Block Decomposition of Large-Scale MINLPs and (Penalty) Alternating Direction Methods

Martin Schmidt

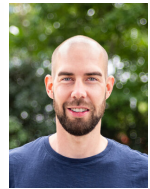
@schmaidt

March 30, 2023 — 3SMINLP in Sevilla

The Team

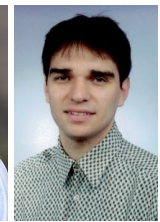
Pablo Burgard | Carina Costa | Björn Geißler | Thomas Kleinert | Dennis Kreber | Richard Krug







Günter Leugering | Alexander Martin | Antonio Morsi | Lars Schewe | Dieter Weninger

Overview

- Mixed-Integer Nonlinear Optimization
- Some Famous Alternating Methods
  - Feasibility Pump
  - k-Means Clustering
- Penalty Alternating Direction Methods
- Applications
  - Feasibility Pump
  - Bilevel Optimization
  - Gas Transport Networks
- Outlook

Mixed-Integer Nonlinear Optimization

Mixed-Integer Nonlinear Optimization

$$
\begin{aligned}
\min_{x,y} \quad & f(x,y) \\
\text{s.t.} \quad & g(x,y) = 0 \\
& h(x,y) \geq 0 \\
& x \in \mathbb{R}^{n_x},\ y \in \mathbb{Z}^{n_y} \cap [l,u]
\end{aligned}
$$

- $f : \mathbb{R}^{n_x} \times \mathbb{Z}^{n_y} \to \mathbb{R}$, $g : \mathbb{R}^{n_x} \times \mathbb{Z}^{n_y} \to \mathbb{R}^m$, $h : \mathbb{R}^{n_x} \times \mathbb{Z}^{n_y} \to \mathbb{R}^k$ can be nonlinear but should be C^2
- Finite bounds $-\infty < l \leq u < +\infty$

Mixed-Integer Nonlinear Optimization

Pro: Extremely versatile modeling tool

- Nonlinearities (economies of scale, physics, ...)
- Integer variables (decision making)

Mixed-Integer Nonlinear Optimization

Pro: Extremely versatile modeling tool

- Nonlinearities (economies of scale, physics, ...)
- Integer variables (decision making)

Cons: Extremely hard to solve

- NP-hard since it includes MILP (Kannan and Monma 1978)
- Even worse, nonconvex integer optimization problems are undecidable in general
  - Jeroslow (1973): example of a quadratically constrained integer program
  - Theorem: no computing device exists that can compute the optimum for all problems in this class
- MINLP software is much less mature than MILP software

Best Subset Selection

- Linear regression: $y = X\beta^0 + \varepsilon$
- Response vector $y \in \mathbb{R}^n$, observed data $X \in \mathbb{R}^{n \times p}$, true predictors $\beta^0 \in \mathbb{R}^p$, errors $\varepsilon \in \mathbb{R}^n$
- β^0 and ε are unknown
- Relevant recorded variables are often unknown; regressors β^0 are sparse
- More formally: $|\mathrm{supp}(\beta^0)| < p$ holds

Best Subset Selection

- Linear regression: $y = X\beta^0 + \varepsilon$
- Response vector $y \in \mathbb{R}^n$, observed data $X \in \mathbb{R}^{n \times p}$, true predictors $\beta^0 \in \mathbb{R}^p$, errors $\varepsilon \in \mathbb{R}^n$
- β^0 and ε are unknown
- Relevant recorded variables are often unknown; regressors β^0 are sparse
- More formally: $|\text{supp}(\beta^0)| < p$ holds

Goal

Select k out of p features, which best describe the relation between X and y

$$\begin{aligned} \min_{\beta \in \mathbb{R}^p} \quad & \|X\beta - y\|_2^2 \\ \text{s.t.} \quad & |\text{supp}(\beta)| \leq k \end{aligned}$$

Best Subset Selection

- Linear regression: $y = X\beta^0 + \varepsilon$
- Response vector $y \in \mathbb{R}^n$, observed data $X \in \mathbb{R}^{n \times p}$, true predictors $\beta^0 \in \mathbb{R}^p$, errors $\varepsilon \in \mathbb{R}^n$
- β^0 and ε are unknown
- Relevant recorded variables are often unknown; regressors β^0 are sparse
- More formally: $|\text{supp}(\beta^0)| < p$ holds

Goal

Select k out of p features, which best describe the relation between X and y

$$\begin{aligned} \min_{\beta \in \mathbb{R}^p} \quad & \|X\beta - y\|_2^2 \\ \text{s.t.} \quad & |\text{supp}(\beta)| \leq k \end{aligned}$$

- NP-hard: Natarajan (1995)
- Very active field of research: Atamtürk, Gómez (2018), Bertsimas, King, Mazumder (2016), Bertsimas, van Parys (2020), Dong, Chen, Linderoth (2015), ...
- Famous heuristic: Lasso (Tibshirani 1996 $\rightarrow$ more than 51 000 citations)

Best Subset Selection

$$\begin{aligned} \min_{\beta \in \mathbb{R}^p} \quad & \|X\beta - y\|_2^2 \\ \text{s.t.} \quad & |\text{supp}(\beta)| \leq k \end{aligned}$$

Best Subset Selection

$$\begin{aligned} \min_{\beta \in \mathbb{R}^p} \quad & \|X\beta - y\|_2^2 \\ \text{s.t.} \quad & |\text{supp}(\beta)| \leq k \end{aligned}$$

How to get to a block structure?

$$\begin{aligned} \min_{\gamma, \delta} \quad & \|X\gamma - y\|_2^2 \\ \text{s.t.} \quad & \gamma = \delta \\ & |\text{supp}(\delta)| \leq k \end{aligned}$$

Best Subset Selection

$$\begin{aligned} \min_{\beta \in \mathbb{R}^p} \quad & \|X\beta - y\|_2^2 \\ \text{s.t.} \quad & |\text{supp}(\beta)| \leq k \end{aligned}$$

How to get to a block structure?

$$\begin{aligned} \min_{\gamma, \delta} \quad & \|X\gamma - y\|_2^2 \\ \text{s.t.} \quad & \gamma = \delta \\ & |\text{supp}(\delta)| \leq k \end{aligned}$$

Block Separability: Variant #1

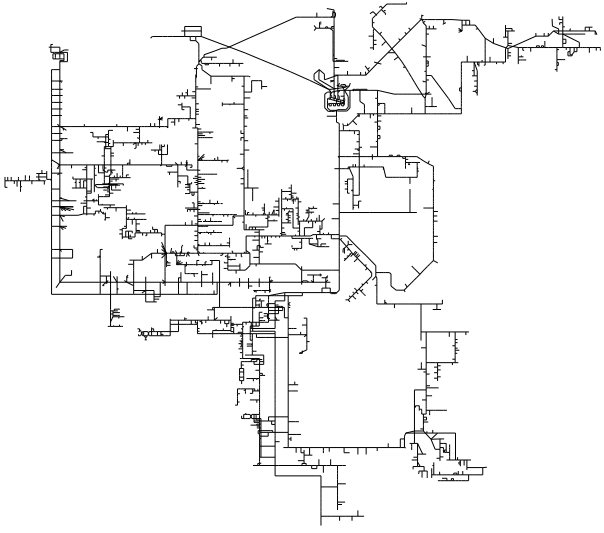
Blocks are determined by qualitatively different modeling aspects

- Nonlinearities
- Integrality constraints

Gas Transport Network Optimization

Given

- Gas transport network
- Contracts with supplying and discharging customers



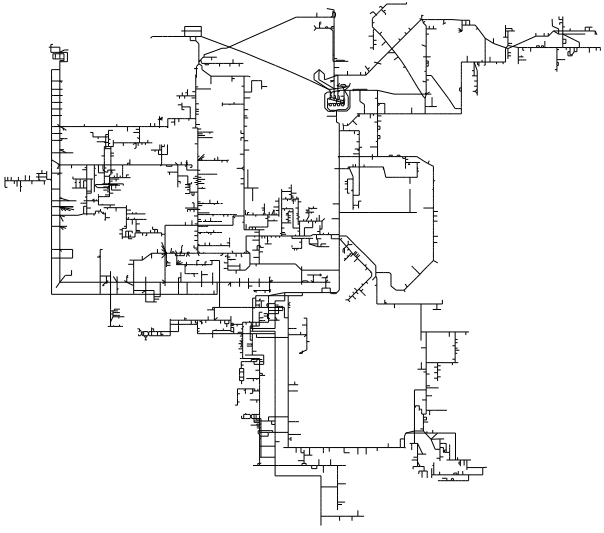
Gas Transport Network Optimization

Given

- Gas transport network
- Contracts with supplying and discharging customers

Question

- Is there a control of the network so that the resulting physical and technical state is feasible and satisfies all contracts?



Gas Transport Network Optimization: Nonlinear Flow Models

Euler equations for compressible fluids in cylindrical pipes

$$
\begin{aligned}
\frac{\partial \rho}{\partial t} + \frac{1}{A}\frac{\partial q}{\partial x} &= 0 \\
\frac{1}{A}\frac{\partial q}{\partial t} + \frac{\partial p}{\partial x} + \frac{1}{A}\frac{\partial (qv)}{\partial x} + g\rho s + \lambda(q)\frac{|v|v}{2D}\rho &= 0 \\
A\rho c_p \left(\frac{\partial T}{\partial t} + v\frac{\partial T}{\partial x}\right) - A\left(1 + \frac{T}{z}\frac{\partial z}{\partial T}\right)\frac{\partial p}{\partial t} - & \\
Av\frac{T}{z}\frac{\partial z}{\partial T}\frac{\partial p}{\partial x} + A\rho v g s + \pi D c_{\mathrm{HT}}\left(T - T_{\mathrm{soil}}\right) &= 0 \\
\frac{p}{R_s z(p,T) T} &= \rho
\end{aligned}
$$

Gas Transport Network Optimization: Nonlinear Flow Models

Euler equations for compressible fluids in cylindrical pipes

$$\frac{\partial \rho}{\partial t} + \frac{1}{A}\frac{\partial q}{\partial x} = 0$$

$$\frac{1}{A}\frac{\partial q}{\partial t} + \frac{\partial p}{\partial x} + \frac{1}{A}\frac{\partial (qv)}{\partial x} + g\rho s + \lambda(q)\frac{|v|v}{2D}\rho = 0$$

$$\frac{p}{R_s z(p,T) T} = \rho$$

Gas Transport Network Optimization: Nonlinear Flow Models

Euler equations for compressible fluids in cylindrical pipes

$$\frac{1}{A}\frac{\partial q}{\partial x} = 0$$

$$\frac{\partial p}{\partial x} + \frac{1}{A}\frac{\partial (qv)}{\partial x} + g\rho s + \lambda(q)\frac{|v|v}{2D}\rho = 0$$

$$\frac{p}{R_s z(p,T)T} = \rho$$

Gas Transport Network Optimization: Nonlinear Flow Models

Euler equations for compressible fluids in cylindrical pipes

$$\frac{1}{A}\frac{\partial q}{\partial x} = 0$$

$$\frac{\partial p}{\partial x} + g\rho s + \lambda(q)\frac{|v|v}{2D}\rho = 0$$

$$\frac{p}{R_s z(p,T)T} = \rho$$

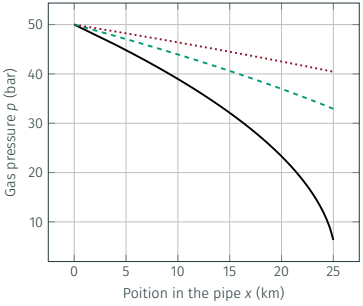
Gas Transport Network Optimization: Nonlinear Flow Models

Euler equations for compressible fluids in cylindrical pipes

$$\frac{1}{A}\frac{\partial q}{\partial x} = 0$$

$$\frac{\partial p}{\partial x} + g\rho s + \lambda(q)\frac{|v|v}{2D}\rho = 0$$

$$\frac{p}{R_s z(p,T)T} = \rho$$



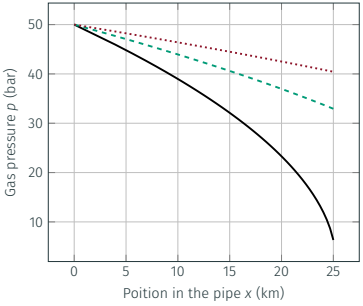
Gas Transport Network Optimization: Nonlinear Flow Models

Euler equations for compressible fluids in cylindrical pipes

$$\frac{1}{A}\frac{\partial q}{\partial x} = 0$$

$$\frac{\partial p}{\partial x} + g\rho s + \lambda(q)\frac{|v|v}{2D}\rho = 0$$

$$\frac{p}{R_s z(p,T) T} = \rho$$



Approximation: Weymouth equation

$$p(x)^2 = \left(p_{\text{in}}^2 - \Lambda|q|q\frac{e^{Sx}-1}{S}\right)e^{-Sx}$$

Gas Transport Network Optimization: Compressor Stations



Possible Modes of Operation

- Bypass mode
- Closed

Gas Transport Network Optimization: Compressor Stations



Possible Modes of Operation

- Active
  - Pressure increase: $p_{\text{out}} > p_{\text{in}}$
  - Discrete control: Which machine is running and which not?
- Bypass mode
- Closed

Gas Transport Network Optimization

- Directed graph $G = (V, A)$
- Nodes $V = V_+ \cup V_- \cup V_0$
- Arcs $A = A_{\text{pi}} \cup A_{\text{cs}} \cup A_{\text{re}} \cup A_{\text{va}} \cup A_{\text{cv}}$

Gas Transport Network Optimization

- Directed graph $G = (V, A)$
- Nodes $V = V_+ \cup V_- \cup V_0$
- Arcs $A = A_{\text{pi}} \cup A_{\text{cs}} \cup A_{\text{re}} \cup A_{\text{va}} \cup A_{\text{cv}}$

Nodes V	Mass balance & mixing
Pipes A_{pi}	Nonlinear differential equations
Compressor stations A_{cs}	Mixed-integer nonlinear model
Resistors A_{re}	Nonlinear pressure drop model
Valves A_{va}	Mixed-integer linear model
Control valves A_{cv}	Mixed-integer linear model

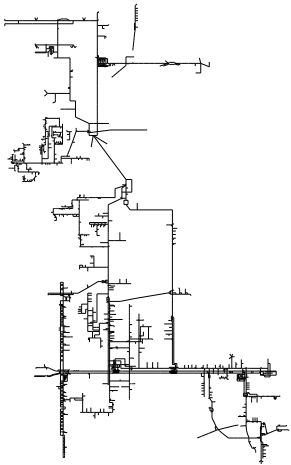
Gas Transport Network Optimization

- Directed graph $G = (V, A)$
- Nodes $V = V_+ \cup V_- \cup V_0$
- Arcs $A = A_{\text{pi}} \cup A_{\text{cs}} \cup A_{\text{re}} \cup A_{\text{va}} \cup A_{\text{cv}}$

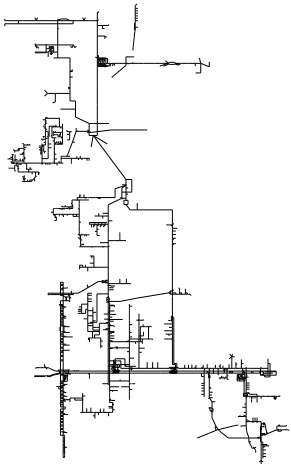
Nodes V	Mass balance & mixing
Pipes A_{pi}	Nonlinear differential equations
Compressor stations A_{cs}	Mixed-integer nonlinear model
Resistors A_{re}	Nonlinear pressure drop model
Valves A_{va}	Mixed-integer linear model
Control valves A_{cv}	Mixed-integer linear model

Large nonconvex MINLP with differential equations

Gas Transport Network Optimization: Block Structure

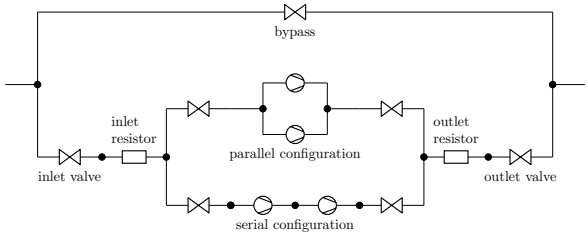


Gas Transport Network Optimization: Block Structure



Block Separability: Variant #2

Blocks are determined by the graph structure



Standalone: nonconvex MINLP!

Supply Network Planning

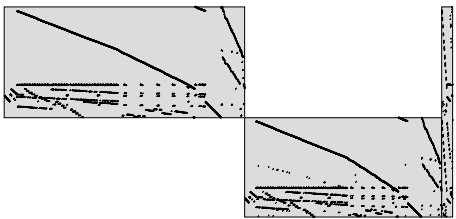
- Goal: satisfaction of demand for certain goods
- Goods need to be produced and transported
- Production requires machines and raw materials
- Produced goods can be stored

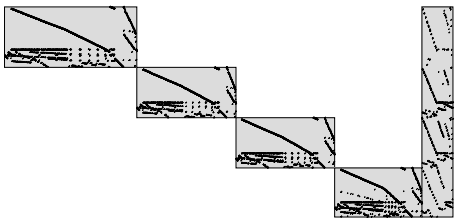
Supply Network Planning

- Goal: satisfaction of demand for certain goods
- Goods need to be produced and transported
- Production requires machines and raw materials
- Produced goods can be stored

Block Separability: Variant #3

Blocks are determined by the linear algebra; induced by a time-discrete model





Linear Bilevel Optimization

$$\min_{x \in \mathbb{R}^n, y \in \mathbb{R}^m} \quad c^\top x + d^\top y \quad \text{s.t.} \quad Ax + By \geq a, \quad y \in \mathcal{S}(x)$$

Linear Bilevel Optimization

$$\min_{x \in \mathbb{R}^n, y \in \mathbb{R}^m} \quad c^\top x + d^\top y \quad \text{s.t.} \quad Ax + By \geq a, \quad y \in \mathcal{S}(x)$$

$\mathcal{S}(x)$ denotes the set of optimal solutions of the x-parameterized linear problem

$$\min_{y} \quad f^\top y \quad \text{s.t.} \quad Dy \geq b - Cx$$

Linear Bilevel Optimization

$$\min_{x \in \mathbb{R}^n, y \in \mathbb{R}^m} \quad c^\top x + d^\top y \quad \text{s.t.} \quad Ax + By \geq a, \quad y \in \mathcal{S}(x)$$

$\mathcal{S}(x)$ denotes the set of optimal solutions of the x-parameterized linear problem

$$\min_{y} \quad f^\top y \quad \text{s.t.} \quad Dy \geq b - Cx$$

- Strongly NP-hard problem in general (Hansen, Jaumard, Savard 1992)
- In what follows: optimistic variant

Linear Bilevel Optimization: KKT Reformulation

The lower-level problem is an LP

$$\min_{y} \quad f^{\top} y \quad \text{s.t.} \quad Dy \geq b - Cx$$

Linear Bilevel Optimization: KKT Reformulation

The lower-level problem is an LP

$$\min_y \quad f^\top y \quad \text{s.t.} \quad Dy \geq b - Cx$$

The KKT conditions

$$\begin{aligned} Cx + Dy &\geq b \\ \lambda \geq 0,\ D^\top \lambda &= f \\ \lambda^\top (Cx + Dy - b) &= 0 \end{aligned}$$

are necessary and sufficient

Linear Bilevel Optimization: KKT Reformulation

The lower-level problem is an LP

$$\min_{y} \quad f^\top y \quad \text{s.t.} \quad Dy \geq b - Cx$$

The KKT conditions

$$\begin{aligned} Cx + Dy &\geq b \\ \lambda \geq 0,\ D^\top \lambda &= f \\ \lambda^\top (Cx + Dy - b) &= 0 \end{aligned}$$

are necessary and sufficient

Single-Level Reformulation

$$\begin{aligned} \min_{x,y,\lambda} \quad & c^\top x + d^\top y \\ \text{s.t.} \quad & Ax + By \geq a, \quad Cx + Dy \geq b \\ & \lambda \geq 0, \quad D^\top \lambda = f \\ & \lambda^\top (Cx + Dy - b) = 0 \end{aligned}$$

Linear Bilevel Optimization: KKT Reformulation

$$
\begin{aligned}
\min_{x,y,\lambda} \quad & c^\top x + d^\top y \\
\text{s.t.} \quad & Ax + By \geq a, \quad Cx + Dy \geq b \\
& \lambda \geq 0, \quad D^\top \lambda = f \\
& \lambda^\top (Cx + Dy - b) = 0
\end{aligned}
$$

Linear Bilevel Optimization: KKT Reformulation

$$
\begin{aligned}
\min_{x,y,\lambda} \quad & c^\top x + d^\top y \\
\text{s.t.} \quad & Ax + By \geq a, \quad Cx + Dy \geq b \\
& \lambda \geq 0, \quad D^\top \lambda = f \\
& \lambda^\top (Cx + Dy - b) = 0
\end{aligned}
$$

$$
\begin{aligned}
\min_{x,y,\lambda,u} \quad & c^\top x + d^\top y \\
\text{s.t.} \quad & Ax + By \geq a, \quad Cx + Dy \geq b \\
& \lambda \geq 0, \quad D^\top \lambda = f \\
& Cx + Dy - b \leq M_{\mathrm{P}}(1 - u) \\
& \lambda \leq M_{\mathrm{D}} u \\
& u \in \{0,1\}^{\ell}
\end{aligned}
$$

Linear Bilevel Optimization: KKT Reformulation

$$\begin{aligned}\min_{x,y,\lambda} \quad & c^\top x + d^\top y \\ \text{s.t.} \quad & Ax + By \geq a, \quad Cx + Dy \geq b \\ & \lambda \geq 0, \quad D^\top \lambda = f \\ & \lambda^\top (Cx + Dy - b) = 0\end{aligned}$$

$$\begin{aligned}\min_{x,y,\lambda,u} \quad & c^\top x + d^\top y \\ \text{s.t.} \quad & Ax + By \geq a, \quad Cx + Dy \geq b \\ & \lambda \geq 0, \quad D^\top \lambda = f \\ & Cx + Dy - b \leq M_P(1 - u) \\ & \lambda \leq M_D u \\ & u \in \{0,1\}^\ell\end{aligned}$$

That's an MILP

Linear Bilevel Optimization: KKT Reformulation

$$\begin{aligned}\min_{x,y,\lambda} \quad & c^\top x + d^\top y \\ \text{s.t.} \quad & Ax + By \geq a, \quad Cx + Dy \geq b \\ & \lambda \geq 0, \quad D^\top \lambda = f \\ & \lambda^\top (Cx + Dy - b) = 0\end{aligned}$$

Block Separability: Variant #4

Blocks are determined by primal vs. dual parts of the model

$$\begin{aligned}\min_{x,y,\lambda,u} \quad & c^\top x + d^\top y \\ \text{s.t.} \quad & Ax + By \geq a, \quad Cx + Dy \geq b \\ & \lambda \geq 0, \quad D^\top \lambda = f \\ & Cx + Dy - b \leq M_P(1 - u) \\ & \lambda \leq M_D u \\ & u \in \{0,1\}^\ell\end{aligned}$$

That's an MILP

Some Famous Alternating Methods

Overview

Mixed-Integer Nonlinear Optimization

Some Famous Alternating Methods

- Feasibility Pump
- k-Means Clustering

Penalty Alternating Direction Methods

Applications

- Feasibility Pump
- Bilevel Optimization
- Gas Transport Networks

Outlook

The Basic Feasibility Pump for 0-1-MIPs

$$
\begin{aligned}
\min_{x} \quad & c^\top x \\
\text{s.t.} \quad & Ax \geq b \\
& x_i \in \mathbb{Z} \quad \text{for all } i \in I
\end{aligned}
$$

The Basic Feasibility Pump for 0-1-MIPs

$$\begin{aligned} \min_{x} \quad & c^\top x \\ \text{s.t.} \quad & Ax \geq b \\ & x_i \in \mathbb{Z} \quad \text{for all } i \in I \end{aligned}$$

Two Main Challenges

1. Optimize over a polyhedron
2. Satisfy integrality conditions

The Basic Feasibility Pump for 0-1-MIPs

$$\begin{aligned} \min_{x} \quad & c^\top x \\ \text{s.t.} \quad & Ax \geq b \\ & x_i \in \mathbb{Z} \quad \text{for all } i \in I \end{aligned}$$

Two Main Challenges

1. Optimize over a polyhedron
2. Satisfy integrality conditions

Compute $\bar{x}^0 \in \arg\min\{c^\top x : x \in P\}$.
if $\bar{x}^0$ is integer feasible **then**
  return $\bar{x}^0$
end if
Set $\tilde{x}^0 = \lceil \bar{x}^0 \rfloor$ and $k \leftarrow 0$.
while not termination condition **do**
  Compute $\bar{x}^{k+1} \in \arg\min\{\|x_I - \tilde{x}_I^k\|_1 : x \in P\}$.
  if $\bar{x}^{k+1}$ is integer feasible **then**
    return $\bar{x}^{k+1}$
  end if
  Set $\tilde{x}^{k+1} = \lceil \bar{x}^{k+1} \rfloor$.
  if algorithm stalls or cycles **then**
    perturb $\tilde{x}^{k+1}$
  end if
  Set $k \leftarrow k + 1$.
end while

Overview

Mixed-Integer Nonlinear Optimization

Some Famous Alternating Methods

- Feasibility Pump
- k-Means Clustering

Penalty Alternating Direction Methods

Applications

- Feasibility Pump
- Bilevel Optimization
- Gas Transport Networks

Outlook

What is the MSSC problem?

Given

- n data points in $\mathbb{R}^d$: $p \in P \subset \mathbb{R}^d$, $|P| = n$
- $2 \leq k \leq n$ clusters

What is the MSSC problem?

Given

- n data points in $\mathbb{R}^d$: $p \in P \subset \mathbb{R}^d$, $|P| = n$
- $2 \leq k \leq n$ clusters

Task

Find ...

- an assignment for every data point to exactly one cluster
- and a representative for every cluster

What is the MSSC problem?

Given

- n data points in $\mathbb{R}^d$: $p \in P \subset \mathbb{R}^d$, $|P| = n$
- $2 \leq k \leq n$ clusters

Task

Find ...

- an assignment for every data point to exactly one cluster
- and a representative for every cluster

Key problem in unsupervised machine learning

A little bit more formal ...

Find k centroids c^j, $j = 1, \ldots, k$, that solve

$$\min_{c} \; h(P, c), \quad c = (c^j)_{j=1,\ldots,k},$$

A little bit more formal ...

Find k centroids c^j, $j = 1, \ldots, k$, that solve

$$\min_c \; h(P, c), \quad c = (c^j)_{j=1,\ldots,k},$$

with

$$h(P, c) = \sum_{j=1}^{k} \sum_{p \in S_j} \|p - c^j\|_2^2$$

A little bit more formal ...

Find k centroids c^j, $j = 1, \ldots, k$, that solve

$$\min_c \; h(P, c), \quad c = (c^j)_{j=1,\ldots,k},$$

with

$$h(P, c) = \sum_{j=1}^{k} \sum_{p \in S_j} \|p - c^j\|_2^2$$

Two Main Challenges

1. Find an assignment of points to clusters
2. Compute a representative point for each cluster

A little bit more formal ...

Find k centroids c^j, $j = 1, \ldots, k$, that solve

$$\min_{c} \; h(P, c), \quad c = (c^j)_{j=1,\ldots,k},$$

with

$$h(P, c) = \sum_{j=1}^{k} \sum_{p \in S_j} \|p - c^j\|_2^2$$

Two Main Challenges

1. Find an assignment of points to clusters
2. Compute a representative point for each cluster

k-Means Clustering

Simply alternate between these two challenges

Penalty Alternating Direction Methods

Problem Statement

$$\begin{aligned} \min_{x,y} \quad & f(x,y) \\ \text{s.t.} \quad & g(x,y) = 0, \quad h(x,y) \geq 0 \\ & x \in X, \quad y \in Y \end{aligned}$$

Problem Statement

$$\begin{aligned} \min_{x,y} \quad & f(x,y) \\ \text{s.t.} \quad & g(x,y) = 0, \quad h(x,y) \geq 0 \\ & x \in X, \quad y \in Y \end{aligned}$$

Assumptions

- Objective function $f : \mathbb{R}^{n_x+n_y} \to \mathbb{R}$ is continuous
- Constraint functions $g : \mathbb{R}^{n_x+n_y} \to \mathbb{R}^m$, $h : \mathbb{R}^{n_x+n_y} \to \mathbb{R}^p$ are continuous
- X and Y non-empty, compact, and disjoint
- Ω: feasible set

Alternating Direction Methods

Choose initial value $(x^0, y^0) \in X \times Y$.

for $k = 0, 1, \ldots$ **do**

Compute

$$x^{k+1} \in \arg\min_x \{f(x, y^k) \colon g(x, y^k) = 0,\ h(x, y^k) \geq 0,\ x \in X\}.$$

Compute

$$y^{k+1} \in \arg\min_y \{f(x^{k+1}, y) \colon g(x^{k+1}, y) = 0,\ h(x^{k+1}, y) \geq 0,\ y \in Y\}.$$

Set $k \leftarrow k + 1$.

end for

Alternating Direction Methods

Choose initial value $(x^0, y^0) \in X \times Y$.
for $k = 0, 1, \ldots$ **do**
Compute

$$x^{k+1} \in \arg\min_x \{f(x, y^k) \colon g(x, y^k) = 0,\ h(x, y^k) \geq 0,\ x \in X\}.$$

Compute

$$y^{k+1} \in \arg\min_y \{f(x^{k+1}, y) \colon g(x^{k+1}, y) = 0,\ h(x^{k+1}, y) \geq 0,\ y \in Y\}.$$

Set $k \leftarrow k + 1$.
end for

- Glowinski, Marroco (1975), Gabay, Mercier (1976):
  Classic ADMs as extensions of Lagrangian methods
- Boyd et al. (2011): Large number of applications like
  machine learning, big data analysis, distributed computing, ...

Alternating Direction Methods: Convergence

A point $(x^*, y^*) \in \Omega$ is a partial minimum if

$$f(x^*, y^*) \leq f(x, y^*) \quad \text{for all } (x, y^*) \in \Omega,$$
$$f(x^*, y^*) \leq f(x^*, y) \quad \text{for all } (x^*, y) \in \Omega$$

holds.

Alternating Direction Methods: Convergence

A point $(x^*, y^*) \in \Omega$ is a partial minimum if

$$f(x^*, y^*) \leq f(x, y^*) \quad \text{for all } (x, y^*) \in \Omega,$$
$$f(x^*, y^*) \leq f(x^*, y) \quad \text{for all } (x^*, y) \in \Omega$$

holds.

Theorem (Gorski et al. 2007)

Let $\{(x^i, y^i)\}_{i=0}^{\infty}$ be a sequence with $(x^{i+1}, y^{i+1}) \in \Theta(x^i, y^i)$ and

$$\Theta(x^i, y^i) := \{(x^*, y^*) \colon f(x^*, y^i) \leq f(x, y^i)\ \forall x \in X,\ f(x^*, y^*) \leq f(x^*, y)\ \forall y \in Y\}.$$

Further assume that the assumptions (above) hold and that the solution of the first problem is always unique. Then, every convergent sub-sequence of $\{(x^i, y^i)\}_{i=0}^{\infty}$ converges to a partial minimum. For two accumulation points z, z' of such sub-sequences it holds $f(z) = f(z')$.

Penalty Extension of the ADM Framework

- Weighted ℓ_1 penalty function

$$\phi_1(x, y; \mu, \rho) := f(x, y) + \sum_{i=1}^{m} \mu_i |g_i(x, y)| + \sum_{i=1}^{p} \rho_i [h_i(x, y)]^-$$

  with

$$[\alpha]^- := \begin{cases} 0, & \text{if } \alpha \geq 0 \\ -\alpha, & \text{if } \alpha < 0 \end{cases}$$

- Penalty parameters $\mu = (\mu_i)_{i=1}^m, \rho = (\rho_i)_{i=1}^p \geq 0$

Penalty Extension of the ADM Framework

- Weighted ℓ_1 penalty function

$$\phi_1(x, y; \mu, \rho) := f(x, y) + \sum_{i=1}^{m} \mu_i |g_i(x, y)| + \sum_{i=1}^{p} \rho_i [h_i(x, y)]^-$$

with

$$[\alpha]^- := \begin{cases} 0, & \text{if } \alpha \geq 0 \\ -\alpha, & \text{if } \alpha < 0 \end{cases}$$

- Penalty parameters $\mu = (\mu_i)_{i=1}^m, \rho = (\rho_i)_{i=1}^p \geq 0$

Penalty problem

$$\min_{x,y} \quad \phi_1(x, y; \mu, \rho) \quad \text{s.t.} \quad x \in X,\ y \in Y$$

ℓ_1 Penalty Alternating Direction Method

1: Choose initial value $(x^{0,0}, y^{0,0}) \in X \times Y$ and initial penalty parameters $\mu^0, \rho^0 \geq 0$.
2: **for** $k = 0, 1, \ldots$ **do**
3: Set $l = 0$.
4: **while** $(x^{k,l}, y^{k,l})$ is not a partial minimum of the penalty problem with $\mu = \mu^k$ and $\rho = \rho^k$ **do**
5: Compute

$$x^{k,l+1} \in \arg\min_x \{\phi_1(x, y^{k,l}; \mu^k, \rho^k) \colon x \in X\}.$$

6: Compute

$$y^{k,l+1} \in \arg\min_y \{\phi_1(x^{k,l+1}, y; \mu^k, \rho^k) \colon y \in Y\}.$$

7: Set $l \leftarrow l + 1$.
8: **end while**
9: Check feasibility and choose new penalty parameters $\mu^{k+1} \geq \mu^k$ and $\rho^{k+1} \geq \rho^k$.
10: **end for**

Penalty ADM: Convergence

Lemma

Let (x^*, y^*) be a partial minimum of $\phi_1(x, y; \mu, \rho)$ for fixed penalty parameters $\mu, \rho \geq 0$ and let (x^*, y^*) be feasible for the original problem. Then, (x^*, y^*) is a partial minimum of the original problem.

Lemma

Let (x^*, y^*) be a partial minimum of $\phi_1(x, y; \mu, \rho)$ for fixed penalty parameters $\mu, \rho \geq 0$ and let (x^*, y^*) be feasible for the original problem. Then, (x^*, y^*) is a partial minimum of the original problem.

Lemma

Suppose "the assumptions" hold and $\mu_i^k \nearrow \infty$ for all $i = 1, \ldots, m$ as well as $\rho_i^k \nearrow \infty$ for all $i = 1, \ldots, p$. Furthermore, let (x^k, y^k) be a sequence of partial minima of the penalty problem (for $\mu = \mu^k$ and $\rho = \rho^k$) with $(x^k, y^k) \to (x^*, y^*)$. Then, there exist weights $\bar{\mu}, \bar{\rho} \geq 0$ such that (x^*, y^*) is a partial minimum of the weighted l_1 feasibility measure

$$\chi_{\bar{\mu},\bar{\rho}}(x, y) := \sum_{i=1}^{m} \bar{\mu}_i |g_i(x, y)| + \sum_{i=1}^{p} \bar{\rho}_i [h_i(x, y)]^-.$$

Penalty ADM: Convergence (Geißler, Morsi, Schewe, S., 2017)

Theorem

Suppose "the assumptions" hold and $\mu_i^k \nearrow \infty$ for all $i = 1, \ldots, m$ and $\rho_i^k \nearrow \infty$ for all $i = 1, \ldots, p$. Furthermore, let (x^k, y^k) be a sequence of partial minima of the penalty problem (for $\mu = \mu^k$ and $\rho = \rho^k$) with $(x^k, y^k) \to (x^*, y^*)$. Then, the following holds:

1. (x^*, y^*) is a partial minimum of the original problem or
2. there exist weights $\bar{\mu}, \bar{\rho} \geq 0$, such that (x^*, y^*) is a partial minimum of the feasibility measure $\chi_{\bar{\mu},\bar{\rho}}$.

Penalty ADM: Convergence (Geißler, Morsi, Schewe, S., 2017)

Theorem

Suppose "the assumptions" hold and $\mu_i^k \nearrow \infty$ for all $i = 1, \ldots, m$ and $\rho_i^k \nearrow \infty$ for all $i = 1, \ldots, p$. Furthermore, let (x^k, y^k) be a sequence of partial minima of the penalty problem (for $\mu = \mu^k$ and $\rho = \rho^k$) with $(x^k, y^k) \to (x^*, y^*)$. Then, the following holds:

1. (x^*, y^*) is a partial minimum of the original problem or
2. there exist weights $\bar{\mu}, \bar{\rho} \geq 0$, such that (x^*, y^*) is a partial minimum of the feasibility measure $\chi_{\bar{\mu},\bar{\rho}}$.

Case 1 can be improved if stronger assumptions are used.

Applications

Overview

Mixed-Integer Nonlinear Optimization

Some Famous Alternating Methods

- Feasibility Pump
- k-Means Clustering

Penalty Alternating Direction Methods

Applications

- Feasibility Pump
- Bilevel Optimization
- Gas Transport Networks

Outlook

The Feasibility Pump for 0-1-MIPs

Compute $\bar{x}^0 \in \arg\min\{c^\top x : x \in P\}$.
if $\bar{x}^0$ is integer feasible **then**
  return $\bar{x}^0$
end if
Set $\tilde{x}^0 = \lceil \bar{x}^0 \rfloor$ and $k \leftarrow 0$.
while not termination condition **do**
  Compute $\bar{x}^{k+1} \in \arg\min\{\|x_I - \tilde{x}_I^k\|_1 : x \in P\}$.
  if $\bar{x}^{k+1}$ is integer feasible **then**
    return $\bar{x}^{k+1}$
  end if
  Set $\tilde{x}^{k+1} = \lceil \bar{x}^{k+1} \rfloor$.
  if algorithm stalls or cycles **then**
    perturb $\tilde{x}^{k+1}$
  end if
  Set $k \leftarrow k + 1$.
end while

- Original Papers
  - Fischetti, Glover, Lodi (2005)
  - Bertacco, Fischetti, Lodi (2007)
- Many follow-up papers/improvements
  - Achterberg and Berthold (2007)
  - Fischetti and Salvagnin (2009)
  - Baena and Castro (2011)
  - Boland et al. (2014)
  - ...
- Only very little theory is/was known
  - Eckstein, Neidak (2007)
  - Boland et al. (2012)
  - De Santis, Lucidi, Rinaldi (2013)
  - De Santis, Lucidi, Rinaldi (2014)
  - Geißler, Morsi, Schewe, S. (2017)

Feasibility Pumps and (P)ADMs

$$\begin{aligned} \min_{x} \quad & c^\top x \\ \text{s.t.} \quad & Ax \geq b \\ & x_i \in \mathbb{Z} \quad \text{for all } i \in I \end{aligned}$$

Feasibility Pumps and (P)ADMs

$$\begin{aligned} \min_{x} \quad & c^\top x \\ \text{s.t.} \quad & Ax \geq b \\ & x_i \in \mathbb{Z} \quad \text{for all } i \in I \end{aligned}$$

Duplicate variables x_I using the new variable vector $y \in \{0,1\}^I$

$$\begin{aligned} \min_{x,y} \quad & c^\top x \\ \text{s.t.} \quad & x \in X := \{x \in \mathbb{R}^n : Ax \geq b,\ x_I \in [0,1]^I\} \\ & y \in Y := \{0,1\}^I, \quad g(x,y) = x_I - y = 0, \end{aligned}$$

$$\begin{aligned} \min_{x} \quad & c^\top x \\ \text{s.t.} \quad & Ax \geq b \\ & x_i \in \mathbb{Z} \quad \text{for all } i \in I \end{aligned}$$

Duplicate variables x_I using the new variable vector $y \in \{0,1\}^I$

$$\begin{aligned} \min_{x,y} \quad & c^\top x \\ \text{s.t.} \quad & x \in X := \{x \in \mathbb{R}^n : Ax \geq b,\ x_I \in [0,1]^I\} \\ & y \in Y := \{0,1\}^I, \quad g(x,y) = x_I - y = 0, \end{aligned}$$

Move coupling constraints to objective function

$$\begin{aligned} \min_{x,y} \quad & \|x_I - y\|_1 \\ \text{s.t.} \quad & x \in X := \{x \in \mathbb{R}^n : Ax \geq b,\ x_I \in [0,1]^I\}, \quad y \in Y := \{0,1\}^I \end{aligned}$$

Lemma

The ADM does not cycle.

Applied to the reformulated MIP, this is equivalent to Proposition 1 of De Santis, Lucidi, Rinaldi (2013).

Lemma

The ADM does not cycle.

Applied to the reformulated MIP, this is equivalent to
Proposition 1 of De Santis, Lucidi, Rinaldi (2013).

Theorem

The idealized feasibility pump terminates at a partial minimum (x^*, y^*) *of the reformulated MIP after a finite number of iterations. If the partial minimum* (x^*, y^*) *has objective value* $\|x_I^* - y^*\|_1 = 0$*, the point* (x^*, y^*) *is feasible for the original MIP.*

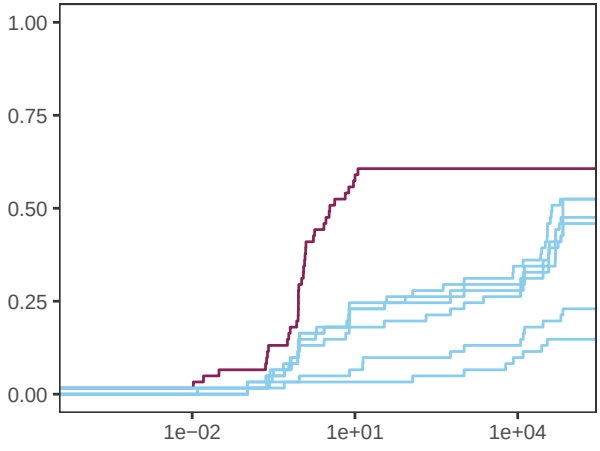
An analogous result can be obtained for convex MINLPs.

PADM & Feasibility Pumps

- Main goal: Get rid of the perturbation step in the feasibility pump
- Perturbation is applied if the classic feasibility pump cycles/stalls
- This is the case if the classic feasibility pump has found a partial minimum
- Replace perturbation by penalty parameter updates!
- The entire PADM theory applies to classic feasibility pumps with perturbation replaced with penalty parameter updates

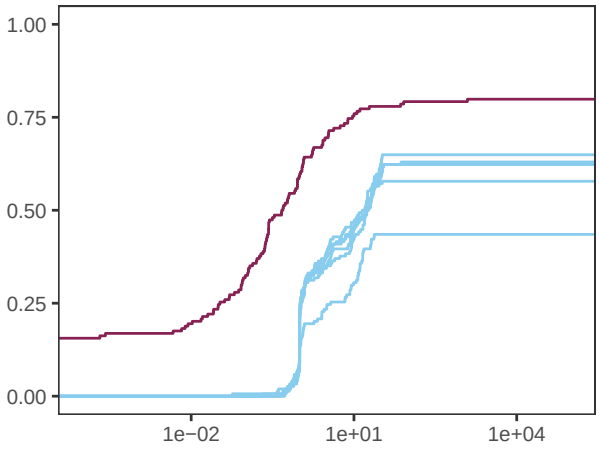
PADM & Feasibility Pumps: Improved Primal-Dual Gaps

Penalty ADM vs. all 6 feasibility pump versions from D'Ambrosio, Frangioni, Liberti, and Lodi (2012)



PADM & Feasibility Pumps: Improved Primal-Dual Gaps

Penalty ADM vs. all 5 feasibility pump versions from Berthold (2014)



PADM in SCIP

Screenshot from “Enabling Research through the SCIP Optimization Suite 8.0” by Bestuzheva et al. (2023)

In SCIP 7.0, the Benders decomposition framework and the heuristic Graph Induced Neighborhood Search were extended to exploit user-provided decompositions, and a first version of the heuristic Penalty Alternating Direction Method (PADM) [25, 55] was introduced. SCIP 8.0 comes with an improvement of PADM and provides another decomposition heuristic Dynamic Partition Search (DPS) [8].

Improvement of Penalty Alternating Direction Method. PADM splits an MINLP into several subproblems according to a given decomposition $\mathcal{D}$ with linking variables only, whereby the linking variables get copied and the differences are penalized. Then the subproblems are alternatingly solved. For faster convergence, the objective function of each subproblem has been replaced by a penalty term, and this replacement can lead to arbitrarily bad solutions. Therefore, PADM has been extended by the option to improve a found solution by reintroducing the original objective function.

Overview

Mixed-Integer Nonlinear Optimization

Some Famous Alternating Methods

- Feasibility Pump
- k-Means Clustering

Penalty Alternating Direction Methods

Applications

- Feasibility Pump
- Bilevel Optimization
- Gas Transport Networks

Outlook

Linear Bilevel Problems: Single-Level Reformulation

Single-level reformulation based on strong duality

$$\begin{aligned} \min_{x,y,\lambda} \quad & c^\top x + d^\top y \\ \text{s.t.} \quad & Ax + By \geq a, \quad Cx + Dy \leq b \\ & D^\top \lambda = e, \quad \lambda \geq 0 \\ & e^\top y - b^\top \lambda + x^\top C^\top \lambda \geq 0 \end{aligned}$$

Linear Bilevel Problems: Single-Level Reformulation

Single-level reformulation based on strong duality

$$\begin{aligned} \min_{x,y,\lambda} \quad & c^\top x + d^\top y \\ \text{s.t.} \quad & Ax + By \geq a, \quad Cx + Dy \leq b \\ & D^\top \lambda = e, \quad \lambda \geq 0 \\ & e^\top y - b^\top \lambda + x^\top C^\top \lambda \geq 0 \end{aligned}$$

Primal-dual split and penalization of coupling

$$\begin{aligned} \min_{x,y,\lambda} \quad & c^\top x + d^\top y + \rho \left[e^\top y - b^\top \lambda + x^\top C^\top \lambda \right]^- \\ \text{s.t.} \quad & Ax + By \geq a, \quad Cx + Dy \leq b \\ & D^\top \lambda = e, \quad \lambda \geq 0 \end{aligned}$$

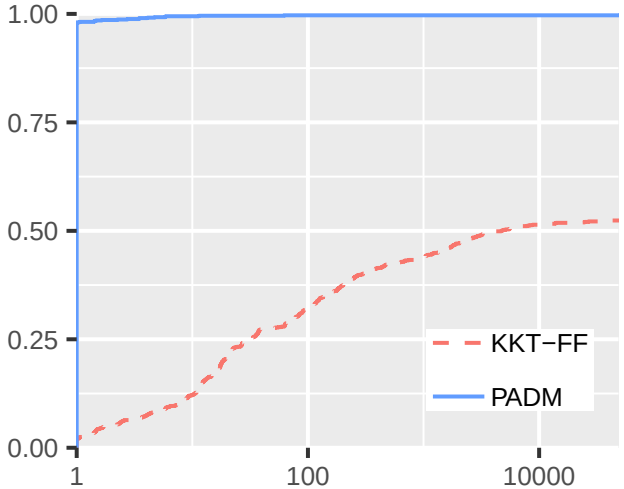
Smoothness of penalty function comes for free

Linear Bilevel Problems: Instances

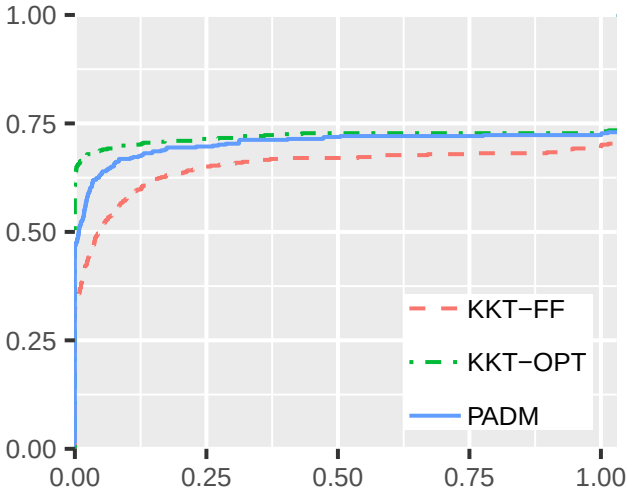
Instance Set	Ref.	Size	Constr. Structure
CLIQUE	Fischetti, Monaci, et al. (2018)	60	0/100
GENERALIZED	Fischetti, Monaci, et al. (2018)	90	50/50
GK	Fischetti, Ljubić, et al. (2019a)	33	0/100, 50/50, 100/0
IMKP	Fischetti, Ljubić, et al. (2019b)	144	0/100, 50/50, 100/0
INT0SUM	Ralphs (2019)	8	50/50
INTER-CLIQUE	Tang et al. (2015)	80	0/100
INTER-FIRE	Baggio et al. (2016)	72	0/100
KP	Fischetti, Monaci, et al. (2018)	450	0/100
MIPLIB	Fischetti, Ljubić, et al. (2016)	60	0/100
MIPLIB2010	Koch et al. (2011)	261	0/100, 50/50, 100/0
MIPLIB2017	MIPLIB 2017 (2018)	654	0/100, 50/50, 100/0
OR	Chu and Beasley (1998)	810	—
XUWANG	Xu and Wang (2014)	101	50/50
XUWANG-LARGE	Fischetti, Ljubić, et al. (2017)	60	50/50

After filtering: 924 LP-LP instances

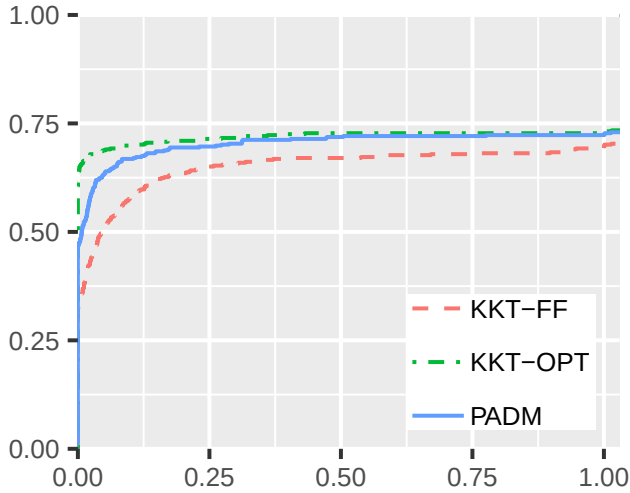
Linear Bilevel Problems: Run Time



Linear Bilevel Problems: Optimality Gaps



Linear Bilevel Problems: Optimality Gaps & Run Times



Solver	Mean	Med
PADM	5.6	0.1
KKT-FF	142.9	50.3
KKT-OPT	570.9	900.0

Overview

Mixed-Integer Nonlinear Optimization

Some Famous Alternating Methods

- Feasibility Pump
- k-Means Clustering

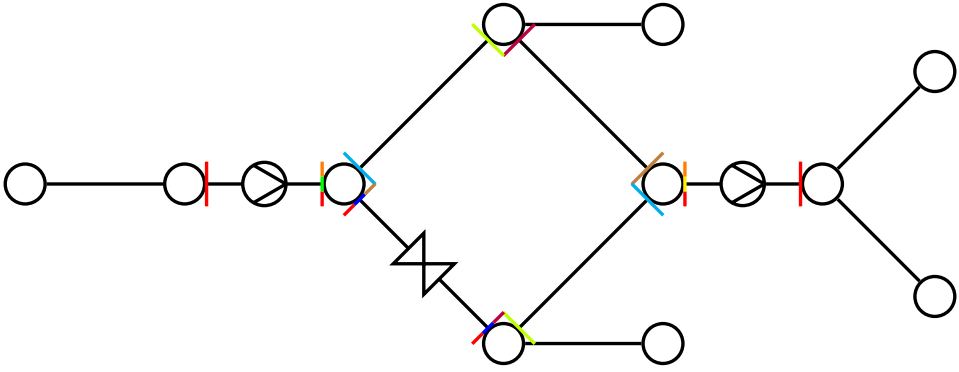
Penalty Alternating Direction Methods

Applications

- Feasibility Pump
- Bilevel Optimization
- Gas Transport Networks

Outlook

Gas Transport Networks: Test Instance



Decompositions of GasLib-11

Active (red), Valve (blue), Left Arm (green), Right Arm (yellow), Both Arms (orange),
1 o'clock (purple), 2 o'clock (brown), 4 o'clock (cyan), 5 o'clock (lime)

Gas Transport Networks: Test Instance

Decomposition	Time (s)	Objective	#outer	#inner
No decomp.	1001.46	265.60613	–	–
Active	42.77	0.58586	7	32
Valve	11.40	0.00294	1	2
Both Arms	163.71	0.00802	3	12
1 o'clock	63.23	0.00189	2	6
2 o'clock	51.34	0.01047	3	11
4 o'clock	18.87	0.00013	1	2
5 o'clock	19.58	0.00004	1	2

Outlook

Well … that’s it

Summary

- Many MINLPs possess a block-separable structure
  - Graph-induced, time-dependent models, primal-dual structures, nonlinearities vs. mixed-integer aspects

Well ... that's it

Summary

- Many MINLPs possess a block-separable structure
  - Graph-induced, time-dependent models, primal-dual structures, nonlinearities vs. mixed-integer aspects
- (P)ADM seems to be a one-fits-many approach
  - Gas networks, bilevel optimization, general MI(N)LPs, supply chain networks, best subset selection, portfolio optimization, clustering, quantile minimization problems

Well … that's it

Summary

- Many MINLPs possess a block-separable structure
  - Graph-induced, time-dependent models, primal-dual structures, nonlinearities vs. mixed-integer aspects
- (P)ADM seems to be a one-fits-many approach
  - Gas networks, bilevel optimization, general MI(N)LPs, supply chain networks, best subset selection, portfolio optimization, clustering, quantile minimization problems
- Classic theory can be extended to the penalty case

Well … that's it

Summary

- Many MINLPs possess a block-separable structure
  - Graph-induced, time-dependent models, primal-dual structures, nonlinearities vs. mixed-integer aspects
- (P)ADM seems to be a one-fits-many approach
  - Gas networks, bilevel optimization, general MI(N)LPs, supply chain networks, best subset selection, portfolio optimization, clustering, quantile minimization problems
- Classic theory can be extended to the penalty case
- Excellent numerical results
  - Extremely fast and very good gaps
  - A "good" decomposition is key

Well … that's it

Summary

- Many MINLPs possess a block-separable structure
  - Graph-induced, time-dependent models, primal-dual structures, nonlinearities vs. mixed-integer aspects
- (P)ADM seems to be a one-fits-many approach
  - Gas networks, bilevel optimization, general MI(N)LPs, supply chain networks, best subset selection, portfolio optimization, clustering, quantile minimization problems
- Classic theory can be extended to the penalty case
- Excellent numerical results
  - Extremely fast and very good gaps
  - A "good" decomposition is key

Outlook

- What's the reason for the small gaps (often even optimal)?
- Is there a canonical dual bounding scheme?
- How to decompose in general? Can we "learn" good decompositions?

Well … that's it

Summary

- Many MINLPs possess a block-separable structure
  - Graph-induced, time-dependent models, primal-dual structures, nonlinearities vs. mixed-integer aspects
- (P)ADM seems to be a one-fits-many approach
  - Gas networks, bilevel optimization, general MI(N)LPs, supply chain networks, best subset selection, portfolio optimization, clustering, quantile minimization problems
- Classic theory can be extended to the penalty case
- Excellent numerical results
  - Extremely fast and very good gaps
  - A "good" decomposition is key

Outlook

- What's the reason for the small gaps (often even optimal)?
- Is there a canonical dual bounding scheme?
- How to decompose in general? Can we "learn" good decompositions?

Please use it and answer the three questions!