IPEPR D'UNE STRATÉGIE NATIONALE
PROJET CIBLÉ
2022

PEPR IA

DOCUMENT PRÉSENTATION PROJET

PDE-AI

Acronyme	**PDE-AI**		
Titre du projet en français	*Analyse numérique, contrôle optimal et transport optimal pour l'IA*		
Titre du projet en anglais	***Numerical analysis, optimal control and optimal transport for AI***		
Mots-clefs	Analyse non-linéaire et équations aux dérivées partielles, analyse numérique, contrôle, optimisation, apprentissage, réseaux de neurones		
Établissement porteur	Université Paris-Dauphine, PSL		
Responsable du projet	**Prénom, Nom, Qualité**		
	Chambolle, Antonin, Directeur de recherches CNRS		
	Courriel	**Téléphone**	
	chambolle@ceremade.dauphine.fr	**0144054621**	
Durée du projet	48 mois		
Aide totale demandée	4 099 811 €	**Coût complet**	14 534 490,57 €

Liste des établissements du consortium :

Établissements d'enseignement supérieur et de recherche	***Secteur(s) d'activité***
Université Paris Sciences et Lettres	*EPSCP - Enseignement Supérieur*
Université Paris-Cité	*EPSCP - Enseignement Supérieur*
Sorbonne Université	*EPSCP - Enseignement Supérieur*
Université Paris-Saclay	*EPSCP - Enseignement Supérieur*
Institut National Polytechnique de Toulouse	*EPSCP - Enseignement Supérieur*
Université de Bordeaux	*EPSCP - Enseignement Supérieur*
CREST-GENEST (ENSAE)	*EPSCP - Enseignement Supérieur*
Université de Strasbourg	*EPSCP - Enseignement Supérieur*

Organismes de recherche	***Secteur(s) d'activité***
CNRS (DR7)	*EPST - Recherche*
INRIA (Nice)	*EPST - Recherche*

Résumé du projet en français (Non Confidentiel – 4000 caractères maximum, espaces inclus)

L'analyse mathématique, en particulier les équations différentielles partielles (EDP), le contrôle optimal et le transport optimal, devient de plus en plus importante dans le domaine de l'apprentissage automatique. En témoignent le nombre croissant d'articles de recherche et de conférences consacrées à ce sujet, ainsi que l'impact des nouvelles architectures et leur analyse théorique. Si des techniques classiques de contrôle optimal sont adaptées avec succès aux réseaux neuronaux, il reste encore beaucoup à faire pour relier les architectures existantes à des modèles mathématiques souvent idéalisés. En outre, les capacités des architectures actuelles à modéliser des structures complexes, à préserver les structures et invariances physiques ou géométriques et à fournir des estimations d'erreur ou de confiance dans leurs résultats sont limitées.

La communauté mathématique française a été active dans la recherche sur l'apprentissage automatique, mais s'est surtout concentrée sur les statistiques et l'optimisation, négligeant des questions importantes liées à la dynamique et à l'architecture de l'apprentissage automatique. L'objectif principal de cette action est de soutenir la création d'un groupe de mathématiciens appliqués spécialisés dans les questions d'apprentissage automatique et de stimuler la fertilisation croisée entre les EDP, le transport et le contrôle, conduisant à de nouvelles architectures, à des schémas d'optimisation stables et à des solveurs et approximations efficaces.

Cette action rassemblera dix groupes d'analystes actifs dans ce domaine, sous la coordination d'A. Chambolle à Paris-Dauphine (PSL). Elle consistera principalement à financer des actions de recherche (projets de post-doc et de thèses) dans trois directions principales : un premier thème portera sur l'analyse des méthodes d'apprentissage (théoriques et pratiques), un second sur les nouvelles architectures profondes (où l'on étudiera en particulier des architectures spécifiques telles que les PINNs, l'analyse numérique et les questions d'approximation, ou le point de vue contrôle optimal sur les grands réseaux), et un dernier angle de recherche se concentrera sur les méthodes génératives et la diffusion d'un point de vue analytique.

Résumé du projet en anglais (Non Confidentiel – 4000 caractères maximum, espaces inclus)

Mathematical analysis, especially Partial Differential Equations (PDEs), optimal control, and optimal transport, is becoming increasingly important in the field of machine learning. This is evidenced by the growing number of research papers and dedicated conferences on the topic, and the impact of new architectures and theoretical analyses. While there have been successful adaptations of classical optimal control techniques to neural networks, there is still much work to be done in connecting existing architectures to idealized mathematical models. Additionally, there are limitations in current architectures' abilities to model complex structures, preserve physical or geometric structures and invariances, and provide error estimates or confidence in their output. The French mathematical community has been active in machine learning research but has focused mostly on statistics and optimization, overlooking important issues related to machine learning dynamics and architecture. The main purpose of this action is to support the creation of a group of applied mathematicians specializing in machine learning issues and to stimulate cross-fertilization between PDEs, transport, and control, leading to new architectures, stable optimization schemes, and efficient solvers and approximations. This action will gather ten groups of analysts which have been active in this field, under the coordination of A. Chambolle at Paris-Dauphine (PSL). It will consists mostly in funding research actions (post-doc and PhD projects) in three main directions: one first theme will focus on the analysis of training methods (theoretically and practically), a second on new deep architecture (where in particular we will study specific networks such as PINNs, numerical analysis and approximation issues, or the optimal control point of view on large networks), and a last angle of research will focus on generative methods and diffusion from an analysis point of view.

Table of contents / Table des matières

1. Context, objectives and previous achievements

1.1. Context, objectives and innovative features of the project

Mathematical analysis, and in particular Partial Differential Equation (PDEs), optimal control and optimal transport, is now at the heart of machine learning. The growing number of analysis and PDE research papers dealing with machine learning issues or incorporating machine learning ideas is testifying for this trend, as well as the recent foundation of dedicated conferences such as the MSML (Mathematical and Scientific Machine Learning) This is especially the case in three key areas: the dynamics of neural networks (NNs) training, the design of very deep architectures and generative model sampling. This resulted over the last few years into a fantastic impact of new architectures and theoretical analyses driven by the interface between machine learning and the mathematical analysis of dynamical processes. This includes for instance the development of neural Ordinary Differential Equations (ODEs) methods (Chen et al., 2018, best paper award Neurips'18) which are direct adaptations of classical optimal control techniques (adjoint equations/states). Other striking examples include the use of optimal transport methods for generative modeling (Arjovsky et al., 2017) and in-depth analyses of the dynamics of training single hidden layer networks (Chizat and Bach, 2018). These mathematical approaches are however at their infancy: there is still a large gap to connect the most successful residual networks or transformer architectures to their idealized mathematical dynamical models. Moreover, while the mentioned architectures have demonstrated excellent empirical performances on vision and natural language processing tasks, they are still limited on many some aspects. Indeed, they struggle to model multimodal probability distributions with complex structures, their ability to preserve physical or geometric structures and invariances is limited and, the lack of error estimates or confidence of their output is critical for some applications. Also, the optimization of these networks remains particularly difficult and unstable. Beside these theoretical bottlenecks, new architectures are still to be discovered by leveraging successful existing approaches in optimal control and transport.

The French mathematical community is already very active in machine learning, but mostly focuses on research in (theoretical and applied) statistics and optimization, overlooking some important issues relative to the dynamics of machine learning systems, their approximation properties or their architecture. On the applied analysis side, the French community is very strong on these themes, but to a large extent is not yet focused on the theoretical questions related to machine learning. The main purpose of this action is to support the creation of a strong group of applied mathematicians specialists of analysis, numerical analysis and control, focusing on machine learning issues. This community will be able to stimulate the interactions between these three types of tools (PDEs, transport and control), and this cross-fertilization will bring major contributions for the theory of machine learning. On the theoretical side, interpreting very deep networks as time evolution and training dynamics as differential equations will open the door to a sharp analysis of the convergence properties as well as a better understanding of the expressiveness of these networks. On the more practical side, the analogy with optimal transport flows and optimal control problems, as well as the need for reliable and efficient solvers and approximations in numerical analysis, will drive the emergence of new class of architectures, more stable accelerated optimization schemes, and sampling techniques that can better cope with high dimensional feature spaces.

1.2. Main previous achievements

The consortium gather ten groups of analysts, numerical analysts and applied mathematicians who have all explored, although for some of them still superficially, machine learning techniques and in

particular NNs and their optimization, either as a topic of study or as a tool for the approximation of functions or probability distributions. Among these, sampling and approximation issues, often based on Wasserstein metrics for probability measure, is a part which is common to many groups of the consortium (Lebrat et al., 2019; Bigot et al., 2017; Delon and Desolneux, 2020; Cazelles et al., 2018; Mérigot et al., 2021; Xu et al., 2022) and will serve as a basis for the research which will be conducted in "Theme 3" of this proposal, see next Section. This is also naturally related to "Theme 1", and even more many groups (notably in Toulouse, Bordeaux, Paris, Nice) have a well-acknowledged experience in optimization for learning or imaging (Aujol et al., 2019; Traonmilin and Aujol, 2020; Caillau et al., 2022; Korba et al., 2020; Xu et al., 2022; Huix et al., 2022; Laumont et al., 2022b, 2023; Bergou et al., 2022; Chambolle and Pock, 2016; Chambolle and Contreras, 2022) and in the design and analysis of accelerated (stochastic) first-order methods. Together with the expertise of some of the other groups (such as Paris-Saclay) on Wasserstein flows and optimal transportation, these knowledge and tools will ensure quick progress on the analysis of the dynamics of NNs training, and on the efficiency of sampling schemes (see (Xu et al., 2022)). Mean-field dynamics and mean-field games (Hu et al., 2021; Conforti et al., 2022; Achdou and Laurière, 2020) are the domain of expertise of a few members of the consortium and will be of great importance (as illustrated in the celebrated reference (Chizat and Bach, 2018)) for the study and development of learning algorithms.

As for "Theme 2", it is more related to our experience and developments in approximation theory and numerical analysis (Després and Ancellin, 2020; Després, 2022; Després, 2022; Bretin et al., 2022; Chambolle and Pock, 2016; Caroccia et al., 2020; Chambolle and Pock, 2021a,b). In this branch, the need for reliable architectures and strategies is a source of considerable innovation in the design of new networks or learning methods for various purposes, whether it be the design of functions/networks with particular properties (Bounded Hessian, Lipschitz, see (Ambrosio et al., 2023) for instance), or as a tool for designing fast solvers for high-dimensional PDEs with convergence guarantees. An important direction developed at Sorbonne-Université is the study of reference solutions for residual and recurrent retworks (RNNs), with a particular emphasis on connection with Finite Elements theory (Després, 2022; Després, 2022). Learning is also of great interest for approximating complex data (such as geometric low dimensional objects in analysis, or graph structures) and some groups (Lyon, Orsay, Paris) wish to collaborate to further develop some of their work on such topics (Digne et al., 2018; Buet and Rumpf, 2022; Chambolle and Pock, 2021a; Bogensperger et al., 2022; Chambolle and Pock, 2021b) (also related to surface evolution using neural-based diffusion-thresholding algorithms (Bretin et al., 2022)) and variational limits (Caroccia et al., 2020) of graphs in data processing.

2. Detailed project description

2.1. Project outline, scientific strategy

This project will develop new mathematical approaches at the interface between numerical analysis, optimal transport and optimal control to study the dynamics of neural networks for training and sampling. These theoretical breakthroughs will also lead to the design of new architectures with mathematical guarantees.

Theme 1 - The dynamics of neural network training. The current workhorse of both supervised and unsupervised learning are gradient-based optimization schemes. A promising approach to understand their favorable convergence properties is to study PDEs emerging as a time-continuous and infinite width (large number of neurons) limit mean field equations. This analysis is however currently mostly restricted to simplistic single hidden layer shallow neural networks. A first area of research of this project is to study the extension of this framework to deeper networks, in particular those

associated to residual networks (ResNet). For sequential data, as in natural language processing, state-of-the-art performance is attained by Transformers networks, which uses attention mechanisms operating over high dimensional point clouds. The mean field dynamics associated with these architectures are highly non-linear and include non-local interactions. Treating these evolutions as Wasserstein gradient flows offers a convenient mathematical framework to study both the convergence of these dynamics and the expressivity of the associated architectures. A second and parallel research thread will go beyond classical Euclidean gradient flows to take into account state of the art solvers such as ADAM, which are usually used as black box optimization schemes with limited theoretical guarantees. Re-casting ADAM as a non-Euclidean stochastic gradient flow will open the door to new mathematical analyses and innovative stabilization methods. This will follows and extends a successful stream of research currently limited to Euclidean accelerated first order methods such as Nesterov schemes (Aujol et al., 2019) and will also benefit from the integration of second order information. ADAM-type solvers are also very successful to solve non-convex min/max problems to train adversarial networks. In the infinite width limit, this draws a link with mean-field games approaches, a far-reaching generalization of optimal transport flows.

Theme 2 - New very deep architectures. The numerical analysis of PDEs has recently emerged as an important source of inspiration to design new very deep residual architectures leveraging stability properties of ODEs and PDEs discretization schemes. Our consortium has a strong expertise in this area and will develop new networks emerging as recurrent networks implementing fixed point equations with strong contraction properties. The exactness of this construction is known to hold in 1D (Després and Ancellin, 2020). and an important avenue of research to study the properties of these constructions in high dimensions (using spectral properties of linearized operators similarly to preconditioning techniques) and to draw links with reduced modeling. The design and analysis of these very deep architectures can also be tackled using continuous time dynamics, which is fully inline with optimal control methods (e.g., Model Predictive Control). They can reduce the memory footprint of backpropagation by solving the adjoint ODEs without storing the activations. This still raises major difficulties to cope with instabilities in high dimensions, and the members of our groups will study various mitigation and approximation techniques strongly related to automatic differentiation and control (Caillau et al., 2022). Our strong expertise in optimal control will also be used to cope with the huge computational time of reinforcement learning problems. In particular, we will focus on reduced modeling approaches which have proved successful for the training of physical devices and will be scaled to high dimensional problems. A last and somewhat complementary question is the study of geometric deep learning, where the goal is to extend these very deep networks to cope with non-Euclidean data, the most iconic examples being graph neural networks. The expertise of our members on optimal transport (Delon and Desolneux, 2020; Cazelles et al., 2018) or geometric measure theory (Caroccia et al., 2020; Chambolle and Pock, 2021a) will be leveraged to propose new geometric loss functions which can be used to train these very deep architectures on point clouds, graphs and surfaces, possibly jointly in several high dimensional feature spaces to cope with multimodal data. Extending differential operators and pooling layers to these non-Euclidean domains will be achieved by leveraging tools and expertise of our researchers in computer graphics (Digne et al., 2018) and optimal control for shape optimization.

Theme 3 - Sampling and generative models. Dynamical models such as normalizing flows are now central for unsupervised learning, leveraging deep models to perform high dimensional sampling. A first fundamental sets of questions are linked with optimal quantization of high dimensional distributions using particle systems. Despite being highly non-convex, optimal transport methods recover some hidden convexity (Mérigot et al., 2021) and show promising global optimality results of the as-

sociated sampling flow (such as Lloyd diffusion). This is however restricted to simple settings, and we will consider problems involving robust cost functions used for Generative Adversarial Networks (GANs) training, barycenter and non-linear PCA computation (Cazelles et al., 2018). This will require the design of a new class of convex potentials that can drive dynamical sampling processes on non-Euclidean domains and manifolds. Alongside these quantization questions, understanding the ability of generative networks (GANs, VAE, normalizing flows and score matching) to model disconnected or multimodal distributions will be at the heart of our research agenda to establish statistical bounds on the optimal transport distance toward the target distribution. In parallel to these theoretical guarantees, we will also develop new sampling strategies leveraging recent classes of optimal transport particle flows (Korba et al., 2020). These new flows will integrate perturbed dynamics (noise injection, momentum, acceleration) and modified geometries (such as Wasserstein Fisher Rao and kernelized Hilbert norms) together with new time discretization to enable provably efficient sampling techniques.

Shaping a community. Beside these three interleaved themes, the promise of this project is to bring together a community which will hybridize and extend tools from different mathematical backgrounds to tackle difficult questions in Artificial Intelligence. We will use this project to organize focused workshops and larger events to help the French mathematical community to gain momentum and make PDEs a central topic in modern machine learning.

2.2. Scientific and technical description of the project

The project is split into three themes, which explore complementary aspects of the design, optimization and sampling of new neural network architectures. Each objective involves a few teams from the consortium specialized on the topics of interest. We will organize on a bi-yearly basis small meetings to exchanges altogether.

Theme 1 – The dynamics of Neural Networks training

In this first part of the proposal, we will explore the interface between gradient based optimization and optimal control to develop new optimization solvers, which are integrated into efficient, GPU-friendly, computational toolbox.

Objective 1.1 – *Non-convex optimization* (participants: Bordeaux, Paris-Saclay, Dauphine) Non-convex optimization is the work horse of most recent developments in machine learning. The success of stochastic gradient-based optimisation method in these settings remains a mathematical challenge. Understanding the benefit some popular methods (such as ADAM), improving these approach, and exploring what happens in very wide and deep network require to develop new mathematical tools. This task will leverage tools from dynamical systems and mean field analysis to gain a better understanding and develop new scalable solvers.

Task 1: Non convex first order optimization. Efficient optimization algorithms have played a key role in the recent development of machine learning, and they are especially paramount to every deep learning approach, from convolutional networks to the most sophisticated GANs (Generative Adversarial Networks) whose training can be boiled down to some high dimensional optimization problem involving complex objective functions. Over the last four years many papers (Apidopoulos et al., 2018; Aujol et al., 2019; Su et al., 2016) proposed to study convex optimization algorithms such as Gradient Descent or Inertial Gradient Descent, through Ordinary Differential Equations (ODEs). This point of view was helpful to understand convergence rates of these algorithms depending on the properties of the function F to minimize. **(i) Non convex optimization:** The quest for new Lyapounov function is related to the problem of being able to consider non-convex functions F. It is a main bottleneck, and

we wish to propose solutions for it. Apart from (Polyak and Shcherbakov, 2017; Apidopoulos et al., 2022), there is always at some point a convexity-like assumption that is made on F to control the Lyapounov function. A key question is then how to extend these works to incorporate a fine geometric assumption on F and to get optimal convergence rates. **(ii) ADAM algorithm:** We will investigate the ADAM algorithm (adaptive moment algorithm) (Kingma and Ba, 2014) which is the most used in deep learning solvers. It is an inertial algorithm with a flavour of stochastic gradient descent. Most of the time, this algorithm is used as a black box. This lead to problems in term of controlling the quality of the obtained results. Moreover, even if the ADAM algorithm is very efficient, it may diverge even in the case of a convex function (Reddi et al., 2019). Many modifications have been proposed (Défossez et al., 2020). But so far understanding the role of inertia to escape saddle points and to compute local minima of nonconvex functional has not been fully understood (Xie et al., 2022).

Task 2: Particles-based NN training. Training a single hidden layer NN can be seen as a convex minimization problem on a space of measures (Bengio, Le Roux, Vincent, Delalleau, and Marcotte, 2005). The dynamics of standard optimization algorithms on this space provide a useful reference point to understand the dynamics of network training, notably the convergence properties (Chizat and Bach, 2018). In practice, however, NNs cannot explore the whole space of measures: they are restricted to finite sums of Dirac measures. On this discrete space, the convergence guarantees are weaker than for general measures; results from (Chizat, 2022), for instance, only apply if the number of Dirac measures is very large. Similar issues arise in the simpler setting of super-resolution, when the goal is to recover a sparse measure from low-frequency measurements. Under weak assumptions, this can be achieved by solving a mass minimization problem on the space of measures (De Castro and Gamboa, 2012). In practice, some form of discretization is necessary to numerically solve the problem. The leading algorithm, *Sliding Frank-Wolfe*, does this with a combination of convex and non-convex steps, the non-convex ones being similar in spirit to NN optimization (Denoyelle, Duval, Peyré, and Soubies, 2019). Our first goal is to develop a refined understanding of the convergence rate of these algorithms which tackle minimization problems on a whole measure space by optimizing only over the subset of finite sums of Dirac measures. We will first focus on *Sliding Frank-Wolfe*. This algorithm is likely to be slightly easier to analyze than its fully non-convex relatives. Indeed, the convex steps help driving the iterates towards a global minimizer, which is hinted at by the numerical observation that *Sliding Frank-Wolfe* does not seem to require overparametrization, while fully non-convex algorithms do. In addition, the setting of super-resolution problems allows for the use of powerful harmonic analysis tools through the machinery of dual certificates. In the longer term, we will try to adapt our findings to other problems such as NN training. This topic will be studied with a post-doc student hired at Paris-Dauphine, working with I. Waldspurger and C. Royer.

Task 3: Mean-field optimization and NN training. Training a neural network can be viewed, in an infinite width limit, as a mean field optimization problem. It can be shown that the solution of this optimization, regularized by relative entropy, is an invariant measure of a system of mean-field Langevin (MFL) equations. This system can be viewed as a continuous-time version of a numerical algorithm for solving the optimization problem. Proving convergence results for the system of MFL equations can thus give a better understanding of convergence properties of algorithms for neural networks. We can generalise the above setting by considering games regularised by relative entropy with players' strategies coupled through a random environment (Conforti et al., 2022). This kind of generalisation allows to describe a wider range of machine learning tasks, in particular, supervised learning or training a generative-adversarial neural network. It may also be informative to consider different regularizers in the optimization problem. The solution to the optimization problem with Fischer information regulariser can be viewed as the invariant measure of mean-field Schrödinger (MFS) equations (Claisse et al.,

2023). However, the solution of MFS being hard to simulate, we plan to explore different algorithms (fictitious play, mirror descent) to approximate the solution of the optimization problem. We expect this study to suggest new algorithms for machine learning tasks, and to give insight into their convergence properties. (PhD thesis or post-doc supervised by A. Kazeykina, Paris-Saclay and Z. Ren and P. Cardaliaguet, Dauphine).

Objective 1.2 – *Control and machine learning: there and back again* (participants: Strasbourg, Nice) A mostly open question in large scale control for learning is to solve problems in which the space of controls (actions) is large and difficult to characterize numerically and thus to explore randomly.

Task 1: Toward new exploration strategies. We propose to investigate gradient and model learning methods in order to build new exploration approaches aiming at generating physically realistic and time regular controls. With standard reinforcement methods (Sutton and Barto, 2018; Lillicrap et al., 2016), due to the random search methods of the controls, the obtained functions are often not very smooth and therefore not easily usable. In particular, we plan to implement such methods on physical systems, especially to solve optimal control problems in quantum control. A PhD in Strasbourg is planned to work on these aspects.

Task 2: Control for neural network analysis. Conversely, control has recently proved very instrumental for the analysis of neural networks, and for their use in classification, *e.g.* When dealing with very deep networks, a relevant approximation is to assume there is a continuum of layers, indexed by time. In this framework, the composition of a finite number of cells can be interpreted as discretising a continuous ODE, called the neural ODE. In turn, for a large number of layers, the continuous model is meaningful to understand the network properties. The weights, labeled by time, are the controls of the system, and there are several issues that can be efficiently addressed through this point of view. First, in the case of supervised learning: learning a classifier is just learning some input-output map, a standard task in control theory. And meeting the requirements tied to the known data is associated with controllability issues, namely *ensemble controllability*, ultimately related to controllability on the group of diffeomorphisms on the ambient manifold (Agrachev and Sarychev, 2022). Lie bracket techniques are in order for these questions, as well as other more constructive approaches that exploit the structure of the nonlinear activation functions of the original network (Zuazua, 2022; Ruiz-Balet and Zuazua, 2022). Another line of active search to contribute to is the analysis of the convergence of training of NN, modelled through by neural ODEs, and its relation to the well known turnpike phenomenon in optimal control (Geshkovski and Zuazua, 2022). This task is closely related to the Task 2 on "Training algorithms for discretized neural ODE" of Objective 2.1 which focuses on the network architecture. Application to renewable energy production (KEEP project, Kite Electrical Energy Production) will be considered as a benchmark for this task. Learning techniques will be compared to more traditional limit cycle estimation for parametric dynamical systems, ensuring that the problem is addressed at least in a control-centric framework. As these questions are rather advanced an require a good background at the crossroads of dynamical systems, optimisation and machine learning, we plan to hire a 24 month postdoc in Nice. Interaction with Strasbourg colleagues on control applications is also expected (Task 1).

Objective 1.3 – *Scalable solvers and softwares* (participants: U. Paris Cité , Nice) From a computational perspective, the success of deep architecture in ML is due to the combination of large scale GPU computations with automatic differentiation library. These approaches however quickly hit a wall in term of memory usage, which is often mitigated by the use of slower computational routine. In this part, we will explore new approach to improve this memory/computation tradeoff, which will be integrated in efficient open source software packages.

Task 1: Large scale GPU primitives with KeOps. A guiding principle of this proposal is the coupling of new efficient numerical solvers (in this theme) with new model architectures (in theme 2), that may combine strong theoretical guarantees with effective performance on real data. In this context, we intend to provide optimal software implementation for our main "numerical tools" that are not supported well by major software packages: the computation of optimal transport plans, operations on non-Euclidean distance matrices, convolutions on (high-dimensional) point clouds and kernel matrix-vector products. This is precisely the goal of the KeOps library ((Charlier et al., 2021; Feydy et al., 2020), https://www.kernel-operations.io), which provides a mathematician-friendly interface to state-of-the-art numerical schemes on the GPU. Its main user interface uses the concept of lazy tensors that allows to use it almost seamlessly in a Python script, as if manipulating actual dense tensors. Since its first release in 2018, it has been downloaded more than 400k times and has been featured in several works in e.g. geometric deep learning (Sverrisson et al., 2021; Kazi et al., 2023), kernel methods(Meanti et al., 2020), Gaussian processes (Gardner et al., 2018), discrete optimal transport (Feydy et al., 2019). We intend to recruit a research engineer for 2 years to help us with further development of the library.

Task 2: Automatic differentiation and control. More and more innovative algorithms in the field of AI and its declinations in optimal control or transport require the use of automatic differentiation (AD): schematically, the minimization of a loss function in machine learning (ML) necessitates the computation of a direction of descent, generally associated with a more or less weak notion of derivative. The estimation of this derivative classically requires to differentiate the calculation produced by an algorithm. Automatically generating the algorithm for calculating this derivative often has many virtues (efficiency, numerical robustness, *etc.*) Although relatively old, the technical ecosystem of AD has undergone an extremely important revival over the last decade with the explosion of problems and techniques associated with machine learning. A quick review of the state of the art allows us to cite historical codes such as Tapenade, developed at Inria Sophia, and numerous more recent achievements such as TensorFlow, Pytorch, JAX or Zygote, directly resulting from the work of the machine learning community. In our academic context, we think it is relevant to cross this theme with another discipline well represented in Nice-Sophia: proof of code in language theory. The numerical aspects being crucial for the quality of the results provided by an algorithm, proving that these differentiation methods are correct will give an enormous added value to this work, which does not yet exist in the academic world. A postdoc recruitment will allow to focus initially on an apprehension and a criticism of the automatic differentiation tools in close connection with the fundamental algorithms of optimal control and optimal transport: this work which mixes numerical analysis, control and transport, is at the heart of this targeted project of the PEPR IA. When relevant, the performance of these tools will be evaluated for a specific and common use case (definition of the use case for the evaluation and benchmarking). At the end of this critical phase, if the evaluation is encouraging, we first plan to apply and optimize the relevant tool in optimal control (we already have many cases thanks to the ongoing developments of the control-toolbox project), optimal transport and their applications in learning. This action will naturally be linked to the work of the MAP5 group (Paris) and the KeOps library. The recruitment of a research engineer, also competent in formal aspects related to the proof of computation, and in language issues (code generation), will complete this work (an overlap of 6 months with the postdoc is planned). Let us note from this point of view the recent importance for automatic differentiation of approaches that operate in a language-agnostic way, directly at the level of an Intermediate Representation (LLVM code in the case of the Enzyme project, for example). We are aware that such engineering profiles are rather rare, that is why this recruitment is planned for the end of the project. The recruited engineer will benefit from the SED (dev center of Centre Inria d'Université Côte d'Azur) environment to get rapidly acquainted with dev techniques.

Theme 2 – New very deep architectures

Deep neural architecture have been proven to be efficient to solve specific tasks, where the structure of the network is tailored to capture the geometry and invariances of the problem under study. This is especially true when dealing with very deep architectures, which are notoriously difficult to train in a stable and scalable way. In this theme, we will focus on the efficient training of very deep architectures, which can integrate geometric constraints routinely encountered in shape optimization problems and the solution of partial differential equations.

Objective 2.1 – *Training and calibrating neural differential architectures* (participants: Dauphine, Strasbourg) Neural differential equations are a recently proposed learning model in which the output is obtained as the solution of a differential equation. Such a model highlights connections between residual architectures and ordinary differential equations (ODEs) on one hand, and between convolutional layers and partial differential equations (PDEs) on the other hand (Chen et al., 2018; Haber and Ruthotto, 2018; Ruthotto and Haber, 2020). Training these neural differential architectures amounts to solving a constrained optimization problem where the constraints are expressed as ODEs or PDEs. Such formulations bear a strong connection with that arising in scientific computing, where one typically minimizes a given cost function under physical constraints represented by ODEs or PDEs (Antil et al., 2016). In fact, recent work suggests that approaches inspired by scientific computing possess an edge over traditional training methods based on stochastic gradient when the underlying data comes from a physical system (Newman et al., 2021). Still, a number of challenges have yet be overcome in order to deploy numerical algorithms used in scientific computing at the deep learning scale. This objective will tackle the development of practical optimization algorithms for training such neural architectures, as well as the calibration of such algorithms in connection with tuning the architecture itself.

Task 1: Training algorithms for implicitly constrained Neural PDE models. Our first task will tackle problems involving Neural PDEs by considering the associated constraints as implicit, i.e. that the objective is implicitly defined through the solution of a PDE. Under this assumption, we will develop second-order algorithms dedicated to solving these problems, in a departure from the first-order methods considered in Objective 1.1. To mitigate the cost of applying these techniques, we will rely matrix-free methods, that do not compute full second-order derivative matrices but rather rely on matrix-vector products (Antil et al., 2016; Biros and Ghattas, 2005; Heinkenschloss and Ridzal, 2014). By leveraging existing work on matrix-free techniques in nonconvex optimization (Curtis et al., 2021), we expect to develop methods with both theoretical guarantees, such as complexity bounds Cartis et al. (2022), as well as practical appeal. Our main goal will be the design of a matrix-free Newton-type method with complexity guarantees for PDE-constrained optimization problems with a stochastic objective, and its empirical validation on training neural differential architectures. Deriving complexity results in the context of implicitly constrained optimization would be a key contribution to the field, while strong numerical evidence of the method's performance would confirm recent findings on the performance of second-order methods Newman et al. (2021).

Task 2: Training algorithms for discretized neural ODEs. In this task, we will consider implementations of neural ordinary differential equations wherein the dynamics are discretized using a numerical approximation. This setup is typically that used in practice, often by relying on toolboxes from numerical analysis Chen et al. (2018); Queiruga et al. (2020). In that context, the problem naturally becomes a highly structured constrained optimization problem, and can be viewed as an optimal control task Haber and Ruthotto (2018). We will leverage this observation to design tailored algorithms for training discretized Neural ODEs. Similarly to task 1, those techniques will rely on second-order information in a matrix-free fashion, and will handle stochasticity in the objective function. We will

investigate several discretization techniques proposed in the literature Lu et al. (2018), with the goal of proposing a method that adapts to the discretized architecture at hand, while being endowed with novel complexity guarantees. By leveraging linear algebra and optimization techniques popular in optimal control Betts (2020), we will also seek practical efficiency in our implementation.

Task 3: Calibration of neural differential architecture training and design. Implementing neural differential architectures requires to set multiple degrees of freedom in the actual design, typically related to the discretization operator used to implement the PDE or ODE calculations. Choosing the best architecture for a given task can be posed as an optimization problem in which the variables are the hyperparameters defining the network (as opposed to the parameters of the network that are learned through training). Meanwhile, training algorithms often involve algorithmic hyperparameters that can greatly influence the performance of the method. Tuning those hyperparameters represents a modern challenge towards automated use of neural networks. In this task, we will develop optimization techniques to calibrate both the hyperparameters defining the training algorithm and those describing the neural differential architecture at hand. Although those two tasks have been studied independently, our algorithmic proposal will be the first to tackle both types of hyperparameters at once. We will rely on blackbox (or derivative-free) optimization techniques, that have been previously used to optimize parameters of discretization schemes arising in scientific computing **?**. Building on recent advances in that field Chen et al. (2018); Bergou et al. (2022), we will develop new algorithms supported by theoretical guarantees such as complexity results and demonstrated practical performance.

The research on Tasks 1 and 2 will be mainly carried on at Paris-Dauphine through the hiring of a PhD student (co-supervision Chambolle/Royer). Possible collaboration is envisaged with I. Mazari, E. Franck, Y. Privat, as the results in this objective may be also interesting for control based methods, *cf* Obj. 4. Task 3 will be lead by C. Royer, and is meant to intersect with the developments and collaborators involving in both Tasks 1 and 2.

Objective 2.2 – *Neural methods and finite elements* (participants: Sorbonne-Univ., Nice) The development and study of references solutions for Recurrent Neural Networks (RNNs) is an emerging field of research (Després and Ancellin, 2020; Després, 2022; Després, 2022). This analysis requires to determine the structure of specific RNNs which fit into the general theory of Neural Networks and for which one can trigger the coefficients in order to define families of solutions with different regularities. Based on our recent research, we will determine the best strategy and to specify an applicative scenario to evaluate the advances and make comparison with the literature. The scientific issue is to develop a comprehensive theory of approximation of polynomial functions and general function based on these principles, and to exploit these exact NN functions in order to investigate the accuracy of training algorithms. More precisely, we will address the following goals.

Task 1: Takagi generated Neural Networks and preconditioning. We will first develop a new theory of Takagi generated Neural Networks, following (Després and Ancellin, 2020; Després, 2022). The exploration of the approximation properties of Neural Networks will be studied in collaboration with A. Cohen and Katharina Schratz (LJLL, Sorbonne U.). Then we will study of precondioning techniques generated with Machine Learning and Neural Networks (w. Frédéric Nataf, LJLL). We aim at exploring the use of NN for the construction of matrices which are a good approximations of the inverse of the matrix of a specific linear system. The interest is that the accuracy of the general procedure is guaranteed by standard results on GMRES type convergence results, with a control of the statistical noise produced by standard NNs. This direction is fundamental yet still underexploited in machine learning, and would help designing fast solvers for complex problems without sacrificing the guaranteed quality of the solutions.

Task 2: Neural finite volume schemes. We will develop new Finite Volume numerical schemes by Machine Learning (VOFML) (Ancellin et al., 2023; Després and Jourdren, 2020). We aim at develop the 3D version of this method, with incorporation into a PDE solver to run multiphase CFD scenario (in collaboration with the industry and in particular the team headed by Stéphane Jaouen at CEA). A dedicated study will be performed on the calibration of the number of moments needed to perform good accuracy for a magnetized plasma kinetic partial differential equation model. The code is currently developed at LJLL and it is expected that IA and NN will help to get a nonlinear control of the number of moments with optimal accuracy.

Task 3: Link with control theory. The last and most ambitious objective is the development and analysis of Neural Networks by means of Control Theory, using the fact that the training is very similar to the calculation of optimal controls. This is developed in Geshkovski and Zuazua (2022) and we expect that Borjan Gueshkovski will join in 2023 the pole LJLL/SCAI/INRIA, so this axis of research will be an important one for the team. We will easily find a good candidate for a PhD thesis on the approximations of polynomials with RNNs (as B. Després is teaching a course on this topic in the Master M2A at Sorbonne U.). We will also assume post-doc students to interact with the team which also involves F. Charles, A. Cohen, K. Schratz, F. Nataf. The dynamics around the SCAI center (at Sorbonne) will be a strong opportunity to establish collaboration with other teams of the consortium involved in *Control for Neural Networks*, in particular we think of joint collaboration with the Nice team and the Strasbourg team.

Objective 2.3 – *Geometric problems, shape optimization and control* (participants: Lyon, Dauphine, Strasbourg) The development of deep neural network solvers for geometric problems is an appealing alternative to traditional numerical approaches. Taking into account complex geometrical constraints formidable challenges which in turn requires to revisit some guiding principle to design neural architectures, which we explore in several steps.

Task 1: Learning techniques and neural networks for the discretization of singular geometric measures, interfaces and their geometric flows. To minimize energies of surfaces, lines, or other geometric structures, one needs an accurate description of these objects, which is sometimes contradictory with the discretization. Very recently, currents or geometric distributions have been approximated by neural functions (Palmer et al., 2022), allowing a very precise resolution of the singularities. Yet this requires a training a network for each new problem. Discretizations which allow for very sharp resolution of singularities have been studied in the context of total variation minimization (Condat, 2017; Chambolle and Pock, 2021a). In particular, (Chambolle and Pock, 2021a) requires the learning of optimal parameters for reconstructing sharp objects, which needs only to be done once for a class of problems, and relies on bi-level optimization. The generalization to complex interfacial energies (eg., Mollenhoff and Cremers (2019)) requires new representations, possibly based on local or convolutional networks, yet maintaining a structure which allows to show consistency. For more complex problems, (Bretin et al., 2022) have shown that it is possible to approximate geometric flows of (diffuse) interfaces using networks whose structure is inspired by phase field models and their numerical discretizations. The experimental results obtained so far have shown the relevance of this approach, for curvature-type flows or to compute minimal structures (both oriented and non-oriented, which is a major contribution compared to conventional methods). This raises many difficulties: how should we design the NN for a given geometric flow? How does it compare with PINNs like approaches (or "geometry aware")? Can we show convergence, stability or comparison for the obtained schemes, or how can we improve the network to ensure these properties? Can one retrieve information on the flow (mobility, surface tension, anisotropy...)? Can we infer new PDE reaction-diffusion models and new energies and can we compute their sharp interface limit? Can we extend this approach to

other representations such as point clouds, curves or parametric surfaces? A post-doc (1yr) will be hired in PSL (w. A. Chambolle) to address the static problems, and a PhD will be hired in Lyon and co-supervised jointly by E. Bretin, S. Masnou and A. Chambolle to investigate the approximation of geometric flows.

Task 2: study and enrichment of PINNs networks for optimal control and shape optimization. PINNs algorithms (Mowlavi and Nabi, 2023; Raissi et al., 2019) are promising because they can cleverly take into account EDO/EDP dynamics and can be used without introducing any mesh (gridless algorithm) and with fairly high dimensional controls. Nevertheless, they require setting parameters, they have been applied so far to rather simple problems and they are not easy to converge. The challenge is to determine new training methods and to include relevant constraints for the control/inverse problems. A complementary issue is to investigate how to combine hybrid methods between training and classical methods (gradient, Newton, fixed point) to improve both the accuracy of the control obtained and the efficiency of the algorithm. PINNs also makes it possible to search for closed-loop controls, without using the Hamilton-Jacobi equations. We also wish to develop them in this direction and study their use in this context. The case of the application in shape optimization constitutes an important challenge: indeed, shape optimization algorithms are known to converge with difficulty, they are not very parallelizable and require the implementation of costly steps at each iteration (resolution of the state, the adjoint, determination of the step). We intend to compare our approaches to classical resolutions (spectral shape optimization problems, Dirichlet energy). We also wish to develop the consideration of constraints that are difficult to handle (geometric/manufacturing constraints). A post-doc will be hired at Strasbourg on these topics.

Task 3: Learning geometries in the context of PDE constrained optimisation. We want to understand the capacity of NNs to learn geometries related to PDE constrained optimisation problems. Recent contributions have shown that NNs of moderate sizes can learn in a very accurate way purely geometric flows (Bretin et al., 2022). These flows are very closely related to thresholding schemes that also appear in connection with PDE constrained optimisation (Chambolle, Mazari, Privat, in preparation). In this setting, a example of question is: what is the optimal location of a pool of resources in order to preserve biodiversity, or where should one fish to reach a compromise between harvest outcome and the preservation of ecosystems? These problems, as they mix geometric (e.g. volume) and analytic (through the use of a PDE) information exhibit a wide range of geometrical features; a typical instance, in the setting of population dynamics, is the fragmentation of optimal location resources for low diffusivities (Mazari et al., 2022).
One focus here will be to train NNs so that, given a fixed domain (where the PDE is set) and the parameters of the PDE under consideration, they return an optimal configuration. This presents a variety of interest: first this would allow to exemplify what type of geometrical properties NNs are able to learn. In other words: can NNs learn fragmentation properties? This is a non-trivial question, as it is known that NNs tend to be biased towards low modes (Rahaman et al., 2018) — which might tilt them towards simpler geometries. Second, this would allow to provide a broader context to recent approaches in NN-based shape functional analysis (Calabrò et al., 2021). Finally, this could serve as a pre-conditioner in bi-level optimisation schemes that are off constant use in multi-players PDE constrained optimisation problems, where one rather aims at determining Nash equilibria (Mazari and Ruiz-Balet, 2022). This will be the topic of a post-doctoral research in PSL (I. Mazari, A. Chambolle), in collaboration with Strasbourg.

*Task 4: shape generative modeling with topological constraints and Neural ODEs.*Today's 3D shape generative models for shapes do not achieve the high level of realism obtained for images (Rombach et al., 2021). Indeed, shapes are non-euclidean data, for which defining an equivariant convolu-

tion remains an open problem. By leveraging flows in the ambient space, continuous normalizing flows (Yang et al., 2019) or more recently denoising diffusion models (Zeng et al., 2022) allow to synthesize new point sets resembling a set of shapes, but they often result in a set of points approximately sampling the underlying manifold. Our goal is to propose a way to synthesize surfaces by deformation of a basis surface of the right topology, by leveraging the power of Neural ODEs (Chen et al., 2018). However, instead of synthesizing surfaces in a Lagrangian formalism, simulating particles motion through the ODE, we propose instead to focus on an Eulerian approach, simulating the flow in the ambient space, and use an implicit definition for the surface so as to guarantee both the shape topology and the shape "manifoldness", while being able to synthesize it at various scales. We will hire one PhD student for this task working with J. Digne, S. Masnou and R. Chaine.

Theme 3 - Sampling and generative models

Sampling and generative modeling aim at approximating an intractable target probability distribution, and eventually generate proxy samples. This task is critical in machine learning and computational statistics. For instance, in Bayesian inference, the target distribution corresponds to the posterior distribution of the parameters, which is known only up to an intractable normalisation constant, and is needed for predictive inference. Another example is the one of neural network optimization for supervised learning, where the target distribution corresponds to the optimal distribution over the parameters of the network, and optimality is defined by minimizing a loss over the data distribution. Also, in generative modeling, one aims at approximating the intractable distribution of high-dimensional, structured data (e.g. images) while having only access to samples of the target. The main challenge is that traditional sampling schemes suffer from the curse of dimensionality, and their efficiency is severely limited for complex probability distributions (in particular those being not log-concave). Improving the state of the art requires to leverage the specific structure and geometry of the problems under study. We aim at achieving this by hybridizing methods from optimization, Monte-Carlo sampling and optimal control and transport.

Objective 3.1 – *High dimensional sampling and optimization* (participants: U. Paris Cité, Sorbonne U., Bordeaux, Lyon, IPP) Sampling can be casted as an optimization problem over the space of probability distributions, where the goal is to minimize some dissimilarity measure to the target distribution, which plays the role of the cost function; e.g. divergences from information theory, optimal transport distances, or weaker but computationally cheaper metrics such as the Maximum Mean Discrepancy (MMD) or Sliced Wasserstein (SW) distance. The goal of this objective of to leverage this point of view to design and study new sampling strategies for high dimensional sampling.

Task 1: quantization through optimization of weak discrepancies. A first fundamental problem is to study the best approximation accuracy (namely quantization) one can obtain when minimizing the cost function on discrete measures; depending on the cost itself, number of atoms, dimension and characteristics of the target distribution. Some results are known for the 2-Wasserstein distance cost that suffers from the curse of dimensionality; but for many weaker and computationally cheaper metrics such as the Maximum Mean Discrepancy or Sliced Wasserstein distances, this remains an open problem. Nevertheless, the latter metrics gained a lot of popularity in the machine learning community, and are often chose for the cost, especially in generative deep learning (Deshpande et al., 2018; Wu et al., 2019; Liutkus et al., 2019). We will address the following questions. First, we intend to derive upper and lower bounds on quantization error, for weaks metrics such as the SW-distance or MMD, by leveraging techniques preliminary investigated in Mérigot et al. (2021); Xu et al. (2022). Second, as these objectives lead to non-convex problems, we aim at investigating the behavior of the optimization dynamics. We intend to hire one Phd (A. Korba and Quentin Mérigot) on this task.

Task 2: Properties of sliced-Wasserstein metrics and Monte Carlo approximations. While many studies have shown statistical guarantees for the use of sliced-Wasserstein (SW) in machine learning (Nadjahi et al., 2020, 2019), little has been proposed to provide optimization guarantees when using SW (or its Monte-Carlo counterparts SW_p) as a loss. Our goal here is to study losses of the form $F : (x_1, \dots, x_m) \to SW(\frac{1}{m}\sum_{i=1}^m \delta_{x_i}, \mu)$ or $F_p : (x_1, \dots, x_m) \to SW_p(\frac{1}{m}\sum_{i=1}^m \delta_{x_i}, \mu)$. In particular, we plan to study the landscape of F_p and F, the existence and properties of their critical points, and the relations between them when $p \to \infty$. We will also provide guarantees on the choice of p (related to the space dimension) to ensure almost sure separation properties for SW_p, which is only a pseudo-distance. Finally, we intend to analyse the different numerical approaches used in the literature to minimize these non convex and non smooth losses (stochastic gradient descent or mini-batch SGD) and to understand their convergence guarantees. We intend to hire one post-doctoral student (J. Delon, N. Gozlan, A. Desolneux) for one year on this task.

Task 3: Efficient direction sampling for sliced-Wasserstein costs. Sliced Wasserstein original formulation consists in projecting these distributions on random straight lines, and operating a very efficient 1-d optimal transport computation on this projection. By summing this calculation over a large number of different straight lines, we arrive at an approximate motion of the optimal transport. However, the convergence of this algorithm when the dimension increases is not established: a large number of methods require an exponential number of data when the dimension of the problem increases. Our objective is to establish better strategies for choosing these random directions, such as importance sampling, or sampling methods by Markov-Chain Monte-Carlo (MCMC) which seek to determine the directions of these lines for which the 1-d optimal transport is maximal, while remaining a stochastic process. This will allow to use this distance as a loss to train Neural Networks efficiently for Wasserstein GAN or in the context of transfer learning (Wu et al., 2019; Rowland et al., 2019). We intend to hire one PhD student on this task co-advised between Université Paris Cité and Université Lyon 1 (J. Delon, F. Santambrogio, N. Bonneel, J. Digne).

Task 4: overcoming challenges in sampling: multimodality and constraints. Sampling methods from MCMC or generative modelling can break down in important cases. The first one is when the target distribution is high-dimensional with many separated modes, which is typically the case for the posterior distributions of neural networks models for instance. In practice, multimodality remains a challenge even for state-of-the-art generative methods which fail at generating samples from all the modes; while theoretical guarantees are still preliminary. The second case is when the target distribution is supported or integrable in a constrained domain, or when we only want samples that satisfy certain inequalities (e.g., fairness constraints in Bayesian inference). A few recent approaches extend classical MCMC methods to constrained domains but they contain expensive numerical subroutines, are limited to specific kind of constraints and come with few theoretical guarantees. Casting sampling as optimization enables to readily investigate novel dynamics and discretization schemes as well as constrained optimization techniques in the sampling setting, and to adopt strategies to derive theoretical guarantees. We intend to improve sampling methods in the multimodal setting by involving perturbed dynamics (noise injection, momentum) and novel geometries (e.g. Wasserstein Fisher Rao) and investigate theoretically the convergence of these methods through the recruitment of a postdoctoral student (co-supervised by Anna Korba and Julie Delon). We also intend to leverage recent constrained optimization techniques to develop numerically efficient sampling schemes for constrained distributions that come with theoretical guarantees, by recruiting one Phd student on this subject (co-supervised by Anna Korba and Arnak Dalalyan at IP Paris).

Objective 3.2 – *Inverse problems and generative models via control and transport* (participants: Strasbourg, U. Paris Cité, Orsay) Generative and sampling methods are core ingredients for the resolution

of challenging inverse problems. The use of advanced geometric notion such as optimal transport is a key aspect to explore these questions and propose new efficient solvers.

Task 1: new architectures for posterior sampling with implicit priors.. The solution of ill-posed inverse problems increasingly requires sampling the whole posterior distribution to perform uncertainty quantification or parameter calibration. We aim at developing new architectures and algorithms for posterior sampling of large families of inverse problems with pretrained neural priors. Current solutions based on standard discretizations of the Langevin SDE (Laumont et al., 2022a) and Tweedie's formula require an extremely large number of iterations to guarantee convergence. Diffusion models (Song et al., 2021) are significantly faster but require coarse approximations (Chung et al., 2023) or assumptions (Meng and Kabashima, 2022). We can obtain finer estimates by leveraging Monte Carlo simulations and Gaussian approximations that should be applicable to more general, non-Gaussian and degenerate likelihoods. We can also reduce inaccuracies in the score function based on uncertainty estimates for the underlying denoiser (Kitichotkul et al., 2021; Gal and Ghahramani, 2016; Gawlikowski et al., 2021). Finally we can further reduce time complexity via semi-implicit discretization with a suitably trained denoiser (Hurault et al., 2022). This work will be carried on by a PhD student in U. Paris Cité.

Task 2: generative models for optimal control and inverse problems. Recently, an optimal control approach (Janner et al., 2022) has been introduced using deep generative models. The objective is to concentrate by learning the probability law (with density) used to perform the control exploration on a restricted area of the action space. For the moment, the use of this approach is limited to the search for controls in small dimensions (search for optimal paths in mazes, parametric control of robots). We will try to consider problems of control/optimization of physical or biological models that involve, by essence, high dimensional constraints. The questions will consist, on the one hand, in determining the initial probability law of the algorithm in relation to the physical/biological model considered. On the other hand, the extension of these algorithms to complex problems is not obvious and will require investigation. We intend to use these algorithms to solve geometrical/topological shape optimization problems. This is a 18 month postdoc research to be supervised in Strasbourg.

Task 3: Optimal transport convexifies optimization landscape. It has been observed that the use of optimal transport data attachment terms tends to "convexify" the optimization landscape in difficult and genuinely non-convex optimization problems from inverse problems (see (Métivier et al., 2016) in the context of full waveform inversion for seismic imaging) or generative modeling. We plan hire a postdoc to study this convexification effect when the model is linear or near-linear in the parameters, and in particular LDDMM (Large deformation diffeomorphic metric mapping) or ResNet, using techniques from Wasserstein gradient flows. For this, we will build on the insights obtained on the (simpler) quantization problem. This will be achieved by a post-doc at Paris-Saclay co-supervised by Q. Mérigot and T. Gallouët (PSL).

Objective 3.3 – *Distributional data analysis and optimal transport* (participants: Toulouse, Bordeaux) In many statistical learning problems, data comes in a more complex form than simple real vectors. This is the case, for example, when using representations of these data in the form of probability measures, which corresponds to a recent field of research in statistics on the analysis of so-called distributional data. Such an approach is interesting when studying specific examples in demography (histograms of age pyramid or age of mortality), text analysis (word clouds for the study of language in *Natural Language Processing*), in computer vision or signal processing. Capturing the intrinsic geometric structure of a measure is however challenging theoretically and numerically, since it can be either viewed as a continuous object or a sets of points. Achieving this is essential to preserve

and capture meaningful information in data analysis and machine learning. We will thus develop new theoretical and computational models by leveraging the both the geometry of information theory and of optimal transport.

Task 1: Data analysis over distributions. We first intend to focus on first and second order analysis of a set of probability measures endowed with the Fisher information or the Wasserstein metric (Cazelles et al., 2018; Bigot et al., 2017) on the one hand, and on the quantization of probability measures on the other hand. *Task 1 :* First, we intend to perform a thorough comparison of the geometry of distributions according to the Fisher information metric, the Wasserstein metric and its extensions (entropy regularised OT, sliced-Wasserstein) for classical families of distributions, such as Gaussian probability measures, more generally elliptic distributions, and mixtures of Gaussians (Delon and Desolneux, 2020). We will study geodesic principal component analysis (PGA) for the Fisher metric. Algorithms to compute exact PGAs in manifolds (Sommer et al., 2014), or more specifically in hyperbolic spaces (Chakraborty et al., 2016), have been proposed. We aim at mixing Wasserstein and Fisher metrics to cope with the difficulty of expressing the Fisher geometry of general families of measures. These different tasks will be implemented in the python package Geomstats, and achieved by a postdoc student already familiar with either Fisher or Wasserstein geometry. The postdoc will be co-supervised by E. Cazelles, A. Le Brigant (SAMM) and T. Klein (ENAC), starting in 2024. *Task 2 :* The approximation of a probability density by means of an optimization problem based on an optimal transport loss is deeply related to the quantization problem and is usually solved with Lloyd's algorithm, a generalization of the k-means algorithm. Lloyd's algorithm, although proposed in the early 1980s (Lloyd, 1982), is still extensively studied (Du et al., 2006; Hamerly and Drake, 2015; Mérigot et al., 2021). The objective of this study will be focused on the numerical aspects by proposing variants of the algorithm (momentum, block variant, and clever initialization), and on a more theoretical analysis of its local convergence properties (via the Kurdyka-Lojasiewicz inequality) and global convergence properties, that are strongly related to the regularity of the c-transform studied in Objective 3.3 - Task 2. In a second phase, we will investigate the ancillary variants integrating a regularization in relation with the quantization problem and its structure (hexagonal patterns). We intend to hire a Phd student (starting in 2023), and he/she will be co-supervised by E. Cazelles (IRIT), E. Pauwels (IRIT) and J. Bolte (TSE).

Task 2: Optimal Transport for large-scale learning. Optimal transportation is a very active research topic with applications in various fields, like economics, machine learning, or image processing. With a general cost function $c(x, y)$, the optimal Kantorovich potentials ϕ, ψ in the dual formulation of an OT problem of transporting a measure μ to a measure ν are related through the so-called c-transform $\phi = \psi^c$ defined as: $\psi^c(x) = \min_y c(x, y) - \psi(y)$. One of our main goals is to supervise a PhD thesis to investigate regularity properties of the optimal c-transform, depending on the regularity of the measures μ, ν. We will focus primarily on the semi-discrete case of optimal transportation (Kitagawa et al., 2017) meaning that μ is absolutely continuous whereas ν is supported on a finite set Y. In this setting, computing the c-transform is essentially a nearest neighbour search. Besides, this setting leads to a finite-dimensional concave problem that can be solved with deterministic and stochastic solvers. However, such stochastic gradient descent approaches do not scale well when the cardinal of Y is large. In contrast, the stochastic approach of (Seguy et al., 2017) based on a parametrisation with neural networks should scale better, but remains to be analysed. The gradient formula in (Houdard et al., 2022) provides a solid foundation for the training of generative models with a deep parametrization of the dual potential, but its integration in a converging stochastic algorithm such as (Sebbouh et al., 2022) remains to be studied. In order to deal with large scale problems, the use of minibatch strategies may also be considered (Fatras et al., 2021). From the analysis of the regularity properties of the optimal dual variables, the PhD student will propose new approximation classes for ϕ, ψ.

She/he will then design new scalable numerical OT solvers relying on a principled parametrisation of the dual variables ϕ, ψ. The developed numerical solutions will be applied to several applications that require large-scale optimal transport, for example domain adaptation Courty et al. (2015), generative networks, color transfer, texture synthesis, or shape analysis. The PhD project will thus focus on leveraging the regularity of optimal dual variables ϕ, ψ to develop new stochastic solver based on a new parametrisation of the dual problem. This in will be applied to large scale imaging and learning problems.

Task 3: Distributional learning. Optimal transport, and the associated Wasserstein distance, is the de-facto standard way to compare distributional data. Yet the cornerstone and the limiting factor in the use of Wasserstein distances for statistical learning is the cost of numerical computation (more precisely its approximation) of the optimal transport between two probability measures. A PhD thesis will address the study of statistical applications of estimators based on the use of a regularized transport distance, in the settings where this regularization effect can improve the performance of estimators based solely on the use of an transport cost between probability measures. In particular, the following issues may be considered:

- the extension to the case of directional data of the method recently proposed in (Bercu et al., 2023) which extend recent approaches proposed in (Chernozhukov et al., 2017; Hallin et al., 2021) on the notion of quantiles for multivariate distributions from the standard theory of unregularized optimal transport.
- the extension of the usual linear regression model to the framework where the co-variables and the response variable can be modeled as probability measures.
- the problem of building time series models for time-indexed data that can be modeled as a stationary series of random probability measures $(\mu_t)_{t\geq 1}$, as well as the search for recursive estimators in this type of model based on the notion of regularized optimal transport.
- the convergence properties of stochastic algorithms to solve the minimax optimization problem that underlies NNs for Wasserstein GAN (Arjovsky et al., 2017) models based on optimal transport. Currently, the most used numerical methods are based on the Adam (Kingma and Ba, 2014) or $\mathrm{RMSProp}$ algorithms which are non-linear second-order algorithms. A challenge would be to achieve a better understanding of the convergence properties of these second-order algorithms in the context of Wasserstein GANs based on a recent work (Bercu et al., 2022) on the convergence of Newton-type stochastic algorithms for regularized optimal transport.

2.3. Scientific and technical description of the project

Objective 1.1. J-F. Aujol, A. Chambolle, I. Waldspurger, A. Kazeykina, C. Royer, Cardaliaguet and Z. Ren
Task 1 *(First order non-convex optimization solvers):* lead by J-F Aujol, participation of a PhD student from Bordeaux University, interactions with C. Royer and A. Chambolle (see Objective 2.1). *Milestone:* Use the interpretation as ODEs of first order optimization schemes to develop new non convex accelerated schemes. The main difficulty arises in the non convexity of the considered cost functions. *Deliverable:* Research paper and open-source implementation, e.g. Python. *Dates:* PhD starts in September/October 2023.
Task 2 *(Particle-based optimization algorithm):* lead by I. Waldspurger, with C. Royer and a post-doc student. *Milestone:* quantitative convergence guarantees for a realistic particle-based optimization

algorithm. *Deliverable:* research papers. *Dates:* 2025.
Task 3 *(Mean-field optimization):* lead by A. Kazeykina with 2 years post-doc or PhD, and P. Cardaliaguet and Z. Ren with Post-doc. *Milestone:* New algorithms for mean-field optimization achieving better performances on ML tasks. *Deliverable:* Theoretical convergence proofs and Python code. *Dates:* 2024-2027.

Objective 1.2. (Control and machine learning: there and back again) Strasbourg (C. Courtès, E. Franck, L. Navoret, Y. Privat) and Nice (J.-B. Caillau, L. Rifford, postdoc)
Task 1 *(Gradient and model learning methods to build new exploration approaches):* lead by E. Franck, Y. Privat, PhD student recruited at Université de Strasbourg. *Milestone:* Development of RL algorithms for regular control generation. Application to quantum control. *Deliverable:* Research paper, Experimental validations of the obtained strategies, in collaboration with the IPCMS (Institut de Physique et de Chimie des Matériaux de Strasbourg). *Dates:* End of 2026.
Task 2 *(Controllability and geometric properties of neural networks):* lead by J.-B. Caillau, participation of L. Rifford, with postdoc (24 months). *Milestone:* Review of controllability methods for neural ODEs and structural analysis of the extremal flow during first postdoc year. Application to energy management during second year. *Deliverable:* Assessment of turnpike property for neural ODEs and limit cycle optimisation for the KEEP (Kite Electrical Energy Production) project. *Dates:* Postdoc starting fall 2025.

Objective 1.3. (Scalable solvers and softwares) Paris Cité (J. Glaunés, B. Charlier, J. Feydy, engineer) and Nice (J.-B. Caillau, T. Kloczko, S. Vaiter, postdoc, engineer)
Task 1 *(Development of the KeOps library):* lead by J. Glaunés, participation of B. Charlier and J. Feydy, with engineer. *Milestone:* Developing new interfaces (Julia, JAX) for the KeOps library, support for a wider range of hardware, and support for approximations strategies. *Deliverable:* open-source release. *Dates:* Engineer starting spring 2024.
Task 2 *(Innovative automatic differentiation tools with code certification for numerical analysis, control and machine learning):* lead by J.-B. Caillau, participation of T. Kloczko and S. Vaiter, with postdoc (24 months) and research engineer (18 months). *Milestone:* Joining the international effort to build modern, efficient language agnostic AD tools for scientific computing, incorporating Inria *savoir faire* in AD and program proving. *Deliverable:* POC of new AD backend in Julia for control / ML applications within the control-toolbox ecosystem. *Dates:* Postdoc starting fall 2023, research engineer spring 2024 (6 months overlap).

Objective 2.1. A. Chambolle, C. Royer.
Task 1 *(Methods for implicitly constrained training):* lead by C. Royer, participation of A. Chambolle and possibly I. Mazari/E. Franck/Y. Privat, PhD student recruited at Université Paris Dauphine-PSL.. *Milestone:* Development of a matrix-free Newton-type method with complexity guarantees for implicitly constrained optimization and demonstrated performance on neural PDE training. *Deliverable:* Research paper and open-source implementation, e.g. Python or Julia. *Dates:* Spring 2024 (Sem. 2).
Task 2 *(Training of discretized architectures):* lead by C. Royer, participation of A. Chambolle and possibly I. Mazari/E. Franck/Y. Privat, PhD student recruited at Université Paris Dauphine-PSL.. *Milestone:* Development of a second-order algorithm exploiting discretized structures with complexity guarantees and numerical evidence of its interest for neural ODE training. *Deliverable:* Research paper and open-source implementation, e.g. Python or Julia. *Dates:* Spring 2025 (Sem. 4).
Task 3 *(Calibration of training methods and architectures):* lead by C. Royer, participation of A. Chambolle and PhD student. *Milestone:* Development of a blackbox optimization technique amenable to optimize both the hyperparameters of a discretized neural architecture and that of a training method

based on the methods from the previous task. *Deliverable:* Research paper and open-source package integrating those from the previous tasks. *Dates:* Spring 2026 (Sem. 6) for the final version. A preliminary one focusing only on the results from Task 1 could appear in Spring 2025 (Sem. 4)..

Objective 2.2. B. Després, Albert Cohen, Katharina Shratz, Frédéric Nataf, Borjan Gueshkovski
Task 1 *(PhD on Takagi generated Neural Network):* lead by Bruno Després, participation of Albert Cohen, PhD student from Sorbonne University. *Milestone:* Development of a full class of Takagi generated Neural Networks, with proof of convergence and study of the complexity of the algorithms. Comparison of the numerical convergence between new Takagi generated Neural Network and training with standard methods (ADAM, ...). The difficult part is to obtain mathematical series with the ability to implemented with controlled complexity.. *Deliverable:* Software written in Tensorflow/Keras with reference solutions described in research articles. *Dates:* starting september 2024.
Task 2 *(postdoctoral work on SciML):* lead by Bruno Després, participation of Frédéric Nataf, postdoctoral student Ruiyang Dai. *Milestone:* use and calibration of the number of moments in a plasma physics code with a ML technique. *Deliverable:* Proof of efficiency by comparison with standard test cases of the literature. In particular the rate of convergence of the linear solvers will be evaluated in combination with the Machine Learning tuning of the number of moments. Collaboration with the Strasbourg group is planed to accelerate the research.. *Dates:* octobre 2023.
Task 3 *(New Finite Volume numerical schemes and numerical methods generated with Machine Learning):* lead by Bruno Després, participation of Stéphane Jaouen (CEA) or Borjan Gueshkovski, one post-doctoral student.. *Milestone:* Development of a fully evolutionary PDE model generated and tuned ith Machine Learning and Neural Networks.. *Deliverable:* Code in 2D or 3D with visualization of the solution and measurement of the improvement of efficiency due to the Machine Learning strategy.. *Dates:* starting any date in 2024.

Objective 2.3. Chambolle, Privat, Digne, Masnou, Mazari, Bretin, Chaine
Task 1 *(learned NNs based discretization of singularities and their flows):* lead by A. Chambolle, S. Masnou, E. Bretin, one post-doc student, one PhD student. *Milestone:* general learned discretization for measures, currents and curvature based energy (postdoc), consistency of NNs based geometric flows, non-oriented flows, minimal surfaces. *Deliverable:* NNs based local optimization of sharp singularities – NN architectures for higher order geometric flows – consistency results. *Dates:* PhD starting 2024 - Postdoc 2024 or 2025.
Task 2 *(PDE constrained NNs):* lead by Y. Privat, E. Franck, participation of J. Aghili, A. Chambolle, I. Mazari, one post-doc student. *Milestone:* Development of hybrid methods between training (in particular PINNS) and classical optimization under PDE constraints methods (gradient, Newton, fixed point) to improve both the accuracy of the control obtained and the efficiency of the algorithm.. *Deliverable:* Paper associated to an open source software for shape optimization. *Dates:* mid 2026.
Task 3 *(Geometric control for PDE constrained optimization):* lead by I. Mazari, participation of A. Chambolle, V. Michel-Dansac, Y. Privat, Post-doc. *Milestone:* Efficient NNs for (preconditioning) shape optimisation in control problems. *Deliverable:* Papers on specific NNs architectures. *Dates:* 2025.
Task 4 *(shape generative modeling with topological constraints and Neural ODEs):* lead by J. Digne, S. Masnou, R. Chaine and a PhD Student. *Milestone:* Development of a new Neural Network architecture for synthesizing non euclidean data, addition of topological constraints and "as topology preserving as possible" constraints in the generative model.. *Deliverable:* Set of codes for shape synthesis, assessment of the solution on standard benchmarks such as ShapeNet.. *Dates:* Dates: PhD starting 2024.

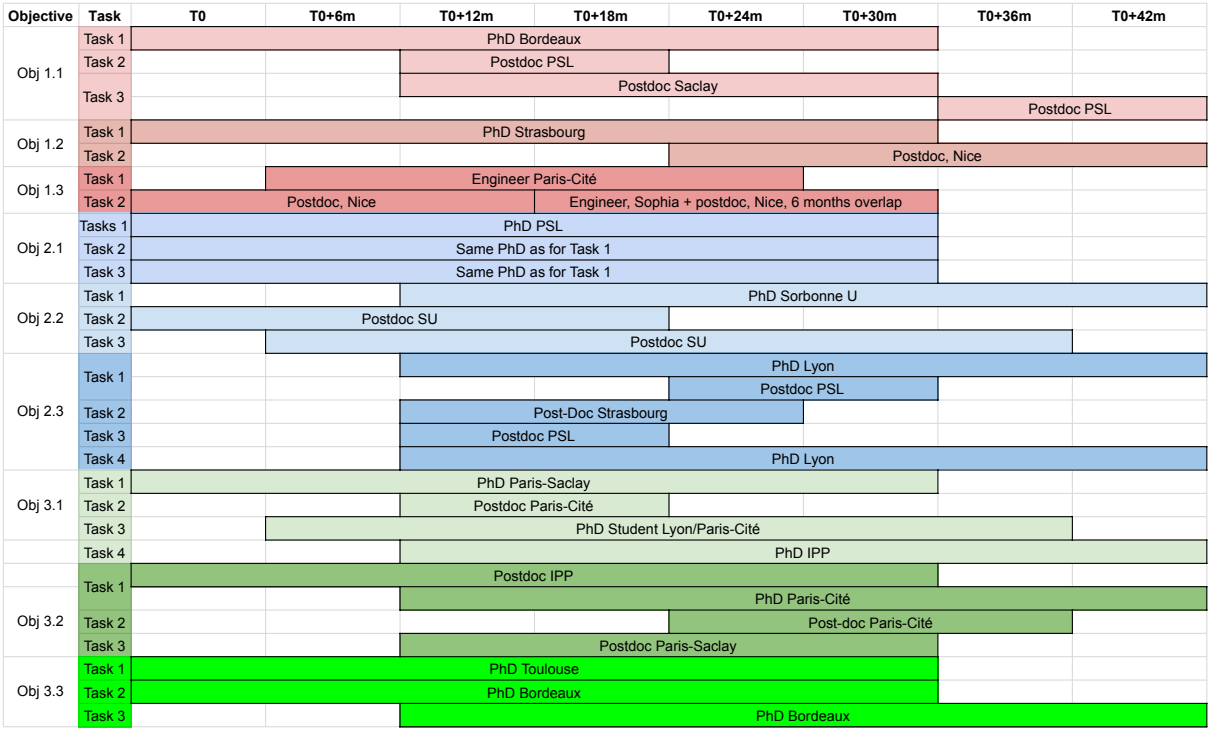
Objective 3.1. A. Korba, Q. Mérigot, J. Delon, N. Gozlan

Task 1 *(quantization through optimization of weak discrepancies):* lead by A. Korba and Q. Mérigot, one Phd student. *Milestone:* Development of optimization schemes of weak discrepancies on measures that come with tight quantization guarantees. *Deliverable:* Research papers, open-source code. *Dates:* Phd starting in 2023.
Task 2 *(Optimization guarantees for Sliced Wasserstein distances):* lead by J. Delon, participation of A. Desolneux and N. Gozlan, with Postdoc. *Milestone:* Study of the landscape of variants of SW as losses, critical points, convergence guarantees of SGD and mini-batch SGD for these losses. *Deliverable:* Research papers, open-source code. *Dates:* Postdoc starting in 2024.
Task 3 *(Efficient direction sampling for sliced-Wasserstein costs):* lead by J. Delon, N. Bonneel, J. Digne, participation of Filippo Santambrogio, one PhD student. *Milestone:* Development of new sampling strategies reducing the computational cost of Sliced Wasserstein.. *Deliverable:* Codes for fast sampling of directions, integration to losses of generative modeling architectures. *Dates:* PhD starting 2024 or 2025.
Task 4 *(overcoming challenges in sampling: multimodality and constraints):* lead by A. Korba, A. Dalalyan, J. Delon, one Phd student and one post-doc student. *Milestone:* Novel sampling algorithms, application sampling from high-dimensional target distributions with constraints, e.g. Bayesian Neural Networks with constraints. *Deliverable:* Research papers and open-source code. *Dates:* Phd starting in 2024 - Post-doc starting in 2023 or 2024.

Objective 3.2. (Paris-Cité, Strasbourg, Paris-Saclay, PSL)
Task 1 *(New architectures for posterior sampling with implicit priors.):* lead by A. Almansa, participation of P. Tan, A. Leclaire, M. Szopos, and a PhD student. *Milestone:* Devlopement of more efficient and faster schemes for posterior sampling and data generation. *Deliverable:* Research papers and release of open-source code. *Dates:* PhD starting 2024.
Task 2 *(Generative models for optimal control):* lead by C. Courtès, J. Delon, E. Franck, V. Michel-Dansac, L. Navoret, Y. Privat, V. Vigon, one post-doc student. *Milestone:* Use of generative models for solving optimal control problems inspired by Physics/Biology. Determination of a systematic and efficient way of constructing the initial probability law of the algorithm in relation to the model considered.. *Deliverable:* Research Paper including an open source code.. *Dates:* mid 2027.
Task 3 *(Optimal transport convexifies optimization landscape):* lead Q. Mérigot, participation of Blanche Buet and Hugo Leclerc, Thomas Gallouët, one post-doc. *Milestone:* Understanding the convexification effect when the model is linear or near-linear in the parameters, and in particular LDDMM (Large deformation diffeomorphic metric mapping) or ResNet, using techniques from Wasserstein gradient flows. *Deliverable:* Research papers and a benchmark comparing the energy landscape under various OT-based data attachment terms. *Dates:* Post-doc starting 2024.

Objective 3.3. Cazelles, Le Brigant, Klein, Pauwels, Bolte, Aujol, Papadakis, Bigot, Bercu
Task 1 *(Data analysis over distributions):* lead by Cazelles, participation of Le Brigant, Klein, Pauwels, Bolte, one postdoc student, on PhD student. *Milestone:* Leverage the geometry of probability distributions for first and second order analysis data analysis through Fisher-Rao and Wasserstein and approximations of measures. *Deliverable:* Implementation of geometric data analytical tools in the Geomstats package, and development and study of fast algorithms for classical Lloyd's method. *Dates:* Postdoc starting 2024 - PhD starting 2023.
Task 2 *(Optimal Transport for large-scale learning):* lead by N. Papadakis, participation of J. Delon and a PhD student of Bordeaux University. *Milestone:* Design new scalable numerical Optimal Transport solvers relying on a principled parametrisation of the dual variables of the problem. The main bottleneck lies in the scaling of the problems with respect to the dimension of the data. Applications to large scale imaging and learning problems. *Deliverable:* Research paper and open-source imple-

mentation, e.g. Python. *Dates:* PhD starts in October 2023.
Task 3 *(Distributional learning):* lead by J. Bigot, participation of B. Bercu and a PhD student from Bordeaux University, interactions with J-F Aujol (see task 1 in Objective 1.1). *Milestone:* Use of regularized transport to compute efficiently approximate Wasserstein distances. The main cornerstone with Wasserstein distances for statistical learning is the cost of numerical computations. Study of convergence properties of stochastic algorithms to solve the associated minimax optimization problem. *Deliverable:* Research paper and open-source implementation, e.g. Python. *Dates:* PhD starts in October 2024.



3. Project organisation and management

3.1. Project manager

Antonin Chambolle is a CNRS senior scientist who has been working for 17 years at CMAP, Ecole Polytechnique, Palaiseau (and teaching at Ecole Polytechnique during the whole period) before moving to Université Paris-Dauphine (CEREMADE) in 2020. He also spent one year with the "Cambridge Image Analysis group" in DAMTP, Centre of Mathematical Sciences, Cambridge, UK, in 2015-16 as a "French Government Fellow" of Churchill College.

Before that, he was trained at Ecole Normale Supérieure in Paris, was hired by CNRS in 1993 after his PhD Thesis with Jean-Michel Morel on Mathematical Image Analysis, and spent one year as a post-doctoral researcher in SISSA, Trieste, Italy.

His research topics include Calculus of Variations, Numerical Optimization, Numerical Analysis, with applications to surface evolution problems, engineering (fracture mechanics) and image and data analysis. He is the (co-)author of about 120 indexed articles in mathematical journals (including Arch. Rational Mech. Anal, Journal of the American Math. Society, Journal of the European Math. Society...), and the co-author of many conference proceedings mostly related to imaging science (such

as 3 contributions to "SSVM 2023", scale-space and variational methods in imaging).

Among his contributions, he introduced the first sound definition of a "crystalline mean curvature flow" in arbitrary dimensions, and proved the first existence result for a variational model for fracture growth introduced in the late 90's. But his most cited papers are in mathematical optimization and the study of the "total variation" regularizer for inverse problems in imaging, which are responsible for most of his more than 23000 citations in Google Scholar (4300 in MathSciNet, which places him among the 5 most cited mathematicians with PhD in 1993). He is in particular a specialist of (accelerated) first order methods in convex optimization and the co-inventor of a celebrated "primal-dual" algorithm for saddle-point problems, widely used thanks to its simplicity. He teaches a course in mathematical optimization at a Master 2 level in Paris-Dauphine (joint with Sorbonne Univ.).

He trained many students and post-docs, some of them being now prominent researchers in analysis or computer science/mathematical imaging (Michael Goldman and Etienne Corman, CNRS, Samuel Vaiter [post-doc], CNRS). He also worked as an expert for many agencies (as a regular reviewer: DFG, Hong Kong, FWF, EPSRC, ERC, NSF, or with a DFG Panel), and was also the head of the scientific council of the "Programme Gaspard Monge pour l'Optimisation et la Recherche Opérationnelle" (PGMO / FMJH / EDF R&D), attributing fundings for small research projects in optimization.

He is currently a member of the Editorial Board of 8 mathematical journals (including, since 2023, the Journal of the European Math. Society (JEMS)), and is an Editor in Chief of "Interfaces and Free Boundaries" (EMS Publishing House). He also was associate editor of the SIAM J. Imaging Science for 9 years – the maximal term).

He gave many talks and seminars, including a few plenary talks in national and international conferences (such as SIAM Imaging Science 2014, Hong Kong) and is an invited speaker at the next ICIAM 2023 in Tokyo (the International Conference of Industrial and Applied Mathematics).

He also was awarded the "prix INRIA - Michel Monpetit" of the French Academy of Science in 2021.

3.2. Organization of the partnership

The project PDE-AI takes the form of a consortium gathering the following institutions, and their respective team leader:

- Univ. Paris Dauphine PSL, CNRS: Antonin Chambolle, calculus of variations, nonsmooth and geometric optimization.

- Univ. Paris-Cité: Julie Delon, numerical optimal transport, inverse problems, stochastic models for image restoration and image editing, high dimensional sampling
- Sorbonne Univ.: Bruno Després, numerical analysis and approximation theory, error analysis, preconditioning,
- Univ. Paris Saclay: Quentin Mérigot, optimal transport, computational geometry, PDEs.
- Univ. Toulouse, CNRS: Elsa Cazelles, machine learning, optimal transport, large-scale convex and non-convex optimisation, inverse problem.
- Univ. Lyon, CNRS: Julie Digne, geometry processing, computer graphics, machine learning, optimal transport.

- Univ. Bordeaux: Jean-François Aujol, dynamical system, optimization, Nesterov acceleration, variational methods, inverse problems, optimal transport.
- Univ. Côte d'Azur, CNRS, Inria: Jean-Baptiste Caillau, geometric control, computational optimal control and applications
- CREST/Institut Polytechnique de Paris: Anna Korba, sampling, dynamical systems, optimal transport, optimization, neural networks
- Univ. Strasbourg: Yannick Privat, shape optimisation and control

The main coordinator will be Antonin Chambolle in Paris-Dauphine. He will be in charge of organizing joint meetings and discussions and encouraging the interactions between the groups forming the consortium, and following the overall progress of the researchs. The list of the researchers implied in the project is shown on table 1. blabla

Participant	Name	Dept.	Status	Employer	% inv.	m-m.
PSL	Antonin Chambolle	CEREMADE	DR	CNRS	40%	19,2
	Irène Waldspurger	CEREMADE	CR	CNRS	20%	9,6
	Thomas Gallouët	Mokaplan	CR	INRIA	20%	9,6
	Idriss Mazari	CEREMADE	MC	Dauphine	20%	9,6
	Zhenjie Ren	CEREMADE	MC	Dauphine	20%	9,6
	Pierre Cardaliaguet	CEREMADE	Prof.	Dauphine	20%	9,6
	Clément Royer	LAMSADE	MC	Dauphine	20%	9,6
U. Paris-Cité	Julie Delon	MAP5	Prof.	U. Paris Cité	40%	19,2
	Andrés Almansa	MAP5	DR	CNRS	20%	9,6
	Joan Glaunés	MAP5	MC	U. Paris Cité	20%	9,6
	Agnès Desolneux	Centre Borelli	DR	CNRS	15%	7,2
	Nathaël Gozlan	MAP5	Prof.	U. Paris Cité	15%	7,2
	Marcela Szopos	MAP5	Prof.	U. Paris Cité	10%	4,8
	Jonathan Vacher	MAP5	MC	U. Paris Cité	10%	4.8
Sorbonne Univ.	Bruno Després	LJLL	Prof.	S.U.	40%	19,2
	Albert Cohen	LJLL	Prof.	S.U.	20%	9,6
	Katharina Schratz	LJLL	Prof.	S.U.	20%	9,6
	Frédéric Nataf	LJLL	DR	CNRS	20%	9,6
U. Paris-Saclay	Quentin Mérigot	LMO	Prof.	U. Paris-Saclay	40%	19,2
	Anna Kazeykina	LMO	MC	U. Paris-Saclay	20%	9,6
	Blanche Buet	LMO	MC	U. Paris-Saclay	20%	9,6
	Luca Nenna	LMO	MC	U. Paris-Saclay	20%	9,6
	Hugo Leclerc	LMO	IR	CNRS	20%	9,6
U. Toulouse	Elsa Cazelles	IRIT	CR	CNRS	40%	19,2
	Édouard Pauwels	IRIT	MC	U. Toulouse	15%	7,2
	Jérôme Bolte	TSE	Prof.	U. Toulouse	10%	4,8
	Alice Le Brigant	SAMM	MC	U. Paris 1	20%	9,6
	Thierry Klein	IMT	Prof.	ENAC	10%	4,8
U. Lyon	Julie Digne	LIRIS	CR	CNRS	40%	19,2
	Nicolas Bonneel	LIRIS	CR	CNRS	20%	9.6
	Simon Masnou	ICJ	PR	Univ Lyon 1	20%	9.6
	Raphaëlle Chaine	LIRIS	PR	Univ Lyon 1	20%	9.6
	Elie Bretin	ICJ	MCF	INSA Lyon	20%	9.6
U. Bordeaux	Jean-François Aujol	IMB	Prof.	U. Bordeaux	40%	19,2
	Jérémie Bigot	IMB	Prof.	U. Bordeaux	20%	9,6
	Bernard Bercu	IMB	Prof.	U. Bordeaux	20%	9,6
	Arthur Leclaire	IMB	MC	U. Bordeaux	20%	9,6
	Nicolas Papadakis	IMB	DR	CNRS	20%	9,6
U. Côte d'Azur	Jean-Baptiste Caillau	LJAD/McTAO	Prof.	U. Côte d'Azur	40%	19,2
	Thibaud Kloczko	SED	IR	Inria	20%	9,6
	Ludovic Rifford	LJAD	Prof.	U. Côte d'Azur	20%	9,6
	Samuel Vaiter	LJAD	CR	CNRS	20%	9,6
IPP	Anna Korba	CREST	Ass. Prof.	GENES	30%	14,4
	Arnak Dalalyan	CREST	Prof.	GENES	20%	9,6
U. Strasbourg	Yannick Privat	IRMA	Prof.	U. Strasbourg	40%	19,2
	Joubine Aghili	IRMA	Ass. Prof.	U. Strasbourg	20%	9,6
	Clémentine Courtès	IRMA	Ass. Prof.	U. Strasbourg	20%	9,6
	Emmanuel Franck	Tonus	CR	INRIA	20%	9,6
	Philippe Helluy	IRMA	Prof.	U. Strasbourg	10%	4,8
	Victor Michel-Dansac	Tonus	ISFP	INRIA	20%	9,6
	Laurent Navoret	IRMA	Ass. Prof.	U. Strasbourg	20%	9,6
	Christophe Prud'homme	IRMA	Prof.	U. Strasbourg	10%	4,8
	Vincent Vigon	IRMA	MCF	U. Strasbourg	20%	9,6

Table 1: Participants (Permanent researchers)

3.3. Management framework

The project will rely on the work of independent teams, working together on some common goals, most of the time in collaboration. On a regular basis, we will organize mini-workshops where each group can discuss its research directions, which will encourage new collaboration. We expect to organize such meetings at least twice a year for each "Theme" independently, and at least each 18 months for the whole consortium. We will also maintain a common (wiki type) website where each team will be encouraged to describe their findings on a more frequent basis.

3.4. Institutional strategy

Since 2017, the French National Centre for Scientific Research (CNRS) has been at the forefront of promoting cutting-edge research in data science, artificial intelligence (AI), and their interaction with other fields. Its proactive efforts have led to the implementation of multiple actions, especially within the framework of the National Strategy in Artificial Intelligence.

Some of the most significant actions include the creation of the GDR IA and TAL, which have enriched the existing groups on the theme (MADICS, ISIS), the establishment of the Jean Zay supercomputer and its associated engineers network , the creation of international networks (IRN) and an international laboratory (IRL ILLS with Canada), and the annual recruitment of researchers in computer science or mathematics. Moreover, the creation of the 3IA institutes and the new AISSAI center (AI for the Sciences, Sciences for AI) has further cemented the CNRS's commitment to the field. In its latest Contract of Objectives and Performance (COP), the CNRS has reaffirmed its commitment to AI as a disciplinary research theme and cross-cutting approach that has been taken up by all CNRS institutes, beyond mathematics and computer science. The CNRS's scientific policy in the field of AI is implemented throughout the country, including in Joint Research Units (UMR) in collaboration with partners in higher education and research. As the co-leader of most of the PEPRs, the CNRS plays a unique role in the national and international landscape as a leading player in cutting-edge research and innovation. Particularly within the framework of PEPR AI, the CNRS promotes research in the mathematical foundations of AI, trustworthy and frugal AI and new computational paradigms of AI.

The consortium gathers CNRS UMRs of many universities and institution which are all involved in research at the best international level. They provide the adequate administrative staff and an excellent environment to their researchers, which will be able to use all of their time to work on the project.

The consortium leader is the PSL research university. Paris Sciences et Lettres University (PSL) is a research-intensive university that prioritizes interdisciplinary research. Research projects in the field of Partial Differential Equations (PDEs) and Artificial Intelligence (AI) have significant potential to contribute to this mission, as they involve cutting-edge mathematical research and the development of innovative AI techniques. PSL University has a strong tradition of excellence in mathematical research, with an important part of its main mathematical research center, the CEREMADE, focusing on PDEs and optimization. To support this project, PSL University can provide access to excellent research facilities, support staff, and collaborative research opportunities. Additionally, PSL's commitment to promoting interdisciplinary research and knowledge transfer provides a platform for researchers to collaborate and exchange knowledge with researchers from different fields.

Artificial intelligence is a priority research area for Université Paris Cité. Indeed, the university has already created a Graduate School of Artificial Intelligence and Data Science and also hosts the Paris Institute of Data Intelligence (dIIP), two initiatives which demonstrate the university's commitment to invest in data science. To support the project, Université Paris Cité will provide research facilities and

administrative support.

The Sorbonne Université part of the national PEPR project will strongly enhance the creation of ongoing interdisciplinarity team LJLL/SCAI focused on the fusion between machine learning and evolution models at Sorbonne University, under the umbrella of SCAI (Sorbonne Center for Artificial Intelligence). The general goal is to establish strong mathematical and applicative connections between Evolutionary Equations (PDEs, EDOS, SPDs, ...) and the new revolution of data analysis.

The research project that combine Partial Differential Equations (PDEs) and Artificial Intelligence (AI) have the potential to significantly advance mathematical modeling in AI. At Université Paris-Saclay, several research laboratories operate at the interface of mathematics and AI, facilitated by the DATAIA convergence institute. The Laboratoire de Mathématique d'Orsay, which will host the PDE-AI project, has developed expertise in optimal transport over the past decade and recently expanded its team of researchers working on the intersection of optimal transport, PDEs, and AI. To further support this area of research, Université Paris-Saclay and Inria are collaborating to establish a new scientific team at LMO, which will focus on these three topics and contribute to achieving the goals of the PDE-AI project while amplifying its impact.

The University of Toulouse, through its multiple interactions between laboratories, engineering schools (ENAC, ISAE, INSA, INP Toulouse) and research centers (Toulouse School of Economics), constitutes one of the structuring forces of the scientific research landscape in southwestern France. In particular, the interdisciplinary cohesion of the research is articulated around three major laboratories: the Institute for Research in Computer Science (IRIT), which is among the largest computer and information sciences laboratory in France, the Institute of Mathematics (IMT), which has a strong tradition in probability theory and stochastic processes, and the Systems Analysis and Architecture Laboratory (LAAS), a CNRS laboratory specializing in optimization in the broad sense. Researchers from these different institutes are grouped around the Toulouse Interdisciplinary Institute of Artificial Intelligence, ANITI, which is one of four interdisciplinary institutes of artificial intelligence (3IA) as part of the Villani Plan's Future Investments Program. ANITI focuses on artificial intelligence and data science, and supports a strong interdisciplinary research activity. ANITI 2.0 has just been renewed for four years and will be launched in September 2023. Through this 3IA institute, as well as the LabEx CIMI (International Center for Mathematics and Computer Science), the university aims to develop and disseminate research in mathematics and computer science in Toulouse. These two programs propose a strategy of recruiting high-level senior and junior scientists and exceptional students, and organize thematic trimesters, conferences and seminars. Thanks to the different actors and platforms of scientific research in Toulouse, we expect to foster synergy between researchers working on theoretical mathematical topics, such as geometry, probability and optimization, and applied research in machine learning, optimal transportation, numerical optimization and computer science.

The University of Lyon is a leading research university in France that places a strong emphasis on fundamental research with applications to important societal challenges. Among others, Artificial Intelligence research has been identified as one of the important challenges of today's research. As such, the PEPR IA-EDP fits right within this strategy. It will provide a new way to develop and explore new aspects of Artificial Intelligence and promote new collaborations between the mathematics and the computer science department.

The University of Bordeaux, is a public research university which offers a range of courses and programs in various fields, including Artificial Intelligence (AI). The AI program at Bordeaux University is part of the engineering and numerical science department, and it covers a wide range of topics

related to AI, such as machine learning, natural language processing, computer vision, robotics, and more. The program is designed to equip students with the necessary skills and knowledge to apply AI techniques in real-world applications.
Students in the AI program at Bordeaux University can pursue undergraduate, graduate, and doctoral degrees; the university also has research groups and labs dedicated to AI research, such as part of the IMB (Mathematics Institute of Bordeaux), the LaBRI (computer science institute of Bordeaux), the IMS (electrical engineering institute of Bordeaux). This offer opportunities for students to conduct research and participate in many cutting-edge projects.
IA is at the core of Bordeaux University new research politics. In 2018, The University has created the "Bordeaux Artificial Intelligence Alliance" (BAIA) to officialise the association between the University of Bordeaux and some of its major scientific partners: Bordeaux INP, CNRS, Inria and INSERM as well as many companies.

Université Côte d'Azur offers a rich and very active environment for research in mathematics and its applications. At the heart of these topics, the Laboratoire J. A. Dieudonné of CNRS is a large and very well established international mathematics center in Nice that benefits from intense collaborations with the Centre Inria d'Université Côte d'Azur. These efforts are also backed by the support of University Academy of Complex Systems, Maison de la Simulation et des Interactions, and 3IA Côte d'Azur Institute for research in AI. There are long standing interactions between mathematicians and other azurean research institutes such as Observatory of the Côte d'Azur, Mines ParisTech, I3S (computer science, CNRS), Inphyni (physics, CNRS), with a growing emphasis on AI and machine learning.

CREST (IPP) is a recognized center of excellence and a leader in the field of statistics. In particular, it has gathered experts in sampling (Nicolas Chopin), in machine learning and optimal transport (Marco Cuturi, Vianney Perchet), or both (Arnak Dalalyan, Anna Korba). It is at the heart of several strategic partnerships. Through ENSAE, it is a member of the Institut Polytechnique de Paris (IP Paris), a group of institutions including the École Polytechnique, the École Nationale Supérieure des Techniques Avancées (ENSTA Paris), Télécom Paris and Télécom SudParis. It is a member of Hi! Paris, the interdisciplinary center dedicated to artificial intelligence and data science, at the service of science, the economy and society, created jointly by IP Paris and HEC. CREST's partnerships with other leading schools and laboratories play an important role in amplifying CREST's strategy for developing its research activities, particularly in the fields of statistics and artificial intelligence; through funding research and organizing scientific events.
This PEPR project is a complementary partnership. It will bring closer researchers from CREST in sampling and machine learning with researchers experts in PDE, optimal control and optimal transport from the consortium; themes that are originally less represented at CREST, but are necessary to overcome fundamental obstacles e.g. in sampling research as presented in this detailed project. The PEPR partnership aims at supporting ambitious, novel and cutting-edge research in statistics and machine learning, in particular in sampling. By enforcing the synergies between the members of the consortium, it will enable to develop competitive research with the best international laboratories and to publish to top-tier machine learning conferences.

Regarding the University of Strasbourg, this project will reinforce the interdisciplinarity of mathematics with other disciplines: computer science, quantum physics and biology, but also medicine. The members plan to collaborate with the technological platform in Cemosis, which will allow to quickly establish academic contacts from other disciplines than Mathematics and has means to support research (research engineers in particular). The project fits perfectly with the strategy of Interdisciplinary Thematic Institutes (ITI) and in particular IRMIA++. The IRMIA++ institute also offers possibilities of

financing the continuation of the contract, as well as means of communication on the various projects.

4. Expected outcomes of the project

Results and impact. This project is transverse in nature: it will develop **theoretical** and **algorithmic** tools which will find applications in other areas covered by the PEPR.

On the **theoretical** side, we will develop methods for a better understanding of machine learning techniques, with new points of view. IA is an incredibly rapidly growing field of research with promising applications (see the recent success of ChatGPT in generating plausible texts and conversation), yet these applications are by now far more advanced than our understanding of IA. The development of robust mathematical foundations for artificial intelligence (AI) is crucial for enhancing the quality, reliability, and range of its applications. Achieving this is of paramount strategic importance for the country, especially if we aim to reduce our reliance on large technology conglomerates such as Google, Amazon, Facebook, Apple, and Microsoft (GAFAM). By investing in the development of a solid mathematical framework for AI, we can unlock the potential of this powerful technology to address critical societal challenges across multiple domains. The aim of this project is to strengthen the mathematical foundations of artificial intelligence (AI) techniques, with a focus on developing safer and more reliable AI systems. This will involve both core AI areas such as computer vision, natural language processing, and robotics, as well as the application of AI to various scientific fields, including physics, chemistry, and biology. In particular, the project will focus on using partial differential equations (PDEs), which are widely used in these fields as a fundamental language to describe complex physical phenomena, to enhance the effectiveness of AI techniques in these areas. By doing so, this project seeks to create a new paradigm for AI research that integrates mathematical rigor with practical application to solve real-world problems in various domains.

On the **algorithmic** side, the main set of tools developed by the project are:

- Novel optimisation schemes, which are at the workhorses of training methods. These schemes makes use of both novel optimization methods and automatic differentiation advances.
- Novel deep architectures, leveraging structures from differential equations, physics and inverse problems.
- Novel sampling schemes, integrating ideas from optimal transport and optimization over the space of measures.

The three classes of algorithmic contributions will be integrated into open-source software, making them accessible to a wide range of users and facilitating their adoption by other projects within the PEPR.

Diffusion and promotion of results. This project is theoretical in nature, so the first source of output will be scientific paper. But very importantly, these scientific paper will be accompanied with open source software to reproduce their result, and be applicable beyond the project to other part of the PEPR. This approach aligns with the project's commitment to openness, transparency, and collaboration, and ensures that the benefits of these contributions can be realized beyond the immediate scope of the project. Additionally, making these contributions available through open-source software will help to promote transparency and accountability in AI development, which is essential for ensuring that these technologies are used ethically and responsibly.

In addition to the development of mathematical foundations for AI and the integration of algorithmic ad-

vancements into open-source software, another crucial aspect of this project is the establishment and growth of a new community at the intersection of machine learning and partial differential equations (PDEs). This community will be instrumental in driving the progress of the project and will be actively involved in organizing workshops and other events supported by the PEPR. By bringing together experts from these two fields, the community will facilitate the exchange of ideas, promote collaboration, and foster a culture of innovation and creativity. The community will also serve as a platform for the dissemination of research findings and the sharing of best practices, enabling researchers and practitioners to stay up-to-date with the latest developments and trends in the field. By nurturing this community, the project aims to create a vibrant and supportive ecosystem of researchers and practitioners who are committed to advancing the frontiers of AI and PDEs, and to addressing some of the most pressing challenges facing our society today.

Relation with the PEPR. The PEPR comprises multiple projects that develop and use mathematical methods related to the core objectives of our project. These projects may include, for example, efforts to develop new numerical methods, statistical inference techniques, or mathematical models for analyzing complex systems.

The HOLIGRAIL, ADAPTING, EMERGENCE and REDEEM projects will each introduce novel constraints on ML optimization methods, with a focus on areas such as quantization/sparsity of parameters, energy consumption, computational models and decentralized computations. These constraints represent important considerations for the development of AI models that are both effective and efficient, as they can have a significant impact on performance and resource utilization. As our project aims to develop novel optimization and sampling routines, it will be important to assess whether some of these constraints can be integrated into our approaches, and to identify any potential theoretical obstacles that may need to be addressed. To this end, our teams will have scientific exchange with the HOLIGRAIL, ADAPTING, and EMERGENCE projects to understand the specific constraints they will introduce, and to explore ways in which they may be incorporated into our optimization and sampling routines. This may involve developing new algorithms or modifying existing ones to better accommodate these constraints, or it may require a deeper theoretical understanding of the underlying mathematical principles involved. In either case, our team will remain focused on developing approaches that are both practical and grounded in rigorous mathematical principles. These cross-fertilization between the projects will be key to advance the state of the art in AI research and development, and to help address some of the most pressing challenges facing our society today.

In addition to optimization-based considerations, our mathematical advances will also play a key role in developing trustworthy AI. For example, the FOUNDRY project will rely on theoretical concepts from game theory and optimal transport, which are also central to our project. As members of our team have expertise in these areas, we expect that collaboration with FOUNDRY will be both natural and highly beneficial. Similarly, the CAUSALIT-AI project will focus on theoretical and numerical aspects of causality for AI, which rely on concepts such as computational graphs. These concepts are also at the core of automatic differentiation, which our team will explore at both a theoretical and numerical level. By developing efficient computational software, our team hopes to facilitate stimulating discussions with the CAUSALIT-AI project. In both cases, the integration of our mathematical advances into these related projects has the potential to yield significant benefits. By leveraging our expertise and collaborating with these projects, we can help to further advance the development of trustworthy AI and promote interdisciplinary research across multiple domains. Moreover, these collaborations can also help to foster a broader understanding of the theoretical foundations underlying AI research and development, which can ultimately lead to more effective and impactful solutions to some of the most pressing challenges facing our society today.

While all these PEPR projects may have different goals and approaches, they share a common interest in advancing the frontiers of mathematical methods for AI research and development. As such, our project will actively encourage and support collaboration and discussion between these projects and our teams. By fostering collaboration and knowledge-sharing among these different projects, we aim to create a synergistic environment that facilitates the development of novel and innovative approaches to AI research and development. This approach is also consistent with the broader goals of the PEPR, which seeks to promote interdisciplinary research and collaboration across multiple domains.

Let us also stress that the PEPR will also organize three ANR calls focused on the mathematics of decision for AI, the geometry and topology of deep learning, and probabilistic methods for AI. While the specific projects selected for these calls have not yet been determined, we anticipate that there will be strong links between these projects and our teams, leading to stimulating exchanges at both the theoretical and practical levels, including algorithmic and software development. These ANR calls provide an important opportunity to expand the scope of our research project and engage with a broader community of scholars working in related areas. By collaborating with other teams through these calls, we can further deepen our understanding of the mathematical foundations of AI and explore new avenues for advancing the state of the art. Additionally, these collaborations can help to promote cross-disciplinary research and facilitate the exchange of ideas and techniques across multiple domains, ultimately leading to more impactful and innovative research outcomes. We look forward to the opportunities presented by these ANR calls and are excited to contribute to the ongoing development of cutting-edge AI research in France and beyond.

5. References

I. Contributions by the PDE-AI consortium

Matthieu Ancellin, Bruno Després, and Stéphane Jaouen. Extension of generic two-component vof interface advection schemes to an arbitrary number of components. *Journal of Computational Physics*, 473:111721, 2023. ISSN 0021-9991. doi: https://doi.org/10.1016/j.jcp.2022.111721. URL https://www.sciencedirect.com/science/article/pii/S0021999122007847.

V. Apidopoulos, J.-F. Aujol, and Ch. Dossal. The differential inclusion modeling FISTA algorithm and optimality of convergence rate in the case $b \leq 3$. *SIAM J on Optimization*, 2018.

V. Apidopoulos, N. Ginatta, and S. Villa. Convergence rates for the heavy-ball continuous dynamics for non-convex optimization, under Polyak–Łojasiewicz condition. *J of G Opt*, 2022.

Jean-Francois Aujol, Charles Dossal, and Aude Rondepierre. Optimal convergence rates for Nesterov acceleration. *SIAM J. Optim.*, 29(4):3131–3153, 2019. ISSN 1052-6234. doi: 10.1137/18M1186757. URL https://doi.org/10.1137/18M1186757.

B. Bercu, J. Bigot, S. Gadat, and E. Siviero. A stochastic gauss-newton algorithm for regularized semi-discrete optimal transport. Information and Inference: A Journal of the IMA, 2022.

Bernard Bercu, Jérémie Bigot, and Gauthier Thurin. Stochastic optimal transport in Banach spaces for regularized estimation of multivariate quantiles. *Pre-print Arxiv:2302.00982*, 2023. doi: 10.48550/ARXIV.2302.00982.

E. Bergou, Y. Diouane, V. Kungurtsev, and C. W. Royer. A stochastic Levenberg-Marquardt method for using random models with complexity results and application to data assimilation. *SIAM/ASA J. Uncertain. Quantif.*, 10:507–536, 2022.

Jérémie Bigot, Raúl Gouet, Thierry Klein, and Alfredo López. Geodesic PCA in the Wasserstein space by convex PCA. *Ann. Inst. Henri Poincaré Probab. Stat.*, 53(1):1–26, 2017. ISSN 0246-0203. doi: 10.1214/15-AIHP706. URL https://doi.org/10.1214/15-AIHP706.

Lea Bogensperger, Antonin Chambolle, and Thomas Pock. Convergence of a piggyback-style method for the differentiation of solutions of standard saddle-point problems. *SIAM J. Math. Data Sci.*, 4(3): 1003–1030, 2022. doi: 10.1137/21M1455887. URL https://doi.org/10.1137/21M1455887.

Elie Bretin, Roland Denis, Simon Masnou, and Garry Terii. Learning phase field mean curvature flows with neural networks. *Journal of Computational Physics*, 470:111579, dec 2022. doi: 10.1016/j.jcp.2022.111579. URL https://doi.org/10.1016%2Fj.jcp.2022.111579.

Jean-Baptiste Caillau, Olivier Cots, and Pierre Martinon. ct: control toolbox – numerical tools and examples in optimal control*. *IFAC-PapersOnLine*, 55(16):13–18, 2022. ISSN 2405-8963. doi: https://doi.org/10.1016/j.ifacol.2022.08.074. URL https://www.sciencedirect.com/science/article/pii/S2405896322011648. 18th IFAC Workshop on Control Applications of Optimization CAO 2022.

Marco Caroccia, Antonin Chambolle, and Dejan Slepčev. Mumford-Shah functionals on graphs and their asymptotics. *Nonlinearity*, 33(8):3846–3888, 2020. ISSN 0951-7715. doi: 10.1088/1361-6544/ab81ee. URL https://doi.org/10.1088/1361-6544/ab81ee.

Elsa Cazelles, Vivien Seguy, Jérémie Bigot, Marco Cuturi, and Nicolas Papadakis. Geodesic PCA versus log-PCA of histograms in the Wasserstein space. *SIAM J. Sci. Comput.*, 40(2):B429–B456, 2018. ISSN 1064-8275. doi: 10.1137/17M1143459. URL https://doi.org/10.1137/17M1143459.

Antonin Chambolle and Juan Pablo Contreras. Accelerated Bregman primal-dual methods applied to optimal transport and Wasserstein barycenter problems. *SIAM J. Math. Data Sci.*, 4(4):1369–1395, 2022. doi: 10.1137/22M1481865. URL https://doi.org/10.1137/22M1481865.

Antonin Chambolle and Thomas Pock. An introduction to continuous optimization for imaging. *Acta Numer.*, 25:161–319, 2016. ISSN 0962-4929. doi: 10.1017/S096249291600009X. URL http://dx.doi.org/10.1017/S096249291600009X.

Antonin Chambolle and Thomas Pock. Learning consistent discretizations of the total variation. *SIAM J. Imaging Sci.*, 14(2):778–813, 2021a. doi: 10.1137/20M1377199. URL https://doi.org/10.1137/20M1377199.

Antonin Chambolle and Thomas Pock. Approximating the total variation with finite differences or finite elements. In *Geometric partial differential equations. Part II*, volume 22 of *Handb. Numer. Anal.*, pages 383–417. Elsevier/North-Holland, Amsterdam, 2021b. doi: 10.1016/bs.hna.2020.10.005. URL https://doi.org/10.1016/bs.hna.2020.10.005.

Benjamin Charlier, Jean Feydy, Joan Alexis Glaunès, François-David Collin, and Ghislain Durif. Kernel operations on the GPU, with autodiff, without memory overflows. *Journal of Machine Learning Research*, 22(74):1–6, 2021. URL http://jmlr.org/papers/v22/20-275.html.

Julien Claisse, Giovanni Conforti, Zhenjie Ren, and Songbo Wang. Mean field optimization problem regularized by fisher information, 2023. URL https://arxiv.org/abs/2302.05938.

Giovanni Conforti, Anna Kazeykina, and Zhenjie Ren. Game on random environment, mean-field langevin system, and neural networks. *Mathematics of Operations Research*, 2022.

F. E. Curtis, D. P. Robinson, C. W. Royer, and S. J. Wright. Trust-region Newton-CG with strong second-order complexity guarantees for nonconvex optimization. *SIAM J. Optim.*, 31:518–544, 2021.

Julie Delon and Agnès Desolneux. A Wasserstein-type distance in the space of Gaussian mixture models. *SIAM J. Imaging Sci.*, 13(2):936–970, 2020. doi: 10.1137/19M1301047. URL https://doi.org/10.1137/19M1301047.

Quentin Denoyelle, Vincent Duval, Gabriel Peyré, and Emmanuel Soubies. The sliding frank–wolfe algorithm and its application to super-resolution microscopy. *Inverse Problems*, 36(1):014001, 2019.

Bruno Després. *Neural networks and numerical analysis*, volume 6 of *De Gruyter Series in Applied and Numerical Mathematics*. De Gruyter, Berlin, 2022. ISBN 978-3-11-078318-6; 978-3-

11-078312-4; 978-3-11-078326-1. doi: 10.1515/9783110783186. URL https://doi.org/10.1515/9783110783186.

Bruno Després. A convergent deep learning algorithm for approximation of polynomials. working paper or preprint, August 2022. URL https://hal.sorbonne-universite.fr/hal-03762185.

Bruno Després and Matthieu Ancellin. A functional equation with polynomial solutions and application to neural networks. *C. R. Math. Acad. Sci. Paris*, 358(9-10):1059–1072, 2020. ISSN 1631-073X. doi: 10.5802/crmath.124. URL https://doi.org/10.5802/crmath.124.

Bruno Després and Hervé Jourdren. Machine learning design of volume of fluid schemes for compressible flows. *Journal of Computational Physics*, 408:109275, 2020. ISSN 0021-9991. doi: https://doi.org/10.1016/j.jcp.2020.109275. URL https://www.sciencedirect.com/science/article/pii/S0021999120300498.

Julie Digne, Sébastien Valette, and Raphaëlle Chaine. Sparse geometric representation through local shape probing. *IEEE Transactions on Visualization and Computer Graphics*, 24(7):2238–2250, 2018. doi: 10.1109/TVCG.2017.2719024.

Jean Feydy, Alexis Glaunès, Benjamin Charlier, and Michael Bronstein. Fast geometric learning with symbolic matrices. In H. Larochelle, M. Ranzato, R. Hadsell, M.F. Balcan, and H. Lin, editors, *Advances in Neural Information Processing Systems*, volume 33, pages 14448–14462. Curran Associates, Inc., 2020. URL https://proceedings.neurips.cc/paper/2020/file/a6292668b36ef412fa3c4102d1311a62-Paper.pdf.

A. Houdard, A. Leclaire, N. Papadakis, and J. Rabin. On the gradient formula for learning generative models with regularized optimal transport costs. *preprint*, 2022.

Kaitong Hu, Zhenjie Ren, David Šiška, and Ł ukasz Szpruch. Mean-field Langevin dynamics and energy landscape of neural networks. *Ann. Inst. Henri Poincaré Probab. Stat.*, 57(4):2043–2065, 2021. ISSN 0246-0203. doi: 10.1214/20-aihp1140. URL https://doi.org/10.1214/20-aihp1140.

Tom Huix, Szymon Majewski, Alain Durmus, Eric Moulines, and Anna Korba. Variational inference of overparameterized bayesian neural networks: a theoretical and empirical study. *arXiv preprint arXiv:2207.03859*, 2022.

Samuel Hurault, Arthur Leclaire, and Nicolas Papadakis. Proximal Denoiser for Convergent Plug-and-Play Optimization with Nonconvex Regularization. In *(ICML) International Conference on Machine Learning*, number 2, jan 2022. URL https://github.com/samuro95/Prox-PnP.

J. Kitagawa, Q. Merigot, and B. Thibert. A Newton algorithm for semi-discrete optimal transport. Journal of the European Math Society, 2017.

Anna Korba, Adil Salim, Michael Arbel, Giulia Luise, and Arthur Gretton. A non-asymptotic analysis for stein variational gradient descent. *Advances in Neural Information Processing Systems*, 33: 4672–4682, 2020.

Rémi Laumont, Valentin De Bortoli, Andrés Almansa, Julie Delon, Alain Durmus, and Marcelo Pereyra. Bayesian Imaging Using Plug & Play Priors: When Langevin Meets Tweedie. *SIAM Journal on Imaging Sciences*, 15(2):701–737, jun 2022a. ISSN 1936-4954. doi: 10.1137/21M1406349. URL https://hal.archives-ouvertes.fr/hal-03161400v2.

Rémi Laumont, Valentin De Bortoli, Andrés Almansa, Julie Delon, Alain Durmus, and Marcelo Pereyra. Bayesian imaging using plug & play priors: when Langevin meets Tweedie. *SIAM J. Imaging Sci.*, 15(2):701–737, 2022b. doi: 10.1137/21M1406349. URL https://doi.org/10.1137/21M1406349.

Rémi Laumont, Valentin De Bortoli, Andrés Almansa, Julie Delon, Alain Durmus, and Marcelo Pereyra. On Maximum a Posteriori Estimation with Plug & Play Priors and Stochastic Gradient Descent. *J. Math. Imaging Vision*, 65(1):140–163, 2023. ISSN 0924-9907. doi: 10.1007/s10851-022-01134-7. URL https://doi.org/10.1007/s10851-022-01134-7.

Idriss Mazari and Domènec Ruiz-Balet. Spatial ecology, optimal control and game theoretical fishing problems. *J. Math. Biol.*, 85(5):Paper No. 55, 61, 2022. ISSN 0303-6812. doi: 10.1007/s00285-022-01829-w. URL https://doi.org/10.1007/s00285-022-01829-w.

Idriss Mazari, Grégoire Nadin, and Yannick Privat. Optimisation of the total population size for logistic diffusive equations: bang-bang property and fragmentation rate. *Comm. Partial Differential Equations*, 47(4):797–828, 2022. ISSN 0360-5302. doi: 10.1080/03605302.2021.2007533. URL https://doi.org/10.1080/03605302.2021.2007533.

Quentin Mérigot, Filippo Santambrogio, and Clément Sarrazin. Non-asymptotic convergence bounds for wasserstein approximation using point clouds. In A. Beygelzimer, Y. Dauphin, P. Liang, and J. Wortman Vaughan, editors, *Advances in Neural Information Processing Systems*, 2021. URL https://openreview.net/forum?id=_6j_jQiYB2c.

Ludovic Métivier, Romain Brossier, Quentin Mérigot, Edouard Oudet, and Jean Virieux. Measuring the misfit between seismograms using an optimal transport distance: Application to full waveform inversion. *Geophysical Supplements to the Monthly Notices of the Royal Astronomical Society*, 205 (1):345–377, 2016.

Yann Traonmilin and Jean-François Aujol. The basins of attraction of the global minimizers of the non-convex sparse spike estimation problem. *Inverse Problems*, 36(4):045003, 2020.

Lantian Xu, Anna Korba, and Dejan Slepcev. Accurate quantization of measures via interacting particle-based optimization. In *International Conference on Machine Learning*, pages 24576–24595. PMLR, 2022.

II. Other references

Yves Achdou and Mathieu Laurière. Mean field games and applications: Numerical aspects. In Pierre Cardaliaguet and Alessio Porretta, editors, *Mean Field Games: Cetraro, Italy 2019*, pages 249–307. Springer International Publishing, Cham, 2020. ISBN 978-3-030-59837-2. doi: 10.1007/978-3-030-59837-2_4. URL https://doi.org/10.1007/978-3-030-59837-2_4.

A. Agrachev and A. Sarychev. Control on the manifolds of mappings with a view to the deep learning. J Dyn Control Syst, 28:989–1008, 2022.

Luigi Ambrosio, Camillo Brena, and Sergio Conti. Functions with bounded Hessian-Schatten variation: density, variational and extremality properties. *arXiv preprint arXiv:2302.12554*, 2023.

H. Antil, D. P. Khouri, M.-D. Lacasse, and D. Ridzal, editors. *Frontiers in PDE-Constrained Optimization*, volume 163 of *The IMA Volumes in Mathematics and its Applications*. Springer, New York, NY, USA, 2016.

Martin Arjovsky, Soumith Chintala, and Léon Bottou. Wasserstein generative adversarial networks. In *Proceedings of the 34th International Conference on Machine Learning - Volume 70*, ICML'17, page 214–223. JMLR.org, 2017.

Yoshua Bengio, Nicolas Le Roux, Pascal Vincent, Olivier Delalleau, and Patrice Marcotte. Convex neural networks. *Advances in neural information processing systems*, 18, 2005.

G. Biros and O. Ghattas. Parallel Lagrange-Newton-Krylov-Schur methods for PDE-constrained optimization. Part I: The Krylov-Schur solver. *SIAM J. Sci. Comput.*, 27:687–713, 2005.

Blanche Buet and Martin Rumpf. Mean curvature motion of point cloud varifolds. *ESAIM Math. Model. Numer. Anal.*, 56(5):1773–1808, 2022. ISSN 2822-7840. doi: 10.1051/m2an/2022047. URL https://doi.org/10.1051/m2an/2022047.

F. Calabrò, S. Cuomo, F. Giampaolo, S. Izzo, C. Nitsch, F. Piccialli, and C. Trombetti. Deep learning for the approximation of a shape functional, 2021.

C. Cartis, N. I. M. Gould, and Ph. L. Toint. *Evaluation Complexity of Algorithms for Nonconvex Op-*

timization: Theory, Computation and Perspectives, volume MO30 of *MOS-SIAM Series on Optimization*. SIAM, 2022.

Rudrasis Chakraborty, Dohyung Seo, and Baba C Vemuri. An efficient exact-pga algorithm for constant curvature manifolds. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3976–3984, 2016.

R. Chen, M. Menickelly, and K. Scheinberg. Stochastic optimization using a trust-region method and random models. *Math. Program.*, 169:447–487, 2018.

R. T. Q. Chen, Y. Rubanova, J. Bettencourt, and D. K. Duvenaud. Neural ordinary differential equations. In *Advances in Neural Information Processing Systems 31*, pages 6571–6583, 2018.

Ricky T. Q. Chen, Yulia Rubanova, Jesse Bettencourt, and David K Duvenaud. Neural ordinary differential equations. In S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 31. Curran Associates, Inc., 2018. URL https://proceedings.neurips.cc/paper/2018/file/69386f6bb1dfed68692a24c8686939b9-Paper.pdf.

Victor Chernozhukov, Alfred Galichon, Marc Hallin, and Marc Henry. Monge-Kantorovich depth, quantiles, ranks and signs. *The Annals of Statistics*, 45(1):223 – 256, 2017. doi: 10.1214/16-AOS1450. URL https://doi.org/10.1214/16-AOS1450.

Lenaic Chizat. Sparse optimization on measures with over-parameterized gradient descent. *Mathematical Programming*, 194(1-2):487–532, 2022.

Lenaic Chizat and Francis Bach. On the global convergence of gradient descent for over-parameterized models using optimal transport. *Advances in neural information processing systems*, 31, 2018.

Hyungjin Chung, Jeongsol Kim, Michael T. Mccann, Marc L. Klasky, and Jong Chul Ye. Diffusion Posterior Sampling for General Noisy Inverse Problems. In *(ICLR) International Conference on Learning Representations*, pages 1–28, sep 2023. URL https://openreview.net/forum?id=OnD9zGAGT0k.

Laurent Condat. Discrete total variation: new definition and minimization. *SIAM J. Imaging Sci.*, 10 (3):1258–1290, 2017. ISSN 1936-4954. URL https://doi.org/10.1137/16M1075247.

N. Courty, R. Flamary, R. Tuia, and A. Rakotomamonjy. Optimal transport for domain adaptation. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015.

Yohann De Castro and Fabrice Gamboa. Exact reconstruction using beurling minimal extrapolation. *Journal of Mathematical Analysis and applications*, 395(1):336–354, 2012.

A. Défossez, L. Bottou, F. Bach, and N. Usunier. A simple convergence proof of ADAM and ADAGRAD. *preprint arXiv:2003.02395*, 2020.

Ishan Deshpande, Ziyu Zhang, and Alexander G Schwing. Generative modeling using the sliced Wasserstein distance. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3483–3491, 2018.

Qiang Du, Maria Emelianenko, and Lili Ju. Convergence of the lloyd algorithm for computing centroidal voronoi tessellations. *SIAM journal on numerical analysis*, 44(1):102–119, 2006.

K. Fatras, T. éjourné, R. Flamary, and N. Courty. Unbalanced minibatch optimal transport; applications to domain adaptation. In International Conference on Machine Learning, 2021.

Jean Feydy, Thibault Séjourné, François-Xavier Vialard, Shun-ichi Amari, Alain Trouvé, and Gabriel Peyré. Interpolating between optimal transport and MMD using sinkhorn divergences. In Kamalika Chaudhuri and Masashi Sugiyama, editors, *The 22nd International Conference on Artificial Intelligence and Statistics, AISTATS 2019, 16-18 April 2019, Naha, Okinawa, Japan*, volume 89 of *Proceedings of Machine Learning Research*, pages 2681–2690. PMLR, 2019. URL http://proceedings.mlr.press/v89/feydy19a.html.

Yarin Gal and Zoubin Ghahramani. Dropout as a Bayesian Approximation: Representing Model Uncertainty in Deep Learning. In *(ICML) International Conference on Machine Learning*, pages 1050—-1059. PMLR, 2016. URL https://proceedings.mlr.press/v48/gal16.html.

Jacob R. Gardner, Geoff Pleiss, Kilian Q. Weinberger, David Bindel, and Andrew Gordon Wilson. Gpytorch: Blackbox matrix-matrix gaussian process inference with GPU acceleration. In Samy Bengio, Hanna M. Wallach, Hugo Larochelle, Kristen Grauman, Nicolò Cesa-Bianchi, and Roman Garnett, editors, *Advances in Neural Information Processing Systems 31: Annual Conference on Neural Information Processing Systems 2018, NeurIPS 2018, December 3-8, 2018, Montréal, Canada*, pages 7587–7597, 2018. URL https://proceedings.neurips.cc/paper/2018/hash/27e8e17134dd7083b050476733207ea1-Abstract.html.

Jakob Gawlikowski, Cedrique Rovile Njieutcheu Tassi, Mohsin Ali, Jongseok Lee, Matthias Humt, Jianxiang Feng, Anna Kruspe, Rudolph Triebel, Peter Jung, Ribana Roscher, Muhammad Shahzad, Wen Yang, Richard Bamler, and Xiao Xiang Zhu. A Survey of Uncertainty in Deep Neural Networks. pages 1–41, jul 2021. URL http://arxiv.org/abs/2107.03342.

B. Geshkovski and E. Zuazua. Turnpike in optimal control pdes, resnets, and beyond. Acta Numer., 31:135–263, 2022.

E. Haber and L. Ruthotto. Stable architectures for deep neural networks. *Inverse Problems*, 34: 014004, 2018.

Marc Hallin, Eustasio del Barrio, Juan Cuesta-Albertos, and Carlos Matran. Distribution and quantile functions, ranks and signs in dimension d: A measure transportation approach. *The Annals of Statistics*, 49(2):1139 – 1165, 2021. doi: 10.1214/20-AOS1996. URL https://doi.org/10.1214/20-AOS1996.

Greg Hamerly and Jonathan Drake. Accelerating lloyd's algorithm for k-means clustering. *Partitional clustering algorithms*, pages 41–78, 2015.

M. Heinkenschloss and D. Ridzal. A matrix-free trust-region SQP method for equality constrained optimization. *SIAM J. Optim.*, 24(3):1507–1541, 2014.

Michael Janner, Yilun Du, Joshua Tenenbaum, and Sergey Levine. Planning with diffusion for flexible behavior synthesis. In Kamalika Chaudhuri, Stefanie Jegelka, Le Song, Csaba Szepesvari, Gang Niu, and Sivan Sabato, editors, *Proceedings of the 39th International Conference on Machine Learning*, volume 162 of *Proceedings of Machine Learning Research*, pages 9902–9915. PMLR, 17–23 Jul 2022. URL https://proceedings.mlr.press/v162/janner22a.html.

Anees Kazi, Luca Cosmo, Seyed-Ahmad Ahmadi, Nassir Navab, and Michael M. Bronstein. Differentiable graph module (dgm) for graph convolutional networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(2):1606–1617, 2023. doi: 10.1109/TPAMI.2022.3170249.

P. Kingma and J. Ba. ADAM: A method for stochastic optimization. *arXiv:1412.6980*, 2014.

Ruangrawee Kitichotkul, Christopher A. Metzler, Frank Ong, and Gordon Wetzstein. Suremap: Predicting Uncertainty in Cnn-Based Image Reconstructions Using Stein's Unbiased Risk Estimate. In *(ICASSP) International Conference on Acoustics, Speech and Signal Processing*, pages 1385–1389. IEEE, jun 2021. ISBN 978-1-7281-7605-5. doi: 10.1109/ICASSP39728.2021.9414306.

Léo Lebrat, Frédéric de Gournay, Jonas Kahn, and Pierre Weiss. Optimal transport approximation of 2-dimensional measures. *SIAM J. Imaging Sci.*, 12(2):762–787, 2019. doi: 10.1137/18M1193736. URL https://doi.org/10.1137/18M1193736.

Timothy P. Lillicrap, Jonathan J. Hunt, Alexander Pritzel, Nicolas Heess, Tom Erez, Yuval Tassa, David Silver, and Daan Wierstra. Continuous control with deep reinforcement learning. In Yoshua Bengio and Yann LeCun, editors, *4th International Conference on Learning Representations, ICLR 2016, San Juan, Puerto Rico, May 2-4, 2016, Conference Track Proceedings*, 2016. URL http:

//arxiv.org/abs/1509.02971.

Antoine Liutkus, Umut Simsekli, Szymon Majewski, Alain Durmus, and Fabian-Robert Stöter. Sliced-wasserstein flows: Nonparametric generative modeling via optimal transport and diffusions. In *International Conference on Machine Learning*, pages 4104–4113. PMLR, 2019.

Stuart Lloyd. Least squares quantization in pcm. *IEEE transactions on information theory*, 28(2): 129–137, 1982.

Y. Lu, A. Zhong, Q. Li, and B. Dong. Beyond finite layer neural networks: Bridging deep architectures and numerical differential equations. In *Proceedings of the 35th International Conference on Machine Learning*, volume 80, pages 3276–3285. PMLR, 2018.

Giacomo Meanti, Luigi Carratino, Lorenzo Rosasco, and Alessandro Rudi. Kernel methods through the roof: Handling billions of points efficiently. In Hugo Larochelle, Marc'Aurelio Ranzato, Raia Hadsell, Maria-Florina Balcan, and Hsuan-Tien Lin, editors, *Advances in Neural Information Processing Systems 33: Annual Conference on Neural Information Processing Systems 2020, NeurIPS 2020, December 6-12, 2020, virtual*, 2020. URL https://proceedings.neurips.cc/paper/2020/hash/a59afb1b7d82ec353921a55c579ee26d-Abstract.html.

Xiangming Meng and Yoshiyuki Kabashima. Diffusion Model Based Posterior Sampling for Noisy Linear Inverse Problems. nov 2022. URL http://arxiv.org/abs/2211.12343.

Thomas Mollenhoff and Daniel Cremers. Lifting vectorial variational problems: A natural formulation based on geometric measure theory and discrete exterior calculus. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2019.

Saviz Mowlavi and Saleh Nabi. Optimal control of PDEs using physics-informed neural networks. *J. Comput. Phys.*, 473:Paper No. 111731, 22, 2023. ISSN 0021-9991. doi: 10.1016/j.jcp.2022.111731. URL https://doi.org/10.1016/j.jcp.2022.111731.

Kimia Nadjahi, Alain Durmus, Umut Simsekli, and Roland Badeau. Asymptotic guarantees for learning generative models with the sliced-wasserstein distance. *Advances in Neural Information Processing Systems*, 32, 2019.

Kimia Nadjahi, Alain Durmus, Lénaïc Chizat, Soheil Kolouri, Shahin Shahrampour, and Umut Simsekli. Statistical and topological properties of sliced probability divergences. *Advances in Neural Information Processing Systems*, 33:20802–20812, 2020.

E. Newman, L. Ruthotto, J. Hart, and B. van Bloemen Waanders. Train like a (var)pro: Efficient training of neural networks by variable projection. *SIAM J. Math. Data Sci.*, 3:1041–1066, 2021.

David Palmer, Dmitriy Smirnov, Stephanie Wang, Albert Chern, and Justin Solomon. Deepcurrents: Learning implicit representations of shapes with boundaries. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 18665–18675, June 2022.

B.T. Polyak and P. Shcherbakov. Lyapunov functions: An optimization theory perspective. *IFAC-PapersOnLine*, 50(1):7456–7461, 2017.

A. F. Queiruga, N. B. Erichson, D. Taylor, and M. W. Mahoney. Continuous-in-depth neural networks. arXiv:2008.02389v1, 2020.

Nasim Rahaman, Devansh Arpit, Aristide Baratin, Felix Draxler, Min Lin, Fred A. Hamprecht, Yoshua Bengio, and Aaron C. Courville. On the spectral bias of deep neural networks. *CoRR*, abs/1806.08734, 2018. URL http://arxiv.org/abs/1806.08734.

M. Raissi, P. Perdikaris, and G. E. Karniadakis. Physics-informed neural networks: a deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *J. Comput. Phys.*, 378:686–707, 2019. ISSN 0021-9991. doi: 10.1016/j.jcp.2018.10.045. URL https://doi.org/10.1016/j.jcp.2018.10.045.

S. Reddi, S. Kale, and S. Kumar. On the convergence of ADAM and beyond. *preprint*

arXiv:1904.09237, 2019.

Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-resolution image synthesis with latent diffusion models, 2021.

Mark Rowland, Jiri Hron, Yunhao Tang, Krzysztof Choromanski, Tamas Sarlos, and Adrian Weller. Orthogonal estimation of wasserstein distances. In *The 22nd International Conference on Artificial Intelligence and Statistics*, pages 186–195. PMLR, 2019.

D. Ruiz-Balet and E. Zuazua. Neural ode control for classification, approximation and transport. SIAM Review, 2022.

L. Ruthotto and E. Haber. Deep neural networks motivated by partial differential equations. *J. Math. Imaging Vision*, 62:352–364, 2020.

O. Sebbouh, M. Cuturi, and G. Peyré. Randomized stochastic gradient descent ascent. In International Conference on Artificial Intelligence and Statistics, 2022.

V. Seguy, B. Bhushan Damodaran, R. Flamary, N. Courty, A. Rolet, and M. Blondel. Large-scale optimal transport and mapping estimation. preprint arXiv:1711.02283, 2017.

Stefan Sommer, François Lauze, and Mads Nielsen. Optimization over geodesics for exact principal geodesic analysis. *Advances in Computational Mathematics*, 40:283–313, 2014.

Yang Song, Jascha Sohl-Dickstein, Diederik P. Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole. Score-Based Generative Modeling through Stochastic Differential Equations. *(ICLR) International Conference on Learning Representations*, pages 1–36, nov 2021. ISSN 23318422. URL https://openreview.net/forum?id=PxTIG12RRHS.

W. Su, S. Boyd, and E. J. Candès. A differential equation for modeling Nesterov's accelerated gradient method: Theory and insights. *J. Mach. Learn. Res.*, 17:1–43, 2016.

Richard S. Sutton and Andrew G. Barto. *Reinforcement learning: an introduction*. Adaptive Computation and Machine Learning. MIT Press, Cambridge, MA, second edition, 2018. ISBN 978-0-262-03924-6.

Freyr Sverrisson, Jean Feydy, Bruno E. Correia, and Michael M. Bronstein. Fast end-to-end learning on protein surfaces. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 15272–15281, June 2021.

Jiqing Wu, Zhiwu Huang, Dinesh Acharya, Wen Li, Janine Thoma, Danda Pani Paudel, and Luc Van Gool. Sliced wasserstein generative models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3713–3722, 2019.

Z. Xie, X. Wang, H. Zhang, I. Sato, and M. Sugiyama. Adaptive inertia: Disentangling the effects of adaptive learning rate and momentum. In *ICML*. PMLR, 2022.

Guandao Yang, Xun Huang, Zekun Hao, Ming-Yu Liu, Serge Belongie, and Bharath Hariharan. Pointflow: 3d point cloud generation with continuous normalizing flows. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 4541–4550, 2019.

Xiaohui Zeng, Arash Vahdat, Francis Williams, Zan Gojcic, Or Litany, Sanja Fidler, and Karsten Kreis. Lion: Latent point diffusion models for 3d shape generation. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2022.

E. Zuazua. Control and machine learning. SIAM News, 55, 2022.

Miscellaneous

J. T. Betts. *Practical Methods for Optimal Control Using Nonlinear Programming*, volume DC36 of *Advances in Design and Control*. SIAM, third edition, 2020.