Part III
Hessian Estimation Schemes

Newton-based algorithms estimate both the gradient and the Hessian of the objective, and are in general seen to be more efficient than gradient-based algorithms as they exhibit fast convergence (in terms of the number of iterates). However, they require more computation than gradient-based schemes, because of the need to project the associated Hessian matrix to the space of positive definite and symmetric matrices at each update epoch and to invert the same. During the course of the last twelve years, there has been significant work done on developing efficient simultaneous perturbation-based Hessian estimators and Newton-based schemes. This part consists of two chapters – on Newton SPSA and Newton SF methods.

In a paper in 2000, Spall presented a Newton-based procedure involving an efficient four-simulation Hessian estimation procedure that relies on two independent perturbation sequences, each of which have similar properties to the one used for obtaining gradient SPSA estimates. This algorithm was presented for the case when the cost is an expectation over noisy cost samples. Subsequently, in a paper in 2005, Bhatnagar obtained three other SPSA-based Hessian estimates, that require three, two and one simulation(s), respectively. Together with the four simulation estimate presented by Spall, Bhatnagar proposed four Newton SPSA algorithms (using the aforementioned estimates in addition) for the long-run average cost objective. We discuss the Newton SPSA schemes in detail in Chapter 7.

Bhatnagar, in a paper in 2007, presented two Newton SF algorithms. He presented both algorithms for the case when the perturbation sequence used is i.i.d. and Gaussian distributed. By convolving the Hessian of the objective function with the N-dimensional Gaussian density and applying the integration-by-parts argument twice, one obtains the convolved Hessian as a convolution of the Hessian with a scaled N-dimensional Gaussian where the scaling matrix is obtained from the components of the N-vector Gaussian. The same simulation (with a Gaussian-perturbed parameter) is seen to estimate both the gradient as well as the Hessian of the objective. A two-simulation version of the same is seen to exhibit better performance in general. We discuss in detail the Newton SF schemes in Chapter 8.