Constrained Structured Optimization and Augmented Lagrangian Proximal Methods

Alberto De Marchi
aldma.github.io

Universität der Bundeswehr München

MOCAO: Mathematics of Computation and Optimisation

VA & OPT Webinar
May 25, 2022

Preprint on arXiv

Constrained Structured Optimization and Augmented Lagrangian Proximal Methods

Alberto De Marchi* Xiaoxi Jia† Christian Kanzow†
Patrick Mehlitz‡

Abstract

We investigate and develop numerical methods for finite-dimensional constrained structured optimization problems. Offering a comprehensive yet simple and expressive language, this problem class provides a modeling framework for a variety of applications. A general and flexible algorithm is proposed that interlaces proximal methods and safeguarded augmented Lagrangian schemes. We provide a theoretical characterization of the algorithm and its asymptotic properties, deriving convergence results for fully nonconvex problems. Adopting a proximal gradient method with an oracle as a formal tool, it is demonstrated how the inner subproblems can be solved by off-the-shelf methods for composite optimization, without introducing slack variables and despite the appearance of set-valued projections. Finally, we describe our open-source matrix-free implementation of the proposed algorithm and test it numerically. Illustrative examples show the versatility of constrained structured programs as a modeling tool, expose difficulties arising in this vast problem class and highlight benefits of the implicit approach developed.

arXiv:2203.05276 [math.OC] 5 Apr 2022

Outline

- Introduction
- Proximal Gradient Method
- Constrained Structured Programs
- **A**ugmented **L**agrangian **P**roximal **S**olver
- Numerical Tests

Introduction

Problem class: expressive and flexible

Constrained Structured Programs

$$\begin{aligned} \text{minimize} \quad & f(x) + g(x) \\ \text{over} \quad & x \in \mathbb{R}^n \\ \text{subject to} \quad & c(x) \in D \end{aligned}$$

Assumptions:

- $f : \mathbb{R}^n \to \mathbb{R}$ and $c : \mathbb{R}^n \to \mathbb{R}^m$ are smooth
- $g : \mathbb{R}^n \to \overline{\mathbb{R}} := \mathbb{R} \cup \{\infty\}$ is proper and lsc
- $D \subset \mathbb{R}^m$ is nonempty and closed

Motivations: nonsmooth structures

Optimal control with switching costs

$$\begin{aligned} \text{minimize} \quad & \int_0^T \ell(t, x(t))\mathrm{d}t + \mathrm{TV}(u) \\ \text{subject to} \quad & \dot{x}(t) = f(t, x(t), u(t)) \\ & b(x(0), x(T)) = 0 \\ & u(t) \in \{0, 1\} \end{aligned}$$

$\rightsquigarrow$ nonsmooth regularization

Vanishing constraints

$$\begin{aligned} \text{minimize} \quad & f(x, y) \\ \text{subject to} \quad & x \geq 0, \quad x > 0 \;\Rightarrow\; y \geq 0 \\ & c(x) \geq 0 \end{aligned}$$

$\rightsquigarrow$ disjunctive constraints: projection and distance

More examples

Nonsmooth Rosenbrock and either-or constraint:

$$\begin{aligned} \text{minimize} \quad & f(x) + |x_1| \\ \text{over} \quad & x \in \mathbb{R}^2 \\ \text{subject to} \quad & x_2 \le -x_1 \;\vee\; x_2 \ge x_1 \end{aligned}$$

$\rightsquigarrow$ nonsmooth cost, nonconvex constraint set

Sparse portfolio optimization:

$$\begin{aligned} \text{minimize} \quad & f(x) + \|x\|_0 \\ \text{subject to} \quad & \mu^\top x \ge \varrho, \quad \|x\|_1 = 1, \quad 0 \le x \le u \end{aligned}$$

$\rightsquigarrow$ discontinuous cost, MIQP reformulation

Constrained Structured Optimization

$$\begin{array}{rl} \text{minimize} & f(x) + g(x) \\ \text{over} & x \in \mathbb{R}^n \\ \text{subject to} & c(x) \in D \end{array}$$

aka Composite Optimization

Assumptions:

- $f : \mathbb{R}^n \to \mathbb{R}$ is smooth
- $g : \mathbb{R}^n \to \overline{\mathbb{R}}$ is proper and lsc

Oracles:

- $x \mapsto f(x)$ and $x \mapsto \nabla f(x)$
- $x, \gamma \mapsto \bar{x} \in \text{prox}_{\gamma g}(x), g(\bar{x})$

Proximal Mapping

$\mathrm{prox}_{\gamma g} : \mathbb{R}^n \rightrightarrows \mathbb{R}^n, \gamma > 0,$

$$\mathrm{prox}_{\gamma g}(x) := \operatorname*{arg\,min}_{z \in \mathbb{R}^n} \left\{ g(z) + \frac{1}{2\gamma} \|z - x\|^2 \right\}$$

Basic properties

- stepsize γ balances minimization of g and distance from x
- often has closed-form solution
- generalizes projection

See proximity-operator.net (MATLAB, Python) and JuliaFirstOrder/ProximalOperators.jl (Julia).

Proximal mapping and Projections

Let $D \subset \mathbb{R}^m$ be a nonempty closed set.

Indicator function $\delta_D : \mathbb{R}^m \to \overline{\mathbb{R}}$

$$\delta_D(v) := \begin{cases} 0 & \text{if } v \in D \\ \infty & \text{otherwise} \end{cases}$$

Projection mapping $\operatorname{proj}_D : \mathbb{R}^m \rightrightarrows \mathbb{R}^m$

$$\operatorname{proj}_D(v) := \arg\min_{z \in D} \|z - v\| = \operatorname{prox}_{\gamma\,\delta_D}(v)$$

Composite Optimization
and
Proximal Gradient Method

Composite Optimization: Necessary Conditions

Problem

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

Minimizer $x^\star \in \arg\min q$.

Stationarity condition ($\Leftarrow$ Optimality)

$$0 \in \partial q(x^\star) \qquad \Leftrightarrow \qquad \operatorname{dist}(\partial q(x^\star), 0) = 0$$

Approximate stationarity, $\varepsilon \geq 0$,

$$\operatorname{dist}(\partial q(x^\star), 0) \leq \varepsilon$$

Composite Optimization: Necessary Conditions

Problem

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

Minimizer $x^\star \in \arg\min q$.

Stationarity condition

$$0 \in \partial q(x^\star)$$

Let $\mathrm{T}_\gamma : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$

$$\mathrm{T}_\gamma(x) := \mathrm{prox}_{\gamma g}(x - \gamma \nabla f(x))$$

combine gradient and proximal steps.

Criticality condition ($\Rightarrow$ Stationarity)

$$\exists \gamma > 0 : \quad x^\star \in \mathrm{T}_\gamma(x^\star)$$

[1] A. Themelis, L. Stella and P. Patrinos. Forward-Backward Envelope for the Sum of Two Nonconvex Functions. SIOPT 28(3), 2018.

Some notation and definitions

Let $h : \mathbb{R}^n \to \overline{\mathbb{R}}$ be a proper and lsc function and $\bar{x} \in \mathbb{R}^n$ a point with $h(\bar{x})$ finite.

$\hat{\partial} h : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ denotes the *regular (Fréchet) subdifferential* of h:

$$v \in \hat{\partial} h(\bar{x}) \quad \overset{def}{\Leftrightarrow} \quad \liminf_{\substack{x \to \bar{x} \\ x \neq \bar{x}}} \frac{h(x) - h(\bar{x}) - \langle v, x - \bar{x} \rangle}{\|x - \bar{x}\|} \geq 0$$

$\partial h : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ denotes the *(limiting) subdifferential* of h:

$$v \in \partial h(\bar{x}) \quad \overset{def}{\Leftrightarrow} \quad \exists \{(x^k, v^k)\} : x^k \overset{h}{\to} \bar{x},\ v^k \in \hat{\partial} h(x^k),\ v^k \to v$$

h-attentive convergence of a sequence $\{x^k\}$:

$$x^k \overset{h}{\to} \bar{x} \quad \overset{def}{\Leftrightarrow} \quad x^k \to \bar{x} \quad \text{with} \quad h(x^k) \to h(\bar{x})$$

[2] R.T. Rockafellar and R.J.B. Wets, Variational Analysis, volume 317, Springer, 1998.

Proximal Gradient Method aka Forward-Backward Splitting

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

Iterative method:

$$x^{k+1} \in \mathrm{T}_\gamma(x^k)$$

where $\mathrm{T}_\gamma(x) := \mathrm{prox}_{\gamma g}(x - \gamma \nabla f(x))$.

Descent lemma: if $\bar{x} \in \mathrm{T}_\gamma(x)$,

$$q(\bar{x}) \leq q(x) - \frac{1 - \gamma L_f}{2\gamma} \|\bar{x} - x\|^2$$

$\rightsquigarrow$ adaptive variant needed: $\gamma \in (0, 1/L_f)$

[3] Proximal Algorithms. N. Parikh and S. Boyd. Foundations and Trends in Optimization, 1(3):127–239, 2014.

Adaptive Proximal Gradient Method

select $x^0 \in \mathbb{R}^n$, $\gamma_0 > 0$, $\alpha \in (0, 1)$
for $k = 0, 1, 2, \ldots$ do
while true **do**
  select $\bar{x}^k \in \mathrm{T}_{\gamma_k}(x^k)$
  if $f(\bar{x}^k) > f(x^k) + \langle \nabla f(x^k), \bar{x}^k - x^k \rangle + \frac{\alpha}{2\gamma_k} \|\bar{x}^k - x^k\|^2$ **then**
    set $\gamma_k \leftarrow \gamma_k / 2$
  else
    break
  end if
end while
set $\gamma_{k+1} \leftarrow \gamma_k$ and $x^{k+1} \leftarrow \bar{x}^k$

Accelerated Proximal Gradient Method

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

$$\text{PGM:} \quad x^{k+1} \in \mathrm{T}_{\gamma_k}(x^k)$$

$\rightsquigarrow$ cheap iterations, slow convergence
✓ nonconvex problems supported

$\rightsquigarrow$ Nesterov acceleration (e.g. FISTA)
. . . usually needs some convexity

$\rightsquigarrow$ Anderson acceleration

Accelerated Proximal Gradient Method: PANOC

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

$$\text{PGM:} \quad x^{k+1} \in \mathrm{T}_{\gamma_k}(x^k)$$

$$0 \in \mathrm{R}_\gamma(x) \quad \Leftrightarrow \quad x \in \mathrm{T}_\gamma(x) \quad \Rightarrow \quad 0 \in \partial q(x)$$

PANOC$^+$
$\rightsquigarrow$ limited-memory quasi-Newton acceleration
$\rightsquigarrow$ globalization
$\rightsquigarrow$ cheap iterations, fast convergence
✓ nonconvex problems
$\rightsquigarrow$ supports ∇f locally Lipschitz continuous

[4] L. Stella, A. Themelis, P. Sopasakis and P. Patrinos. A simple and efficient algorithm for nonlinear model predictive control, IEEE CDC, 2017.
[5] A. D.M. and A. Themelis. Proximal gradient algorithms under local Lipschitz gradient continuity, 2021. arXiv:2112.13000

Constrained Structured Programs

Constrained Structured Programs

$$\begin{aligned} \text{minimize} \quad & q(x) := f(x) + g(x) \\ \text{over} \quad & x \in \mathbb{R}^n \\ \text{subject to} \quad & c(x) \in D \end{aligned}$$

Assumptions:

- $f : \mathbb{R}^n \to \mathbb{R}$ and $c : \mathbb{R}^n \to \mathbb{R}^m$ are smooth
- $g : \mathbb{R}^n \to \overline{\mathbb{R}} := \mathbb{R} \cup \{\infty\}$ is proper and lsc
- $D \subset \mathbb{R}^m$ is nonempty and closed

Oracles:

- $x \mapsto f(x)$ and $x \mapsto \nabla f(x)$
- $x, \gamma \mapsto \bar{x} \in \text{prox}_{\gamma g}(x), g(\bar{x})$
- $x \mapsto c(x)$ and $x, v \mapsto \nabla c(x)^\top v$
- $v \mapsto z \in \text{proj}_D(v)$

Feasibility and Stationarity Concepts

A point $x^\star \in \mathbb{R}^n$ is called *feasible* if $x^\star \in \operatorname{dom} g$ and $c(x^\star) \in D$.

A feasible point $x^\star$ is called *M-stationary* if there exists $y^\star \in \mathbb{R}^m$ such that

$$\begin{aligned} 0 &\in \partial q(x^\star) + \nabla c(x^\star)^\top y^\star \\ y^\star &\in \mathcal{N}_D^{\lim}(c(x^\star)) \end{aligned}$$

Asymptotic Stationarity Concept

A feasible point $x^\star$ is called *AM-stationary* if there exist sequences $\{x^k\}, \{\eta^k\} \subset \mathbb{R}^n$ and $\{y^k\}, \{\nu^k\}, \{\zeta^k\} \subset \mathbb{R}^m$ such that

$$
\begin{aligned}
& x^k \xrightarrow{q} x^\star \\
& \eta^k \to 0, \quad \nu^k \to 0, \quad \zeta^k \to 0 \\
& \eta^k \in \partial q(x^k) + \nabla c(x^k)^\top y^k \\
& \nu^k \in \mathcal{N}_D^{\lim}(c(x^k) + \zeta^k) - y^k
\end{aligned}
$$

$\rightsquigarrow$ attentive convergence
$\rightsquigarrow$ tolerances η^k, ν^k, ζ^k

Theorem

Let $x^\star$ be a local minimizer. Then, $x^\star$ is an AM-stationary point.

Augmented **L**agrangian **P**roximal **S**olver

Back to our Comfort Zone

Lifted reformulation

$$\begin{aligned} \text{minimize} \quad & q(x) + \delta_D(s) \\ \text{over} \quad & x \in \mathbb{R}^n,\ s \in \mathbb{R}^m \\ \text{subject to} \quad & c(x) - s = 0 \end{aligned}$$

Augmented Lagrangian function, $\mu > 0$, $y \in \mathbb{R}^m$,

$$\begin{aligned} \mathcal{L}_\mu^{\mathrm{S}}(x, s, y) &:= q(x) + \delta_D(s) + \langle y, c(x) - s \rangle + \frac{1}{2\mu}\|c(x) - s\|^2 \\ &= q(x) + \delta_D(s) + \frac{1}{2\mu}\|c(x) + \mu y - s\|^2 - \frac{\mu}{2}\|y\|^2 \end{aligned}$$

AL subproblem, with slack variables

$$\begin{aligned} &\underset{x,s}{\text{minimize}} \quad \mathcal{L}_\mu^{\mathrm{S}}(x, s, \hat{y}) \\ \Leftrightarrow \quad &\underset{x,s}{\text{minimize}} \quad \underbrace{f(x) + \frac{1}{2\mu}\|c(x) - s + \mu\hat{y}\|^2}_{\tilde{f}(\tilde{x})} + \underbrace{g(x) + \delta_D(s)}_{\tilde{g}(\tilde{x})} \end{aligned}$$

⇝ composite optimization subproblem
⇝ adaptive (accelerated) proximal gradient

$$\begin{aligned} \tilde{x} &:= (x, s) \\ \nabla \tilde{f}(\tilde{x}) &= \begin{pmatrix} \nabla f(x) + \frac{1}{\mu}\nabla c(x)^\top [c(x) - s + \mu\hat{y}] \\ -\frac{1}{\mu}[c(x) - s + \mu\hat{y}] \end{pmatrix} \\ \operatorname{prox}_{\gamma\tilde{g}}(\tilde{x}) &= \begin{pmatrix} \operatorname{prox}_{\gamma g}(x) \\ \operatorname{proj}_D(s) \end{pmatrix} \end{aligned}$$

Safeguarded AL outer loop

select $\mu_0 > 0$, $\theta, \kappa \in (0,1)$ and $Y \subset \mathbb{R}^m$ bounded
for $k = 0, 1, 2, \ldots$ **do**
  select $\hat{y}^k \in Y$ from y^{k-1} and $\varepsilon_k \geq 0$ such that $\varepsilon_k \to 0$
  find (x^k, s^k) an ε_k-stationary point of $\mathcal{L}^{\mathrm{S}}_{\mu_k}(\cdot, \cdot, \hat{y}^k)$ from x^{k-1}
  set $\zeta^k \leftarrow s^k - c(x^k)$ and $y^k \leftarrow \hat{y}^k - \zeta^k / \mu_k$
  if $k = 0$ or $\|\zeta^k\| \leq \theta \|\zeta^{k-1}\|$ or $\|\zeta^k\| \leq \epsilon$ **then**
    set $\mu_{k+1} \leftarrow \mu_k$
  else
    select $\mu_{k+1} \in (0, \kappa\mu_k]$
  end if
end for

- safeguarded dual estimate $\hat{y}$
- primal and dual warm-starting
- inexact sub-solutions and relaxed update criteria

AL subproblem as a regularization

$$
\begin{aligned}
& \underset{x,s}{\text{minimize}} \quad \mathcal{L}^{\mathrm{S}}_{\mu}(x, s, \hat{y}) \\
\Leftrightarrow \quad & \underset{x,s}{\text{minimize}} \quad q(x) + \frac{1}{2\mu}\|c(x) - s + \mu\hat{y}\|^2 + \delta_D(s) \\
\Leftrightarrow \quad & \underset{x,s,\zeta}{\text{minimize}} \quad q(x) + \frac{1}{2\mu}\|\zeta - \mu\hat{y}\|^2 + \delta_D(s) \\
& \text{subject to} \quad s = c(x) + \zeta \\
\Leftrightarrow \quad & \underset{x,\zeta}{\text{minimize}} \quad q(x) + \frac{1}{2\mu}\|\zeta - \mu\hat{y}\|^2 \\
& \text{subject to} \quad c(x) + \zeta \in D
\end{aligned}
$$

AL subproblem: approximate stationarity

$$\begin{aligned} &\underset{x,s}{\text{minimize}} \quad \mathcal{L}^{\mathrm{S}}_{\mu}(x, s, \hat{y}) \\ \Leftrightarrow \quad &\underset{x,s}{\text{minimize}} \quad q(x) + \frac{1}{2\mu}\|c(x) - s + \mu\hat{y}\|^2 + \delta_D(s) \end{aligned}$$

ε-stationarity of (x, s):

$$\operatorname{dist}(\partial_{(x,s)}\mathcal{L}^{\mathrm{S}}_{\mu}(x, s, \hat{y}), 0) \leq \varepsilon$$

dual estimate y:

$$0 \in \partial_x \mathcal{L}^{\mathrm{S}}_{\mu}(x, s, \hat{y}) = \partial q(x) + \nabla c(x)^\top \underbrace{\left[\hat{y} + \frac{c(x) - s}{\mu}\right]}_{y}$$

Convergence Analysis

Lagrange multipliers vanish for inactive constraints

Let $x^\star$ be an accumulation point of $\{x^k\}$ and $\{x^k\}_K$ subsequence such that $x^k \to_K x^\star$. If $c(x^\star) \in \operatorname{int} D$, then $y^k = 0$ for all $k \in K$ large enough.

$\leadsto$ Easily refined by exploiting the separable structure of D, if any.

Sufficient conditions for feasibility

Each accumulation point $x^\star$ of $\{x^k\}$ is feasible if one of the following conditions is satisfied:

- $\{\mu_k\}$ is bounded away from zero, or
- $\{\mathcal{L}^{S}_{\mu_k}(x^k, s^k, \hat{y}^k)\}$ is bounded from above.

Accumulation points are good candidates

Consider a sequence of iterates generated by ALPS with $\varepsilon_k \to 0$. Let $x^\star$ be an accumulation point of $\{x^k\}$ and $\{x^k\}_K$ subsequence such that $x^k \overset{q}{\to}_K x^\star$. Then, $x^\star$ is a AM-stationary.

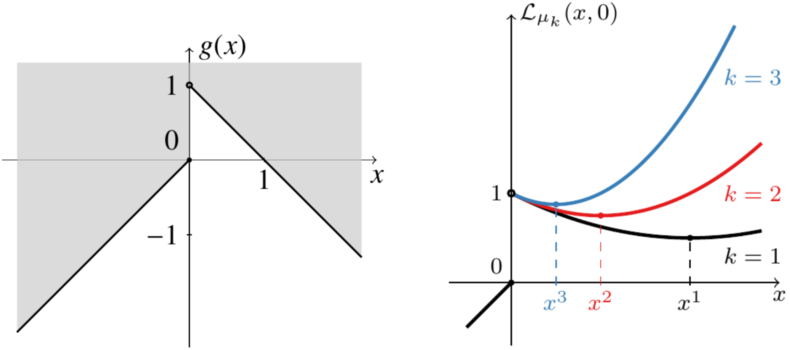
$\rightsquigarrow$ attentive convergence

Non-attentive convergence? An example

Consider

$$\underset{x\in\mathbb{R}}{\text{minimize}} \quad g(x) \quad \text{subject to} \quad x \leq 0$$

where $g(x) := x$ if $x \leq 0$, $g(x) := 1 - x$ otherwise.



Although $x^\star := 0$ is the global maximizer, it is not AM-stationary.

Applying ALPS, we have $x^k \to x^\star$ but not $x^k \overset{q}{\to} x^\star$.

Termination criteria for ALPS

AM-stationarity

$$x^k \xrightarrow{q} x^\star$$
$$\eta^k \to 0, \quad \nu^k \to 0, \quad \zeta^k \to 0$$
$$\eta^k \in \partial q(x^k) + \nabla c(x^k)^\top y^k$$
$$\nu^k \in \mathcal{N}_D^{\lim}(c(x^k) + \zeta^k) - y^k$$

One may stop with (x^k, s^k, y^k) if

$$\operatorname{dist}(\partial q(x^k) + \nabla c(x^k)^\top y^k, 0) \le \varepsilon_k \le \epsilon$$

and

$$\|c(x^k) - s^k\| = \|\zeta^k\| \le \epsilon$$

But...

$$y^k \in \mathcal{N}_D^{\lim}(s^k)?$$

Proximal Gradient with an Oracle

AL subproblem: easy wrt slack variables!

$$\begin{aligned} &\underset{x,s}{\text{minimize}} \quad \mathcal{L}_\mu^{\mathrm{S}}(x, s, \hat{y}) \\ \Leftrightarrow \quad &\underset{x,s}{\text{minimize}} \quad q(x) + \frac{1}{2\mu}\|c(x) - s + \mu\hat{y}\|^2 + \delta_D(s) \end{aligned}$$

$$\begin{aligned} \mathcal{S}_\mu(x, \hat{y}) &:= \arg\min_s \mathcal{L}_\mu^{\mathrm{S}}(x, s, \hat{y}) \\ &= \arg\min_s \|c(x) - s + \mu\hat{y}\|^2 + \delta_D(s) \\ &= \operatorname{proj}_D(c(x) + \mu\hat{y}) \end{aligned}$$

explicit, exact minimization:

$$s \in \mathcal{S}_\mu(x, \hat{y}) \quad \Rightarrow \quad 0 \in \partial_s \mathcal{L}_\mu^{\mathrm{S}}(x, s, \hat{y})$$

Slack variables and complementarity

$$y \in \mathcal{N}_D^{\lim}(s)? \quad \textit{yes}$$

$$s \in \mathcal{S}_\mu(x, \hat{y}) = \mathrm{proj}_D(c(x) + \mu\hat{y})$$

Observe that, for all $v, z \in \mathbb{R}^m$:

$$z \in \mathrm{proj}_D(v) \quad \Rightarrow \quad v - z \in \mathcal{N}_D^{\lim}(z)$$

$$z \leftrightarrow s \qquad\qquad v \leftrightarrow c(x) + \mu\hat{y} \qquad\qquad v - z \leftrightarrow c(x) + \mu\hat{y} - s = \mu y$$

AL without slack variables

Augmented Lagrangian function, $\mu > 0$, $y \in \mathbb{R}^m$,

$$\begin{aligned}\mathcal{L}_\mu(x, \hat{y}) &:= \min_s \mathcal{L}_\mu^{\mathrm{S}}(x, s, \hat{y}) \\ &= \mathcal{L}_\mu^{\mathrm{S}}(x, \mathcal{S}_\mu(x, \hat{y}), \hat{y}) \\ &= q(x) + \frac{1}{2\mu}\,\mathrm{dist}_D^2\,(c(x) + \mu\hat{y}) - \frac{\mu}{2}\|\hat{y}\|^2\end{aligned}$$

AL subproblem without slack variables

$$\begin{aligned} &\underset{x}{\text{minimize}} \quad \mathcal{L}_\mu(x, \hat{y}) \\ \Leftrightarrow \quad &\underset{x}{\text{minimize}} \quad \underbrace{f(x) + \frac{1}{2\mu}\,\text{dist}_D^2\,(c(x) + \mu\hat{y})}_{\psi(x)} + g(x) \end{aligned}$$

But $\psi : \mathbb{R}^n \to \mathbb{R}$ is not smooth, but don't worry!

$$\partial\left(\frac{1}{2}\,\text{dist}_D^2(v)\right) = v - \text{proj}_D(v)$$

Just select any element of $\partial\psi$ as a proxy for $\nabla\psi$!

$$\begin{aligned} &s \in \mathcal{S}_\mu(x, \hat{y}) \\ &\nabla\psi(x) \leftarrow \nabla f(x) + \frac{1}{\mu}\nabla c(x)^\top\,[c(x) + \mu\hat{y} - s] \end{aligned}$$

Adaptive Proximal Gradient Method with an Oracle

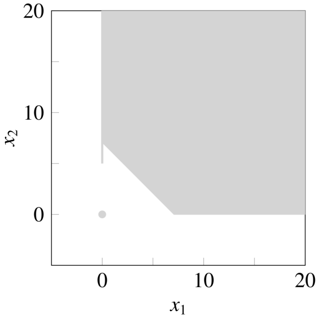
select $x^0 \in \mathbb{R}^n$, $\gamma_0 > 0$, $\alpha \in (0,1)$
for $k = 0, 1, 2, \dots$ do
while true **do**
    select $s^k \in \mathcal{S}_\mu(x^k, \hat{y})$, evaluate $\psi(x^k)$ and $\nabla\psi(x^k)$
    select $\bar{x}^k \in \mathrm{T}_{\gamma_k}(x^k)$
    select $\bar{s}^k \in \mathcal{S}_\mu(\bar{x}^k, \hat{y})$ and evaluate $\psi(\bar{x}^k)$
    if $\psi(\bar{x}^k) > \psi(x^k) + \langle \nabla\psi(x^k), \bar{x}^k - x^k \rangle + \frac{\alpha}{2\gamma_k}\|\bar{x}^k - x^k\|^2$ **then**
        set $\gamma_k \leftarrow \gamma_k / 2$
    else
        break
    end if
end while
set $\gamma_{k+1} \leftarrow \gamma_k$ and $x^{k+1} \leftarrow \bar{x}^k$

✓ also for PANOC$^+$

Numerical Results

Academic MPVC

$$\begin{aligned}
\underset{x}{\text{minimize}} \quad & 4x_1 + 2x_2 \\
\text{subject to} \quad & x_1 \geq 0, \qquad x_2 \geq 0 \\
& x_1 > 0 \quad \Rightarrow \quad x_1 + x_2 \geq 5\sqrt{2} \\
& x_2 > 0 \quad \Rightarrow \quad x_1 + x_2 \geq 5
\end{aligned}$$



Common NLP reformulation

$$\begin{aligned}
x_1 \geq 0, \qquad x_2 \geq 0 \\
x_1(x_1 + x_2 - 5\sqrt{2}) \geq 0 \\
x_2(x_1 + x_2 - 5) \geq 0
\end{aligned}$$

Global minimum at $x^\star = (0, 0)$ and local minimum at $x^\diamond = (0, 5)$.

[6] T. Hoheisel. Mathematical Programs with Vanishing Constraints. PhD thesis, 2009.

Academic MPVC: *convex* reformulation

Set of VCs:

$$\begin{aligned} D_{\mathrm{VC}} &:= \{(a,b) \mid a \geq 0, ab \geq 0\} \\ &= \{(a,b) \mid a = 0\} \cup \{(a,b) \mid a \geq 0, b \geq 0\} \end{aligned}$$

Introducing slacks, convex constraint set *D*:

$$\begin{aligned} f(x) &:= 4x_1 + 2x_2 \\ g(x) &:= \delta_{D_{\mathrm{VC}}}(x_1, x_3) + \delta_{D_{\mathrm{VC}}}(x_2, x_4) \\ c(x) &:= \begin{pmatrix} x_1 + x_2 - 5\sqrt{2} - x_3 \\ x_1 + x_2 - 5 - x_4 \end{pmatrix} \\ D &:= \{(0,0)\} \end{aligned}$$

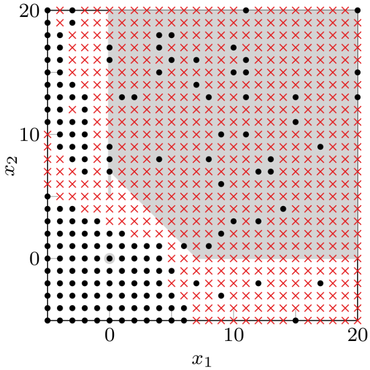
Academic MPVC: *nonconvex* reformulation

Without slacks, nonconvex constraint set D:

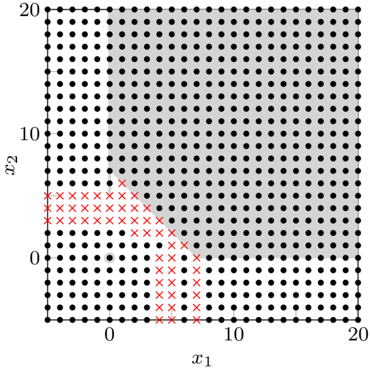
$$
\begin{aligned}
f(x) &:= 4x_1 + 2x_2 \\
g(x) &:= \delta_{\mathbb{R}_+}(x) \\
c(x) &:= \begin{pmatrix} x_1 \\ x_1 + x_2 - 5\sqrt{2} \\ x_2 \\ x_1 + x_2 - 5 \end{pmatrix} \\
D &:= D_{\mathrm{VC}} \times D_{\mathrm{VC}}
\end{aligned}
$$

Academic MPVC: results

Convex D:



Nonconvex D:



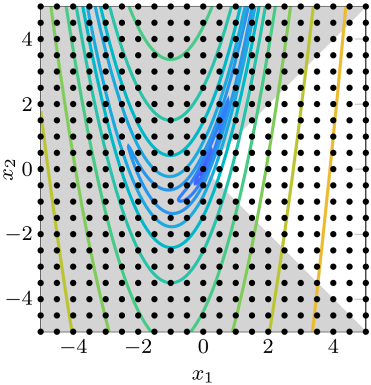
Global minimizer $x^\star = (0, 0)$ (black dot)
Local minimizer $x^\diamond = (0, 5)$ (red cross)

implicit formulations

Nonsmooth Rosenbrock with Or-Constraints

$$\begin{aligned} \underset{x}{\text{minimize}} \quad & 10(x_2 + 1 - (x_1+1)^2)^2 + |x_1| \\ \text{subject to} \quad & x_2 \le -x_1 \ \vee \ x_2 \ge x_1 \end{aligned}$$

Unique (global) minimizer
$x^\star = (0, 0)$.
The feasible set is nonconvex
and connected.



Nonsmooth Rosenbrock with Or-Constraints

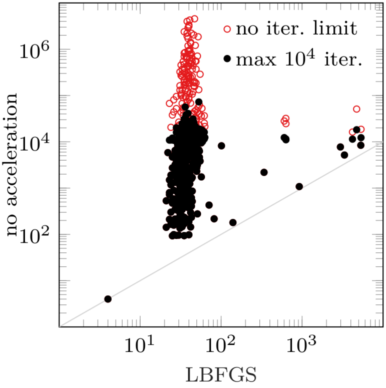
Set of either-or constraint:

$$D_{\mathrm{EO}} := \{(a,b) \mid a \geq 0 \vee b \geq 0\} = \{(a,b) \mid a \geq 0\} \cup \{(a,b) \mid b \geq 0\}$$

Nonconvex reformulation

$$f(x) := 10(x_2 + 1 - (x_1+1)^2)^2$$
$$g(x) := |x_1|$$
$$c(x) := \begin{pmatrix} -x_1 - x_2 \\ -x_1 + x_2 \end{pmatrix}$$
$$D := D_{\mathrm{EO}}$$

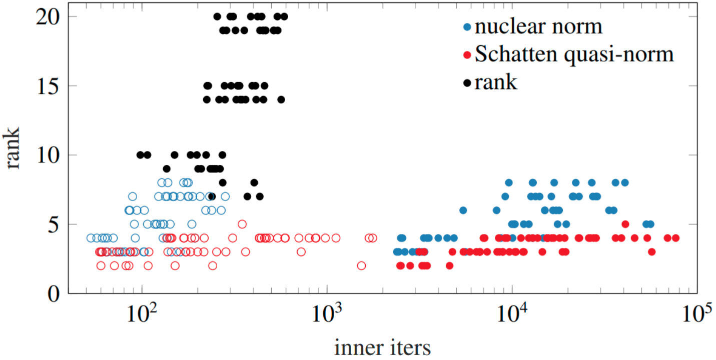
Global minimizer $x^\star = (0,0)$ found in all cases.



accelerated subsolver

Find more in the paper

- Sparse Portfolio Optimization
- Mixed-Integer Optimal Control with Switching Cost
- Matrix Completion with Minimum Rank



Conclusions

ALPS

Augmented Lagrangian Proximal Solver for constrained structured programs

stationarity concepts	attentive convergence
(regularity and CQs)	implicit formulations

simple | first order oracles | matrix free

PREPRINT: ARXIV 2203.05276

JULIA CODE: BAZINGA.JL

Alberto De Marchi
UniBw M
aldma.github.io

Thank you!
Questions?

References

Andreas Themelis, Lorenzo Stella, and Panagiotis Patrinos.
Forward-backward envelope for the sum of two nonconvex functions: Further properties and nonmonotone linesearch algorithms.
SIAM Journal on Optimization, 28(3):2274–2303, 2018.

R. Tyrrell Rockafellar and Roger J.B. Wets.
Variational Analysis, volume 317.
Springer, 1998.

Neal Parikh and Stephen Boyd.
Proximal algorithms.
Foundations and Trends in Optimization, 1(3):127–239, 2014.

Lorenzo Stella, Andreas Themelis, Pantelis Sopasakis, and Panagiotis Patrinos.
A simple and efficient algorithm for nonlinear model predictive control.
In *2017 IEEE 56th Annual Conference on Decision and Control (CDC)*, pages 1939–1944, 2017.

Alberto De Marchi and Andreas Themelis.
Proximal gradient algorithms under local Lipschitz gradient continuity.
Journal of Optimization Theory and Applications, 2022.
accepted.

Tim Hoheisel.
Mathematical Programs with Vanishing Constraints.
PhD thesis, University of Würzburg, 2009.