Nonlinear Techniques to Solve Linear Optimization Problems: Think of Interior Point Methods

Jacek Gondzio

Email: `J.Gondzio@ed.ac.uk`
URL: `https://www.maths.ed.ac.uk/~gondzio`

Outline

- Roger Fletcher and Michael Powell
- Quasi-Newton (QN) Methods
- Interior Point Methods (IPMs)
- Use Quasi-Newton Methods within IPMs

Part I: Two Influential Scholars

Roger Fletcher
1939-2016

Michael Powell
1936-2015

Biographical Memoirs of the Fellows of the Royal Society

Roger Fletcher. 29 January 1939—15 July 2016
by N. I. M. Gould and J. A. J. Hall
Published: 14 May 2025
https://doi.org/10.1098/rsbm.2024.0037

Michael J. D. Powell. 29 July 1936—19 April 2015
M. D. Buhmann, R. Fletcher, A. Iserles and P. Toint
Published: 31 January 2018
https://doi.org/10.1098/rsbm.2017.0023

All photos used in this talk come from the Royal Society archives.

Giants of Optimization & Numerical Analysis

- Wrote foundation papers
  $\rightarrow$ launched the field of **nonlinear optimization**
- Made seminal and ground-breaking contributions
- Recognized as recipients of multiple prizes
- Educated excellent students
- Left impressive legacy of works,
  which continue to influence our field today

PhD students of Roger Fletcher

- Shirley Lill, now Carter
  one of the co-founders of NAG Ltd.
- Sven Leyffer
  President of SIAM.
- Julian Hall
  creator of HiGHS software.
- Robert Womersley
- Paul Matthews
- Mehi Al-Baali
- Suliman Al-Homidan

PhD students of Michael Powell

- Philippe Toint
  former Chair of the Math. Optimization Society
- Ya-Xiang Yuan
  member of the Chinese Academy of Sciences
- Hans Martin Gutmann
  Lead OR Consultant at British Airways
- Martin Dietrich Buhmann
- Coralia Cartis
- Ioannis Demetriou

Davidon-Fletcher-Powell (DFP) method

In 1962 Fletcher and Powell independently studied a tech report of William Davidon from Argonne National Lab.

Powell gave a talk about it in Leeds.
Fletcher already knew the method and had a working code. They joined forces and wrote:

R. Fletcher and M.J.D. Powell (1963).
A rapidly convergent descent method for minimization,
The Computer Journal 6 (1963), 163-168.

It is today known as the **Davidon-Fletcher-Powell (DFP)** method.

Davidon, Fletcher and Powell

Quasi-Newton method(s)

The Hessian approximations in DFP method might become badly conditioned.

In 1970 four scholars (Broyden, Fletcher, Goldfarb and Shanno) independently arrived at the similar conclusion although they came from different motivations.

Fletcher applied the DFP method to the *inverse of the Hessian* and proposed a new approximation formula.

The resulting **Broyden-Fletcher-Goldfarb-Shanno (BFGS)** method is more stable and usually performs much better in practice than the DFP one.

Broyden, Fletcher, Goldfarb and Shanno

Part II: Quasi-Newton Methods

Quadratic Models

Let $f : \mathcal{R}^n \mapsto \mathcal{R}$ be a twice continuously differentiable but not necessarily convex function.

Its quadratic model

$$m_k(x_k + p) = f(x_k) + \nabla f(x_k)^T p + \frac{1}{2} p^T \nabla^2 f(x_k)\, p$$

is an approximation of f in the neighbourhood of x_k. Optimization of this quadratic model gives
Newton direction

$$p_k = -(\nabla^2 f(x_k))^{-1} \nabla f(x_k).$$

Quadratic Models (cont'd)

If

- $\nabla^2 f(x_k)$ is not positive definite, or
- $\nabla^2 f(x_k)$ is very expensive to compute, or
- the linear system with $\nabla^2 f(x_k)$ is too hard,

then instead of using exact $\nabla^2 f(x_k)$, we can use its **positive definite approximation**, B_k.

The quadratic model is then

$$m_k(x_k + p) = f(x_k) + \nabla f(x_k)^T p + \frac{1}{2} p^T B_k p$$

and its optimization gives the direction

$$p_k = -(B_k)^{-1} \nabla f(x_k).$$

Quasi-Newton Methods: Idea

These methods build (and continuously update) the approximation of $(\nabla^2 f(x_k))^{-1}$.

Let $\boldsymbol{H_k}$ be such an approximation.

The iteration k has the structure

(i) Set $p_k = -H_k g_k$.

(ii) Search along $x_k + \alpha p_k$ to give $x_{k+1} = x_k + \alpha_k p_k$.

(iii) Update H_k to give H_{k+1}.

Quasi-Newton Methods (cont'd)

To make such an approach viable, the approximations H_k should be such that

- The multiplication operation $H_k g_k$ and the update H_k to H_{k+1} are "easy";
- H_k gradually becomes a more and more accurate approximation of $(\nabla^2 f(x_k))^{-1}$.

Approximate $(\nabla^2 f(x_k))^{-1}$ using $\nabla f(x_k)$.

Use **symmetric low rank updates**.

Quasi-Newton Condition

f is twice continuously differentiable so

$$\nabla f(x_k + p) = \nabla f(x_k) + \nabla^2 f(x_k)\, p + o(\|p\|).$$

Therefore for $p = x_{k+1} - x_k$ we write

$$\nabla f(x_{k+1}) - \nabla f(x_k) \approx \nabla^2 f(x_k)(x_{k+1} - x_k).$$

In other words,

$$(\nabla^2 f(x_k))^{-1} \gamma_k \approx \delta_k,$$

where $\gamma_k = \nabla f_{k+1} - \nabla f_k$ and $\delta_k = x_{k+1} - x_k$.
The **quasi-Newton** method builds the approximations which satisfy the condition

$$\boldsymbol{H_{k+1}\, \gamma_k = \delta_k.}$$

Quasi-Newton: Hereditary Property

The quasi-Newton method builds the approximations which satisfy the **hereditary property**

$$H_{k+1}\gamma_j = \delta_j, \qquad j = 1, \ldots, k.$$

In other words, in iteration k the quasi-Newton condition holds not just for the pair of vectors $\{\delta_k, \gamma_k\}$ but also for the pairs $\{\delta_j, \gamma_j\}$, $j = 1, \ldots, k-1$ obtained in all previous iterations.

Different Quasi-Newton Updates

The **rank one** formula (1968)

$$H^{(new)} = H + \frac{(\delta - H\gamma)(\delta - H\gamma)^T}{(\delta - H\gamma)^T\gamma}.$$

The **DFP** formula (1963)

$$H^{(new)}_{\text{DFP}} = H + \frac{\delta\delta^T}{\delta^T\gamma} - \frac{H\gamma\gamma^T H}{\gamma^T H\gamma}.$$

The **BFGS** formula (1970)

$$H^{(new)}_{\text{BFGS}} = H + \left(1 + \frac{\gamma^T H\gamma}{\delta^T\gamma}\right)\frac{\delta\delta^T}{\delta^T\gamma} - \left(\frac{\delta\gamma^T H + H\gamma\delta^T}{\delta^T\gamma}\right).$$

The DFP and BFGS formulae guarantee that $H^{(new)} \succeq 0$.

Impact of Quasi-Newton Methods

- Powerful techniques to solve
  - nonlinear optimization problems
  - systems of nonlinear equations
- Limited Memory QN Methods (Nocedal et al.)
  $\rightarrow$ well-suited to very large scale optimization
- Provide potent preconditioners (Bergamaschi et al.)
- Extended to
  - nondifferentiable functions (Overton)
  - stochastic case (Nocedal, Curtis)

Part III: Interior Point Methods

Interior Point Methods

- re-born in 1984
- **Narendra Karmarkar**, AT&T Bell Labs
  "the simplex [method] is complex"

Surprising mathematical concept:

Take *linear* optimization problem
and add *nonlinear* function to the objective.

A step against common sense and
the centuries of mathematical practice:

"nonlinearize" the linear problem

Practical Optimization Problems

LP/QP:

$$\begin{array}{ll} \min & c^T x + \frac{1}{2} x^T Q x \\ \text{s.t.} & Ax = b, \\ & x \geq 0. \end{array}$$

SOCP:

$$\begin{array}{ll} \min & c^T x \\ \text{s.t.} & Ax = b, \\ & x \in \mathcal{K}, \qquad \mathcal{K} \ \text{ is a cone.} \end{array}$$

SDP:

$$\begin{array}{ll} \min & C \bullet X \\ \text{s.t.} & \mathcal{A} X = b, \\ & X \succeq 0, \quad X \in \mathcal{SR}^{n \times n}. \end{array}$$

Logarithmic Barrier Functions

For $x \geq 0$, use

$$f(x) = \begin{cases} -\ln(x) & \text{if } x > 0 \\ +\infty & \text{otherwise.} \end{cases}$$

For the quadratic cone

$$K_q = \{(x,t) : x \in \mathcal{R}^{n-1}, t \in \mathcal{R},\ t^2 \geq \|x\|^2,\ t \geq 0\},$$

use

$$f(x,t) = \begin{cases} -\ln(t^2 - \|x\|^2) & \text{if } \|x\| < t \\ +\infty & \text{otherwise.} \end{cases}$$

For the cone of positive definite matrices, use

$$f(X) = \begin{cases} -\ln\det X & \text{if } X \succ 0 \\ +\infty & \text{otherwise.} \end{cases}$$

LP: Replace $x \geq 0$ with $-\mu \sum_{j=1}^n \ln x_j$.

SDP: Replace $X \succeq 0$ with $-\mu \sum_{j=1}^n \ln \lambda_j = -\mu \ln(\prod_{j=1}^n \lambda_j)$.

Self-concordant Barrier

Def: Let $C \in \mathcal{R}^n$ be an open nonempty convex set. Let $f : C \mapsto \mathcal{R}$ be convex, 3 times continuously diff'able. A function f is called **self-concordant** if there exists a constant $p > 0$ such that

$$|\nabla^3 f(x)[h,h,h]| \leq 2p^{-1/2}(\nabla^2 f(x)[h,h])^{3/2},$$

$\forall x \in C, \forall h : x + h \in C.$ (f is p-self-concordant).

A self-concordant function is well approximated by the quadratic model because the error of such an approximation can be bounded by the 3/2 power of $\nabla^2 f(x)[h,h]$.

Nesterov and Nemirovskii,
Interior Point Polynomial Algorithms in Convex Programming: Theory and Applications, SIAM, 1994.

Logarithmic Barrier in Action $\rightarrow$ IPMs

From LP to Barrier LP

LP

$$\begin{array}{ll} \min & c^T x \\ \text{s.t.} & Ax = b, \\ & x \geq 0; \end{array}$$

Barrier LP

$$\begin{array}{ll} \min & c^T x - \mu \sum_{j=1}^{n} \ln x_j \\ \text{s.t.} & Ax = b. \end{array}$$

First Order Optimality Conditions

LP

$$\begin{aligned} Ax &= b \\ A^T y + s &= c \\ XSe &= 0, \end{aligned}$$

Barrier LP

$$\begin{aligned} Ax &= b \\ A^T y + s &= c \\ XSe &= \boldsymbol{\mu e}, \end{aligned}$$

$X = diag\{x_1, \cdots, x_n\}$, $S = diag\{s_1, \cdots, s_n\}$, $e = (1, \cdots, 1) \in \mathcal{R}^n$.

Complementarity $\quad x_j \cdot s_j = 0 \quad \forall j = 1, 2, ..., n.$

Simplex Method guesses an optimal partition:

For *basic* variables, $s_B = 0$ and

$$(x_B)_j \cdot (s_B)_j = 0 \quad \forall j \in \mathcal{B}.$$

For *non-basic* variables, $x_N = 0$ hence

$$(x_N)_j \cdot (s_N)_j = 0 \quad \forall j \in \mathcal{N}.$$

Interior Point Method uses ε-mathematics:

Replace $\quad x_j \cdot s_j = 0 \quad \forall j = 1, 2, ..., n$

by $\quad x_j \cdot s_j = \boldsymbol{\mu} \quad \forall j = 1, 2, ..., n.$

Force convergence $\boldsymbol{\mu} \rightarrow 0.$

Overarching Feature of IPMs

IPMs possess an unequalled ability to identify
the **“essential subspace”**
in which the optimal solution is hidden.

Worst-Case Complexity Results for LP/QP/SOCP/SDP

- *Feasible* methods converge in $\mathcal{O}(\sqrt{n})$ iterations
- *Infeasible* methods converge in $\mathcal{O}(n)$ iterations

Part IV:
Use Quasi-Newton Methods within Interior Point Methods

joint work with Francisco Sobral

IPM for LP

Primal-Dual Pair of Linear Programs

Primal

$$\begin{array}{rrcl} \min & c^T x & & \\ \text{s.t.} & Ax & = & b, \\ & & x \geq 0; & \end{array}$$

Dual

$$\begin{array}{rrcl} \max & b^T y & & \\ \text{s.t.} & A^T y + s & = & c, \\ & & s \geq 0. & \end{array}$$

First Order Opt Conditions for Barrier LP

$$\begin{aligned} Ax &= b, \\ A^T y + s &= c, \\ XSe &= \boldsymbol{\mu e}, \\ (x, s) &> 0, \end{aligned}$$

Define function $F : \mathcal{R}^{2n+m} \to \mathcal{R}^{2n+m}$

$$F(x, y, s) = (A^T y + s - c, Ax - b, XSe),$$

where $X = \mathrm{diag}(x)$, $S = \mathrm{diag}(s) \in \mathcal{R}^{n\times n}$ and $e \in \mathcal{R}^n$ is the vector of ones.
We seek a solution of

$$F(x, y, s) = (0, 0, \sigma\mu e),$$

with $0 < \sigma < 1$.

At iteration k, IPM computes the Newton direction:

$$\underbrace{\begin{bmatrix} 0 & A^T & I \\ A & 0 & 0 \\ S_k & 0 & X_k \end{bmatrix}}_{J_k = J(x_k, y_k, s_k)} \begin{bmatrix} \Delta x_k \\ \Delta y_k \\ \Delta s_k \end{bmatrix} = \begin{bmatrix} c - s_k - A^T y_k \\ b - Ax_k \\ \sigma_k \mu_k e - X_k S_k e \end{bmatrix}.$$

The same in QN notation

Let $J_k = J(x_k, y_k, s_k)$, $H_k = (J_k)^{-1}$ and write

$$\begin{bmatrix} \Delta x_k \\ \Delta y_k \\ \Delta s_k \end{bmatrix} = H_k \begin{bmatrix} c - s_k - A^T y_k \\ b - A x_k \\ \sigma_k \mu_k e - X_k S_k e \end{bmatrix}.$$

New H should satisfy

$$\delta_k = H_{k+1} \gamma_k,$$

where

$$\delta_k = \alpha_k \begin{bmatrix} \Delta x_k \\ \Delta y_k \\ \Delta s_k \end{bmatrix} \quad \text{and} \quad \gamma_k = \begin{bmatrix} \alpha_k (A^T \Delta y_k + \Delta s_k) \\ \alpha_k A \Delta x_k \\ X_{k+1} S_{k+1} e - X_k S_k e \end{bmatrix}.$$

Broyden “bad” rank-one approximation

The Broyden “bad” method uses the following update:

$$H_{k+1} = H_k + \frac{(\delta_k - H_k\gamma_k)\gamma_k^T}{\gamma_k^T\gamma_k}$$

$$= H_k V_k + \frac{\delta_k\gamma_k^T}{\rho_k},$$

where $V_k = \left(I - \frac{\gamma_k\gamma_k^T}{\rho_k}\right)$ and $\rho_k = \gamma_k^T\gamma_k$.

It is a rank-1 update:
H_{k+1} matrix satisfies the secant equation, and
H_{k+1} is the closest to H_k in the Frobenius norm.

A sequence of ℓ updates

We want to solve $J_k r = v$ or $r = H_k v$.

Given a perfect approximation $H_{k-\ell}$ of $J_{k-\ell}^{-1}$, after ℓ rank-one updates

$$H_k v = \underbrace{H_{k-\ell}}_{J_{k-\ell}^{-1}} \left(\prod_{j=k-\ell}^{k-1} V_j \right) v + \sum_{i=1}^{\ell} \left(\frac{\delta_{k-i}\gamma_{k-i}^T}{\rho_{k-i}} \prod_{j=k-i+1}^{k-1} V_j \right) v$$

where

$$V_j = \left(I - \frac{\gamma_j \gamma_j^T}{\rho_j} \right), \qquad \rho_j = \gamma_j^T \gamma_j$$

$$\delta_j = (x_{j+1}, y_{j+1}, s_{j+1}) - (x_j, y_j, s_j), \qquad \gamma_j = F_{j+1} - F_j$$

Computing

$$\begin{bmatrix} \Delta x_k \\ \Delta y_k \\ \Delta s_k \end{bmatrix} = H_k v \qquad \left(\approx J_k \begin{bmatrix} \Delta x_k \\ \Delta y_k \\ \Delta s_k \end{bmatrix} = v \right)$$

is equivalent to solving

$$\begin{bmatrix} 0 & A^T & I \\ A & 0 & 0 \\ S_{k-\ell} & 0 & X_{k-\ell} \end{bmatrix} \begin{bmatrix} \Delta x_k \\ \Delta y_k \\ \Delta s_k \end{bmatrix} =$$

$$v + \sum_{i=1}^{\ell} \beta_i \left(\begin{bmatrix} 0 & A^T & I \\ A & 0 & 0 \\ S_{k-\ell} & 0 & X_{k-\ell} \end{bmatrix} \delta_{k-i} - \gamma_{k-i} \right) =$$

$$v + \sum_{i=1}^{\ell} \beta_i \begin{bmatrix} 0 \\ 0 \\ S_{k-\ell}\delta_{k-i,x} + X_{k-\ell}\delta_{k-i,s} - \gamma_{k-i,\mu} \end{bmatrix}$$

Conceptual Quasi-Newton IPM

Input: F, J and (x_0, y_0, s_0)
for $k = 0, 1, \ldots$ **do**
  if k is *odd* **then**
    Use Quasi-Newton direction $(\Delta x_k, \Delta y_k, \Delta s_k)$
  else
    Use Newton direction $(\Delta x_k, \Delta y_k, \Delta s_k)$
  end if
  Make step:
  $(x_{k+1}, y_{k+1}, s_{k+1}) = (x_k, y_k, s_k) + \alpha_k(\Delta x_k, \Delta y_k, \Delta s_k)$
  for $\alpha_k \in [0, 1]$, such that $x_{k+1}, s_{k+1} > 0$
end for

→ Alternate between Newton and Quasi-Newton steps

Worst-Case Complexity Analysis

Example: Feasible algorithm, $\mathcal{N}_s$ neighbourhood:

Theorem.

Suppose that $(x_0, y_0, s_0) \in \mathcal{N}_s(\gamma)$ and $\gamma \geq \sigma_{min}/2$, where $0 < \sigma_{min} \leq \sigma_{max} < 1$ are user-defined constants such that $\sigma_k \in [\sigma_{min}, \sigma_{max}]$ for all iterations k of the algorithm.

If $\mu_0 = \epsilon^\kappa$, for a given $\epsilon > 0$ and a constant κ, then there exists $K > 0$, $\boldsymbol{K = O(n^3)}$, such that

$$\boldsymbol{\mu_k \leq \epsilon}$$

for all $\boldsymbol{k \geq K}$.

Summary of Theoretical Results

Method	Feasible			Infeasible	
Neighbourhood	$\mathcal{N}_2$	$\mathcal{N}_s$	$\mathcal{N}_{-\infty}$	$\mathcal{N}_s$	$\mathcal{N}_{-\infty}$
Newton	$O(\sqrt{n})$	$O(n)$	$O(n)$	-	$O(n^2)$
Quasi-Newton	$O(n)$	$O(n^3)$	$O(n^4)$	$O(n^5)$	$O(n^5)$

Gondzio and Sobral,
Quasi-Newton approaches to interior point methods for quadratic problems, *COAP* 74 (2019) pp 93–120.
`https://doi.org/10.1007/s10589-019-00102-z`
→ Implementation for (convex) QP.

Gondzio and Sobral,
Polynomial worst-case iteration complexity of quasi-Newton primal-dual interior point algorithms for linear programming, *COAP* 91 (2025) pp 649–681.
`https://doi.org/10.1007/s10589-024-00584-6`
→ Complete theoretical results.

Implementation: LPs and QPs

Theory: *one* QN update
Practice: ℓ QN updates

Problem	IPM		qnIPM		qnIPM-mcc	
	F	CPUt	F	CPUt	F	CPUt
dfl001	53	56.97	24	36.89	21	36.12
pilot87	31	4.24	10	3.28	11	3.58
qap8	12	0.66	5	0.44	4	0.40
qap12	20	34.23	12	23.05	9	19.12
tai12a	24	39.17	10	20.82	8	20.89
tai17a	24	1015.65	15	836.89	6	314.27

Roger Fletcher and Michael Powell

left an impressive legacy of works which continue
to influence the field of optimization.

Thank You!