SOME NEWS ON AUGMENTED LAGRANGIAN METHODS

Mikhail Solodov

(IMPA, Rio)

(mostly) joint work with

Damián Fernández

(Univ. de Córdoba)

Grenoble, November 2014

Outline

- The augmented Lagrangian method
  (the method of multipliers)
- Previous rate of convergence results:
  dual Q-linear and primal R-linear if
  - LICQ, strict complementarity, SOSC
- New rate of convergence results:
  primal-dual Q-linear and primal Q-linear if
  - SOSC only (no SC, no LICQ)
- Flavor of the analysis
  (Newtonian analysis of non-Newtonian method!)
- Numerical comparisons (6 solvers, MacMPEC)

The augmented Lagrangian algorithm

For the problem

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

given the multiplier estimates (λ^k, μ^k),
the method of multipliers computes x^{k+1} by

$$\begin{aligned} \min_x\ L_{c_k}(x, \lambda^k, \mu^k) := \quad & f(x) + \langle \lambda^k, h(x) \rangle + \frac{c_k}{2} \|h(x)\|^2 \\ & + \frac{1}{2c_k} \Big(\| \max\{0, \mu^k + c_k g(x)\} \|^2 - \|\mu^k\|^2 \Big) \end{aligned}$$

$$\lambda^{k+1} = \lambda^k + c_k h(x^{k+1}), \quad \mu^{k+1} = \max\Big\{0, \mu^k + c_k g(x^{k+1})\Big\}.$$

LANCELOT and ALGENCAN software; Hestenes, Powell, Rockafellar, Bertsekas, ...

The augmented Lagrangian algorithm

More generally, for the problem

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

the linear constraints are handled directly:

$$\min_{Ax\le a}\ L_{c_k}(x,\lambda^k,\mu^k) := \ f(x)+\langle\lambda^k,h(x)\rangle+\frac{c_k}{2}\|h(x)\|^2 + \frac{1}{2c_k}\left(\|\max\{0,\mu^k+c_k g(x)\}\|^2-\|\mu^k\|^2\right)$$

$$\lambda^{k+1}=\lambda^k+c_k h(x^{k+1}),\qquad \mu^{k+1}=\max\left\{0,\mu^k+c_k g(x^{k+1})\right\}.$$

The augmented Lagrangian algorithm

To shorten notation, define

$$Q = \{\nu \in \mathbf{R}^m \mid \nu_i \in \mathbf{R},\ i = 1, \dots, l;\ \nu_i \geq 0,\ i = l+1, \dots, m\},$$

then the problem becomes

$$\min\ f(x) \quad \text{s.t.} \quad g(x) \in Q^\circ,$$

and x^{k+1} is given by

$$\min_x\ L_{c_k}(x, \mu^k) := f(x) + \frac{1}{2c_k}\left(\left\|\Pi_Q\left(\mu^k + c_k g(x)\right)\right\|^2 - \|\mu^k\|^2\right),$$

with the new multiplier estimates given by

$$\mu^{k+1} = \Pi_Q\left(\mu^k + c_k g(x^{k+1})\right).$$

Standard rate of convergence statement

$$\|\mu^{k+1} - \bar{\mu}\| \leq \frac{q}{c_k}\|\mu^k - \bar{\mu}\| \quad \text{(dual } Q\text{-linear)}$$

$$\|x^{k+1} - \bar{x}\| \leq \frac{q}{c_k}\|\mu^k - \bar{\mu}\| \quad \text{(primal } R\text{-linear)}$$

if μ^0 close to $\bar{\mu}$, $c_k > \bar{c}$, and

- LICQ: active constraints gradients linearly idependent ($\bar{\mu}$ is unique),
- Strict complementarity: $\bar{\mu}_i > 0$ for all active inequality constraints,
- SOSC: $\left\langle \frac{\partial^2 L}{\partial x^2}(\bar{x}, \bar{\mu})d, d \right\rangle > 0 \quad \forall d \in C(\bar{x}) \setminus \{0\}$.

Some considerations

- These assumptions are strong
  (not much more could possibly be asked...)
- Not quite clear what for are LICQ and strict complementarity...
  (other than to employ certain proof techniques)
- SOSC is needed indeed...
  (gives local coercivity of $L_{c_k}(\cdot, \mu^k)$, which ensures the existence of the minimizer x^{k+1})
- Perhaps SOSC is all that is really needed? Yes!

New rate of convergence results

If μ^0 close to $\bar{\mu} \in M(\bar{x})$ with SOSC at $\bar{\mu}$, $c_k > \bar{c}$, then

$$\{(x^k, \mu^k)\} \to (\bar{x}, \hat{\mu}), \quad \hat{\mu} \in M(\bar{x}),$$

$$\frac{\|(x^{k+1}, \mu^{k+1}) - (\bar{x}, \hat{\mu})\|}{\|(x^k, \mu^k) - (\bar{x}, \hat{\mu})\|} \leq \frac{q}{c_k}$$ (primal-dual Q-linear)

$$\frac{\|x^{k+1} - \bar{x}\|}{\|x^k - \bar{x}\|} \leq \theta \, \frac{\|(x^{k+1}, \mu^{k+1}) - (\bar{x}, \hat{\mu})\|}{\|(x^k, \mu^k) - (\bar{x}, \hat{\mu})\|}$$ (primal Q-linear)

No LICQ, no strict complementarity!

Flavor of the analysis

The method does the following

$$\left\| \frac{\partial L_{c_k}}{\partial x}(x^{k+1}, \mu^k) \right\| \leq \epsilon_k, \qquad \mu^{k+1} = \Pi_Q(\mu^k + c_k g(x^{k+1})).$$

Observe

$$\begin{aligned} \frac{\partial L_{c_k}}{\partial x}(x^{k+1}, \mu^k) &= f'(x^{k+1}) + (g'(x^{k+1}))^\top \Pi_Q(\mu^k + c_k g(x^{k+1})) \\ &= \frac{\partial L}{\partial x}(x^{k+1}, \mu^{k+1}), \end{aligned}$$

so that

$$0 \in \frac{\partial L}{\partial x}(x^{k+1}, \mu^{k+1}) + B(0, \epsilon_k).$$

Flavor of the analysis

Therefore, the method is

$$0 \in \frac{\partial L}{\partial x}(x^{k+1}, \mu^{k+1}) + B(0, \epsilon_k), \qquad \mu^{k+1} = \Pi_Q(\mu^k + c_k g(x^{k+1})).$$

From the projection condition,

$$\mu^k + c_k g(x^{k+1}) - \mu^{k+1} \in N_Q(\mu^{k+1}).$$

And, puting together, the iteration satisfies the GE

$$0 \in \begin{bmatrix} \dfrac{\partial L}{\partial x}(x^{k+1}, \mu^{k+1}) \\ -g(x^{k+1}) + \dfrac{1}{c_k}(\mu^{k+1} - \mu^k) \end{bmatrix} + \begin{bmatrix} B(0, \epsilon_k) \\ N_Q(\mu^{k+1}) \end{bmatrix}.$$

Flavor of the analysis

Thus, the method is

$$0 \in \begin{bmatrix} \dfrac{\partial L}{\partial x}(x^{k+1}, \mu^{k+1}) \\ -g(x^{k+1}) + \dfrac{1}{c_k}(\mu^{k+1} - \mu^k) \end{bmatrix} + \begin{bmatrix} B(0, \epsilon_k) \\ N_Q(\mu^{k+1}) \end{bmatrix}.$$

In other words, (x^{k+1}, μ^{k+1}) solves (perturbed) GE

$$0 \in \begin{bmatrix} \dfrac{\partial L}{\partial x}(x, \mu) \\ -g(x) \end{bmatrix} + \begin{bmatrix} 0 \\ N_Q(\mu) \end{bmatrix} + \begin{bmatrix} B(0, \epsilon_k) \\ \dfrac{1}{c_k}(\mu^{k+1} - \mu^k) \end{bmatrix},$$

i.e., KKT optimality conditions + perturbation

For the problem

$$\min\ f(x) \quad \text{s.t.} \quad g(x) \in\ Q^{\circ},\ Ax \leq a,$$

we do

$$\min\ L_{c_k}(x, \mu^k) \quad \text{s.t.} \quad Ax \leq a,$$

$$\mu^{k+1} = \Pi_Q \left(\mu^k + c_k g(x^{k+1}) \right).$$

This means

$$0 \in \frac{\partial L_{c_k}}{\partial x}(x^{k+1}, \mu^k) + A^{\top} \nu^{k+1} + B(0, \epsilon_k),$$

$$Ax^{k+1} - a \leq 0, \quad \nu^{k+1} \geq 0, \quad \langle \nu^{k+1}, Ax^{k+1} - a \rangle = 0,$$

$$g(x^{k+1}) - \frac{1}{c_k}(\mu^{k+1} - \mu^k) \in N_Q(\mu^{k+1}).$$

Thus, for the problem

$$\min\ f(x) \quad \text{s.t.} \quad g(x) \in Q^\circ,\ Ax \le a,$$

$(x^{k+1}, \mu^{k+1}, \nu^{k+1})$ solves (perturbed) GE

$$0 \in \begin{bmatrix} \dfrac{\partial L}{\partial x}(x, \mu, \nu) \\ -g(x) \\ -(Ax - a) \end{bmatrix} + \begin{bmatrix} 0 \\ N_Q(\mu) \\ N_{\mathbf{R}^s_+}(\nu) \end{bmatrix} + \begin{bmatrix} B(0, \epsilon_k) \\ \dfrac{1}{c_k}(\mu^{k+1} - \mu^k) \\ 0 \end{bmatrix},$$

i.e., KKT conditions + (same) perturbation

Flavor of the analysis

Analysis of perturbed generalized equations

$$\mathrm{SOL}(p) = \{(x, \mu) \mid 0 \in G(x, \mu) + p + N(x, \mu)\}.$$

Under SOSC, the KKT GE is upper-Lipschitzian

$$\mathrm{SOL}(p) \cap \Delta \subset \mathrm{SOL}(0) + \tau_1 \|p\| B(0, 1) \qquad \text{for } p \text{ “small”}.$$

For KKT GE, the latter property is equivalent to

- Local error bound given by KKT residual
- The multiplier $\bar{\mu} \in M(\bar{x})$ is noncritical

A. Izmailov and M. Solodov, Math. Programming, Vol. 133, 2012.

Upper-Lipschitzian behavior of perturbed KKT GE

$$0 \in \begin{bmatrix} \dfrac{\partial L}{\partial x}(x,\mu) \\ -g(x) \end{bmatrix} + \begin{bmatrix} 0 \\ N_Q(\mu) \end{bmatrix} + p$$

$$\Updownarrow$$

$$\mathrm{dist}((x,\mu),\mathrm{SOL}) \le \beta \left\| \begin{bmatrix} \dfrac{\partial L}{\partial x}(x,\mu) \\ \mu - \Pi_Q(\mu + g(x)) \end{bmatrix} \right\|$$

$$\Updownarrow$$

$d = 0$ is the unique stationary point of

$$\min \left\langle \frac{\partial^2 L}{\partial x^2}(\bar{x},\bar{\mu})d, d \right\rangle \quad \text{s.t.} \quad d \in C(\bar{x})$$

Flavor of the analysis

is Newtonian:

- Prove existence of solutions (x^{k+1}, μ^{k+1}) of GE, given any μ^k close to $\bar{\mu}$;
- Prove that the primal-dual step is "small" (so that (x^{k+1}, μ^{k+1}) stays in the neighborhood):

$$\|(x^{k+1}, \mu^{k+1}) - (x^k, \mu^k)\| \le \tau_2 \mathrm{dist}((x^k, \mu^k), \mathrm{SOL});$$

- Then upper-Lipschitzian property gives

$$\mathrm{dist}((x^{k+1}, \mu^{k+1}), \mathrm{SOL}) \le \tau_1 \|p^k\| \le \ldots.$$

Somewhat related to Stabilized SQP, D. Fernández & M. Solodov, Math. Programming, 2010.

Flavor of the analysis

So we have (simplifying with $\epsilon_k = 0$)

$$\mathrm{dist}((x^{k+1}, \mu^{k+1}), \mathrm{SOL}) \le \frac{\tau_1}{c_k} \|\mu^{k+1} - \mu^k\|,$$

and

$$\|(x^{k+1}, \mu^{k+1}) - (x^k, \mu^k)\| \le \tau_2 \mathrm{dist}((x^k, \mu^k), \mathrm{SOL}).$$

From the latter,

$$\|\mu^{k+1} - \mu^k\| \le \tau_2 \mathrm{dist}((x^k, \mu^k), \mathrm{SOL}),$$

and hence,

$$\mathrm{dist}((x^{k+1}, \mu^{k+1}), \mathrm{SOL}) \le \frac{\tau_1 \tau_2}{c_k} \mathrm{dist}((x^k, \mu^k), \mathrm{SOL})$$

Finalizing,

$$\text{dist}((x^{k+1}, \mu^{k+1}), \text{SOL}) \le \frac{\tau_1 \tau_2}{c_k} \text{dist}((x^k, \mu^k), \text{SOL})$$

gives Q-linear convergence of the distance to the primal-dual solution set.

Further analysis shows

- Convergence of $\{(x^k, \mu^k)\}$ to $(\bar{x}, \hat{\mu})$, $\hat{\mu} \in M(\bar{x})$;
- Q-linear rate of convergence of $\{x^k\}$.

Boundedness of the penalty parameters

In ALGENCAN, for $\theta \in (0,1)$ and $\rho > 1$,

$$c_{k+1} = \begin{cases} c_k, & \text{if } \sigma(x^{k+1}, \mu^{k+1}) \le \theta \sigma(x^k, \mu^k), \qquad (*) \\ \rho c_k, & \text{otherwise,} \end{cases}$$

where

$$\sigma(x,\mu) = \left\| \begin{bmatrix} \dfrac{\partial L}{\partial x}(x,\mu) \\ \mu - \Pi_Q(\mu + g(x)) \end{bmatrix} \right\|$$

As a by-product of our analysis,

rate of $\sigma(x^k, \mu^k) \to 0$ is inversely proportionate to c_k.

Hence, the linear test (*) holds once c_k large enough.

On problems with complementarity constraints

$$\min \; f(x) \quad \text{s.t.} \quad H(x) \geq 0, \; G(x) \geq 0, \; \langle H(x), G(x) \rangle = 0$$

- no MFCQ (or LICQ) at any feasible point (hence, multiplier set always unbounded)
- RCPLD (and even CPLD) can hold... but only for artificial examples
- Augmented Lagrangian methods need no CQs... Attractive for MPCC? At least worth a try!

---

joint work with A. Izmailov and E. Uskov

Theoretical global convergence on MPCC

If MPCC-LICQ holds at a feasible primal accumulation point $\bar{x}$,

- Then $\bar{x}$ is strongly stationary (KKT) when a certain part of the dual sequence is bounded;
- Otherwise, it is at least C-stationary.

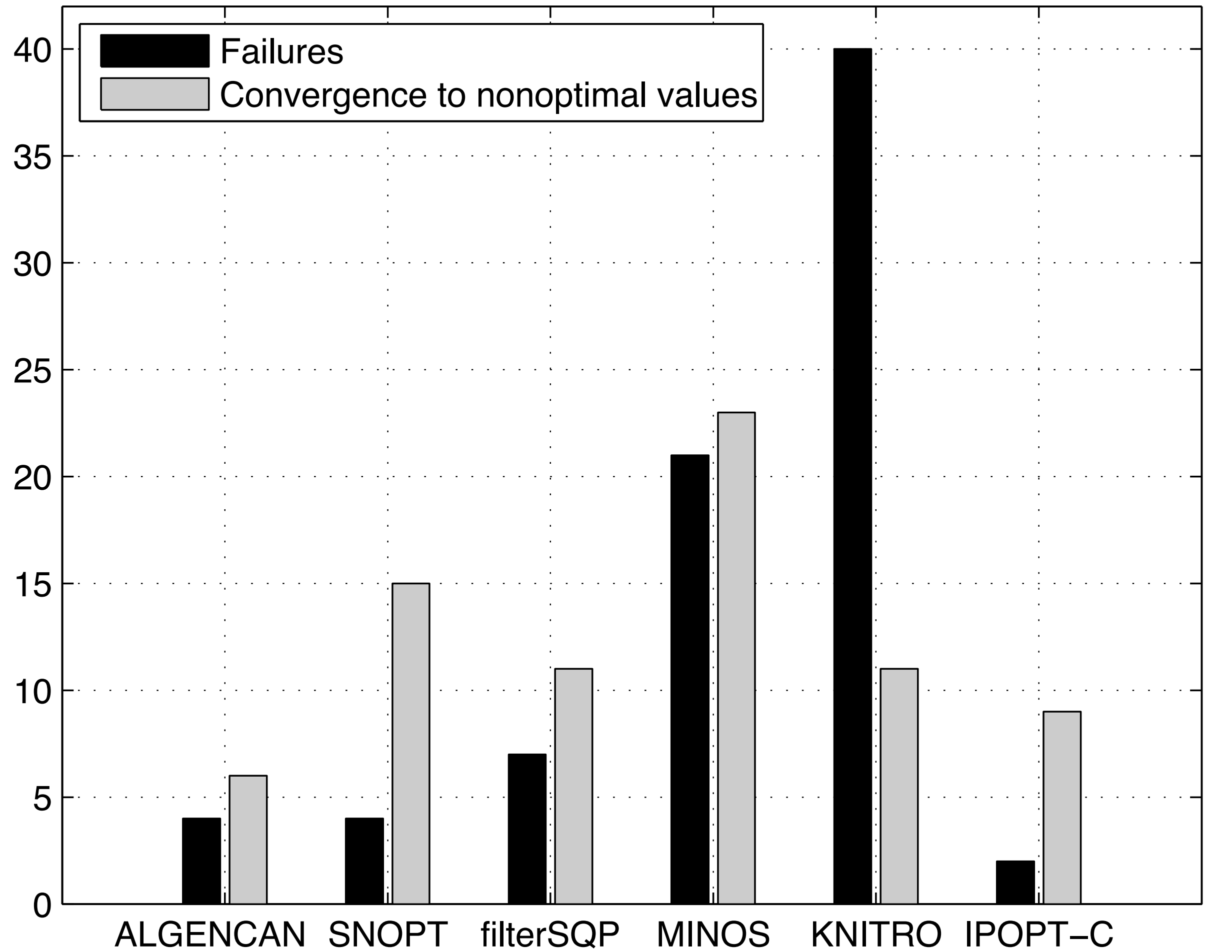
Considerations:

- Ideally, want strong stationarity. However...
- Seems impossible with usual NLP tools (without expensive combinatorial overhead)
- The stated is competitive with state-of-the-art (SQP can converge to arbitrary feasible point)
- In practice, augmented Lagrangian usually achieves strong stationarity.
  - many problems do not have C-stationary non-strongly stationary points
  - augmented Lagrangian has an intrinsic dual stabilization property which tends to keep dual sequences bounded (even for MPCC!)

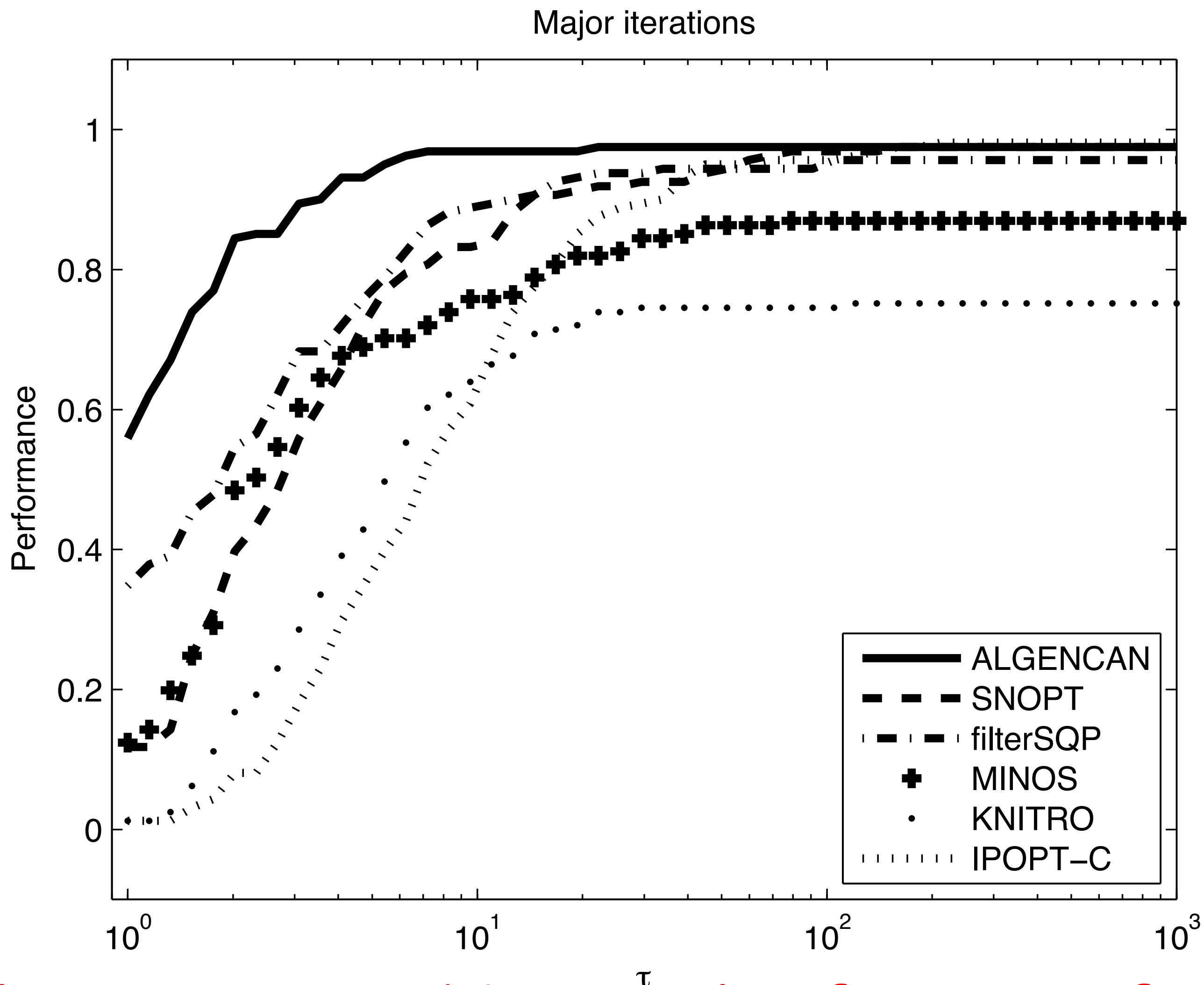
Comparisons on MacMPEC

6 solvers:

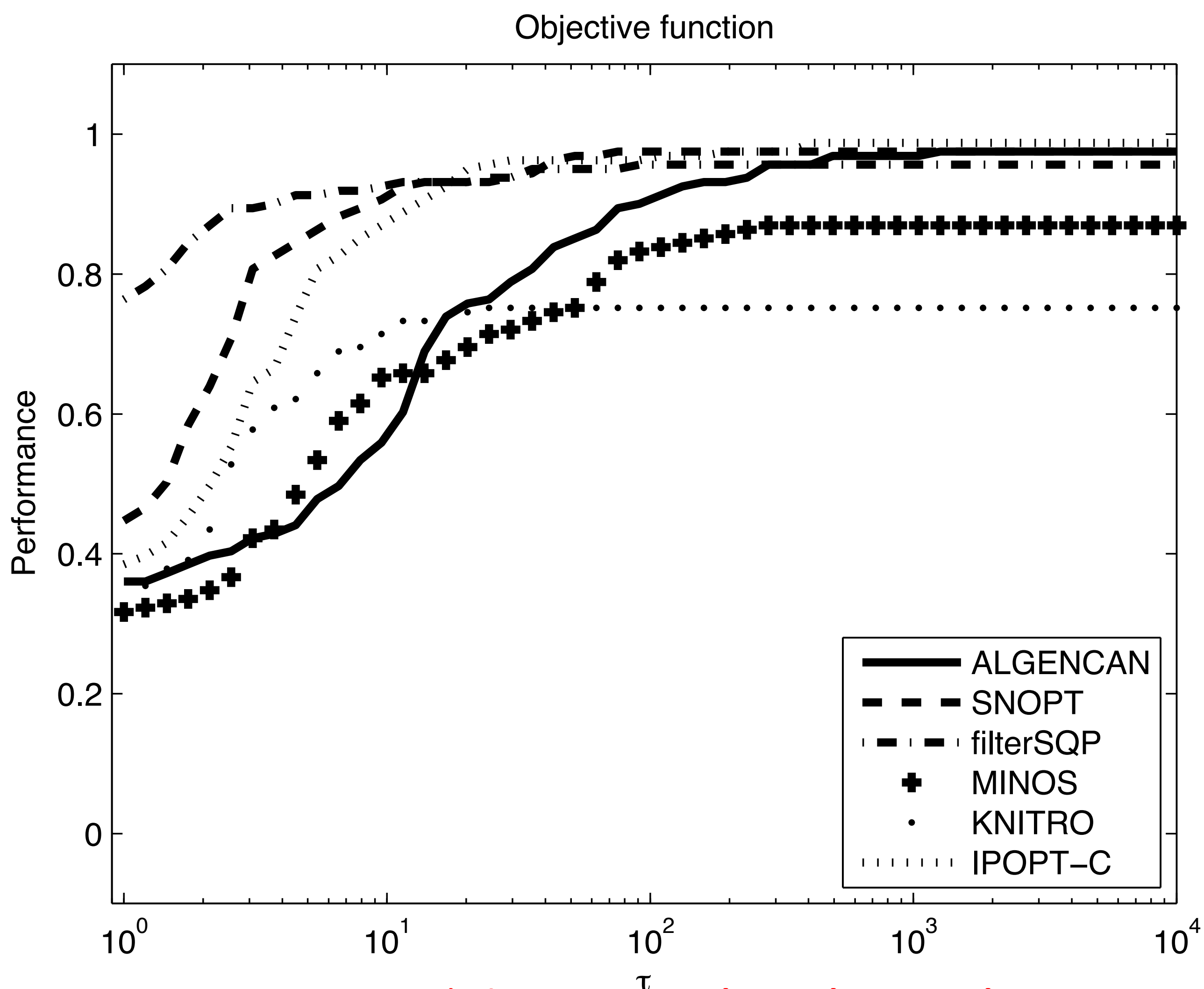
- ALGENCAN (augmented Lagrangian)
- SNOPT (elastic mode linesearch SQP)
- filterSQP (trust-region filter SQP)
- MINOS (linearized augmented Lagrangian)
- KNITRO (interior points, with MPCC option)
- IPOPT-C (interior points, with MPCC option)



Augmented Lagrangian is robust;
computes good solutions!



Confirms augmented Lagrangian fast rate of convergence results!



However, augmented Lagrangian iterations are expensive...

Conclusions

The augmented Lagrangian algorithm
(method of multipliers)

- Requires only SOSC
  (equality constraints – noncriticality) for
  - Primal-dual Q-linear rate of convergence
  - Primal Q-linear rate of convergence
  - The rates are inversely proportional to c_k
- Not needed:
  - LICQ, Strict complementarity
- Reliably computes good MPCC solutions

Details:

SIAM Journal on Optimization, Vol. 22, 384–407
(Rates of convergence theory , with D. Fernández)

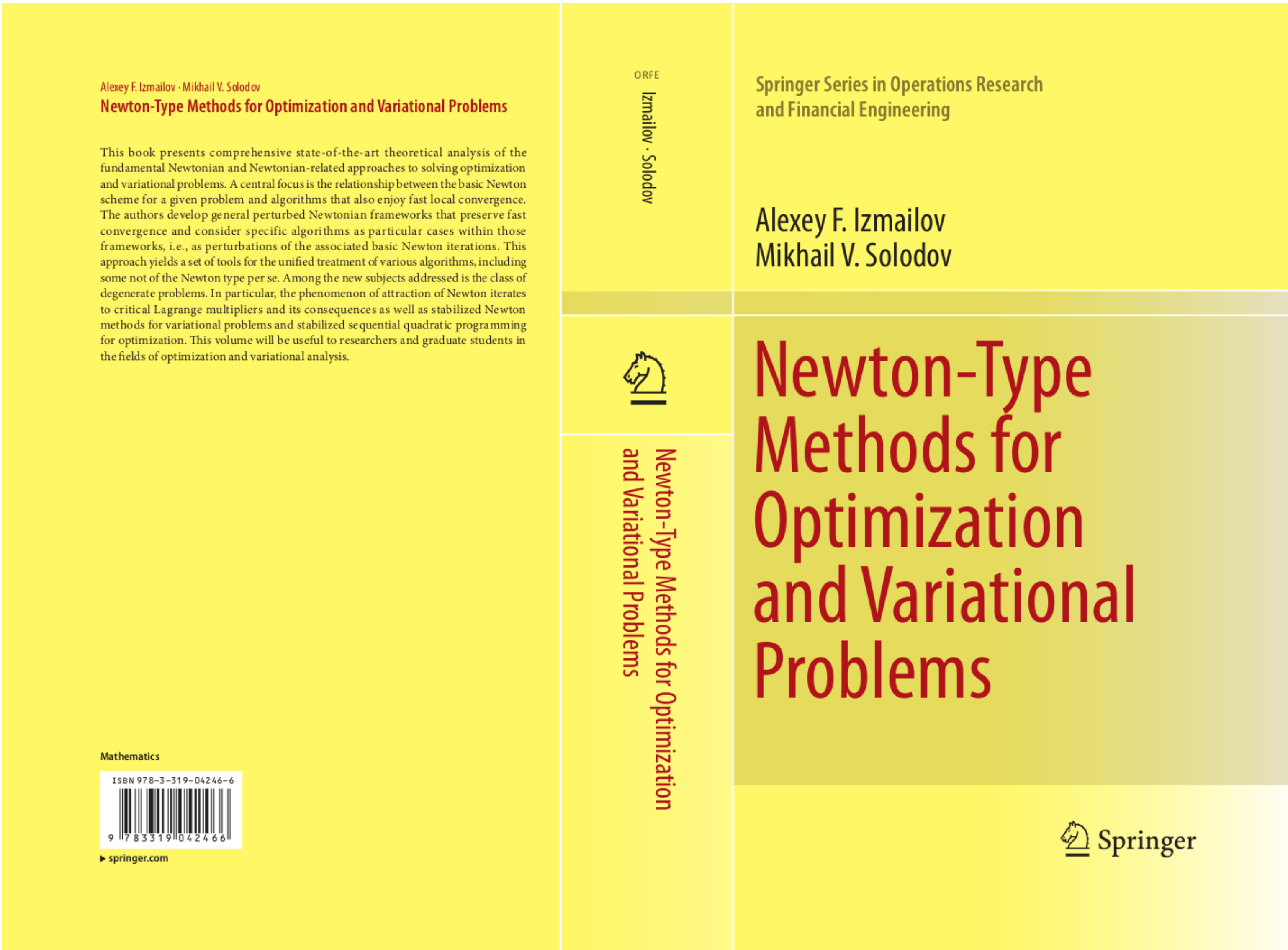
SIAM Journal on Optimization, Vol. 22, 1579–1606
(MPCC and numerics, with Uskov & Izmailov)

Computational Optimization and Applications, 2014
(Equality constraints, with Kurennoy & Izmailov)

or

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

(600 pages approx.)



Alexey F. Izmailov · Mikhail V. Solodov

Newton-Type Methods for Optimization and Variational Problems

This book presents comprehensive state-of-the-art theoretical analysis of the fundamental Newtonian and Newtonian-related approaches to solving optimization and variational problems. A central focus is the relationship between the basic Newton scheme for a given problem and algorithms that also enjoy fast local convergence. The authors develop general perturbed Newtonian frameworks that preserve fast convergence and consider specific algorithms as particular cases within those frameworks, i.e., as perturbations of the associated basic Newton iterations. This approach yields a set of tools for the unified treatment of various algorithms, including some not of the Newton type per se. Among the new subjects addressed is the class of degenerate problems. In particular, the phenomenon of attraction of Newton iterates to critical Lagrange multipliers and its consequences as well as stabilized Newton methods for variational problems and stabilized sequential quadratic programming for optimization. This volume will be useful to researchers and graduate students in the fields of optimization and variational analysis.

Mathematics

ISBN 978-3-319-04246-6

▸ springer.com

ORFE

Izmailov · Solodov

Newton-Type Methods for Optimization and Variational Problems

Springer Series in Operations Research and Financial Engineering

Alexey F. Izmailov

Mikhail V. Solodov

Newton-Type Methods for Optimization and Variational Problems

Springer

Thanks!