate how to extend this results to the non-linear case, the stochastic case and possibly infinite dimension, which will be the subject of our future works.*

Acknowledgments.
Theses short notes were written as a side project.

References

[1] Cauchy A. Méthode générale pour la résolution des systèmes d'équations simultanées. *Comptes Rendus de l'Académie des Sciences de Paris*, 25:536–538, January 1847.

[2] H. Attouch and J. Fadili. From the ravine method to the nesterov method and vice versa: A dynamical system perspective. *SIAM Journal on Optimization*, 32(3):2074–2101, 2022.

[3] H. Brezis. Asymptotic behaviour of some evolution systems. In Michael G. Crandall, editor, *Nonlinear Evolution Equations*, pages 141–154. Academic Press, 1978.

[4] J.-M. Coron. *Control and Nonlinearity*. American Mathematical Society, USA, 2007.

[5] S. M. Mordukhovich. *Variational Analysis and Applications*. Springer Monographs in Mathematics. Springer Cham, Berlin, Heidelberg, 2018.

[6] N. Raissi, C. Sanogo, and M. Serhani. Differential game model for sustainability multi-fishery. *Trends in Biomathematics: Mathematical Modeling for Health, Harvesting, and Population Dynamics: Selected works presented at the BIOMAT Consortium Lectures, Morocco 2018*, pages 119–137, 2019.

[7] B. Recht. *A tour of reinforcement learning : the view from continuous*. arxiv:1806.09460v2, 2018.

[8] R. T. Rockafellar. *Convex Analysis*. Princeton University Press, 1970.

[9] R. T. Rockafellar and R. J. B. Wets. *Variational Analysis*, volume 317 of *Grundlehren der mathematischen Wissenschaften*. Springer Berlin Heidelberg, Berlin, Heidelberg, 1998.

[10] E. Trélat. *Contrôle optimal: théorie & applications*. Mathématiques concrètes. Vuibert, 2005.