

## **Decent gradient is all you need**

**Yoram Singer, Princeton University**

### **Abstract**

The talk reviews old and new gradient-based learning algorithms and their use in machine learning. We start with a simple and intuitive view of gradient descent as a proximal operator. We then discuss its use for online learning and stochastic optimization of empirical objectives. Next, the notion of proximal methods is generalized to data-dependent operators and analyzed in the regret bound model. Last, we examine the generalization properties of gradient methods using simple stability properties. Travel-size description of gradient-based learning of convex and non-convex (a.k.a deep) models for speech, translation, neuro-engineering, and whatnot will be provided free of charge.