Gradient descent method

# Introduction

The gradient of at , denoted , if it is not a zero vector, is orthogonal to the tangent vector to an arbitrary smooth curve passing through on the level set . Showed as the picture below:

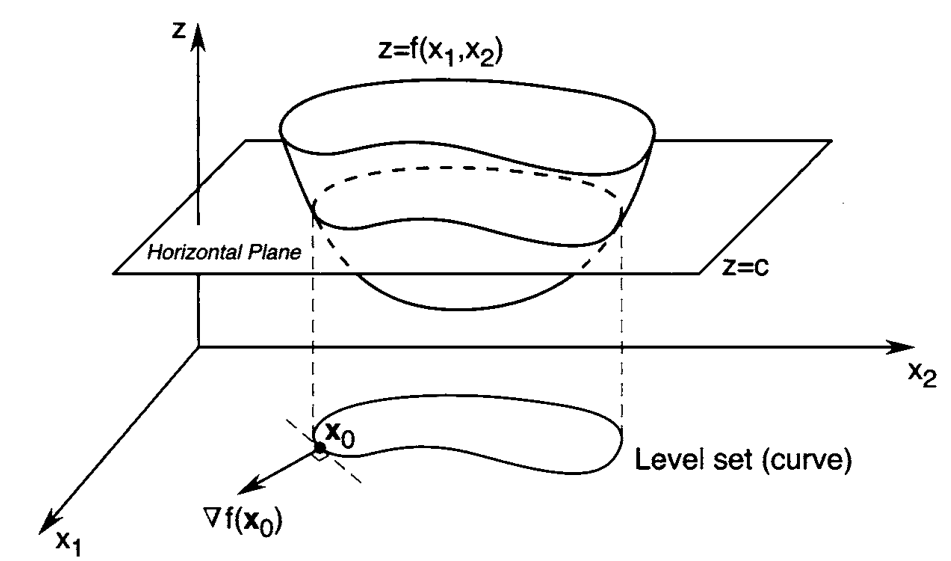


Figure Constructing a level set corresponding to level c for f

Thus, the direction of maximum rate of increase of a real-valued differentiable function at a point is orthogonal to the level set of the function through that point. In other words, the gradient acts in such a direction that for a given small displacement, the function  increases more in the direction of the gradient than in any other direction.

Proof:

Recall that , is the rate of increase of in the direction at the point . By the Cauchy-Schwarz inequality,



Because . But if , then



Thus, the direction in which points is the direction of maximum rate of increase of at . The direction in which  points is the direction of maximum rate of decrease of at . Hence, the direction of negative gradient is a good direction to search if we want to find a function minimizer.

Let be a starting point, and consider the point . Then, by Taylor’s theorem, we obtain



Thus, if , then for sufficiently small , we have



This means the point is an improvement over the point if we are searching for a minimizer.

To formulate an algorithm that implements this idea, suppose that we are given a point .To find the next point , we start at and move by an amount where is a positive scalar called the *step size*. This procedure leads to the following iterative algorithm:



We refer to this as a gradient descent algorithm (or simply a gradient algorithm). The gradient varies as the search proceeds, tending to zero as we approach the minimizer.