



# A study of progressive hedging for stochastic integer programming

Jeffrey Christiansen<sup>1</sup> · Brian Dandurand<sup>1</sup> · Andrew Eberhard<sup>1</sup>  · Fabricio Oliveira<sup>2</sup>

Received: 8 July 2022 / Accepted: 19 September 2023 / Published online: 11 October 2023  
© The Author(s) 2023

## Abstract

Motivated by recent literature demonstrating the surprising effectiveness of the heuristic application of progressive hedging (PH) to stochastic mixed-integer programming (SMIP) problems, we provide theoretical support for the inclusion of integer variables, bridging the gap between theory and practice. We provide greater insight into the following observed phenomena of PH as applied to SMIP where optimal or at least feasible convergence is observed. We provide an analysis of a modified PH algorithm from a different viewpoint, drawing on the interleaving of (split) proximal-point methods (including PH), Gauss–Seidel methods, and the utilisation of variational analysis tools. Through this analysis, we show that under mild conditions, convergence to a feasible solution should be expected. In terms of convergence analysis, we provide two main contributions. First, we contribute insight into the convergence of proximal-point-like methods in the presence of integer variables via the introduction of the notion of persistent local minima. Secondly, we contribute an enhanced Gauss–Seidel convergence analysis that accommodates the variation of the objective function

---

Jeffrey Christiansen, Brian Dandurand, Andrew Eberhard and Fabricio Oliveira have contributed equally to this work.

---

This work was supported by the Australian Research Council (ARC) Grant No. (ARC DP140100985).

---

✉ Andrew Eberhard  
andy.eberhard@rmit.edu.au

Jeffrey Christiansen  
jeffrey.christiansen@rmit.edu.au

Brian Dandurand  
bcdandurand@gmail.com

Fabricio Oliveira  
fabricio.oliveira@aalto.fi

<sup>1</sup> Mathematics and Statistics, RMIT University, 124 La Trobe Street, 3000 Melbourne, VIC, Australia

<sup>2</sup> Mathematics and Systems Analysis, Aalto University, Otakaari 1, 02150 Espoo, Uusimaa, Finland

under mild assumptions. We provide a practical implementation of a modified PH and demonstrate its convergent behaviour with computational experiments in line with the provided analysis.

**Keywords** Stochastic mixed integer programming · Progressive hedging · Variational analysis · Gauss–Seidel methods

**Mathematics Subject Classification** 68Q25 · 68R10 · 68U05

## 1 Introduction

Stochastic mixed integer programming (SMIP) models are, in essence, large-scale mixed-integer programming (MIP) models in which the uncertain nature of the input parameters is modelled by means of a finite set of discrete scenarios [1]. This general framework allows one to model a broad class of decision problems, as can be attested from the wealth of publications from diverse areas of science and engineering. Important applications employing SMIP models include unit commitment [2], hydro-thermal generation scheduling [3], military operations [4], vaccination planning [5], air traffic flow management [6], forestry management and forest fire response [7], supply chain and logistics planning [8], and other applications referred to on the SIPLIB website [9]. The practical and theoretical development of stochastic programs (SP) (without integer variables) preceded SMIP and has influenced its development. The Progressive Hedging (PH) algorithm [10] for solving SP problems is well studied and theoretically supported for convex problems with no integer-constrained variables. Even without this theoretical support in the setting with integer-constrained variables, PH as a heuristic often demonstrates effectiveness for providing both upper and lower bounds [11] and often feasible solutions. Motivated by the limited theoretical support of PH for its application to SMIP and the observed success of PH heuristics for SMIP, our objective is to develop a theoretical framework and demonstrate convergence in numerical experiments.

The large scale of the deterministic equivalent of SMIP models proves to be challenging for off-the-shelf solvers that do not utilise the decomposable structure inherent in the extensive deterministic forms of SMIP models. By contrast, more promising solution methods utilise the SMIP's decomposable structure. The PH algorithm [10] addresses the decomposable structure as a variant of the alternating direction method of multipliers (ADMM) [12] where the non-anticipativity constraint is relaxed into an augmented Lagrangian (AL) reformulation. One of the earliest detailed treatments of its convergence was based on variational analysis techniques [10], where nonsmooth analysis also provided important tools for the study of convergence with respect to the satisfaction of optimality conditions. Augmented Lagrangian duality, which plays a fundamental role in this work, also appeared in subsequent works to provide duality theorems for very general nonconvex problems, including cases encompassing integer constraints [13, Chapter 11, section K], [14]. These publications have attracted the attention of the integer programming community and resulted in a body of literature focussing specifically on the application of augmented Lagrangian duality to

mixed-integer programming (MIP) [15, 16]. This in turn motivated researchers to state, analyse, and test a version of the PH method containing nonsmooth augmentations [17]. Concurrently, researchers have also explored the combination of PH with the Frank-Wolfe algorithm [18] to obtain provably convergent dual bounding methods for SMIP based on the Lagrangian relaxation of the non-anticipativity constraint [19]. Other researchers produced primal (heuristic) methods where the quadratic subproblems of PH were replaced by mixed integer quadratic programs (MIQP) [11]. These approaches were shown to produce excellent solutions as long as the penalty parameter was chosen judiciously, and so have remained an enigma, lacking any theoretical convergence result. In this paper, we show that variational analysis techniques will further draw back the curtain on this enigma and so explain what actually underpins the success of PH with a MIQP when applied to SMIP.

Under reasonable assumptions, we analyse the convergence of PH applied to SMIPs, where we allow the penalty parameters to vary in a less restricted fashion than is typically required for PH and related approaches, while also applying Lagrangian multiplier updates requiring a special rule due to the required satisfaction of convergence/boundedness criteria that may not be automatically satisfied when PH is applied to SMIPs. Furthermore, our approach allows for more generality in the type of augmented Lagrangian terms. In our analysis, we view the PH method as an application of Gauss–Seidel iterations with penalty and Lagrange terms allowed to vary between iterations. In this setting, we contribute insight into when PH generates a sequence of solutions that converges to a feasible point of the SMIP. Our approach may also be viewed as interfacing some seemingly distinct solution methodologies found in proximal point methods such as PH, Gauss–Seidel (GS) methods, and (mixed-integer) augmented Lagrangian duality [15–17, 20]. Furthermore, connection with feasibility pump (FP) primal heuristics is evident in the same spirit as contributed in [21].

Some of the conditions assumed for the penalty and/or augmented Lagrangian term that are required to achieve an exact penalty effect in [15, 16] (e.g., [16, Theorem 5]) require the penalty functions to be non-differentiable, which can impede the analysis of Gauss–Seidel methods [22]. Thus, in this paper, we set out to develop this theory from another direction that allows for a differentiable penalty term, in line with that typically used for analysing progressive hedging-like methods. To compensate for the loss of the exact penalty effect shown in [15, 16], we provide an analysis describing the effect of (potentially) letting the penalty coefficient go to infinity in order to achieve feasibility. In particular, we analyse a SMIP solution method inspired by the FP, PH, and Gauss–Seidel convergence analysis, that for short, will be denoted FPPH, which in practice is similar to the use of PH as an heuristic [11], except we allow for greater generality in the updating of Lagrange multipliers, the changing of penalty coefficients, and in the allowable forms of the augmented Lagrangian penalty function itself. Successful convergence of the method allows for (but is not predicated on) the unbounded increase of penalty parameters. To be clear, our analysis does not promise both primal and dual optimal convergence as is provided for PH in the convex, continuous setting. Rather, we address convergence goals similar to those of feasibility pump methods, where high-quality feasible solutions are sought, and the main challenge is avoiding either non-feasible convergence or cycling.

Our experimental results will demonstrate the effectiveness of FPPH. As with all FP approaches, one needs to develop heuristics for updating the penalty parameters to encourage the methods to locate the best possible feasible solution and hence the strongest primal bound. As a general conclusion, the FPPH presents promising performance relative to Progressive Hedging in terms of quickly obtaining good feasible solutions for SMIPs with pure integer first-stage variables.

This paper is structured as follows. In Sect. 2, we set up the assumptions on the regularisation and the conceptual framework on which the analysis rests. In Sect. 3, further results are developed on how we may decompose the regularisation into its “cross-sections” where integer variables are fixed, which provides a foundation for insight into the local minima of the (whole) regularisation. Section 4 introduces the concept of *persistent local minima* and their relationship to feasibility for SMIP (1). The convergence analysis of the associated Gauss–Seidel algorithm is carried out in Sect. 5. In Sect. 6, we present computational results illustrating the employment of variants of FPPH to find high-quality feasible solutions to SMIP instances. In Sect. 7, we provide concluding remarks and directions for future developments.

## 2 Fundamental concepts and conceptual algorithmic framework

Denote  $\mathbf{x} = (x_s)_{s \in S}$  where  $x_s \in \mathbb{X}_d := \mathbb{R}^{n-q} \times \mathbb{Z}^q \subseteq \mathbb{X} := \mathbb{R}^n$ . Similarly  $\mathbf{y} = (y_s)_{s \in S}$  where  $y_s \in \mathbb{Y}_d := \mathbb{R}^{m-r} \times \mathbb{Z}^r \subseteq \mathbb{Y} := \mathbb{R}^m$ . We state the SMIP in the following split-variable deterministic formulation (see, e.g., [1])

$$\zeta^{SMIP} = \min_{\mathbf{x} \in \mathbb{X}_d^{|S|}, \mathbf{y} \in \mathbb{Y}_d^{|S|}, z \in \mathbb{X}, \mathbf{w} \in \mathbb{Y}^{|S|}} \sum_{s \in S} f_s(x_s, y_s) \tag{1a}$$

$$\text{s.t. } (z - x_s, w_s - y_s) = (0, 0), \quad (x_s, y_s) \in K_s, \quad s \in S, \tag{1b}$$

where

$$f_s(x_s, y_s) := p_s \left( c^\top x_s + d_s^\top y_s \right), \quad s \in S \tag{1c}$$

$$K_s := \{(x, y) \in \mathbb{X}_d \times \mathbb{Y}_d \mid x \in X, y \in Y_s(x)\}, \quad s \in S \tag{1d}$$

$$X := \{x \in \mathbb{X}_d \mid Ax \leq b\} \tag{1e}$$

$$Y_s(x) := \{y \in \mathbb{Y}_d \mid T_s x + W_s y \leq h_s\}, \quad s \in S. \tag{1f}$$

Note that the constraints  $x_s \in X$  that hold only for the first-stage decision variables  $x_s$  are identical for all  $s \in S$ .

We denote the extended real values by  $\mathbb{R}_{+\infty} := \mathbb{R} \cup \{+\infty\}$ . For each scenario  $s \in S$  copy of first-stage variables  $x_s$  and separately for each scenario  $s \in S$  second-stage variables  $y_s$  we assume that the integer variable component indices ( $\mathcal{I}$ ) always follow the real variable component indices ( $\mathcal{R}$ ). That is,  $x_s := (x_{s,\mathcal{R}}, x_{s,\mathcal{I}})$  and  $y_s := (y_{s,\mathcal{R}}, y_{s,\mathcal{I}})$  for each  $s \in S$ . Define the projection  $\text{proj}_{\mathbb{X},\mathcal{I}} : \mathbb{X}_d \rightarrow \mathbb{Z}^q$  by  $\text{proj}_{\mathbb{X},\mathcal{I}}((x_{\mathcal{R}}, x_{\mathcal{I}})) = x_{\mathcal{I}}$  (with a similar definition for  $y_{\mathcal{I}}$  projection  $\text{proj}_{\mathbb{Y},\mathcal{I}}$ ). As the first-stage consensus variable  $z$  components should match those for each  $x_s$ , due to

the non-anticipativity constraints  $z - x_s = 0, s \in S$ , the same first-stage distinction between real and integer components  $z := (z_{\mathcal{R}}, z_{\mathcal{I}})$  apply. Corresponding distinctions for the second-stage consensus  $w_s := (w_{s,\mathcal{R}}, w_{s,\mathcal{I}}), s \in S$ , apply as well. Note that  $z$  is *not explicitly* constrained to lie within the discrete feasible set  $X$ . Nor are  $w_s, s \in S$ , explicitly constrained to lie within  $Y_s(x_s)$  or  $Y_s(z)$ . Thus, strictly speaking,  $z \in \mathbb{X}$  and  $w_s \in \mathbb{Y}$  vary freely within their respective spaces. Denote  $\mathbf{w} := (w_1, \dots, w_{|S|})$  and similarly for  $(\mathbf{x}, \mathbf{y})$  and when needed we denote  $\mathbf{z} := (z, \dots, z) \in \mathbb{R}^{n \times |S|}$ .

Since the second-stage non-anticipativity variables are independent for each outcome scenario and otherwise unconstrained, the non-anticipativity constraints  $w_s - y_s = 0, x \in S$ , have no practical effect on the feasibility of the second-stage decisions  $y_s$  in SMIP (1). Nevertheless, this formulation aids the subsequent analysis by allowing the incorporation of all variables (regardless of stage) into regularisation terms in a symmetric fashion. The second stage feasibility is propagated from  $\mathbf{y}$  to  $\mathbf{w}$  via this constraint, while  $\mathbf{w}$  remains unconstrained. The formulation (1) is also conducive to generalising our results for two-stage SMIPs to multi-stage problems in which all stages except the last have active non-anticipativity constraints. In the practical application of developed algorithms to two-stage problems, the use of  $\mathbf{w}$  may be suppressed, as it is in the description of the computational experiments of Sect. 6.

Throughout our developments, we assume the following assumptions to hold regarding our SMIP (1). We explicitly assume the existence of an optimal solution, which could be replaced by the standard assumption of rationality of the data defining the problem.

**Assumption 1** We make the following standard SMIP assumptions:

1. **Stochasticity of  $p_s$ :** for each  $s \in S$ , we have  $p_s > 0$  and  $\sum_{s \in S} p_s = 1$ .
2. **Non-emptiness:**  $K_s, s \in S$ , is a non-empty set of feasible decisions constructed with linear constraints and integrality constraints on the  $x_s$  and  $y_s$  variables. (This also implies that  $K_s$  is closed.)
3. **Boundedness and Optimality:** The optimal value of the SMIP (1) is bounded from below. Also, the feasible sets  $K_s, s \in S$ , are bounded. Furthermore,  $\zeta^{SMIP}$  is feasible and possesses an optimal solution.
4. **Relatively complete recourse:** The SMIP model has relatively complete recourse;  $\forall x \in X, \forall s \in S$ , we have  $Y_s(x) \neq \emptyset$ : that is, first-stage decisions  $x$  that satisfy the first-stage specific constraints  $x \in X$  have at least one second-stage decision solution  $(y_s)_{s \in S}$  for which  $(x, y_s) \in K_s$  for all scenarios  $s \in S$ .

Of interest is the dual function  $\zeta : \Lambda \times \mathbb{R}_{>0}^{|S|} \times \mathbb{R}_{>0} \rightarrow \mathbb{R}_{+\infty}$  defined by

$$\zeta(\lambda, \pi, \rho) := \min\{\varphi^{\lambda, \rho, \pi}(z, \mathbf{w}) \mid (z, \mathbf{w}) \in \mathbb{X} \times \mathbb{Y}^{|S|}\}, \tag{2}$$

where

$$\varphi^{\lambda, \rho, \pi}(z, \mathbf{w}) := \sum_{s \in S} \varphi_s^{\lambda, \rho, \pi}(z, w_s) \text{ and } \Phi^{\lambda, \rho, \pi}(z, \mathbf{w}) := \prod_{s \in S} \Phi_s^{\lambda, \rho, \pi}(z, w_s) \tag{3a}$$

with, for each  $s \in S$ ,

$$\varphi_s^{\lambda, \rho, \pi}(z, w_s) := \min_{(x_s, y_s) \in K_s} f_s(x_s, y_s) - \lambda_s^\top (z - x_s) + \rho \pi_s \psi(z - x_s, w_s - y_s) \tag{3b}$$

$$\Phi_s^{\lambda, \rho, \pi}(z, w_s) := \operatorname{argmin}_{(x_s, y_s) \in K_s} f_s(x_s, y_s) - \lambda_s^\top (z - x_s) + \rho \pi_s \psi(z - x_s, w_s - y_s). \tag{3c}$$

We assume that  $K_s$  is defined as in (1d), and that the usual dual feasibility  $\lambda \in \Lambda := \{\lambda \mid \sum_{s \in S} \lambda_s = 0\}$  holds. For each scenario  $s \in S$ , the penalty function  $\psi$  output value is scaled by a penalty scaling parameter  $\rho > 0$ , and scenario  $s \in S$  specific penalty weighting parameters  $\pi_s > 0$  (for which  $\sum_{s \in S} \pi_s = 1$ ) to be specified. Note that under the assumption that  $\lambda \in \Lambda$ , the summation  $\sum_{s \in S} \lambda_s z$  conveniently vanishes and so these terms may be dropped in subsequent developments.

Each instance of problem (2) is a continuous optimisation problem over the space  $\mathbb{X} \times \mathbb{Y}^{|S|}$ , and for nontrivial instances of SMIP (1),  $\varphi^{\lambda, \rho, \pi}$  is nonconvex with multiple isolated local minima. Under assumptions in [15, 16], we have  $\zeta^{SMIP} = \zeta(\lambda, \pi, \rho)$  for sufficiently large but finite  $\rho$ . Properties of locally optimal solutions to the minimisation of  $\varphi^{\lambda, \rho, \pi}$ , and how these local minimisers relate to the solutions to the original SMIP (1), are of special interest in this paper’s subsequent analysis.

As mentioned earlier, the nonsmoothness of penalty functions  $\psi$  that support the exact penalty properties discussed in [15, 16] prevents the support of convergence theory provided by Gauss–Seidel approaches. For this reason, we modify the properties assumed in [15, 16] for the penalty function  $\psi$  to the conditions stated in Assumption 2. In particular, we assume that the penalty is strongly convex and differentiable from the outset (departing markedly from [15, 16]), as this is required for a Gauss–Seidel approach to be applied with desirable convergence properties (see Lemma 21).

**Assumption 2** For our smooth penalty function  $\psi : \mathbb{X} \times \mathbb{Y} \rightarrow \mathbb{R}$ , we make the following **integer compatible regularisation function (ICRF)** assumptions:

1.  $\psi(u, v) \geq 0$  for all  $(u, v)$  and  $(u, v) = 0$  if and only if  $\psi(u, v) = 0$ .
2. If  $\gamma \in [0, 1)$  then  $\psi(\gamma u, v) < \psi(u, v)$  for all  $u \neq 0$  and  $\psi(u, \gamma v) < \psi(u, v)$  for all  $v \neq 0$ .
3. Strong convexity holds with modulus  $m > 0$  i.e.

$$\psi(u, v) \geq \psi(u^0, v^0) + \left\langle \nabla \psi(u^0, v^0), \begin{bmatrix} u - u^0 \\ v - v^0 \end{bmatrix} \right\rangle + \frac{m}{2} \left\| \begin{bmatrix} u - u^0 \\ v - v^0 \end{bmatrix} \right\|^2. \tag{4}$$

We note that Assumption 2 implies  $(0, 0) = \nabla \psi(0, 0)$  and thus (4) implies

$$\psi(u, v) \geq \frac{m}{2} \|(u, v)\|^2, \text{ for all discrepancies } (u, v). \tag{5}$$

**Remark 1** In the theoretical development, we partition the discrepancies into  $u$  and  $v$  components to correspond to the special treatment of early-stage variables against late-stage variables. For a two-stage problem,  $u$  corresponds to first-stage discrepancies, and  $v$  corresponds to second-stage discrepancies. To allow for versatility in how the theoretical development informs algorithmic approaches, especially for application to multi-stage problems, we carry the development with the distinction between  $u$  and  $v$  discrepancies through Sect. 5.

**Remark 2** In our computational developments in Sect. 6, we use a weighted squared 2-norm penalty function  $\psi(u, v) = \frac{1}{2} (\sum_{i=1}^n [\bar{\mu}_i u_i^2] + \|v\|^2)$  with weights  $\bar{\mu}_i > 0, i = 1, \dots, n$ . In general, the strong convexity with modulus  $m$  is equivalent to the convexity of the function  $(u, v) \mapsto \psi(u, v) - \frac{m}{2} \|(u, v)\|^2$ . For the algorithmic manifestation as presented in Sect. 6,  $v$  may furthermore be set identically to value zero.

### 2.1 Preliminary application of Gauss–Seidel iterations

We define the following notation based on the assumption that the Lagrange multipliers  $\lambda^n \in \Lambda$  and the penalty parameters  $\rho^n > 0, \pi^n > 0, \sum_{s \in S} \pi_s^n = 1$ , vary with each iteration  $n \geq 0$ :

$$\varphi^n(z, \mathbf{w}) := \sum_{s \in S} \varphi_s^n(z, w_s) \quad \text{where} \quad \varphi_s^n(z, w_s) := \varphi_s^{\lambda^n, \rho^n, \pi^n}(z, w_s). \tag{6}$$

One iterative solution approach for finding locally optimal solutions for SMIP (1) starting with initial  $z^0 \in \mathbb{X}$  is based on Gauss–Seidel (GS) iterations  $n \geq 0$  of the form

$$w_s^{n+1} \leftarrow \underset{w \in \mathbb{Y}}{\operatorname{argmin}} \varphi_s^n(z^n, w) \quad \text{for all } s \in S, \tag{7a}$$

$$z^{n+1} \leftarrow \underset{z \in \mathbb{X}}{\operatorname{argmin}} \varphi^n(z, \mathbf{w}^{n+1}). \tag{7b}$$

The  $z$  update (7b) is not easily computable, but the  $w$  update (7a) is so, as demonstrated in the following proposition.

**Proposition 3** *Let  $(z, w) \in \mathbb{X} \times \mathbb{Y}$ .*

1. *For each  $s \in S, w \in \operatorname{argmin}_{w'} \varphi_s^{\lambda, \rho, \pi}(z, w')$  implies  $w \in \operatorname{proj}_{\mathbb{Y}}(\Phi_s^{\lambda, \rho, \pi}(z, w))$ .*
2. *Moreover, given  $z^n, \mathbf{w}^{n+1} \in \operatorname{argmin}_{w'} \varphi^n(z^n, w')$  may be computed by solving for each  $s \in S$*

$$(x_s^{n+1}, y_s^{n+1}) \in \operatorname{arg} \min_{(x_s, y_s) \in K_s} f_s(x_s, y_s) + (\lambda_s^n)^\top x_s + \rho^n \pi_s^n \psi(z^n - x_s, 0) \tag{8}$$

*and then setting  $\mathbf{w}^{n+1} = \mathbf{y}^{n+1}$ .*

**Proof** We show the contrapositive. Assume for some  $s \in S$  that  $w \in \operatorname{argmin}_{w'} \varphi_s^{\lambda, \rho, \pi}(z, w')$ , but that  $w \notin \operatorname{proj}_{\mathbb{Y}} \left( \Phi_s^{\lambda, \rho, \pi}(z, w) \right)$ . Then as  $(x_s, y_s) \in \Phi_s^{\lambda, \rho, \pi}(z, w)$  with  $w_s \neq y_s$ , we have

$$\begin{aligned} \varphi_s^{\lambda, \rho, \pi}(z, w_s) &= f_s(x_s, y_s) + (\lambda_s)^\top x_s + \rho \pi_s \psi(z - x_s, w_s - y_s) \\ &> f_s(x_s, y_s) + (\lambda_s)^\top x_s + \rho \pi_s \psi(z - x_s, 0) \quad (\text{due to Assumption (2)}) \\ &= f_s(x_s, y_s) + (\lambda_s)^\top x_s + \rho \pi_s \psi(z - x_s, y_s - y_s) \\ &\geq \min_{(x'_s, y'_s) \in K_s} \left\{ f_s(x'_s, y'_s) + (\lambda_s)^\top x'_s + \rho \pi_s \psi(z - x'_s, y_s - y'_s) \right\} \\ &= \varphi_s^{\lambda, \rho, \pi}(z, y_s) \end{aligned}$$

which implies the contradiction that  $w_s \notin \operatorname{argmin}_w \varphi_s^{\lambda, \rho, \pi}(z, w_s)$ . To show the claim of Part 2, assume that  $w^{n+1}$  computed from (8) does not satisfy  $w_s^{n+1} \in \operatorname{argmin}_{w'} \varphi_s^n(z^n, w')$  for at least one  $s \in S$ . Let  $\hat{w}_s^{n+1} \in \operatorname{argmin}_{w'} \varphi_s^n(z^n, w')$ . By Part 1, there exists  $\hat{x}_s^{n+1}$  with  $(\hat{x}_s^{n+1}, \hat{w}_s^{n+1}) \in \Phi_s^n(z^n, \hat{w}_s^{n+1})$  such that

$$\begin{aligned} &f_s(\hat{x}_s^{n+1}, y_s^{n+1}) + (\lambda_s^n)^\top \hat{x}_s^{n+1} + \rho^n \pi_s^n \psi(z^n - \hat{x}_s^{n+1}, 0) \\ &< f_s(x_s^{n+1}, y_s^{n+1}) + (\lambda_s^n)^\top x_s^{n+1} + \rho^n \pi_s^n \psi(z^n - x_s^{n+1}, 0), \end{aligned}$$

which would contradict the optimality in (8). □

Computing the update  $z^{n+1} \in \operatorname{argmin}_z \varphi^n(z, w^{n+1})$  given fixed  $w^{n+1}$  corresponds to an infimal convolution of  $(x_s, y_s) \mapsto f_s(x_s, y_s) + \delta_{K_s}(x_s, y_s) + (\lambda_s^n)^\top x_s$  and  $(u, v) \mapsto \rho^n \pi_s^n \psi(u, v)$ , for each  $s \in S$ , where we denote the indicator function of a set  $K_s$  by  $\delta_{K_s}(x, y)$  that takes the value zero if  $(x, y) \in K_s$  and  $+\infty$  otherwise. The infimal convolution is well-studied [13, Chapter 1, section H], and later we make use of certain convex “cross-sections” of this infimal convolution. However, the calculation culminating in  $z^{n+1} \in \operatorname{argmin}_z \varphi^n(z, w^{n+1})$  is still not easily computable, as it requires the solution of a MIP of comparable difficulty to the original SMIP (1). Nevertheless, this problem  $z^{n+1} \in \operatorname{argmin}_z \varphi^n(z, w^{n+1})$  is useful from a *theoretical* standpoint, as it links the consensus problem to the Gauss–Seidel step of the continuous regularisation.

A more practical approach to the  $z$  update takes the form of descent steps using the usual consensus update, i.e.

$$z^{n+1} \in \operatorname{argmin}_{z \in \mathbb{X}} \sum_{s \in S} \pi_s^n \psi(z - x_s^{n+1}, 0), \tag{9}$$

where  $w^{n+1} - y^{n+1} = 0$  follows from Proposition 3. From Assumption 2(3) with  $u_s^0 = z^{n+1} - x_s^{n+1}$ ,  $u_s = z - x_s^{n+1}$  and  $v_s^0 = v_s = 0$  for all  $s \in S$ , and the optimality condition associated with the  $z^{n+1}$  update  $\sum_{s \in S} \pi_s^n \nabla_z \psi(z^{n+1} - x_s^{n+1}, 0) = 0$  we



have that

$$\sum_{s \in S} \pi_s^n \psi(z - x_s^{n+1}, 0) \geq \sum_{s \in S} \pi_s^n \psi(z^{n+1} - x_s^{n+1}, 0) + \frac{m}{2} \|z - z^{n+1}\|^2 \quad (10)$$

and so the  $z^{n+1}$  update (9) must be unique.

Using this observation, we can devise a Gauss–Seidel algorithm that is guaranteed to produce non-ascent steps while the stabilisation  $(z^n, w^{n+1}) = (z^{n+1}, w^{n+1})$  is not achieved, which is given in Algorithm 1.

---

**Algorithm 1** Modified block GS method for SIP

---

- 1: **initialise**  $z^0 \in \mathbb{X}, \rho^0 > 0, \lambda_s^0 \leftarrow 0, s \in S$
  - 2: **for**  $n = 0, \dots, n_{\max} - 1$  **do**
  - 3:   Compute the following:
  - 4:    $w_s^{n+1} \in \arg \min_w \varphi_s^n(z^n, w)$
  - 5:    $(x_s^{n+1}, y_s^{n+1}) \in \Phi_s^n(z^n, w_s^{n+1})$  for all  $s \in S$
  - 6:    $z^{n+1} \leftarrow \arg \min_z \left\{ \sum_{s \in S} \pi_s^n \psi(z - x_s^{n+1}, w_s^{n+1} - y_s^{n+1}) \right\}$
  - 7:   Update  $\lambda_s^{n+1} \in \Lambda, \pi_s^{n+1} > 0$  for all  $s \in S$ , and  $\rho^{n+1} \geq \rho^n > 0$
  - 8: **end for**
  - 9: **return**  $(z^{n_{\max}}, w^{n_{\max}})$
- 

Algorithm 1 describes a two-block Gauss–Seidel iterative approach on the two blocks  $(x_s, y_s, w_s)_{s \in S}$  and  $z$ , where the mixed-integer constraints only appear in the block  $(x_s, y_s, w_s)_{s \in S}$  subproblem implicitly referred to in Lines 4–5 of Algorithm 1. In the following sections, we analyse the convergence properties of certain embedded subsequences of (mid-)iterations  $(x^{n_k+1}, y^{n_k+1}, w^{n_k+1}, z^{n_k})$  generated by Algorithm 1 for penalty coefficient values  $\rho^n > 0$ , penalty weights  $\pi_s^n, s \in S$ , and Lagrangian multipliers  $\lambda_s^n, s \in S$ , that vary with iteration  $n \geq 1$ . (It is convenient to maintain that  $\sum_{s \in S} \lambda_s^n = 0$  for all  $s \in S$  and all  $n \geq 0$ .) We must also assume the solution  $(x^{n_k+1}, y^{n_k+1}, w^{n_k+1})$  is globally optimal in order to carry out our convergence analysis in Sect. 5.1. Here we assume the existence of Fréchet subdifferentials at the minimising points, and this is assured for any global minima. Furthermore, when  $\psi$  is a quadratic form, a global minimum may be computed in practice using a MIQP solver.

**Remark 3** The classical progressive hedging algorithm is realised by taking  $\pi_s^n = p_s$  and so  $\rho^n = \sum_{s \in S} p_s \rho^n$  (as  $\sum_{s \in S} p_s = 1$ ). Then for  $\psi(\cdot, 0) = \frac{1}{2} \|\cdot\|^2$  we have  $z^{n+1} = \sum_{s \in S} p_s x_s^{n+1}$  and assuming dual feasibility  $\sum_{s \in S} \lambda_s^n = 0$  one can also assert for all  $s \in S$  that

$$(x_s^{n+1}, y_s^{n+1}) \in \Phi_s^n(z^n, w^{n+1}) = \underset{(x_s, y_s) \in K_s}{\operatorname{argmin}} f_s(x_s, y_s) + (\lambda_s^n)^\top x_s + \rho^n p_s \psi(z^n - x_s, 0)$$

along with the multiplier update that retains dual feasibility of the multipliers i.e.  $\lambda_s^{n+1} = \lambda_s^n - p_s \rho^n (z^{n+1} - x_s^{n+1})$ . Moreover, the dual feasibility allows one to assert

that the same  $z^{n+1}$  solves the minimisation with respect to  $z$  in the full augmented Lagrangian. Penalties  $\rho^n$  between iterations with progressive hedging are usually left unchanged or are updated in such a manner as to realise stabilisation. (See, e.g., [12, Section 3.4.1].)

Next, we build on that development where Algorithm 1 is viewed as an approximate two-block GS iterative approach within the *continuous* optimisation framework of successively minimising  $\varphi^n$  in  $z$  (approximately) and  $w$  (globally and exactly) with Lagrange multipliers  $\lambda^n \in \Lambda$ , penalty coefficient values  $\rho^n > 0$  and penalty weights  $\pi_s^n, s \in S$ , varying between iterations  $n \geq 0$  under certain assumptions.

We conclude this section by noting that the above algorithm is essentially that of [11], with alterations to the Lagrangian multiplier and penalty parameter updates. In particular, we consider what happens when Lagrange multipliers  $\lambda^n \in \Lambda$  and penalty weights  $\pi^n$  stop changing after a finite number of iterations, while penalty parameters  $\{\rho^n\}$  may increase without bound. The latter feature requires us to consider the limiting behaviour of the regularisation  $\varphi^n$  as  $\rho^n \rightarrow \infty$ . Such an analysis is facilitated by analysing the level curves of the sequence of functions, denoted by  $\text{lev}_c \varphi^{\lambda, \rho, \pi} := \{(z, w) \mid \varphi^{\lambda, \rho, \pi}(z, w) \leq c\} = \{(z, w) \mid \frac{1}{\rho} \varphi^{\lambda, \rho, \pi}(z, w) \leq \frac{c}{\rho}\} = \text{lev}_{\frac{c}{\rho}} \varphi^{\lambda, \rho, \pi}$ , prompting the use of epi-convergence as a tool in our analysis as this is associated with the convergence of level sets.

### 3 Properties of the SMIP regularisation $\varphi^{\lambda, \rho, \pi}$

The continuous regularisation  $\varphi^{\lambda, \rho, \pi}$  of SMIP (1) has properties that allow for feasible points of SMIP (1) to be associated with *certain local minima* of  $\varphi^{\lambda, \rho, \pi}$ . To gain insight into these properties of  $\varphi^{\lambda, \rho, \pi}$ , we first note some additional properties of  $\psi$  that follow from the properties listed in Assumption 2.

**Proposition 4** *Assume  $\psi$  satisfies Assumption 2. Then, for all  $(z, w) \in \mathbb{X} \times \mathbb{Y}^{|S|}$ ,  $\rho > 0$ ,  $\pi_s > 0, s \in S$ , and  $\lambda \in \Lambda$ , the following properties hold for each  $s \in S$ :*

1. *The set of solutions for problem (3), that is  $\Phi_s^{\lambda, \rho, \pi}(z, w)$ , is non-empty.*
2. *The function  $\rho \mapsto \varphi_s^{\lambda, \rho, \pi}(z, w_s)$  is non-decreasing.*
3. *If in addition  $(z, w_s) \in K_s$ , then  $\varphi_s^{\lambda, \rho, \pi}(z, w_s) \leq f_s(z, w_s)$ . If  $(z, w) \in K := \prod_{s \in S} K_s$  (with  $z := (z, z, \dots, z)$ ) then*

$$\varphi^{\lambda, \rho, \pi}(z, w) = \sum_{s \in S} \varphi_s^{\lambda, \rho, \pi}(z, w_s) \leq \sum_{s \in S} f_s(z, w_s) < +\infty. \tag{11}$$

4. *The function  $(z, w_s) \mapsto \varphi_s^{\lambda, \rho, \pi}(z, w_s)$  is locally Lipschitz continuous over*

$$K_s^\Delta := \{(z, w_s) \in \mathbb{X} \times \mathbb{Y}_s : (z, w_s) \in \text{conv}(K_s) + B_\Delta(0, 0)\}, \tag{12}$$

*with modulus  $\pi_s \rho L_s^\Delta$ , where  $L_s^\Delta$  depends on the diameter of  $\text{conv}(K_s) + B_\Delta(0, 0)$ . Taking  $L^\Delta := \max\{L_s^\Delta\}$ , we also have  $\varphi^{\lambda, \rho, \pi}$  is Lipschitz continuous with modulus  $\rho L^\Delta$ .*

**Proof** See Appendix A. □

**Definition 1** Denote  $\text{proj}_{\mathcal{I}}(\cdot) := \text{proj}_{\mathbb{X},\mathcal{I}}(\cdot) \times \text{proj}_{\mathbb{Y},\mathcal{I}}(\cdot)$ , integer-component projection and, for each  $(\bar{x}_{\mathcal{I}}, \bar{y}_{\mathcal{I}}) \in \text{proj}_{\mathcal{I}}(K) := \prod_{s \in S} \text{proj}_{\mathcal{I}}(K_s)$ , denote

$$K_s^{(\bar{x}_{\mathcal{I}}, \bar{y}_{\mathcal{I}})} := \{(x_s, y_s) \in K_s \mid (\bar{x}_{\mathcal{I}}, \bar{y}_{\mathcal{I}}) = \text{proj}_{\mathcal{I}}(x, y)\}.$$

Note that this corresponds to a polyhedral subset of  $K_s$  once we have removed the integrality constraint by fixing the integer variables at a specific integer value. We now consider the behaviour of  $\varphi^{\lambda, \rho, \pi}$  within neighbourhoods having progressively additional structure imposed. In preparation thereof, we introduce notation that facilitates the view of  $\varphi^{\lambda, \rho, \pi}$  in terms of its finitely numbered specific cross-sections over  $(x_{\mathcal{I}}, y_{\mathcal{I}}) \in \text{proj}_{\mathcal{I}}(K)$ .

Definition 1 induces the following notation for proximal cross-sections for each  $(x_{\mathcal{I}}^0, y_{\mathcal{I}}^0) \in \text{proj}_{\mathcal{I}}(K)$ .

**Definition 2** For  $(\lambda, \rho, \pi) \in \Lambda \times \mathbb{R}_{>0} \times \mathbb{R}_{>0}^{|S|}$ , the proximal cross-sectional values are defined by

$$\begin{aligned} \varphi_s^{\lambda, \rho, \pi} \left( z, w_s \mid x_{\mathcal{I}}^0, y_{\mathcal{I}}^0 \right) &:= \inf_{(x_s, y_s)} \{ f_s(x_s, y_s) + \lambda_s^\top x_s + \delta_{K_s^{(x_{\mathcal{I}}^0, y_{\mathcal{I}}^0)}}(x_s, y_s) \\ &\quad + \rho \pi_s \psi(z - x_s, w_s - y_s) \}. \end{aligned} \tag{13a}$$

$$\varphi^{\lambda, \rho, \pi} \left( z, w \mid x_{\mathcal{I}}^0, y_{\mathcal{I}}^0 \right) := \sum_{s \in S} \varphi_s^{\lambda, \rho, \pi} \left( z, w_s \mid x_{\mathcal{I}}^0, y_{\mathcal{I}}^0 \right), \tag{13b}$$

and the the set of arguments realising the proximal cross-sectional values are defined by

$$\begin{aligned} \Phi_s^{\lambda, \rho, \pi} \left( z, w_s \mid x_{\mathcal{I}}^0, y_{\mathcal{I}}^0 \right) &:= \underset{x, y}{\text{argmin}} \{ f_s(x_s, y_s) + \lambda_s^\top x_s + \delta_{K_s^{(x_{\mathcal{I}}^0, y_{\mathcal{I}}^0)}}(x_s, y_s) \\ &\quad + \rho \pi_s \psi(z - x_s, w_s - y_s) \}. \end{aligned} \tag{14a}$$

$$\Phi^{\lambda, \rho, \pi} \left( z, w \mid x_{\mathcal{I}}^0, y_{\mathcal{I}}^0 \right) := \prod_{s \in S} \Phi_s^{\lambda, \rho, \pi} \left( z, w_s \mid x_{\mathcal{I}}^0, y_{\mathcal{I}}^0 \right). \tag{14b}$$

For each  $s \in S$  and  $(x_{s,\mathcal{I}}, y_{s,\mathcal{I}}) \in \text{proj}_{\mathcal{I}}(K_s)$ , the properties of Assumption 2 for  $\psi$  allow for the following properties of the cross-sections to be established.

**Lemma 5** For  $(x_{\mathcal{I}}^0, y_{\mathcal{I}}^0) \in \text{proj}_{\mathcal{I}}(K)$  the mapping  $(z, w_s) \mapsto \varphi_s^{\lambda, \rho, \pi} \left( z, w_s \mid x_{\mathcal{I}}^0, y_{\mathcal{I}}^0 \right)$  is convex over  $\mathbb{R}^n \times \mathbb{R}^m$  for each  $s \in S$ .

**Proof** This function can be represented as the infimal convolution of two closed, convex functions

$$\begin{aligned} (x_s, y_s) &\mapsto f_s(x_s, y_s) + \lambda_s^\top x_s + \delta_{K_s^{(x_{\mathcal{I}}^0, y_{\mathcal{I}}^0)}}(x_s, y_s) \\ (u_s, v_s) &\mapsto \rho \pi_s \psi(u_s, v_s) \end{aligned}$$

with  $(z, w_s) = (x_s, y_s) + (u_s, v_s)$ . The compactness of  $K_s$  ensures that of  $K_s^{(x_s^0, y_s^0)}$ , which in turn ensures that the infimal convolution is bounded away from  $-\infty$ . As the strict epi-graph of an infimal convolution equals the sum of strict epi-graphs of the constituent functions, convexity follows [13, Exercise 1.28].  $\square$

Note that  $\varphi_s^{\lambda, \rho, \pi}(z, w_s) = \min_{(x_s, \mathcal{I}, y_s, \mathcal{I}) \in \text{proj}_{\mathcal{I}}(K_s)} \varphi_s^{\lambda, \rho, \pi}(z, w_s \mid x_s, \mathcal{I}, y_s, \mathcal{I})$ , for each  $s \in S$ , is a minimum of a finite number of convex functions, but  $\varphi_s^{\lambda, \rho, \pi}$  itself is not guaranteed to be convex or differentiable on its entire domain  $\mathbb{X} \times \mathbb{Y}^{|S|}$ . Nevertheless,  $\varphi_s^{\lambda, \rho, \pi}$  is locally convex and differentiable on open neighbourhoods  $N$  where, for all  $(z, w_s) \in N$ ,  $\varphi_s^{\lambda, \rho, \pi}(z, w_s) = \varphi_s^{\lambda, \rho, \pi}(z, w_s \mid x_s, \mathcal{I}, y_s, \mathcal{I})$  holds for exactly one  $(x_s, \mathcal{I}, y_s, \mathcal{I}) \in \text{proj}_{\mathcal{I}}(K_s)$ .

**Lemma 6** Assume  $\psi$  satisfies the Assumption 2, with parameter  $m$  as in Assumption 2(3). For each fixed  $D > 0$ , there exists a  $\tilde{\delta} > 0$  such that if a discrepancy  $(u^0, v^0)$  satisfies  $\|(u^0, v^0)\| < \tilde{\delta}$ , then  $\psi(u^0, v^0) < \psi(u, v)$  for all discrepancies  $(u, v)$  satisfying  $\|(u - u^0, v - v^0)\| > D$ .

**Proof** See Appendix A.  $\square$

For  $\psi$  defined by  $\psi(u, v) = \frac{1}{2} \|(u, v)\|^2$  and any fixed  $D > 0$ , we may identify  $\tilde{\delta} = \frac{1}{2}D$  (since  $m = 1$  and  $\nabla\psi(u, v) = (u, v)$ ), so that if  $D = 1/2$  for example, we have that for all  $(u^0, v^0)$  such that  $\|(u^0, v^0)\| < \frac{1}{4}$ , then  $\psi(u^0, v^0) < \psi(u, v)$  for all  $(u, v)$  such that  $\|(u - u^0, v - v^0)\| > \frac{1}{2}$ . This observation will have practical value in terms of separating values for different cross-sections of  $\varphi$ .

**Proposition 7** Assume  $\psi$  satisfies Assumption 2 and  $(z^0, w_s^0) \in K_s$  for all  $s \in S$ . If there is at least one scenario  $s \in S$  such that  $(x_s, \mathcal{I}, y_s, \mathcal{I}) \in \text{proj}_{\mathcal{I}}(K_s)$  with  $(z_{\mathcal{I}}^0, w_{s, \mathcal{I}}^0) \neq (x_s, \mathcal{I}, y_s, \mathcal{I})$ , then there exists a finite threshold penalty coefficient  $\tilde{\rho} > 0$  and a threshold  $\tilde{\delta} > 0$  such that for all  $\rho > \tilde{\rho}$  and  $0 < \delta < \tilde{\delta}$ , the strict inequality  $\varphi_s^{\lambda, \rho, \pi}(z, w_s) < \varphi_s^{\lambda, \rho, \pi}(z, w_s \mid x_s, \mathcal{I}, y_s, \mathcal{I})$  hold over  $(z, w_s) \in B_{\delta}(z^0, w_s^0)$ .

**Proof** Assuming for some  $s \in S$  that we have  $(z_{\mathcal{I}}^0, w_{s, \mathcal{I}}^0) \neq (x_s, \mathcal{I}, y_s, \mathcal{I})$ , then identifying  $(u_s, v_s) = (z - x_s, w_s - y_s)$  and  $(u_s^0, v_s^0) = (z - z^0, w_s - w_s^0)$  for each  $s \in S$ , we have  $(u_s - u_s^0, v_s - v_s^0) = (z^0 - x_s, w_s^0 - y_s)$  and  $\|(u_s - u_s^0, v_s - v_s^0)\| > \frac{1}{2}$  for all  $(x_s, y_s) \in K_s^{(x_{\mathcal{I}}, y_{\mathcal{I}})}$ . Thus, using Lemma 6 with  $D = \frac{1}{2}$ , we have a  $\tilde{\delta} > 0$  for which  $(z, w_s)$  satisfying  $\|(z - z^0, w_s - w_s^0)\| < \tilde{\delta}$  implies that

$$\psi(z - z^0, w_s - w_s^0) < \psi(z - x_s, w_s - y_s) \tag{15}$$

for all  $(x_s, y_s) \in K_s^{(x_{\mathcal{I}}, y_{\mathcal{I}})}$  given that  $(z_{\mathcal{I}}^0, w_{s, \mathcal{I}}^0) \neq (x_s, \mathcal{I}, y_s, \mathcal{I})$ .

Due to the compactness of  $K_s^{(x_{\mathcal{I}}, y_{\mathcal{I}})}$ , defining for fixed  $(z^0, w_s^0), s \in S$ ,

$$\Delta_s(z, w_s) := \min_{(x, y)} \left\{ \psi(z - x_s, w_s - y_s) - \psi(z - z^0, w_s - w_s^0) \mid (x_s, y_s) \in K_s^{(x_{\mathcal{I}}, y_{\mathcal{I}})} \right\},$$

we have  $\Delta_s(z, w_s) > 0$ . It follows for each  $s \in S' := \left\{s \mid (z_{\mathcal{I}}^0, w_{s,\mathcal{I}}^0) \neq (x_{s,\mathcal{I}}, y_{s,\mathcal{I}})\right\}$  that  $\psi(z - z^0, w_s - w_s^0) + \frac{\Delta_s(z, w_s)}{2} < \psi(z - x_s, w_s - y_s)$  for all  $(x_s, y_s) \in K_s^{(x_{\mathcal{I}}, y_{\mathcal{I}})}$ .

Again due to the compactness of  $K_s^{(x_{\mathcal{I}}, y_{\mathcal{I}})}$ , we have still for each  $s \in S'$  that

$$\begin{aligned} \frac{f_s(z^0, w_s^0) + \lambda_s^\top z^0}{\rho} + \pi_s \psi(z - z^0, w_s - w_s^0) + \frac{\pi_s \Delta_s(z, w_s)}{2} \\ < \frac{f_s(x_s, y_s) + \lambda_s^\top x_s}{\rho} + \pi_s \psi(z - x_s, w_s - y_s) \end{aligned}$$

for all  $(x_s, y_s) \in K_s^{(x_{\mathcal{I}}, y_{\mathcal{I}})}$  when  $\rho > \tilde{\rho} > 0$  sufficiently large. Thus, for  $\|(z - z^0, w_s - w_s^0)\| < \tilde{\delta}$  and  $\rho > \tilde{\rho} > 0$ , we have

$$\varphi_s^{\lambda, \rho, \pi}(z, w_s) < \varphi_s^{\lambda, \rho, \pi}(z, w_s \mid \mathbf{x}_{\mathcal{I}}, \mathbf{y}_{\mathcal{I}}) \tag{16}$$

for all  $\rho > \tilde{\rho}$  and  $(z, w_s) \in B_{\tilde{\delta}}(z^0, w_s^0)$  whenever  $s \in S'$ . (Otherwise, for  $s \in S \setminus S'$ ,  $\varphi_s^{\lambda, \rho, \pi}(z, w_s) = \varphi_s^{\lambda, \rho, \pi}(z, w_s \mid \mathbf{x}_{\mathcal{I}}, \mathbf{y}_{\mathcal{I}})$  holds.) Summing over  $s \in S$ , the same holds then for  $\varphi^{\lambda, \rho, \pi}(z, \mathbf{w}) < \varphi^{\lambda, \rho, \pi}(z, \mathbf{w} \mid \mathbf{x}_{\mathcal{I}}, \mathbf{y}_{\mathcal{I}})$ .  $\square$

We note that for each  $(z, w_s) \in B_{\tilde{\delta}}(z^0, w_s^0)$ , with  $s \in S'$ , we have  $\Delta_s(z, w_s) > 0$  and so the gap between the left and right hand sides of the inequality in (16) can only grow with increasing  $\rho$ . Recall that we seek elements of the set of feasible (non-anticipative) solutions is given by  $F := \{(z, \mathbf{w}) \mid (z, w_s) \in K_s \text{ for all } s \in S\}$ , which is distinct from the set  $K$ . The next result follows immediately.

**Corollary 8** *Assume  $\psi$  satisfies Assumption 2. If  $(z^0, w^0) \in F$ , then there exists a  $\tilde{\rho} > 0$  and a  $\tilde{\delta} > 0$  such that for  $\rho \geq \tilde{\rho}$  and  $0 < \delta < \tilde{\delta}$  we have*

$$\varphi^{\lambda, \rho, \pi}(z, \mathbf{w}) = \varphi^{\lambda, \rho, \pi}(z, \mathbf{w} \mid \mathbf{z}_{\mathcal{I}}^0, \mathbf{w}_{\mathcal{I}}^0) \text{ for all } (z, \mathbf{w}) \in B_{\delta}(z^0, \mathbf{w}^0).$$

Hence, for  $(z, w_s) \in B_{\delta}(z^0, w_s^0)$ ,  $s \in S$ , with  $0 < \delta < \tilde{\delta}$ , the function  $\varphi_s^{\lambda, \rho, \pi}$  coincides with a convex function for all  $\rho \geq \tilde{\rho}$ .

### 4 The theory of persistent local minima

For iteration indices  $n \geq 0$ , let  $(\lambda^n, \pi^n, \rho^n) \in \Lambda \times \mathbb{R}_{>0}^{|\mathcal{S}|} \times \mathbb{R}_{>0}$  and define for  $(\lambda^n, \pi^n, \rho^n) \rightarrow_{n \rightarrow \infty} (\lambda, \pi, \infty)$ :

$$f^n(\mathbf{x}, \mathbf{y}, \mathbf{w}, z) := \sum_{s \in S} f_s(x_s, y_s) + (\lambda_s^n)^\top x_s + \rho^n \pi_s^n \psi(z - x_s, y_s - w_s), \tag{17a}$$

$$\varphi_s^n := \varphi^{\lambda^n, \pi^n, \rho^n}; \varphi^n := \varphi^{\lambda^n, \pi^n, \rho^n}; \Phi_s^n := \Phi_s^{\lambda^n, \pi^n, \rho^n}; \Phi^n := \Phi^{\lambda^n, \pi^n, \rho^n}, \tag{17b}$$

$$\Phi^\infty(z, \mathbf{w}) := \prod_{s \in S} \arg \min_{(x_s, y_s)} \{\pi_s \psi(z - x_s, w_s - y_s) \mid (x_s, y_s) \in K_s\}. \tag{17c}$$

In this section, we consider sequences  $\{(\tilde{\mathbf{x}}^n, \tilde{\mathbf{y}}^n, \tilde{\mathbf{w}}^n, \tilde{z}^n)\}_{n=0}^\infty$  where we have  $(\tilde{\mathbf{x}}^n, \tilde{\mathbf{y}}^n) \in \Phi^n(\tilde{z}^n, \tilde{\mathbf{w}}^n)$ . Furthermore, we assume  $\lim_{n \rightarrow \infty} \rho^n = \infty$  and we single out a specific class of sequences of local minima  $\{(\tilde{z}^n, \tilde{\mathbf{w}}^n)\}_{n=0}^\infty$  for  $\varphi^n$ , which we call *persistent*, and which we show to be closely related to the feasible points of the underlying SMIP (1). We assume the following.

**Assumption 9** Solution Sequence Assumptions (SSA) on  $\{(\tilde{\mathbf{x}}^n, \tilde{\mathbf{y}}^n, \tilde{\mathbf{w}}^n, \tilde{z}^n)\}_{n=0}^\infty$  and  $\{(\lambda^n, \pi^n, \rho^n)\}_{n=0}^\infty$ , indexed by integers  $n \geq 0$ :

1. Penalty coefficients are non-decreasing  $\rho^{n+1} \geq \rho^n > 0$  for  $n \geq 0$  and increase without bound  $\lim_{n \rightarrow \infty} \rho^n = \infty$ .
2. Dual feasibility  $\lambda^n \in \Lambda$  is satisfied, and boundedness  $\limsup_{n \rightarrow \infty} \|\lambda^n\| < \infty$  holds.
3. Each  $\tilde{z}^n$  is a local minimum of the function  $z \mapsto \inf_w \varphi^n(z, w)$ .
4. The extracted sequence  $\{\tilde{z}^n\}_{n=0}^\infty$  converges to  $\bar{z}$ .
5. Each  $\tilde{\mathbf{w}}_s^n, s \in S$ , is *globally* optimal  $\tilde{\mathbf{w}}_s^n \in \arg \min_w \varphi_s^n(\tilde{z}^n, w_s)$ ; thus  $(\tilde{\mathbf{x}}^n, \tilde{\mathbf{y}}^n, \tilde{\mathbf{w}}^n) \in \arg \min_{\mathbf{w}, (x, y) \in K} \sum_{s \in S} f_s^n(x_s, y_s, w_s, \tilde{z}^n)$  is globally optimal with  $z = \tilde{z}^n$  fixed.

We also consider the following assumption on penalty weights separately.

**Assumption 10** Penalty Weight Assumptions (PWA): We assume that  $\sum_{s \in S} \pi_s^n = 1$  and  $\pi_s^n > 0, s \in S$ . We assume in addition that the applied update rule for generating penalty weights over iterations  $n \geq 0$  ensures that we have  $\pi_s^n \geq \xi$ , for some fixed  $\xi > 0$ , for all but a finite number of iterations  $n \geq 0$ , and for each  $s \in S$  such that  $\tilde{x}_{s, \mathcal{I}}^n \neq \tilde{z}_{\mathcal{I}}^n$  holds infinitely often in  $n$ . Furthermore, for  $n \geq 0$  for which the set  $S_n := \{s \in S \mid \tilde{x}_{s, \mathcal{I}}^n \neq \tilde{z}_{\mathcal{I}}^n\}$  is empty, we assume the penalty weight update rule also ensures that  $\pi_s^{n+1} = \pi_s^n$  for all  $s \in S$ .

If  $S_n$  is empty for all but a finite number of iterations  $n$ , then consensus  $\tilde{\mathbf{x}}_{s, \mathcal{I}}^n = \tilde{z}_{\mathcal{I}}^n$  has been reached and the above assumption is trivially satisfied. When  $S_n = \emptyset$  occurs, then  $\tilde{z}^n \in X$  and by relative complete recourse there exists  $y_s \in Y_s(\tilde{z}^n)$  for all  $s \in S$  so that  $(\tilde{z}^n, \mathbf{y})$  is feasible for SMIP (1).

In the context of Assumptions 9 and 10, we examine when it holds that  $\bar{z} \in X$ . Under Assumptions 9 and 10, local minimisers  $\tilde{z}^n$  of  $z \mapsto \inf_w \varphi^n(z, \mathbf{w})$  can be peculiar in the sense that  $\inf_w \varphi^n(\tilde{z}^n, \mathbf{w})$  can increase without bound as  $\rho^n \rightarrow \infty$ , while the maximal neighbourhoods of the local minimiser  $\bar{z}$  verifying local optimality of  $\tilde{z}^n$  for  $z \mapsto \inf_w \varphi^n(z, \mathbf{w})$  vanish in measure as  $n \rightarrow \infty$ . The local minimisers  $\tilde{z}^n$  that do *not* suffer from these issues are those that we wish to isolate, in that  $\inf_w \varphi^n(\tilde{z}^n, \mathbf{w})$  remains bounded at the local minimum  $\tilde{z}^n$  despite  $\rho^n \rightarrow \infty$ .

**Definition 3** Let  $\{(\tilde{\mathbf{x}}^n, \tilde{\mathbf{y}}^n, \tilde{\mathbf{w}}^n, \tilde{z}^n)\}_{n=0}^\infty$  and  $\{(\lambda^n, \pi^n, \rho^n)\}_{n=0}^\infty$  satisfy Assumption 9.

1. The sequence  $\{(\tilde{z}^n, \tilde{\mathbf{w}}^n)\}_{n=0}^\infty$  is *persistent* if  $\limsup_{n \rightarrow \infty} \varphi^n(\tilde{z}^n, \tilde{\mathbf{w}}^n) < \infty$ .
2. If  $\lim_{n \rightarrow \infty} (\tilde{z}^n, \tilde{\mathbf{w}}^n) = (\bar{z}, \bar{\mathbf{w}})$ , we say that  $(\bar{z}, \bar{\mathbf{w}})$  is a *persistent limit*.

**Remark 4** Clearly when we have a convergent sequence of local minima  $\tilde{z}^n \rightarrow \bar{z}$  for  $z \mapsto \inf_w \varphi^n(z, w), n \geq 0$ , then for any  $\tilde{w}^n$  with  $\inf_w \varphi^n(\tilde{z}^n, w) = \varphi^n(\tilde{z}^n, \tilde{w}^n), n \geq 0$ , we have any convergent subsequence  $\{(\tilde{z}^{n_k}, \tilde{w}^{n_k})\}_{k=0}^\infty$  converging to a persistent limit  $(\bar{z}, \bar{w})$ .

The subdifferential analysis of  $\varphi^{\lambda, \rho, \pi}$  requires addressing its nonconvexity and non-differentiability. A notion of differentiation suitable for this purpose is Fréchet subdifferentiability as defined in [13].

**Definition 4** The function  $\varphi : \mathbb{X} \times \mathbb{Y} \rightarrow \mathbb{R} \cup \{\infty\}$  is **Fréchet subdifferentiable** at  $(z^0, w^0)$  if there exists a **Fréchet subderivative**  $(\zeta, \omega)$  such that

$$\liminf_{(z-z^0, w-w^0) \rightarrow 0} \frac{\varphi(z, w) - \varphi(z^0, w^0) - \langle (\zeta, \omega), (z - z^0, w - w^0) \rangle}{\|(z - z^0, w - w^0)\|} \geq 0.$$

We denote the collection of all such subderivatives by  $\widehat{\partial}\varphi(z^0, w^0)$ , the **Fréchet subdifferential** of  $\varphi$  at  $(z^0, w^0)$ . A point  $(z^0, w^0)$  is **Fréchet stationary point** of  $\varphi$  if  $(0, 0) \in \widehat{\partial}\varphi(z^0, w^0)$ . The limiting, or **Mordukhovich subdifferential**, of  $\varphi$  is denoted  $\partial\varphi(\bar{z}, \bar{w})$ , where  $(\bar{\zeta}, \bar{\omega}) \in \partial\varphi(\bar{z}, \bar{w})$  if there exists a sequence  $\{(\zeta^n, \omega^n) \in \widehat{\partial}\varphi(z^n, w^n)\}_{n=0}^\infty$  such that  $(z^n, w^n) \rightarrow (\bar{z}, \bar{w})$  and  $(\zeta^n, \omega^n) \rightarrow (\bar{\zeta}, \bar{\omega})$ .

The first part of the following Lemma is a modest restatement of the cited result which we shall use to deduce differentiability whenever the Fréchet subdifferential is non-empty. In the second part we obtain local minimality from stationarity for structured functions.

**Lemma 11** Let  $\varphi : \mathbb{X} \times \mathbb{Y} \rightarrow \mathbb{R}_{+\infty}$  be a function defined by  $\varphi(\bar{z}, \bar{w}) := \min_{i \in I} \varphi_i(\bar{z}, \bar{w})$  where  $\{\varphi_i \mid \mathbb{X} \times \mathbb{Y} \rightarrow \mathbb{R}_{+\infty}\}_{i \in I}$  is a finite family of proper, convex, lower semicontinuous functions.

1. Then

$$\widehat{\partial}\varphi(\bar{z}, \bar{w}) = \bigcap_{i \in I(\bar{z}, \bar{w})} \widehat{\partial}\varphi_i(\bar{z}, \bar{w})$$

where  $I(\bar{z}, \bar{w}) := \{i \in I \mid i \in \arg \min_{i \in I} \varphi_i(\bar{z}, \bar{w})\}$ . If each function  $\varphi_i, i \in I$ , is differentiable then  $\widehat{\partial}\varphi(\bar{z}, \bar{w}) \neq \emptyset$  implies  $\partial\varphi(\bar{z}, \bar{w}) = \{\nabla\varphi(\bar{z}, \bar{w})\}$ .

2. If, in particular, Fréchet stationarity  $(0, 0) \in \widehat{\partial}\varphi_i(\bar{z}, \bar{w})$  is satisfied for all  $i \in I(\bar{z}, \bar{w})$ , then  $(\bar{z}, \bar{w})$  is a local minimum of  $\varphi$  with  $(0, 0) \in \widehat{\partial}\varphi(\bar{z}, \bar{w})$ . Moreover, if at least one of the  $\varphi_i, i \in I(\bar{z}, \bar{w})$  is differentiable then we also have  $\widehat{\partial}\varphi(\bar{z}, \bar{w}) = \nabla\varphi(\bar{z}, \bar{w}) = 0$ .

**Proof Part 1:** Follows due to [23, Theorem 1 via Theorem 10].

**Part 2:** Due to the convexity of  $\varphi_i, i \in I$ , we have  $(0, 0) \in \widehat{\partial}\varphi_i(\bar{z}, \bar{w})$  being a subgradient in both the Fréchet and classical sense for  $i \in I(\bar{z}, \bar{w})$ , and furthermore,  $(\bar{z}, \bar{w})$  is a globally optimal solution for each  $\varphi_i, i \in I(\bar{z}, \bar{w})$ . The membership  $(0, 0) \in \widehat{\partial}\varphi(\bar{z}, \bar{w})$  would follow immediately from Part 1. We claim that there exists  $\delta > 0$  such that  $\varphi(\bar{z}, \bar{w}) \leq \varphi(z, w)$  for all  $(z, w) \in B_\delta(\bar{z}, \bar{w})$ . For otherwise, for some

$i' \notin I(\bar{z}, \bar{w})$ , we have for  $(z, w)$  arbitrarily close to  $(\bar{z}, \bar{w})$  that  $\varphi(\bar{z}, \bar{w}) > \varphi_{i'}(z, w)$ . But since  $i' \notin I(\bar{z}, \bar{w})$ , the inequality  $\varphi(\bar{z}, \bar{w}) < \varphi_{i'}(\bar{z}, \bar{w})$  holds, and so the lower semicontinuity of  $\varphi_{i'}$  is contradicted. Thus, we have the local minimality  $\varphi(z, w) \geq \varphi(\bar{z}, \bar{w})$  for all  $(z, w) \in B_\delta(\bar{z}, \bar{w})$ . Moreover, if  $(0, 0) \in \widehat{\partial}\varphi(\bar{z}, \bar{w})$  and for at least one  $i \in I(\bar{z}, \bar{w})$  we have  $\widehat{\partial}\varphi_i(\bar{z}, \bar{w}) = \{(0, 0)\}$ , then by Part 1, we must also have  $\widehat{\partial}\varphi(\bar{z}, \bar{w}) = \{(0, 0)\}$  and so  $\nabla\varphi(\bar{z}, \bar{w}) = (0, 0)$  exists.  $\square$

The following motivates our set of assumptions on the penalty parameter update.

**Proposition 12** *Assume that SMIP (1) satisfies Assumption 1. Let  $\psi$  satisfy Assumption 2. Suppose we have a persistent local minima sequence  $(\tilde{z}^n, \tilde{w}^n) \rightarrow (\bar{z}, \bar{w})$  for  $\rho^n \rightarrow \infty$  (and hence Fréchet stationarity  $0 \in \widehat{\partial}\varphi^n(\tilde{z}^n, \tilde{w}^n)$  for each  $n$ ). If the PWA Assumption 10 holds, then for  $n$  sufficiently large and for all  $(\tilde{x}^n, \tilde{y}^n) \in \Phi^n(\tilde{z}^n, \tilde{w}^n)$ , we have  $\tilde{z}_I^n = \tilde{x}_{s,I}^n$  for all  $s \in S$  i.e. consensus holds in the integral components at a fixed value  $\tilde{z}_I^n = \bar{z}_I$ , and furthermore, the Fréchet stationarity  $0 \in \widehat{\partial}f^n(\tilde{x}^n, \tilde{y}^n, \tilde{w}^n, \tilde{z}^n)$  holds.*

**Proof** In general, the Fréchet stationarity  $0 \in \widehat{\partial}\varphi^n(\tilde{z}^n, \tilde{w}^n)$ ,  $n \geq 0$ , is a much stronger notion in that it allows us to deduce the Fréchet stationarity  $0 \in \widehat{\partial}f^n(\tilde{x}^n, \tilde{y}^n, \tilde{w}^n, \tilde{z}^n)$  via standard subderivative inclusions for marginal mappings (see [13, Theorem 10.13], Lemma 11 and elsewhere). Hence the Fréchet stationarity  $0 \in \widehat{\partial}\varphi^n(\tilde{z}^n, \tilde{w}^n)$  implies the cross-sectional Fréchet stationarity  $0 \in \widehat{\partial}\varphi^n(\tilde{z}^n, \tilde{w}^n \mid \mathbf{x}_I, \mathbf{y}_I)$  for all optimal cross-sections  $(\mathbf{x}_I, \mathbf{y}_I) \in \text{proj}_I(\Phi^n(\tilde{z}^n, \tilde{w}^n))$ . Identifying in terms of Lemma 11 (Part 2)  $\varphi$  with  $\varphi^n$ , the  $\varphi_i$ ,  $i \in I$ , with the finite number of cross-sections  $\varphi^n(\cdot, \cdot \mid \mathbf{x}_I, \mathbf{y}_I)$  with  $(\mathbf{x}_I, \mathbf{y}_I) \in \text{proj}_I(K)$ , and  $(\bar{z}, \bar{w})$  with  $(\tilde{z}^n, \tilde{w}^n)$ , we have a local minimum of  $\varphi^n$  at  $(\tilde{z}^n, \tilde{w}^n)$  and by the definition of  $\varphi^n$  we have the Fréchet stationarity  $0 \in \widehat{\partial}f^n(\tilde{x}^n, \tilde{y}^n, \tilde{w}^n, \tilde{z}^n)$ . From this, we show that all such solutions have a common set of integral values for  $n$  sufficiently large. As  $\{(\tilde{z}^n, \tilde{w}^n)\}_{n=0}^\infty$  is persistent there exists  $\kappa > 0$  such that  $\varphi^n(\tilde{z}^n, \tilde{w}^n) \leq \kappa$  for all  $n$  sufficiently large. Hence for each optimal cross-section  $(\tilde{x}_I^n, \tilde{y}_I^n) \in \text{proj}_I(\Phi^n(\tilde{z}^n, \tilde{w}^n))$  we have, using (5),

$$\begin{aligned} & \frac{1}{\rho^n} [\kappa + (\|c\| + \|d\| + \limsup_{n' \rightarrow \infty} \|\lambda^{n'}\|) (\sup_{(x,y) \in K} \max\{\|x\|, \|y\|\})] \\ & \geq \sum_{s \in S} \pi_s^n \psi(\tilde{z}^n - \tilde{x}_s^n, 0) \geq \frac{m}{2} \sum_{s \in S} \pi_s^n \|\tilde{z}_I^n - \tilde{x}_{s,I}^n\|^2 \end{aligned} \tag{18}$$

The left-hand side of (18) tends to zero as  $\rho^n \rightarrow \infty$  and  $\pi_s^n \geq \xi$  for all  $s \in S_n$ . After choosing a small  $0 < \delta < \frac{1}{2|\mathcal{S}|}$  we conclude that  $\|\tilde{x}_{s,I}^n - \tilde{z}_I^n\| < \frac{1}{2}$  for all  $s \in S_n$  and so  $\tilde{x}_{s',I}^n = \tilde{x}_{s,I}^n = \tilde{z}_I^n$  for all  $s, s' \in S_n$ . As  $\tilde{z}_I^n = \tilde{x}_{s,I}^n$  for all  $s \notin S_n$  already, we have equality for all  $s \in S$  and as  $\tilde{x}_{s,I}^n = \bar{x}_{s,I} = \bar{z}_I$  are fixed independent of  $\rho^n$  for  $n$  sufficiently large.  $\square$

Feasibility may also be shown to hold, as stated in Lemma 13. Furthermore, in Proposition 14 we state the relationships between persistency and feasibility.

**Lemma 13** *Let the problem SMIP (1) satisfy the SMIP Assumption 1, and let penalty function  $\psi$  satisfy Assumption 2. If a sequence  $\{(\tilde{x}^n, \tilde{y}^n, \tilde{w}^n, \tilde{z}^n)\}_{n=0}^\infty$  given*



$\{(\lambda^n, \pi^n, \rho^n)\}_{n=0}^\infty$  with integer index  $n \geq 0$  satisfies the Assumption 9, then  $\tilde{y}_s^n = \tilde{w}_s^n \in Y_s(\tilde{x}_s^n)$ ,  $s \in S$ , and  $\tilde{z}^n \in \arg \min_z \sum_{s \in S} \pi_s^n \psi(z - \tilde{x}_s^n, 0)$ . Furthermore, for each  $n \geq 0$  for which  $\tilde{z}^n \in X$  holds, we have  $\varphi^n(\tilde{z}^n, \tilde{w}^n) = \inf_w \varphi^n(\tilde{z}^n, w)$  bounded from above independent of the specific value of  $\rho^n$ .

**Proof** It follows from Lemma 4 that for all  $n \geq 0$ , we have the existence of  $(\tilde{x}^n, \tilde{y}^n) \in K$  that attains the infimum in the definition of  $\varphi^n(\tilde{z}^n, \tilde{w}^n)$ . Because  $\tilde{w}_s^n, s \in S$ , is a global optimum for  $w_s \mapsto \varphi_s^n(\tilde{z}^n, w)$ , the claim  $\tilde{y}_s^n = \tilde{w}_s^n, s \in S$  follows readily from Proposition 3.

To establish that  $\tilde{z}^n \in \arg \min_z \sum_{s \in S} \pi_s^n \psi(z - \tilde{x}_s^n, 0)$ , assume that  $\tilde{z}^n \notin \arg \min_z \sum_{s \in S} \pi_s^n \psi(z - \tilde{x}_s^n, 0)$ . For any  $\bar{z}^n \in \arg \min_z \sum_{s \in S} \pi_s^n \psi(z - \tilde{x}_s^n, 0)$ , using the convexity of  $\psi$ , and  $\eta \in (0, 1)$  we have

$$\begin{aligned} & \sum_{s \in S} \inf_w \varphi_s^n(\eta \bar{z}^n + (1 - \eta)\tilde{z}^n, w_s) \\ & \leq \sum_{s \in S} p_s \left[ c^\top \tilde{x}_s^n + d_s^\top \tilde{w}_s^n \right] + (\lambda_s^n)^\top \tilde{x}_s^n + \rho^n \pi_s^n \psi([\eta \bar{z}^n + (1 - \eta)\tilde{z}^n] - \tilde{x}_s^n, 0) \\ & \leq \eta \sum_{s \in S} p_s \left[ c^\top \tilde{x}_s^n + d_s^\top \tilde{w}_s^n \right] + (\lambda_s^n)^\top \tilde{x}_s^n + \rho^n \pi_s^n \psi(\bar{z}^n - \tilde{x}_s^n, 0) \\ & \quad + (1 - \eta) \sum_{s \in S} p_s \left[ c^\top \tilde{x}_s^n + d_s^\top \tilde{w}_s^n \right] + (\lambda_s^n)^\top \tilde{x}_s^n + \rho^n \pi_s^n \psi(\tilde{z}^n - \tilde{x}_s^n, 0) \\ & < \varphi^n(\tilde{z}^n, \tilde{w}^n), \end{aligned}$$

which would contradict the local optimality of  $\tilde{z}^n$  for  $z \mapsto \inf_w \varphi^n(z, w)$ . Thus,  $\tilde{z}^n \in \arg \min_z \sum_{s \in S} \pi_s^n \psi(z - \tilde{x}_s^n, 0)$  for all  $n \geq 0$ .

Furthermore, it also follows that, when  $\tilde{z}^n \in X$  (a compact set):

$$\begin{aligned} \varphi^n(\tilde{z}^n, \tilde{w}^n) &= \sum_{s \in S} p_s^n \left[ c^\top \tilde{x}_s^n + d_s^\top \tilde{y}_s^n \right] + (\lambda_s^n)^\top \tilde{x}_s^n + \rho^n \pi_s^n \psi(\tilde{z}^n - \tilde{x}_s^n, \tilde{w}_s^n - \tilde{y}_s^n) \\ &\leq \sup_n \inf_{\tilde{w} \in Y(\tilde{z}^n)} \sum_{s \in S} p_s [c^\top \tilde{z}^n + d_s^\top \tilde{w}_s] + (\lambda_s^n)^\top \tilde{z}^n \leq \Gamma < \infty. \end{aligned}$$

where, after noting that  $\sum_{s \in S} (\lambda_s^n)^\top \tilde{z}^n = 0$  vanishes due to  $\lambda \in \Lambda$ , we have that  $\Gamma < \infty$  can be chosen to hold regardless of the specific realisations of  $\rho^n > 0$  and  $\tilde{z}^n \in X$  due to the boundedness properties of the SMIP Assumptions; the finiteness of  $\Gamma$  also requires the assumed relatively complete recourse. □

**Proposition 14** Assume  $\psi$  satisfies Assumption 2. If  $(\bar{z}, \bar{w})$  is a persistent limit for a persistent sequence  $\{(\tilde{z}^n, \tilde{w}^n)\}_{n=0}^\infty$  then

1.  $(\bar{z}, \bar{w}) \in F$ ; namely  $\bar{z} \in X$  and  $\bar{w}_s \in Y_s(\bar{z})$ ;
2. there is a fixed neighbourhood  $B_\delta(\bar{z}, \bar{w})$  with  $\delta > 0$  on which  $(\tilde{z}^n, \tilde{w}^n)$  is locally optimal for  $\varphi^n$  for all  $n$  large enough (i.e., for all  $\rho^n$  larger than some threshold  $\tilde{\rho}$ ).

If we furthermore assume that the PWA Assumption 10 holds, then we have  $\tilde{z}_T^n = \tilde{x}_{s,T}^n$  for all  $s \in S$  for all  $(\tilde{x}^n, \tilde{y}^n) \in \Phi^n(\tilde{z}^n, \tilde{w}^n)$  for  $n$  large enough.

**Proof** To prove (1), suppose  $(\bar{z}, \bar{w}) \notin F$ . Then there exists  $\delta > 0$  such that  $\inf_{(x_s, y_s) \in K_s} \|(\bar{z} - x_s, \bar{w}_s - y_s)\|^2 \geq 2\delta$  for at least one scenario  $s \in S$ . As  $(\tilde{z}^n, \tilde{w}^n) \rightarrow (\bar{z}, \bar{w})$  we have, for  $n$  (and thus  $\rho^n$ ) large enough that,  $\inf_{(x, y) \in K_s} \|(\tilde{z}^n - x_s, \tilde{w}_s^n - y_s)\|^2 \geq \delta$  for some  $s \in S_n$  (in which case  $\tilde{z}^n \neq \tilde{x}_s^n$  since by Proposition 3 we have  $\tilde{w}^n = \tilde{y}^n$ ). We now use Assumption 2 (3) to bound the penalty values below. By the differentiability assumed in Assumption 2 we apply the inequality (5) for each  $s \in S$  to get  $\psi(\tilde{z}^n - x_s, \tilde{w}_s^n - y_s) \geq \frac{m}{2} \|(\tilde{z}^n - x_s, \tilde{w}_s^n - y_s)\|^2$ . It follows that, as  $\lim_{n \rightarrow \infty} \rho^n = \infty$ , we have  $\liminf_n \pi_s^n \geq \xi$  so

$$\begin{aligned} \min_{(x_s, y_s) \in K_s} \left\{ \rho^n \sum_{s \in S} \pi_s^n \psi(\tilde{z}^n - x_s, \tilde{w}_s^n - y_s) \right\} \\ \geq \frac{m\rho^n\xi}{2} \inf_{(x_s, y_s) \in K_s} \left\{ \|(\tilde{z}^n - x_s, \tilde{w}_s^n - y_s)\|^2 \right\} \geq \frac{m\rho^n\xi}{2} \delta \end{aligned}$$

which is unbounded. This contradicts the assumption that  $\{\varphi^n(\tilde{z}^n, \tilde{w}^n)\}_{n=0}^\infty$  is bounded above as required by the persistency assumption on  $\{\tilde{z}^n\}_{n=0}^\infty$ . Thus  $(\bar{z}, \bar{w}) \in F$ .

Having shown  $(\bar{z}, \bar{w}) \in F$ , it follows from definitions that  $\bar{w}_s \in Y_s(\bar{z})$  and so claim (2) follows readily from Corollary 8 using the same critical  $\tilde{\rho}$  and  $\tilde{\delta}$  that apply regardless of the choice of  $\bar{z} \in X$ , and so it is established that  $(z, w) \mapsto \varphi^n(z, w)$  is convex over  $B_\delta(\bar{z}, \bar{w})$  for all  $\rho^n > \tilde{\rho}$  with  $n$  sufficiently large. By Remark 4 we have the same neighbourhood associated with the local minimum at  $(\bar{z}, \bar{w})$  also associated with a persistent local minimum at some  $(\bar{z}, \bar{w})$  and thus  $B_\delta(\bar{z}, \bar{w})$  serves as the fixed neighbourhood verifying local optimality of  $(\tilde{z}^n, \tilde{w}^n)$  for  $(z, w) \mapsto \varphi^n(z, w)$  for each  $n$  large enough. The last claim follows from Proposition 12.  $\square$

The following contains a version of the strong augmented duality result for augmented Lagrangian. Notice that this result is more general than those in [15], in that it allows for the consideration of an inexact penalty that may be differentiable everywhere. When we have pure integer variables we see that all feasible solutions are persistent and we have a stronger form of duality.

**Theorem 15** *Suppose problem SMIP (1) satisfies the SMIP Assumption 1 and  $\psi$  is an ICRF.*

1. *If problem SMIP (1) has pure integer variable in both stages and feasible point  $(\bar{z}, \bar{w}) \in F$  satisfies  $\limsup_n \varphi^n(\bar{z}, \bar{w}) < +\infty$  for  $\lim_{n \rightarrow \infty} \rho^n = \infty$ , then  $(\bar{z}, \bar{w}) \in F$  is a local minimum of  $\varphi^n$  for  $n$  large enough.*
2. *For any sequence  $\{(\tilde{z}^n, \tilde{w}^n)\}_{n=0}^\infty$  of global minimisers of  $\varphi^n$  with  $\lim_{n \rightarrow \infty} \rho^n = \infty$  and  $\lambda^n, n \geq 0$ , satisfying Assumption 9(2), and  $\pi_s > 0, s \in S$ , then its limit points  $(\bar{z}, \bar{w})$  are globally optimal solutions to SMIP (1). That is, there exists at least one globally optimal solution  $(\bar{z}, \bar{w})$  to SMIP (1) that is a persistent limit. Moreover, for any  $\{\rho^n\}_{n=0}^\infty$  with  $\rho^n \rightarrow \infty$  there exists a persistent local minimum sequence  $\{(\tilde{z}^n, \tilde{w}^n)\}_{n=0}^\infty$  for which  $\lim_{n \rightarrow \infty} \varphi^n(\tilde{z}^n, \tilde{w}^n) = \zeta^{SMIP}$*

3. We have for any  $\lambda \in \Lambda, \pi_s > 0, s \in S,$

$$\sup_{\rho > 0} \min_{(z, \mathbf{w}) \in \mathbb{X} \times \mathbb{Y}^{|S|}} \varphi^{\lambda, \rho, \pi}(z, \mathbf{w}) = \zeta^{SMIP}. \tag{19}$$

Moreover, for a pure integer SMIP a finite value of  $\bar{\rho} > 0$  exists for which  $\min_{(z, \mathbf{w}) \in \mathbb{X} \times \mathbb{Y}^{|S|}} \varphi^{\lambda, \rho, \pi}(z, \mathbf{w}) = \zeta^{SMIP}$  for  $\rho \geq \bar{\rho}.$

**Proof 1):** Suppose  $(\bar{z}, \bar{\mathbf{w}}) \in F.$  Using Lemma 7 and Corollary 8 we have a locally convex function

$$(z, \mathbf{w}) \mapsto \varphi^n(z, \mathbf{w}) = \varphi^n(z, \mathbf{w} \mid \bar{z}_I, \bar{\mathbf{w}}_I)$$

for all  $(z, \mathbf{w}) \in B_\delta(\bar{z}, \bar{\mathbf{w}})$  for some fixed  $\delta > \tilde{\delta} > 0$  and  $\rho^n > \tilde{\rho} > 0$  with  $n$  large enough. Moreover for all  $(z', \mathbf{w}') \in F$  with  $(\bar{z}_I, \bar{\mathbf{w}}_I) \neq (z'_I, \mathbf{w}'_I)$  we have  $\varphi^n(z, \mathbf{w} \mid z'_I, \mathbf{w}'_I) > \varphi^n(z, \mathbf{w} \mid \bar{z}_I, \bar{\mathbf{w}}_I)$  for  $(z, \mathbf{w}) \in B_\delta(\bar{z}, \bar{\mathbf{w}})$  for some fixed  $\delta > \tilde{\delta} > 0$  and  $\rho^n > \tilde{\rho} > 0$  with  $n$  large enough. Supposing  $(\bar{z}, \bar{\mathbf{w}}) \in F$  is pure integer, then we have  $(\bar{z}_I, \bar{\mathbf{w}}_I) = (\bar{z}, \bar{\mathbf{w}})$  and hence  $(\bar{z}, \bar{\mathbf{w}})$  is a local minimum of  $\varphi^n$  with  $\varphi^n(\bar{z}, \bar{\mathbf{w}}) \leq \sum_{s \in S} c^\top \bar{z} + d_s^\top \bar{w}_s < +\infty,$  due to the boundedness assumptions for the SMIP (and dual feasibility of any sequence  $\{\lambda^n\}_{n=0}^\infty).$

2): Let  $\{(\tilde{z}^n, \tilde{\mathbf{w}}^n)\}_{n=0}^\infty$  be a sequence where each  $(\tilde{z}^n, \tilde{\mathbf{w}}^n), n \geq 0,$  is a global minimiser of  $\varphi^n$  which implies  $\varphi^n(\tilde{z}^n, \tilde{\mathbf{w}}^n) \leq \zeta^{SMIP},$  so that  $\{(\tilde{z}^n, \tilde{\mathbf{w}}^n)\}_{n=0}^\infty$  is a persistent sequence. Its limit points  $(\bar{z}, \bar{\mathbf{w}})$  thus satisfy  $(\bar{z}, \bar{\mathbf{w}}) \in F$  by Proposition 14. Furthermore  $(\tilde{x}_s^n, \tilde{y}_s^n) \in \Phi_s^n(\tilde{z}^n, \tilde{w}_s^n)$  and as  $\Phi^\infty(\bar{z}, \bar{\mathbf{w}}) = \{(\bar{z}, \bar{\mathbf{w}})\}$  we have (after passing to the subsequence)  $\lim_{n \rightarrow \infty} (\tilde{x}_s^n, \tilde{y}_s^n) = (\bar{z}, \bar{w}_s) \in K_s.$  (For if not, the boundedness of  $\{\lambda^n\}_{n=0}^\infty, \rho^n \rightarrow \infty, \pi_s > 0, s \in S,$  and the minorisation (5) would imply that  $\limsup_{n \rightarrow \infty} \varphi^n(\tilde{z}^n, \tilde{\mathbf{w}}^n) = \infty.$ ) Furthermore, since  $\varphi^n(\tilde{z}^n, \tilde{\mathbf{w}}^n) \leq \zeta^{SMIP},$  we must have also  $\sum_{s \in S} f_s(\tilde{x}_s^n, \tilde{y}_s^n) + (\lambda_s^n)^\top \tilde{x}_s^n \leq \zeta^{SMIP}$  and so  $\sum_{s \in S} f_s(\bar{z}, \bar{w}_s) = \zeta^{SMIP}$  (due to the boundedness and dual feasibility  $\lambda^n \in \Lambda$  and  $\tilde{x}_s^n \rightarrow \bar{z}$  for all  $s \in S,$  we have  $\limsup_{n \rightarrow \infty} \sum_{s \in S} \lambda_s^n \tilde{x}_s^n = 0$ ) and thus,  $(\bar{z}, \bar{\mathbf{w}})$  must be optimal for the original SMIP (1).

3). Denote

$$\xi_\rho^{SMIP} := \min\{\varphi^{\lambda, \rho', \pi}(z^{\rho'}, \mathbf{w}^{\rho'}) \mid \{(z^{\rho'}, \mathbf{w}^{\rho'})\} \text{ are persistent for } \rho' \geq \rho\}.$$

If  $(\bar{z}, \bar{\mathbf{w}})$  is a persistent limit, Lemma 14 implies  $(\bar{z}, \bar{\mathbf{w}}) \in F$  and by Proposition 11  $\varphi^{\lambda, \rho, \pi}(\bar{z}, \bar{\mathbf{w}}) \leq \sum_{s \in S} p_s [c^\top \bar{z} + d_s^\top \bar{w}_s].$  It follows that:

$$\lim_{\rho \rightarrow \infty} \xi_\rho^{SMIP} \leq \min_{(\bar{z}, \bar{\mathbf{w}})} \left\{ \sum_{s \in S} p_s [c^\top \bar{z} + d_s^\top \bar{w}_s] \mid (\bar{z}, \bar{\mathbf{w}}) \text{ are persistent limits} \right\} \leq \zeta^{SMIP},$$

where the last inequality follows from the existence of global solutions that are limits of persistent local minima. Let  $\rho^n \rightarrow \infty$  and  $\{(\tilde{z}^n, \tilde{\mathbf{w}}^n)\}_{n=0}^\infty$  be a sequence of persistent local minima, globally minimising  $\varphi^n$  with  $\varphi^n(\tilde{z}^n, \tilde{\mathbf{w}}^n) \rightarrow \zeta^{SMIP}.$  Via global

optimality  $\varphi^n(\tilde{z}^n, \tilde{w}^n) \leq \xi_{\rho^n}^{SMIP}$ , from which it follows that  $\sup_{\rho>0} \xi_{\rho}^{SMIP} = \zeta^{SMIP}$ . As all global minima are eventually persistent we are finished.

When we have a pure integer SMIP then by Part 1, we have the existence of a  $\bar{\rho} > 0$  such that  $\varphi^{\lambda,\rho,\pi}(z^\rho, w^\rho) = \varphi^{\lambda,\rho,\pi}(\bar{z}, \bar{w})$  for all  $\rho \geq \bar{\rho}$ , where  $(\bar{z}, \bar{w})$  is a global minimum of the SMIP. Hence, a global minimum is achieved for a finite  $\rho$ .  $\square$

We now investigate the role of the fixed neighbourhoods verifying the local minima for  $\varphi^n, n \geq 0$ . Indeed, for limiting points that are not persistent, we show that such a neighbourhood does not exist.

**Assumption 16** The sequence  $\{(\tilde{x}^n, \tilde{y}^n, \tilde{w}^n, \tilde{z}^n), (\lambda^n, \pi^n, \rho^n)\}_{n=0}^\infty$  satisfies the *joint PH assumptions* (joint PHA) when:

1. The problem SMIP (1) satisfies the SMIP Assumption 1,
2. The penalty function  $\psi$  meets the Integer Compatibility Regularisation Functions Assumptions, (ICRF) given in Assumption 2, and
3. the sequences  $\{(\tilde{x}^n, \tilde{y}^n, \tilde{w}^n, \tilde{z}^n)\}_{n=0}^\infty$  and  $\{(\lambda^n, \pi^n, \rho^n)\}_{n=0}^\infty$  with integer index  $n \geq 0$  satisfy the Solution Sequence Assumptions (SSA) in Assumption 9 and Penalty Weighting (PWA) given in Assumption 10.

**Proposition 17** Assume  $\{(\tilde{x}^n, \tilde{y}^n, \tilde{z}^n, \tilde{w}^n), (\lambda^n, \pi^n, \rho^n)\}_{n=0}^\infty$  satisfies the joint PHA Assumption 16. If the radii  $\delta_n, n \geq 0$ , on which  $(\tilde{z}^n, \tilde{w}^n)$  are locally optimal for  $\varphi^n$  satisfies  $\liminf_{n \rightarrow \infty} \delta_n = \bar{\delta}$  for some  $\bar{\delta} > 0$ , then

1.  $\lim_{n \rightarrow \infty} \|\tilde{z}^n - \tilde{x}_s^n\| = 0$  for all  $s \in S$  for which  $\limsup_{n \rightarrow \infty} \pi_s^n > 0$ . Thus,  $\bar{z} \in X, \lim_{n \rightarrow \infty} \sum_{s \in S} \pi_s^n \tilde{x}_s^n \rightarrow \bar{z}$  and for  $n$  sufficiently large, we have  $\tilde{z}_I^n = \tilde{x}_{s,I}^n$  for all  $s \in S$ , and  $\sum_{s \in S} \pi_s \tilde{x}_s^n \in X$ . (When  $\psi = \frac{1}{2} \|\cdot\|^2$  we have  $\tilde{z}^n = \sum_{s \in S} \pi_s \tilde{x}_s^n \in X$  for all  $n \geq 0$ .)
2. For all  $n$  we have  $\tilde{w}_s^n = \tilde{y}_s^n \in Y_s(\tilde{x}_s^n)$  and limit points  $\bar{w}_s$  of  $\{\tilde{w}_s^n\}_{n=0}^\infty$  satisfy  $\bar{w}_s \in Y_s(\bar{z})$  for all  $s \in S$ , with  $\limsup_{n \rightarrow \infty} \pi_s > 0$ .

**Proof** See Appendix A.  $\square$

The previous analysis allows us to pose the following result which confirms that the “basis of attraction” of non-persistent local minima has no interior in the limit. The next result follows immediately as contra-positives of Proposition 17.

**Corollary 18** Let  $\{(x^n, y^n, w^n, z^n), (\lambda^n, \pi^n, \rho^n)\}_{n=0}^\infty$  satisfy the joint PH Assumption 16. If any one of the following is true:

1.  $\bar{z} \notin X$ ,
2. There exist arbitrarily large  $n$  such that  $z_I^n \neq x_{s,I}^n$  for at least one  $s \in S$ ,
3.  $\lim_{n \rightarrow \infty} x_s^n \neq \bar{z}$  or  $\lim_{n \rightarrow \infty} x_s^n$  does not exist for at least one  $s \in S$  for which  $\limsup_{n \rightarrow \infty} \pi^n > 0$  and
4. when  $\psi = \frac{1}{2} \|\cdot\|^2, z^n \not\rightarrow \bar{z}$ ,

then  $\lim_{n \rightarrow \infty} \delta^n = 0$  for the radii  $\delta^n > 0, n \geq 0$ , on which the local optimality of each  $\tilde{z}^n$  for  $z \mapsto \inf_w \varphi^n(z, w)$  is verified.

**Example 1** Consider an augmented Lagrangian reformulation of a simple split variable extensive form of a two-stage SMIP

$$\min_{x,y,z,w} \left\{ \sum_{s=1}^2 [c^T x_s + d_s^T y_s] + \rho \pi_s \psi(z - x_s, w_s - y_s) \mid (x_1, y_1) \in K_1, (x_2, y_2) \in K_2, \right. \\ \left. x_1 = z, x_2 = z, y_1 = w_1, y_2 = w_2, x_1, x_2 \in \{0, 1\}, y_1, y_2 \in [0, 1] \right\} \tag{20}$$

with penalty coefficient  $\rho > 0$  where  $K_1 = \{(x, y) \in \{0, 1\} \times [0, 1] \mid x \leq y\} \supset \{(0, 0), (1, 1)\}$ ,  $K_2 = \{(x, y) \in \{0, 1\} \times [0, 1] \mid 1 - x \leq y\} \supset \{(0, 1), (1, 0)\}$ ,  $c = 0$ ,  $d_1 = 1 = d_2$ , and  $\psi(u, v) = \|(u, v)\|^2$  for  $s = 1, 2$ . We assume  $p_1 = p_2 = \pi_1 = \pi_2 = \frac{1}{2}$ , and  $\lambda = 0$  throughout this example. For any  $\{\rho^n\}_{n=0}^\infty$  with  $\rho^n > 0, n \geq 0$ , one may verify that the local minimisers  $\tilde{z}^n, \tilde{w}^n$  and local optimal values as parameterised by  $\rho^n, n \geq 0$ , are as follows

The locally optimal solutions  $(\tilde{z}^n, \tilde{w}^n)$  are the same for all  $\rho > 0$ , so that  $(\tilde{z}^n, \tilde{w}^n) = (\bar{z}, \bar{w})$  for  $n \geq 0$  for each locally optimal  $(\bar{z}, \bar{w})$ . Here we see that the two globally optimal solutions for  $\varphi^{\lambda, \rho, \pi}$  are the persistent solutions with either  $\bar{z} = \bar{x}_1 = \bar{x}_2 = 0$  or  $\bar{z} = \bar{x}_1 = \bar{x}_2 = 1$ , which both satisfy non-anticipativity. The non-persistent solution has  $\bar{z} = 0.5$  with  $0 = \bar{x}_1 \neq \bar{x}_2 = 1$ ; it only stays optimal over an ever shrinking neighbourhood  $B_\delta(\bar{z}, \bar{w})$  with radius  $\delta = 1/\rho^n$  vanishing as  $\rho^n \rightarrow \infty$ .

Solution $(\tilde{z}^n, \tilde{y}^n, \tilde{z}^n, \tilde{w}^n)$	Value $\varphi^n(\tilde{z}^n, \tilde{w}^n)$	Locally Optimal Over $B_\delta(\bar{z}, \bar{w})$ $(\bar{z}, \bar{w})$	$\delta$	Persistent?
$\left( \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix}, 0, \begin{pmatrix} 0 \\ 1 \end{pmatrix} \right)$	1	$\left( 0, \begin{pmatrix} 0 \\ 1 \end{pmatrix} \right)$	$\frac{1}{2} - \frac{1}{\rho^n}$	Yes
$\left( \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \frac{1}{2}, \begin{pmatrix} 0 \\ 0 \end{pmatrix} \right)$	$\frac{\rho^n}{4}$	$\left( \frac{1}{2}, \begin{pmatrix} 0 \\ 0 \end{pmatrix} \right)$	$\frac{1}{\rho^n}$	No
$\left( \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \end{pmatrix}, 1, \begin{pmatrix} 1 \\ 0 \end{pmatrix} \right)$	1	$\left( 1, \begin{pmatrix} 1 \\ 0 \end{pmatrix} \right)$	$\frac{1}{2} - \frac{1}{\rho^n}$	Yes

### 5 Analysis of the block Gauss–Seidel sequence

Block Gauss–Seidel iterations are most easily analysed for differentiable optimisation problems. However, we need to perform Gauss–Seidel iterations on nonsmooth functions with varying parameterisations and, hence, we develop the necessary theory to facilitate this analysis. We start with statements of elementary definitions and properties of Gauss–Seidel iterations that apply under general assumptions on the function and their domain sets.

**Definition 5** Let  $G : \mathbb{X} \times \mathbb{Y} \rightarrow \mathbb{R}_{+\infty}$  be a continuous function over a closed subset of  $\mathbb{X} \times \mathbb{Y}$ . A solution  $(z^*, w^*) \in \mathbb{X} \times \mathbb{Y}$  is a *partial minimum* of  $G$  if

$$G(z^*, w) \geq G(z^*, w^*) \quad \text{for all } w \in \mathbb{Y} \quad \text{and} \tag{21a}$$

$$G(z, w^*) \geq G(z^*, w^*) \quad \text{for all } z \in \mathbb{X}. \tag{21b}$$

For general non-smooth  $G$ , partial minimality does not imply (joint) minimality. Under suitable assumptions of convexity and (additive) separability of non-smoothness in  $G$ , we may recover joint minimality as described in Lemma 21

**Assumption 19 Separability and Convexity Assumptions (SCA)** on  $G : \mathbb{X} \times \mathbb{Y} \rightarrow \mathbb{R}_{+\infty}$ :

1.  $G$  is bounded from below and its level sets are bounded.
2.  $G$  has the form  $G(z, w) = Q(z, w) + h(w)$  where
  - (a)  $Q : \mathbb{X} \times \mathbb{Y} \rightarrow \mathbb{R}$  is convex and continuously differentiable over  $\mathbb{X} \times \mathbb{Y}$ ;
  - (b)  $h : \mathbb{Y} \rightarrow \mathbb{R}_{+\infty}$  is proper, lower semicontinuous, and convex.

The following properties follow immediately from Assumption 19.

**Lemma 20** Let  $G : \mathbb{X} \times \mathbb{Y} \rightarrow \mathbb{R}_{+\infty}$  satisfy SCA given in Assumption 19.

1.  $G$ ,  $Q$ , and  $h$  are regular functions (due to the assumed convexity). Thus,  $\widehat{\partial}G = \partial G$  exist; and likewise with  $\widehat{\partial}Q = \partial Q = \nabla Q$ ,  $\widehat{\partial}h = \partial h$ .
2. Calculus rules (e.g., [13, Exercise 8.8(c)]) imply that for any  $(z, w)$

$$\partial G(z, w) = \{\nabla_z Q(z, w)\} \times \{\nabla_w Q(z, w) + \partial_w h(w)\} \neq \emptyset, \quad (22a)$$

$$\widehat{\partial}G(z, w) = \{\nabla_z Q(z, w)\} \times \{\nabla_w Q(z, w) + \widehat{\partial}_w h(w)\} \neq \emptyset. \quad (22b)$$

**Lemma 21** Assume that  $G$  satisfies Assumption 19. If  $(z^*, w^*) \in \mathbb{X} \times \mathbb{Y}$  is a partial minimum of  $G$  as in Definition 5 (so that  $\nabla_z G(z^*, w^*) = 0$  and  $0 \in \widehat{\partial}_w G(z^*, w^*)$ ), then we have the Fréchet stationarity  $(0, 0) \in \widehat{\partial}G(z^*, w^*)$ .

**Proof** Follows as an application of [22, Theorem 4.1]. □

## 5.1 On the stationarity of Gauss–Seidel limit points

Two views of framing the Gauss–Seidel step of Sect. 2.1 now apparent are: 1) via continuous block  $z$  and block  $w$  partial minimisation updates of the continuous “regularisation function”  $\varphi^{\lambda, \rho, \pi}$ , and 2) continuous consensus block  $z$  minimisation updates and mixed-integer block  $(x, y, w)$  minimisation updates applied directly to augmented Lagrangian reformulations of SMIP (1). The former approach still requires an analysis of the  $(x, y)$  update, but in a hidden form. On the other hand, the latter relies on the fact that the iterates will eventually fall into a region where the integer variable will become fixed in value, and thus subproblem optimisations are local and associated with *continuous (and convex)* parts of the problem.

Motivated by the use of Lagrangian- and penalty-based solution approaches, we furthermore assume that  $G = G^k$  varies across iterations  $k \geq 0$  subject to the following assumptions.

**Assumption 22 Structural Assumptions (SA):** Given  $\{G^k : \mathbb{X} \times \mathbb{Y} \rightarrow \mathbb{R}_{+\infty}\}_{k=0}^{\infty}$ , let Assumption 19 hold for each  $G^k$ ,  $k \geq 0$ . We assume that  $\{G^k\}_{k=0}^{\infty}$  epi-converges to  $\overline{G} : \mathbb{X} \times \mathbb{Y} \rightarrow \mathbb{R}_{+\infty}$ ,  $\{Q^k\}_{k=0}^{\infty}$  epi-converges to  $\overline{Q} : \mathbb{X} \times \mathbb{Y} \rightarrow \mathbb{R}$  and that  $\{h^k\}_{k=0}^{\infty}$  epi-converges to  $\overline{h} : \mathbb{X} \times \mathbb{Y} \rightarrow \mathbb{R}_{+\infty}$ .

In Sect. 5.2, we identify the sequences  $\{G^k, (z^k, w^k)\}_{k=0}^\infty$  with a subsequence of GS (mid-)iterations associated with the application of Algorithm 1. For now, we deliberately detach the analysis of  $\{G^k, (z^k, w^k)\}_{k=0}^\infty$  from its intended algorithmic identification. The convexity of  $Q^k$  and  $h^k$  in Assumptions 19 and 22 allows for  $\{\partial G^k\}_{k=0}^\infty$  to converge in graph [13, Theorem 12.35]. This assumption will not prove restrictive in the integration of the present analysis with the convergence properties of Algorithm 1 even though the underlying problem has mixed-integer constraints.

**Lemma 23** *Let  $\{G^k : \mathbb{X} \times \mathbb{Y} \rightarrow \mathbb{R}_{+\infty}\}_{k=0}^\infty$  epi-converge to  $\overline{G} : \mathbb{X} \times \mathbb{Y} \rightarrow \mathbb{R}_{+\infty}$  satisfy Assumption 22. For  $\{(z^k, w^k)\}_{k=0}^\infty \rightarrow (\bar{z}, \bar{w})$  we have  $0 \notin \widehat{\partial G}(\bar{z}, \bar{w})$  if and only if  $\liminf_{k \rightarrow \infty} \inf_{(\zeta, \omega) \in G^k(z^k, w^k)} \|(\zeta, \omega)\| = \gamma > 0$ .*

**Proof** Given that  $0 \notin \partial \overline{G}(\bar{z}, \bar{w})$ , it follows from [13, Theorem 12.35(b)] that the sequence  $\{\widehat{\partial} G^k(z^k, w^k)\}_{k=0}^\infty$  must be strictly bounded away from zero in that  $\liminf_{k \rightarrow \infty} \inf_{(\zeta, \omega) \in \widehat{\partial} G^k(z^k, w^k)} \|(\zeta, \omega)\| > 0$ . □

**Assumption 24 Stationarity of  $w$  ( $w$ -stat)** on the (sub)sequence indexed by  $k$ : For each  $k \geq 1$ ,  $0 \in \widehat{\partial}_w G^k(z^k, w^k)$ .

**Lemma 25** *Assume  $\{G^k(z^k, w^k)\}_{k=0}^\infty \rightarrow (\overline{G}, \bar{z}, \bar{w})$  satisfies the SA Assumption 22. If  $0 \notin \widehat{\partial G}(\bar{z}, \bar{w})$  and  $w$ -stat (Assumption 24 holds), then  $\|\nabla_z \overline{Q}(\bar{z}, \bar{w})\| \neq 0$  and  $\liminf_{k \rightarrow \infty} \|\nabla_z Q^k(z^k, w^k)\| = \gamma > 0$ .*

**Proof** Under Assumption 24 ( $w$ -stat), we must have for the  $w$  subgradient components  $0 \in \{\nabla_w Q^k(\bar{z}, \bar{w}) + \widehat{\partial}_w h^k(\bar{w})\} \neq \emptyset$  for  $k \geq 0$ , and so the hypothesis  $0 \notin \widehat{\partial G}(\bar{z}, \bar{w})$  and the calculus rules of Lemma 20 and Lemma 23 imply the intended result. □

For the following results, we introduce an Armijo descent step rule for the  $z$  step to aid in the convergence analysis.

**Algorithm 2** Computing an Armijo rule step length  $\alpha > 0$ .

**Preconditions:**  $\beta, \sigma \in (0, 1)$ ;  $G$  satisfies Assumption 19;  $d$  satisfies  $\nabla_z G(z, w) d < 0$ .

```

1: function ARMIJOSTEP( $G, z, w, d, \beta, \sigma$ )
2:    $\alpha \leftarrow 1$ 
3:   while  $G(z + \alpha d, w) - G(z, w) > \alpha \sigma \nabla_z G(z, w) d$  do
4:      $\alpha \leftarrow \beta \alpha$ 
5:   end while
6:   return  $\alpha$ 
7: end function
    
```

**Assumption 26  $z$ -Descent Assumption ( $z$ -DA):** Given  $\beta, \sigma \in (0, 1)$  and given subsequences  $\{d^k\}_{k=0}^\infty$ ,  $\{G^k\}_{k=0}^\infty$ , and  $\{(z^k, w^k)\}_{k=0}^\infty$  such that  $\nabla_z Q^k(z^k, w^k) d^k < 0$  for  $k \geq 1$ ,  $z$ -DA is satisfied if

$$\lim_{k \rightarrow \infty} G^k(z^k + \alpha^k d^k, w^k) - G^k(z^k, w^k) = 0$$

where  $\alpha^k$  is computed with Algorithm 2 given  $z = z^k$  and  $w = w^k$ ,  $k \geq 1$ .

The  $z$ -DA Assumption 26 itself makes no assumption on *how* the sequence  $\{(z^k, w^k), d^k\}_{k=0}^\infty$  is constructed. Subsequently stated identifications of the sequence  $\{(z^k, w^k)\}_{k=0}^\infty$  with subsequences generated by Algorithm 1 will guarantee the satisfaction of  $z$ -DA Assumption 26 under mild assumptions on the implementation of Algorithm 1. The Armijo Step of Algorithm 2 is not actually used in our implementation of Algorithm 1. Rather, it is merely a theoretical tool in what follows.

The proof of the following lemma is based on ideas from [24, Proposition 1.2.1], [25, Proposition 3.2], and [26, Technical Lemmas, Appendix A].

**Lemma 27** *Assume that 1)  $\{G^k : \mathbb{X} \times \mathbb{Y} \rightarrow \mathbb{R}_{+\infty}\}_{k=0}^\infty$  epi-converging to  $\bar{G} : \mathbb{X} \times \mathbb{Y} \rightarrow \mathbb{R}_{+\infty}$  satisfies SA (Assumption 22); 2)  $\{(w^k, z^k)\}_{k=0}^\infty$  converges to  $(\bar{w}, \bar{z})$  and satisfies  $w$ -stat (Assumption 24); and 3)  $\nabla_z Q^k(z^k, w^k) \neq 0$  for each  $k \geq 1$ .*

*If, for some  $\beta, \sigma \in (0, 1)$ , the  $z$ -DA (Assumption 26) holds for  $d^k = -\nabla_z Q^k(z^k, w^k)$ , then  $0 \in \partial \bar{G}(\bar{z}, \bar{w})$ . Moreover,  $\bar{G}$  is regular at  $(\bar{z}, \bar{w})$  and so we have also  $0 \in \partial \bar{G}(\bar{z}, \bar{w})$ .*

**Proof** From (22), the satisfaction of  $w$ -stat Assumption 24, and Lemma 21, we only need to show that  $\lim_{k \rightarrow \infty} \|\nabla_z Q^k(z^k, w^k)\| = 0$ . Due to SA Assumption 22 and  $z$ -DA Assumption 26, we have

$$\begin{aligned} 0 &= \lim_{k \rightarrow \infty} G^k(z^k + \alpha^k d^k, w^k) - G^k(z^k, w^k) = \lim_{k \rightarrow \infty} Q^k(z^k + \alpha^k d^k, w^k) - Q^k(z^k, w^k) \\ &\leq \lim_{k \rightarrow \infty} \alpha^k \sigma \nabla_z Q^k(z^k, w^k) d^k \leq 0. \end{aligned}$$

Thus,  $\lim_{k \rightarrow \infty} \alpha^k \sigma \nabla_z Q^k(z^k, w^k) d^k = 0$ .

We consider two cases: 1)  $\limsup_{k \rightarrow \infty} \alpha^k > 0$ , and 2)  $\limsup_{k \rightarrow \infty} \alpha^k = 0$ . Due to the assumed continuity of  $\nabla_z Q^k$ , and given that  $d^k = -\nabla_z Q^k(z^k, w^k)$ , the first case implies that

$$\lim_{k \rightarrow \infty} \nabla_z Q^k(z^k, w^k) d^k = \lim_{k \rightarrow \infty} -\|\nabla_z Q^k(z^k, w^k)\|^2 = 0.$$

and so  $\lim_{k \rightarrow \infty} \nabla_z Q^k(z^k, w^k) = 0$ . Otherwise, assuming that  $\limsup_{k \rightarrow \infty} \alpha^k = \lim_{k \rightarrow \infty} \alpha^k = 0$ , we have for some large enough  $k \geq \bar{k} \geq 0$  that  $\alpha^k \leq \beta < 1$  and so we have  $\bar{\alpha}^k = \alpha^k / \beta \leq 1$ —that is, the state of  $\alpha^k$  at the penultimate iteration of Algorithm 2—for which it holds that

$$G^k(z^k + \bar{\alpha}^k d^k, w^k) - G^k(z^k, w^k) > \bar{\alpha}^k \sigma \nabla_z Q^k(z^k, w^k) d^k,$$

which implies

$$\frac{Q^k(z^k + \bar{\alpha}^k d^k, w^k) - Q^k(z^k, w^k)}{\bar{\alpha}^k} > \sigma \nabla_z Q^k(z^k, w^k) d^k.$$

Applying the Mean Value Theorem at each  $k$ , we have

$$\nabla_z Q^k(z^k + \bar{\alpha}^k d^k, w^k) d^k > \sigma \nabla_z Q^k(z^k, w^k) d^k.$$



for some  $\tilde{\alpha}^k \in [0, \bar{\alpha}^k]$ . Using the continuity of  $\nabla_z Q^k$  (and the Cauchy-Schwartz inequality), we have for arbitrarily small  $\epsilon > 0$  that there exists  $\delta < 0$  where for large enough  $k$ , we have  $\tilde{\alpha}^k \leq \bar{\alpha}^k < \delta$  so that

$$\epsilon \|d^k\| + \nabla_z Q^k(z^k, w^k)d^k > \nabla_z Q^k(z^k + \tilde{\alpha}^k d^k, w^k)d^k > \sigma \nabla_z Q^k(z^k, w^k)d^k$$

holds for sufficiently large  $k$ .

Recalling that  $d^k = -\nabla_z Q^k(z^k, w^k) \neq 0$ , we then have

$$\epsilon \|d^k\| > (\sigma - 1)\nabla_z Q^k(z^k, w^k)d^k = (1 - \sigma)\|\nabla_z Q^k(z^k, w^k)\|_2^2$$

and so  $\epsilon > (1 - \sigma)\|\nabla_z Q^k(z^k, w^k)\|_2$  holds for sufficiently large  $k$ . In the limit, we have  $0 \geq (1 - \sigma)\|\nabla_z \bar{Q}(\bar{z}, \bar{w})\|_2$ , which is a contradiction since  $(1 - \sigma) > 0$  and  $\|\nabla_z \bar{Q}(\bar{z}, \bar{w})\|_2 > 0$  as established from Lemma 25 and the SA Assumption 22. Thus,  $0 \in \partial \bar{G}(\bar{z}, \bar{w})$ , and since  $\bar{G}$  is regular, then  $0 \in \widehat{\partial G}(\bar{z}, \bar{w})$  holds also.  $\square$

In order to apply Lemma 27 to a convergence analysis of Algorithm 1, we need to establish the satisfaction of the SA Assumption 22,  $w$ -stat Assumption 24, and especially the  $z$ -DA Assumption 26 requiring  $\lim_{k \rightarrow \infty} G^k(z^k + \alpha^k d^k, w^k) - G^k(z^k, w^k) = 0$  given an appropriate identification with Algorithm 1 subsequence iterations  $\{n_k\}_{k=0}^\infty$ .

### 5.2 Interleaving analysis and algorithm

We analyse subsequences  $\{(x^{n_k+1}, y^{n_k+1}, w^{n_k+1}, z^{n_k})\}_{k=0}^\infty$  from the iterations generated by the application of Algorithm 1 applied to problem (1) that converge to  $(\bar{x}, \bar{y}, \bar{w}, \bar{z})$ . (Such limit points with respect to the entire sequence in  $n$  exist due to inf-compactness (1) that will be demonstrated in this subsection.) This analysis depends on establishing that the Sect. 5.1 assumptions hold under the appropriate identifications with Algorithm 1 (i.e., the SA Assumption 22, the  $w$ -stat Assumption 24, and the  $z$ -DA Assumption 26).

Given the assumed subsequence convergence, we may take the subsequence  $\{(x^{n_k+1}, y^{n_k+1}, w^{n_k+1}, z^{n_k})\}_{k=0}^\infty$  so that integer component values for  $x_{\mathcal{I}} = \bar{x}_{\mathcal{I}}$  and  $y_{\mathcal{I}} = \bar{y}_{\mathcal{I}}$  are fixed. With respect to Algorithm 1, we apply the identifications given GS iterations indexed by  $n \geq 1$  and subsequence iterations indexed with  $n_k, k \geq 1$ :

#### Assumption 28 Algorithm 1 Identifications:

1. Variables:  $z^k \leftarrow z^{n_k}$  and  $w^k \leftarrow (x_s^{n_k+1}, y_s^{n_k+1}, w_s^{n_k+1})_{s \in S}$
2. Expressions:  $Q^k(z^k, w^k) \leftarrow \sum_{s \in S} \pi_s^{n_k} \psi(z^{n_k} - x_s^{n_k+1}, w_s^{n_k+1} - y_s^{n_k+1})$  and

$$h^k(w^k) \leftarrow \sum_{s \in S} \left[ \frac{1}{\rho^{n_k}} \left( f_s(x_s^{n_k+1}, y_s^{n_k+1}) - (\lambda_s^{n_k})^\top x_s^{n_k+1} \right) + \delta_{K_s(\bar{x}_{\mathcal{I}}, \bar{y}_{\mathcal{I}})}(x_s^{n_k+1}, y_s^{n_k+1}) \right]$$

Recalling  $f^n(x, y, w, z) = \sum_{s \in S} f_s(x_s, y_s) + (\lambda_s^n)^\top x_s + \rho^n \pi_s^n \psi(z - x_s, w_s - y_s)$  defined in (17), define for each  $n \geq 0$

$$g^n(x, y, w, z) := \frac{1}{\rho^n} f^n(x, y, w, z). \tag{23}$$

we have also

$$\begin{aligned}
 G^k(z^k, w^k) &\leftarrow g^{n_k}(x^{n_k+1}, y^{n_k+1}, w^{n_k+1}, z^{n_k}) + \sum_{s \in S} \delta_{K_s(\bar{x}_T, \bar{y}_T)}(x_s^{n_k+1}, y_s^{n_k+1}) \\
 &= g^{n_k}(x^{n_k+1}, y^{n_k+1}, w^{n_k+1}, z^{n_k})
 \end{aligned}$$

where the last equality is by construction (fixed integral values) of the subsequence.

Furthermore, to guarantee the set of assumptions: SA,  $w$ -Stat, and  $z$ -DA, we assume the following of the GS (sub)sequences:

**Assumption 29 Algorithm Assumptions:** In the application of Algorithm 1 to problem (1), the following hold:

1. **SMIP assumptions:** Assumption 1 holds for problem (1).
2. **Penalty function assumptions:**  $\psi$  satisfies the Assumption 2. Furthermore, we subsequently note special implications that hold in the cases where  $\psi$  takes the weighted squared 2-norm form with weights  $\bar{\mu}_i > 0$  such that

$$\psi(z - x_s, 0) = \frac{1}{2} \sum_{s \in S} \sum_{i=1}^n \bar{\mu}_i (z_i - x_{s,i})^2. \tag{24}$$

3. **Global optimality:** Each  $(x^{n+1}, y^{n+1}, w^{n+1})$  is globally optimal given fixed  $z^n$  in that

$$(x^{n+1}, y^{n+1}, w^{n+1}) \in \arg \min_{x,y,w} f^n(x, y, w, z^n)$$

(hence limit points  $(\bar{x}, \bar{y}, \bar{w})$  are globally optimal given fixed limit point  $\bar{z}$  by known results [27, Propositions 1.3.5 and 1.3.6]). Also,  $z^{n+1} \in \arg \min_z \sum_{s \in S} \pi_s^n \psi(z - x_s^{n+1}, 0)$ ,  $n \geq 1$ , are globally optimal. Furthermore, under the additional assumption that  $\psi$  is of the weighted squared 2-norm form (24), we have (independent of weights  $\bar{\mu}_i > 0, i = 1, \dots, n$ )

$$z^{n+1} \leftarrow \sum_{s \in S} \pi_s^n x_s^{n+1} \in \arg \min_z \sum_{s \in S} \pi_s^n \psi(z - x_s^{n+1}, 0).$$

4. **Generation of Lagrange multipliers:**  $\lambda_s^0 = 0$  for  $s \in S$ . If  $\psi$  is *not* of the weighted squared 2-norm form (24), then we may assume  $\lambda_s^n \equiv 0, n \geq 1$ , identically. Otherwise, if  $\psi$  is of the squared 2-norm form (24), For each subsequent iteration  $n$ , either  $\lambda_s^n$  is left unchanged via  $\lambda_s^{n+1} \leftarrow \lambda_s^n$  for all  $s \in S$ , or

$$\lambda_{s,i}^{n+1} \leftarrow \lambda_{s,i}^n - \rho^n \pi_s^n \bar{\mu}_i (z_i^{n+1} - x_{s,i}^{n+1}) \quad \text{for all } s \in S.$$

Under these assumptions on  $\lambda$ , it follows that for  $n \geq 0$ , we have vanishing sums  $\sum_{s \in S} \lambda_s^n = 0$ . (i.e., dual feasibility maintained and the absence of the  $\sum_{s \in S} \lambda_s^n z$

terms in the Lagrangian is thus justified.) Non-trivial  $\lambda$  updates between iterations are suppressed as necessary to ensure in the limit that  $\sum_{n=1}^{\infty} \|\lambda_s^n - \lambda_s^{n+1}\| < \infty$  hold. (In practice this will usually entail only a finite number of nontrivial updates.)

**5. Update of penalty parameters:** We assume the following.

- (a) Penalty coefficients are nondecreasing  $\rho^{n+1} \geq \rho^n > 0, n \geq 0$ .
- (b)  $0 < \pi_s^n \rho^n \leq \pi_s^{n+1} \rho^{n+1}, s \in S, n \geq 0$ . ( $\pi_s^n \leq \pi_s^{n+1}$  does *not* hold in general.)
- (c) For each  $n \geq 0$  and  $s \in S$ , we have  $\pi_s^n > 0$  and  $\sum_{s \in S} \pi_s^n = 1$ , and  $\{\pi_s^n\}_{n=0}^{\infty}$  converges to  $\pi_s > 0$  for all  $s \in S$  such that  $\sum_{n=1}^{\infty} |\pi_s^{n+1} - \pi_s^n| < \infty$ . Initially,  $\pi_s^0 \leftarrow p_s$ .

The algorithm does not necessarily adjust  $\pi_s^n$  and  $\rho^n$  parameters separately. Instead, it may apply penalty updates in a scenario-specific manner to  $\rho_s^n := \pi_s^n \rho^n$  with  $\rho^n = \sum_{s \in S} \rho_s^n$ . Under the condition that  $s \in S_n := \{s \in S \mid z_{\mathcal{I}} \neq x_{s, \mathcal{I}}\}$ , we place  $\rho_s^{n+1} = \gamma_n \rho_s^n$  with  $\gamma_n > 1$ .

**Lemma 30** *Under the Algorithmic Identifications 28 with  $G^k = h^k + Q^k$  and Assumption 29, we have*

1.  $\{G^k\}_{k=0}^{\infty}$  epi-converges to  $\bar{G} = \bar{h} + \bar{Q}$  and satisfies the SA Assumption 22.
2.  $0 \in \partial_w G^k(z^k, w^k)$  and  $0 \in \partial_w \bar{G}(\bar{z}, \bar{w})$ .

**Proof** By regularity of  $\bar{G}$  due to its convexity, the (limiting) Mordukhovich and Fréchet subdifferentials coincide.

We argue that we have epi-convergence of  $\{G_k\}_{k=0}^{\infty}$  to  $\bar{G}$ , whenever  $\{(\pi_s^{n_k})_{s \in S}\}_{k=0}^{\infty}$  converges to  $\{\pi_s\}_{s \in S}$ . As

$$h^k(w) := \frac{1}{\rho^{n_k}} \left( \sum_{s \in S} p_s \{c^\top x_s + d_s^\top y_s\} - (\lambda_s^{n_k})^\top x_s \right) + \delta_{K(\bar{x}_{\mathcal{I}}, \bar{y}_{\mathcal{I}})}(x, y).$$

we have  $h^k$  convex and converging both monotonically point-wise and uniformly to  $\delta_{K(\bar{x}_{\mathcal{I}}, \bar{y}_{\mathcal{I}})}$ . This is because we have uniform convergence to zero of  $\frac{1}{\rho^{n_k}} (\sum_{s \in S} p_s \{c^\top x_s + d_s^\top y_s\} - (\lambda_s^{n_k})^\top x_s)$  on the compact and convex polyhedral set  $K(\bar{x}_{\mathcal{I}}, \bar{y}_{\mathcal{I}})$ . Thus we have epi-convergence of  $\{h^k\}_{k=0}^{\infty}$  to  $\delta_{K(\bar{x}_{\mathcal{I}}, \bar{y}_{\mathcal{I}})}$ . Whenever  $\{(\pi_s^{n_k})_{s \in S}\}_{k=0}^{\infty}$  converges to  $\pi_s$ , for  $s \in S$  we then have a family of convex functions  $\{Q^k := \sum_{s \in S} \pi_s^{n_k} \psi\}_{k=0}^{\infty}$ , uniformly converging on compact sets and hence also epi-convergent to  $\bar{Q} = \sum_{s \in S} \pi_s \psi_s(\cdot, \cdot)$ . Applying [27, Theorem 7.1.5] or [13, Theorem 7.46], we know that the sum of a uniformly convergent sequence and an epi-convergent sequence must epi-converge. Thus we may deduce that we have  $\{G^k := h^k + Q^k\}_{k=0}^{\infty}$  epi-converges to  $\bar{G} = \bar{h} + \bar{Q}$ . Now we may apply [13, Theorem 12.35] to deduce that the convex subdifferentials  $\{\partial_w G^k\}_{k=0}^{\infty}$  converge in graph to  $\partial_w \bar{G}$ . As  $w^k$  is a global minimiser for  $w \mapsto G^k(z^k, w)$ , then we have (by definitions) that  $0 \in \partial_w G^k(z^k, w^k)$  and hence also over the subsequence indexed by  $k$ . Thus by graphical convergence  $0 \in \partial_w \bar{G}(\bar{z}, \bar{w})$ . □

**Definition 6** Let *intervening GS iterations* between  $n_k$  and  $n_{k+1}$  be denoted  $n_k + 1, n_k + 2, \dots, n_{k+1} - 2, n_{k+1} - 1$ , etc.

**Lemma 31** *If  $(\mathbf{x}^n, \mathbf{y}^n, \mathbf{w}^n, z^n)$ ,  $n \geq 1$ , are computed with the GS iterations of Algorithm 1, then for each fixed  $n \geq 1$  and positive integer  $j$ , we have*

$$\begin{aligned}
 &g^{n+j}(\mathbf{x}^{n+j+1}, \mathbf{y}^{n+j+1}, \mathbf{w}^{n+j+1}, z^{n+j}) - g^n(\mathbf{x}^{n+1}, \mathbf{y}^{n+1}, \mathbf{w}^{n+1}, z^n) \\
 &\quad + \sum_{i=0}^{j-1} \left[ g^{n+i}(\mathbf{x}^{n+i+1}, \mathbf{y}^{n+i+1}, \mathbf{w}^{n+i+1}, z^{n+i+1}) \right. \\
 &\quad \left. - g^{n+i+1}(\mathbf{x}^{n+i+1}, \mathbf{y}^{n+i+1}, \mathbf{w}^{n+i+1}, z^{n+i+1}) \right] \\
 &\leq g^n(\mathbf{x}^{n+1}, \mathbf{y}^{n+1}, \mathbf{w}^{n+1}, z^n + \alpha^n d^n) - g^n(\mathbf{x}^{n+1}, \mathbf{y}^{n+1}, \mathbf{w}^{n+1}, z^n)
 \end{aligned}$$

where  $z^{n+i+1} \in \arg \min_z \{g^{n+i}(\mathbf{x}^{n+i+1}, \mathbf{y}^{n+i+1}, \mathbf{w}^{n+i+1}, z)\}$  for  $i = 0, \dots, j - 1$  as is consistent with an iteration of Algorithm 1.

**Proof** See Appendix A. □

**Corollary 32** *Given  $n_k, k \geq 0$ , a subsequence index, and  $j_k$  the positive integer such that  $n_k + j_k = n_{k+1}$ , if  $(\mathbf{x}^{n+1}, \mathbf{y}^{n+1}, \mathbf{w}^{n+1}, z^n)$ ,  $n \geq 1$ , are computed with GS iterations, then*

$$\begin{aligned}
 &\sum_{i=0}^{j_k-1} \left[ g^{n_k+i}(\mathbf{x}^{n_k+i+1}, \mathbf{y}^{n_k+i+1}, \mathbf{w}^{n_k+i+1}, z^{n_k+i+1}) \right. \\
 &\quad \left. - g^{n_k+i+1}(\mathbf{x}^{n_k+i+1}, \mathbf{y}^{n_k+i+1}, \mathbf{w}^{n_k+i+1}, z^{n_k+i+1}) \right] \\
 &\quad + G^{k+1}(z^{k+1}, w^{k+1}) - G^k(z^k, w^k) \leq G^k(z^k + \alpha^k d^k, w^k) - G^k(z^k, w^k)
 \end{aligned}$$

**Proof** By the construction of the subsequence,  $\delta_{K_s(\bar{x}_s, \mathcal{I}, \bar{y}_s, \mathcal{I})}(x_s^{n_k+1}, y_s^{n_k+1}) = 0$  and so any potential discrepancy between  $g^{n_k}(\mathbf{x}^{n_k+1}, \mathbf{y}^{n_k+1}, \mathbf{w}^{n_k+1}, z^{n_k})$  and  $G^k(z^k, w^k)$  is avoided. □

**Lemma 33** *Under Assumption 28 and Assumption 29, we have*

$$\begin{aligned}
 &\lim_{k \rightarrow \infty} \sum_{i=0}^{j_k-1} \left[ g^{n_k+i}(\mathbf{x}^{n_k+i+1}, \mathbf{y}^{n_k+i+1}, \mathbf{w}^{n_k+i+1}, z^{n_k+i+1}) \right. \\
 &\quad \left. - g^{n_k+i+1}(\mathbf{x}^{n_k+i+1}, \mathbf{y}^{n_k+i+1}, \mathbf{w}^{n_k+i+1}, z^{n_k+i+1}) \right] = 0.
 \end{aligned}$$

**Proof** See Appendix A. □

**Corollary 34** *Assume the Algorithm Identifications 28. If the Algorithm Assumptions 29 hold, then z-DA Assumption 26 holds under any allowable realisation of its assumptions on  $\beta, \sigma, d^k$ , etc. (Thus, the intended z-DA condition will hold for any convergent subsequence  $\{(z^k, w^k)\}_{k=0}^\infty = \{(\mathbf{x}^{n_k+1}, \mathbf{y}^{n_k+1}, \mathbf{w}^{n_k+1}, z^{n_k})\}_{k=0}^\infty$  with  $d^k = -\nabla_z Q^k(z^k, w^k)$ , and  $\alpha^k$  computed with the Armijo rule for any  $\beta, \sigma \in (0, 1)$ .)*

**Proof** Given that  $G^k(z^k + \alpha^k d^k, w^k) - G^k(z^k, w^k) \leq 0$  already holds per the Armijo step, the satisfaction of  $z$ -DA Assumption 26 follows from Lemma 33 and Corollary 32 once it is noted that  $\lim_{k \rightarrow \infty} G^{k+1}(z^{k+1}, w^{k+1}) - G^k(z^k, w^k) = 0$  follows from the SA epi-convergence of Assumption 22 [13, Theorem 12.35].  $\square$

**Definition 7** Under the epi-convergence of SA Assumption 22, we define the *limiting regularisation*

$$\phi^\infty(z, \mathbf{w}) := \lim_{\rho \rightarrow \infty} \frac{1}{\rho} \varphi^{\lambda, \rho, \pi}(z, \mathbf{w}) = \min_{x, y} \left\{ \sum_{s \in S} \pi_s \psi(z - x_s, w_s - y_s) \mid (x_s, y_s) \in K_s \right\}$$

We now state one of our main results. Before doing so, we denote the following

**Definition 8** To accommodate both possibilities  $\lim_{n \rightarrow \infty} \rho^n = \bar{\rho} < \infty$  or  $\lim_{n \rightarrow \infty} \rho^n = \infty$  disjunctively, we define

$$\begin{aligned} g^*(\mathbf{x}, \mathbf{y}, \mathbf{w}, z) &:= \lim_{n \rightarrow \infty} g^n(\mathbf{x}, \mathbf{y}, \mathbf{w}, z) \\ &= \lim_{n \rightarrow \infty} \sum_{s \in S} \left[ \frac{1}{\rho^n} \left( f_s(x_s, y_s) - (\lambda_s^n)^\top x_s \right) + \pi_s^n \psi_s(z - x_s, w_s - y_s) \right] \end{aligned} \tag{25}$$

(recalling the definition (23)). From (25), we define the limiting regularisation

$$\begin{aligned} \phi^*(z, \mathbf{w}) &:= \lim_{n \rightarrow \infty} \frac{1}{\rho^n} \varphi^n(z, \mathbf{w}) = \min_{x, y} g^*(\mathbf{x}, \mathbf{y}, \mathbf{w}, z) \\ \phi^*(z, \mathbf{w} \mid \bar{\mathbf{x}}_{\mathcal{I}}, \bar{\mathbf{y}}_{\mathcal{I}}) &:= \min_{x, y} g^*(\mathbf{x}, \mathbf{y}, \mathbf{w}, z) + \delta_{K(\bar{\mathbf{x}}_{\mathcal{I}}, \bar{\mathbf{y}}_{\mathcal{I}})}(\mathbf{x}, \mathbf{y}) \end{aligned}$$

The corresponding set of solutions  $(\mathbf{x}, \mathbf{y})$  realising these values given  $(z, \mathbf{w})$  is denoted

$$\begin{aligned} \Phi^*(z, \mathbf{w}) &:= \arg \min_{x, y} g^*(\mathbf{x}, \mathbf{y}, \mathbf{w}, z) \\ \Phi^*(z, \mathbf{w} \mid \bar{\mathbf{x}}_{\mathcal{I}}, \bar{\mathbf{y}}_{\mathcal{I}}) &:= \arg \min_{x, y} g^*(\mathbf{x}, \mathbf{y}, \mathbf{w}, z) + \delta_{K(\bar{\mathbf{x}}_{\mathcal{I}}, \bar{\mathbf{y}}_{\mathcal{I}})}(\mathbf{x}, \mathbf{y}). \end{aligned}$$

**Proposition 35** Let  $(\bar{\mathbf{x}}, \bar{\mathbf{y}}, \bar{\mathbf{w}}, \bar{z})$  satisfy  $(\bar{\mathbf{x}}, \bar{\mathbf{y}}) \in K$ . The following implications hold.

1. If the Fréchet stationarity  $0 \in \widehat{\partial} g^*(\bar{\mathbf{x}}, \bar{\mathbf{y}}, \bar{\mathbf{w}}, \bar{z})$  holds, then  $(0, 0) \in \widehat{\partial} \phi^*(\bar{z}, \bar{\mathbf{w}} \mid \bar{\mathbf{x}}_{\mathcal{I}}, \bar{\mathbf{y}}_{\mathcal{I}})$  so that  $(\bar{z}, \bar{\mathbf{w}})$  is a minimum of  $\phi^*(\bar{z}, \bar{\mathbf{w}} \mid \bar{\mathbf{x}}_{\mathcal{I}}, \bar{\mathbf{y}}_{\mathcal{I}})$ .
2.  $(\bar{z}, \bar{\mathbf{w}})$  is a local minimum of  $\phi^*$  with  $(0, 0) \in \widehat{\partial} \phi^*(\bar{z}, \bar{\mathbf{w}}) = \nabla \phi^*(\bar{z}, \bar{\mathbf{w}})$  if and only if  $(\bar{z}, \bar{\mathbf{w}})$  is a local minimum of  $\phi^*(\cdot, \cdot \mid \mathbf{x}_{\mathcal{I}}, \mathbf{y}_{\mathcal{I}})$  with  $(0, 0) \in \widehat{\partial} \phi^*(\bar{z}, \bar{\mathbf{w}} \mid \mathbf{x}_{\mathcal{I}}, \mathbf{y}_{\mathcal{I}}) = \partial \phi^*(\bar{z}, \bar{\mathbf{w}} \mid \mathbf{x}_{\mathcal{I}}, \mathbf{y}_{\mathcal{I}})$  for all  $(\mathbf{x}_{\mathcal{I}}, \mathbf{y}_{\mathcal{I}}) \in \text{proj}_{\mathcal{I}}(\Phi^*(\bar{z}, \bar{\mathbf{w}}))$ .
3. If  $\bar{z} = \bar{x}_s$  for all  $s \in S$ , then  $(0, 0) \in \widehat{\partial} \phi^*(\bar{z}, \bar{\mathbf{w}})$  and so  $(0, 0) = \nabla \phi^*(\bar{z}, \bar{\mathbf{w}})$ . Furthermore,  $(\bar{z}, \bar{\mathbf{w}})$  is also a (persistent) local minimum for  $\phi^*$ .
4. In the specific case where  $\lim_{n \rightarrow \infty} \rho^n = \infty$  (so that  $\phi^* = \phi^\infty$ ), the reverse of the previous implication also holds, where  $(\bar{z}, \bar{\mathbf{w}})$  being a local minimum for  $\phi^*$  with  $(0, 0) \in \widehat{\partial} \phi^*(\bar{z}, \bar{\mathbf{w}})$  implies that  $\bar{z} = \bar{x}_s, s \in S$ .

**Proof** Part 1: Given  $0 \in \widehat{\partial}g^*(\bar{x}, \bar{y}, \bar{w}, \bar{z})$  and the structure of  $g^*$  as the sum of a linear function and an indicator function for a polyhedral set with integer cross-sections, we have that  $g^*(x, y, w, z) \geq g^*(\bar{x}, \bar{y}, \bar{w}, \bar{z})$  for all  $(x, y) \in K(\bar{x}_{\mathcal{I}}, \bar{y}_{\mathcal{I}})$  and  $(z, w) \in \mathbb{X} \times \mathbb{Y}$  and more particularly,  $g^*(x, y, \bar{w}, \bar{z}) \geq g^*(\bar{x}, \bar{y}, \bar{w}, \bar{z})$  for all  $(x, y) \in K(\bar{x}_{\mathcal{I}}, \bar{y}_{\mathcal{I}})$ . Thus,

$$g^*(\bar{x}, \bar{y}, \bar{w}, \bar{z}) = \phi^*(\bar{z}, \bar{w} \mid \bar{x}_{\mathcal{I}}, \bar{y}_{\mathcal{I}}).$$

Furthermore, since  $g^*(x, y, w, z) \geq g^*(\bar{x}, \bar{y}, \bar{w}, \bar{z})$  for all  $(x, y) \in K(\bar{x}_{\mathcal{I}}, \bar{y}_{\mathcal{I}})$  and  $(z, w) \in \mathbb{X} \times \mathbb{Y}$ , we have

$$\phi^*(z, w \mid \bar{x}_{\mathcal{I}}, \bar{y}_{\mathcal{I}}) \geq \phi^*(\bar{z}, \bar{w} \mid \bar{x}_{\mathcal{I}}, \bar{y}_{\mathcal{I}})$$

and so by the convexity of  $\phi^*(\cdot, \cdot \mid \bar{x}_{\mathcal{I}}, \bar{y}_{\mathcal{I}})$ , we have also that  $(0, 0) \in \widehat{\partial}\phi^*(\bar{z}, \bar{w} \mid \bar{x}_{\mathcal{I}}, \bar{y}_{\mathcal{I}})$ .

Part 2: We have  $(0, 0) \in \widehat{\partial}\phi^*(\bar{z}, \bar{w} \mid x_{\mathcal{I}}, y_{\mathcal{I}})$  (in both the Fréchet and classical sense) due to  $(\bar{z}, \bar{w}) \in \arg \min_{z, w} \phi^*(z, w \mid x_{\mathcal{I}}, y_{\mathcal{I}})$  and as  $\phi^*(z, w) = \min_{(x_{\mathcal{I}}, y_{\mathcal{I}}) \in \text{proj}_{\mathcal{I}}(K)} \phi^*(z, w \mid x_{\mathcal{I}}, y_{\mathcal{I}})$  where each  $(z, w) \mapsto \phi^*(z, w \mid x_{\mathcal{I}}, y_{\mathcal{I}})$  is convex, we may invoke Lemma 11 Part 2 to obtain both directions of the implication after identifying  $\varphi$  with  $\phi^*$  and  $\varphi_i, i \in I$ , with  $\phi^*(\cdot, \cdot \mid x_{\mathcal{I}}, y_{\mathcal{I}})$  for  $(x_{\mathcal{I}}, y_{\mathcal{I}}) \in \text{proj}_{\mathcal{I}}(\Phi^*(z, w))$ .

Part 3: The fact that  $\bar{z} = \bar{x}_s$  for all  $s \in S$  implies by Corollary 8 that  $\phi^*(\bar{z}, \bar{w}) = \phi^*(\bar{z}, \bar{w} \mid x_{\mathcal{I}}, y_{\mathcal{I}})$  for just one  $(x_{\mathcal{I}}, y_{\mathcal{I}}) = (\bar{x}_{\mathcal{I}}, \bar{w}_{\mathcal{I}})$  only, and so the claim follows. The persistency follows from the fact that inequality (11) implies a bound independent of  $\rho$ .

Part 4: Knowing that  $(\bar{z}, \bar{w})$  is a local minimum for  $\phi^\infty$ , we form the cleared instance of the SMIP (1) by clearing first- and second-stage coefficients  $c = d = 0$ , and for all  $n \geq 0$ , clearing  $\lambda^n = 0$  and setting  $\pi^n = \pi$ . Thus, for all  $\rho > 0$ , we have that  $\frac{1}{\rho}\phi^{\lambda, \rho, \pi} \equiv \phi^\infty$  and so  $(\tilde{z}^n, \tilde{w}^n) \equiv (\bar{z}, \bar{w}), n \geq 0$ , forms a sequence with limit  $(\bar{z}, \bar{w}) \equiv (\tilde{z}^n, \tilde{w}^n), n \geq 0$ . Each  $(\tilde{z}^n, \tilde{w}^n) \equiv (\bar{z}, \bar{w})$  is a local minimum for  $\varphi^n$  over a fixed neighbourhood  $B_\delta(\bar{z}, \bar{w})$  for some fixed  $\delta = \bar{\delta} > 0$ , and since the SSA Assumption 9 and the PWA Assumption 10 therefore hold, by Proposition 17 applied to this sequence associated with this cleared instance of SMIP (1), we have  $\bar{z} \in X$  and  $\bar{w} \in Y(\bar{z})$ , which applies with respect to the original (non-cleared) instance of SMIP (1) also. □

**Theorem 36** *Assume that problem (1) satisfies the SMIP Assumption 1, to which Algorithm 1 is applied to generate a sequence  $\{(x^n, y^n, w^n, z^n)\}_{n=0}^\infty$ . If the Algorithm Assumption 29 is satisfied, then there exists a limit point  $(\bar{x}, \bar{y}, \bar{w}, \bar{z})$  of the mid-iteration sequence  $\{(x^{n+1}, y^{n+1}, w^{n+1}, z^n)\}_{n=0}^\infty$ , and each such limit point  $(\bar{x}, \bar{y}, \bar{w}, \bar{z})$  is a Fréchet stationary point for the problem*

$$\min_{z, x, y, w} g^*(x, y, w, z) \tag{26}$$

and in either limiting case, the cross-sectional optimality  $(\bar{z}, \bar{w}) \in \arg \min_{z, w} \phi^*(z, w \mid \bar{x}_{\mathcal{I}}, \bar{y}_{\mathcal{I}})$  holds. Thus, the following implications hold:

1.  $(\bar{z}, \bar{w})$  is a local minimum of  $\phi^*$  if and only if  $(\bar{z}, \bar{w})$  is a local minimum of  $\phi^*(\cdot, \cdot \mid \mathbf{x}_{\mathcal{I}}, \mathbf{y}_{\mathcal{I}})$  for all  $(\mathbf{x}_{\mathcal{I}}, \mathbf{y}_{\mathcal{I}}) \in \text{proj}_{\mathcal{I}}(\Phi^*(\bar{z}, \bar{w}))$ .
2. If  $\bar{z} = \bar{x}_s$  for all  $s \in S$  (so that  $(\bar{\mathbf{x}}, \bar{\mathbf{y}})$  is feasible and locally optimal for SMIP (1)), then  $(\bar{z}, \bar{w})$  is a (persistent) local minimum of  $\phi^*$ .
3. In the specific case where  $\lim_{n \rightarrow \infty} \rho^n = \infty$  (so that  $\phi^* = \phi^\infty$ ), the reverse of the previous implication also holds, where  $(\bar{z}, \bar{w})$  being a local minimum of  $\phi^*$  implies that  $\bar{z} = \bar{x}_s, s \in S$ , so that  $(\bar{\mathbf{x}}, \bar{\mathbf{y}})$  is feasible and locally optimal for SMIP (1).

**Proof** Under the SMIP Assumption that  $K_s, s \in S$ , are compact and penalty  $\psi$  satisfies Assumptions 2, it follows that the level sets of  $g$  are compact, and so the sequence  $\{(\mathbf{x}^{n+1}, \mathbf{y}^{n+1}, \mathbf{w}^{n+1}, z^n)\}_{n=0}^\infty$  will have limit points  $(\bar{\mathbf{x}}, \bar{\mathbf{y}}, \bar{\mathbf{w}}, \bar{z})$  to which an associated subsequence  $\{(\mathbf{x}^{n_k+1}, \mathbf{y}^{n_k+1}, \mathbf{w}^{n_k+1}, z^{n_k})\}_{k=0}^\infty$  converges. For  $k$  large enough, we have  $(\mathbf{x}_{\mathcal{I}}^{n_k+1}, \mathbf{y}_{\mathcal{I}}^{n_k+1}, \mathbf{w}_{\mathcal{I}}^{n_k+1}) = (\bar{\mathbf{x}}_{\mathcal{I}}, \bar{\mathbf{y}}_{\mathcal{I}}, \bar{\mathbf{w}}_{\mathcal{I}})$  becoming fixed. Therefore, it is only the  $z^{n_k}$  and the real-valued components  $(\mathbf{x}_{\mathcal{R}}^{n_k+1}, \mathbf{y}_{\mathcal{R}}^{n_k+1}, \mathbf{w}_{\mathcal{R}}^{n_k+1})$  that are still changing throughout the (sub)sequence tail. After passing to a convergent subsequence with integer components fixed, the required SA Assumption 22 and the  $w$ -stat Assumption 24 applies to Assumption 28 identified with  $\{(G^k, z^k, w^k)\}_{k=0}^\infty$  by Lemma 30. Under the same assumptions, the  $z$ -DA Assumption 26 is satisfied due to Corollary 34. Thus, Lemma 27 may be applied to the Assumption 28 identification sequence  $\{(G^k, z^k, w^k)\}_{k=0}^\infty$  to establish the stationarity properties which, after dereferencing the identifications back to the Algorithm 1 context, yield intended results.

The cross-sectional optimality  $(\bar{z}, \bar{w}) \in \arg \min_{z, w} \phi^*(z, w \mid \bar{\mathbf{x}}_{\mathcal{I}}, \bar{\mathbf{y}}_{\mathcal{I}})$  holds by Proposition 35 Part 1. The proof of the three implications follows, respectively, from implications 2–4 of Proposition 35. □

From Theorem 36 we know that the GS limit points  $(\bar{z}, \bar{w})$  will be optimal for at least one cross-section  $\varphi^{\lambda, \rho, \pi}(\cdot, \cdot \mid \mathbf{x}_{\mathcal{I}}, \mathbf{y}_{\mathcal{I}})$  or  $\phi^\infty(\cdot, \cdot \mid \mathbf{x}_{\mathcal{I}}, \mathbf{y}_{\mathcal{I}})$ .

A simple example demonstrates the possibility that the above GS procedure produces a limit point  $(\bar{z}, \bar{w})$  for which  $\widehat{\partial} \phi^\infty(\bar{z}, \bar{w})$  is empty.

**Example 2** We revisit a rescaled version of the augmented Lagrangian problem of (20) defined for Example 1, where the objective function is rescaled by a factor of  $\frac{1}{\rho}$ .

Of note is the locally optimal solution  $(\tilde{z}^n, \tilde{\mathbf{w}}^n) = (\frac{1}{2}, [0, 0]^T)$ , which for  $0 < \rho^n < \infty$  is clearly a local minimum for  $\varphi^n$  over  $|\bar{z} - z| < 1/\rho^n$ . Furthermore for  $\rho^n < \infty$ ,  $\widehat{\partial} \varphi^n(\tilde{z}^n, \tilde{\mathbf{w}}^n) = \{(0, [0, 0]^T)\}$  is non-empty. However, in the limit as  $\rho \rightarrow \infty$ , we have clearly that  $(\bar{z}, \bar{\mathbf{w}}) = (\frac{1}{2}, [0, 0]^T)$  realises the value  $\phi^\infty(\bar{z}, \bar{\mathbf{w}}) = \frac{1}{4}$  over all cross-sections, but  $\widehat{\partial}_z \phi^\infty(\bar{z}, \bar{\mathbf{w}} \mid \mathbf{x}_{\mathcal{I}}, \mathbf{y}_{\mathcal{I}}) \not\supset \{0\}$  for two of the four cross sections, and furthermore their intersection  $\bigcap_{(\mathbf{x}_{\mathcal{I}}, \mathbf{y}_{\mathcal{I}})} \widehat{\partial}_z \phi^\infty(\bar{z}, \bar{\mathbf{w}} \mid \mathbf{x}_{\mathcal{I}}, \mathbf{y}_{\mathcal{I}}) = \emptyset$  is empty and so  $\widehat{\partial} \phi^\infty(\bar{z}, \bar{\mathbf{w}}) = \emptyset$ . Also note that while this  $(\bar{z}, \bar{\mathbf{w}})$  is a *partial* minimum for  $\phi^\infty$ , it is not (even) a local minimum over  $(z, \mathbf{w})$  jointly.

We have demonstrated through example (but never observed in our experiments) a pathological case where a partial minimum is encountered in the limit, but local optimality or feasibility for SMIP is not achieved. This lack of local optimality is due to a partial minimum being found for  $\phi^*$  where Fréchet subdifferentiability fails, a problem foreshadowed in Lemma 21 (indeed the existence of a non-empty subdifferential ensures stationarity from which local optimality follows, Lemma 11). This

$\mathbf{x}_{\mathcal{I}}$	Cross-Section $\frac{1}{\rho^n} \inf_{\mathbf{w}} \varphi^n(z, \mathbf{w} \mid \mathbf{x}_{\mathcal{I}})$	Gradient w.r.t $z$ $\partial_z \varphi^n(z, \mathbf{w})$	Value $\frac{1}{\rho^n} \varphi^n(\bar{z}^n, \bar{\mathbf{w}}^n)$	Loc. opt. over $B_\delta(\bar{z}, \bar{\mathbf{w}})$ $(\bar{z}^n, \bar{\mathbf{w}}^n)$	$\delta$	$\Phi^*$ $(z^*, \mathbf{w}^*)$
$\begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \\ 0 \\ 1 \\ 1 \\ 1 \end{bmatrix}$	$\frac{1}{\rho^n} + \ z\ ^2$	$2z$	$\frac{1}{\rho^n}$	$\begin{pmatrix} 0 \\ 1 \end{pmatrix}$	$\frac{1}{2} - \frac{1}{\rho^n}$	$\{0\} \times \mathbb{R}^2$
$\begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \\ 1 \\ 1 \\ 1 \end{bmatrix}$	$\frac{2}{\rho^n} + \frac{1}{2} (\ z\ ^2 + \ z - 1\ ^2)$	$2z - 1$	$\frac{2}{\rho^n} + \frac{1}{4}$	$\begin{pmatrix} 1 \\ \frac{1}{2}, 1 \end{pmatrix}$	0	$\left\{ \frac{1}{2} \right\} \times \mathbb{R}^2$
$\begin{bmatrix} 0 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$	$\frac{1}{2} (\ z\ ^2 + \ z - 1\ ^2)$	$2z - 1$	$\frac{1}{4}$	$\begin{pmatrix} 1 \\ \frac{1}{2}, 0 \end{pmatrix}$	$\frac{1}{\rho^n}$	$\left\{ \frac{1}{2} \right\} \times \mathbb{R}^2$
$\begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$	$\frac{1}{\rho^n} + \ z - 1\ ^2$	$2z - 2$	$\frac{1}{\rho^n}$	$\begin{pmatrix} 1 \\ 1, 0 \end{pmatrix}$	$\frac{1}{2} - \frac{1}{\rho^n}$	$\{1\} \times \mathbb{R}^2$



failure of subdifferentiability occurs only when our solution is minimising some (not all) of active sections defining  $\phi^*$ , see Lemma 11. Furthermore, for any solution to problem (26) that satisfies consensus, this lack of Fréchet differentiability is ruled out, see Theorem 36 (2) and we are then assured of obtaining a persistent local minimum. A partial converse of this may be found in Proposition 12 where integer consensus is ensured for a persistent minimum. Such pathological limit points are unstable in the sense that they are mere partial minima but not even locally minimal jointly in  $(z, w)$  for  $\phi^*$ . Consequently, an apt minor perturbation of  $(\bar{z}, \bar{w})$  (suggested by Corollary 18) may be employed to get the iterative FPPH approach unstuck.

## 6 Computational results

### 6.1 Algorithm

Algorithm 3 presents a modified version of Algorithm 1, in which we explicitly consider the initialisation steps and the rules for updating Lagrange and penalty parameters between successive iterations. We use an algorithm formulation consistent with the typical presentation of Progressive Hedging but allowing for differences in how (and when) Lagrange and penalty parameters are updated. Also, since the second-stage discrepancies are always zero in the context of a two-stage SMIP, we omit the second component of the penalty function (the  $v$  component in Assumption 2).

The required properties of the penalty function specified in the ICRF Assumption 2 gives us flexibility in choosing  $\psi$ . As described next, we compute weights for a weighted squared 2-norm form of the penalty function  $\psi$  during the initialisation with the aim of accelerating convergence to a reasonably high-quality feasible solution. Subsequently, we describe update schemes and the termination condition used in our computational experiments.

#### 6.1.1 Initialisation

The penalty function weights in determining the weighted squared 2-norm penalty function  $\psi$  are denoted  $\bar{\mu}_i, i = 1, \dots, n$ , which do not change between iterations  $n \geq 0$ . The iteration  $n \geq 0$  penalty coefficients  $\rho_s^n = \pi_s^n \rho^n$  denote the weighting of the iteration  $n$  penalty magnitude  $\rho^n$  by penalty weight  $\pi_s^n$ . Initially,  $\rho_s^0 = \pi_s^0 = p_s$ . Once the *PenaltyUpdateCondition* is satisfied, the  $\rho_s$  terms are increased by the *PenaltyUpdate* function to modify the penalty applied to each scenario. These functions are defined in Sects. 6.1.2 and 6.1.3. The initial  $z^0$  of Algorithm 3 Line 3 may be computed by  $z_i^0 = \sum_{s \in S} p_s x_{s,i}^0$  for all  $i \in 1, \dots, n$ . Having  $z^0$ , the values  $\mu_i$  for  $i = 1, \dots, n$  which are required to form the penalty weights for initialising the penalty function  $\psi$  are computed as given in [11] for applying Progressive Hedging to SMIPs:

$$\mu_i = \frac{c_i}{\max \left\{ \sum_{s' \in S} p_{s'} \left| x_{s',i}^0 - z_i^0 \right|, 1 \right\}} \quad \text{for each } i \in 1, \dots, n$$

**Algorithm 3** FPPH algorithm for SMIP

---

```

1:  $n \leftarrow 0$ 
2:  $(x_s^0, y_s^0) \in \arg \min_{(x_s, y_s) \in K_s} [c^\top x_s + d_s^\top y_s]$  for all  $s \in S$ 
3:  $z^0 \leftarrow \sum_{s \in S} p_s x_s^0$ 
4: Use  $x^0$  and  $z^0$  to compute  $\{\bar{\mu}_i\}_{i=1}^n$ , set  $\psi(u) := \frac{1}{2} \sum_{i=1}^n \bar{\mu}_i u_i^2$ 
5:  $\rho_s^0 \leftarrow p_s$  for  $s \in S$ 
6: UpdateDualFlag  $\leftarrow$  True
7:  $\lambda_{s,i}^0 \leftarrow -\rho_s^0 \bar{\mu}_i (z_i^0 - x_{s,i}^0)$  for all  $s \in S$  and  $i = 1, \dots, n$ 
8: while  $n < n_{\max}$  and  $x_{s,\mathcal{I}}^n \neq z_{\mathcal{I}}^n$  for at least one  $s \in S$  do
9:    $n \leftarrow n + 1$ 
10:  for all  $s \in S$  do
11:     $(x_s^n, y_s^n) \in \arg \min_{(x_s, y_s) \in K_s} p_s [c^\top x_s + d_s^\top y_s] + (\lambda_s^{n-1})^\top x_s + \rho_s^{n-1} \psi(z^{n-1} - x_s)$ 
12:  end for
13:   $w^n \leftarrow y^n$ 
14:   $z^n \leftarrow (\sum_{s \in S} \rho_s^{n-1} x_s^n) / (\sum_{s \in S} \rho_s^{n-1})$ 
15:  if UpdateDualFlag and PenaltyUpdateCondition( $\lambda$ ) then
16:    UpdateDualFlag  $\leftarrow$  False
17:  end if
18:  if UpdateDualFlag then
19:     $\rho_s^n \leftarrow \rho_s^{n-1}$ ,  $s \in S$ 
20:     $\lambda_{s,i}^n \leftarrow \lambda_{s,i}^{n-1} - \rho_s^{n-1} \bar{\mu}_i (z_i^n - x_{s,i}^n)$  for all  $s \in S$  and  $i = 1, \dots, n$ 
21:  else
22:     $(\rho_s^n)_{s \in S} \leftarrow \text{PenaltyUpdate}((\rho_s^{n-1})_{s \in S}, x^n, z^n)$ 
23:     $\lambda^n \leftarrow \lambda^{n-1}$ 
24:  end if
25: end while
26: return  $(z^n, w^n)$ 

```

---

when variable  $i$  is continuous, and

$$\mu_i = \frac{c_i}{\left(\max_{s' \in S} x_{s',i}^0\right) - \left(\min_{s' \in S} x_{s',i}^0\right) + 1} \quad \text{for each } i \in 1, \dots, n$$

when variable  $i$  is discrete.

We employ a slightly modified version of this scheme, where penalty parameters that would be set to zero by this rule are instead set to the smallest non-zero value  $\mu'$  among all other penalty parameters  $\mu_i$ ,  $i = 1, \dots, n$ . We denote the modified penalty function weights as  $\bar{\mu}_i := \max\{\mu_i, \mu'\}$  for each  $i = 1, \dots, n$ . This modification did not materially affect the performance of Progressive Hedging and provides a guarantee that in the penalty-updating algorithms, all penalties can grow to be arbitrarily large, as required by Assumption 10.

This choice of penalty initialisation has been made to allow as direct a comparison as possible between Progressive Hedging (using a set of parameters established to be reasonable for that algorithm) and the penalty-updating variations of Algorithm 3. Having computed  $\bar{\mu}_i$  for all  $i = 1, \dots, n$ , we set the penalty function for our computational experiments as  $\psi(u) = \frac{1}{2} \sum_{i=1}^n \bar{\mu}_i u_i^2$  and initialise  $\rho_s^0 = \pi_s^0 = p_s$  for all  $s \in S$ . (Note that  $\rho_s^n := \pi_s^n \rho^n$  for  $n \geq 0$ .) With the definition of  $\psi$ , the  $z$  update step

on Line 14 of Algorithm 3 can be written in the form

$$z_i^n \leftarrow \sum_{s \in S} \pi_s^{n-1} x_{s,i}^n \quad \text{for all } i \in 1, \dots, n,$$

where  $\pi_s^n = \frac{\rho_s^n}{\sum_{s' \in S} \rho_{s'}^n}$  for each  $s \in S, n \geq 0$ . (Note that  $\bar{\mu}_i$  does not influence this update step.) Furthermore, the dual multiplier update based on this definition of  $\psi$  is given at each iteration  $n \geq 0$  as

$$\lambda_{s,i}^n \leftarrow \lambda_{s,i}^{n-1} - \rho_s^{n-1} \bar{\mu}_i (z_i^n - x_{s,i}^n) \quad \text{for all } s \in S \text{ and } i = 1, \dots, n.$$

One may verify that this dual multiplier update maintains the feasibility condition  $\sum_{s \in S} \lambda_s^n = 0$  for each  $n \geq 0$ .

### 6.1.2 Penalty update condition

We consider three update-type conditions:

1. *PenaltyUpdateCondition* always returns False, meaning that the algorithm performs dual updates at every iteration and never increases the penalty parameters. This is equivalent to the **Progressive Hedging** algorithm for SMIP. This update condition does **not** satisfy Assumption 10, since the penalty parameters do not become arbitrarily large.
2. *PenaltyUpdateCondition* always returns True, meaning that the algorithm does not update the dual variables again after the initialisation step and instead increases the penalty parameters. This is designated as the **Penalty Only** variant of FPPH.
3. Track the degree of change in the dual variables

$$\Delta_k = \sum_{s \in S} \sum_{i=1}^n |\lambda_{s,i}^{k-1} - \lambda_{s,i}^k|$$

at each iteration  $k$ . If in the current iteration  $k'$  the condition

$$\Delta_{k'} < \beta \frac{\Delta_1 + \max_k \Delta_k}{2} - \gamma \tag{27}$$

is satisfied, *PenaltyUpdateCondition* returns True so that no further dual updates are performed; otherwise it returns False. We set the parameters  $\beta$  and  $\gamma$  to 0.5 and  $10^{-3}$  respectively. This is designated as the **Dual Step Length** variant of FPPH.

As a simple guarantee that the Dual Step Length method satisfies Assumption 10 we could specify a specific number of iterations after which *PenaltyUpdateCondition* must return True. However, in our computational tests with this update condition either (27) or integer-variable consensus was always satisfied after a reasonably small number of iterations.

### 6.1.3 Penalty update scheme

We gradually increase the penalty parameter for the scenario whose first-stage variables are furthest from consensus with the following method. For each scenario  $s \in S$ , we calculate its distance from consensus  $D_s^n = \|z^n - x_s^n\|_2$ . Then, update the penalty multipliers as follows:

$$\rho_s^n \leftarrow \left(1 + \alpha |S| \frac{D_s^n}{\sum_{s \in S} D_s^n}\right) \rho_s^{n-1} \quad \text{for all } s \in S.$$

We set the parameter  $\alpha$  to 0.1. This rule is intended to prioritise increasing the penalty parameters corresponding to the scenarios whose first-stage variables are furthest from consensus. Assuming that *PenaltyUpdateCondition* returns True after a finite number of iterations, this update scheme satisfies Assumption 10.

### 6.1.4 Termination condition and $n_{max}$

Termination of each computational test is conditioned on attaining consensus  $z_{\mathcal{I}}^n = x_{s,\mathcal{I}}$ , for all  $s \in S$  in all integer variables. For the instances with pure integer first-stage variables, this condition is the same as requiring first-stage consensus. For the instances with mixed integer first-stage variables, we generally do not have full consensus in the continuous variables at this point. To obtain feasible solutions, we take each unique first-stage solution  $x_s$  and find the corresponding optimal second-stage decisions  $y$ . We then report the best solution value found among these candidate solutions.

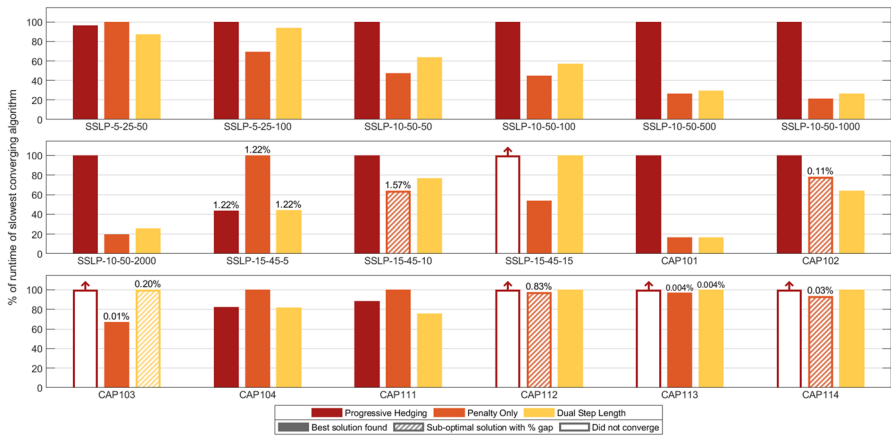
Our motivation for applying this convergence criterion to mixed integer first-stage instances is that when allowed to run beyond achieving integer consensus, the FPPH variants typically satisfied the convergence criterion  $\sqrt{\sum_{s \in S} p_s \|x_s^n - z^n\|_2^2} < 10^{-3}$  within 100 iterations but with very poor solution quality, whereas PH failed to satisfy this criterion given even 200 iterations. Any potential method for finding a high-quality solution quickly given a fixed value for the first-stage integer variables could be applied to the solutions produced by both PH and FPPH, but implementing and tuning such a method is outside the scope of this paper.

We set  $n_{max} = 100$  since both FPPH variants generally converge well within this iteration limit, and in cases where PH does not it is already clearly slower than FPPH in terms of both runtime and iteration count.

## 6.2 Computation environment

The experiments in this section were conducted with a C++ implementation of Algorithm 3 using CPLEX 22.1 [28] as the solver. For reading SMPS files into scenario-specific subproblems and for their interface with CPLEX, we used modified versions of the COIN-OR [29] Smi and Osi libraries to instantiate appropriate C++ class instances of the subproblems directly.

The computing environment is the Gadi cluster maintained by Australia's National Computing Infrastructure (NCI) and supported by the Australian government [30]. To



**Fig. 1** Comparison between our implementation of Progressive Hedging and variants of FPPH, applied to instances with pure integer first stage (SSLP and CAP). Bar height indicates the time required for convergence compared to the slowest converging algorithm. Solid bars indicate the best quality solution found among the three algorithms. Tinted bars indicate convergence to a lower-quality solution. Suboptimal solutions are indicated by a percentage optimality gap. A solid bar with no percentage gap indicates the optimal solution was found. Empty bars indicate non-convergence within 100 iterations; the arrow signifies that these 100 iterations took much longer than the slowest converging algorithm

maintain a comparable environment, experiments were performed on a single CPU using one thread per CPLEX solve for both algorithms.

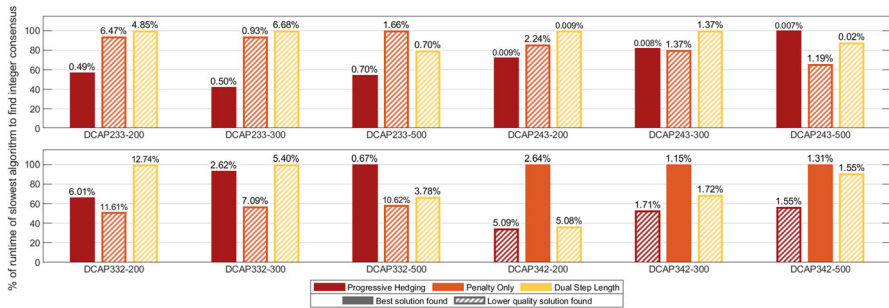
The PH and FPPH algorithms are deterministic in terms of the solutions produced, but the time required for CPLEX to solve the subproblems at each iteration has some variation. Therefore, for each test, we ran each algorithm three times on each instance and report the average runtime.

### 6.3 Computational experiments: Pure integer first-stage instances

We first consider the application of FPPH and our implementation of Progressive Hedging to the CAP instance set [31] using the first 250 scenarios for each instance, and the SSLP instance set [32]. To evaluate algorithm performance we compare it to the known IP optimal solution. To obtain the integer feasible optimal solutions for the CAP instances we used CPLEX to directly solve the MIP reformulation of each instance. The integer feasible optimal solutions for the SSLP instances are provided by SIPLIB [9].

The computational results are summarised in Fig. 1. These figures compare both the wall-clock time required for convergence (compared to the slowest algorithm to achieve convergence) and the quality of the feasible solutions obtained at termination. A more detailed summary of our results, including absolute runtime and solution values, is provided in the supplementary material (Table B1).

When applied to the SSLP instances, all three algorithms typically find the same solution, and it is often optimal. The Dual Step Length variant of FPPH outperforms PH in terms of runtime for all instances except for SSLP-15-45-5, where they require an



**Fig. 2** Comparison between our implementation of Progressive Hedging and variants of FPPH, applied to instances with mixed integer first stage (DCAP). Bar height indicates the time required for integer variable consensus compared to the slowest converging algorithm. Solid bars indicate the best quality solution found among the three algorithms. Tinted bars indicate lower-quality solutions. The percentage gap between found solution and the known upper bound is given in each case

equal (and small) amount of time. The Penalty Only variant of FPPH often outperforms both the Dual Step Length variant and PH, but fails to find the optimal solution of SSLP-15-45-10 and is a little slower when applied to SSLP-5-25-50 and SSLP-15-45-5. PH fails to converge to a feasible solution within 100 iterations when applied to SSLP-15-45-15.

When applied to the CAP instances, PH fails to converge to a feasible solution within 100 iterations for four of the eight instances and is again consistently outperformed in terms of runtime and matched in solution quality by the Dual Step Length variant of FPPH even when it does converge. There is not a clear favourite between the Penalty Only and Dual Step Length variants when applied to the CAP instances; each variant finds a higher-quality solution than the other variant for at least one instance, and converges faster than the other variant for several instances.

#### 6.4 Computation experiments: Mixed-integer first-stage instances

We also compared the performance of FPPH and our implementation of Progressive Hedging applied to the DCAP instance set [33, 34]. In this case, we compare with the known upper bounds given by SIPLIB [9]. These results are summarised in Fig. 2, with further detail in the supplementary material (Table B2).

For these instances, PH consistently obtains consensus in the integer variables within 100 iterations and generally outperforms the Dual Step Length variant of FPPH in terms of runtime and solution quality. The Penalty Only variant of FPPH obtains better solution quality than PH when applied to DCAP342 (with 200, 300 and 500 scenarios) but finds considerably worse solutions when applied to the other DCAP instances.

## 7 Conclusions

We have shown that the tools and techniques of variational analysis are well-suited for the analysis of the progressive hedging algorithm as applied to SMIP. Indeed the analysis interfaces well with the “just MIP it” approach to the development heuristics in this field. It allows for a new study of augmented Lagrangians and Gauss–Seidel methods, specifically recognising where the presence of smoothness is essential for the success of algorithmic approaches. The theory is able to shed light on how critical parameters need to be updated to ensure convergence.

Our computational results demonstrate that the FPPH algorithm, which is motivated by the above theory, has the potential to outperform PH in terms of quickly and reliably converging to high-quality feasible solutions for SMIP instances, particularly those with pure integer first-stage variables. By contrast, PH tended to outperform FPPH when applied to the DCAP instances which have mixed-integer first-stage variables, though the variant of FPPH performing no dual variable updates found higher-quality solutions than PH for the DCAP342 subclass. Further testing on a wider variety of instance classes is needed for a deeper understanding of how the structure of SMIP instances influences the relative performance of PH and FPPH, and how to set the penalty update rules of FPPH for the best performance on a given class of instances.

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1007/s10589-023-00532-w>.

**Funding** Open Access funding enabled and organized by CAUL and its Member Institutions. This work was supported by the Australian Research Council (ARC) through Discovery Program under Grants DP140100985.

**Data availability** The problem instances are based on the data from three distinct sets of problems, namely the capacitated facility location problem (CAP) from [31], the server location under uncertainty problem (SSLP), introduced in [32] and the dynamic capacity allocation problem (DCAP) available in [33]. (SSLP & DCAP) <https://www2.isye.gatech.edu/~sahmed/siplib/> (CAP) The instances are based on the data from: <http://people.brunel.ac.uk/~mastjjb/jeb/orlib/capinfo.html> Data was provided by the authors of [31] on request.

## Declarations

**Conflict of interest** The authors declare that they have no conflict of interest.

**Ethical approval** Not applicable.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## Appendix A: Proofs of selected results

### A.1 Proof of Proposition 4

**Proof** The proofs of all four claims are the same for each  $s \in S$ . The first three claims are obvious from the compactness of the constraint sets  $K_s$ , the coercivity of  $\psi$  and its other properties as implied by Assumptions 2, and the role of  $\rho$  in the objective function  $(x_s, y_s, w_s, z) \mapsto f_s(x_s, y_s) + \lambda_s^\top x_s + \rho \pi_s \psi(z - x_s, w_s - y_s)$ . To show claim (4), let  $(z^1, w_s^1), (z, w_s) \in \mathbb{X} \times \mathbb{Y}_s, \|(z^1 - z, w_s^1 - w_s)\| < \delta$ , and let  $(x_s^\rho, y_s^\rho) \in \Phi_s^{\lambda_s, \rho, \pi}(z, w_s)$ . Via convexity [13, Example 9.14], we may choose,  $L_s^\Delta \geq 1$  to be the local Lipschitz continuity modulus associated with the finite valued function  $\psi$  on  $\overline{B}_\Delta(0, 0) := \{(u, v) : \|(u, v)\|_2 \leq \Delta\}$ , with  $\Delta = \|(z - x_s^\rho, w_s - y_s^\rho)\| + \delta$ . Then using dual feasibility ( $\sum_{s \in S} \lambda_s = 0$ ) we have

$$\begin{aligned} \varphi_s^{\lambda_s, \rho, \pi}(z^1, w_s^1) - \varphi_s^{\lambda_s, \rho, \pi}(z, w_s) &\leq \sum_{s \in S} \lambda_s (z^1 - x_s^\rho - (z - x_s^\rho)) \\ &\quad + \pi_s \rho [\psi(z^1 - x_s^\rho, w_s^1 - y_s^\rho) - \psi(z - x_s^\rho, w_s - y_s^\rho)] \\ &\leq \pi_s \rho L^\Delta \|(z^1 - z, w_s^1 - w_s)\| \end{aligned}$$

for  $(z, w_s), (z^1, w_s^1) \in K_s^\Delta$ , thus establishing the Lipschitz modulus  $\pi_s \rho L^\Delta$ . It readily follows by summation that  $\varphi^{\lambda, \rho, \pi}$  is Lipschitz continuous with modulus  $\rho L^\Delta$  for  $(z, w_s), (z^1, w_s^1) \in K_s^\Delta, s \in S$ .  $\square$

### A.2 Proof of Lemma 6

**Proof** By the continuity of  $\nabla \psi$ , we have for each fixed  $D > 0$  that there exists  $\tilde{\delta} > 0$  for which  $\|\nabla \psi(u^0, v^0)\| < \frac{m}{2} D$  whenever  $\|(u^0, v^0)\| < \tilde{\delta}$ . For any discrepancy  $(u, v)$  for which  $\|(u - u^0, v - v^0)\| > D$ , we have  $\|\nabla \psi(u^0, v^0)\| < \frac{m}{2} \|(u - u^0, v - v^0)\|$ , which then implies by the Cauchy-Schwartz inequality

$$\langle \nabla \psi(u^0, v^0) + \frac{m}{2}(u - u^0, v - v^0), (u - u^0, v - v^0) \rangle > 0$$

which implies

$$\langle \nabla \psi(u^0, v^0), (u - u^0, v - v^0) \rangle + \frac{m}{2} \|(u - u^0, v - v^0)\|^2 > 0.$$

Adding the last inequality to the inequality

$$\psi(u, v) \geq \psi(u^0, v^0) + \nabla \psi(u^0, v^0), (u - u^0, v - v^0) + \frac{m}{2} \|(u - u^0, v - v^0)\|^2$$

obtained due to Assumption 2(3), we have the intended inequality  $\psi(u^0, v^0) < \psi(u, v)$ .  $\square$



### A.3 Proof of proposition 17

**Proof** Proof of 1: Assume for sake of contradiction that  $\limsup_{n \rightarrow \infty} \|\tilde{z}^n - \tilde{x}_s^n\| > 0$  for at least one  $s \in S$ . As  $\tilde{z}^n \rightarrow \bar{z}$ ,  $\bar{\delta} > 0$ , and the boundedness of  $X$ , so there exists a fixed  $0 < \eta < 1$  such that for all sufficiently large  $n \geq 0$  and any  $\bar{x}^n \in X$  we have  $z^{\eta,n} := \eta\tilde{z}^n + (1 - \eta)\bar{x}^n \in B_{\bar{\delta}}(\tilde{z}^n)$  (indeed there exists a fixed  $\kappa > 0$  such that  $\kappa B_{\bar{\delta}}(\tilde{z}^n) \supseteq X$  for  $n$  large).

For any given  $n$ , there exists  $s' \in \arg \min_{s \in S} \psi(\eta(\tilde{z}^n - \tilde{x}_s^n), 0)$ . Set  $\bar{x}^n := \tilde{x}_{s'}^n \in X$ . Then

$$\sum_{s \in S} \pi_s^n \psi(\eta(\tilde{z}^n - \bar{x}^n), 0) = \psi(\eta(\tilde{z}^n - \bar{x}^n), 0) \leq \sum_{s \in S} \pi_s^n \psi(\eta(\tilde{z}^n - \tilde{x}_s^n), 0) \tag{A1}$$

By the relative recourse assumption, there exists  $\bar{y}^n \in Y_s(\bar{x}^n)$  for all  $s \in S$ .

Let  $z^{\eta,n} = \eta\tilde{z}^n + (1 - \eta)\bar{x}^n$  and note that for any  $0 < \eta < 1$  we have by (A1)

$$\begin{aligned} \sum_{s \in S} \pi_s^n \psi(z^{\eta,n} - \bar{x}^n, 0) &= \sum_{s \in S} \pi_s^n \psi(\eta(\tilde{z}^n - \bar{x}^n), 0) \leq \sum_{s \in S} \pi_s^n \psi(\eta(\tilde{z}^n - \tilde{x}_s^n), 0) \\ &\leq \sum_{s \in S} \pi_s^n \psi(\tilde{z}^n - \tilde{x}_s^n, 0) - \frac{m(1 - \eta)^2}{2} \sum_{s \in S} \pi_s^n \|\tilde{z}^n - \tilde{x}_s^n\|^2. \end{aligned}$$

We used the fact that the gradient descent term  $\langle \nabla \psi_z(\eta(\tilde{z}^n - \tilde{x}_s^n), 0), (1 - \eta)(\tilde{z}^n - \tilde{x}_s^n) \rangle$  of the ICRF strong convexity assumption 2(3) is guaranteed to be positive due to the convexity of  $\psi$  and the strict increasing property in ICRF assumption 2(2). Define

$$\begin{aligned} M := \limsup_{n \rightarrow \infty} \max_{x, x', y, y'} \{ &\sum_{s \in S} p_s [(c + \frac{\lambda_s^n}{p_s})^\top (x_s - x'_s) + d_s^\top (y - y')] \\ &| (x, y_s), (x', y'_s) \in K_s, s \in S \}, \end{aligned}$$

which is guaranteed to be finite due to the boundedness of  $K_s, s \in S$  and the PWA assumption 10 boundedness of  $\{\lambda^n\}_{n=0}^\infty$ . We have for each  $n \geq 0$  and  $\rho^n$  sufficiently large that

$$\begin{aligned} \sum_{s \in S} \inf_w \varphi_s^n(z^{\eta,n}, w_s) &\leq \sum_{s \in S} p_s \left[ \left( c + \frac{\lambda_s^n}{p_s} \right)^\top \bar{x}^n + \inf_{y \in Y_s(\bar{x}^n)} \{d^\top y_s\} \right] + \rho^n \pi_s^n \psi(z^{\eta,n} - \bar{x}^n, 0) \\ &\leq \sum_{s \in S} p_s [(c + \frac{\lambda_s^n}{p_s})^\top \bar{x}^n + \inf_{y \in Y_s(\bar{x}^n)} \{d^\top y_s\}] + \rho^n \pi_s^n \psi(\tilde{z}^n - \tilde{x}_s^n, 0) \\ &\quad - \frac{m(1 - \eta)^2 \rho^n}{2} \sum_{s \in S} \pi_s^n \|\tilde{z}^n - \tilde{x}_s^n\|^2 \\ &\leq \sum_{s \in S} \inf_w \{\varphi_s^n(\tilde{z}^n, w_s)\} + M - \frac{m(1 - \eta)^2 \rho^n}{2} \sum_{s \in S} \pi_s^n \|\tilde{z}^n - \tilde{x}_s^n\|^2 < \sum_{s \in S} \inf_w \varphi_s^n(\tilde{z}^n, w_s). \end{aligned}$$

The last inequality follows from the assumptions that  $\limsup_{n \rightarrow \infty} \|\tilde{z}^n - \tilde{x}_s^n\| > 0$  for at least one  $s \in S$  and the assumption that the corresponding  $\pi_s^n$  values are bounded away from zero as due to the PWA assumption 10(2). But since  $z^{n,n} \in B_{\delta}(\bar{z})$ , for  $n$  large enough, we have a contradiction of the local optimality of  $\tilde{z}^n$  for  $z \mapsto \inf_w \varphi^n(z, w)$  for some  $n$  arbitrarily large. Therefore,  $\lim_{n \rightarrow \infty} \|\tilde{z} - \tilde{x}_s\| = 0$  for all  $s \in S$ , as long as  $\pi_s^n \not\rightarrow 0$  as in the PWA assumptions 10(2), and hence  $\tilde{x}_s^n \rightarrow \bar{z}$ . In particular  $\sum_{s \in S} \pi_s^n \tilde{x}_s^n \rightarrow \bar{z}$ . Taking into account the definition of  $S_n$  we have  $\tilde{x}_{s,\mathcal{I}}^n = \tilde{z}_{\mathcal{I}}^n$  for all  $s \in S$  and  $n$  sufficiently large. (When  $\psi = \frac{1}{2} \|\cdot\|^2$  we have from Lemma 13 that  $\tilde{z}^n \in \operatorname{argmin}_z \frac{1}{2} \sum_{s \in S} \|z - \tilde{x}_s^n\|^2$  so  $\tilde{z}^n = \sum_{s \in S} \pi_s \tilde{x}_s^n$ . Moreover, in general  $\sum_{s \in S} \pi_s \tilde{x}_s^n \in X$  for  $n$  large enough due to the fact that all  $\tilde{x}_s^n \in X$ ,  $\tilde{z}_{\mathcal{I}}^n = \tilde{x}_{s,\mathcal{I}}^n$  for all  $s \in S$  and  $\sum_{s \in S} \pi_s = 1$ .)

Proof of 2: By Proposition 3 have the first assertion of 4. holding. The second follows from the closed graph of  $x_s \mapsto Y_s(x_s)$ . □

### A.4 Proof of lemma 31

**Proof** The proof is by induction, with the key observations that, for  $j = 1$

$$\begin{aligned} &g^n(x^{n+1}, y^{n+1}, w^{n+1}, z^n + \alpha^n d^n) - g^n(x^{n+1}, y^{n+1}, w^{n+1}, z^n) \\ &\geq g^n(x^{n+1}, y^{n+1}, w^{n+1}, z^{n+1}) - g^n(x^{n+1}, y^{n+1}, w^{n+1}, z^n) \\ &= g^n(x^{n+1}, y^{n+1}, w^{n+1}, z^{n+1}) - g^{n+1}(x^{n+1}, y^{n+1}, w^{n+1}, z^{n+1}) \\ &\quad + g^{n+1}(x^{n+1}, y^{n+1}, w^{n+1}, z^{n+1}) - g^n(x^{n+1}, y^{n+1}, w^{n+1}, z^n) \\ &\geq g^n(x^{n+1}, y^{n+1}, w^{n+1}, z^{n+1}) - g^{n+1}(x^{n+1}, y^{n+1}, w^{n+1}, z^{n+1}) \\ &\quad + g^{n+1}(x^{n+2}, y^{n+2}, w^{n+2}, z^{n+1}) - g^n(x^{n+1}, y^{n+1}, w^{n+1}, z^n) \end{aligned}$$

and, more generally for  $j > 1$ , the satisfaction of the inequality

$$\begin{aligned} &g^n(x^{n+1}, y^{n+1}, w^{n+1}, z^n + \alpha^n d^n) - g^n(x^{n+1}, y^{n+1}, w^{n+1}, z^n) \\ &\geq \sum_{i=0}^{j-2} \left[ g^{n+i}(x^{n+i+1}, y^{n+i+1}, w^{n+i+1}, z^{n+i+1}) \right. \\ &\quad \left. - g^{n+i+1}(x^{n+i+1}, y^{n+i+1}, w^{n+i+1}, z^{n+i+1}) \right] \\ &\quad + g^{n+j-1}(x^{n+j}, y^{n+j}, w^{n+j}, z^{n+j-1}) - g^n(x^{n+1}, y^{n+1}, w^{n+1}, z^n) \end{aligned}$$

implies

$$g^n(x^{n+1}, y^{n+1}, w^{n+1}, z^n + \alpha^n d^n) - g^n(x^{n+1}, y^{n+1}, w^{n+1}, z^n)$$

$$\begin{aligned}
 &\geq \sum_{i=0}^{j-2} \left[ g^{n+i}(x^{n+i+1}, y^{n+i+1}, w^{n+i+1}, z^{n+i+1}) \right. \\
 &\quad \left. - g^{n+i+1}(x^{n+i+1}, y^{n+i+1}, w^{n+i+1}, z^{n+i+1}) \right] \\
 &\quad + g^{n+j-1}(x^{n+j}, y^{n+j}, w^{n+j}, z^{n+j}) - g^{n+j}(x^{n+j}, y^{n+j}, w^{n+j}, z^{n+j}) \\
 &\quad + g^{n+j}(x^{n+j}, y^{n+j}, w^{n+j}, z^{n+j}) - g^n(x^{n+1}, y^{n+1}, w^{n+1}, z^n) \\
 &\geq \sum_{i=0}^{j-2} \left[ g^{n+i}(x^{n+i+1}, y^{n+i+1}, w^{n+i+1}, z^{n+i+1}) \right. \\
 &\quad \left. - g^{n+i+1}(x^{n+i+1}, y^{n+i+1}, w^{n+i+1}, z^{n+i+1}) \right] \\
 &\quad + g^{n+j-1}(x^{n+j}, y^{n+j}, w^{n+j}, z^{n+j}) - g^{n+j}(x^{n+j}, y^{n+j}, w^{n+j}, z^{n+j}) \\
 &\quad + g^{n+j}(x^{n+j+1}, y^{n+j+1}, w^{n+j+1}, z^{n+j}) - g^n(x^{n+1}, y^{n+1}, w^{n+1}, z^n) \\
 &= \sum_{i=0}^{j-1} \left[ g^{n+i}(x^{n+i+1}, y^{n+i+1}, w^{n+i+1}, z^{n+i+1}) \right. \\
 &\quad \left. - g^{n+i+1}(x^{n+i+1}, y^{n+i+1}, w^{n+i+1}, z^{n+i+1}) \right] \\
 &\quad + g^{n+j}(x^{n+j+1}, y^{n+j+1}, w^{n+j+1}, z^{n+j}) - g^n(x^{n+1}, y^{n+1}, w^{n+1}, z^n)
 \end{aligned}$$

□

### A.5 Proof of lemma 33

**Proof** As aid, denote  $\rho^{n_k+i} = \rho^{n_k} \eta_i^k$  for  $i = 0, 1, \dots, j_k - 1$ , where it is evident for all  $k \geq 0$  that  $\eta_0^k = 1$  and  $\eta_i^k \geq 1$  for  $i > 1$ . Thus, given fixed  $k$  with  $n_{k+1} = n_k + j_k$ , the sum of differences becomes (setting  $j = j_k$ )

$$\sum_{s \in S} \sum_{i=0}^{j-1} \frac{\left( \frac{1}{\eta_i^k} - \frac{1}{\eta_{i+1}^k} \right) f_s(x_s^{n_k+i+1}, y_s^{n_k+i+1})}{\rho^{n_k}} \tag{A2a}$$

$$- \left( \frac{\lambda_s^{n_k+i}}{\rho^{n_k+i}} - \frac{\lambda_s^{n_k+i+1}}{\rho^{n_k+i+1}} \right)^\top x_s^{n_k+i+1} \tag{A2b}$$

$$+ \left( \pi_s^{n_k+i} - \pi_s^{n_k+i+1} \right) \psi(z^{n_k+i+1} - x_s^{n_k+i+1}, w_s^{n_k+i+1} - y_s^{n_k+i+1}). \tag{A2c}$$

We address the vanishing limits of the three summation term groupings (A2a)–(A2c) separately. Because  $\frac{1}{\eta_i^k} - \frac{1}{\eta_{i+1}^k} \geq 0$  and due to the boundedness

$$-M \leq \sum_{s \in S} \left[ f_s(x_s^{n_k+i+1}, y_s^{n_k+i+1}) \right] \leq M$$

of the inner summation over  $s \in S$  in (A2a) implied by the SMIP assumption 1 and assumption 29, we have

$$\begin{aligned}
 -\frac{M}{\rho_n} \sum_{i=0}^{j-1} \left[ \frac{1}{\eta_i^k} - \frac{1}{\eta_{i+1}^k} \right] &\leq \frac{1}{\rho_n} \sum_{i=0}^{j-1} \left[ \frac{1}{\eta_i^k} - \frac{1}{\eta_{i+1}^k} \right] \sum_{s \in S} \left[ f_s(x_s^{n_k+i+1}, y_s^{n_k+i+1}) \right] \\
 &\leq \frac{M}{\rho_n} \sum_{i=0}^{j-1} \left[ \frac{1}{\eta_i^k} - \frac{1}{\eta_{i+1}^k} \right]
 \end{aligned}$$

which implies (via telescoping)

$$\begin{aligned}
 -\frac{M}{\rho_n} \left[ \frac{1}{\eta_0^k} - \frac{1}{\eta_{jk}^k} \right] &\leq \frac{1}{\rho_n} \sum_{i=0}^{j-1} \left[ \frac{1}{\eta_i^k} - \frac{1}{\eta_{i+1}^k} \right] \sum_{s \in S} \left[ f_s(x_s^{n_k+i+1}, y_s^{n_k+i+1}) \right] \\
 &\leq \frac{M}{\rho_n} \left[ \frac{1}{\eta_0^k} - \frac{1}{\eta_{jk}^k} \right]
 \end{aligned}$$

If  $\lim_{n \rightarrow \infty} \rho_n = \infty$ , then  $\lim_{n \rightarrow \infty} \frac{M}{\rho_n} = 0$  while  $\left[ \frac{1}{\eta_0^k} - \frac{1}{\eta_j^k} \right] \leq 1$  for all  $k \geq 0$ , and so the summation (A2a) vanishes. Otherwise, if  $\lim_{n \rightarrow \infty} \rho_n = \bar{\rho} < \infty$ , then  $\lim_{n \rightarrow \infty} \frac{M}{\rho_n} = \frac{M}{\bar{\rho}}$  and  $\lim_{k \rightarrow \infty} \left[ \frac{1}{\eta_0^k} - \frac{1}{\eta_j^k} \right] = 0$  will vanish instead (as  $\eta_{jk}^k = 1$  eventually). Therefore, the summation (A2a) vanishes in either case as  $k \rightarrow \infty$ .

To demonstrate the vanishing of the second summation (A2b) we note that when  $\rho^n \rightarrow \infty$  then this term vanishes since  $x_s \in X$  being compact implies  $\|x_s\| \leq M < +\infty$ . Otherwise if  $\rho^n = \bar{\rho}$  eventually we have the inner sum of the term eventually bounded by  $\frac{M}{\bar{\rho}} (\sum_{i=0}^{j_n} \|\lambda_s^{n_k} - \lambda_s^{n_{k+i}}\|)$ . Now  $\lim_{k \rightarrow \infty} (\sum_{i=0}^{j_n} \|\lambda_s^{n_k} - \lambda_s^{n_{k+i}}\|) = 0$  as the tail of a convergent series converges to 0.

To demonstrate the vanishing of the third summation (A2c) as  $k \rightarrow \infty$ , the satisfaction of assumption 29(3) implies that it only needs to be shown that the sequence

$$\left\{ \psi_s(z^{n_k+i} - x_s^{n_k+i+1}, w_s^{n_k+i+1} - y_s^{n_k+i+1}) \right\}_{k=0}^{\infty} \tag{A3}$$

is bounded in magnitude by  $M$  for all  $n$ . Indeed if this is the case then the inner summation of the third term may be bounded in both cases: we have

$$\begin{aligned}
 \sum_{i=0}^{j-1} (\pi_s^{n_k+i} - \pi_s^{n_k+i+1}) \psi \left( z^{n_k+i+1} - x_s^{n_k+i+1}, w_s^{n_k+i+1} - y_s^{n_k+i+1} \right) \\
 \leq M \sum_{i=0}^{j-1} \left\| \pi_s^{n_k+i} - \pi_s^{n_k+i+1} \right\|.
 \end{aligned}$$

Now  $\lim_{k \rightarrow \infty} \sum_{i=0}^{j-1} \left\| \pi_s^{n_k+i} - \pi_s^{n_k+i+1} \right\| = 0$  as the tail of a convergent series converges to 0. The boundedness of (A3) follows from the strong convexity assumption

on  $\psi$  and the compactness of  $K$ . Indeed it is enough to observe that under these assumptions we have  $\liminf_{\|z\| \rightarrow \infty} \inf_{x \in X} \{\psi(z - x, 0)\} = +\infty$ . But the convergence  $\lim_{k \rightarrow \infty} z^{n_k} = \bar{z}$  while  $(x, y)$  are bounded due to the compactness of  $K$ , and so the boundedness of (A3) must hold.  $\square$

## References

1. Birge, J.R., Louveaux, F.: Introduction to stochastic programming. Springer, Berlin (2011)
2. Montero, L., Bello, A., Reneses, J.: A review on the unit commitment problem: approaches, techniques, and resolution methods. *Energies* **15**(4), 1296 (2022)
3. Parvez, I., Shen, J., Cheng, C., Parvez, I., Shen, J., Khan, M., Cheng, C.: Modeling and solution techniques used for hydro generation scheduling. *Water* **11**(7) (2019)
4. Turan, H., Kahagalage, S., El Sawah, S., Jalalvand, F.: A multi-objective simulation-optimization for a joint problem of strategic facility location, workforce planning, and capacity allocation: a case study in the royal Australian navy. *Expert Syst. Appl.* **186**, 115751 (2021)
5. Fadaki, M., Abareshi, A., Lee, P., Far, S.: Multi-period vaccine allocation model in a pandemic: a case study of covid-19 in Australia. *Trans. Res. Part E: Logist. Trans. Rev.* **161**, 102689 (2022)
6. Shone, R., Glazebrook, K., Zografos, K.: Applications of stochastic modeling in air traffic management: methods, challenges and opportunities for solving air traffic problems under uncertainty. *Eur. J. Oper. Res.* **292**(1), 1–26 (2021)
7. Badilla Veliz, F., Watson, J.-P., Weintraub, A., Wets, R.J.-B., Woodruff, D.L.: Stochastic optimization models in forest planning: a progressive hedging solution approach. *Ann. Oper. Res.* **232**, 259–274 (2015)
8. Fan, Y., Schwartz, F., Voß, S., Woodruff, D.L.: Stochastic programming for global supply chain planning under uncertainty: an outline. In: Bektaş, T., Coniglio, S., Martinez-Sykora, A., Voß, S. (eds.) *Computational Logistics*, pp. 437–451. Springer, Cham (2017)
9. Ahmed, S., Garcia, R., Kong, N., Ntaimo, L., Parija, G., Qiu, F., Sen, S.: SIPLIB: A stochastic integer programming test problem library (2015). <http://www.isye.gatech.edu/sahmed/siplib>
10. Rockafellar, R.T., Wets, R.J.-B.: Scenarios and policy aggregation in optimization under uncertainty. *Math. Oper. Res.* **16**(1), 119–147 (1991)
11. Watson, J.-P., Woodruff, D.L.: Progressive hedging innovations for a class of stochastic mixed-integer resource allocation problems. *CMS* **8**(4), 355–370 (2011)
12. Boyd, S., Parikh, N., Chu, E., Peleato, B., Eckstein, J.: Distributed optimization and statistical learning via the alternating direction method of multipliers. *Found. Trends Mach. Learn.* **3**(1), 1–122 (2011)
13. Rockafellar, R.T., Wets, R.J.-B.: *Variational Analysis*. Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], vol. 317, p. 733. Springer, Berlin (1998)
14. Burachik, R.S., Rubinov, A.: Abstract convexity and augmented Lagrangians. *SIAM J. Optim.* **18**(2), 413–436 (2007)
15. Boland, N.L., Eberhard, A.C.: On the augmented Lagrangian dual for integer programming. *Math. Program.* **150**(2, Ser. A), 491–509 (2015)
16. Feizollahi, M.J., Ahmed, S., Sun, A.: Exact augmented Lagrangian duality for mixed integer linear programming. *Mathematical Programming*, 1–23 (2016)
17. Oliveira, F., Christiansen, J., Dandurand, B., Eberhard, A.: Combining penalty-based and gauss-seidel methods for solving stochastic mixed-integer problems. *Int. Trans. Operat. Res.* **27**(1), 494–524 (2018)
18. Frank, M., Wolfe, P.: An algorithm for quadratic programming. *Naval Res. Logist. Quart.* **3**(1–2), 95–110 (1956)
19. Boland, N., Christiansen, J., Dandurand, B., Eberhard, A., Linderoth, J., Luedtke, J., Oliveira, F.: Combining progressive hedging with a frank-wolfe method to compute Lagrangian dual bounds in stochastic mixed-integer programming. *SIAM J. Optim.* **28**(2), 1312–1336 (2018)
20. Boland, N.L., Eberhard, A.C., Engineer, F., Tsoukalas, A.: A new approach to the feasibility pump in mixed integer programming. *SIAM J. Optim.* **22**(3), 831–861 (2012)
21. Geißler, B., Morsi, A., Schewe, L., Schmidt, M.: Penalty alternating direction methods for mixed-integer optimization: a new view on feasibility pumps. *SIAM J. Optim.* **27**(3), 1611–1636 (2017)

22. Tseng, P.: Convergence of a block coordinate descent method for nondifferentiable minimization. *J. Optim. Theory Appl.* **109**, 475–494 (2001)
23. Roshchina, V.: Exact calculus of fréchet subdifferentials for hadamard directionally differentiable functions. *Nonlinear Anal.* **69**, 1112–1124 (2008)
24. Bertsekas, D.P.: *Nonlinear Programming*. Athena Scientific, Nashua, NH (1999)
25. Bonettini, S.: Inexact block coordinate descent methods with application to non-negative matrix factorization. *IMA J. Numer. Anal.* **31**(4), 1431–1452 (2011)
26. Boland, N., Christiansen, J., Dandurand, B., Eberhard, A., Oliveira, F.: A parallelizable augmented Lagrangian method applied to large-scale non-convex-constrained optimization problems. *Math. Progr.* **175**, 503–536 (2018)
27. Beer, G. (1993) *Topologies on closed and closed convex sets*. mathematics and its applications, vol. 268, Kluwer Academic Publishers Group, Dordrecht , p. 340
28. IBM Corporation: IBM ILOG CPLEX V22.1. IBM Corporation. Last accessed 27 Jun 2022. <https://www.ibm.com/docs/en/icos/22.1.0>
29. COmputational INfrastructure for Operations Research. Last accessed 29 March, 2020. <http://www.coin-or.org/>
30. National Computing Infrastructure (NCI): NCI Website. National Computing Infrastructure (NCI). Last accessed 19 November 2016. <http://www.nci.org.au>
31. Bodur, M., Dash, S., Günlük, O., Luedtke, J.: Strengthened benders cuts for stochastic integer programs with continuous recourse. Technical Report RC25452, IBM Research Report, 2014. Available as Optimization Online 2014-03-4263 (2014)
32. Ntaimo, L., Sen, S.: The million-variable “march” for stochastic combinatorial optimization. *J. Global Optim.* **32**(3), 385–400 (2005)
33. Ahmed, S., Garcia, R.: Dynamic capacity acquisition and assignment under uncertainty. *Ann. Oper. Res.* **124**, 267–283 (2003)
34. Ahmed, S., Tawarmalani, M., Sahinidis, N.V.: A finite branch-and-bound algorithm for two-stage stochastic integer programs. *Math. Program.* **100**(2(Ser. A)), 355–377 (2004)

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.