Convex Optimization with Random Pursuit

A thesis submitted to attain the degree of
Doctor of Sciences of ETH Zurich
(Dr. sc. ETH Zurich)

presented by
Sebastian Urban Stich

MSc ETH Mathematics
born on 21 May 1985
citizen of Kleinlützel SO, Switzerland

accepted on the recommendation of
Prof. Dr. Bernd Gärtner, examiner
Prof. Dr. Yurii Nesterov, co-examiner
Prof. Dr. Emo Welzl, co-examiner
Dr. Christian L. Müller, co-examiner

2014
Abstract

Optimization problems are ubiquitous in science and engineering. In this thesis, we study unconstrained black-box optimization problems that can only be accessed by an oracle that returns the function value at a query point. The theory of convex optimization problems is well-developed and such problems are typically solved with gradient-based methods. For non-convex problems, there is no unifying theoretical treatment and one has to rely on, typically gradient-free, search heuristics. Here, we analyze gradient-free optimization algorithms on convex functions.

In the first part of this thesis, we study Random Pursuit algorithms. These are iterative search schemes, where each iteration consists of two steps: (i) the generation of a (random) search direction and (ii) performing a step along this direction. We present a general framework to study such algorithms and prove convergence on smooth convex and strongly convex functions. The convergence rates depend on a sufficient decrease condition that measures the quality of the generated steps. This condition is for instance met by schemes that use a line search to generate the steps. For line search algorithms, we extend the convergence analysis to functions that are not necessarily everywhere strongly convex, but only at the optimum. Line search algorithms do not need any problem specific parameterization as input and are invariant under strictly monotone transformations of the objective functions. They thus enjoy identical convergence behavior on a wider function class. We discuss several kinds of random search directions and provide estimates for the expected convergence rates.

In the second part, we present three, at first sight seemingly unrelated, optimization algorithms that can be analyzed in the Random Pursuit framework. The examples comprise (i) solving linear systems with Kaczmarz’ method and (ii) Hessian learning with Leventhal and
Lewis’ estimation algorithm. Both these algorithms are instances of Random Pursuit algorithms with exact line search. We show this by demonstrating that these algorithms do only require the computation of scalar products, which in turn (iii) amounts to special Random Pursuit algorithms in Hilbert spaces, that have a simple geometric interpretation. We provide exact rates for the expected convergence. The Hessian learning scheme has a specific application: it can be used to estimate the underlying metric of an optimization problem which helps to accelerate the subsequent optimization with Random Pursuit. We do not only derive precise convergence rates, we also show that a specific implementation of this combined scheme converges equally fast on all quadratic functions, i.e. it is affine invariant.

In the last chapter, we review Nesterov’s gradient-based accelerated random search scheme. Each iteration of this scheme comprises two steps: (i) a simple search step like the one in the Random Pursuit algorithms and (ii) a model building step that allows for acceleration. We show that the step (i) can harmlessly be replaced with a line search, whereas the situation in step (ii) is more delicate. We cannot show that implementing step (ii) with a line search still yields acceleration, however, the resulting scheme does not diverge and converges on quadratic functions at least as fast as the simple Random Pursuit—with the possibility to accelerate.
Zusammenfassung


gen und schätzen für jede Verteilung die erwartete Konvergenzrate ab.


Im letzten Kapitel diskutieren wir Nesterovs gradientenbasierte Beschleunigungstechnik für zufallsgesteuerte Suchverfahren. Jede Iteration dieses Verfahrens besteht aus zwei Schritten: (i) einer einfachen Suche nach einem besseren Suchpunkt, wie in den Zufallsjagdverfahren, und (ii) der Aktualisierung einer Schätzung der Zielfunktion (Modell), welche die Grundlage bildet für die schnellere Konvergenz. Wir zeigen, dass in Schritt (i) gefährlos eine Liniensuche verwendet werden kann, aber wir können dies im allgemeinen Fall nicht auch für Schritt (ii) zeigen. Auf quadratischen Funktionen konvergiert das beschleunigte Verfahren mit Liniensuche mindestens gleich schnell wie die einfachen Zufallsjagdverfahren—möglicherweise aber deutlich schneller.