Linearized augmented Lagrangian methods for nonconvex minimization

Ion Necoara

Universitatea Națională de Știință și Tehnologie POLITEHNICA BUCUREȘTI

ISMMA

Univ. Politehnica Bucharest & Institute Appl. Math. Romanian Academy

Workshop: Optimization for Learning and Control, Lucca, Italy, June 2025

Content

- Problem formulation and assumptions
- Linearized Augmented Lagrangian algorithm
- Convergence analysis
- Extension: Linearized ADMM alg. for structured optimization
- Numerical simulations

L. El Bourkhissi, I. Necoara, *Complexity of a linearized augmented Lagrangian method for nonconvex minimization with nonlinear equality constraints*, arxiv, 2025.

L. El Bourkhissi, I. Necoara, *Convergence rates for an inexact linearized ADMM for nonsmooth nonconvex optimization with nonlinear equality constraints*, COAP, 2024.

Problem formulation

Consider nonlinear optimization problem:

$$\min_{x \in \mathbb{R}^n} \quad f(x)$$
$$\text{s.t.} \quad F(x) = 0,$$

- $f : \mathbb{R}^n \to \mathbb{R}$ and $F(x) \triangleq (f_1(x), ..., f_m(x))^T$, with $f_i : \mathbb{R}^n \to \mathbb{R}$ for all $i = 1 : m$.
- $f, f_i \in \mathcal{C}^1$ for all $i = 1 : m$ and F is nonlinear.
- $\nabla f(x) \in \mathbb{R}^n$ denotes gradient; $J_F(x) \in \mathbb{R}^{m \times n}$ denotes Jacobian.

Definition: x_ϵ^* is an ϵ-first-order solution if $\exists \lambda_\epsilon^* \in \mathbb{R}^m$ such that:

$$\|\nabla f(x_\epsilon^*) + J_F(x_\epsilon^*)^T \lambda_\epsilon^*\| \leq \epsilon \quad \text{and} \quad \|F(x_\epsilon^*)\| \leq \epsilon.$$

Motivation:

1. Phase retrieval
2. Control
3. Training DNNs
4. Inverse problems
5. Nonlinear programming, etc

Problem formulation: main assumptions

Consider nonlinear optimization problem:

$$\min_{x \in \mathbb{R}^n} \quad f(x)$$
$$\text{s.t.} \quad F(x) = 0,$$

Assumptions:

For any compact set $\mathcal{S} \subseteq \mathbb{R}^n$, there exist positive constants $M_f, M_F, \sigma, L_f, L_F$ such that f and F satisfy the following conditions:

1. $\|\nabla f(x)\| \leq M_f, \ \ \|\nabla f(x) - \nabla f(x')\| \leq L_f \|x - x'\|$ for all $x, x' \in \mathcal{S}$.
2. $\|J_F(x)\| \leq M_F, \ \ \sigma_{\min}(J_F(x)) \geq \sigma > 0$ for all $x \in \mathcal{S}$.
3. $\|J_F(x) - J_F(x')\| \leq L_F \|x - x'\|$ for all $x, x' \in \mathcal{S}$

Remark:

Our assumptions allow general classes of problems.

- first conditions hold if e.g., f is differentiable and $\nabla f(\cdot)$ is *locally* Lipschitz continuous on a neighborhood of $\mathcal{S}$.
- second conditions hold when e.g., F is differentiable on a neighborhood of $\mathcal{S}$ and satisfies an LICQ condition over $\mathcal{S}$ (hence, $m \leq n$).
- third condition holds if e.g., J_F is *locally* Lipschitz continuous on $\mathcal{S}$.

Obs.: Note that any twice continuously differentiable function is locally Lipschitz and locally smooth on a compact set.

Notations

Consider nonlinear optimization problem:

$$\min_{x \in \mathbb{R}^n} \quad f(x)$$
$$\text{s.t.} \quad F(x) = 0,$$

- Augmented Lagrangian function associated to our problem:
$$\mathcal{L}_\rho(x, \lambda) = f(x) + \langle \lambda, F(x) \rangle + \frac{\rho}{2} \|F(x)\|^2$$
- We use the notations:
$$l_f(x; \bar{x}) := f(\bar{x}) + \langle \nabla f(\bar{x}), x - \bar{x} \rangle, \quad l_F(x; \bar{x}) := F(\bar{x}) + J_F(\bar{x})(x - \bar{x}) \;\; \forall x, \bar{x}$$
- Denote **quadratic** function derived from linearization of objective and functional constraints in a Gauss-Newton fashion, at a given point $\bar{x}$:
$$\bar{\mathcal{L}}_\rho(x, \lambda; \bar{x}) = l_f(x; \bar{x}) + \langle \lambda, l_F(x; \bar{x}) \rangle + \frac{\rho}{2} \|l_F(x; \bar{x})\|^2$$
(in contrast to pure linearization of $\mathcal{L}_\rho(\cdot, \lambda)$)!
- Introduce Lyapunov function:
$$P(x, \lambda, \bar{x}, \gamma) = \mathcal{L}_\rho(x, \lambda) + \frac{\gamma}{2} \|x - \bar{x}\|^2$$
- Evaluation of Lyapunov function along iterates is denoted by:
$$P_k = P\left(x_k, \lambda_k, x_{k-1}, \frac{\beta_k}{2}\right) \quad \forall k \geq 0,$$
- We also denote: $\Delta x_k = x_k - x_{k-1}$ and $\Delta \lambda_k = \lambda_k - \lambda_{k-1}$ $\forall k \geq 0$

History

Consider nonlinear optimization problem:

$$\min_{x\in\mathbb{R}^n} \quad f(x)$$
$$\text{s.t.} \quad F(x) = 0,$$

Literature on augmented Lagrangian methods is vast:

- Augmented Lagrangian methods (Rockafellar'76, Bertsekas'82, Birgin&Martinez group ('08,...,'20), etc...):

AL alg.
Given x_0, $\lambda_0, \Lambda = [\Lambda_{\min}, \Lambda_{\max}]$ and $\rho_0 > 0, \gamma > 1, \tau > 0$. For $k \geq 0$ do:

i. find approximate solution: $x_{k+1} \approx \arg\min_x \mathcal{L}_{\rho_k}(x, \lambda_k)$

ii. update
$\lambda_{k+1} \leftarrow \mathrm{Pr}_\Lambda(\lambda_k + \rho_k F(x_{k+1}))$, $\rho_{k+1} \leftarrow \gamma\rho_k$ if $\|F(x_{k+1})\|_\infty > \tau\|F(x_k)\|_\infty$

- Proximal augmented Lagrangian methods (Rockafellar'76, Hajinezhad'19, Sahin'19, Wright'20, etc...):

Proximal AL alg.
Given x_0, λ_0 and $\rho, \beta > 0$. For $k \geq 0$ do:

ii. find approximate solution: $x_{k+1} \approx \arg\min_x \mathcal{L}_\rho(x, \lambda_k) + \frac{\beta}{2}\|x - x_k\|^2$

iii. update: $\lambda_{k+1} \leftarrow \lambda_k + \rho F(x_{k+1})$

- Why AL type alg's are attractive? shown that when subproblem is solved to approximate global optimality, limit points are global solution of original problem $\Rightarrow$ (higly nonconvex) subproblem solved with Newton CG, etc...

Our goal

Consider nonlinear optimization problem:

$$\min_{x \in \mathbb{R}^n} \quad f(x)$$
$$\text{s.t.} \quad F(x) = 0,$$

- (Proximal) Augmented Lagrangian methods enjoy nice convergence properties and have good practical behavior
- However they require solving highly nonconvex subproblems at each iteration (of the form):

$$x_{k+1} \approx \arg\min_x \mathcal{L}_\rho(x, \lambda_k) + \frac{\beta}{2}\|x - x_k\|^2$$

- one needs to call complicated subroutines (such as Newton CG, gradient) to solve nonconvex subproblem

$$\Downarrow$$

Goal: derive (proximal) linearized AL methods (subproblem easy - e.g. convex)

Linearized augmented Lagrangian algorithm (L-AL)

We consider the following algorithm:

Algorithm L-AL

Given $x_{-1} = x_0,\ \lambda_0$ and $\rho \geq 1,\ \beta_0 \geq \underline{\beta} > 0$.

For $k \geq 0$ do:

find $\beta_{k+1} \geq \underline{\beta}$ such that the points:

$$x_{k+1} \leftarrow \arg\min_{x} \bar{\mathcal{L}}_{\rho}(x, \lambda_k; x_k) + \frac{\beta_{k+1}}{2}\|x - x_k\|^2$$

$$\lambda_{k+1} \leftarrow \lambda_k + \rho\left(F(x_k) + J_F(x_k)(x_{k+1} - x_k)\right)$$

satisfy *descent inequality*

$$P_{k+1} - P_k \leq \frac{3}{2\rho}\|\Delta\lambda_{k+1}\|^2 - \frac{\beta_{k+1}}{4}\|\Delta x_{k+1}\|^2 - \frac{\beta_k}{4}\|\Delta x_k\|^2$$

Discussion:

- objective function in subproblem is unconstrained, quadratic and strongly convex $\Rightarrow$ finding a solution is equivalent to solving a linear system
- update of dual multipliers is different from literature, i.e., instead $\lambda_{k+1} = \lambda_k + \rho F(x_{k+1})$, we evaluate the linearization of F at x_k in the new point x_{k+1} and update $\lambda_{k+1} = \lambda_k + \rho(F(x_k) + J_F(x_k)(x_{k+1} - x_k))$.
- Is algorithm well-posed ($\exists \beta_{k+1}$)?

Convergence analysis

Lemma (Bound for $\|\Delta\lambda_{k+1}\|$)

Consider algorithm L-AL. Suppose that for a fixed $k \geq 1$ our assumption holds for some set $\mathcal{S}$ and that $x_{k-1}, x_k \in \mathcal{S}$. Then,

$$\|\Delta\lambda_{k+1}\|^2 \leq c(\beta_{k+1})\|\Delta x_{k+1}\|^2 + c(\beta_k)\|\Delta x_k\|^2,$$

where $c(\beta) = \frac{4(1+3\mu)^2(L_f M_F + M_f L_F)^2}{\sigma^4} + \frac{4(1+3\mu)^2 M_F^2}{\sigma^4}(\beta - \mu L_f)^2$ *and $\mu > 1$ (from the line search procedure).*

Lemma (Existence of β_{k+1})

Consider algorithm L-AL. Suppose that for a fixed $k \geq 0$, our assumption holds for some set $\mathcal{S}$ and that $x_k, x_{k+1} \in \mathcal{S}$ together with $\lambda_k \in \Lambda$, where Λ is a compact set of $\mathbb{R}^m$. Aditionally assume $f(x) \geq \underline{f}$ for all $x \in \mathbb{R}^n$. If β_{k+1} is chosen to satisfy:

$$\beta_{k+1} \geq L_f + L_F\sqrt{2\rho}\sqrt{\mathcal{L}_\rho(x_k, \lambda_k) + \frac{1}{2\rho}\|\lambda_k\|^2 - \underline{f}}$$

then descent inequality holds, i.e., algorithm L-AL is well-posed.

Remark: Note that β depends on $\sqrt{\rho}$ not on ρ (crucial in our analysis)!

Usually β is determined through a standard line search at each iteration k.

Convergence analysis cont.

Lemma (Boundedness of primal-dual sequence)

If $\bar{f} \geq f(x) \geq \underline{f}$ *for all* $x : \|F(x)\| \leq 1$ *(left) and respectively* $x \in \mathbb{R}^n$ *(right),* $\|F(x_0)\|^2 \leq \min\left\{1, \frac{c_0}{\rho}\right\}$ *for some* $c_0 > 0$ *and our assumption hold on:*

$$\mathcal{S} = \{x : \ f(x) + \frac{\rho_0}{2}\|F(x)\|^2 \leq \bar{P}\} \text{ and } \bar{P} = \bar{f} + c_0 + 4\|\lambda_0\|^2 + 2.$$

If $\rho \geq \textit{expression}(\rho_0, M_f, L_f, M_F, L_F, \sigma, \bar{P})$*, then for* $k \geq 1$ *the following holds:*

$$\beta_k \leq \bar{\beta}, \quad P_k \leq \bar{P},$$
$$x_k \in \mathcal{S}, \quad \|\lambda_k\|^2 \leq 2\bar{\gamma}(\rho - \rho_0),$$
$$P_{k+1} - P_k \leq -\frac{\beta_{k+1}}{8}\|\Delta x_{k+1}\|^2 - \frac{\beta_k}{8}\|\Delta x_k\|^2.$$

Remark:

- main challenge when using (augmented) Lagrangian lies in simultaneously ensuring feasibility and optimality $\Rightarrow$ common approach assumes *boundedness of dual iterates* and/or progressively increasing penalty parameter ρ (Teboulle'22; Sahin'19; Birgin'20,...)
- boundedness assumption presents limitation, as it's imposed on algorithm's sequence rather than being an inherent property of problem itself
- boundedness of multiplier sequence in nonconvex setting is a difficult matter because coercivity arguments do not apply directly
- proper assumptions and analysis allow to bound x_k and λ_k, while keeping ρ constant (depending on problem's data), e.g., no need $\rho \approx \epsilon^{-1}$!

Convergence analysis cont.

Theorem (Limit points are KKT points)

Under the assumptions of previous lemma, any limit point (x^*, λ^*) *of sequence* $\{(x_k, \lambda_k)\}_{k\geq 1}$ *generated by algorithm L-AL is a KKT point of our problem, i.e.:*

$$\nabla f(x^*) + J_F(x^*)^T \lambda^* = 0, \qquad F(x^*) = 0.$$

If additionally Lyapunov function $P(\cdot)$ *satisfies the KL property, then the whole sequence* $\{(x_k, \lambda_k)\}_{k\geq 1}$ *converges to a KKT point of our problem.*

Theorem (First-order complexity)

Under the assumptions of previous lemma and considering $\epsilon > 0$*, sequence* $\{(x_k, \lambda_k)\}_{k\geq 1}$ *generated by algorithm L-AL yields an* ϵ*-first-order solution of our problem after* $K = \mathcal{O}(\sqrt{\rho}\epsilon^{-2})$ *Jacobian evaluations, i.e.:*

$$\|\nabla f(x_K) + J_F(x_K)^T \lambda_K\| \leq \epsilon \quad \text{and} \quad \|F(x_K)\| \leq \epsilon.$$

Convergence analysis for L-AL: takeaways

- *obtained optimal complexity* (?) - $\mathcal{O}(\sqrt{\rho}\epsilon^{-2})$ - in the context of augmented Lagrangian and penalty-based methods for smooth nonconvex constrained optimization problems, as penalty parameter ρ enters under square root and desired accuracy ϵ enters quadratically in algorithm's complexity

$$\text{recall}: \quad \beta = \mathcal{O}(\sqrt{\rho})!$$

- our convergence rate *improves existing complexity results* for augmented Lagrangian (measured through Jacobian evaluations), on the same class of problems: e.g., $\mathcal{O}(\epsilon^{-5.5})$ in Xie&Wright'21; $\mathcal{O}(\epsilon^{-4})$ in Sahin'19;...
- another key advantage lies in its avoidance of calling complicated subroutines, as *unconstrained* subproblem in L-AL algorithm has a *quadratic* strongly convex objective function, compared to Xie&Wright'21; Sahin'19; Birgin'20; ... where subproblem is higly nonconvex

New problem formulation: nonsmooth objective

In previous problem, objective function was smooth. Consider now a nonsmooth separable nonlinear optimization problem:

$$\min_{x\in\mathbb{R}^n, y\in\mathbb{R}^p} f(x) + g(x) + h(y) + \mathbf{1}_{\mathcal{Y}}(y)$$
$$\text{s.t.: } F(x) + Gy = 0$$

- $\mathcal{Y}$ easy subset of $\mathbb{R}^p$, e.g., admiting an easy projection
- matrix $G \in \mathbb{R}^{m\times p}$ *has full row rank*
- functions $f : \mathbb{R}^n \to \mathbb{R}, g : \mathbb{R}^n \to \bar{\mathbb{R}}, h : \mathbb{R}^p \to \mathbb{R}$, and $F \triangleq (f_1, ..., f_m)^T$, with $f_i : \mathbb{R}^n \to \mathbb{R}$ for all $i \in \{1, ..., m\}$, are nonlinear functions
- assume f, h, f_i, for all $i = 1, ..., m$, are continuously differentiable functions, f, h possibly nonconvex and g proper lower semi-continuous and prox-bounded function relative to its domain dom$g$

Remark (compared to previous model):

1. Objective function is separable, but nonsmooth
2. This model allows additional constraints on x and y via $g(\cdot)$ and $\mathbf{1}_{\mathcal{Y}}(\cdot)$
3. Nonlinear equality constraints have a particular structure (see also Teboulle'22)

New problem formulation: nonsmooth objective

Consider nonsmooth separable nonlinear optimization problem:

$$\min_{x\in\mathbb{R}^n, y\in\mathbb{R}^p} f(x) + g(x) + h(y) + \mathbf{1}_{\mathcal{Y}}(y)$$
$$\text{s.t.: } F(x) + Gy = 0$$

Motivation:
For example, any constrained composite optimization problem frequently appearing in nonlinear optimal control (Diehl'21):

$$\min_{x\in\mathcal{X}} f(x) + h(F(x)) \quad \text{s.t.} \quad F(x) \in \mathcal{Y},$$

can be easily recast in the form of our optimization problem by defining $F(x) = y$, then $G = -I_m$ and g the indicator function of the set $\mathcal{X}$ and thus having constraints on both block variables.

In the context of optimal control, f, h are quadratic functions; F describes the nonlinear dynamical system.

New problem formulation: nonsmooth objective

Consider nonsmooth separable nonlinear optimization problem:

$$\min_{x\in\mathbb{R}^n, y\in\mathbb{R}^p} f(x) + g(x) + h(y) + \mathbf{1}_{\mathcal{Y}}(y)$$
$$\text{s.t.: } F(x) + Gy = 0$$

Definition: $(x_\epsilon^*, y_\epsilon^*) \in \text{dom } g \times \mathcal{Y}$ is an ϵ-first-order solution if $\exists \lambda_\epsilon^* \in \mathbb{R}^m$ s.t.:

$$\text{dist}\Big(-\nabla f(x_\epsilon^*) - \nabla F(x_\epsilon^*)^T \lambda_\epsilon^*, \partial g(x_\epsilon^*)\Big) \leq \epsilon, \quad \text{dist}\Big(-\nabla h(y_\epsilon^*) - G^T \lambda_\epsilon^*, N_{\mathcal{Y}}(y_\epsilon^*)\Big) \leq \epsilon,$$
$$\|F(x_\epsilon^*) + G y_\epsilon^*\| \leq \epsilon.$$

Assumptions:
For any compact sets $\mathcal{S}_x \subseteq \text{dom } g$ and $\mathcal{S}_y \subseteq \mathcal{Y}$, there exist positive constants σ, L_f, L_h, L_F such that f, h and F satisfy the following conditions for all $x, x' \in \mathcal{S}_x$ and for all $y, y' \in \mathcal{S}_y$:

1. $\|\nabla f(x) - \nabla f(x')\| \leq L_f \|x - x'\|$
2. $\|\nabla h(y) - \nabla h(y')\| \leq L_h \|y - y'\|$
3. $\|J_F(x) - J_F(x')\|_2 \leq L_F \|x - x'\|$
4. $\sigma_{\min}(G) \geq \sigma > 0$

Remark: In previous model we assumed LICQ $\sigma_{\min}(J_F(x)) > 0$ for all $x \in \mathcal{S}$; now we only require $\sigma_{\min}(G) > 0$! Hence, a condition easier to check.

Other assumptions related to smoothness are similar to first part.

Notations

Consider nonsmooth separable nonlinear optimization problem:

$$\min_{x\in\mathbb{R}^n, y\in\mathbb{R}^p} f(x) + g(x) + h(y) + \mathbf{1}_{\mathcal{Y}}(y)$$
$$\text{s.t.:}\ \ F(x) + Gy = 0$$

- Augmented Lagrangian function associated to our problem:

$$\begin{aligned}\mathcal{L}_\rho(x,y,\lambda) &= f(x) + g(x) + h(y) + \langle \lambda, F(x) + Gy\rangle + \frac{\rho}{2}\|F(x) + Gy\|^2 \\ &= g(x) + h(y) + \psi_\rho(x,y,\lambda),\end{aligned}$$

- where smooth part w.r.t. x is denoted

$$\psi_\rho(x,y,\lambda) = f(x) + \langle \lambda, F(x) + Gy\rangle + \frac{\rho}{2}\|F(x) + Gy\|^2.$$

- Denote the function derived from linearization of objective and functional constraints in a Gauss-Newton fashion, at a given point $\bar{x}$:

$$\begin{aligned}&\bar{\mathcal{L}}_\rho(x,y,\lambda;\bar{x},\bar{y}) \\ &= l_f(x;\bar{x}) + g(x) + l_h(y;\bar{y}) + \langle \lambda, l_F(x;\bar{x}) + Gy\rangle + \frac{\rho}{2}\|l_F(x;\bar{x}) + Gy\|^2\end{aligned}$$

Inexact Linearized ADMM (iL-ADMM)

We consider the following algorithm:

Algorithm iL-ADMM

Given x_0, y_0, λ_0 and $\rho, \underline{\beta}, \underline{\theta}, \alpha > 0$.

For $k \geq 0$ do:

1. find a proximal parameter $\beta_{k+1} \geq \underline{\beta}$ such that

$$x_{k+1} \approx \arg\min_{x} \bar{\mathcal{L}}_\rho(x, y_k, \lambda_k; x_k, y_k) + \frac{\beta_{k+1}}{2}\|x - x_k\|^2$$

satisfies an inexact stationary condition and a descent, i.e.:

$$\exists s_{k+1} \in \partial_x \left(\bar{\mathcal{L}}_\rho(x, y_k, \lambda_k; x_k, y_k) + \frac{\beta_{k+1}}{2}\|x - x_k\|^2 \right) \Big|_{x = x_{k+1}}$$

such that

$$\|s_{k+1}\| \leq \alpha \|x_{k+1} - x_k\|$$

$$\psi_\rho(x_{k+1}, y_k, \lambda_k) - l_{\psi_\rho}(x_{k+1}, y_k, \lambda_k; x_k) \leq \frac{\beta_{k+1}}{4}\|x_{k+1} - x_k\|^2$$

2. find a proximal parameter $\theta_{k+1} \geq \underline{\theta}$ such that

$$y_{k+1} \leftarrow \arg\min_{y \in \mathcal{Y}} \bar{\mathcal{L}}_\rho(x_{k+1}, y, \lambda_k; x_{k+1}, y_k) + \frac{\theta_{k+1}}{2}\|y - y_k\|^2$$

satisfies the following inequality:

$$h(y_{k+1}) - l_h(y_{k+1}; y_k) \leq \frac{\theta_{k+1}}{4}\|y_{k+1} - y_k\|^2.$$

3. Update

$$\lambda_{k+1} \leftarrow \lambda_k + \rho\left(F(x_{k+1}) + G y_{k+1}\right).$$

Survey paper on ADMM: Boyd'11.

Inexact Linearized ADMM (iL-ADMM): discussion

- Dominant steps in algorithm iL-ADMM are Step 1 and Step 2
- Step 1 involves nonsmooth function g in addition to a quadratic term. When g is convex or weakly convex, the objective function of the subproblem in Step 1 is usually strongly convex
- Moreover, subproblem in Step 1 is solved inexactly
- In contrast, subproblem in Step 2 has always a strongly convex quadratic function, even if h is nonconvex, and a feasible set $\mathcal{Y}$
- Regularization (proximal) parameters β_{k+1} and θ_{k+1} are dynamically chosen and are well defined since ψ_ρ and h are smooth functions (to determine them, one can use a standard line search procedure)
- Dual variables are updated in Step 3 using the conventional update of dual multipliers in traditional augmented Lagrangian methods

Additional Assumption:
(i) Sequence $\{(x_k, y_k, \lambda_k)\}_{k\geq 0}$ generated by algorithm iL-ADMM is bounded.
(ii) Set $\mathcal{Y}$ admits a Lipschitz continuous normal cone mapping[1].

Remark:
Previously we proved boundedness of primal-dual sequence generated by L-AL. Now, when we have additional constraints, we assume their boundedness.

[1] $\mathrm{dist}_H\left(\bar{N}_{\mathcal{Y}}(y), \bar{N}_{\mathcal{Y}}(y')\right) \leq \kappa \|y - y'\| \quad \forall y, y' \in \mathcal{Y}$, where e.g., $\bar{N}_{\mathcal{Y}}(y) = N_{\mathcal{Y}}(y) \cap \mathbb{B}_r$

Convergence analysis

As for L-AL algorithm, we use a similar Lyapunov function:

$$P(x, y, \lambda, \bar{y}, \gamma) = \mathcal{L}_\rho(x, y, \lambda) + \frac{\gamma}{2}\|y - \bar{y}\|^2$$

and define

$$P_k = P(x_k, y_k, \lambda_k, y_{k-1}, \theta_k/4)$$

Theorem (Limit points are KKT points)

(i) Let $\{z_k := (x_k, y_k, \lambda_k\}_{k\geq 1}$ be generated by algorithm iL-ADMM. If assumptions on smoothness and boundedness hold and $\rho \geq expression(\underline{\theta}, L_h, \sigma)$, then any limit point $z^* := (x^*, y^*, \lambda^*)$ of $\{z_k\}_{k\geq 1}$ is a KKT point of our problem.

(ii) If additionally Lyapunov function $P(\cdot)$ satisfies KL property, then whole sequence $\{z_k := (x_k, y_k, \lambda_k\}_{k\geq 1}$ converges to a KKT point of our problem.

Theorem (First-order complexity)

(i) Let $\{z_k := (x_k, y_k, \lambda_k)\}_{k\geq 1}$ be generated by algorithm iL-ADMM. If assumptions on smoothness and boundedness hold and $\rho \geq expression(\underline{\theta}, L_h, \sigma)$, then for any $\epsilon > 0$, algorithm iL-ADMM yields an ϵ-first-order solution after $K = \mathcal{O}(\epsilon^{-2})$ Jacobian evaluations.

(ii) If additionally f, g, h, F are semi-algebraic ($\iff$ $P(\cdot)$ semi-algebraic), improved rates can be derived.

Convergence analysis for iL-ADMM: some conclusions

- *obtained complexity* - $\mathcal{O}(\epsilon^{-2})$ - best rate in the context of augmented Lagrangian and penalty-based methods for nonconvex constrained optimization problems.
- our optimization problem allows additional (inequality) constraints on both block variables x and y compared to e.g., Teboulle'22 which has only constraints on x
- another key advantage lies in its avoidance of calling complicated subroutines, as subproblems in iL-ADMM algorithm usually have *quadratic* strongly convex objective function and simple constraints (inexact solutions of subproblems are also possible in our framework)

Comparisson L-AL and iL-ADMM

Algorithms L-AL and iL-ADMM have common & different features

L-AL $\longleftrightarrow$ "Let us go together"

iL-ADMM $\longleftrightarrow$ "Let us go to get her"

- We considered general nonconvex problems: nonconvex - nonsmooth objective and nonlinear equality constraints
- However, in first part everything was smooth; second part allowed nonmooth terms in objective function (but some separability and special nonlinear equality constraints)
- Proposed augmented Lagrangian-based algorithms using linearization of (smooth part of) objective and of functional constraints in a Gauss-Newton fashion: L-AL and iL-ADMM
- Iterates in L-AL and iL-ADMM are simple to compute: convex subproblems that are easy to solve (even inexact)
- Penalty parameter ρ in L-AL and iL-ADMM depends on parameters of problem's functions, no need to depend on ϵ
- Derived (optimal) global convergence rates for L-AL and iL-ADMM: both alg's enjoy $\mathcal{O}(\epsilon^{-2})$ Jacobian evaluations to get an ϵ-first-order solution

Extensions and future work

- Extension of L-AL algorithm to nonsmooth problems:

$$\min_x f(x) \ \text{ s.t. } \ F(x) = 0 \ \Rightarrow \ \min_x f(x) + g(x) \ \text{ s.t. } \ F(x) = 0,$$

  where g is nonsmooth[2].

- before we used LICQ, for this new problem we need to define a proper constraint qualification condition (see [BNPT]), i.e.:

$$\sigma_{\min}(J_F(x)) \geq \sigma > 0 \ \text{ for all } x \in \mathcal{S}$$

  versus

$$\sigma \|F(x)\| \leq \operatorname{dist}\left(-J_F(x)^T F(x), \partial^\infty g(x)\right) \ \text{ for all } x \in S$$

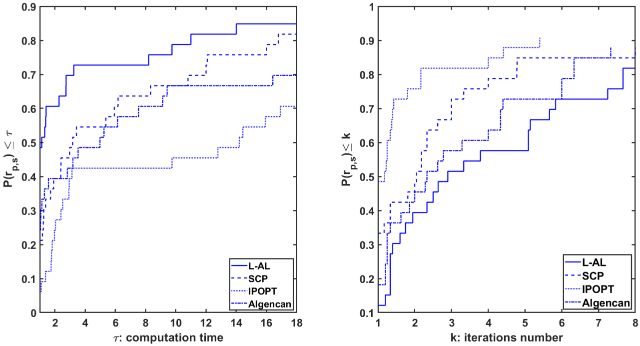
- consider also "linearized" penalty methods[3]

[2][BNPT]: L. El Bourkhissi, I. Necoara, P. Patrinos, Q. Tran-Dinh, *Complexity of linearized perturbed augmented Lagrangian methods for nonsmooth nonconvex optimization with nonlinear equality constraints*, arxiv, 2025.

[3][BN]: L. El Bourkhissi, I. Necoara, *Convergence analysis of linearized ℓ_q penalty methods for nonconvex optimization with nonlinear equality constraints*, UPB Scientific Bulletin, 2025.

Numerical simulations: L-AL

- we numerically compare algorithms L-AL, SCP (Diehl'21), IPOPT and Algencan (Birgin'20) - which is also an augmented Lagrangian method.
- we stop algorithms when difference between two consecutive values of objective is less than 10^{-3} & norm of constraints is less than 10^{-5}
- large-scale real-world problems with nonlinear equality constraints selected from CUTEst collection
- performance profile for computation time and number of iterations: L-AL is the fastest, but needs many (simple) iterations!



Numerical simulations: results for L-AL

Alg / (n,m)	L-AL		SCP		IPOPT		Algencan	
	# iter / f^*	cpu / $\|F\|$	# iter / f^*	cpu / $\|F\|$	# iter / f^*	cpu / $\|F\|$	# iter / f^*	cpu / $\|F\|$
OPTCTRL3	56	**19.57**	24	105.08	11	26.95	11	102.68
(4499,3000)	74465.03	1.76e-08	74465.03	6.87e-09	74465.03	1.09e-08	74470	8.66 e-09
DTOC4	4	**46.91**	4	566.80	3	146.73	18	149.66
(14997,9998)	2.87	3.40e-07	2.87	1.05e-07	2.86	4.49e-09	2.86	7.27e-09
DTOC5	23	**42.40**	24	799.07	3	75.25	18	48.27
(9998,4999)	1.54	2.19e-07	1.54	1.96e-07	1.53	2.49e-07	1.53	3.31 e-07
ORTHREGA	58	**65.72**	-	-	20	71.78	40	68.61
(8197,4096)	22647.84	1.83e-07	-	-	22674.84	1.86e-09	22674.84	6.32e-08
MSS1	70	**1.23**	12	0.15	53	13.52	15	0.53
(90, 73)	-15.99	8.11e-06	-8.71e-08	1.76e-06	-16.00	4.17e-08	-15.00	3.29 e-08
MSS2	58	**21.99**	21	8.05	7	14.65	-	-
(756, 703)	-123.99	3.11e-06	-2.53e-10	6.12e-06	-26.97	5.96e-08	-	-
MSS3	58	**106.79**	22	135.15	-	-	-	-
(2070, 1981)	-338.91	9.42e-07	-5.29e-09	7.76e-06	-	-	-	-
OPTCTRL6	56	19.03	24	**13.46**	13	27.42	11	101.34
(4499, 3000)	74465.03	1.85e-08	74465.03	3.27e-09	74465.03	2.32e-09	74470	8.47 e-09
OPTCDEG3	9	**9.01**	43	22.64	11	21.33	25	84.92
(4499, 3000)	12.13	8.28e-06	12.13	6.12e-06	12.13	4.61e-07	12.13	7.04e-07
ORTHREGC	28	**20.09**	29	20.93	16	31.14	26	17.82
(5005, 2500)	94.81	9.92e-06	94.81	8.42e-06	94.81	7.52e-07	94.81	3.07e-07
EIGENC2	6	**1.95**	6	4.68	13	24.93	6	32.02
(2652, 1326)	0.01	5.98e-06	11162.75	4.64e-16	0.00	8.43e-10	0.00	3.51e-10
EIGENACO	5	2.43	8	1.75	-	-	2	**1.87**
(2550, 1275)	0.01	4.22e-06	22425.04	2.37e-18	-	-	0.00	3.21e-09
EIGENBCO	7	**3.37**	5	1.23	9	19.58	-	-
(2550, 1275)	0.01	1.45e-06	49.50	5.79e-16	0.00	3.18e-17	-	-
DTOC1NA	29	54.24	4	3.86	5	12.08	5	**0.23**
(5994, 3996)	4.14	3.09e-06	47.66	5.03e-13	4.15	7.44e-11	4.14	8.03e-10
SPINOP	101	**76.31**	-	-	-	-	-	-
(1327, 1325)	150.50	9.34e-06	-	-	-	-	-	-
ROBOTARM	131	**106.41**	-	-	7	109.20	23	377.26
(4400, 3202)	7.84	9.62e-06	-	-	9.14	2.05e-08	9.14	1.20e-08

Numerical simulations: iL-ADMM

- we numerically compare alg. iL-ADMM, dynamic linearized alternating direction method of multipliers (DAM) (Teboulle'22) and IPOPT
- we stop algorithms when approximate KKT conditions are less than 10^{-3}
- nonlinear model predictive control problems for several nonlinear systems: inverted pendulum (IP), single machine infinite bus (SMIB), lane tracking (LT), four tanks (4T) and free-flying robot (FFR)
- numerical results for iL-ADMM, DAM and IPOPT on solving $N_{\text{sim}} = 50$ nonlinear MPC problems for 5 dynamical systems of different dimensions (average and standard deviation results for #iter and cpu).

Algorithm / System (i_d, s_d)	iL-ADMM			DAM			IPOPT		
	E(# iter)	E(cpu)	f^*	E(# iter)	E(cpu)	f^*	E(# iter)	E(cpu)	f^*
	σ(# iter)	σ(cpu)	$\|F\|$	σ(# iter)	σ(cpu)	$\|F\|$	σ(# iter)	σ(cpu)	$\|F\|$
SMIB	495.36	**0.90**	0.3088	3621.65	9.42	0.3089	87.28	1.12	0.3087
(2,4)	0.06	3.44e-4	9.95e-7	0.19	0.01	9.97e-7	6e-2	1e-3	2.87e-8
IPOC	135.12	**0.31**	166.60	497.76	16.23	166.60	28.7	0.57	166.60
(1,4)	0.02	1.08e-4	8.74e-7	0.08	4.12e-3	9.39e-7	2.3e-2	1.04e-3	3.69e-8
4T	658.31	**1.93**	87.95	2846.79	7.50	87.94	187.64	2.79	87.94
(2,6)	0.04	5.04e-4	9.32e-7	0.17	0.03	9.94e-7	2.66e-3	4.36e-2	3.68e-8
LT	102.07	**0.53**	6.99	675.63	7.10	7.00	32.94	1.03	6.98
(2,7)	0.02	2.29e-4	8.98e-7	0.15	0.01	9.65e-7	0.005	1.24e-4	4.93e-8
FFR	121.24	**1.57**	1066.27	1734.66	11.94	1066.32	23.94	2.88	1065.87
(2,6)	0.01	1.37e-4	9.90e-7	0.21	0.009	9.96e-7	0.02	1.69e-4	1.87e-7

References

- L. El Bourkhissi, I. Necoara, *Complexity of a linearized augmented Lagrangian method for nonconvex minimization with nonlinear equality constraints*, arxiv, 2025.
- L. El Bourkhissi, I. Necoara, *Convergence rates for an inexact linearized ADMM for nonsmooth nonconvex optimization with nonlinear equality constraints*, Comp. Opt. Applic., 2024.
- E. Cohen, N. Hallak and M. Teboulle, *A Dynamic alternating direction of multipliers for nonconvex minimization with nonlinear functional equality constraints*, J. Optim. Theory Applic., 2022.
- F. Messerer, K. Baumgärtner and M. Diehl, *Survey of sequential convex programming and generalized Gauss-Newton methods*, ESAIM: Proceedings and Surveys, 2021.
- E.G. Birgin and J.M. Martínez, *Complexity and performance of an augmented Lagrangian algorithm*, Optim. Met. Soft., 2020.
- M.F. Sahin, A. Eftekhari, A. Alacaoglu, F.L. Gomez and V. Cevher, *An inexact augmented Lagrangian framework for nonconvex optimization with nonlinear constraints*, Neural Information Processing Systems, 2019.

Thank you for your attention!

Q&A