A.VIII. Optimization Algorithms

Olivier Scaillet

University of Geneva and Swiss Finance Institute

Outline

Introduction

The estimation of the parameters is usually obtained by

- minimization (nonlinear least squares, GMM)

or

- maximization (maximum likelihood)

of some criterion.

These criteria are usually highly *nonlinear* in the parameters, and numerical procedures are needed to compute the minimum or maximum.

Introduction

Let us focus on the maximization problem. Consider the twice differentiable objective function $Q(\theta)$ of the parameter θ

$\max Q(\theta) = \min -Q(\theta)$, so no generality is lost
Gradient: $\frac{\partial Q(\theta)}{\partial \theta}$, Hessian: $\frac{\partial^2 Q(\theta)}{\partial \theta \partial \theta'}$

A local maximum $\hat{\theta}$ is obtained by solving the first order condition (FOC):

$$\frac{\partial Q(\hat{\theta})}{\partial \theta} = 0$$

while the Hessian $\frac{\partial^2 Q(\hat{\theta})}{\partial \theta \partial \theta'}$ is negative definite.

All local maxima need to be found in order to determine among them the global maximum.

Gradient Methods

A numerical optimization algorithm is an iterative procedure that allows generation of sequence of approximations $\theta^{(0)}$, $\theta^{(1)}$, ..., $\theta^{(k)}$, ..., $\theta^{(K)}$ to the desired local maximum $\hat{\theta}$.

A numerical optimization algorithm is defined by

1. the method for choosing the *initial condition* $\theta^{(0)}$.
2. the *iteration formula* describing how the approximation $\theta^{(k+1)}$ at the $(k+1)$th step is computed from the approximation $\theta^{(k)}$ obtained in the preceding step.
3. the *stopping rule* determining the maximum number of iterations.

Gradient Methods

The central piece of an algorithm is the iteration formula, which takes the generic form

$$\theta^{(k+1)} = \theta^{(k)} + s^{(k)} d^{(k)}$$

The value $\theta^{(k+1)}$ is determined in two steps:

1. a *direction* $d^{(k)}$ of search is chosen for the function Q.
2. a *step size* $s^{(k)}$ along this direction is chosen so as to maximize the increase in the function Q.

In gradient methods, the search direction $d^{(k)}$ is proportional to the gradient, while the step size often involves the products of the gradient and the Hessian.

To stop the algorithm, a maximum number K of iterations is fixed indirectly by verifying whether several conditions are satisfied simultaneously

Gradient Methods: Stopping Conditions

1. Numerical convergence of the sequence, i.e., the difference between iterates $\theta^{(k)}$ and $\theta^{(k+1)}$ is sufficiently small

$$\left\| \theta^{(k+1)} - \theta^{(k)} \right\| \leq \varepsilon_1$$

2. Numerical convergence of the criterion function, i.e., the difference between $Q(\theta^{(k)})$ and $Q(\theta^{(k+1)})$ is small

$$\left| Q(\theta^{(k+1)}) - Q(\theta^{(k)}) \right| \leq \varepsilon_2$$

3. The value of the gradient $g(\theta^{(k)})$ is close to zero

$$\left\| g(\theta^{(k)}) \right\| \leq \varepsilon_3$$

$\varepsilon_1, \varepsilon_2, \varepsilon_3$ are tolerance levels determining precision.

Grid Search

Grid search:
Direct search methods evaluate the criterion function at various points and retain the point that gives the largest value of the criterion function.

This way of proceeding may be computer intensive if the dimension of parameter is large but avoids the computation of derivatives (gradient and Hessian)

The idea is to take a grid of values for θ, compute the function for these grid values, and determine the highest value of the function.

Then we refine the grid near that optimum and so on.