Noname manuscript No.
(will be inserted by the editor)

...

...

Received: date / Accepted: date

Abstract Keywords augmented Lagrangian method · proximal bundle method · nonlinear block Gauss-Seidel method · simplicial decomposition method · parallel computing

Mathematics Subject Classification (2000) 90-08, 90C06, 90C11, 90C15, 90C25, 90C26, 90C30, 90C46

1 Introduction and Background

The problem of interest has the form

$$f^* := \min_{x,z} \left\{ f(x) : Qx = z, x \in X, z \in Z \right\}, \tag{1}$$

where f is convex and continuously differentiable, $Q \in \mathbb{R}^{q \times n}$ is a block-diagonal matrix determining linear constraints $Qx = z$, $X \subset \mathbb{R}^n$ is a closed and bounded set, and $Z \subset \mathbb{R}^q$ is a linear subspace. The vector $x \in X$ of decision variables is derived from the original decisions associated with a problem, while the vector $z \in Z$ of auxiliary variables are introduced to effect a decomposable structure in (1). The block diagonal components of Q are denoted $Q_i \in \mathbb{R}^{q_i \times n_i}$, $i = 1, \ldots, m$. Problem (1) is general enough to subsume, for example, the split-variable deterministic reformulation of a stochastic optimization problem with potentially multiple stages, as defined, for example, in [1], while it can also model the case where f is nonlinear (and convex) and/or X is any compact (but not necessarily convex) set.

We develop a branch-and-bound approach to solving (1) based on solving Lagrangian relaxations due to relaxing $Qx = z$. At each node, we use the recently developed algorithm SDM-GS-ALM to solve two equivalent characterizations of the Lagrangian dual problem. The Lagrangian dual function is defined

$$\phi(\omega) := \min_{x,z} \left\{ f(x) + \omega^\top (Qx - z) : x \in X, z \in Z \right\}.$$

Under the dual feasibility condition $\omega \in Z^\perp$, we have

$$\phi(\omega) = \min_{x} \left\{ f(x) + \omega^\top Qx : x \in X \right\}.$$

When f is not linear, we also need to refer to the convexified Lagrangian dual funciton $\phi^C(\omega) := \min_{x,z} \left\{ f(x) + \omega^\top (Qx - z) : x \in \text{conv}(X), z \in Z \right\}$, for which analogously, we have

$$\phi^C(\omega) = \min_{x} \left\{ f(x) + \omega^\top Qx : x \in \text{conv}(X) \right\} \tag{2}$$

when $\omega \in Z^\perp$. Optimal solutions to problem (2) will be referred to generically as $\widehat{x}(\omega)$.

Remark 1 Under the assumption that $\text{conv}(X)$ is not known beforehand by any characterization, direct evaluation of ϕ^C or any of its subgradients at any $\omega \in Z^\perp$ is not possible. This dual function is only treated indirectly in the following algorithms.

The (conviexified) Lagrangian dual problem is:

$$\phi_C^* := \max_{\omega \in \Omega} \phi^C(\omega). \tag{3}$$

Its well-known primal characterization is

$$\phi_C^* = \min_{x,z} \left\{ f(x) : Qx = z, x \in \text{conv}(X), z \in Z \right\}, \tag{4}$$

It is straightforward from the definitions that $\phi^C(\omega) \leqslant \phi(\omega)$ for all dual feasible $\omega \in Z^\perp$. In the case when f is linear, we have $\phi^C(\omega) = \phi(\omega)$ for all $\omega \in Z^\perp$ and so $\phi^* = \phi_C^*$. But in the general case where f is nonlinear, the dual (**??**) can be "weaker" than (**??**), where $\phi_C^* < \phi^*$ can occur, which we see in the following example. Let $f : \mathbb{R}^2 \mapsto \mathbb{R}$ be defined by $f(x) = (x_1 - 0.5)^2 + (x_2 - 0.5)^2$, $X = \{0,1\} \times \{0,1\}$, and let $Qx = z$ be defined to model the constraints $x_1 - z_1 = 0$ and $x_2 - z_2 = 0$ where $Z = \{(z_1, z_2) : z_1 = z_2\} \subset \mathbb{R}^2$. We see trivially that $\phi_C^* = 0$, which is verified with the saddle point $x_1^* = x_2^* = z_1^* = z_2^* = 0.5$ and $\omega^* = (0,0)$. However, $\phi^* = 0.5$, which is verified with either of the saddle points $x_1^* = x_2^* = z_1^* = z_2^* = 0$ and $\omega^* = (0,0)$, or $x_1^* = x_2^* = z_1^* = z_2^* = 1$ and $\omega^* = (0,0)$. Thus, $\phi_C^* < \phi^*$.

Dual optimal solutions to (3), when they exist, are generically denoted by ω^*, and corresponding optimal primal solution verifying the value $\phi_C(\omega^*)$ is denoted by $\widehat{x}(\omega^*)$. The optimal solutions to the second characterization (4), when they exist, are denoted (x^*, z^*). We shall denote the set

$$X_Z := \{ x \in X : \exists\, z \in Z \text{ s.t. } Qx = z \}.$$

It is straightforward to show that $\text{conv}(X_Z) = \{x \in \text{conv}(X) : \exists\, z \in Z \text{ s.t. } Qx = z\}$ under the assumption that Z is convex.

We shall use both characterizations (3) and (4) to inform different ways of branching on variables and branching values.

1. Viewing the node subproblems as instances of problem (4), we have optimal solutions (x^*, z^*) to problem (4) that satisfy

$$(x^*, z^*) \in \{(x, z) : x \in \text{conv}(X), z \in Z, Qx = z\},$$

but $x^* \notin X$ in the case of strict duality gap $\phi_C^* < f^*$. Thus we project x^* onto X, denoting the projection by $P_X(x^*)$. (When X is defined by polyhedral and mixed-integer constraints, this projection can just take the form of rounding the components of x with integer restrictions to their nearest integer values.) We would consider the discrepancies $x^* - \mathrm{P}_X(x^*)$ and their dispersions. Under this paradigm, bounding is applied to X.

2. Viewing the node subproblems as instances of problem (3), assuming we have a dual optimal solution ω^*, we have corresponding sets of primal certificate solutions $\widehat{x}(\omega^*)$. For these solutions, we have $Q\widehat{x}(\omega^*) \cap Z = \varnothing$ in the case of duality gap $\phi_C^* < f^*$. We consider one of two projections:
   (a) Take the projection $\mathrm{P}_Z(Q\widehat{x}(\omega^*))$. Conceptually, we want to accumulate the dispersions in some set-wise or point-wise fashion. The important thing to note is that under this line of thought, we do not consider the dispersion of only one $\widehat{x} \in \widehat{x}(\omega^*)$, but the entire set $\widehat{x}(\omega^*)$, which is contained in the highest dimensional face of $\text{conv}(X)$ for which z^* is in its relative interior.
   (b) Rather than projection, we use the information from x^*, z^* and ω^*. (Or, more practically, x^k, z^k and ω^k.) We still work with the set $Q\widehat{x}(\omega^k)$, whose vertex description is available from computing $\widehat{x}^t \in \widehat{x}(\omega^t)$ at each iteration $t = 1, \ldots, k$.

   Under this paradigm bounds are applied (conceptually) to Z, but in implementational practice, still on X.

Due to the large number of variables that would be branched on under the first approach, we do not consider branching based on the first approach, although we do use the dispersions from the first approach discrepancies to inform branching based on the second approach. Thus, conceptually, we branch in the ambient space of Z. Branching occurs on components z_i of z that either violate integrality constraints, or that show dispersion from some consensus value associated with other violated constraints. The constraints added due to branching are denoted $Qx \in B$. Based on this constraint, define $X_B := \{x \in X : Qx \in B\}$, so that we define, respectively,

$$f_B^* := \min_{x,z} \{f(x) : Qx = z, x \in X_B,\ z \in Z\}, \tag{5}$$

$$\phi_B^{C*} := \max_{\omega \in \Omega} \phi_B^C(\omega). \tag{6}$$

with

$$\phi_B^C(\omega) = \min_x \left\{ f(x) + \omega^\top Qx : x \in \text{conv}(X_B) \right\} \tag{7}$$

and

$$\phi_B^{C*} = \min_{x,z} \left\{ f(x) : Qx = z, x \in \operatorname{conv}(X_B),\ z \in Z \right\}, \tag{8}$$

Note that we may use either the constraints $z \in B$ or $Qx \in B$ interchangeably as the context warrants.

During each node processing of the branch-and-bound, an attempt is made to find a feasible solution to the node-specific instance of (5), which provides a finite upper bound, which we shall refer to as the *incumbent value* and denote as $v \in \mathbb{R}$. Furthermore, we use an iterative approach, referred to as the *oracle*, to solving each instance of (6) and (8) simultaneously, that is guaranteed to converge optimally. During the processing of each node, one of three things will happen:

1. The oracle terminates due to the objective value exceeding the incumbent value (fathoming due to bound), or due to subproblem infeasibility.
2. The oracle terminates due to optimality of (5) within pre-specified precision. That is $\|x^* - \mathrm{P}_X(x^*)\| + \|Qx^* - z^*\| < \epsilon$. The node is fathomed after testing whether $f_B^* < v$ and updating $v \leftarrow f_B^*$ if that is the case.
3. Otherwise, branchings (in $\mathbb{R}^m$) need to be determined.

```
Definitions:
    $\mathcal{S} := \{f, Q, X, Z, B\}$, $\mathcal{P} := \{\rho, \gamma, \epsilon, t_{max}, k_{max}\}$
    $\upsilon$ is an incumbent value, and
    $(\widetilde{x}, \widetilde{z})$ is a corresponding incumbent feasible solution for (1) for which $\upsilon = F(\widetilde{z})$
Precondition:
    It is assumed that the value of the current node has been checked to not exceed the
incumbent value $\upsilon$
function PROCESS($\mathcal{S}$, $\mathcal{P}$, $\omega^0$, $\upsilon$, $(\widetilde{x}, \widetilde{z})$)
    $\left(\widehat{x}^*, x^*, z^*, \omega^*, \breve{\phi}^*\right) \leftarrow$ bound($\mathcal{S}$, $\mathcal{P}$ $\omega^0$)
    if $\breve{\phi}^* > \upsilon$ or $\breve{\phi}^* = \infty$ then
        fathom node
        $\mathcal{B} \leftarrow \varnothing$
        return $(\upsilon, \widetilde{z}, \mathcal{B})$
    end if
    $\mathcal{B} \leftarrow$ findBranchings$(\widehat{x}^*, z^*)$
    $(\widetilde{z}, \upsilon) \leftarrow$ findFeasibleSolution$(z^*, \omega^*, \mathcal{S}, \mathcal{P})$ (Note that any solution $\widetilde{z}$ found will
satisfy $\widetilde{z} \in B$.)
    if $\mathcal{B} \neq \varnothing$ then
        node needs to branch
    else
        fathom node (Due to optimality)
    end if
    return $(\upsilon, \widetilde{z}, \mathcal{B})$
end function
function BOUND($\mathcal{S}$, $\mathcal{P}$ $\omega^0$)
    $(D^0, x^0, z^0, \omega^0, \breve{\phi}^0) \leftarrow$ Initialize$(\mathcal{S}, \mathcal{P}, \omega^0)$
    $k \leftarrow 0$, term $\leftarrow$ false
    while $\neg(\breve{\phi}^k > \upsilon$ or term) do
        $(\widehat{x}^{k+1}, x^{k+1}, z^{k+1}, \omega^{k+1}, \breve{\phi}^{k+1},$ term$) \leftarrow$ PSCG$(\mathcal{S}, \mathcal{P}, D^k, x^k, z^k, \omega^k, \breve{\phi}^k)$
        $k \leftarrow k + 1$
    end while
    return $\left(\widehat{x}^k, x^k, z^k, \omega^k, \breve{\phi}^k\right)$
end function
```

1.1 Parallel Stabilized Column Generation (PSCG)

In applying the AL method to problem (**??**), the continuous master problem for fixed $\omega \in Z^\perp$ takes the form

$$\phi_\rho^{AL}(\omega) := \min_{x,z} \{L_\rho(x, z, \omega), x \in \text{conv}(X), z \in Z\} \tag{9}$$

where the augmented Lagrangian (AL) relaxes $Qx = z$ and is defined by

$$L_\rho(x, z, \omega) := f(x) + \omega^\top Qx + \frac{\rho}{2} \|Qx - z\|_2^2 . \tag{10}$$

In the algorithm that follows, we use the following approximation $\widehat{\phi} : \mathbb{R}^q \times \mathbb{R}^n \times \mathbb{R}^q \mapsto \mathbb{R}$ of ϕ^C centered at (x^k, z^k), $k \geqslant 0$, in place the cutting plane model:

$$\widehat{\phi}(\omega, x^k, z^k) := L_\rho(x^k, z^k, \omega) + \frac{\rho}{2} \left\|Qx^k - z^k\right\|_2^2 .$$

The convex hull $\mathrm{conv}(X)$ is not known explicitly, and so ϕ^C cannot be evaluated directly. Consequently, we additionally make use of the following minorization $\breve{\phi}$ of ϕ^C that can be evaluated. For $x^k \in \mathrm{conv}(X)$, $k \geqslant 0$, define $\breve{\phi}(\omega, x^k)$ as follows:

$$\breve{\phi}(\omega, x^k) := \min_x \left\{ f(x^k) + \nabla_x f(x^k)(x - x^k) + \omega^\top Q x : x \in X \right\}. \tag{11}$$

Observe that, due to the linearity of the objective function with respect to x in (11), the use of constraint sets X and $\mathrm{conv}(X)$ are interchangeable, and so in evaluating $\breve{\phi}$, an explicit description of $\mathrm{conv}(X)$ is not required. Furthermore, from the definition of ϕ^C, the convexity of f over $\mathbb{R}^n$, and the interchangeability of X and $\mathrm{conv}(X)$ in (11), it is clear that for all $x^k \in \mathbb{R}^n$, $k \geqslant 0$, we have $\phi^C(\omega) \geqslant \breve{\phi}(\omega, x^k)$. Furthermore, when f is linear, we have $\phi^C(\omega) \equiv \breve{\phi}(\omega, x^k)$ for all x^k, $k \geqslant 0$; the two functions collapse into the same function with the centering at x^k of the latter function now irrelevant.

In [**?**], we developed an efficiently parallelizable iterative procedure to solving problem (**??**). In this paper, we incorporate this procedure as part of the bounding mechanism within a branch-and-bound approach....

For review purposes, the iterative dual procedure is given as follows.

Algorithm 1 A regular iteration of PSCG.

```
Preconditions:
    $x^k \in \mathrm{conv}(X)$, $z^k \in \mathrm{argmin}_z \left\{\|Qx - z\|^2 : z \in Z \cap B\right\}$, $\omega^k \in Z^\perp$,
    $\breve{\phi}^k = \breve{\phi}(\omega^k, x^k)$, $\left\{x^k + \alpha(\widehat{x} - x^k) : \alpha \in [0,1]\right\} \subseteq D \subseteq \mathrm{conv}(X)$ where
        $\widehat{x} \in \mathrm{argmin}_x \left\{\nabla_x L_\rho(x^k, z^k, \omega^k)(x - x^k) : x \in X\right\}$.
    $\mathcal{S} := \{f, Q, X, Z, B\}$, $\mathcal{P} := \{\rho, \gamma, \epsilon, t_{max}, k_{max}\}$
function PSCG($\mathcal{S}$, $\mathcal{P}$, $D^k$, $x^k$, $z^k$, $\omega^k$, $\breve{\phi}^k$ )
    for $k = 1, 2, \ldots, k_{max}$ do
        Initialize $\omega^{k+1} \leftarrow \omega^k$, $\breve{\phi}^{k+1} \leftarrow \breve{\phi}^k$        ▷ (Default, null-step updates)
        $(\widehat{x}^{k+1}, x^{k+1}, z^{k+1}, D^{k+1}, \Gamma) \leftarrow$ SDM-GS($L_\rho(\cdot, \cdot, \omega^k)$, $X$, $Z \cap B$, $D^k$, $x^k$, $z^k$, $t_{max}$)

        if $L_\rho(x^{k+1}, z^{k+1}, \omega^k) + \frac{\rho}{2}\left\|Qx^{k+1} - z^{k+1}\right\|_2^2 - \breve{\phi}^k \leqslant \epsilon$ then
            return $(x^{k+1}, z^{k+1}, \omega^{k+1}, \breve{\phi}^{k+1}, D^{k+1}, \mathbf{true})$
        end if
        $\tilde{\phi} \leftarrow L_\rho(x^{k+1}, z^{k+1}, \omega^k) + \frac{\rho}{2}\left\|Qx^{k+1} - z^{k+1}\right\|_2^2 - \Gamma$
        $\tilde{\gamma} \leftarrow \frac{\tilde{\phi} - \breve{\phi}^k}{L_\rho(x^{k+1}, z^{k+1}, \omega^k) + \frac{\rho}{2}\left\|Qx^{k+1} - z^{k+1}\right\|_2^2 - \breve{\phi}^k}$
        if $\tilde{\gamma} \geqslant \gamma$ then
            set $\omega^{k+1} \leftarrow \omega^k + \rho(Qx^{k+1} - z^{k+1})$, $\breve{\phi}^{k+1} \leftarrow \tilde{\phi}$
        end if
        Possibly update $\rho$, e.g., $\rho \leftarrow \frac{1}{\min\left\{\max\left\{(2/\rho)(1-\tilde{\gamma}), 1/(10\rho), 10^{-4}\right\}, 10/\rho\right\}}$ as in [2]
    end for
    return $(\widehat{x}^{k+1}, x^{k+1}, z^{k+1}, \omega^{k+1}, \breve{\phi}^{k+1}, D^{k+1}, \mathbf{false})$
end function

Preconditions: $\tilde{x} \in \mathrm{conv}(X)$, $\tilde{z} \in \mathrm{argmin}_z \{F(\tilde{x}, z) : z \in Z\}$, $D \subseteq \mathrm{conv}(X)$
function SDM-GS($F$, $X$, $Z$, $D$, $\tilde{x}$, $\tilde{z}$, $t_{max}$)
    for $t = 1, \ldots, t_{max}$ do
        $\tilde{x} \leftarrow \mathrm{argmin}_x \{F(x, \tilde{z}) : x \in D\}$
        $\tilde{z} \leftarrow \mathrm{argmin}_z \{F(\tilde{x}, z) : z \in Z\}$
    end for
    $\widehat{x} \in \mathrm{argmin}_x \{\nabla_x F(\tilde{x}, \tilde{z})(x - \tilde{x}) : x \in X\}$
    Reconstruct $D$ to be any set such that
        $\{\tilde{x} + \alpha(\widehat{x} - \tilde{x}) : \alpha \in [0,1]\} \subseteq D \subseteq \mathrm{conv}(X)$
    Set $\Gamma \leftarrow -\nabla_x F(\tilde{x}, \tilde{z})(\widehat{x} - \tilde{x})$
    return $(\widehat{x}, \tilde{x}, \tilde{z}, D, \Gamma)$
end function
```

Algorithm PSCG addresses the solution to an alternative dual problem which is equivalent to (**??**) when f is linear, but in general provides a weaker dual bound otherwise. This dual problem is used to address the more general setting where f is convex but possibly nonlinear.

Proposition 1 *Let $\left\{(x^k, z^k, \omega^k)\right\}$ be a sequence generated by Algorithm 1 applied to problem* (1) *with X compact, Z a linear subspace, $\omega^0 \in Z^\perp$, B closed and convex, $\rho > 0$, $\gamma \in (0,1)$, $\epsilon = 0$ and $k_{max} = \infty$. If there exists a dual optimal solution ω^* to the dual problem* (**??**)*, then either*

1. *$\omega^k = \overline{\omega}$ is fixed and optimal for* (**??**) *for $k \geqslant \bar{k}$ for some finite $\bar{k}$; or*
2. *ω^k is never optimal for* (**??**) *for any finite $k \geqslant 1$, but $\lim_{k\to\infty} \omega^k = \overline{\omega}$ is optimal,*

and the sequence $\left\{(x^k, z^k)\right\}$ *has limit points* $(\overline{x}, \overline{z})$*, each of which are optimal for problem* (**??**).

1.2 Parallelization and workload

The opportunities for parallelization and distribution of the computational workload in PSCG, as stated in Algorithm 1, are not immediately apparent. This subsection explicitly indicates which update problems may be solved in parallel, and the nature of the required communication between the parallel computational nodes.

The bulk of computational work, parallelization, and parallel communication occurs within the SDM-GS method stated in Algorithm **??**, where for the problems of interest, the following decomposable structures apply: $X = \prod_{i=1}^m X_i$, $D = \prod_{i=1}^m D_i$, and $F(x, z) = \sum_{i=1}^m F(x_i, z)$. In the larger context of Algorithm 1, the subproblem of Line 3 in Algorithm **??** can be solved in parallel given fixed $\widetilde{z} \in Z$ and $\omega \in Z^\perp$ along the block indices $i = 1, \dots, m$ as

$$\min_x \left\{ f_i(x) + (\omega_i)^\top Q_i x + \frac{\rho}{2} \left\| Q_i x - \widetilde{z}_i \right\|_2^2 : x \in D_i \right\}, \tag{12}$$

while the subproblem of Line 6 is solved as

$$\min_x \left\{ \nabla_x f_i(\widetilde{x}_i) + (\omega_i + \rho(Q_i \widetilde{x}_i - \widetilde{z}_i))^\top Q_i x : x \in X_i \right\}.$$

Remark 2 In the setting where problem (1) is a large-scale mixed-integer linear optimization problem, the subproblems of Line 3 are continuous convex quadratic optimization problems for each block $i = 1, \dots, m$, which can be solved independently of one another and in parallel. In the same setting, the Line 6 subproblems are mixed-integer optimization problems for each block $i = 1, \dots, m$, which can also be solved independently of one another and in parallel. Additionally, the reconstruction of D occurring in Line 8 can be done in parallel for each D_i along the indices $i = 1, \dots, m$.

Parallel communication is needed for the computation of the z update. In the larger context of Algorithm 1, this takes the form of solving

$$\min_z \left\{ \sum_{i=1}^m \left\| Q_i \widetilde{x}_i - z_i \right\|_2^2 : z \in Z \right\}.$$

This is solved as an averaging that requires the reduce-sum type parallel communication. The computation of values required to compute γ^k in Line 9 in Algorithm 1 also requires a reduce-sum type parallel communication. For implementation purposes, the computation of these values, including the computation of $\mathit{\Gamma}$ from the SDM-GS call, can be combined into one reduce-sum communication. In total, each iteration of Algorithm 1 requires two reduce-sum type communications, one for computing the z-update, and one combined reduce-sum communication to compute scalars associated with the Lagrangian bounds and the critical values for the termination conditions. The storage and updates of x^k and ω^k and D can also be done in parallel, while z^k and γ^k need to be computed and stored by every processor at each iteration k.

2 Branch-and-bound with SCG-GS-ALM as bounding procedure

In order to inform branching decisions, we need a sense of dispersion based on some violation of feasibility, and a value with which to bound.

2.1 Measure of feasibility

A criterion for fathoming by optimality may be given by

$$\|x^* - \mathrm{P}_X(x^*)\| + \|Qx^* - z^*\| < \epsilon. \tag{13}$$

for some suitably small tolerance $\epsilon > 0$. If the feasibility criterion (13) is not met, then branching must occur, and the branching should be informed by the dispersions measured in the two norm terms of (13). It will be convenient to to refine the dispersions, where the vector

$$\sigma_X := [\|x_i^* - \mathrm{P}_{X_i}(x_i^*)\|]_{i=1,\dots,m}$$

is defined. The diagonal matrix whose diagonal entries are taken from σ_X will be denoted Σ_X.

2.2 Branching decisions

We have the following approaches at measuring dispersion from the constraint $Qx = z$ at iteration k:

1. Given a dual feasible solution ω^k, for some $\widehat{x}^k \in \widehat{x}(\omega^k)$, setting $\widehat{x}^k = (\widehat{x}_i^k)_{i=1,\dots,m}$, we take an average $\widehat{z}^k = (1/m)\sum_{i=1}^m Q_i\widehat{x}_i^k$. Based on this average, we compute a component-wise dispersion

   $$\sigma_j^k := \left\| \left[(Q_i\widehat{x}_i^k)_j - \widehat{z}_j^k \right]_{i=1,\dots,m} \right\| + \left\| Qx^k - z^k \right\|$$

   for each component index $j = 1,\dots,n_x$ using norm $\|\cdot\| = \|\cdot\|_1$, $\|\cdot\|_2$, or $\|\cdot\|_\infty$, and denote $\sigma := (\sigma_j)_{j=1,\dots,n_x}$. This is the approach [in Caroe and Schultz]. One quirk with this approach is that, at the optimal dual solution ω^*, any of the previously generated solutions $\widehat{x} \in \widehat{x}(\omega^k)$ may also satisfy $\widehat{x} \in \widehat{\omega}^*$. This is especially true in the typically satisfied case that $c \in Z$. Also, the approach does not use information from the dispersions $\|x^* - \mathrm{P}_X(x^*)\|$. The latter issue is addressed with the following modification to σ_j^k:

   $$\sigma_j^k := \left\| \left[(([\sigma_X]_i + \epsilon)[Q_i\widehat{x}_i^k) - \widehat{z}_i^k]_j \right]_{i=1,\dots,m} \right\| + \left\| Qx^k - z^k \right\|$$

2. Let $\widehat{x}^t \in \widehat{x}(\omega^t)$ for $t = 0, 1, \ldots, k$. Then another form of dispersion is given by:

$$\sigma_j := \sum_{i=1,\ldots,m} \sum_{t=1,\ldots,k} \alpha_t \left| Q_i \left(\widehat{x}^t_{i,j} - x^k_{i,j} \right) \right| + \rho_{i,j} \left| Q_i x^k_{i,j} - z^k_j \right|,$$

   where $0 \leqslant \alpha_t \leqslant 1$ for $t = 1, \ldots, T_i$ and $\sum_{t=1,\ldots,T_i} \alpha_t = 1$ are the convex multipliers used to form a convex combination $x^k = \sum_{t=1,\ldots,k} \alpha_t \widehat{x}^t$.

Remark 3 We also considered using a measure of the dispersion of the dual solutions ω^k, taking

$$\sigma_j := \left\| [\omega_{i,j}]_{i=1,\ldots,m} \right\|.$$

This approach is (somewhat) sensible for when restricting the branching to variables with integrality restrictions, or when the disjunctions are otherwise strictly separated. It is otherwise an ill-defined approach, in that optimal dual solutions are typically non-unique, especially when bounding constraints are being added. Any further thoughts on this matter must take into account the set of dual optimal solutions.

3 Computational results

4 Conclusion and future work

References

1. Birge, J.R., Louveaux, F.: Introduction to Stochastic Programming. Springer Science & Business Media (2011)
2. Kiwiel, K.C.: Approximations in proximal bundle methods and decomposition of convex programs. Journal of Optimization Theory and Applications **84**(3), 529–548 (1995)