

# Notes on Super Acceleration

## 1 Optimization Algorithms

Gradient descent

$$\theta_{i+1} = \theta_i - \eta \nabla L(\theta_i), \quad (1)$$

where  $\eta$  is the step length?

Momentum gradient descent

$$m_i = gm_{i-1} - (1 - g)\eta \nabla L(\theta_i) \quad (2)$$

$$\theta_{i+1} = \theta_i + m_i, \quad (3)$$

where  $g = 0.9$ .

Nesterov / Super acceleration

$$m_i = gm_{i-1} - (1 - g)\eta \nabla L(\theta_i + \nu m_{i-1}) \quad (4)$$

$$\theta_{i+1} = \theta_i + m_i, \quad (5)$$

where  $\nu$  is the acceleration parameter.

## 2 Model

We will focus our analysis first on a parabolic, one-dimensional function,  $L(x) = 1/2x^2$ . Then 4 reads

$$m_i = (g - (1 - g)\eta\nu)m_{i-1} - (1 - g)\eta x_i \quad (6)$$

$$x_{i+1} = x_i + m_i \quad (7)$$