IEEE Control Systems Letters paper presented at
2023 62nd IEEE Conference on Decision and Control (CDC)
December 13-15, 2023. Marina Bay Sands, Singapore

Optimal Control of Distributed Ensembles with Application to Bloch Equations

Roman Chertovskih, Nikolay Pogodaev, Maxim Staritsyn, and A. Pedro Aguiar,

***Abstract*— Motivated by the problem of designing robust composite pulses for Bloch equations in the presence of natural perturbations, we study an abstract optimal ensemble control problem in a probabilistic setting with a general nonlinear performance criterion. The model under study addresses mean-field dynamics described by a linear continuity equation in the space of probability measures. For the resulting optimization problem, we derive an exact representation of the increment of the cost functional in terms of the flow of the driving vector field. Using this representation, a descent method is designed that is free of any internal line search. The method is applied to solve new optimal control problems for distributed ensembles of Bloch equations.**

I. MOTIVATION

Consider a population of homotypic individuals labeled by the points ω of some set Ω. The state of the ωth object at the time moment t, $x(t,\omega) \in \mathbb{R}^n$, evaluates on a given time interval $I \doteq [0,T]$ under the action of the parameterized vector field $V : \mathbb{R}^n \times \mathbb{R}^s \times U \to \mathbb{R}^n$, starting from a given position $x_0(\omega) \in \mathbb{R}^n$:

$$\left\{ \begin{array}{l} \partial_t x(t,\omega) = V_u\left(x(t,\omega), \eta(\omega)\right) \\ x(0,\omega) = x_0(\omega) \end{array} \right| \quad \omega \in \Omega. \qquad (1)$$

The dynamics (1) involves two types of structural "parameters": the function $\eta : \Omega \to \mathbb{R}^s$ manifests disturbances or structural variations of the underlying model, while an exogenous signal u with values in a given set $U \subseteq \mathbb{R}^m$ models the control action.

In the simplest case, the parameterization space Ω is just a finite set of indexes, and (1) reduces to a multi-agent system of non-interacting units. In a more general setup, we deal with the continuum of individuals moving in a discoordinated way.

Commonly, in such models, Ω is a simply organized compact subset of $\mathbb{R}^s$, $l = s$ and η is the identity mapping $\Omega \to \Omega$.

Problems of ensemble control arise when one has to design a control signal in a "broadcast" way, i.e., such that it acts simultaneously on all individual trajectories $x(\cdot,\omega)$, $\omega \in \Omega$, to force them towards a desired behavior; this means that u should be a function $t \mapsto u(t)$ of time variable only (independent of ω).

A canonical example is the problem of designing external excitations of quantum ensembles. Pioneering works in this area [1], [2] were focused on the famous Bloch equation, which models the macroscopic evolution of bulk magnetization in a population of non-interacting nuclear spins immersed in an intense static magnetic field, which is modulated by the radio frequency (rf-) field (for some extensions, see, e.g., [3] and citations therein). In nuclear magnetic resonance (NMR) experiments, the strength of the applied magnetic field is subject to unavoidable perturbations (static- and/or rf-field inhomogeneity), while the spin ensembles demonstrate perceptible variations in their dissipation rates and/or natural frequencies (Larmor dispersion). The related problem of control engineering is to design robust signals (so-called composite pulses) compensating for the mentioned disturbances; mathematically, this task can be formalized as a problem of optimal ensemble control, see, e.g. [4]. In NMR spectroscopy, the designed pulse sequences are typically desired to be selective, i.e., some sub-populations (with prescribed Larmor frequencies) have to be excited, while the other ones should remain intact or saturated [5]; such are, e.g., contrast problems in NMR imaging [6], [7]. In the language of ensemble control, this means to drive several uncoupled populations of spins by a common magnetic field.

A. Probabilistic Setup. Distributed Ensembles

In contrast to [6]–[8], our approach stems from the probabilistic interpretation of the ensemble dynamics, assuming that Ω is endowed with the structure of probability space $(\Omega, \mathcal{A}, \mathbb{P})$ with a specified σ-algebra $\mathcal{A} \subset 2^\Omega$ and a canonical probability measure $\mathbb{P}$ on $(\Omega, \mathcal{F})$ (we shall write $\mathbb{P} \in \mathcal{P}(\Omega)$).

This interpretation is motivated by practical applications, in which the individual states $x(\cdot,\omega)$ can not be measured directly, and all the available information is based on some "observables" – measurement outputs accompanying the dynamics (1) and involving certain statistical characteristics, see, e.g., [9].

In the probabilistic setup, the map $(t,\omega) \mapsto (x(t,\omega), \eta(\omega))$ is naturally viewed as a random process, and the behavior of the random variable $\omega \mapsto (x(t,\omega), \eta(\omega))$ can be analyzed by investigating the time-evolution of its law

$$\varrho_t = \left(x(t,\cdot), \eta(\cdot)\right)_\sharp \mathbb{P} \in \mathcal{P}(\mathbb{R}^{n+s}). \qquad (2)$$

Hereinafter, the operator $F_\sharp : \mathcal{P}(\mathcal{X}) \to \mathcal{P}(\mathcal{Y})$ denotes the pushforward of a measure $\mu \in \mathcal{P}(\mathcal{X})$ through a (Borel) map $F : \mathcal{X} \to \mathcal{Y}$ between two measurable spaces that acts on functions $\varphi : \mathcal{Y} \to \mathbb{R}$ with the property $\varphi \circ F \in L_1(\mathcal{X}; \mu)$ by the rule (for the other properties, see, e.g., [10])

$$\int_{\mathcal{Y}} \varphi \, \mathrm{d}(F_\sharp \mu) = \int_{\mathcal{X}} \varphi \circ F \, \mathrm{d}\mu. \qquad (3)$$

Under the standard regularity of the map $(x,\eta) \mapsto V_v(x,\eta)$, the measure-valued curve $t \mapsto \varrho_t$ is a unique distributional solution of the continuity equation [11]

$$\partial_t \varrho_t + \nabla_x \cdot \left(V_{u(t)}\, \varrho_t\right) = 0, \quad \varrho_0 = (x_0(\cdot), \eta(\cdot))_\sharp \mathbb{P}; \qquad (4)$$

∇_x denotes the gradient w.r.t. $x \in \mathbb{R}^n$, and "$\cdot$" means the scalar product.

The discussed interpretation of (1) postulates a passage from the microscopic model (a random process satisfying the ODE with uncertainty) to a macroscopic one (a deterministic distributed process described by the PDE) called the mean field; systems (1) and (4) are the so-called Lagrangian and Eulerian forms of the mean-field dynamics, respectively [12].

Remark that, as a result of this passage, a nonlinear finite-dimensional object is replaced by an infinite-dimensional but state-linear (ϱ-linear) one. The linearity of the reduced model plays a vital role in our study as it gives rise to an exact representation of the increment (∞-order variation) of the cost functional in the corresponding optimal control problem to be presented in Sec. IV.

Finally, observe that PDE (4) can be viewed as a family $\eta \mapsto \mu_t^\eta$ of $\mathcal{P}(\mathbb{R}^n)$-valued curves solving, Ξ-a.e., the "sliced" continuity equation with the vector field $V^\eta \doteq V(\cdot, \eta)$ and initial condition $\mu_0^\eta = \vartheta^\eta \in \mathcal{P}(\mathbb{R}^n)$, where the map $\eta \mapsto \vartheta^\eta$ is obtained by disintegrating [13] the distribution ϱ_0 w.r.t. the projection $\Xi \doteq ((x, \eta) \mapsto \eta)_\sharp \varrho_0 \in \mathcal{P}(\mathbb{R}^s)$. We call such a family the distributed ensemble; this concept separates two types of uncertainty: dispersion in the initial data $\omega \mapsto x_0(\omega)$ is converted to the mean field, while fluctuations of the dynamics, $\eta \mapsto V^\eta$, are treated independently.

B. Contribution and Novelty

This work has two lines of contribution: the theoretical part (Secs. II-IV) presents a general approach to the numerical solution of optimal ensemble control problems on the space of probability measures, based on the exact (∞-order) variation of the cost functional. We elaborate on a new technique relying on the concept of intrinsic derivative of a function of probability measure, and the language of flows of vector fields. This technique enables us to compute explicitly (to our knowledge, for the first time in the existing literature) the ∞-order variation of an abstract nonlinear functional along the measure-valued solution of the continuity equation, which essentially generalizes our recent results [14], [15] obtained for problems of the linear-quadratic structure (using the duality argument that is inapplicable in the general nonlinear case). Another line of contribution is the application to the optimal control of quantum ensembles (Sec. V). Our interpretation captures the natural probabilistic flavor of ensemble control problems, and enables us to improve the quality of designed control signals since it takes into account the available statistical information, and in this way allows us to concentrate the "resource" of feasible control options around relevant values of ω.

A key result is a descent algorithm for optimal ensemble control originating from an exact variation. In contrast to the familiar indirect methods [16], [17] based on the first variation (i.e. on Pontryagin's maximum principle, PMP), our approach is free of additional parameters, like step length in the line search. Typically, this property brings a significant gain in computational performance: such algorithms converge [15] to a PMP extremal faster than PMP-based methods. In our computational practice, we observe that solving similar problems by the gradient descent with backtracking [17] requires hundreds of iterations, while the method presented here converges in several ones (Sec. V).

II. OPTIMAL CONTROL PROBLEM

First, we introduce some necessary notations: Let $\mathcal{X}$ be a metric space, and $I \doteq [0, T]$. We denote by $\mathcal{C}(I; \mathcal{X})$ the spaces of continuous maps $I \mapsto \mathcal{X}$ with the usual sup-norm. If $\mathcal{X} \subseteq \mathbb{R}^n$, $\mathcal{C}^1(\mathcal{X})$ denotes the space of continuously differentiable functions $\mathcal{X} \to \mathbb{R}$, and $\mathcal{C}_c^\infty(\mathcal{X})$ the space of smooth functions with a compact support in $\mathcal{X}$; $L_p(I; \mathbb{R}^m)$, $p = 1, \infty$, the Lebesgue spaces of summable and bounded measurable functions $I \mapsto \mathbb{R}^m$, respectively. $\mathcal{P}(\mathcal{X})$ the set of probability measures on $\mathcal{X}$, and $\mathcal{P}_c(\mathcal{X}) \subseteq \mathcal{P}(\mathcal{X})$ the set of measures having compact support in $\mathcal{X}$; $\mathcal{P}_c(\mathbb{R}^n)$ is a dense subset of a complete separable metric space as it is endowed with any p-Kantorovich (Wasserstein) distance W_p, $p \geq 1$. Among all measures on $\mathbb{R}^n$, we mark out two specific ones – the usual Lebesgue measure, $\mathcal{L}^n$, and a Dirac point-mass measure concentrated at $x \in \mathbb{R}^n$, δ_x.

A. General Problem Statement

Our prototypic mathematical object is the following optimization problem on $\mathcal{P}_c(\mathbb{R}^n)$:

$$
\begin{aligned}
(P) \quad \min \quad & \mathcal{I}[u] = \ell(\mu_T[u]) \\
\text{subject to} \quad & \partial_t \mu_t + \nabla_x \cdot (V_u \, \mu_t) = 0, && (5) \\
& t \in I \doteq [0, T]; \quad \mu_0 = \vartheta; && (6) \\
& u(\cdot) \in \mathcal{U} \doteq L_\infty(I; U), \; U \subset \mathbb{R}^m, && (7)
\end{aligned}
$$

where $\ell : \mathcal{P}_c(\mathbb{R}^n) \to \mathbb{R}$ is a given performance criterion, and $V : \mathbb{R}^n \times U \to \mathbb{R}^n$ a control vector field. Despite its probabilistic appearance, (P) is a deterministic optimal control problem, in which the trajectories are measure-valued functions $\mu \in \mathcal{C}(I, \mathcal{P}_c(\mathbb{R}^n))$, and the control signals are usual functions $u \in L_\infty(I, U)$. This problem can be specified to the case of distributed ensembles as follows:

$$
(\widetilde{P}) \quad \min \int_{\mathbb{R}^s} \ell(\mu_T^\eta[u]) \, \mathrm{d}\Xi(\eta), \tag{8}
$$

where $t \mapsto \mu_t^\eta[u]$ solves the linear PDE (5), (6) with $V_u = V_u^\eta$ for Ξ-a.a. $\eta \in \mathbb{R}^s$. Note that problems (P) and $(\widetilde{P})$ are essentially equivalent, up to taking the expectation w.r.t. Ξ. This fact is due to the specific structure of the ensemble dynamics (1) which does not contain the evolution w.r.t. η.

We make the following standard regularity hypotheses:

- (A_1) the map $(x, \upsilon) \mapsto V_\upsilon(x)$ is continuous, continuously differentiable in x and satisfies the sublinear growth condition: there exists a constant $M > 0$ such that $V_\upsilon(x) \leq M(1 + |x|)$ for all $x \in \mathbb{R}^n$ and $\upsilon \in U$.
- (A_2) The set U is convex and compact.
- (A_3) $\vartheta \in \mathcal{P}_c(\mathbb{R}^n)$, and $\ell : \mathcal{P}_c(\mathbb{R}^n) \to \mathbb{R}$ is continuous.
- (A_4) $\ell \in \mathcal{C}^1$ in the sense of intrinsic derivative.

(A_1) is the standard set of assumptions to guarantee the well-posedness of the PDE (5) [11]. (A_1)–(A_3) imply the

existence of a minimizer for problem (P) [18, Theorem 3.2]; under these assumptions, the solution μ_t of (5), (6) is supported in a ball whose radius depends only on the problem data [18, Lemma A2]. Hence, $\mu_t \in \mathcal{P}_c(\mathbb{R}^n)$ for all $t \in I$.

B. Problem Specification

Below, we provide some examples of the performance criterion ℓ that cover typical optimization tasks in the area of ensemble control.

Targeting: In several NMR applications, the guide is supposed to transfer the ensemble (as close as possible) to a target profile $x(T,\cdot) = \hat{x}(\cdot)$, which means to minimize the quantity

$$\int_\Omega |x[u](T,\omega) - \hat{x}(\omega)|^2 \,\mathrm{d}\mathbb{P}(\omega).$$

Such are problems of selective spin excitation, see [8], [19] and the bibliography therein.

This problem is formulated in our setting by using (2), the definition of μ^η, and the change of variable formula (3):

$$\min \int_{\mathbb{R}^s} \ell(\mu^\eta_T; x_T[u](\eta)) \,\mathrm{d}\Xi(\eta), \tag{9}$$

where $\ell(\mu; x) \doteq \displaystyle\int_{\mathbb{R}^n} |y - x|^2 \,\mathrm{d}\mu(y)$. Typically, the map x_T is chosen to be constant, which means that the ensemble is assumed to be aggregated around some given position.

Statistical Tracking: In some cases [15], [20], the previous performance criterion could be too rigid. Instead of matching the desired profile in average, one may require that the target distribution has prescribed statistical characteristics, for instance, its expectation and variance approach some desired values. The cost functional can be reset in the language of distributed ensembles as follows:

$$\int_{\mathbb{R}^s} \Big[\psi_1 \left(\mathcal{E}(\mu^\eta_T[u]) - \hat{\mathcal{E}} \right) + \psi_2 \left(\mathcal{V}(\mu^\eta_T[u]) - \hat{\mathcal{V}} \right) \Big] \,\mathrm{d}\Xi(\eta), \tag{10}$$

where $\mathcal{E}(\mu)$ and $\mathcal{V}(\mu)$ denote the expectation and variance of $\mu \in \mathcal{P}(\mathbb{R}^n)$, respectively, $\hat{\mathcal{E}} \in \mathbb{R}^n$ and $\hat{\mathcal{V}} \in \mathbb{R}$ are target values of the statistical characteristics, and $\psi_1 : \mathbb{R}^n \to \mathbb{R}$ and $\psi_2 : \mathbb{R} \to \mathbb{R}$ are given penalty functions.

Minimum-Energy Control: In many applications, the discussed cost functionals are accompanied by the energy term

$$\frac{\alpha}{2} \int_0^T |u(t)|^2 \,\mathrm{d}t \tag{11}$$

with some weight $\alpha > 0$. In particular, this produces a sort of regularization of the underlying problem.

III. PRELIMINARIES

In this section, we provide the necessary theoretical background and collect some auxiliary results.

A. Flows of Vector Fields. Transport Equation

Let $V : I \times \mathbb{R}^n \to \mathbb{R}^n$ be a time-dependent vector field generating a flow, i.e. a map $X : I \times I \times \mathbb{R}^n \to \mathbb{R}^n$ such that, for all $s \in \mathbb{R}$ and $x \in \mathbb{R}^n$, the function $t \mapsto X_{s,t}(x)$ is a solution of the ODE

$$\partial_t X_{s,t} = V_t \circ X_{s,t}, \quad X_{s,s} = \mathbf{id}, \tag{12}$$

where $\mathbf{id}$ stands for the identical map $\mathbb{R}^n \to \mathbb{R}^n$. In view of the semigroup property $X_{t_0,t_2} = X_{t_1,t_2} \circ X_{t_0,t_1}$ $\forall\, t_0, t_1, t_2$, the inverse of $X_{s,t}$ is the map $X_{t,s}$.

Fixed s, abbreviate $P_t = X_{s,t}$ and $Q_t = X_{t,s}$. Then, by the chain rule, $0 = \partial_t(\mathbf{id}) = \partial_t(Q_t \circ P_t) = (\partial_t Q_t + D_x Q_t \, V_t) \circ P_t$. Since the expression in the brackets vanishes for all values $P_t(x)$, and therefore, for any $x \in \mathbb{R}^n$, we conclude that the inverse flow should satisfy the linear operator equation

$$\partial_t Q_t + D_x Q_t \, V_t = 0, \quad Q_s = \mathbf{id}\,. \tag{13}$$

Returning to the X-notation, and recalling that the Jacobian $J_{t,s} \doteq D_x X_{t,s}$ satisfies [21, Ths. 2.2.3 and 2.3.2] the linear problem

$$\partial_t J_{t,s} = -J_{t,s} \, (D_x V_t \circ X_{t,s}), \quad J_{s,s} = E, \tag{14}$$

where E denotes the identity matrix, we express the derivative of the inverse flow w.r.t. t as follows:

$$\partial_t X_{t,s} = -J_{t,s} \, V_t. \tag{15}$$

Note that operators P and Q refer to the concepts of the left and right chronological exponents in the tradition of geometric control theory [22].

B. Continuity Equation

Recall that the continuity equation (5) on the space $\mathcal{P}_c(\mathbb{R}^n)$ is understood in the weak (distributional) sense. A function $\mu : t \to \mu_t$ is said to be a weak solution of (5) iff the following equality holds for all $\varphi \in \mathcal{C}^\infty_c((0,T) \times \mathbb{R}^n)$:

$$0 = \int_0^T \mathrm{d}t \int \left(\partial_t \varphi_t + \nabla_x \varphi_t \cdot V_{u(t)} \right) \mathrm{d}\mu_t. \tag{16}$$

Hereinafter, we abbreviate $\int = \int_{\mathbb{R}^n}$. (A_1) guarantees the existence of a unique weak solution to the Cauchy problem (5), (6) [11, Propositions 2.10, 2.11]; this solution admits the following representation [11, Proposition 2.12] in terms of the characteristic flow (12): $\mu_t = (X_t)_\sharp \vartheta$, where $X_t \doteq X_{0,t}$.

Clearly, the notion of weak solution applies "pointwise" to the distributed ensemble $\eta \mapsto \mu^\eta$, for Ξ-a.a. $\eta \in \mathbb{R}^s$.

C. Differentiation w.r.t. the Probability Measure

Since $\mathcal{P}_c$ is merely a metric space and does not have a linear structure, standard concepts of the directional derivative are not applicable here (there are simply no "directions" in common sense). At the same time, there is an option to differentiate a function $F : \mathcal{P}_c(\mathbb{R}^n) \to \mathbb{R}$ at some $\mu \in \mathcal{P}_c(\mathbb{R}^n)$ in the "direction" of a (Borel measurable and locally bounded) vector field $f\colon \mathbb{R}^n \to \mathbb{R}^n$ pushing the measure μ: $\frac{d}{d\lambda}\Big|_{\lambda=0} F\left((\mathbf{id} + \lambda f)_\sharp \mu\right)$. Under some reasonable

regularity [23] of the map F, this derivative does exist and takes the form: $\int D_\mu F(\mu)\cdot f\,\mathrm{d}\mu$, where the linear map $D_\mu F\colon \mathcal{P}_c(\mathbb{R}^n)\times\mathbb{R}^n\to\mathbb{R}^n$, called the intrinsic derivative, can be calculated as follows

$$D_\mu F(\mu)(x) = D_x \lim_{h\downarrow 0}\frac{1}{h}\big(F\left(\mu+h(\delta_x-\mu)\right)-F(\mu)\big). \quad (17)$$

The expression under the sign of D_x is called the flat derivative of F (typically denoted by $\frac{\delta F}{\delta \mu}$). Note that the notions of intrinsic and flat derivatives are naturally connected to another useful concept of derivative on $\mathcal{P}_c(\mathbb{R}^n)$, the so-called localized Wasserstein derivative [24].

In contrast to the other concepts of derivative in the space of measures, the quantity (17) can be computed explicitly (and rather easily) for many functionals arising in practice, in particular, for those specified in § II-B. Below, we shall utilize this advantage.

IV. INCREMENT FORMULA

Given two controls $\bar u, u\in\mathcal{U}$, where $\bar u$ is an initial (reference) one, and $u\neq\bar u$ is the target one, we abbreviate by $\bar X$ and X the flows of the vector fields $\bar V_t\doteq V_{\bar u(t)}$ and $V_t\doteq V_{u(t)}$, respectively, and by $\mu:t\mapsto\mu_t[u]=(X_t)_\sharp\vartheta$ and $\bar\mu:t\mapsto\mu_t[\bar u]=(\bar X_t)_\sharp\vartheta$ the corresponding solutions to the Cauchy problem (5), (6). Consider the increment $\Delta_u\mathcal{I}[\bar u]\doteq\mathcal{I}[u]-\mathcal{I}[\bar u]\doteq\ell(\mu_T)-\ell(\bar\mu_T)$ of the cost functional. The base of our approach is the following result proved in Appendix A.

Theorem 1 (Increment formula): Assume that (A_1)–(A_4) hold. Then, the following representation is valid:

$$\Delta_u\mathcal{I}[\bar u]= \quad (18)$$
$$\int_0^T \mathrm{d}t\int D_\mu\ell^*\big|_{(\bar X_{t,T})_\sharp\mu_t}\circ\bar X_{t,T}\,\bar J_t\,\left(V_t-\bar V_t\right)\,\mathrm{d}\mu_t.$$

Here, $\bar J$ is a solution to the linear problem (14) corresponding to $u=\bar u$ and $s=T$; * stands for the matrix transposition. Observe that (18) is an exact variation of $\mathcal{I}$ at the point $\bar u$ w.r.t. any other admissible signal $u\in\mathcal{U}$.

Remark 1: The representation (18) (and the consequent numeric method) can be literally adopted to the case of distributed ensembles by replacing $(\bar V,\bar X,\bar J,X,V)$ with $(\bar V^\eta,\bar X^\eta,\bar J^\eta,X^\eta,V^\eta)$ and taking the expectation w.r.t. Ξ.

A. Control Improvement

The main consequence of the increment formula is the structure of controls of potential decrease from the reference point $\bar u$ provided by minimizers $w_t[\mu]$ in the problem

$$\min_{v\in U}\int D_\mu\ell^*\Big|_{\left(\bar X_{t,T}\right)_\sharp\mu}\circ\bar X_{t,T}\,\bar J_t\,V_v\,\mathrm{d}\mu \quad (19)$$

viewed as μ-feedback controls of the PDE (4). Indeed, if $t\mapsto\check\mu_t$ is a well-defined solution to an initial value problem (4), (6) with a backfed nonlocal vector field $\check V_t\doteq V_{w_t[\check\mu_t]}$, and $u(t)\doteq w_t[\check\mu_t]$, then, obviously, $\Delta_u\mathcal{I}[\bar u]\le 0$. Thus, the cost of open-loop controls u generated by the feedbacks (19) does not exceed (potentially, smaller than) the one of $\bar u$.

B. Numeric Algorithm

A pitfall in the discussed control-update rule is due to the (generic) discontinuity of the map $x\mapsto\check V_t(x)$ that makes the Cauchy problem (12) ill-posed. To resolve this issue, one can employ the classical semi-discrete Krasovskii-Subboting sampling scheme [25] with a time discretization (partition) $\pi_I^N=\{0=t_0<t_1<\ldots<t_N=T\}\subset I$.

Let u^k, $k\in 0,1,\ldots$, be given/computed. On the conceptual level, an iteration of the announced iterative method consists of just three steps:

i) integration of the ODE (12) together with the linearized system (14), for $\bar u=u^k$ and various initial conditions over some mesh $\pi^M_{\operatorname{spt}\vartheta}=\{y^k\}_{k=0}^M\subseteq\operatorname{spt}\vartheta$, to obtain (X^k,J^k),

ii) numeric solution of the PDE (4), (6) backfed by (19) with $(\bar X,\bar J)=(X^k,J^k)$, to obtain μ^{k+1}, and

iii) control update $u^{k+1}:=w_t[\mu_t^{k+1}]$.

Arguments similar to [15, Appendix B] show that this iterative method converges in the residual of Pontryagin's maximum principle [16] for the convexified problem (P) as $\max\limits_i|t_i-t_{i-1}|+\max\limits_k\|y^k-y^{k-1}\|\to 0$ over $\pi_I^N\times\pi^M_{\operatorname{spt}\vartheta}$.

V. APPLICATION: BLOCH EQUATIONS

We now apply the algorithm from § IV-B to a nonstandard problem of designing composite pulses in a multipopulation of nuclear spins, mentioned in the Introduction. Consider a family of Bloch equations, parameterized by the (dimensionless) resonance offset η.

For simplicity, we focus on the non-dissipative case and rewrite the Bloch equations in spherical polar coordinates in the rotating frame [26]:

$$\begin{pmatrix}\dot\theta\\ \dot\varphi\end{pmatrix}=V_u^\eta(\theta,\varphi)\doteq u\begin{pmatrix}\cot\varphi\cos\theta\\ \sin\theta\end{pmatrix}-\eta\begin{pmatrix}1\\0\end{pmatrix}. \quad (20)$$

Here, $\theta\in[0,2\pi]$ and $\varphi\in[0,\pi]$ are the azimuthal and polar angles identifying the position on the Bloch sphere, $(x_1,x_2,x_3)\doteq(\cos\theta\sin\varphi,\sin\theta\sin\varphi,\cos\varphi)$; control input $t\mapsto u(t)$ is the envelope of the actuating rf-field.

Remark 2: It may be apt to stress that the Bloch equations are not really of the quantum feature. These phenomenological ODEs describe the dynamics of an averaged nuclear magnetization in a macroscopic sample, and are inapplicable to an individual nuclear magnetic moment. In other words, each ODE (20) already represents the dynamic ensemble. One can say that, in this example, we actually deal with an "ensemble of ensembles".

A canonical task in NMR experiments is to transfer the bulk magnetization vector from an equilibrium position (aligned with the static magnetic field) to the excited state $(\theta_T,\varphi_T)=(0,\pi/2)$ (so-called $\pi/2$-transfer). In practice, the static field is inhomogeneous, which gives rise to probability distributions $\mu_0^\eta\in\mathcal{P}([0,2\pi]\times[0,\pi])$ in the initial values (θ_0,φ_0), and leads to an optimal control problem of type (9). We assume that μ_0^η are absolutely continuous with a common density function $\rho_0(\theta,\varphi)$, and consider a more delicate performance criterion similar to (10) by incorporating a

variance-like term and the energy cost (11). This problem is adapted to the framework of distributed ensembles as

$$\min \mathcal{I}[u] = \int_{-1}^{1} \ell(\mu_T^\eta)\,\mathrm{d}\Xi(\eta) + \frac{\alpha}{2}\int_0^T u^2(t)\,\mathrm{d}t. \qquad (21)$$

Here, $\ell(\mu) \doteq \int g(\cdot,\cdot,\theta_T,\varphi_T)\,\mathrm{d}\mu + \frac{\beta}{2}\iint g\,\mathrm{d}(\mu\otimes\mu)$, $g(\theta,\varphi,\theta',\varphi') = 1/2[(\sin\theta-\sin\theta')^2+(\sin\varphi-\sin\varphi')^2+(\cos\theta-\cos\theta')^2+(\cos\varphi-\cos\varphi')^2] \doteq 2-\cos(\theta-\theta')-\cos(\varphi-\varphi')$, the integral $\int$ is computed over $[0,2\pi]\times[0,\pi]$, $\otimes$ denotes the tensor product of measures, and $\alpha,\beta>0$ are given parameters. To specify the feedback control (19), we compute: $D_\mu\ell(\mu)(\theta,\varphi) = D_{(\theta,\varphi)}f(\theta,\varphi) + \beta\int D_{(\theta,\varphi)}g(\theta,\varphi,\cdot)\,\mathrm{d}\mu$.

We performed a numerical case study for the initial density ρ_0 (Fig. 1, top panel) on a uniform grid with a spacing of 0.01 for both angles, and for the distribution Ξ chosen to be uniform on $[-0.55,-0.45]$ (due to the lack of space, the results are presented for the mean value $\eta=-0.5$). The standard Lax-Friedrichs numerical integration scheme was implemented (for integration in time from 0 to $T=2$ with the constant time step 10^{-4}). To exclude the singularity at the poles, the problem was solved for $\varphi\in[0.05,0.095\pi]$, assuming the boundaries for θ to be periodic, and vanishing normal derivative for φ. The following values of the parameters in the cost function (21) were taken: $\alpha=0.25$ and $\beta=0.5$. It turned out that the simulations are time-consuming, thus, the code was parallelized for multiprocessor computers with shared memory. The simulations are also memory demanding – storage in memory of a large four-dimensional array (a function of (t,θ,φ,η)) is required (about 150 GB for the parameter values described above).

The initial control is constant, $u^0\equiv 0.1$, with the cost $\mathcal{I}[u^0]\approx 0.88$. Computing the iterations of the proposed algorithms, the cost $\mathcal{I}$ was observed to decrease monotonically (as it is expected): $\approx$ 0.59, 0.49, 0.46, 0.44, 0.43 and then stagnating at a value ≈ 0.43. Terminal density ρ_T and the corresponding control u computed after five iterations are shown in Fig. 1 (middle and bottom panels, respectively).

Our numerical experience can be summarized as follows: the suggested nonlocal algorithm generates a $\mathcal{I}$-monotone control sequence and typically takes a few (2-5) iterations to reach an acceptable solution. Free of any intrinsic parametric optimization, the method can be a "lifeline" for computationally demanding problems (like the presented one).

For future simulations, in order to increase the computational efficiency of our codes and fix the pole problem, we plan to implement the pseudospectral methods using spherical harmonics.

VI. CONCLUSION

Although the proposed approach has a fairly wide scope of application, there are significant restrictions. For instance, the concept of flat derivative (and, as a consequence, intrinsic derivative) does not apply if the functional is undefined ($=+\infty$) for measures, singular w.r.t. a reference one (e.g. $\mathcal{L}^n$); this makes it impossible to treat the performance criteria such as entropy functionals of the Kullback-Leibler type [11].

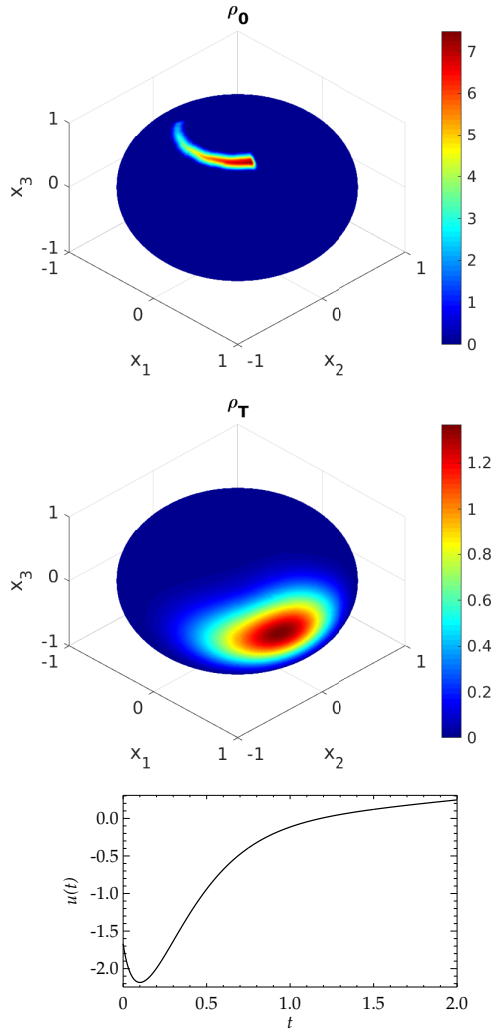

Fig. 1. Initial, for $t=0$, (top), final, for $t=T$, (top) density ρ_T and the corresponding control u (bottom panel) after five iterations of the algorithm.

APPENDIX

A. Proof of Theorem 1

Let $t\mapsto\mu_t$ and $t\mapsto\bar\mu_t$ denote the weak solutions of the PDE (5) with initial condition $\mu_0=\vartheta$, corresponding to control inputs u and $\bar u$, respectively. Recall that $\mu_t = (X_{0,t})_\sharp\vartheta$, $\bar\mu_t=(\bar X_{0,t})_\sharp\vartheta$, and $X_{s,s}=\bar X_{s,s}=\mathbf{id}\ \forall s\in\mathbb{R}$, where X and $\bar X$ are the corresponding characteristic flows.

Denote $\mathcal{F}_t \doteq \bar X_{t,T}\circ X_{0,t}$. Since the map $x\mapsto\mathcal{F}_t(x)$ is a composition of two bijections, it is invertible. Standard arguments from the ODE theory imply that under assumptions (A_1)–(A_3), the maps $t\mapsto\bar X_{t,T}(x)$ and $t\mapsto X_{0,t}(x)$ are Lipschitz on I, for any $x\in\mathbb{R}^n$. Hence, for any $x\in\mathbb{R}^n$, the function $t\mapsto\mathcal{F}_t(x)$ is absolutely continuous on I as a composition of Lipschitz maps; in particular it is $\mathcal{L}^1$-a.e. differentiable: $\mathcal{F}_{t+\lambda} = \mathcal{F}_t + \lambda\,\partial_t\mathcal{F}_t + o(\lambda) \approx (\mathbf{id}+\lambda\,\partial_t\mathcal{F}_t\circ\mathcal{F}_t^{-1})\circ\mathcal{F}_t$. Assumption (A_4) guarantees that, for any Borel measurable, locally bounded map $f\colon\mathbb{R}^n\to$

$\mathbb{R}^n$, the function $\lambda \mapsto \ell\left((\mathbf{id}+\lambda f)_\sharp \mu\right)$ is differentiable at zero, and $\dfrac{d}{d\lambda}\Big|_{\lambda=0} \ell\left((\mathbf{id}+\lambda f)_\sharp \mu\right) = \displaystyle\int D_\mu \ell(\mu)\cdot f \,\mathrm{d}\mu$, where $D_\mu \ell$ stands for the intrinsic derivative. Thus,

$$\begin{aligned}
\partial_t \;\; \ell\left((\mathcal{F}_t)_\sharp\, \vartheta\right) &= \frac{d}{d\lambda}\Big|_{\lambda=0} \ell\left((\mathcal{F}_{t+\lambda})_\sharp\, \vartheta\right) \\
&= \frac{d}{d\lambda}\Big|_{\lambda=0} \ell\left(\left(\mathbf{id}+\lambda\, \partial_t \mathcal{F}_t \circ \mathcal{F}_t^{-1}\right)_\sharp \left((\mathcal{F}_t)_\sharp\, \vartheta\right)\right) \\
&= \int D_\mu \ell\left((\mathcal{F}_t)_\sharp\, \vartheta\right) \cdot \partial_t \mathcal{F}_t \circ \mathcal{F}_t^{-1} \,\mathrm{d}\big((\mathcal{F}_t)_\sharp\, \vartheta\big) \\
&= \int D_\mu \ell\big|_{\left(\bar{X}_{t,T}\circ X_t\right)_\sharp \vartheta} \circ \left(\bar{X}_{t,T}\circ X_{0,t}\right) \cdot \\
&\qquad \partial_t(\bar{X}_{t,T}\circ X_{0,t})\,\mathrm{d}\vartheta. \qquad (22)
\end{aligned}$$

In the last expression, the partial derivative in t is represented by the chain rule as

$$\partial_t(\bar{X}_{t,T}\circ X_{0,t}) = \left[\partial_\tau \bar{X}_{t,T}\circ X_{0,\tau} + \partial_\tau \bar{X}_{\tau,T}\circ X_{0,t}\right]\big|_{\tau=t},$$

where $\partial_\tau\big|_{\tau=t}\bar{X}_{t,T}\circ X_{0,\tau} = \left(D_x \bar{X}_{t,T}\, V_t\right)\circ X_{0,t} \doteq \left(\bar{J}_{t,T}\, V_t\right)\circ X_{0,t}$ by direct computation, and $\partial_\tau\big|_{\tau=t}\bar{X}_{\tau,T} = -\bar{J}_{t,T}\,\bar{V}_t$ by (15). Plugging these expressions to (22), we obtain

$$\begin{aligned}
\partial_t \ell\left((\mathcal{F}_t)_\sharp\, \vartheta\right) \doteq \int \Big[& D_\mu \ell^*\big|_{\left(\bar{X}_{t,T}\circ X_t\right)_\sharp \vartheta} \circ \bar{X}_{t,T} \\
& \bar{J}_{t,T}(V_t - \bar{V}_t)\Big] \circ X_{0,t}\,\mathrm{d}\vartheta. \qquad (23)
\end{aligned}$$

Now, the cost increment is represented as follows:

$$\begin{aligned}
\Delta_u \mathcal{I}[\bar{u}] &\doteq \ell(\mu_T) - \;\ell(\bar{\mu}_T) \\
&= \ell\left((\bar{X}_{T,T}\circ X_{0,T})_\sharp \vartheta\right) - \ell\left((\bar{X}_{T,T}\circ \bar{X}_{0,T})_\sharp \vartheta\right) \\
&\quad \underbrace{- \left[\ell\left((\bar{X}_{0,T}\circ X_{0,0})_\sharp \vartheta\right) - \ell\left((\bar{X}_{0,T}\circ \bar{X}_{0,0})_\sharp \vartheta\right)\right]}_{\equiv 0} \\
&= \int_0^T \partial_t \left[\ell\left((\bar{X}_{t,T}\circ X_{0,t})_\sharp \vartheta\right) - \ell\left((\bar{X}_{t,T}\circ \bar{X}_{0,t})_\sharp \vartheta\right)\right] \mathrm{d}t.
\end{aligned}$$

By the semigroup property, $\bar{X}_{t,T}\circ \bar{X}_{0,t} = \bar{X}_{0,T}$, which implies that the second term under the sign of the time derivative in the latter expression is, in fact, independent of t, and therefore, $\Delta_u \mathcal{I}[\bar{u}]$ equals

$$\int_0^T \partial_t \ell\left(\left(\bar{X}_{t,T}\circ X_{0,t}\right)_\sharp \vartheta\right)\,\mathrm{d}t = \int_0^T \partial_t \ell\left((\mathcal{F}_t)_\sharp\, \vartheta\right)\,\mathrm{d}t.$$

To complete the proof, it remains to combine the latter expression with (23) and use the representation formula $\mu_t = (X_{0,t})_\sharp \vartheta$.

REFERENCES

[1] J.-S. Li and N. Khaneja, "Ensemble control of bloch equations," *IEEE Transactions on Automatic Control*, vol. 54, no. 3, pp. 528–536, 2009.

[2] ——, "Control of inhomogeneous quantum ensembles," *Phys. Rev. A*, vol. 73, p. 030302, Mar 2006.

[3] R. Robin, N. Augier, U. Boscain, and M. Sigalotti, "Ensemble qubit controllability with a single control via adiabatic and rotating wave approximations," *Journal of Differential Equations*, vol. 318, pp. 414–442, 2022.

[4] T. E. Skinner, T. O. Reiss, B. Luy, N. Khaneja, and S. J. Glaser, "Application of optimal control theory to the design of broadband excitation pulses for high-resolution NMR," *Journal of Magnetic Resonance*, vol. 163, no. 1, pp. 8–15, 2003.

[5] S. Conolly, D. Nishimura, and A. Macovski, "Optimal control solutions to the magnetic resonance selective excitation problem," *IEEE Transactions on Medical Imaging*, vol. 5, no. 2, pp. 106–115, 1986.

[6] B. Bonnard, O. Cots, S. J. Glaser, M. Lapert, D. Sugny, and Y. Zhang, "Geometric optimal control of the contrast imaging problem in nuclear magnetic resonance," *IEEE Transactions on Automatic Control*, vol. 57, no. 8, pp. 1957–1969, 2012.

[7] B. Bonnard, A. Jacquemard, and J. Rouot, "Optimal control of an ensemble of Bloch equations with applications in mri," in *2016 IEEE 55th Conference on Decision and Control (CDC)*, 2016, pp. 1608–1613.

[8] S. Wang and J.-S. Li, "Free-endpoint optimal control of inhomogeneous bilinear ensemble systems," *Automatica*, vol. 95, pp. 306–315, 2018.

[9] X. Chen, "Ensemble observability of bloch equations with unknown population density," *Automatica*, vol. 119, p. 109057, 2020.

[10] V. Bogachev, *Measure Theory*. Springer Berlin Heidelberg, 2007, vol. 1.

[11] L. Ambrosio and G. Savaré, "Gradient flows of probability measures," in *Handbook of differential equations: evolutionary equations. Vol. III*, ser. Handb. Differ. Equ. Elsevier/North-Holland, Amsterdam, 2007, pp. 1–136.

[12] G. Cavagnari, A. Marigonda, K. T. Nguyen, and F. S. Priuli, "Generalized control systems in the space of probability measures," *Set-Valued and Variational Analysis*, vol. 26, no. 3, pp. 663–691, 2018.

[13] J. T. Chang and D. Pollard, "Conditioning as disintegration," *Statistica Neerlandica*, vol. 51, no. 3, pp. 287–317, 1997.

[14] M. Staritsyn, N. Pogodaev, R. Chertovskih, and F. L. Pereira, "Feedback maximum principle for ensemble control of local continuity equations: An application to supervised machine learning," *IEEE Control Systems Letters*, vol. 6, pp. 1046–1051, 2022.

[15] M. Staritsyn, N. Pogodaev, and F. L. Pereira, "Linear-quadratic problems of optimal control in the space of probabilities," *IEEE Control Systems Letters*, vol. 6, pp. 3271–3276, 2022.

[16] B. Bonnet, C. Cipriani, M. Fornasier, and H. Huang, "A measure theoretical approach to the mean-field maximum principle for training neurodes," *Nonlinear Analysis*, vol. 227, pp. 113–161, 2023.

[17] N. Pogodaev, M. Staritsyn, and R. Chertovskih, "Optimal control of nonlocal continuity equations: numerical solution (submitted to Applied Mathematics & Optimization)," 2022. [Online]. Available: https://arxiv.org/abs/2212.06608

[18] N. Pogodaev and M. Staritsyn, "Impulsive control of nonlocal transport equations," *Journal of Differential Equations*, vol. 269, no. 4, pp. 3585–3623, 2020.

[19] J.-S. Li, I. Dasanayake, and J. Ruths, "Control and synchronization of neuron ensembles," *IEEE Transactions on Automatic Control*, vol. 58, no. 8, pp. 1919–1930, 2013.

[20] E. Zuazua, "Averaged control," *Automatica*, vol. 50, no. 12, pp. 3077–3087, 2014.

[21] A. Bressan and B. Piccoli, *Introduction to the mathematical theory of control*, ser. AIMS Series on Applied Mathematics. American Institute of Mathematical Sciences (AIMS), Springfield, MO, 2007, vol. 2.

[22] A. Agrachev and Y. Sachkov, *Control Theory from the Geometric Viewpoint*, ser. Control theory and optimization. Springer, 2004.

[23] P. Cardaliaguet, F. Delarue, J.-M. Lasry, and P.-L. Lions, *The Master Equation and the Convergence Problem in Mean Field Games*, ser. Ann. Math. Stud. Princeton, NJ: Princeton University Press, 2019, vol. 201.

[24] B. Bonnet and H. Frankowska, "Necessary optimality conditions for optimal control problems in Wasserstein spaces," *Applied Mathematics & Optimization*, vol. 84, no. 2, pp. 1281–1330, Dec 2021.

[25] A. Krasovskii and N. Krasovskii, *Control Under Lack of Information*, ser. Systems & Control: Foundations & Applications. Birkhäuser Boston, 2012.

[26] B. Tahayori, L. Johnston, I. Mareels, and P. Farrell, "Novel insight into magnetic resonance through a spherical coordinate framework for the bloch equation," vol. 7258, 02 2009.