NEWTON–TYPE METHODS: A BROADER VIEW

Mikhail Solodov

(IMPA, Rio)

joint work with

Alexey Izmailov

(Moscow State Univ.)

Terry Fest, Limoges, May 2015

Outline

- Newton method for (generalized) equations; Sequential Quadratic Programming (SQP)
- Other "fast" methods must be related? How?
- Perturbed Josephy–Newton framework for GE (perturbations = differences w.r.t. Newton)
  - Perturbed SQP framework for optimization
    - Various important SQP modifications (quasi-Newton, truncated, second-order corrections, composite-step, stabilized)
    - Linearly constrained Lagrangian methods
    - Inexact restoration methods
- Augmented Lagrangian methods (SOSC only)

The classical Newton method

For the (nonlinear) equation

$$\Phi(z) = 0,$$

with $\Phi : \mathbf{R}^\nu \to \mathbf{R}^\nu$ smooth,

in the Newton method, z^{k+1} is a solution of

$$\Phi(z^k) + \Phi'(z^k)(z - z^k) = 0.$$

- Requires $\Phi'(\bar{z})$ to be nonsingular

SQP for equality constraints

Consider the problem (where f, h are C^2)

$$\min\ f(x)\ \text{ s.t. }\ h(x) = 0.$$

In the SQP algorithm, x^{k+1} is a stationary point of

$$\begin{array}{ll} \min_x & \langle f'(x^k), x - x^k\rangle + \frac{1}{2}\langle L''_{xx}(x^k, \lambda^k)(x - x^k), x - x^k\rangle \\ \text{s.t.} & h(x^k) + h'(x^k)(x - x^k) = 0, \end{array}$$

and λ^{k+1} is an associated Lagrange multiplier.

This subproblem's optimality conditions are

$$f'(x^k) + L''_{xx}(x^k, \lambda^k)(x^{k+1} - x^k) + (h'(x^k))^\top \lambda^{k+1} = 0,$$

$$h(x^k) + h'(x^k)(x^{k+1} - x^k) = 0.$$

Re-writing the subproblem's optimality conditions

$$f'(x^k) + L''_{xx}(x^k, \lambda^k)(x^{k+1} - x^k) + (h'(x^k))^\top \lambda^{k+1} = 0,$$

$$h(x^k) + h'(x^k)(x^{k+1} - x^k) = 0,$$

as

$$\begin{pmatrix} L'_x(x^k, \lambda^k) \\ h(x^k) \end{pmatrix} + \begin{pmatrix} L''_{xx}(x^k, \lambda^k) & (h'(x^k))^\top \\ h'(x^k) & 0 \end{pmatrix} \begin{pmatrix} x^{k+1} - x^k \\ \lambda^{k+1} - \lambda^k \end{pmatrix} = 0,$$

we see that SQP is precisely the Newton iteration for

$$\Phi(z) = \big(L'_x(x,\, \lambda),\, h(x)\big) = 0.$$

- LICQ: $h'(\bar{x})$ has full rank (hence $\bar{\lambda}$ is unique)
- SOSC: $\langle L''_{xx}(\bar{x}, \bar{\lambda})d, d\rangle > 0 \quad \forall\, d \in \ker h'(\bar{x}) \setminus \{0\}$

LICQ + SOSC $\Rightarrow$ $\Phi'(\bar{z})$ nonsingular at $\bar{z} = (\bar{x}, \bar{\lambda})$, ...

SQP for equality/inequality constraints

Consider the problem

$$\min\ f(x) \ \text{ s.t. } \ h(x) = 0,\ g(x) \le 0,$$

where f, h, g are C^2.

In the SQP algorithm, x^{k+1} is a stationary point of

$$\begin{array}{ll} \min_x & \langle f'(x^k), x - x^k\rangle + \frac{1}{2}\langle L''_{xx}(x^k, \lambda^k, \mu^k)(x - x^k), x - x^k\rangle \\ \text{s.t.} & h(x^k) + h'(x^k)(x - x^k) = 0,\ g(x^k) + g'(x^k)(x - x^k) \le 0, \end{array}$$

and $(\lambda^{k+1}, \mu^{k+1})$ is an associated Lagrange multiplier.

Local primal-dual Q-superlinear convergence if

- SMFCQ: multiplier $(\bar{\lambda}, \bar{\mu})$ exists and is unique;
- SOSC: $\langle L''_{xx}(\bar{x}, \bar{\lambda}, \bar{\mu})d, d\rangle > 0 \quad \forall\, d \in C(\bar{x}) \setminus \{0\}$

SQP and the Josephy–Newton method

Consider the generalized equation (GE)

$$\Phi(z) + N(z) \ni 0,$$

where Φ is smooth and N is set-valued.

In the Josephy–Newton method (JNM),

z^{k+1} is a solution of

$$\Phi(z^k) + \Phi'(z^k)(z - z^k) + N(z) \ni 0.$$

SQP is a case of Josephy–Newton method, taking

$$\Phi(z) = \left(L'_x(x,\,\lambda,\,\mu),\, h(x),\, -g(x)\right),$$

$$N(z) = \begin{cases} \{0\} \times \{0\} \times \{y \in \mathbf{R}^m_+ \mid \langle \mu,\, y\rangle \le 0\}, & \text{if } \mu \ge 0; \\ \emptyset, & \text{otherwise.} \end{cases}$$

Josephy–Newton method $\rightarrow$ sharp results for SQP

JNM converges locally superlinearly to $\bar{z}$,

solution of GE,

if $\bar{z}$ is semistable + hemistable

(solvability of subproblems + distance estimate)

SMFCQ + SOSC

$\Downarrow$

$\bar{z} = (\bar{x}, \bar{\lambda}, \bar{\mu})$ is semistable and hemistable in GE-KKT

$\Downarrow$

SQP converges superlinearly

Note: weaker than LICQ, no strict complementarity!

J.F. Bonnans (AMOPT 1994)

Some considerations

- JNM covers SQP and gives these strong results
- JNM does not cover other methods...
  (JNM is precisely SQP)
- What about other methods? Especially, not explicitly Newtonian? (but “fast”)
- Relate other methods to JNM/SQP “a posteriori”
- Introduce perturbations in JNM/SQP
- These perturbations account for differences between other methods and JNM/SQP

Perturbed Josephy–Newton framework

$$\Phi(z) + N(z) \ni 0$$

The perturbed Josephy–Newton method (pJNM) is

$$\Phi(z^k)+\Phi'(z^k)(z-z^k)+\boxed{\Omega(z^k, z - z^k)}+N(z)+\boxed{\omega(z^k)} \ni 0.$$

- Ω represents structural perturbation, i.e., the difference between a given method and JNM.
- ω accounts for inexact solution of subproblems, e.g., truncation, etc.

pJNM converges superlinearly under appropriate assumptions about $\bar{z}$, Ω and ω.

A. Izmailov and M. Solodov (COAP 2010; Springer book 2014)

Perturbed SQP framework

Associated to pJNM is perturbed SQP (pSQP)

$$\begin{cases} L'_x(x^k,\, \lambda^k,\, \mu^k) + L''_{xx}(x^k,\, \lambda^k,\, \mu^k)(x - x^k) + \boxed{\Omega_L^k} = 0, \\ h(x^k) + h'(x^k)(x - x^k) + \boxed{\Omega_h^k} = 0, \\ \mu \geq 0, \quad g(x^k) + g'(x^k)(x - x^k) + \boxed{\Omega_g^k} \leq 0, \\ \langle \mu,\, g(x^k) + g'(x^k)(x - x^k) + \boxed{\Omega_g^k} \rangle = 0. \end{cases}$$

For $\Omega^k = 0$, this becomes KKT for usual SQP:

$$\min \quad \langle f'(x^k), x - x^k \rangle + \frac{1}{2} \langle L''_{xx}(x^k, \lambda^k, \mu^k)(x - x^k), x - x^k \rangle$$

s.t. $\quad h(x^k) + h'(x^k)(x - x^k) = 0,\; g(x^k) + g'(x^k)(x - x^k) \leq 0.$

Important: subproblems in pSQP need not be QPs!

A. Izmailov and M. Solodov (SIAM J. Optim 2010, Math Progr 2010; Springer book 2014) convergence of perturbed Josephy–Newton method

$\Downarrow$

convergence of perturbed SQP

(SMFCQ + SOSC;

perturbations must be “smooth” and “small”)

$\Downarrow$

convergence of specific algorithms

(often, under “better-than-usual” assumptions)

Specific algorithms, partial list

- Clearly Newtonian:
  - Quasi-Newton SQP, truncated SQP, with second-order corrections, stabilized SQP, ...
- Not-clearly Newtonian:
  - Linearly-constrained Lagrangian methods
  - Quadratically-constrained quadratic programming
- Not Newtonian-looking at all:
  - Inexact restoration methods
  - Augmented Lagrangian methods (methods of multipliers)

Quasi-Newton SQP

In quasi-Newton SQP, x^{k+1} is a stationary point of

$$\begin{array}{ll} \min & \langle f'(x^k),\, x - x^k\rangle + \frac{1}{2}\langle \boxed{\;H_k\;}(x - x^k),\, x - x^k\rangle \\ \text{s.t.} & h(x^k) + h'(x^k)(x - x^k) = 0, \\ & g(x^k) + g'(x^k)(x - x^k) \leq 0, \end{array}$$

and $(\lambda^{k+1}, \mu^{k+1})$ is an associated multiplier.

<u>Quasi-Newton SQP is a case of pSQP</u> with

$$\Omega_L^k = \Big(H_k - L''_{xx}(x^k,\, \lambda^k, \mu^k)\Big)\, d^k, \quad \Omega_h^k = 0,\ \Omega_g^k = 0.$$

Important: Can handle H_k being the Hessian of the Augmented Lagrangian!

Sharp Dennis–Moré type results

Given convergence, for primal superlinear rate

- Required assumptions (in our framework) are SOSC + Dennis–Moré condition
- No constraint qualifications! (previous literature needed LICQ here...)

Furthermore, for basic SQP with $L''_{xx}(x^k, \lambda^k, \mu^k)$, under SMFCQ + SOSC convergence of SQP had been established and Dennis–Moré condition is automatic. Thus, primal convergence rate of basic SQP is Q-superlinear (also a new result).

D. Fernández, A. Izmailov, and M. Solodov (SIAM J. Optim 2010)

Linearly constrained (augmented) Lagrangian methods

In LCL, x^{k+1} is a stationary point of

$$\min_x \quad \boxed{f(x) + \langle \lambda^k, h(x) \rangle + \frac{c_k}{2}\|h(x)\|^2}$$

$$\text{s.t.} \quad h(x^k) + h'(x^k)(x - x^k) = 0, \; x \geq 0,$$

$\lambda^{k+1} = \lambda^k + \eta^k$, η^k multiplier for equality constraint.

Note: subproblem is not a QP!

LCL is a case of pSQP with $\Omega_h^k = 0$, $\Omega_g^k = 0$ and

$$\begin{aligned}\Omega_L^k \;\; &= \;\; L'_x(x^k + d^k, \lambda^k) - L'_x(x^k, \lambda^k) - L''_{xx}(x^k, \lambda^k)d^k \\ &+ c_k(h'(x^k + d^k))^\top (h(x^k + d^k) - h(x^k) - h'(x^k)d^k)\end{aligned}$$

S.M. Robinson (MP 1972); B.A. Murtagh and M.A. Saunders (MPS 1982), MINOS software

Perturbed SQP $\rightarrow$ sharp results for LCL

SMFCQ + SOSC

$$\Downarrow$$

- Local primal-dual Q-superlinear convergence
- Local primal Q-superlinear convergence
- Inexact solution of subproblems (not QPs!)

Previous literature:

- Strict complementarity + LICQ + SOSC
- No primal Q-rate

A. Izmailov and M. Solodov (Springer book 2014)

Inexact Restoration Algorithms

Start with (conceptual) "Exact Restoration" scheme

Feasibility phase: π^k is a global solution of

$$\min_{\pi} \|\pi - x^k\| \quad \text{s.t.} \quad h(\pi) = 0,\ \pi \geq 0.$$

Optimality phase: x^{k+1} is a stationary point of

$$\begin{aligned} &\min\nolimits_x \quad f(x) + \langle \lambda^k, h(x) \rangle \\ &\text{s.t.} \quad h'(\pi^k)(x - \pi^k) = 0,\ x \geq 0, \end{aligned}$$

$\lambda^{k+1} = \lambda^k + \eta^k$, η^k multiplier for equality constraint.

Does not look Newtonian?

(two steps, general nonlinearities, ...)

J.M. Martínez et. al. (JOTA 2000, ..., SIOPT 2013)

Inexact Restoration Algorithms

$$\min \ \|\pi - x^k\| \quad \text{s.t.} \quad h(\pi) = 0, \ \pi \geq 0;$$

$$\min \ f(x) + \langle \lambda^k, \, h(x) \rangle \quad \text{s.t.} \quad h'(\pi^k)(x - \pi^k) = 0, \ x \geq 0.$$

"Exact Restoration" is a case of pSQP with $\Omega_g^k = 0$,

$$\begin{aligned} \Omega_L^k \ &= \ L'_x(x^k + d^k, \, \lambda^k) - L'_x(x^k, \, \lambda^k) - L''_{xx}(x^k, \, \lambda^k) d^k \\ &\quad + (h'(\pi^k) - h'(x^k))^\top (\lambda^{k+1} - \lambda^k), \end{aligned}$$

$$\Omega_h^k = (h'(\pi^k) - h'(x^k)) d^k + h(\pi^k) - h(x^k) - h'(\pi^k)(\pi^k - x^k),$$

Then,

Inexact Restoration is just "Exact Restoration",
with inexactness in solving subproblems in both phases!

A. Izmailov, A. Kurennoy, M. Solodov (Optim. Methods & Software 2014)

The Augmented Lagrangian algorithm

For the problem

(no inequality constraints here for simplicity only)

$$\min \ f(x) \quad \text{s.t.} \quad h(x) = 0,$$

in the method of multipliers x^{k+1} is given by

$$\min_x \ f(x) + \langle \lambda^k, h(x) \rangle + \frac{c_k}{2} \|h(x)\|^2,$$

and then the new multipliers estimate is

$$\lambda^{k+1} = \lambda^k + c_k h(x^{k+1}).$$

Does not look Newtonian at all?

(no part of problem data is being approximated)

Hestenes, Powell, Rockafellar, Bertsekas, LANCELOT and ALGELCAN software, ...

What does this have to do with Newton?

If solving the subproblem exactly, we have

$$0 = f'(x^{k+1}) + (h'(x^{k+1}))^\top(\lambda^k + c_k h(x^{k+1})) = L'_x(x^{k+1}, \lambda^{k+1}).$$

Informally speaking, this can only be "better" than SQP (which uses quadratic model of $L(\cdot, \lambda^k)$).

From the multipliers update,

$$\begin{aligned}\frac{1}{c_k}(\lambda^{k+1} - \lambda^k) &= h(x^{k+1}) \\ &= h(x^k) + h'(x^k)(x^{k+1} - x^k) + o(x^{k+1} - x^k),\end{aligned}$$

which is the perturbed SQP constraint.

Convergence of the augmented Lagrangian algorithm

Under SOSC only,

- Local primal-dual Q-linear for c_k large enough; superlinear for $c_k \to \infty$;
- Primal Q-rate is at least as fast is primal-dual.

Previous literature:

- Strict complementarity + LICQ + SOSC (or LICQ + strong SOSC)
- No primal Q-rate (only weaker R-rate)

D. Fernández and M. Solodov (SIAM J. Optim. 2012)

Conclusions

- A unified line of convergence analysis for
  - Newtonian methods (explicitly SQP related)
  - and not-so-Newtonian methods
    * Linearly constrained Lagrangian methods
    * Sequential qudratically constrained quadratic programming
  - and not-Newtonian-looking at all
    * Inexact Restoration methods
    * Augmented Lagrangian methods
- Often leads to improved convergence results.

Details:

http://www.impa.br/~optim/solodov.html

or

solodov@impa.br

Thanks!