Efficiently Solving Stochastic Mixed-Integer Problems combining Gauss-Siedel and Penalty-Based methods

Fabricio Oliveira
joint work with:
Brian Dandurand
Jeffrey Christiansen
Prof Andrew Eberhard

Mathematical Sciences
School of Science, RMIT University

Introduction

We are interested in solving problems of the form

$$\begin{aligned} \zeta^{SIP} := \; & \min_{x,y} \; c^\top x + \sum_{s \in S} p_s (q_s^\top y_s) \\ \text{s.t.: } & x \in X \\ & y_s \in Y_s(x), \; \forall s \in S, \end{aligned}$$

Introduction

We are interested in solving problems of the form

$$\zeta^{SIP} := \min_{x,y} c^\top x + \sum_{s \in S} p_s(q_s^\top y_s)$$
$$\text{s.t.: } x \in X$$
$$y_s \in Y_s(x), \ \forall s \in S,$$

$x \in \mathbb{R}^{n_x}, y \in \mathbb{R}^{n_y \times |S|}$ are *decision variables*;

Introduction

We are interested in solving problems of the form

$$
\begin{aligned}
\zeta^{SIP} := \; & \min_{x,y} \; c^\top x + \sum_{s \in S} p_s(q_s^\top y_s) \\
\text{s.t.: } & x \in X \\
& y_s \in Y_s(x), \; \forall s \in S,
\end{aligned}
$$

$x \in \mathbb{R}^{n_x}, y \in \mathbb{R}^{n_y \times |S|}$ are *decision variables*;

$c \in \mathbb{R}^{n_x}$ and $q \in \mathbb{R}^{n_y \times |S|}$ are input *parameters*, and $p_s \in \mathbb{R}^{|S|}$ scenario probabilities;

Introduction

We are interested in solving problems of the form

$$\zeta^{SIP} := \min_{x,y} \; c^\top x + \sum_{s \in S} p_s(q_s^\top y_s)$$
$$\text{s.t.: } x \in X$$
$$y_s \in Y_s(x), \; \forall s \in S,$$

$x \in \mathbb{R}^{n_x}, y \in \mathbb{R}^{n_y \times |S|}$ are *decision variables*;

$c \in \mathbb{R}^{n_x}$ and $q \in \mathbb{R}^{n_y \times |S|}$ are input *parameters*, and $p_s \in \mathbb{R}^{|S|}$ scenario probabilities;

Sets $X \subset \mathbb{R}^{n_x}$ and $Y_s(x) \subset \mathbb{R}^{n_y \times |S|}$ define *linear constraints and integrality restrictions* on x and y.

Deterministic equivalent (re)formulation

An alternative equivalent formulation allow us to exploit the block-angular structure that this problem naturally presents.

Deterministic equivalent (re)formulation

An alternative equivalent formulation allow us to exploit the block-angular structure that this problem naturally presents.

$$\begin{aligned}
\zeta^{SIP} := \quad & \min_{x,y,z} \sum_{s \in S} p_s(c^\top x_s + q_s^\top y_s) \\
& \text{s.t.: } x_s - z = 0, \ \forall s \in S \ (\text{NAC}) \\
& \qquad x_s \in X, \ \forall s \in S \\
& \qquad y_s \in Y_s(x_s), \ \forall s \in S.
\end{aligned}$$

Deterministic equivalent (re)formulation

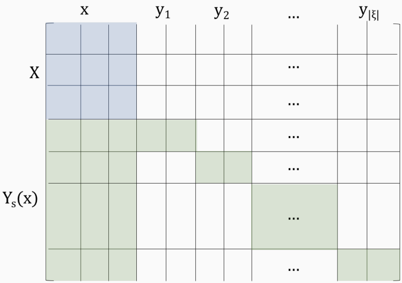
An alternative equivalent formulation allow us to exploit the block-angular structure that this problem naturally presents.

$$\zeta^{SIP} := \min_{x,y,z} \sum_{s \in S} p_s(c^\top x_s + q_s^\top y_s)$$

$$\text{s.t.: } x_s - z = 0, \ \forall s \in S \text{ (NAC)}$$

$$x_s \in X, \ \forall s \in S$$

$$y_s \in Y_s(x_s), \ \forall s \in S.$$



Lagrangian Relaxation

The most straightforward approach is to relax NAC. First, let

$$\zeta^{LR}(\omega) := \min_{x,y,z} \sum_{s \in S} p_s L_s(x_s, y_s, z, \omega)$$
$$\text{s.t.: } x_s \in X, \ \forall s \in S$$
$$y_s \in Y_s(x_s), \ \forall s \in S,$$

where $\omega := (\omega_s)_{s \in S} \in \Omega := \{\omega \mid \sum_{s \in S} p_s^\top \omega_s = 0\}$ and

$$L_s(x_s, y_s, z, \omega_s) := c^\top x_s + q_s^\top y_s + \omega_s^\top (x_s - z), \ \forall s \in S$$
$$:= (c + \omega_s)^\top x_s + q_s^\top y_s, \ \forall s \in S.$$

Lagrangian Relaxation

The most straightforward approach is to relax NAC. First, let

$$\zeta^{LR}(\omega) := \min_{x,y,z} \sum_{s \in S} p_s L_s(x_s, y_s, z, \omega)$$
$$\text{s.t.: } x_s \in X, \ \forall s \in S$$
$$y_s \in Y_s(x_s), \ \forall s \in S,$$

where $\omega := (\omega_s)_{s \in S} \in \Omega := \{\omega \mid \sum_{s \in S} p_s^\top \omega_s = 0\}$ and

$$L_s(x_s, y_s, z, \omega_s) := c^\top x_s + q_s^\top y_s + \omega_s^\top (x_s - z), \ \forall s \in S$$
$$:= (c + \omega_s)^\top x_s + q_s^\top y_s, \ \forall s \in S.$$

Then, we focus on solving the Lagrangian Dual

$$\zeta^{LD} := \max_{\omega \in \Omega} \zeta^{LR}(\omega).$$

Introduction

Due to the presence of integer variables, we typically have a duality gap i.e. $\zeta^{LD} < \zeta^{SIP}$.

Due to the presence of integer variables, we typically have a duality gap i.e. $\zeta^{LD} < \zeta^{SIP}$.

Augmented Lagrangian Relaxation

Alternatively, we can define the (augmented) Lagrangian as

$$\begin{aligned} \zeta_\rho^{LR+}(\omega) := \ & \min_{x,y,z} \sum_{s \in S} p_s L^s(x_s, y_s, z, \omega) + \psi_\rho^s(x_s - z) \\ & \text{s.t.: } x_s \in X, \ \forall s \in S \\ & \quad\quad y_s \in Y_s(x_s), \ \forall s \in S, \end{aligned}$$

where $\omega = (\omega_s)_{s \in S} \in \Omega := \{\omega \mid \sum_s p_s \omega_s = 0\}$ and $\psi_\rho^s : \mathbb{R}^{n_x} \mapsto \mathbb{R}$ is an appropriate penalty function that depends on the penalty parameter ρ.

Augmented Lagrangian Relaxation

Most common choice: $\psi_\rho^s(u_s) := \frac{\rho}{2}||u_s||_2^2$ for each $s \in S$, which provides *smoothness* to the original Lagrangian dual function.

Introduction

Augmented Lagrangian Relaxation

Most common choice: $\psi_\rho^s(u_s) := \frac{\rho}{2}||u_s||_2^2$ for each $s \in S$, which provides *smoothness* to the original Lagrangian dual function.

Main motivations

[Feizollahi et al., 2016, Boland and Eberhard, 2015] have shown that the augmented Lagrangian dual is capable of asymptotically achieving zero duality gap only if the weight ρ is allowed to go to infinity.

Augmented Lagrangian Relaxation

Most common choice: $\psi^s_\rho(u_s) := \frac{\rho}{2}||u_s||_2^2$ for each $s \in S$, which provides *smoothness* to the original Lagrangian dual function.

Main motivations

[Feizollahi et al., 2016, Boland and Eberhard, 2015] have shown that the augmented Lagrangian dual is capable of asymptotically achieving zero duality gap only if the weight ρ is allowed to go to infinity.

However, it is possible to circumvent this drawback if the augmentation *uses a norm as the penalty function*. In this case, the theory suggests that it is possible to attain strong duality for a finite value of ρ.

The Framework: Combining Penalty-based and Gauss-Siedel Approaches

Combining Penalty-based and Gauss-Siedel Approaches

How it all started...

Given discrepancy a vector $u := (u_s)_{s \in S} \in \prod_{s \in S} \mathbb{R}^{n_x}$, we define for each scenario s the penalty function

$$\psi^s_\rho(u_s) := \underline{\rho}_s^\top [u_s]^- + \overline{\rho}_s^\top [-u_s]^-,$$

where $\rho = (\underline{\rho}_s, \overline{\rho}_s)_{s \in S} \in \mathbb{R}^{2n_x|S|}_{>0}$ and $[v]^- := -\min\{0, v\}$ (performed component wise), where in this case $v \in \mathbb{R}^{n_x}$. Then we define

$$\psi_\rho(u) := \sum_{s \in S} \psi^s_\rho(u_s) = \left(\sum_{s \in S} \underline{\rho}_s^\top [u_s]^- + \sum_{s \in S} \overline{\rho}_s^\top [-u_s]^- \right).$$

Combining Penalty-based and Gauss-Siedel Approaches

The definition of a good penalty function is tied to this result:

Theorem 1 (based on [Feizollahi et al., 2016, Thm. 5])

Consider a feasible MIP problem given whose problem data is formed from rational entries and with its optimal value bounded. If $\psi : \prod_{s \in S} \mathbb{R}^{n_x} \mapsto \mathbb{R}$ is a summed augmenting function $\psi(u) := \sum_{s \in S} \psi^s_\rho(u_s)$ for problem $\zeta^{LR+}_\rho(\omega)$ such that

1. $\psi(0) = 0$
2. $\psi(u) \geq \delta > 0, \forall u \notin V$
3. $\psi(u) \geq \gamma ||u||_\infty, \forall u \in V$

for some open neighbourhood V of 0, and positive scalars $\delta, \gamma > 0$, then there exists a finite ρ such that $\zeta^{LD+}_\rho = \zeta^{LR+}_\rho(\omega) = \zeta^{SIP}$, for any $\omega \in \Omega$.

Combining Penalty-based and Gauss-Siedel Approaches

We propose a class of augmenting functions based on the use of positive basis [Davis, 1954].

Definition 4

We say a set of vectors $\{n_1, \ldots,\ n_l\}$, where $m + 1 \leq l \leq 2m$, is a positive basis for $\mathbb{R}^m$ if and only if every $u \in \mathbb{R}^m$ can be expressed as a non-negative combination of these vectors, i.e., there exists $\alpha_i \geq 0$ for $i = 1, \ldots, l$ for which $u = \sum_{i=1}^{l} \alpha_i n_i$.

Combining Penalty-based and Gauss-Siedel Approaches

We propose a class of augmenting functions based on the use of positive basis [Davis, 1954].

Definition 4

We say a set of vectors $\{n_1, \ldots, n_l\}$, where $m + 1 \leq l \leq 2m$, is a positive basis for $\mathbb{R}^m$ if and only if every $u \in \mathbb{R}^m$ can be expressed as a non-negative combination of these vectors, i.e., there exists $\alpha_i \geq 0$ for $i = 1, \ldots, l$ for which $u = \sum_{i=1}^{l} \alpha_i n_i$.

The relationship between *positive basis* and *norms* is somewhat straightforward, being the latter a more general form.

$$\psi_\infty(u) := \|u\|_\infty = \max_{i=1,\ldots,m} \{\pm e_i^\top u\}, \text{ and} \tag{1}$$

$$\psi_1(u) := \|u\|_1 = \sum_{i=1}^{m} \max\{+e_i^\top u, 0\} + \sum_{i=1}^{m} \max\{-e_i^\top u, 0\} \tag{2}$$

where e_i, $i = 1, \ldots, m$ are unit vectors $\mathbb{R}^m$ with entry i equals 1 and all other 0.

Combining Penalty-based and Gauss-Siedel Approaches

The reason for using positive basis is because it encompasses a set of properties that allows us to arrive at the following result:

Corollary 8 (highlights)

For any positive basis $N := \{\mathbf{n}_1, \ldots, \mathbf{n}_l\}$, the optimal value of the augmented Lagrangian dual problem ζ_ρ^{LD+} using an augmenting function of the form of

$$\psi_1^N(u) := \sum_{i=1}^{l} \max\{\mathbf{n}_i^\top u, 0\}$$

is equal to the optimal value of ζ^{SIP} for some finite ρ; that is,

$$\zeta_\rho^{LD+} = \zeta_\rho^{LR+}(\omega) = \zeta^{SIP}$$

for any $\omega \in \Omega$.

Combining Penalty-based and Gauss-Siedel Approaches

Cor 8 gives us hope because it is easy to show that, picking a positive basis to be $N_\rho = \{\overline{\rho}_{s,i} e_{i+(s-1)n_x} \mid s \in S, i \in \{1, \ldots, n_x\}\} \cup \{-\underline{\rho}_{s,i} e_{i+(s-1)n_x} \mid s \in S, i \in \{1, \ldots, n_x\}\}$, we have

$$\begin{aligned}
\psi_1^{N_\rho}(u) &= \sum_{s \in S} \sum_{i=1,\ldots,n_x} \overline{\rho}_{s,i} \max\{0, u_{s,i}\} \\
&\quad + \sum_{s \in S} \sum_{i=1,\ldots,n_x} \underline{\rho}_{s,i} \max\{0, -u_{s,i}\} \\
&= \left(\sum_{s \in S} \underline{\rho}_s^\top [u_s]^- + \sum_{s \in S} \overline{\rho}_s^\top [-u_s]^- \right) \\
&= \sum_{s \in S} \psi_\rho^s(u) = \psi_\rho(u)
\end{aligned}$$

Combining Penalty-based and Gauss-Siedel Approaches

Building upon what we have so far, we end-up with the following augmented Lagrangian problem

$$
\begin{aligned}
\zeta_\rho^{LR+}(\omega): \ \min_{x,y,z} & \sum_{s\in S} p_s([c+\omega]^\top x_s + q_s^\top y_s) \\
& + \sum_{s\in S} \underline{\rho}_s^\top [x_s - z]^- + \sum_{s\in S} \overline{\rho}_s^\top [z - x_s]^- \\
\text{s.t.: } & x_s \in X, \ \forall s \in S \\
& y_s \in Y_s(x_s), \ \forall s \in S,
\end{aligned}
$$

which, with the help of Thm 1 and Cor 8, can be written as

$$
\begin{aligned}
\zeta_\rho^{LR+}(0): \ \min_{x,y,z} & \sum_{s\in S} p_s(c^\top x_s + q_s^\top y_s) + \sum_{s\in S} \underline{\rho}_s^\top [x_s - z]^- + \sum_{s\in S} \overline{\rho}_s^\top [z - x_s]^- \\
\text{s.t.: } & x_s \in X, \ \forall s \in S \\
& y_s \in Y_s(x_s), \ \forall s \in S.
\end{aligned}
$$

Nonlinear Block Gauss-Siedel Method

In principle, one can only solve ζ_{ρ}^{LR+} individually scenario wise if the variable z has a known (or fixed) value. This motivates the application of nonlinear block Gauss-Siedel (GS) approach.

Combining Penalty-based and Gauss-Siedel Approaches

Nonlinear Block Gauss-Siedel Method

In principle, one can only solve ζ_{ρ}^{LR+} individually scenario wise if the variable z has a known (or fixed) value. This motivates the application of nonlinear block Gauss-Siedel (GS) approach.

Suppose we want to solve:

$$\zeta := \min_{x,z} f(x,z)$$
$$\text{s.t.: } x \in X, z \in Z,$$

with f is convex, but not necessarily differentiable and sets X and Z closed, but not necessarily convex.

The idea is to use the following algorithmic setting.

Algorithm 1 A block GS method

initialise $(x^0, z^0) \in X \times Z$
for $k = 1, \dots, k_{\text{max}}$ **do**
    $x^k \leftarrow \text{argmin}_x \left\{ f(x, z^{k-1}) : x \in X \right\}$
    $z^k \leftarrow \text{argmin}_z \left\{ f(x^k, z) : z \in Z \right\}$
    $k \leftarrow k + 1$
end for
return $(x^{k_{\text{max}}}, z^{k_{\text{max}}})$

Proposition 11

For problem ζ, let f be continuous and bounded from below, and let X and Z be compact. Then the limit points (x^*, z^*) of the sequence $\{(x^k, z^k)\}$ generated by iterations of Algorithm 1 are partial minima, i.e.

$$f(x^*, z^*) \leq f(x, z^*), \; \forall x \in X,$$
$$f(x^*, z^*) \leq f(x^*, z), \; \forall z \in Z.$$

Combining Penalty-based and Gauss-Siedel Approaches

Proposition 11

For problem ζ, let f be continuous and bounded from below, and let X and Z be compact. Then the limit points (x^*, z^*) of the sequence $\{(x^k, z^k)\}$ generated by iterations of Algorithm 1 are partial minima, i.e.

$$f(x^*, z^*) \leq f(x, z^*), \ \forall x \in X,$$
$$f(x^*, z^*) \leq f(x^*, z), \ \forall z \in Z.$$

The tricky bit: In the more general setting where f is non-differentiable and/or X and Z are non-convex, a partial minimum is not necessarily a global, or even a local, minimum.

The algorithm

Implementation aspects

Let $\phi^\rho(x, y, z, \rho) := \sum_{s \in S} p_s \phi_s^\rho(x_s, y_s, z, \mu_s)$, where

$$\phi_s^\rho(x_s, y_s, z, \mu_s) := \left\{ c^\top x_s + q_s^\top y_s + \underline{\mu}_s^\top [x_s - z]^- + \overline{\mu}_s^\top [z - x_s]^- \right\}$$

and $(\underline{\mu}_s, \overline{\mu}_s) := (\frac{1}{p_s} \underline{\rho}_s, \frac{1}{p_s} \overline{\rho}_s)$ for each $s \in S$.

For a given $\rho_s^k = (\underline{\rho}_s^k, \overline{\rho}_s^k)_{s \in S}$ and an initial $z^{0,0}$, we iterate between subproblems $l = 0, 1, \ldots, l_{\max}$:

$$\begin{aligned}
(x^{k,l+1}, y^{k,l+1})_{s \in S} \leftarrow \underset{x,y}{\operatorname{argmin}} \ & \phi^\rho(x, y, z^{k,l}, \rho^k) \\
\text{s.t.:} \ & x_s \in X, \ \forall s \in S \\
& y_s \in Y_s(x_s), \ \forall s \in S; \\
z^{k,l+1} \leftarrow \underset{z}{\operatorname{argmin}} \ & \phi^\rho(x^{k,l+1}, y^{k,l+1}, z, \rho^k),
\end{aligned}$$

followed by $l = l + 1$ and successive repetition until *partial convergence* is approximately achieved.

Proposed Algorithm

Algorithm 2 Alternating direction method for SMIP

initialise $\rho^0 = (\underline{\rho}^0, \overline{\rho}^0), \hat{z}^0, \epsilon, \gamma, \beta, l_{\max}, k_{\max}$
for $k = 1, \ldots, k_{\max}$ **do**
  $z^{k,0} \leftarrow \hat{z}^{k-1}$
  for $l = 1, \ldots, l_{\max}$ **do**
    for $s \in S$ **do**
      $(x_s^{k,l}, y_s^{k,l}) \leftarrow \text{argmin}_{x,y} \left\{\phi^{\rho,k}(x_s, y_s, z^{k,l-1}, \rho^k) : x_s \in X, y_s \in Y_s(x_s)\right\}$
    end for
    $z^{k,l} \leftarrow \text{argmin}_z \ \phi^{\rho,k}(x^{k,l}, y^{k,l}, z, \rho^k)$
    $\Gamma \leftarrow \phi^{\rho,k}(x^{k,l-1}, y^{k,l-1}, z^{k,l-1}, \rho^k) - \phi^{\rho,k}(x^{k,l}, y^{k,l}, z^{k,l}, \rho^k)$
    if $\Gamma \leq \epsilon$ **or** $l = l_{\max}$ **then**
      $(\hat{x}_s^k, \hat{y}_s^k) \leftarrow (x_s^{k,l}, y_s^{k,l})$ for all $s \in S$
      $\hat{z}^k \leftarrow z^{k,l}$
      break
    end if
    $l \leftarrow l + 1$
  end for
  if $||\hat{x}^k - \hat{z}^k||_2^2 \leq \epsilon$ **or** $k = k_{\max}$ **then**
    return $((\hat{x}_s^k, \hat{y}_s^k)_{s\in S}, \hat{z}^k)$
  else
    $\underline{\rho}_s^k = \underline{\rho}_s^{k-1} + \gamma[\hat{x}_s^k - \hat{z}^k]^-$ for all $s \in S$
    $\overline{\rho}_s^k = \overline{\rho}_s^{k-1} + \gamma[\hat{z}^k - \hat{x}_s^k]^-$ for all $s \in S$
end if
$k \leftarrow k + 1$
end for

Computational Experiments and Numerical Results

Computational Experiments

Experimental setting

- Three classes of problems from literature: SSLP, CAP, and DCAP;
- 50 random instances for 2 problems in each class: a total of 300 instances;
- The performance was compared Progressive Hedging (PH) method (which combines GS and $\psi^s_\rho(u_s) := \frac{\rho}{2}||u_s||_2^2$);
- instances were solved with three parameter choices for PH (different choices of ρ) and 12 combinations of parameter choices for PBGS (different choices of ρ^0, β and γ);
- Hardware: Intel i7 CPU with 3.40GHz and 8GB of RAM; software: AIMMS 3.14 using CPLEX 12.6.3.

Experimental Results

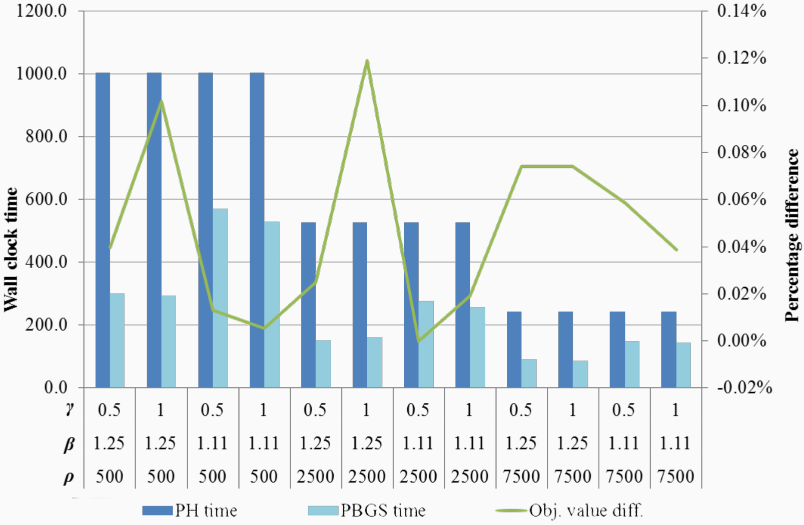


Figure 1: CAP111

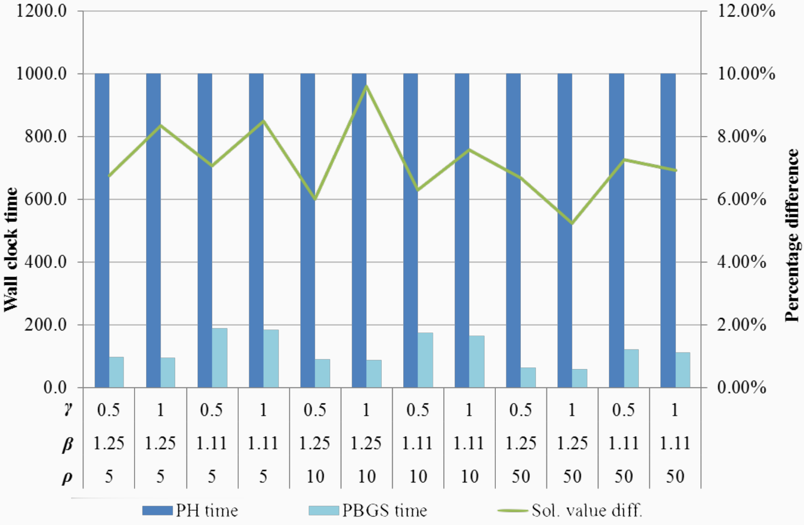


Figure 2: DCAP342

Experimental Results

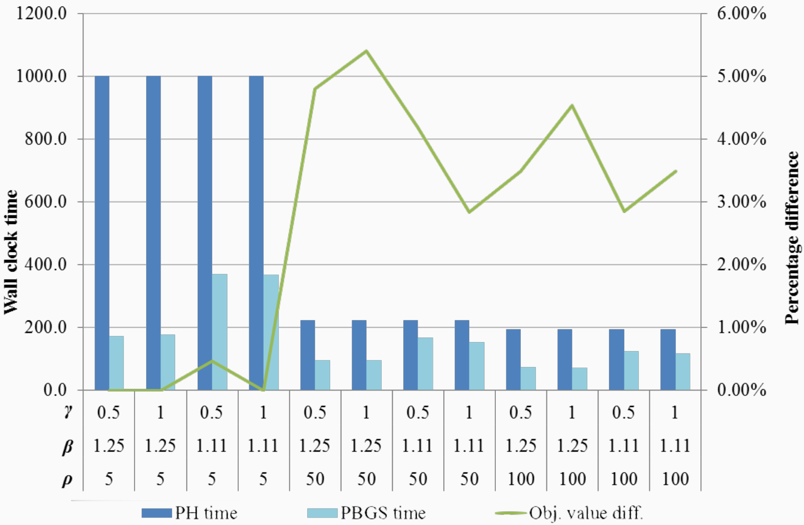


Figure 3: SSLP10-50

Take-aways: Conclusions and Future Directions

Conclusions 1/2

Key points

- A competitive approach in terms of computational efficient;
- Readily amenable to parallelisation, which is a key point for dealing with large-scale SMIPs;
- Theoretical results are encouraging regarding alternative frameworks for defining penalty functions.

Conclusions 1/2

Key points

- A competitive approach in terms of computational efficient;
- Readily amenable to parallelisation, which is a key point for dealing with large-scale SMIPs;
- Theoretical results are encouraging regarding alternative frameworks for defining penalty functions.

Future directions

- Extensions of the block Gauss-Seidel approach into non-smooth non-separable problems.
- Better understanding of how updates of the penalty coefficients can improve the block Gauss-Seidel behaviour.
- Evaluate the proposed approach in contexts other than SMIPs and considering its extension to the multi-stage case.

Conclusions 2/2

Currently underway

A few spin-off ideas that we have been working on related to this topic.

- Instead of using a nonsmooth penalty term, rely on convexification ideas to obtain convergence (Preliminary results in [Boland et al., 2016]).
- Borrow theoretical tools from Bundle methods literature to stabilise the selection of penalty parameter (Preliminary results in [Boland et al., 2017]).
- Ideas for guaranteeing convergences of block Gauss-Siedel via smoothing of norm-like penalty functions (work in progress).

References I

Main reference: [Oliveira et al., 2016], available at

`optimizationonline.org/DB_FILE/2016/10/5687.pdf`

Boland, N., Christiansen, J., Dandurand, B., Eberhard, A., Linderoth, J., Luedtke, J., and Oliveira, F. (2016).
Combining progressive hedging with a frank-wolfe method to compute lagrangian dual bounds in stochastic mixed-integer programming.
submitted to SIOPT.

References II

Boland, N., Christiansen, J., Dandurand, B., and Oliveira, F. (2017).
A parallelizable augmented lagrangian method applied to large-scale non-convex-constrained optimization problems.
submitted to Math. Programming.

Boland, N. L. and Eberhard, A. C. (2015).
On the augmented Lagrangian dual for integer programming.
Math. Program., 150(2, Ser. A):491–509.

References III

- Davis, C. (1954).
  Theory of positive linear dependence.
  American Journal of Mathematics, 76(4):733–746.
- Feizollahi, M. J., Ahmed, S., and Sun, A. (2016).
  Exact augmented Lagrangian duality for mixed integer linear programming.
  Mathematical Programming, pages 1–23.
- Oliveira, F., Christiansen, J., Dandurand, B., and Eberhard, A. (2016).
  Combining penalty-based and gauss-seidel methods for solving stochastic mixed-integer problems.
  submitted to ITOR.