PRACTICAL AUGMENTED LAGRANGIAN METHODS FOR NONCONVEX PROBLEMS

José Mario Martínez

www.ime.unicamp.br/∼martinez

UNICAMP, Brazil

August 2, 2011

Collaborators

Ernesto Birgin (Computer Science - USP)
Laura Schuverdt (Applied Math - UNICAMP)
Roberto Andreani (Applied Math - UNICAMP)
Lucas Garcia Pedroso (Applied Math - UNICAMP)
Maria Aparecida Diniz (Applied Math - UNICAMP)
Marcia Gomes-Ruggiero (Applied Math - UNICAMP)
Sandra Santos (Applied Math - UNICAMP)

Outline

The Nonlinear Programming Problem

$$\text{Minimize } f(x)$$

subject to

$$h(x) = 0, g(x) \leq 0,$$

$$x \in \Omega,$$

where $x \in \mathbb{R}^n$, $h(x) \in \mathbb{R}^m, g(x) \in \mathbb{R}^p$.

PHR Augmented Lagrangian

Definition

$$L_\rho(x,\lambda,\mu) = f(x) + \frac{\rho}{2}\left[\left\|h(x)+\frac{\lambda}{\rho}\right\|^2 + \left\|\left(g(x)+\frac{\mu}{\rho}\right)_+\right\|^2\right]$$

$$(a_+ = \max\{0,a\}, \lambda \in \mathbb{R}^m, \mu \in \mathbb{R}^p_+)$$

Conceptual Algorithm based on PHR

- Outer Iteration

  "Minimize" $L_\rho(x,\lambda,\mu)$ subject to $x \in \Omega$

- Update Multipliers λ, μ and Penalty Parameter ρ

Penalty and Shifting

Penalty Strategy (ρ)

The punishment must be proportional to the constraint violation

$$L_\rho(x, \lambda, \mu) = f(x) + \frac{\rho}{2}\left[\left\|h(x) + \frac{\lambda}{\rho}\right\|^2 + \left\|\left(g(x) + \frac{\mu}{\rho}\right)_+\right\|^2\right]$$

Shift Strategy (λ/ρ and μ/ρ)

"Better" than increasing the penalty parameter, is to "pretend" that the tolerance to constraint violation is "stricter" than it is. Punish with respect to suitably shifted constraint violations.

$$L_\rho(x, \lambda, \mu) = f(x) + \frac{\rho}{2}\left[\left\|h(x) + \frac{\lambda}{\rho}\right\|^2 + \left\|\left(g(x) + \frac{\mu}{\rho}\right)_+\right\|^2\right]$$

Non-PHR Augmented Lagrangians

R. R. Allran and S. E. J. Johnsen, 1970.

A. Auslender, M. Teboulle and S. Ben-Tiba 1999.

Ben Tal, I. Yuzefovich and M. Zibulevsky, 1992.

A. Ben-Tal and M. Zibulevsky, 1997.

D. P. Bertsekas, 1982.

R. A. Castillo, 1998.

C. C. Gonzaga and R. A. Castillo, 2003.

C. Humes and P. S. Silva, 2000.

A. N. Iusem, 1999.

B. W. Kort and D. P. Bertsekas, 1973.

B. W. Kort and D. P. Bertsekas, 1976.

L. C. Matioli, 2001.

F. H. Murphy, 1974.

H. Nakayama , H. Samaya and Y. Sawaragi, 1975.

R. A. Polyak, 2001.

P. Tseng and D. Bertsekas, 1993.

Exact Penalty Connections

G. Di Pillo and L. Grippo, 1979, 1987, 1988.
A. De Luca and G. Di Pillo, 1987.

Stability Issues

J-P Dussault, 1995, 1998, 2003.
C. G. Broyden and N. F. Attia, 1983, 1988.
N. I. M. Gould, 1986.
Z. Dostál, A. Friedlander and S. A. Santos, 1999, 2002.

Hybrid Augmented Lagrangian Algorithms

L. Ferreira-Mendonça, V. L. R. Lopes and J. M. Martínez, 2006.
E. G. Birgin and J. M. Martínez, 2006.
M. Friedlander and S. Leyffer, 2007.

The PHR approach

M. Hestenes, 1969.
M. J. D. Powell, 1969.
R. T. Rockafellar, 1974.
A. R. Conn, N. I. M. Gould, Ph. L. Toint, 1991 (LANCELOT).
A. R. Conn, N. I. M. Gould, A. Sartenaer, Ph. L. Toint, 1996.
Contributions of our team.

Reasons for not abandoning the Augmented Lagrangian approach in practical Nonlinear Programming

Exploit structure of simple subproblems

The lower-level set may be arbitrary.
Augmented Lagrangian methods proceed by sequential resolution of simple problems. Progress in the analysis and implementation of simple-problem optimization procedures produces an almost immediate positive effect on the effectiveness of Augmented Lagrangian algorithms. Box-constrained optimization is a dynamic area of practical optimization from which we can expect Augmented Lagrangian improvements.

Reasons for not abandoning the Augmented Lagrangian approach in practical Nonlinear Programming

Global Minimization

Global minimization of the subproblems implies convergence to global minimizers of the Augmented Lagrangian method. There is a large field for research on global optimization methods for box-constraint optimization. When the global box-constraint optimization problem is satisfactorily solved in practice, the effect on the associated Augmented Lagrangian method for Nonlinear Programming problem is immediate.

Reasons for not abandoning the Augmented Lagrangian approach in practical Nonlinear Programming

Global minimization in practice

Most box-constrained optimization methods are guaranteed to find stationary points. In practice, good methods do more than that. Extrapolation and magical steps steps enhance the probability of convergence to global minimizers. As a consequence, the probability of convergence to Nonlinear Programming global minimizers of a practical Augmented Lagrangian method is enhanced too.

Reasons for not abandoning the Augmented Lagrangian approach in practical Nonlinear Programming

Non-smoothness and global minimization

The Convergence-to-global-minimizers theory of Augmented Lagrangian methods does not need differentiability of the functions that define the Nonlinear Programming problem. In practice, the Augmented Lagrangian approach may be successful in situations were smoothness is "dubious".

Reasons for not abandoning the Augmented Lagrangian approach in practical Nonlinear Programming

Derivative-free

The Augmented Lagrangian approach can be adapted to the situation in which analytic derivatives are not computed. Derivative-free Augmented Lagrangian methods preserve theoretical convergence properties.

Reasons for not abandoning the Augmented Lagrangian approach in practical Nonlinear Programming

Hessian-Lagrangian structurally dense

In many practical problems the Hessian of the Lagrangian is structurally dense (in the sense that any entry may be different from zero at different points) but generally sparse (given a specific point in the domain, the particular Lagrangian Hessian is a sparse matrix). The sparsity pattern of the matrix changes from iteration to iteration. This difficulty is almost irrelevant for the Augmented Lagrangian approach if one uses a low-memory box-constraint solver.

Reasons for not abandoning the Augmented Lagrangian approach in practical Nonlinear Programming

Hessian-Lagrangian poorly structured

Independently of the Lagrangian Hessian density, the structure of the KKT system may be very poor for sparse factorizations. This is a serious difficulty for Newton-based methods but not for suitable implementations of the Augmented Lagrangian PHR algorithm.

Reasons for not abandoning the Augmented Lagrangian approach in practical Nonlinear Programming

Many inequality constraints

Nonlinear Programming problem has many inequality constraints: many additional variables if one uses slack variables. There are several approaches to reduce the effect of the presence of many slacks, but they may not be as effective as not using slacks at all. The price of not using slacks is the absence of continuous second derivatives in L_ρ. In many cases, this does not seem to be a serious practical inconvenience

AL Algorithm with arbitrary lower-level constraints

Initialization

$k \leftarrow 1$, $\|V^0\| = \infty$, $\gamma > 1 > \tau$, $\lambda^1 \in \mathbb{R}^m$, $\mu^1 \in \mathbb{R}^p_+$.

Step 1: Solving the Subproblem

Compute $x^k \in \mathbb{R}^n$ an approximate solution of

$$\text{Minimize } L_{\rho_k}(x, \lambda^k, \mu^k) \text{ subject to } x \in \Omega.$$

Step 2: Update penalty parameter and multipliers

Define $V_i^k = \max\left\{g_i(x^k), -\frac{\mu_i^k}{\rho_k}\right\}$.

If $\max\{\|h(x^k)\|_\infty, \|V^k\|_\infty\} \leq \tau \ \max\{\|h(x^{k-1})\|_\infty, \|V^{k-1}\|_\infty\}$,
define $\rho_{k+1} = \rho_k$. Else, $\rho_{k+1} = \gamma \rho_k$.
Compute $\lambda^{k+1} \in [\lambda_{\min}, \lambda_{\max}]^m$, $\mu^{k+1} \in [0, \mu_{\max}]^p$.
Set $k \leftarrow k+1$ and go to Step 1.

Convergence to Global Minimizers

Theorem

Assume that the problem is feasible and that each subproblem is considered as approximately solved when $x^k \in \Omega$ is found such that

$$L_{\rho_k}(x^k, \lambda^k, \mu^k) \leq L_{\rho_k}(y, \lambda^k, \mu^k) + \varepsilon_k$$

for all $y \in \Omega$, where $\{\varepsilon_k\}$ is a sequence of nonnegative numbers that converge to $\varepsilon \geq 0$.
Then, every limit point x^* of $\{x^k\}$ is feasible and

$$f(x^*) \leq f(y) + \varepsilon$$

for all feasible point y.

Approximate local solution of the subproblem

General form of Lower-Level Constraints

$$\Omega = \{x \in \mathbb{R}^n \mid \underline{h}(x) = 0, \underline{g}(x) \leq 0\},$$

Approximate local solution of the subproblem

Lower-Level ε_k- KKT Conditions

$$\|\nabla L_{\rho_k}(x^k, \lambda^k, \mu^k) + \sum_{i=1}^{\underline{m}} v_i^k \nabla \underline{h}_i(x^k) + \sum_{i=1}^{\underline{p}} u_i^k \nabla \underline{g}_i(x^k)\| \leq \varepsilon_k,$$

$$u_i^k \geq 0, \underline{g}_i(x^k) \leq \varepsilon_k \text{ for all } i,$$

$$\underline{g}_i(x^k) < -\varepsilon_k \Rightarrow u_i^k = 0 \text{ for all } i,$$

$$\|\underline{h}(x^k)\| \leq \varepsilon_k.$$

First-order Choice of Multipliers Estimates

For Equality Upper Level Constraints

$$\lambda_i^{k+1} = \max\{\lambda_{\min}, \min\{\lambda_i^k + \rho_k h_i(x^k), \lambda_{\max}\}\}$$

For Inequality Upper Level Constraints

$$\mu_i^{k+1} = \max\{0, \min\{\mu_i^k + \rho_k g_i(x^k), \mu_{\max}\}\}.$$

Positive linear dependence

Positive linear dependent gradients of active constraints

Assume that the feasible set of a nonlinear programming problem is given by $\overline{h}(x) = 0, \overline{g}(x) \leq 0$. Let $I(x)$ be the set of indices of the active inequality constraints at the feasible point x. Let $I_1 \subset \{1, \ldots, \overline{m}\}$, $I_2 \subset I(x)$. The subset of gradients of active constraints that correspond to the indices $I_1 \cup I_2$ is said to be positively linearly dependent if there exist multipliers λ, μ such that

$$\sum_{i \in I_1} \lambda_i \nabla \overline{h}_i(x) + \sum_{i \in I_2} \mu_i \nabla \overline{g}_i(x) = 0,$$

with $\mu_i \geq 0$ for all $i \in I_2$ and $\sum_{i \in I_1} |\lambda_i| + \sum_{i \in I_2} \mu_i > 0$.

Otherwise, we say that these gradients are positively linearly independent.

Constraint Qualifications

Regularity (LICQ)

The gradients of the active constraints are linearly independent.

STRONGER (MORE RESTRICTIVE) THAN:

Mangasarian-Fromovitz

The gradients of the active constraints are positively linearly independent.

STRONGER (MORE RESTRICTIVE) THAN:

CPLD Constraint Qualification

Constant Positive Linear Dependence (CPLD)

If a subset of gradients of active constraints is positive linear dependent, the same subset of gradients remains linear dependent in a neighborhood of the point.

(Qi & Wei, Andreani, J.M.M. & Schuverdt)

Convergence to feasible points

Theorem

Let x^* be a limit point of $\{x^k\}$. Then, if the sequence of penalty parameters $\{\rho_k\}$ is bounded, the limit point x^* is feasible. Otherwise, at least one of the following possibilities hold:

(i) x^* is a KKT point of the problem

$$\text{Minimize } \frac{1}{2}\left[\sum_{i=1}^{m} h_i(x)^2 + \sum_{i=1}^{p} [g_i(x)_+]^2\right] \text{ subject to } x \in \Omega.$$

(ii) x^* does not satisfy the CPLD constraint qualification associated with Ω.

Convergence to KKT points

Theorem

Assume that x^* is a feasible limit point of $\{x^k\}$ that satisfies the CPLD constraint qualification related to set of all the constraints. Then, x^* is a KKT point of the problem.

Boundedness of Penalty Parameter

Conditions under which ρ_k is bounded

- $\lim_{k\to\infty} x^k = x^*$ and x^* is feasible.
- LICQ holds at x^*. ($\Rightarrow KKT$).
- The Hessian of the Lagrangian is positive definite in the orthogonal subspace to the gradients of active constraints.
- $\lambda_i^* \in (\lambda_{\min}, \lambda_{\max})$, $\mu_j^* \in [0, \mu_{\max})$ for all i, j.
- For all i such that $g_i(x^*) = 0$, we have $\mu_i^* > 0$. (Strict complementarity in the upper level.)
- There exists a sequence $\eta_k \to 0$ such that $\varepsilon_k \leq \eta_k \max\{\|h(x^k)\|, \|V^k\|\}$ for all $k = 0, 1, 2 \ldots$.

Second Order Optimality Condition

Weak Second Order Condition (SOC)

The Hessian of the Lagrangian is positive semi-definite on the orthogonal subspace to the gradients of active constraints.

Regularity and SOC

Text books: At a local minimizer:

$$\text{LICQ} \Rightarrow \text{SOC}$$

Second Order Optimality Condition

May LICQ be weakened?

to Mangasarian-Fromovitz?

Answer: Mangasarian-Fromovitz is not enough. Counterexample Polyak, Anitescu.

Second Order Optimality Condition

Weak Constant Rank Condition WCR

We say that WCR is satisfied at the feasible point x^* if the rank of the matrix formed by the gradients of the active constraints at x^* remains constant (does not increase) in a neighborhood of x_*.

Theorem

At a local minimizer

Mangasarian-Fromovitz $+$ WCR $\Rightarrow$ SOC

Consequences for the Augmented Lagrangian Method

Assume that we implement the Augmented Lagrangian method (with $\Omega = \mathbb{R}^n$) in such a way that, at the solutions of the subproblems, we have:

Stopping Criterion at the Subproblems

$$v^T \nabla^2 L_{\rho_k}(x^k, \lambda^k, \mu^k) v \geq -\varepsilon_k \|v\|^2$$

for all $v \in \mathbb{R}^n$.

where

$$\nabla^2 \left[\max\left(0, g_i(x) + \frac{\mu_i}{\rho}\right) \right]^2 = \nabla^2 \left(g_i(x) + \frac{\mu_i}{\rho} \right)^2 \text{ if } g_i(x) + \frac{\mu_i}{\rho} = 0,$$

Augmented Lagrangian and SOC

Theorem

If the Augmented Lagrangian Method with the approximate second order stopping criterion on the subproblems converges to a feasible point x^* that satisfies Mangasarian-Fromovitz and Weak-Constant-Rank, then x^* satisfies the second order condition SOC.

Derivative-free Augmented Lagrangian

(Ω = a box)

Stopping Criterion for the Subproblems

$$L_{\rho_k}(x^k, \lambda^k, \mu^k) \leq L_{\rho_k}(x^k \pm \varepsilon_k e_j, \lambda^k, \mu^k)$$

for all $j = 1, \ldots, n$, whenever $x^k \pm \varepsilon_k e_j \in \Omega$.

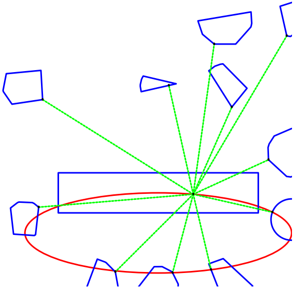
Derivative-free Augmented Lagrangian

Results

- Every limit point is a Stationary point of the quadratic infeasibility measure
- Every feasible limit point that satisfies CPLD is stationary
- Under "additional assumptions", boundedness of penalty parameters.

Example of LA with very structured lower-level constraints

Find the point in the Rectangle but not in the Ellipse such that the sum of the distances to the polygons is minimal.
Upper-level constraints: (All points) $\notin$ Ellipse
Lower-level constraints: Central Point $\in$ Rectangle, Polygon Points $\in$ Polygons.



Example of LA with very structured lower-level constraints

1,567,804 polygons
3,135,608 variables, 1,567,804 upper-level constraints, 12,833,106 lower-level constraints
Convergence in
10 outer iterations, 56 inner iterations, 133 function evaluations, 185 seconds

Reasons for this behavior

We use, in this case, the Spectral Projected Gradient method SPG for convex constrained minimization for solving the subproblems, which turns out to be very efficient because computing projections, in this case, is easy.

ALGENCAN

ALGENCAN is the Augmented Lagrangian algorithm with lower level constraints $x \in \Omega$, where Ω is a box.

Solver for the subproblems: GENCAN

The box-constraint solver GENCAN uses:

- Active set strategy
- Inexact-Newton within the faces
- Spectral Projected Gradient (SPG) to leave faces
- Extrapolation and Magical steps.

"Modest Claim" about ALGENCAN

ALGENCAN is efficient when:

- Many inequality constraints
- Hard KKT-Jacobian structure

Example: Hard-Spheres problem

Find n_p points on the unitary sphere of $\mathbb{R}^{n_d}$ maximizing the minimal pairwise distances.

NLP Formulation

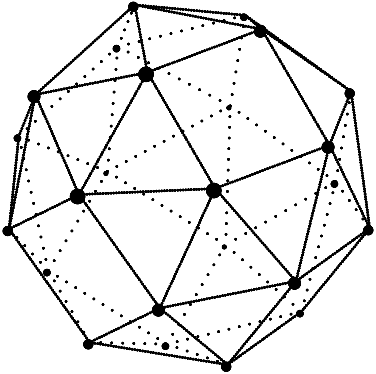
$$\begin{array}{ll} \text{Minimize}_{\,p^i,z} & z \\ \text{subject to} & \|p^i\|^2 = 1,\ i = 1,\ldots,n_p, \\ & \langle p^i, p^j\rangle \le z,\ i = 1,\ldots,n_p - 1,\ j = i+1,\ldots,n_p, \end{array}$$

where $p^i \in \mathbb{R}^{n_d}$ for all $i = 1,\ldots,n_p$. This problem has $n_d \times n_p + 1$ variables, n_p equality constraints and

$$n_p \times (n_p - 1)/2$$

inequality constraints.

Hard-Spheres problem, $n_d = 3$, $n_p = 24$



Behavior of ALGENCAN in Hard-Spheres

Hard-Spheres (3,162)

	Final infeasibility	Final f	Iterations	Time
ALGENCAN	3.7424E-11	9.5889E-01	10	40.15
IPOPT	5.7954E-10	9.5912E-01	944	1701.63

Enclosing-Ellipsoid problem

Find the Ellipsoid with smallest volume that contains n_p given points in $\mathbb{R}^{n_d}$.

$$\begin{array}{ll} \text{Minimize}_{\ l_{ij}} & -\sum_{i=1}^{n_d} \log(l_{ii}) \\ \text{subject to} & (p^i)^T L L^T p^i \leq 1, i = 1, \ldots, n_p, \\ & l_{ii} \geq 10^{-16}, i = 1, \ldots, n_d, \end{array}$$

where $L \in \mathbb{R}^{n_d \times n_d}$ is a lower-triangular matrix. The number of variables is $n_d \times (n_d + 1)/2$ and the number of inequality constraints is n_p (plus the bound constraints).

Enclosing Ellipsoid

Bratu

Discretized three-dimensional Bratu-based problem:

$$\begin{array}{ll} \text{Minimize}_{\,u(i,j,k)} & \sum_{(i,j,k)\in S}[u(i,j,k) - u_*(i,j,k)]^2 \\ \text{subject to} & \phi_\theta(u,i,j,k) = \phi_\theta(u_*,i,j,k), \;\; i,j,k = 2,\ldots,n_p-1, \end{array}$$

where

$$\phi_\theta(v,i,j,k) = -\Delta v(i,j,k) + \theta e^{v(i,j,k)},$$

and

$$\Delta v(i,j,k) = \frac{v(i\pm 1,j,k) + v(i,j\pm 1,k) + v(i,j,k\pm 1) - 6v(i,j,k)}{h^2},$$

The number of variables is n_p^3 and the number of equality constraints is $(n_p-2)^3$. We set $\theta = -100$, $h = 1/(n_p-1)$ and $|S| = 7$. This problem has no inequality constraints.

Characteristics of Hard-Spheres, Enclosing-Ellipsoid and Bratu

Hard-Spheres and Enclosing-Ellipsoid have many inequality constraints.
Bratu-based problem has a difficult KKT structure.

Enclosing Ellipsoid test

6 variables, 20000 inequality constraints.

Enclosing-Ellipsoid (3,20000)

	Final infeasibility	Final f	Iterations	Time
ALGENCAN	8.3449E-09	3.0495E+01	28	1.90
IPOPT	1.1102E-15	3.0495E+01	41	9.45

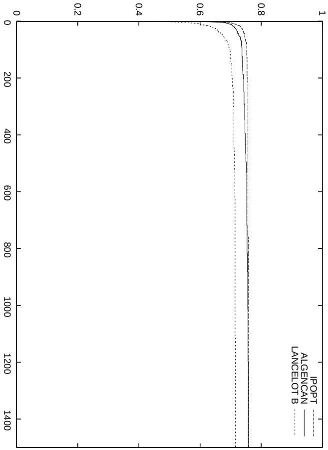
Bratu-based test

$n_p = 20$, $n = 8000$, number of constraints: 5832.

Bratu-based $(20, \theta = -100, \#S = 7)$

	Final infeasibility	Final f	Iterations	Time
ALGENCAN	6.5411E-09	2.2907E-17	3	5.12
IPOPT	2.7311E-08	8.2058E-14	5	217.22

Low-Precision (10^{-4}) Performance-Profiles



Dealing with slow convergence

In spite of the “modest claims” one wishes that ALGENCAN should behave reasonably in “all” the problems.
However: ultimate convergence of the AL method may be slow.
ALGENCAN may converge slowly in problems where “Newton’s method” applied to the KKT system is very effective.

Remedy: Newton-acceleration of ALGENCAN

ALGENCAN + NEWTON

- Run ALGENCAN up to some modest precision.
- Run a moderate number of Newton-KKT iterations.
- Repeat.

Convergence

Global as ALGENCAN , Fast as NEWTON .

Non-standard Problems

Bilevel

$$\text{Minimize } f(x, y)$$

subject to

$$y \text{ solves } P(x),$$

where $P(x)$ is a constrained nonlinear programming problem.

Augmented Lagrangian Strategy:

$$\text{Minimize } f(x, y)$$

subject to

$$y \text{ solves } P(x, \rho_k, \lambda^k, \mu^k),$$

where $P(x, \rho, \lambda, \mu)$ is an unconstrained nonlinear programming problem.

Non-standard Problems

Problem

Minimize $f(x)$ subject to "At least q constraints are satisfied"

Augmented Lagrangian Strategy:

Outer iteration:

$$\text{Minimize } f(x) + \frac{\rho}{2}\left[\sum q \text{ smaller } \left(g_i(x) + \frac{\mu_i}{\rho}\right)_+^2\right]$$

Conclusions

- There are many reason for not abandoning classical PHR Augmented Lagrangian methods.
- The Augmented Lagrangian Method with arbitrary lower-level constraints admits a nice global optimization theory, without "assumptions on the algorithm" and weak constraint qualifications (even second-order)
- Taking advantage of good algorithms for lower-level constraints may be very effective.
- ALGENCAN (Ω= a box) is effective with many inequality constraints and bad KKT-Jacobian structure.
- ALGENCAN can be accelerated with Newton-KKT.
- Non-standard problems
- See the Tango site www.ime.usp.br/∼egbirgin/tango.

Main references of this talk

- R. Andreani, E. G. Birgin, J. M. Martínez and M. L. Schuverdt. On Augmented Lagrangian methods with general lower-level constraints. To appear in *SIAM Journal on Optimization* .
- E. G. Birgin and J. M. Martínez. Improving ultimate convergence of an Augmented Lagrangian method. To appear in *Optimization Methods and Software*.
- R. Andreani, J. M. Martínez and M. L. Schuverdt. On second-order optimality conditions for nonlinear programming. To appear in *Optimization*.
- L. Ferreira-Mendonça, V. L. Lopes and J. M. Martínez. Quasi-Newton acceleration for equality constrained minimization. To appear in *Computational Optimization and Applications*.

Main References of this talk

- E. G. Birgin and J. M. Martínez. Structured Minimal-Memory Inexact Quasi-Newton method and secant preconditioners. To appear in *Computational Optimization and Applications*.
- R. Andreani, E. G. Birgin, J. M. Martínez and M. L. Schuverdt. Augmented Lagrangian methods under the Constant Positive Linear Dependence constraint qualification. *Mathematical Programming* 111, pp. 5-32 (2008).
- R. Andreani, J. M. Martínez and M. L. Schuverdt, On the relation between the Constant Positive Linear Dependence condition and quasinormality constraint qualification, *Journal of Optimization Theory and Applications* 125, pp. 473–485, 2005.
- E. G. Birgin, C. Floudas and J. M. Martínez. Global Optimization using an Augmented Lagrangian method with variable lower-level constraints.