Lesson 19

1. Can the Conjugate Gradient method be used to find the minimum of a general non-linear function?

   The basic conjugate gradient method is well-suited for finding the minimum of a quadratic form. However with some modifications the conjugate gradient method can be used for general nonlinear functions $f(\mathbf{x})$, provided the gradient $\nabla f(\mathbf{x})$ of the function $f(\mathbf{x})$ can be calculated.

2. What are the major modifications to the Conjugate Gradient method necessary for solving a general non-linear problem?

   The residual can no longer be evaluated as $\mathbf{r}_i = -\mathbf{A}\mathbf{e}_i$. The residual would need to be evaluated as:

   $$r_{i+1} = \nabla f(\mathbf{x}) - \nabla f(\mathbf{x}_{i+1}) = -\nabla f(\mathbf{x}_{i+1})$$

   In addition, it is no longer possible to find a closed form expression for the step-size. Instead we need to find the step size by solving a one-dimensional minimization problem. The procedure for finding the Gramm-Schmidt coefficient β_{ij} will also need to be modified, since in the absence of a constant coefficient matrix **A**, how to find β_{ij} becomes an open question.

3. What are some of the alternative methods for finding the Gramm-Schmidt coefficient β_{ij}? Do these methods guarantee convergence?

   Some alternative formulas for calculating the Gramm-Schmidt coefficients include the forms proposed by Polak-Ribiere as well as Fletcher-Reeves. The Polak-Ribiere formula is found to converge in almost all cases, though in rare situations it can result in the algorithm going into an infinite loop and failing to converge. The Fletcher – Reeves formula is guaranteed to converge if the starting point is close to the solution. However if the starting point is far from the minimum, convergence is not guaranteed.

4. Can preconditioning improve the performance of the conjugate gradient method for non-linear problems?

   As with the linear conjugate gradient method, convergence for the nonlinear algorithm can be improved by preconditioning. The diagonal matrix comprising the diagonal elements of the Hessian $(\nabla^2 f(\mathbf{x}_i))$ is sometimes used as a pre-conditioner, if all the diagonal elements are positive, since for convergence the preconditioner must be symmetric and positive definite.

5. Is the conjugate gradient method guaranteed to converge for a non-linear function?

   Convergence guarantees for the linear conjugate gradient algorithm are not valid for non-linear problems: the less quadratic the function $f(\mathbf{x})$ the faster is the loss of conjugacy of the search directions. Since a general nonlinear function can have many local minima – thus the conjugate gradient algorithm may converge to any of the local minima, depending on the starting point.