AN INTERVIEW WITH ROGER FLETCHER

Yu-Hong Dai

Roger Fletcher is one of the pioneers and leading figures in nonlinear optimization. He is the "F" of both DFP and BFGS quasi-Newton methods, and the "Fletcher" of Fletcher-Reeves nonlinear conjugate gradient method. Not to mention his many other sustained and fundamental achievements, his recent pioneering "filter" approach with Sven Leyffer renovated the algorithm design in nonlinear optimization once again. His famous books, *Practical Methods of Optimization (Volumes I and II)*, have deeply influenced both the optimization community and the applied sciences. Due to his illustrious accomplishments, he was awarded the prestigious Dantzig Prize in 1997 jointly by SIAM and the Mathematical Programming Society. He is also a Fellow of the Royal Society and of the Institute of Mathematics and its Applications. It is really my pleasure to interview him particularly about the topic of nonlinear conjugate gradient and quasi-Newton methods, thanks to the idea of Jorge Nocedal.

1. *As a start, can you briefly recall the situation of function minimization before 1960's?*

   Remember, I only finished my PhD in 1963, mostly not doing optimization. But there were the Simplex method for LP, steepest descent, sequential penalty functions, and Levenberg-Gauss-Newton.

2. *"The Computer Journal" published many early and important papers on optimization algorithms, including Fletcher and Powell (1963), Fletcher and Reeves (1964) and Fletcher (1970). In the same journal and also in 1963, however, you and your PhD supervisor, C. M. Reeves, published another paper on algebraic differentiation. My questions are, when and where did you obtain your degrees and what were your majors? Did this background of yours provide helps in the invention of algorithms?*

   I received my first degree from Cambridge University in 1960, in Natural Sciences, majoring in Theoretical Physics. I find this background helps a lot when talking to users of optimization, and has influenced my opinions on what aspects of mathematics are 'useful' and what are less so. I received my PhD in 1963 from Leeds University which had one of the

few early computer laboratories to be established in a university, centered around a Ferranti Pegasus computer. Colin Reeves supervised my work on a project to develop methods for computing molecular wavefunctions. It involved many interesting topics such as numerical eigensolutions, symbolic automatic differentiation, tensor algebra and groups, and latterly optimization. The resulting code was one of the forerunners of the large scale packages now used in quantum chemistry. Of course, the ideas were fed to me by Colin Reeves; my input was mainly in getting the programs to work! Certainly this helped my subsequent work in writing optimization codes.

3. *The circulation of the paper of Davidon (1959) was not good at that time. How did you get its copy and how did you feel its potential?*

Davidon issued his famous Argonne National Laboratories report in 1959. I don't know who received copies, but I was lucky that Colin Reeves somehow got one and passed it on to me. I coded it and realized it was able to solve problems substantially faster than steepest descent. Mike Powell also got a copy and also realized its potential. Davidon's report was presented in a very unusual way, and Mike was able to extract the essential feature that was involved. We pooled our experience and added some more theory, leading to the DFP paper.

4. *When and where did you meet Mike Powell and then you started the cooperation that led to the famous paper in 1963? How did you know Powell at that time?*

Mike presented a seminar at Leeds, and changed his title at the last moment to describe his experiences with Davidon's method. He found that I was also working on the method, hence the subsequent cooperation. I didn't know Mike previous to that.

5. *In his paper of symmetric rank-one in 1968, Davidon said, "The author is indebted to R. Fletcher and M. J. D. Powell for their contributions to the development of these basic ideas and for their having made the original method more accessible to others." On the other hand, yourself and Powell each dedicated a paper to Davidon on the occasion of his 70th birthday (see Math. Program., Ser. B, Vol. 87, No. 2, 2000). Would you like to talk more about Powell and Davidon?*

I have the greatest respect for both of them. Mike's influence on optimization and approximation theory has been immense, both in developing algorithms, and in establishing the underlying theory. I like very much that he is firmly concentrated on the ultimate aim of finding what works best and on making it available to users. Davidon brought the intuition

of a theoretical physicist to bear, and his ideas were very innovative. I enjoyed the company of both of them, and remember many happy outings on the hills in Scotland.

6. *The year 1970 witnessed the prosperity of quasi-Newton study. In this year, C. G. Broyden in Wales, yourself (at Harwell at that time), D. Goldfarb in New York and D. F. Shanno in Chicago proposed one quasi-Newton method simultaneously, which was called the BFGS method later and is now generally believed to be the most efficient quasi-Newton method. Can you briefly describe the approaches by which four of you proposed the method? When did BFGS get its name?*

Broyden saw that the method was one of a one parameter family of methods. I found the method from the Sherman-Morrison formula via a symmetry (duality) argument. Goldfarb established it by a variational principle. I don't remember Shanno's reasoning, but it was certainly different again. I don't remember when the acronym BFGS first came into use.

7. *Can you concisely explain to us why BFGS outperforms DFP?*

I think the paper by Mike Powell on "How bad are the BFGS and DFP methods when the function is quadratic?" sheds the most light on this issue. But my (very limited) experience with formulae outside the convex class indicates that one can possibly improve on BFGS by small amounts, and still retain positive definite Hessians.

8. *How is the relationship between you and the B, G and S?*

I know them all well and respect their work greatly. But I have not done any joint research with them.

9. *Now, my questions turn to conjugate gradient (CG). At first, could you briefly mention the CG study before your another seminal work jointly with C. M. Reeves in 1964?*

There were various publications on CG for $Ax = b$, in particular the celebrated Hestenes and Stiefel paper.

10. *By which chance did you and C. M. Reeves, your PhD supervisor, melt the two important contents, linear CG method and line search technique, resulting in the birth of the Fletcher-Reeves nonlinear CG method?*

Colin Reeves was writing lecture notes on CG for $Ax = b$ and realised that the line search aspect of the DFP method could be used to extend CG to solve nonquadratic optimization problems. Since I had a line search code I was able to follow this idea up for him by making some computations.

11. *Can you talk about your PhD supervisor, Professor C. M. Reeves?*

He was a student of Boys at Cambridge who worked on molecular wave-functions. He worked for a research group at ICI on theoretical chemistry for a short time, before coming to Leeds shortly before I did. He later moved into computer science at Keele University and I rather lost touch. He was a very innovative person and his move to CS was a great loss to numerical analysis.

12. *How did you look at the Fletcher-Reeves method after its proposition?*

It was clear early on that DFP and BFGS were superior to FR on small and medium sized problems. Both the FR and DFP methods were a significant improvement on steepest descent, and so attracted favourable comments.

13. *What comments have you about the nonlinear CG method currently, particularly for large problems?*

I have very little experience in using CG on large problems. Many large problems are solved more effectively using sparse matrix factors. Also CG doesn't fit comfortably with inequality constrained problems. Preconditioning can help, but finding a good preconditioner is not straightforward. I can only envisage using (preconditioned) CG when nothing else is practicable. This means large problems, for which the Hessian is usually ill-conditioned. Getting even six figures of accuracy in the gradients is usually quite a challenge. But the relationship $\delta = G^{-1}\gamma$ indicates that the accuracy in the variables might be much less. Unfortunately there's no easy way to check this out since if one could determine the solution accurately, one would use this technique in preference to CG. On the other hand, the occurrence of superlinear convergence in a Newton or quasi-Newton method is a good indication of an accurate solution.

14. *When I typed "Roger Fletcher" in the search engine of scholar google, the first item appeared is your work on conjugate gradient methods for indefinite linear systems in 1976, which has highly been cited by the community of numerical linear algebra. Can you list some of your most important papers or books?*

I have over 120 publications in my CV. Most of them seemed interesting at the time. Fashions change, and what seems important today may not be so in ten years time.

15. *Very recently, for the better use of filter technique (another original technique proposed by yourself and Sven Leyffer in 2002) in sequential quadratic programming, you turned back to quasi-Newton and proposed a hybrid of*

BFGS and SR1 with interesting properties (see Fletcher 2005). To which extent are you not satisfied with the current state of nonlinear CG and quasi-Newton? Would you like to comment about or predict what the picture of nonlinear CG and quasi-Newton should be twenty years later?

There is some evidence that the SR1 method converges faster than BFGS, especially when line searches are not used (say in a trust region context). However there is the problem of retaining a positive definite Hessian with SR1. My proposal in 2005 enables one to stay closer in a sense to SR1, whilst retaining definiteness. The low rank aspect enables the update to be used efficiently for large scale NLP when the dimension of the null space is small. However, I find it very hard to envisage significant new ideas in nonlinear CG and quasi-Newton methods.

16. *To broaden the topic to the whole mathematical programming, which directions or unsolved problems are you mostly interesting in presently?*

Nonlinear Complementarity and MPECs has interested me recently, but currently I'm still working on methods for NLP. I'd like to get back to working more on applications at some time.

17. *In addition to invent high-performance algorithms, you were also deeply involved in the production of high-quality software at AERE Harwell. Can you recall the circumstances when you were at Harwell? When and how did you come to Dundee?*

I worked at Harwell from 1969 to 1973, in company with Mike Powell and others. But Harwell was becoming commercialised so I took the opportunity to move back into academia, and join the very strong NA group at Dundee. I have been very happy here.

18. *Would you like to present a full list of your PhD students? How many PhD students from Asia do you have?*

I don't keep records on this, but here is a list of students with whom I have published joint papers.

Bill Bradbury (MSc), Tony McCann (MSc), Mike Hebden, Shirley Lill, Mike Jackson, Jim Sinclair, Rob Womersley, Paul Matthews, Mehi Al-Baali, Chenxian Xu, Eduardo Sainz de la Maza, Julian Hall, Sven Leyffer, Suliman Al-Homidan, Eric Chin.

Al-Baali is Syrian, Xu is Chinese, Al-Homidan is Saudi and Chin is Malaysian.

19. *Would you like to talk about your own family background? What kind of influence did you receive from your family so that you became a scientist?*

I was brought up in the austere times of the second world war and its aftermath. My father was killed in the war. My family was not well off, but were very supportive for me to make the most of my opportunities in life. I was lucky to attend a very good state (grammar) school which encouraged students to progress into university. I gained a state scholarship which enabled me to do this.

20. *What are your hobbies? Some people nicknamed optimization algorithms hill-climbing algorithms. Do you think whether your hobby of hill-climbing is helpful to your optimization research?*

I used to play chess at Cambridge and before that I played top board for the England junior side on one occasion. But I gave it up when I found I was being beaten by young players of half my age! I took up with bridge during my PhD and still play regularly. I have long been interested in music, and can play guitar, piano and clarinet, not very well, and not in public! Yes, I'm very keen on hill-walking and Scotland and the North of England is the ideal place to do it. Familiarization with maps, contours, local maxima and saddle points certainly helps in visualizing optimization techniques. But it doesn't help very much in understanding the complexity of high dimensional space.

21. *To round off the interview, what are the most important spirits for doing scientific researches do you think?*

Famous Grouse?! Alternatively, treat everything you read about with some scepticism, and be prepared to follow your own intuition. But be willing to change your mind when it becomes clear that other ideas have been demonstrated to be superior. And in Numerical Analysis, watch for what the numbers are telling you.

The above interview with Roger Fletcher has provided us with many details about the invention of nonlinear CG and quasi-Newton algorithms, together with his own comments about the past, present and future of the two kinds of classical optimization methods. The physics background of the DFP and Fletcher-Reeves methods is fascinating, which has deeply influenced Roger Fletcher's opinions on what aspects of mathematics are 'useful' and what are less so. From the interview, we also learn that although not coming from a wealthy family, he was able to progress to very good school and university education through his own efforts and the far sighted policy of the government and finally became a famous finger of varies powers. Besides his illustrious academic accomplishments, his subtlety of discerning things, his manner in doing researches, his way of getting with professions and his tact in answering questions all are very impressive and worth borrowing.

Acknowledgements.

References

[1] G.C. Broyden, The convergence of a class of double rank minimization algorithms: 2. the new algorithms, *J. Inst. Math. Appl.* 6 (1970) 222–231.

[2] W.C. Davidon, Variable metric method for minimization, Argonne National Laboratory, Report ANL-5990 (Rev.), Argonne, Illinois. (reprinted, with a new preface, in *SIAM J. Optimization* 1 (1991) 1–17).

[3] W.C. Davidon, Variance algorithms for minimization. *Computer J.* 10 (1968) 406–410.

[4] R. Fletcher, A new approach to variable metric algorithms, *Computer J.* 13 (1970) 317–322.

[5] R. Fletcher, Conjugate gradient methods for indefinite linear systems, *Lecture Notes in Math.* 506, Spring-Verlag, Berlin, Heildelberg, New York, 1976, pp. 73-89.

[6] R. Fletcher, *Practical Methods of Optimization, Vol. 1, Unconstrained Optimization*, John Wiley & Sons, Chichester, 1980.

[7] R. Fletcher, *Practical Methods of Optimization, Vol. 2, Constrained Optimization*, John Wiley & Sons, Chichester, 1981.

[8] R. Fletcher, Stable reduced Hessian updates for indefinite quadratic programming, *Math. Program., Series B* 87:2 (2000) 251–264.

[9] R. Fletcher, A new low rank quasi-Newton update scheme for nonlinear programming, Numerical Analysi Report NA/223, University of Dundee, 2005.

[10] R. Fletcher and S. Leyffer, Nonlinear programming without a penalty function, *Math. Program.* 91 (2002) 239–269.

[11] R. Fletcher and M.J.D. Powell, A rapidly convergent descent method for minimization, *Computer J.* 6 (1963) 163–168.

[12] R. Fletcher and C.M. Reeves, A mechanization of algebraic differentiation and the automatic generation of formulae for molecular integrals of Gaussian orbitals, *Computer J.* 3 (1963) 287–292.

[13] R. Fletcher and C.M. Reeves, Function minimization by conjugate gradients, *Computer J.* 7 (1964) 149–154.

[14] D. Goldfarb, A family of variable metric method derived by variation mean, *Math. Comput.* 23 (1970) 23–26.

[15] M.R. Hestenes and E. Stiefel, Methods of conjugate gradients for solving linear systems, *J. Res. Nat. Bur. Stand.* 49 (1952) 409–436.

[16] M.J.D. Powell, How bad are the BFGS and DFP methods when the objective function is quadratic? *Math. Program.* 34 (1986) 34–47.

[17] M.J.D. Powell, On the convergence of the DFP algorithm for unconstrained optimization when there are only two variables, *Math. Program., Series B* 87:2 (2000) 281–301.

[18] D.F. Shanno, Conditioning of quasi-Newton methods for function minimization, *Math. Comput.* 24 (1970) 647–656.

revised 16 December 2005

YU-HONG DAI
State Key Laboratory of Scientific and Engineering Computing,
Institute of Computational Mathematics and Sci./Engr. Computing,
Academy of Mathematics and Systems Science, Chinese Academy of Sciences,
P. O. Box 2719, Beijing 100080, P. R. China
E-mail address: dyh@lsec.cc.ac.cn