

1. Record Nr.	TD20018896
Autore	COLOMBO, TOMMASO
Titolo	Optimization techniques for large scale finite sum problems [Tesi di dottorato]
Lingua di pubblicazione	Inglese
Formato	Tesi di dottorato
Livello bibliografico	Monografia
Note	diritti: info:eu-repo/semantics/openAccess In relazione con info:eu-repo/semantics/altIdentifier/hdl/11573/1340957
Sommario	<p>With the explosion of machine learning and artificial intelligence applications, the need for optimization methods specialized in the training of such models has been steadily growing for the last 10-20 years. Indeed, given the big data regime and the special structure of the optimization problems to be solved in these settings, a number of new, efficient optimization methods have been developed. A large amount of these new methods strongly rely on the finite sum structure of the objective function to be minimized, where the indices <math>i=1, \dots, N</math> often refer to the availability of <math>N</math> input-output pairs on which the model should be trained, i.e. the training set. Nevertheless, this is not the only application where a finite sum structure of the objective function appears. Indeed, beyond the training of Neural Networks (NN) and Support Vector Machines (SVM), which depend by definition on a dataset of input-output pairs, a finite sum structure can also be recognized in Reinforcement Learning (RL) applications, due to the need of estimating expected values by sample approximation. In all these cases, <math>N</math> is usually huge, in the order of millions, or even billions, therefore making the exact computation of the function and gradient infeasible for many real life applications. This is one of the reasons why the field has seen a flourishing of publications from the most diverse</p>

communities, beyond the operations research one, for example the dynamical control, computer science, stochastic optimization ones. Many new methods have been developed by these communities, both deterministic and stochastic algorithms, although their comparison is made difficult by the different approaches coming from the different communities the new algorithms belong to. Due to the above considerations, the focus of this dissertation is on how to solve optimization problems where the function is structured as a finite sum of component functions. In this finite sum setting, a function  $f_i$  can be referred to as a component function, and its gradient  $\nabla f_i$  as a component gradient. In particular, a deep investigation of the algorithms developed so far to solve such problems is carried on, with a specific interest in showing the similarities and differences of the convergence analysis when it is developed in the deterministic vs stochastic cases. The target of the investigation is the case when the component gradients are continuously differentiable, and easily computable, like in many machine learning settings (e.g., neural networks training). In this framework, dynamic minibatching schemes are addressed. These are employed to determine the size of the sample to be used during the optimization process, especially in gradient-based methods, when the gradient is estimated by subsampling the component gradients, namely, when it is estimated based on a subset of the indices  $1, \dots, N$ . The aim of dynamic minibatching schemes is to dynamically test the quality of the gradient approximation, and consequently suggest if the sample size should grow or not. A new technique is proposed, based on statistical analysis of the gradient estimates. The new technique is based on the well-known Analysis of Variance (ANOVA) test, and the convergence of a subsampled gradient-based method is proved when such technique is employed. Numerical experiments are reported on standard machine learning tasks, like (nonlinear) regression and binary classification. Then, the derivative free setting is explored, i.e. the setting where the component functions come from a black-box-like process and the component gradients are not directly available. An example of such setting is policy optimization for reinforcement learning, where only sample approximations of the stochastic reward function are available. Therefore, in literature, Derivative Free Optimization (DFO) methods have been applied to solve this problem, in particular by trying to estimate the gradient by computing only sample approximations of the function. An analysis of the convergence guarantees of stochastic optimization methods in this setting is performed, showing that approximating the gradient by only computing sample-based estimates of the function brings a further approximation error, leading to poorer theoretical results. The special case of policy optimization for reinforcement learning is analysed, showing that such application is even harder, since the sample approximation of the function, in general, does not have continuity guarantees. Finally, a new class of distributed algorithms is introduced to solve linearly constrained, convex problems, with potential application to the dual formulation of the support vector machines training problem. This employs augmented Lagrangian and primal-dual theory to develop a simple, distributable and parallelizable class of algorithms to solve convex problems with simple bound and hard (i.e. coupling all the variables), linear constraints. Such class of algorithms is of particular interest for training support vector machines, since it allows to fully distribute the data, i.e. the input-output pairs, to the available parallel processes, simplifying the (often infeasible) storage of such large

amount of data.

---

Localizzazioni e accesso

<http://memoria.depositolegale.it/>\*/<http://hdl.handle.net/11573/1340957>

---