WEEKLY PARTICIPATION 8

Consider the following adaptive gradient methods:

(1) Heavy Ball Method

(2) Nesterov Accelerated Gradient Descent

(3) AdaGrad

Given an estimate $\boldsymbol{\omega}_t$ of the minimizer, these algorithms all propose a search direction to move in to construct $\boldsymbol{\omega}_{t+1}$. Which of these algorithms use different search directions from the one gradient descent uses?