Proximal Methods in Numerical Optimization

Lecture IV – Nonsmooth Problems with Constraints

Alberto DE MARCHI
University of the Bundeswehr Munich
alberto.demarchi@unibw.de
aldma.github.io

Povo, UniTN – March 5, 2026

⚠ these slides are under development: please email me for corrections and suggestions ⚠

Outline

Introduction

Why Constrained Nonsmooth Optimization?

Many problems in science and engineering are naturally:

- **Nonconvex**: motion planning, phase retrieval, neural networks
- **Nonsmooth**: ℓ_1 regularization, matrix rank, contact mechanics
- **Constrained**: feasibility sets, system dynamics, logical conditions

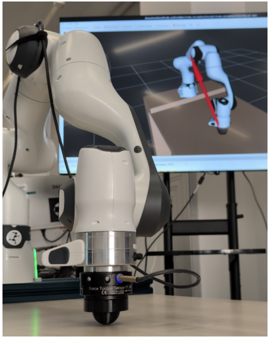
Motivating applications

Signal recovery

Optimal control

Machine learning

Robotics



Justin Carpentier's team @ INRIA Paris

General Problem Template

Constrained Composite Program

$$\underset{x \in \mathcal{X}}{\text{minimize}} \quad f(x) + g(x) \quad \text{subject to} \quad c(x) \in \mathcal{D}$$

$f : \mathbb{R}^n \to \mathbb{R}$	smooth
$g : \mathbb{R}^n \to \overline{\mathbb{R}}$	nonsmooth, extended-real-valued
$c : \mathbb{R}^n \to \mathbb{R}^m$	smooth constraint map
$\mathcal{D} \subset \mathbb{R}^m$	closed set
$\mathcal{X} \subset \mathbb{R}^n$	closed set

Special cases

- $g = 0$, $\mathcal{X} = \mathbb{R}^n$, $\mathcal{D} = \{0\}$: equality-constrained smooth NLP
- $g = \delta_C$ (indicator of a convex set C)

Subdiffferential Calculus

Convex subdifferential

$\hat{\partial}\phi(x) := \{s \in \mathbb{R}^n : \phi(y) \geq \phi(x) + \langle s, y - x \rangle, \ \forall y\}$

Limiting (Mordukhovich) subdifferential

$\partial\phi(x) := \{\limsup_{y \to x} \hat{\partial}\phi(y)\}.$

Optimality conditions

- Unconstrained: $0 \in \partial\phi(x^*)$.
- Constrained ($\min \phi$ s.t. $x \in C$, smooth ϕ): $-\nabla\phi(x^*) \in N_C(x^*)$.
- KKT (smooth NLP): stationarity + complementarity + primal feasibility.

Stationarity notions for nonsmooth nonconvex problems: Clarke, Mordukhovich, proximal, approximately stationary (inexact) ...

Variational analysis, a whole new world...

"Unconstrained" Minimization: Recap

Problem

Given f smooth and g prox-friendly,

$$\underset{x \in \mathbb{R}^n}{\text{minimize}} \quad \varphi(x) := f(x) + g(x)$$

Proximal Gradient Step

$$x_{k+1} \in \text{prox}_{\gamma_k g}\left(x_k - \gamma_k \nabla f(x_k)\right)$$

Derivation: Linearize f around x_k, add proximal term, minimize:

$$x_{k+1} \in \arg\min_x \left\{ f(x_k) + \langle \nabla f(x_k), x - x_k \rangle + g(x) + \frac{1}{2\gamma_k} \|x - x_k\|^2 \right\},$$

check progress, backtrack if needed, repeat.

Approaches to Constraints

Exact penalty / indicator	Augmented Lagrangian	Interior point / barrier	Composite steps / SQP

Each approach makes a different trade-off between

- feasibility maintenance vs. asymptotic feasibility.
- smoothness of subproblems.
- scalability and per-iteration cost.

Constrained structured optimization

$$\underset{x}{\text{minimize}} \quad f(x) + g(x) \qquad \text{subject to} \quad c(x) \leq 0 \tag{P}$$

Assumptions:

- f and c are smooth
- g is prox-friendly
- no convexity assumptions

Applications:

- (nonsmooth) regularization
- indicators CC, VC, SC, ...
- cover $c(x) \in D$ and $g(c(x))$

PIPA *A proximal interior point algorithm with applications to image processing* [CCP20]

IPPROX *An interior proximal gradient method for nonconvex optimization* [DMT24], also [DMT25]

RALMS *An adaptive Lagrangian-based scheme for nonconvex composite optimization* [HT23], also [BST18, CHT22, CT25]

ALPS *Constrained composite optimization and augmented Lagrangian methods* [DMJKM23], also [DM24, DMM24, DMHM25]

KKT-like optimality

Relative to the problem

$$\underset{x}{\text{minimize}} \quad q(x) \qquad \text{subject to} \quad c(x) \leq 0$$

a point $\bar{x} \in \mathbb{R}^n$ is called

- **feasible** if it satisfies $\bar{x} \in \operatorname{dom} q$ and $c(\bar{x}) \leq 0$;
- **strictly feasible** if it satisfies $\bar{x} \in \operatorname{dom} q$ and $c(\bar{x}) < 0$;
- **KKT optimal** if there exists $\bar{y} \in \mathbb{R}^m$ such that

$$0 \in \partial q(\bar{x}) + c'(\bar{x})^\top \bar{y} \qquad c(\bar{x}) \leq 0$$
$$\bar{y} \geq 0 \qquad \bar{y}_i c_i(\bar{x}) = 0 \quad \forall i;$$

- **ε-KKT optimal** (with $\varepsilon > 0$) if there exists $\bar{y} \in \mathbb{R}^m$ such that

$$\operatorname{dist}(-c'(\bar{x})^\top \bar{y}, \partial q(\bar{x})) \leq \varepsilon \qquad c(\bar{x}) \leq \varepsilon$$
$$\bar{y} \geq 0 \qquad \min\{\bar{y}_i, -c_i(\bar{x})\} \leq \varepsilon \quad \forall i.$$

- **AKKT optimal** if feasible and there exist sequences $\{x^k\}$ and $\{y^k\}$ such that $x^k \to \bar{x}$, $y^k \geq 0$ and

$$\operatorname{dist}(0, \partial q(x^k) + c'(x^k)^\top y^k) \to 0, \qquad y_i^k c_i(x^k) \to 0 \quad \forall i;$$

Barrier methods

Structured optimization and barrier idea

We know how to solve

$$\underset{x}{\text{minimize}} \quad \varphi(x) := f(x) + g(x)$$

Now, we may include also a barrier term

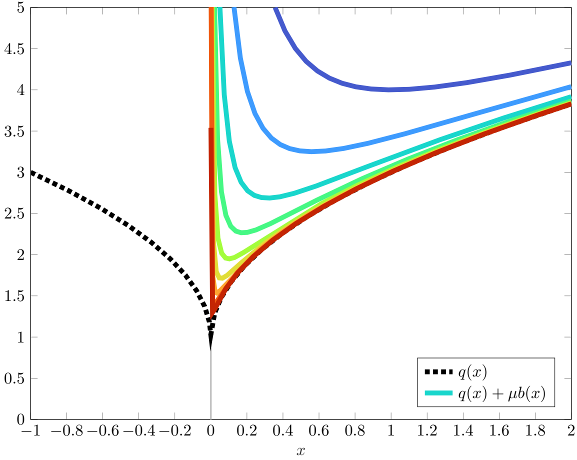
$$\underset{x}{\text{minimize}} \quad \varphi(x) - \log x$$

... to model constraints

$$\underset{x}{\text{minimize}} \quad \varphi(x) \qquad \text{subject to} \quad x \geq 0$$

Illustration

- smooth term $(f + \mu b)$ has not full domain!
- strong motivation to extend PGM **from globally to locally Lipschitz gradient**



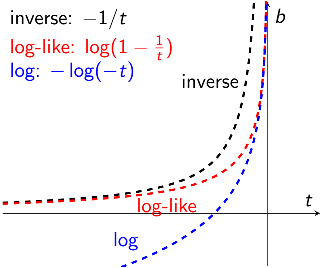
$x \in \mathbb{R}$, $f(x) := 1$, $g(x) := \|x\|_{1/2}^{1/2}$, $b(x) := 1/x$ for $x > 0$, $\mu \in [0.001, 1]$ $\qquad$ $\min_x q(x)$ s.t. $x \geq 0$

Interior Point Subproblem

$$\min_x \quad \varphi(x) \qquad \text{subject to} \quad c(x) \le 0 \qquad \text{(P)}$$

Barrier $b\colon (-\infty, 0) \to [0, \infty)$

- $b' > 0$ on its domain,
- b convex,
- $b(t) \to \infty$ as $t \to 0^-$.

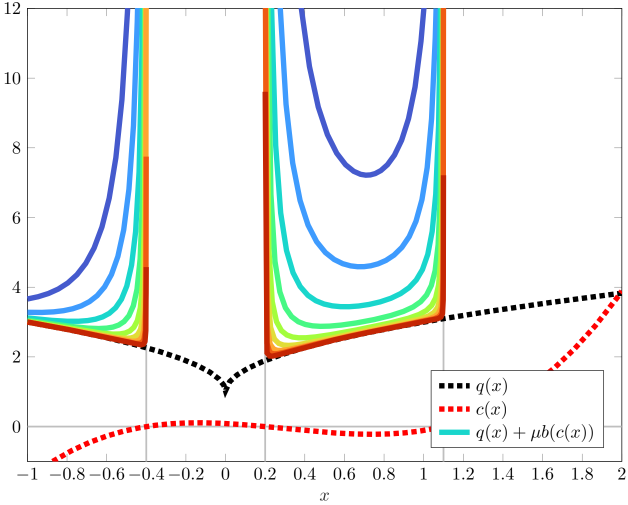


Sequential unconstrained optimization, with barrier parameter $\mu > 0$:

$$\min_x \quad \underbrace{f(x) + \mu \sum_i b(c_i(x))}_{\text{smooth } F_\mu} + g(x) \qquad (\mathrm{P}_\mu)$$

The smooth term $F_\mu := f + \mu \sum_i b \circ c_i$ has not full domain…

Illustration: barrier subproblems



$x \in \mathbb{R}$, $b(x) := -1/x$ for $x < 0$, $\mu \in [0.001, 1]$, $c(x) = \text{cubic}(x)$

$\min_x q(x)$ s.t. $c(x) \leq 0$

Interior Point Framework

Basic Scheme

$$F_\mu(x) := f(x) + \mu \sum_i b(c_i(x))$$

1: find a strictly feasible point x^0 for (P)
2: choose parameters $\mu_0 > 0$ and $\theta_\mu \in (0, 1)$
3: **for** $k = 0, 1, 2, \dots$ **do**
4: find $x^{k+1} \in \arg\min(F_{\mu_k} + g)$
5: set $\mu_{k+1} \leftarrow \theta_\mu \mu_k$
6: **end for**

$\rightsquigarrow$ ill-conditioning as $\mu \searrow 0$
- $\rightsquigarrow$ warm-starting from x_k

$\rightsquigarrow$ affordable subsolver (only descent needed, no global minima)
- $\rightsquigarrow$ inexact subsolves

Algorithm: IPPROX [DMT24]

find a strictly feasible point x^0 for (P)
select tolerances $\epsilon_p, \epsilon_d > 0$
choose parameters $\varepsilon_0, \mu_0 > 0$ and $\theta_\varepsilon, \theta_\mu \in (0, 1)$
for $k = 0, 1, 2, \ldots$ **do**
    find an ε_k-stationary point x^{k+1} of $F_{\mu_k} + g$ starting from x^k
    set $y_i^{k+1} \leftarrow \mu_k b'(c_i(x^{k+1})) \quad \forall i$
    if $\varepsilon_k \leq \epsilon_d$ and $\| \min\{y^{k+1}, -c(x^{k+1})\}\| \leq \epsilon_p$ **then**
        return $(x^*, y^*) \leftarrow (x^{k+1}, y^{k+1})$
    end if
    set $\mu_{k+1} \leftarrow \theta_\mu \mu_k$
    set $\varepsilon_{k+1} \leftarrow \max\{\epsilon_d, \theta_\varepsilon \varepsilon_k\}$
end for

- IPFB: a barrier-friendly proximal gradient method

Assumptions for IPPROX

$$\min_{x} \quad \varphi(x) := f(x) + g(x) \qquad \text{subject to} \quad c(x) \leq 0 \tag{P}$$

- $f : \mathbb{R}^n \to \mathbb{R}$ has locally Lipschitz-continuous gradient.
- $g : \mathbb{R}^n \to \overline{\mathbb{R}} := \mathbb{R} \cup \{\infty\}$ is prox-friendly.
- $c : \mathbb{R}^n \to \mathbb{R}^m$ has locally Lipschitz-continuous Jacobian.
- $\inf\{\varphi(x) \mid c(x) \leq 0\} \in \mathbb{R}$.
- strict feasibility, namely

$$\mathcal{F} := \operatorname{dom} \varphi \cap \{x \mid c(x) < 0\} \neq \emptyset.$$

- some $x^0 \in \mathcal{F}$ is known.

Interior-Point Proximal Method: IPPROX [DMT24]

Finite termination:
For any strictly feasible starting point x^0 and tolerance $\epsilon > 0$,
IPPROX returns in finitely many steps
an ϵ-KKT optimal point $x^\star$ for (P) satisfying $\varphi(x^\star) \leq \varphi(x^0)$.

Properties:

- All iterates are **strictly feasible** ($c(x^k) < 0$) by design
- As $\mu \to 0$, minimizers converge to the solution of the original problem
- Ill-conditioned as $\mu \to 0$
- Slow tail convergence...

Key features:

- **Warm starting**: exploits previous feasible point
- Objective values **decrease monotonically**
- **Nonconvex** f: convergence to ϵ-KKT points in finite iterations

We have to deal with equalities too

Can we solve

$$\underset{x}{\text{minimize}} \quad \varphi(x) \quad \text{subject to} \quad c(x) = 0$$

by means of IPPROX applied to

$$\underset{x}{\text{minimize}} \quad \varphi(x) \quad \text{subject to} \quad \begin{pmatrix} c(x) \\ -c(x) \end{pmatrix} \leq 0 \quad ?$$

No, because there is no interior / strictly feasible point. We need to relax!

Penalty methods

Indicator and Exact Penalty

Problem:

$$\underset{x}{\text{minimize}} \quad \varphi(x) \quad \text{subject to} \quad c(x) = 0$$

Indicator instead of constraints? Not really easier...

$$\underset{x}{\text{minimize}} \quad \varphi(x) + \delta_0(c(x))$$

Exact penalty: nonsmooth [Zangwill 1967, Pietrzykowski 1969], often not easy:

$$\underset{x}{\text{minimize}} \quad \varphi(x) + \eta\|c(x)\|_p$$

with $p \in \{1, 2\}$ and $\eta \nearrow \infty$.

Drawback: ill-conditioning for large η; infeasible iterates.

Forgotten for decades, but now they are coming back [EFOS20, DGO26].
Why? Because ϱ can stay bounded, at least for regular problems—$\bar{\eta} = \|y^\star\|_\infty$.

Quadratic Penalty

Problem:

$$\underset{x}{\text{minimize}} \quad \varphi(x) \quad \text{subject to} \quad c(x) = 0$$

Quadratic penalty yields some smoothness again:

$$\underset{x}{\text{minimize}} \quad \varphi(x) + \varrho\|c(x)\|^2$$

with $\varrho \nearrow \infty$.

Advantage: address subproblem with any reasonable PGM.

$$x_{k+1} \in \arg\min_{x} \left\{ \underbrace{f(x) + \varrho_k\|c(x)\|^2}_{\text{smooth}} + g(x) \right\}$$

Drawback: ill-conditioning for large ϱ; infeasible iterates.

Need $\varrho \nearrow \infty$, even for regular problems!

Very often, the quadratic penalty leaves the smooth part with **only locally Lipschitz gradient**. Again, good motivation for relaxing assumptions in PGM.

Augmented Lagrangian methods

Augmented Lagrangian Method (ALM)

Problem:

$$\underset{x}{\text{minimize}} \quad \varphi(x) \quad \text{subject to} \quad c(x) = 0$$

Augmented Lagrangian function

$$\mathscr{L}_\varrho(x, y) := \varphi(x) + \langle y, c(x) \rangle + \frac{\varrho}{2} \|c(x)\|^2$$

ALM scheme

1: Given x^0, y^0, $\rho > 0$
2: **for** $k = 0, 1, 2, \ldots$ **do**
3: $\quad x^{k+1} \approx \arg\min_x \mathscr{L}_\varrho(x, y^k)$ *(proximal inner solver)*
4: $\quad y^{k+1} = y^k + \varrho c(x^{k+1})$ *(dual ascent)*
5: **end for**

Key advantages: warm-start friendly;
multiplier update improves constraint satisfaction at each iteration;
avoids $\varrho \to \infty$ (for regular problems).

Augmented Lagrangian Method (ALM)

Problem:

$$\underset{x}{\text{minimize}} \quad \varphi(x) \quad \text{subject to} \quad c(x) = 0$$

Augmented Lagrangian function

$$\mathcal{L}_\varrho(x, y) := \varphi(x) + \langle y, c(x) \rangle + \frac{\varrho}{2} \|c(x)\|^2$$

Dual update rule:

$$y^{k+1} = y^k + \varrho c(x^{k+1})$$

Why?

$$\begin{aligned} \partial_x \mathcal{L}_{\varrho_k}(x^{k+1}, y^k) &= \partial \varphi(x^{k+1}) + J_c(x^{k+1})^\top \left(y^k + \varrho_k c(x^{k+1}) \right) \\ &= \partial \varphi(x^{k+1}) + J_c(x^{k+1})^\top y^{k+1} \\ &= \partial_x \mathcal{L}(x^{k+1}, y^{k+1}) \end{aligned}$$

Control **outer** stationarity with **inner** stationarity.
Inexact subsolves with $\varepsilon_k \to \epsilon_d$ for the minimization of $\mathcal{L}_{\varrho_k}(\cdot, y^k)$.

ALM: Convergence Theory

Convex case [Hes69, Pow69, Roc76]:

- Global convergence to a KKT point (even with constant ϱ)
- Convergence rates as for PPA

Nonconvex case [CGT91, BM14, DMJKM23]:

- Convergence to first-order stationary points (under CQs)
- Local linear convergence near strict local minima
- Requires **adaptive** ϱ and **safeguarding** [KS17]

$\rightsquigarrow$ Adaptive ϱ: penalty strong enough.
$\rightsquigarrow$ Safeguarding: multiplier estimate not too crazy.

ALM with adaptive penalty

ALM scheme

Given x^0, y^0, $\rho_0 > 0$, $\kappa > 1$, $\theta \in (0,1)$
for $k = 0, 1, 2, \dots$ **do**
    $x^{k+1} \approx \arg\min_x \mathscr{L}_{\varrho_k}(x, y^k)$ *(proximal inner solver)*
    $y^{k+1} \leftarrow y^k + \varrho_k c(x^{k+1})$ *(dual ascent)*
    if $\|c(x^{k+1})\| \leq \max\{\epsilon_p, \theta\|c(x^{k})\|\}$ **then**
        $\varrho_{k+1} \leftarrow \varrho_k$
    else
        $\varrho_{k+1} \leftarrow \kappa\varrho_k$ *(stronger penalty)*
    end if
end for

- $\{\varrho_k\}$ bounded $\implies$ $c(x^k) \to 0$ (if $\epsilon_p = 0$). Feasible, yey!
- $\varrho_k \nearrow \infty$: since

$$x^{k+1} \in \arg\min_x \varphi(x) + \frac{\varrho_k}{2}\left\|c(x) + \frac{y^k}{\varrho_k}\right\|^2,$$

we **need also** $y^k/\varrho_k \to 0$ to obtain least-infeasible points. This is safeguarding.

ALM with safeguards

ALM iteration — Conn, Gould & Toint [CGT91]

1: $x^{k+1} \approx \arg\min_x \mathscr{L}_{\varrho_k}(x, y^k)$ *(proximal inner solver)*
2: **if** $\|c(x^{k+1})\| \leq \max\{\epsilon_p, \theta\|c(x^{k+1})\|\}$ **then**
3: $\quad y^{k+1} \leftarrow y^k + \varrho_k c(x^{k+1})$, $\varrho_{k+1} \leftarrow \varrho_k$ *(dual ascent)*
4: **else**
5: $\quad \varrho_{k+1} \leftarrow \kappa\varrho_k$, $y^{k+1} \leftarrow y^k$ *(stronger penalty)*
6: **end if**

ALM iteration — Birgin & Martinez [BM14]

1: $\hat{y}^k \leftarrow \mathrm{proj}_Y(y^k)$, given a compact set Y
2: $x^{k+1} \approx \arg\min_x \mathscr{L}_{\varrho_k}(x, \hat{y}^k)$ *(proximal inner solver)*
3: $y^{k+1} \leftarrow y^k + \varrho_k c(x^{k+1})$ *(dual ascent)*
4: **if** $\|c(x^{k+1})\| \leq \max\{\epsilon_p, \theta\|c(x^{k+1})\|\}$ **then**
5: $\quad \varrho_{k+1} \leftarrow \varrho_k$
6: **else**
7: $\quad \varrho_{k+1} \leftarrow \kappa\varrho_k$ *(stronger penalty)*
8: **end if**

Elastic safeguard [DMHM25]

Constrained Structured Programs [DMJKM23]

Problem (P)

$$\underset{x}{\text{minimize}} \quad \varphi(x) \quad \text{subject to} \quad c(x) \in \mathcal{D}$$

$\varphi := f + g$, f and c smooth, g nonsmooth, $\mathcal{D}$ closed simple set.

How can we use all the **ALM machinery for equality constraints**?

Just add some variables and split!

$$\begin{aligned} \underset{x,z}{\text{minimize}} \quad & \varphi(x) + \delta_{\mathcal{D}}(z) \\ \text{subject to} \quad & c(x) - z = 0 \end{aligned}$$

QP with geometric constraints: exploiting structure

Geometric / Structured Constraints

Many practical constraints have special **structure**:

- **Polyhedral**: boxes, simplices, linear equalities/inequalities
- **Conic**: second-order cone, PSD matrices
- **Nonconvex**: cardinality, rank, logical/combinatorial conditions
- **Geometric (set membership)**: $x \in \mathcal{C}$, $\mathcal{C}$ possibly nonconvex

Problem

$$\underset{x}{\text{minimize}} \quad \frac{1}{2}x^\top Qx + q^\top x \quad \text{subject to} \quad Ax \in \mathcal{C}$$

$Q \succeq 0$, $\mathcal{C}$ closed (possibly nonconvex), projection onto $\mathcal{C}$ available.

Examples for $\mathcal{C}$:

- Box + cardinality: $\{x : \|x\|_0 \leq s,\ l \leq x \leq u\}$ (sparse solutions)
- Union of sets (logical conditions, integer-like)
- Polytopes (control limits, allocation constraints)

Why this problem?

Convex quadratic programs (QPs) with polyhedral constraints are standard in signal processing, control, and decision-making.

Real applications need more: **Disjunctive constraints**

- *Logical conditions* — e.g. at most one of two actuators fires at a time
- *Cardinality / sparsity constraints* — e.g. portfolio selection
- *Obstacle avoidance* — feasible region is a union of halfspaces

These constraints are **nonconvex**, but projections onto them are tractable.

Constraint type	$\mathcal{C}$
Box / polyhedron	$\mathcal{C} = \{z : \ell \leq z \leq u\}$
Complementarity	$\mathcal{C} = \{z : z_1 \geq 0,\ z_2 \geq 0,\ z_1 z_2 = 0\}$
Either-or (EOC)	$\mathcal{C} = \{z : \lvert z_1 \rvert \leq u_{\max},\ \lvert z_2 \rvert \leq u_{\max},\ z_1 z_2 = 0\}$
Cardinality	$\mathcal{C} = \{z : \lVert z \rVert_0 \leq s\}$

The Problem Class

Geometric QP

$$\underset{x \in \mathbb{R}^n}{\text{minimize}} \quad \frac{1}{2} x^\top Q x + q^\top x \qquad \text{subject to} \quad Ax \in \mathcal{C} \tag{P}$$

- $Q \succeq 0$, $A \in \mathbb{R}^{m \times n}$, closed (possibly nonconvex) set $\mathcal{C} \subset \mathbb{R}^m$.
- Only requirement on $\mathcal{C}$: a **projection oracle** $\text{proj}_{\mathcal{C}}$ is available.

Question: Can we exploit quadratic cost and projections onto $\mathcal{C}$ efficiently?

Strategy: augmented Lagrangian and structure-exploitation with

- subproblem reformulation via condensing [ZAB+19, DM25]
- splitting and subproblem solver [MBJ+25]

Reduce computational cost significantly while preserving ALM convergence guarantees.

Augmented Lagrangian (AL) Approach

Introduce slack $z \in \mathcal{C}$, split (P) as:

$$\underset{x,\,z}{\text{minimize}} \quad f(x) + \delta_{\mathcal{C}}(z) \qquad \text{subject to} \quad Ax - z = 0$$

The (proximal) **augmented Lagrangian** function

$$\mathscr{L}_{\varrho}(x, z; \hat{x}, \hat{y}) := f(x) + \delta_{\mathcal{C}}(z) + \hat{y}^\top (Ax - z) + \frac{\varrho}{2} \|Ax - z\|^2 + \frac{\sigma}{2} \|x - \hat{x}\|^2$$

Algorithm: Safeguarded ALM

Repeat until $\|Ax_k - z_k\|_\infty \le \epsilon_p$ and stationarity $\varepsilon_k \le \epsilon_d$:

1. **Solve** $(x_k, z_k) \approx \arg\min_{x,\,z} \mathscr{L}_{\varrho_k}(x, z, \hat{x}_k, \hat{y}_k)$
2. **Update** dual $y_{k+1} \leftarrow \hat{y}_k + \varrho_k (Ax_k - z_k)$
3. Adapt parameters $\sigma_{k+1}, \varrho_{k+1}$ and tolerance ε_{k+1}

Guarantees: feasible accumulation points are KKT points; infeasible ones are stationary for the feasibility measure.

Subproblem Structure and Condensing

$$\mathscr{L}_{\varrho}(x, z; \hat{x}, \hat{y}) := f(x) + \delta_{\mathcal{C}}(z) + \hat{y}^{\top}(Ax - z) + \frac{\varrho}{2}\|Ax - z\|^2 + \frac{\sigma}{2}\|x - \hat{x}\|^2$$

For fixed $z \in \mathcal{C}$, the subproblem is *unconstrained* and *strictly convex* in x:

$$\begin{aligned} X_k(z) &:= \arg\min_x \mathscr{L}_{\varrho_k}(x, z; \hat{x}_k, \hat{y}_k) \\ &= \arg\min_x \left\{ f(x) + \hat{y}_k^{\top}(Ax - z) + \frac{\varrho_k}{2}\|Ax - z\|^2 + \frac{\sigma_k}{2}\|x - \hat{x}_k\|^2 \right\} \end{aligned}$$

$$0 = Qx + q + A^{\top}\hat{y}_k + \varrho_k A^{\top}(Ax - z) + \sigma_k(x - \hat{x}_k)$$

Solved by a **single linear system** (matrix independent of z, factorize once):

$$\underbrace{(Q + \sigma_k \mathbb{I} + \varrho_k A^{\top} A)}_{\succ 0,\ \text{const. w.r.t. } z} x = \sigma_k \hat{x}_k - q + A^{\top}(\varrho_k z - \hat{y}_k)$$

The **condensed subproblem** is then $\text{minimize}_{z \in \mathcal{C}} \quad V_k(z)$.
The **marginal function** $V_k := z \mapsto [\mathscr{L}_{\varrho_k}(X_k(z), z; \hat{x}_k, \hat{y}_k) - \delta_{\mathcal{C}}(z)]$ is smooth!

What Condensing Achieves

Dimension reduction

Variables: from $\mathbb{R}^{n+m}$ to $\mathbb{R}^m$

Built-in preconditioning

Construction preconditions the subproblem automatically

Efficient gradient

$\nabla V_k(z)$ is a cheap backsolve away

Solver-agnostic

Any proximal method applies to minimize$_{z \in \mathcal{C}} V_k(z)$

Hard equality constraints

Lower-level $A_{\text{eq}} x = b_{\text{eq}}$ kept *exactly* in condensed step via KKT system:

$$\begin{pmatrix} Q + \sigma_k \mathbb{I} & A^\top & A_{\text{eq}}^\top \\ A & -\varrho_k \mathbb{I} & 0 \\ A_{\text{eq}} & 0 & 0 \end{pmatrix} \begin{pmatrix} x \\ \lambda \\ \lambda_{\text{eq}} \end{pmatrix} = \begin{pmatrix} \dots \\ \dots \\ b_{\text{eq}} \end{pmatrix}$$

Retains sparsity and symmetry; avoids ill-conditioning from penalizing well-structured constraints

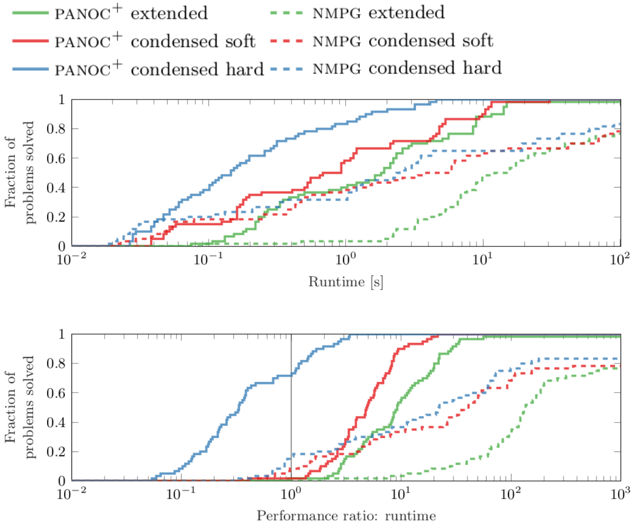
Numerical results

Obstacle benchmark: Discretized 1-D Laplace with obstacle set:

$$\begin{aligned} \underset{x_x, x_y, x_z \in \mathbb{R}^N}{\text{minimize}} \quad & \frac{1}{2}\|x_x\|^2 + \frac{1}{2}\|x_y\|^2 - 1^\top x_y \\ \text{subject to} \quad & x_x \geq 0, \quad 0 \leq x_y \perp x_z \geq 0 \\ & x_x + A x_y = x_z. \end{aligned}$$

Degenerate solution $x_x = x_y = x_z = 0$.
Six solver configurations: NMPG and PANOC$^+$, extended and condensed, soft and hard. $\varepsilon_p = \varepsilon_d = 10^{-6}$; 100 s time limit

- Scales to $N = 256$; [JKMW23] caps at $N = 64$.
- Iters reduced from *millions* to *thousands* using `panoc+`.
- Degeneracy handled robustly.



- condensed-hard ≻ condensed-soft ≻ extended.

More variables, more splitting: divide and conquer

Geometric QP

$$\underset{x}{\text{minimize}} \quad \frac{1}{2}x^\top Q x + q^\top x \quad \text{subject to} \quad x \in \mathcal{X}, \quad Ax \in \mathcal{C} \tag{P}$$

Let's add $z \in \mathcal{C}$

$$\underset{x,z}{\text{minimize}} \quad f(x) + \delta_{\mathcal{X}}(x) + \delta_{\mathcal{C}}(z) \quad \text{subject to} \quad Ax = z$$

and then $(\tilde{x}, \tilde{z}) \in \mathcal{E} := \{(x, z) \,|\, Ax = z\}$

$$\underset{x,z,\tilde{x},\tilde{z}}{\text{minimize}} \quad f(\tilde{x}) + \delta_{\mathcal{E}}(\tilde{x}, \tilde{z}) + \delta_{\mathcal{X}}(x) + \delta_{\mathcal{C}}(z) \quad \text{subject to} \quad x = \tilde{x}, \quad z = \tilde{z}$$

Menager et al.'s "Contact-Implicit Inverse Dynamics" (2025)

What now? ALM

$$\underset{x,z,\tilde{x},\tilde{z}}{\text{minimize}} \quad f(\tilde{x}) + \delta_{\mathcal{E}}(\tilde{x},\tilde{z}) + \delta_{\mathcal{X}}(x) + \delta_{\mathcal{C}}(z) \quad \text{subject to} \quad x = \tilde{x}, \quad z = \tilde{z}$$

Augmented Lagrangian function

$$\begin{aligned}\mathscr{L}_{\varrho}(x,z,\tilde{x},\tilde{z},y_x,y_z) := f(\tilde{x}) + \delta_{\mathcal{E}}(\tilde{x},\tilde{z}) + \delta_{\mathcal{X}}(x) + \delta_{\mathcal{C}}(z) \\ + y_x^\top(x-\tilde{x}) + y_z^\top(z-\tilde{z}) + \frac{\varrho}{2}\|x-\tilde{x}\|^2 + \frac{\varrho}{2}\|z-\tilde{z}\|^2\end{aligned}$$

AL subproblem

$$\underset{x,z,\tilde{x},\tilde{z}}{\text{minimize}} \quad \mathscr{L}_{\varrho}(x,z,\tilde{x},\tilde{z},\hat{y}_x,\hat{y}_z)$$

Alternating subsolver: (x, z)

AL subproblem

$$\underset{x,z,\tilde{x},\tilde{z}}{\text{minimize}} \quad \mathscr{L}_\varrho(x, z, \tilde{x}, \tilde{z}, \hat{y}_x, \hat{y}_z)$$

with respect to (x, z) becomes

$$\begin{aligned}
&\underset{x,z}{\arg\min} \quad y_x^\top(x-\tilde{x})+y_z^\top(z-\tilde{z})+\frac{\varrho}{2}\|x-\tilde{x}\|^2+\frac{\varrho}{2}\|z-\tilde{z}\|^2 \quad \text{subject to} \quad x \in \mathcal{X}, \quad z \in \mathcal{C} \\
\iff \quad &\underset{x,z}{\arg\min} \quad \frac{\varrho}{2}\|x-\tilde{x}+y_x/\varrho\|^2+\frac{\varrho}{2}\|z-\tilde{z}+y_z/\varrho\|^2 \quad \text{subject to} \quad x \in \mathcal{X}, \quad z \in \mathcal{C} \\
\iff \quad & x \in \text{proj}_{\mathcal{X}}(\tilde{x} - y_x/\varrho), \qquad z \in \text{proj}_{\mathcal{C}}(\tilde{z} - y_z/\varrho)
\end{aligned}$$

Alternating subsolver: $(\tilde{x}, \tilde{z})$

AL subproblem

$$\underset{x,z,\tilde{x},\tilde{z}}{\text{minimize}} \quad \mathscr{L}_\varrho(x, z, \tilde{x}, \tilde{z}, \hat{y}_x, \hat{y}_z)$$

with respect to $(\tilde{x}, \tilde{z})$ becomes

$$\arg\min_{\tilde{x},\tilde{z}} \quad f(\tilde{x}) + y_x^\top (x - \tilde{x}) + y_z^\top (z - \tilde{z}) + \frac{\varrho}{2}\|x - \tilde{x}\|^2 + \frac{\varrho}{2}\|z - \tilde{z}\|^2 \quad \text{subject to} \quad A\tilde{x} = \tilde{z}$$

$$\iff \begin{cases} 0 = Q\tilde{x} + q - y_x + \varrho(\tilde{x} - x) + A^\top \lambda \\ 0 = -y_z + \varrho(\tilde{z} - z) - \lambda \\ A\tilde{x} = \tilde{z} \end{cases}$$

$$\iff \begin{bmatrix} Q + \varrho \mathbb{I} & \cdot & A^\top \\ \cdot & \varrho \mathbb{I} & -\mathbb{I} \\ A & -\mathbb{I} & \cdot \end{bmatrix} \begin{pmatrix} \tilde{x} \\ \tilde{z} \\ \lambda \end{pmatrix} = \begin{pmatrix} -q + y_x + \varrho x \\ y_z + \varrho z \\ 0 \end{pmatrix}$$

$$\iff \begin{bmatrix} Q + \varrho \mathbb{I} & A^\top \\ A & -\frac{1}{\varrho}\mathbb{I} \end{bmatrix} \begin{pmatrix} \tilde{x} \\ \lambda \end{pmatrix} = \begin{pmatrix} -q + y_x + \varrho x \\ z + \frac{1}{\varrho} y_z \end{pmatrix}$$

Alternating minimization subsolver

Instead of invoking PGM (with associated stepsizes), we can

$$\underset{x,z,\tilde{x},\tilde{z}}{\text{minimize}} \quad \mathscr{L}_{\varrho}(x, z, \tilde{x}, \tilde{z}, \hat{y}_x, \hat{y}_z)$$

by selecting an initial pair $(\tilde{x}, \tilde{z})$ and then repeating the steps

- $(x, z) \leftarrow \arg\min \mathscr{L}_{\varrho}(\cdot, \cdot, \tilde{x}, \tilde{z}, \hat{y}_x, \hat{y}_z)$,
- $(\tilde{x}, \tilde{z}) \leftarrow \arg\min \mathscr{L}_{\varrho}(x, z, \cdot, \cdot, \hat{y}_x, \hat{y}_z)$,

over and over again until convergence to a stationary point of $\mathscr{L}_{\varrho}(\cdot, \cdot, \cdot, \cdot, \hat{y}_x, \hat{y}_z)$.

Simple steps, no need to call an inner prox-grad solver!

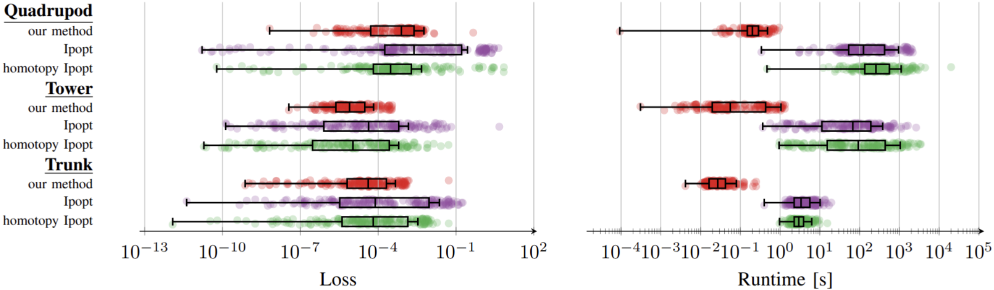
QPCC for robotics

$$\underset{x_a, x_c, x_\sigma}{\text{minimize}} \quad \frac{1}{2}\|Ax_a + Bx_c + b\|^2$$
tracking cost

$$\text{subject to} \quad x_\sigma = Cx_a + Dx_c + g$$
kinematics / dynamics

$$\mathcal{K}_\mu \ni x_c \perp x_\sigma + \Gamma(x_\sigma) \in \mathcal{K}_\mu^*$$
frictional contact

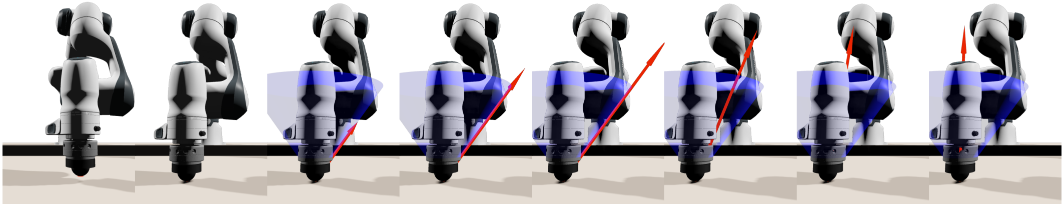
$(0 \le x_c \perp x_\sigma \ge 0)$



Comparison on NCP problems across robots, considering a nonlinear programming method (Ipopt), a regularized scheme with homotopy, and our method (ALM-based approach). [MBJ+25]

Frictional contact-implicit inverse dynamics

$$
\begin{aligned}
\underset{x_a, x_c, x_\sigma}{\text{minimize}} \quad & \frac{1}{2}\|Ax_a + Bx_c + b\|^2 \\
\text{subject to} \quad & x_\sigma = Cx_a + Dx_c + g \\
& \mathcal{K}_\mu \ni x_c \perp x_\sigma + \Gamma(x_\sigma) \in \mathcal{K}_\mu^*
\end{aligned}
$$



Justin Carpentier's team @ INRIA Paris

Franka sanding task: contact cone (blue) and ground reaction force (red arrow). The robot first makes contact with the ground, then starts sliding which makes the reaction force lie in the boundary of the friction cone. Then, the robot comes to a rest on the surface, and the contact force vector is now in the interior of the friction cone. [MBJ+25]

References I

Ernesto G. Birgin and José Mario Martínez.
Practical Augmented Lagrangian Methods for Constrained Optimization.
Society for Industrial and Applied Mathematics, Philadelphia, PA, 4 2014.

Jérôme Bolte, Shoham Sabach, and Marc Teboulle.
Nonconvex Lagrangian-based optimization: Monitoring schemes and global convergence.
Mathematics of Operations Research, 43(4):1210–1232, 2018.

Emilie Chouzenoux, Marie-Caroline Corbineau, and Jean-Christophe Pesquet.
A proximal interior point algorithm with applications to image processing.
Journal of Mathematical Imaging and Vision, 62(6):919–940, 2020.

Andrew R. Conn, Nicholas I. M. Gould, and Philippe L. Toint.
A globally convergent augmented Lagrangian algorithm for optimization with general constraints and simple bounds.
SIAM Journal on Numerical Analysis, 28(2):545–572, 4 1991.

Eyal Cohen, Nadav Hallak, and Marc Teboulle.
A dynamic alternating direction of multipliers for nonconvex minimization with nonlinear functional equality constraints.
Journal of Optimization Theory and Applications, 193(1):324–353, 2022.

References II

Eyal Cohen and Marc Teboulle.
Alternating and parallel proximal gradient methods for nonsmooth, nonconvex minimax: A unified convergence analysis.
Mathematics of Operations Research, 50(1):141–168, 2025.

Youssef Diouane, Maxence Gollier, and Dominique Orban.
Nonsmooth exact penalty methods for equality-constrained optimization: complexity and implementation.
SIAM Journal on Optimization, 2026.

Alberto De Marchi.
Implicit augmented Lagrangian and generalized optimization.
Journal of Applied and Numerical Optimization, 6(2):291–320, 2024.

Alberto De Marchi.
A condensing approach for linear-quadratic optimization with geometric constraints, 2025.

Alberto De Marchi, Tim Hoheisel, and Patrick Mehlitz.
Augmented Lagrangian methods for fully convex composite optimization, 2025.

Alberto De Marchi, Xiaoxi Jia, Christian Kanzow, and Patrick Mehlitz.
Constrained composite optimization and augmented Lagrangian methods.
Mathematical Programming, 201(1):863–896, 9 2023.

References III

Alberto De Marchi and Patrick Mehlitz.
Local properties and augmented Lagrangians in fully nonconvex composite optimization.
Journal of Nonsmooth Analysis and Optimization, 5, 2024.

Alberto De Marchi and Andreas Themelis.
An interior proximal gradient method for nonconvex optimization.
Open Journal of Mathematical Optimization, 5:1–22, 2024.

Alberto De Marchi and Andreas Themelis.
A penalty barrier framework for nonconvex constrained optimization.
Journal of Nonsmooth Analysis and Optimization, 5(14585), 2025.

Ron Estrin, Michael P. Friedlander, Dominique Orban, and Michael A. Saunders.
Implementing a smooth exact penalty function for general constrained nonlinear optimization.
SIAM J. Sci. Comput., 42(3):1836–1859, 2020.

Magnus R. Hestenes.
Multiplier and gradient methods.
Journal of Optimization Theory and Applications, 4(5):303–320, 11 1969.

Nadav Hallak and Marc Teboulle.
An adaptive Lagrangian-based scheme for nonconvex composite optimization.
Mathematics of Operations Research, 48(4):2337–2352, 2023.

References IV

Xiaoxi Jia, Christian Kanzow, Patrick Mehlitz, and Gerd Wachsmuth.
An augmented Lagrangian method for optimization problems with structured geometric constraints.
Mathematical Programming, 199(1):1365–1415, 2023.

Christian Kanzow and Daniel Steck.
An example comparing the standard and safeguarded augmented Lagrangian methods.
Operations Research Letters, 45(6):598–603, 2017.

Etienne Ménager, Antoine Bambade, Wilson Jallet, Alberto de Marchi, and Justin Carpentier.
Contact-implicit inverse dynamics, August 2025.
working paper or preprint.

Michael J. D. Powell.
A method for nonlinear constraints in minimization problems, pages 283–298.
Academic Press, 1969.

Ralph Tyrrell Rockafellar.
Augmented Lagrangians and applications of the proximal point algorithm in convex programming.
Mathematics of operations research, 1(2):97–116, 5 1976.

Peng Zheng, Travis Askham, Steven L. Brunton, J. Nathan Kutz, and Aleksandr Y. Aravkin.
A unified framework for sparse relaxed regularized regression: SR3.
IEEE Access, 7:1404–1423, 2019.