Descent direction

Samuel Vaiter

Created: 2023-11-08.
Last update: 2023-11-08.
Status: inprogress.

(anti-)Gradient descent

A. Cauchy (1789–185

Gradient Descent algorithm

Require: Initialization $x^{(0)} \in \mathbb{R}^d$,
step-size policy $\eta^{(t)} > 0$.

1: **for** $t = 1, \cdots$ **do**

$$x^{(t+1)} = x^{(t)} - \eta^{(t)} \nabla f(x^{(t)}) \qquad \text{(GD)}$$

ANALYSE MATHÉMATIQUE. — *Méthode générale pour la résolution des systèmes d'équations simultanées; par* M. AUGUSTIN CAUCHY.

« Étant donné un système d'équations simultanées qu'il s'agit de résoudre, on commence ordinairement par les réduire à une seule, à l'aide d'éliminations successives, sauf à résoudre définitivement, s'il se peut, l'équation résultante. Mais il importe d'observer, 1° que, dans un grand nombre de cas, l'élimination ne peut s'effectuer en aucune manière; 2° que l'équation résultante est généralement très-compliquée, lors même que les équations données sont assez simples. Pour ces deux motifs, on conçoit qu'il serait

(anti-)Gradient descent

A. Cauchy (1789–185

Gradient Descent algorithm

Require: Initialization $x^{(0)} \in \mathbb{R}^d$,
step-size policy $\eta^{(t)} > 0$.

1: **for** $t = 1, \cdots$ **do**

$$x^{(t+1)} = x^{(t)} - \eta^{(t)} \nabla f(x^{(t)}) \qquad \text{(GD)}$$

ANALYSE MATHÉMATIQUE. — *Méthode générale pour la résolution des systèmes d'équations simultanées; par* M. AUGUSTIN CAUCHY.

« Étant donné un système d'équations simultanées qu'il s'agit de résoudre, on commence ordinairement par les réduire à une seule, à l'aide d'éliminations successives, sauf à résoudre définitivement, s'il se peut, l'équation résultante. Mais il importe d'observer, 1° que, dans un grand nombre de cas, l'élimination ne peut s'effectuer en aucune manière; 2° que l'équation résultante est généralement très-compliquée, lors même que les équations données sont assez simples. Pour ces deux motifs, on conçoit qu'il serait

Elephant in the room: choice of $\eta^{(t)}$?

$_{\text{(anti-)}}$Gradient descent

A. Cauchy (1789–185

Gradient Descent algorithm

Require: Initialization $x^{(0)} \in \mathbb{R}^d$,
step-size policy $\eta^{(t)} > 0$.

1: **for** $t = 1, \cdots$ **do**

$$x^{(t+1)} = x^{(t)} - \eta^{(t)} \nabla f(x^{(t)}) \qquad \text{(GD)}$$

ANALYSE MATHÉMATIQUE. — *Méthode générale pour la résolution des systèmes d'équations simultanées; par* M. Augustin Cauchy.

« Étant donné un système d'équations simultanées qu'il s'agit de résoudre, on commence ordinairement par les réduire à une seule, à l'aide d'éliminations successives, sauf à résoudre définitivement, s'il se peut, l'équation résultante. Mais il importe d'observer, 1° que, dans un grand nombre de cas, l'élimination ne peut s'effectuer en aucune manière; 2° que l'équation résultante est généralement très-compliquée, lors même que les équations données sont assez simples. Pour ces deux motifs, on conçoit qu'il serait

Elephant in the room: choice of $\eta^{(t)}$? 3 main choices (discussed later)

- Constant step-size $\eta^{(t)} = \eta$, or predetermined function
- "Oracle"
- Backtracking

Gradient descent: illustration on a quadratic problem