Stochastic Gradient Techniques for Optimization and Learning

Felisa J. Vázquez-Abad and Bernd Heidergott

University of Melbourne 2019

[Version: March 6, 2019]

Contents

Preface

Throughout this monograph we assume that random variables are defined on a common underlying probability space $(\Omega, \mathcal{F}, P)$. Furthermore, we assume that $\mathcal{F}$ contains all Null stets with respect to P.

By convention, we equip discrete spaces with the discrete topology and the real numbers with the usually topology. Product spaces are equipped with the product topology and, unless stated otherwise, measurable spaces are equipped with the corresponding Borel fields. If not stated otherwise, random variables are real-valued and, in line with the aforementioned conventions, measurable mappings from $(\Omega, \mathcal{F}, P)$ onto $(\mathbb{R}, \mathbb{B})$, with $\mathbb{B}$ denoting the Borel field on $\mathbb{R}$. Expectation of a random variable X with respect to $\mathbb{P}$ is denoted by $\mathbb{E}$, i.e., we write $\mathbb{E}[X] = \int_\Omega X(\omega)\mathbb{P}(d\omega)$. To simplify notation we suppress denoting ω when this causes no confusion. We use the symbol "$\sim$" to relate random variables and their corresponding distribution, i.e., we write $X \sim F$ if X has cumulative distribution function F. We use this notation in a similar way for measures. Equality in distribution of two random variables, say, X and Y is denoted by $X \stackrel{\mathcal{L}}{=} Y$.

We use the following abbreviations through:

- cdf for "cumulative distribution function"
- pdf for "probability density function"
- iid for "independently identically distributed"