The following papers are classics in the field. Although many of them cover topics outside the material we shall have time to cover, they are all worth reading.

**Early quasi-Newton methods**

These methods were introduced by


and championed by


Although the so-called DFP method has been superseded by the more reliable BFGS method, it paved the way for a number of classes of important updates.

**More modern quasi-Newton methods**

Coincidently, all of the papers


appeared in the same year. The aptly-named BFGS method has stood the test of time well, and is still regarded as possibly the best secant updating formula.

**Quasi-Newton methods for large problems**

Limited memory methods are secant-updating methods that discard old information so as to reduce the amount of storage required when solving large problems. The methods first appeared in

Secant updating formulae proved to be less useful for large-scale computation, but a successful generalization, applicable to what are known as partially separable functions, was pioneered by A. Griewank and Ph. Toint, “Partitioned variable metric updates for large structured optimization problems”, *Numerische Mathematik* **39** (1982) 119:137, see also 429:448, as well as A. Griewank and Ph. Toint, “On the unconstrained optimization of partially separable functions”, in *Nonlinear Optimization 1981* (Powell, M., ed.) Academic Press (1982)

**Conjugate gradient methods for large problems**

Generalizations of Conjugate Gradient methods for non-quadratic minimization were originally proposed by

- R. Fletcher and C. M. Reeves, “Function minimization by conjugate gradients”, *Computer J.* (1964) 149:154, and

An alternative is to attempt to solve the (linear) Newton system by a conjugate-gradient like method. Suitable methods for terminating such a procedure while still maintaining fast convergence were proposed by


**Non-monotone methods**

While it is usual to think of requiring that the objective function decreases at every iteration, this isn’t actually necessary for convergence so long as there is some overall downward trend. The first method along these lines was by


**Trust-region methods**

The earliest methods that might be regarded as trust-region methods are those by

for the solution of nonlinear least-squares problems, although they are motivated from the perspective of modifying indefinite Hessians rather than restricting the step. Probably the first “modern” interpretation is by


Certainly, the earliest proofs of convergence are given by


while a good modern introduction is by


You might want to see our book


for a comprehensive history and review of the large variety of articles on trust-region methods.

**Trust-region subproblems**

Almost all you need to know about solving small-scale trust-region subproblems is contained in the seminal paper


Likewise


provides the basic truncated conjugate-gradient approach used so successfully for large-scale problems. More recently


show how to improve on Steihaug’s approach by moving around the trust-region boundary. A particularly nice new paper by


proves that Steihaug’s approximation gives at least 50% of the optimal function decrease when applied to convex problems.

**The Symmetric Rank-One quasi-Newton approximation**

Since trust-region methods allow non-convex models, perhaps the simplest of all Hessian approximation methods, the Symmetric Rank-One update, is back in fashion. Although it is unclear who first suggested the method,

---

1I would hate to claim “seminal” status for one of my own papers!

is the earliest reference that I know of. Its revival in fortune is due to


and it has now taken its place alongside the BFGS method as the pre-eminent updating formula.

**Non-monotone methods**

Non-monotone methods have also been proposed in the trust-region case. The basic reference here is the paper by


**Barrier function methods**

Although they appear to have originated in a pair of unpublished University of Oslo technical reports by K. Frisch in the mid 1950s, (logarithmic) barrier function were popularized by


A full early history is given in the book


The worsening conditioning of the Hessian was first highlighted by


although recent work by


See previous footnote . . .
demonstrates that this “defect” is far from fatal.

**Interior-point methods**

The interior-point revolution was started by


It didn’t take long for


to realize that this radical “new” approach was actually something that nonlinear programmers had tried (but, most unfortunately, discarded) in the past.

**SQP methods**

The first SQP method was proposed in the overlooked 1963 Harvard Master’s thesis of R. Wilson. The generic linesearch SQP method is that of


while there is a much larger variety of trust-region SQP methods, principally because of the constraint incompatibility issue.

**Merit functions for SQP**

The first use of an exact penalty function to globalize the SQP method was by

S. Han, “A globally convergent method for nonlinear programming”, *J. Optimization Theory and Appl*., 22 (1977) 297:309, and


The fact that such a merit function may prevent full SQP steps was observed N. Maratos in his 1978 U. of London Ph. D. thesis, while methods for combating the Maratos effect were subsequently proposed by


An SQP method that avoids the need for a merit function altogether by staying feasible is given by

### Hessian approximations

There is a vast literature on suitable Hessian approximations for use in SQP methods. Rather than point at individual papers, a good place to start is


but see also our paper


### Trust-region SQP methods

Since the trust-region and the linearized constraints may be incompatible, almost all trust-region SQP methods modify the basic SQP method in some way. The S\_1QP method is due to


Methods that relax the constraints include those proposed by


as well as a method that appeared in the 1989 U. of Colorado at Boulder Ph. D. thesis of E. Omojokun, supervised by R. Byrd. The highly original Filter-SQP approach was proposed by


while the analysis of a typical algorithm may be found in


### Modern methods for nonlinear programming

Many modern methods for nonlinearly constrained optimization tend to be SQP-interior-point hybrids. A good example is due to


and forms the basis for the excellent KNITRO package.