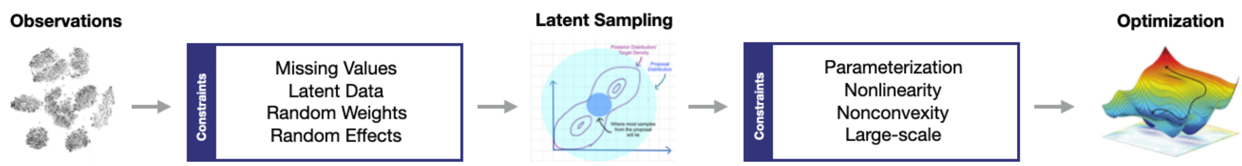
Research Statement

Belhal Karimi
Baidu Research USA

Throughout my research, I focus on developing *training*, also known as *optimization*, methods for large-scale datasets. There are several specificities to my work.
The broad panel of my work has applications on various problems, datasets and domains. To name a few, such learning task as stated above is crucial while fitting complex nonlinear models (mixed models, deep neural networks, mixture models) on tabular, image, textual data to tackle problems encountered in computer vision, drug development or natural language processing.
Based on the principled approach that consists of *observing* the world, *designing* a model describing the best those observations and *training* it on the latter, my main focal point in the realm of machine learning resides in the *training*, or *learning*, phase. With the sheer size of data and the high nonconvexity of the modern models, such as mutlilayer nerual network, used to describe complex human tasks, there is a rising interest and need for scalable, faster learning methods and their rigorous theoretical understanding.
Up to some observations, either fixed or streaming, and a well designed model, the definition of a loss/cost function and its optimization (minimization) are at the heart of this training phase. Continuously improving those optimization algorithms is key for *machine learning* in order to sustain the rapid growth in dimension, compositionality of the models and the high variety of input observations (sound, image, LIDAR, etc ...).



While my work provides *novel* methods for particularly deep neural networks (DNNs), one special case of the setting above, is when the input-output relationship of a phenomena is not completely characterized by the observations. A set of latent variables is thus needed and the loss function accepts the latter as a third argument.
Illustrative example of latent data model: During clinical trials, the kinetics and dynamics of a drug being tested are modeled using nonlinear functions (or systems of ordinary differential equations) and observations from patients which comprise for instance their gender, height, the concentration of the drug after injection. While those observed covariates are necessary, they are not sufficient to describe well the biological phenomena. A set of latent variables are used to quantify what can not be measured. In the special case of pharmacology, those latent variables describe the inter-individual variability among patients of a population (this is what makes us all different other than measurable signals). Therefore, the loss function, here the likelihood, is completed by simulations of those random effects and are then used to complete the observations before final optimization.
Thus, part of my research is at *the intersection* of **sampling** and **optimization**, bridging the gap between sampling methods such as Markov Chain Monte Carlo (MCMC) or Variational Inference and optimization method such as gradient-based learning algorithms or maximum likelihood estimation. My research has been published in top-tier conferences in machine learning such as NeurIPS, COLT, BAYSM, and made the object of contribution in statistics Journal such as CSDA. I also received a collection of awards from those conferences and a Jacques Hadamard grant for a summer visit to the Russian leading group in Bayesian Deep Learning called *BayesGroup*.