Citation:
Zhang, M. and Hou, L. and Sun, J. and Yan, A. 2020. A Model of Multistage Risk-Averse Stochastic Optimization and its Solution by Scenario-Based Decomposition Algorithms. Asia-Pacific Journal of Operational Research. 37 (4): ARTN 2040004.
http://doi.org/10.1142/S0217595920400047

A Model of Multistage Risk-Averse Stochastic Optimization and its Solution by Scenario-Based Decomposition Algorithms

This paper is dedicated to Professor Minyi Yue, a founder of operations research in China, in celebration of his 100th birthday.

Min Zhang

State Key Laboratory of Desert and Oasis Ecology,
Xinjiang Institute of Ecology and Geography,
Chinese Academy of Sciences, Urumqi 830011, China and
University of Chinese Academy of Sciences,
Beijing 100049, China.
zhangmin1206@ms.xjb.ac.cn

Liangshao Hou

Department of Mathematics, Hong Kong Baptist University, Hong Kong.
16484096@life.hkbu.edu.hk

Jie Sun

School of EECMS, Curtin University, Australia and
Institute of Mathematics, Hebei University of Technology, China.
jie.sun@curtin.edu.au

Ailing Yan*

Institute of Mathematics, Hebei University of Technology, China.
ailing-yan78@163.com

Stochastic optimization models based on risk-averse measures are of essential importance in financial management and business operations. This paper studies new algorithms for a popular class of these models, namely the mean-deviation models in multistage decision making under uncertainty. It is argued that this type of problems enjoys a scenario-decomposable structure, which could be utilized in an efficient progressive hedging procedure. In case that linkage constraints arise in reformulations of the original problem, a Lagrange progressive hedging algorithm could be utilized to solve the reformulated problem. Convergence results of the algorithms are obtained based on the recent development of the Lagrangian form of stochastic variational inequalities. Numerical results are provided to show the effectiveness of the proposed algorithms.

Keywords: Progressive hedging algorithm; risk-aversion; stochastic optimization.

*Corresponding author.

1. Introduction

Multistage stochastic optimization models aim at determining optimal responses to information as it becomes available over a finite horizon of N stages. Consider a *finite* set Ξ of scenarios $\xi = (\xi_1, \ldots, \xi_N) \in \mathbb{R}^{m_1} \times \cdots \times \mathbb{R}^{m_N}$, where ξ_k is revealed at the end of stage k, after a decision $x_k(\xi_1, ..., \xi_{k-1})$ is made but before the decision $x_{k+1}(\xi_1, ..., \xi_k)$ is made. Each scenario ξ has a known probability $p(\xi) > 0$, and these probabilities add up to one. In this way Ξ is a probability space. Our attention is directed to such decision-mappings that designate responses to the scenarios in Ξ, i.e.,

$$x(\cdot) : \xi \mapsto x(\xi) = \big(x_1(\xi), \ldots, x_N(\xi)\big) \in \mathbb{R}^{n_1} \times \cdots \times \mathbb{R}^{n_N} \triangleq \mathbb{R}^n,$$

where $x_k(\xi)$ is the decision made at stage $k = 1, \ldots, N$.

The linear space $\mathscr{H}_n$ consisting of all such mappings $x(\cdot)$ from Ξ to $\mathbb{R}^n$ is a finite-dimensional Hilbert space equipped with the expectational inner product

$$\langle x(\cdot), w(\cdot)\rangle = \mathbb{E}\Big(x(\xi)^T w(\xi)\Big) = \sum_{\xi\in\Xi} p(\xi) \sum_{k=1}^{N} x_k(\xi)^T w_k(\xi), \tag{1}$$

where T means the transpose.

The objective function of a multistage stochastic optimization is usually taken as $\mathbb{E}\big(f(x(\xi), \xi)\big)$, where $\mathbb{E}$ stands for the expectation and $f(x(\xi), \xi)$ is the total cost of N stages under scenario ξ. This objective uses the expectation as a risk measure, which is often criticized for being risk-neutral because it treats the cost (the cases of $f(x(\xi), \xi) > 0$) and the gain (the cases of $f(x(\xi), \xi) < 0$) indifferently. A more reasonable approach, which is a focal point of recent research, is to use a more flexible measure $\mathcal{R}\big(f(x(\xi), \xi)\big)$, which is risk-averse (defined later), as the objective function in the multistage stochastic optimization models.

A popular risk-averse measure in finance and business is the mean-deviation function

$$\mathcal{R}_{\text{md}}(\eta) = \mathbb{E}(\eta) + \beta\sigma(\eta),$$

where η is a random variable, $\beta > 0$ is a constant, and σ stands for the standard deviation. This function takes standard deviation as a safeguard of the decision. However, for practical applications, this safeguard may not be good enough. For instance, in portfolio optimization, Markowitz (1959) noted the shortcoming of using variance and proposed using the semivariance of a portfolio to control risk. Bawa (1975), Bawa and Lindenberg (1977) and Fishburn (1977) introduced a class of downside risk measure known as the lower partial moment (LPM) to better suit different risk profiles of the investors. Since the LPM can be used to control the loss of portfolio, it has become a popular risk measure (see for instances Chen et al. (2011), Grootveld and Hallerbach (1999), Harlow (1991), Ling et al. (2019) and Yu et al. (2006)). Let t be a constant. The LPM of random variable η with respect to t is defined as

$$\text{LPM}_m(\eta, t) = \mathbb{E}(\eta - t)_+^m, \text{ where } m \geq 0 \text{ and } (\eta - t)_+ = \max\{0, \eta - t\}.$$

In particular,

$$\mathrm{LPM}_0(\eta, t) = \mathbb{P}(\eta \geq t) \text{ (Probability of } \eta \geq t),$$

and $\mathrm{LPM}_1(\eta, t)$ is the expected shortfall of η falling beyond t.

The corresponding risk measure to LPM (MLPM for short) can be expressed as

$$\mathcal{R}_{\mathrm{m}}(\eta, t) = \mathbb{E}(\eta) + \beta \mathrm{LPM}_m(\eta, t). \tag{2}$$

Let us specify the meaning of "averse risk measure" formally.

Definition 1.1. Let η be a random variable. A risk measure $\mathcal{R}(\eta)$ is called averse if it is a closed convex function of η and satisfies

$$\mathcal{R}(\eta) > \mathbb{E}(\eta) \text{ for all non-constant } \eta.$$

It can be verified that both $\mathcal{R}_{\mathrm{md}}$ and $\mathcal{R}_{\mathrm{m}}$ $(m \geq 1)$ are averse risk measures. This paper is concerned with a computational scheme that can solve a constrained optimization problem which uses $\mathcal{R}_{\mathrm{m}}$ as the objective function.

In addition to this objective function, we assume that every admissible decision $x(\cdot)$ must satisfy a set of constraints and the constraints generally depend on ξ that imposes a constraint on $x(\cdot)$. We write this fact in the form of

$$x(\cdot) \in \mathcal{C} \subset \mathscr{H}_n, \text{ which means } x(\xi) \in C(\xi) \ \forall\ \xi \in \Xi. \tag{3}$$

It is easy to see that $\mathcal{C}$ is convex and closed if $C(\xi)$ is convex and closed for all $\xi \in \Xi$.

In addition to (3), an important constraint to a multi-stage stochastic optimization problem is that the mappings $x(\cdot)$ must be *nonanticipative* in the sense that the response $x_k(\xi)$ at stage k depends only on the portion $(\xi_1, \ldots, \xi_{k-1})$ of the scenario ξ realized in earlier stages, i.e.

$$x(\xi) = \big(x_1, x_2(\xi_1), x_3(\xi_1, \xi_2), \ldots, x_N(\xi_1, \xi_2, \ldots, \xi_{N-1})\big).$$

This format of $x(\cdot)$ imposes that $x(\cdot)$ belongs to the so-called *nonanticipativity subspace* $\mathcal{N}$ of $\mathscr{H}_n$, where

$$\mathcal{N} = \{x(\cdot) : x_k(\xi_1, \ldots, \xi_N) \text{ does not depend on } \xi_k, \ldots, \xi_N \ \forall\ k\}. \tag{4}$$

The complementary linear subspace of $\mathcal{N}$ with respect to the expectational inner product (1) is denoted by $\mathcal{M} = \mathcal{N}^{\perp}$, which will contain the dual sequence generated by the algorithms we are going to introduce.

In summary, the multistage MLPM minimization problem accounts for finding an optimal response function $x(\cdot) \in \mathscr{H}_n$ for the following problem

$$\begin{aligned} \min \ & \mathcal{R}_{\mathrm{m}}\big(f(x(\xi), \xi), t\big) \\ \text{s.t.} \ & x(\cdot) \in \mathcal{C} \cap \mathcal{N}, \end{aligned} \tag{5}$$

in which for every ξ, $f(x(\xi), \xi)$ is continuously differentiable and convex in $x(\xi)$ and $C(\xi)$ is nonempty, closed and convex.

Note that the objective function of Problem (5) is

$$H\big(x(\cdot)\big) \triangleq \mathcal{R}_{\mathrm{m}}\big(f(x(\xi),\xi),t\big) = \mathbb{E}\big(f(x(\xi),\xi)\big) + \beta\mathbb{E}\big[f(x(\xi),\xi) - t\big]_+^m.$$

The convexity of $H\big(x(\cdot)\big)$ depends on the definition of t. If t is given as an independent parameter, then $H\big(x(\cdot)\big)$ is convex in $x(\cdot)$. If t is a function of $x(\cdot)$, say $t = \mathbb{E}\big(f(x(\xi),\xi)\big)$, then $H\big(x(\cdot)\big)$ may not be convex in $x(\cdot)$ when $f\big(x(\xi),\xi\big)$ is not affine in $x(\xi)$.

We shall develop a practical numerical method for solving Problem (5) for either fixed t, or $t = \mathbb{E}\big(f(x(\xi),\xi)\big)$. There are several difficult points in developing such a method; i.e.,

- Even if there is an analytic definition for each $C(\xi)$ via equality-inequality systems, there is no analytic representation for set $\mathcal{C}$.
- There is generally no matrix representation for subspace $\mathcal{N}$.
- The LPM term in the objective function brings in a linkage constraint that disrupts the required separability in ξ (i.e. $C(\xi)$ depends on a single realization of ξ, not all realizations of ξ), which creates an additional obstacle for decomposition-based algorithms.

We propose to use a revision of the scenario aggregation (now called progressive hedging) idea of Rockafellar and Wets (1991) for solving Problem (5). To get around the linkage constraint, a scenario based decomposition algorithm called the Lagrangian progressive hedging algorithm (LPHA for short) is designed for the Lagrangian form of (5). The advantage of the LPHA is its decomposability: At each iteration, one first solves K subproblems of dimension n, ignoring the nonanticipativity, where K is the cardinality of the set Ξ. The second step of each iteration is to restore nonanticipativity, by projecting the primal solution obtained in the first step onto the space $\mathcal{N}$ and updating the dual solution obtained in last iteration in the space $\mathcal{M}$ — this is called the projection step. Overall, by taking advantage of the scenario-decomposition approach, we reduce the amount of computation significantly because it solves K problems of size n, rather than a single problem of size nK in each iteration.

The paper is organized as follows. In next section, we briefly introduce the original progressive hedging algorithm (PHA for short). Then in Section 3 we study how to reformulate Problem (5) into a suitable format for PHA. In particular, we introduce the saddle point format of (5) and the LPHA for the saddle point format of (5). We also discuss convergence properties of the LPHA. A numerical example will be presented in Section 4, which demonstrates the effectiveness of the LPHA. The paper is concluded in Section 5.

2. The original PHA in a nutshell

The PHA for stochastic variational inequalities, studied in Rockafellar and Sun (2019), aims at solving a stochastic variational inequality (SVI) as follows.

$$\text{Find } x(\cdot) \in \mathscr{H}_n \text{ such that } -\mathcal{F}(x(\cdot)) \in N_{\mathcal{C}\cap\mathcal{N}}(x(\cdot)), \tag{6}$$

where $\mathcal{C}$ and $\mathcal{N}$ are defined as (3) and (4), respectively, $N_{\mathcal{C}\cap\mathcal{N}}\big(x(\cdot)\big)$ stands for the normal cone of set $\mathcal{C}\cap\mathcal{N}$ at $x(\cdot)$ in the sense of convex analysis (Rockafellar, 1970), and $\mathcal{F} : \mathscr{H}_n \rightrightarrows \mathscr{H}_n$ defined as

$$\mathcal{F}(x(\cdot)) :\ \xi \ \mapsto\ F(x(\xi), \xi), \tag{7}$$

where $F(\cdot, \xi) : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ is a given mapping $\forall \xi$. Under the following constraint qualification (CQ for short) condition

$$\text{Either } \mathcal{C} \text{ is a convex polyhedron and } \mathcal{C}\cap\mathcal{N} \neq \emptyset \text{ or } \mathrm{ri}\,\mathcal{C}\cap\mathcal{N} \neq \emptyset, \tag{8}$$

one has

$$N_{\mathcal{C}\cap\mathcal{N}}\big(x(\cdot)\big) = N_{\mathcal{C}}\big(x(\cdot)\big) + N_{\mathcal{N}}\big(x(\cdot)\big) = N_{\mathcal{C}}\big(x(\cdot)\big) + \mathcal{M}.$$

Therefore, together with the decomposable structure (7) of mapping $\mathcal{F}$ and the definition (3) of $\mathcal{C}$, Problem (6) is equivalent to

$$\text{Find } x(\cdot) \in \mathcal{N}, w(\cdot) \in \mathcal{M} \text{ such that } -F\big(x(\xi), \xi\big) - w(\xi) \in N_{C(\xi)}\big(x(\xi)\big)\ \forall \xi. \tag{9}$$

The PHA for solving Problem (9) consists of the following iteration steps.

Algorithm 1. The PHA for Problem (9)

Given $x^\nu(\cdot) \in \mathcal{N}$, $w^\nu(\cdot) \in \mathcal{M}$, and $r > 0$,

Step 1. Obtain $\hat{x}(\xi)$ for every ξ via solving

$$-F\big(\hat{x}(\xi), \xi\big) - w^\nu(\xi) - r|\hat{x}(\xi) - x^\nu(\xi)| \in N_{C(\xi)}\big(\hat{x}(\xi)\big);$$

Step 2. Update

$$x^{\nu+1}(\cdot) = P_{\mathcal{N}}\big(\hat{x}(\cdot)\big),\ w^{\nu+1}(\cdot) = w^\nu(\cdot) + rP_{\mathcal{M}}\big(\hat{x}(\cdot)\big);$$

$\nu := \nu + 1$, **repeat**.

$P_{\mathcal{N}}\big(\hat{x}(\cdot)\big)$ and $P_{\mathcal{M}}\big(\hat{x}(\cdot)\big)$ in Step 2 are the projections of $\hat{x}(\cdot)$ onto subspaces $\mathcal{N}$ and $\mathcal{M}$, respectively. It should be noted that the calculation of $P_{\mathcal{N}}\big(\hat{x}(\cdot)\big)$ is rather simple — it accounts to computation of certain conditional expectations of $\hat{x}(\xi)$. Specifically, in two-stage problems, this projection is simply to calculate the usual expectation of $\hat{x}(\xi)$, i.e., $x^{\nu+1}(\xi) = \mathbb{E}\big(\hat{x}(\xi)\big)\ \forall \xi$. For more details, see Rockafellar and Wets (1991). Besides, $P_{\mathcal{M}}\big(\hat{x}(\cdot)\big)$ can be easily obtained by

$$P_{\mathcal{M}}\big(\hat{x}(\cdot)\big) = \hat{x}(\cdot) - P_{\mathcal{N}}\big(\hat{x}(\cdot)\big) = \hat{x}(\cdot) - x^{\nu+1}(\cdot).$$

We summarize the convergence properties of Algorithm 1 in the theorem below.

Theorem 2.1 (Theorem 2 of Rockafellar and Sun (2019)). *Suppose that (i) $F(x(\xi),\xi)$ is monotone for all ξ, (ii) CQ (8) is satisfied, and (iii) Problem (6) has a solution. Then the sequence generated by Algorithm 1 converges to a primal-dual solution pair of Problem (9), $\big(x^*(\cdot),w^*(\cdot)\big)$, namely*

$$\big(x^\nu(\cdot),w^\nu(\cdot)\big)\to\big(x^*(\cdot),w^*(\cdot)\big)\ \text{as}\ \nu\to+\infty,$$

and the primal part $x^*(\cdot)$ is a solution of Problem (6).

Moreover, if $F(x(\xi),\xi)$ is affine in $x(\xi)$ and $C(\xi)$ is polyhedral for all ξ, then the sequence $\{\big(x^\nu(\cdot),w^\nu(\cdot)\big)\}$ converges at q-linear rate with respect to the r-norm defined as

$$\|\big(x(\cdot)-x^*(\cdot),w(\cdot)-w^*(\cdot)\big)\|_r=\sqrt{\|x(\cdot)-x^*(\cdot)\|^2+\frac{1}{r^2}\|w(\cdot)-w^*(\cdot)\|^2}.$$

Based on the theory of variational inequalities, under CQ (8), a necessary condition for $x(\cdot)$ to be an optimal solution to Problem (5) is

$$-\partial H\big(x(\cdot)\big)\in N_{\mathcal{C}\cap\mathcal{N}}\big(x(\cdot)\big),\tag{10}$$

where ∂H is the subdifferential mapping of H, and condition (10) will be also sufficient if Problem (5) is convex. Clearly, when t is a fixed parameter, Problem (5) is equivalent to Problem (6) with $\mathcal{F}\big(x(\cdot)\big)=\partial H\big(x(\cdot)\big)$, and $\mathcal{F}$ has the decomposable structure in this case, thus under CQ (8) one can apply Algorithm 1 to obtain a solution to Problem (5). However, a complication arises when $t=\mathbb{E}\big(f(x(\xi),\xi)\big)$ and f is not affine in $x(\xi)$, which makes MLPM minimization problem (5) to be generally nonconvex and hence $\partial H\big(x(\cdot)\big)$ to be non-monotone. In the next section, we build a convex relaxation of Problem (5) in this case and extend the PHA to a Lagrange form for handling non-decomposability of the relaxed problem.

3. The LPHA for MLPM minimization

3.1. *Conversion of Problem (5) to a decomposable form of (9)*

Case 1: t is an independent parameter in Problem (5).

Under the assumption that $f(x(\xi),\xi)$ is convex and continuously differentiable in $x(\xi)$ for every ξ, one can calculate the subdifferential of H, $\partial H\big(x(\cdot)\big):\mathscr{H}_n\rightrightarrows\mathscr{H}_n$, via $\partial H\big(x(\cdot)\big):\xi\mapsto\partial h(x(\xi),\xi)$, where $\partial h(x(\xi),\xi)$ is defined as follows.

- When $m=1$,

$$\partial h(x(\xi),\xi)=\begin{cases}\nabla f(x(\xi),\xi)+\beta\nabla f(x(\xi),\xi), & \text{if } f(x(\xi),\xi)>t,\\ \nabla f(x(\xi),\xi)+\lambda\beta\nabla f(x(\xi),\xi),\lambda\in[0,1], & \text{if } f(x(\xi),\xi)=t,\\ \nabla f(x(\xi),\xi), & \text{if } f(x(\xi),\xi)<t.\end{cases}$$

- When $m>1$, $\partial h(x(\xi),\xi)=\nabla f(x(\xi),\xi)+\beta m(f(x(\xi),\xi)-t)_+^{m-1}\nabla f(x(\xi),\xi)$.

Since Problem (5) is a convex program in this case, thus, under constraint qualification, Problem (5) can be solved by Algorithm 1 with $F(x(\xi),\xi)=\partial h(x(\xi),\xi)$.

Case 2: $t = \mathbb{E}\big(f(x(\xi), \xi)\big)$ **in Problem (5).**

By introducing an auxiliary variable $s(\cdot)$, Problem (5) can be converted to

$$\begin{aligned} \min_{x(\cdot), s(\cdot)} \quad & H_1\big(x(\cdot), s(\cdot)\big) \triangleq \mathbb{E}\big(f(x(\xi), \xi) + \beta s(\xi)^m\big) \\ \text{s.t.} \quad & \big(x(\cdot), s(\cdot)\big) \in \mathcal{C}_1 \cap \mathcal{N}_1, \end{aligned} \tag{11}$$

where

$$\mathcal{C}_1 = \left\{ \big(x(\cdot), s(\cdot)\big) \in \mathscr{H}_n \times \mathscr{H}_1 : \begin{array}{l} x(\xi) \in C(\xi),\ s(\xi) \geq 0, \\ \mathbb{E}\big(f(x(\xi), \xi)\big) \geq f(x(\xi), \xi) - s(\xi), \end{array} \forall \xi \right\}$$

and

$$\begin{aligned} \mathcal{N}_1 &= \text{The nonanticipative subspace of } \big(x(\cdot), s(\cdot)\big) \\ &= \big\{ \big(x(\cdot), s(\cdot)\big) \in \mathscr{H}_n \times \mathscr{H}_1 : x(\cdot) \in \mathcal{N}, s(\cdot) \text{ free} \big\}. \end{aligned}$$

Note that $\partial H_1\big(x(\cdot), s(\cdot)\big)$ has decomposable structure like (7), however the constraints in $\mathcal{C}_1$ are not decomposable in ξ due to the existence of $\mathbb{E}\big(f(x(\xi), \xi)\big)$. To overcome this obstacle, we introduce a single auxiliary variable y and consider a convex relaxation of (11) as follows.

$$\begin{aligned} \min_{y, x(\cdot), s(\cdot)} \quad & \mathbb{E}\big(y + \beta s(\xi)^m\big) \\ \text{s.t.} \quad & x(\xi) \in \mathcal{C}(\xi),\ s(\xi) \geq 0,\ f(x(\xi), \xi) - s(\xi) - y \leq 0,\ \forall \xi \\ & \mathbb{E}\big(f(x(\xi), \xi)\big) \leq y,\ \big(y, x(\cdot), s(\cdot)\big) \in \mathbb{R} \times \mathcal{N}_1. \end{aligned} \tag{12}$$

Proposition 3.1. *Problem (12) is a convex relaxation of Problem (11), which means that the optimal value of Problem (12) is a lower bound of the optimal value of Problem (11). Moreover, if $\big(y^*, x^*(\cdot), s^*(\cdot)\big)$ is a solution to Problem (12) and $y^* = \mathbb{E}\big(f(x^*(\xi), \xi)\big)$, then $\big(x^*(\cdot), s^*(\cdot)\big)$ is a solution to Problem (11).*

Furthermore, if $f(x(\xi), \xi)$ is affine in $x(\xi)$ for all ξ, then Problem (11) is equivalent to a convex program that is obtained by replacing the inequality $\mathbb{E}(f(x(\xi), \xi)) \leq y$ in (12) with the equality $\mathbb{E}(f(x(\xi), \xi)) = y$.

Proof. Since Problem (11) can be equivalently reformulated as

$$\begin{aligned} \min_{y, x(\cdot), s(\cdot)} \quad & \mathbb{E}\big(y + \beta s(\xi)^m\big) \\ \text{s.t.} \quad & x(\xi) \in C(\xi),\ s(\xi) \geq 0,\ \forall \xi \\ & f(x(\xi), \xi) - s(\xi) \leq \mathbb{E}(f(x(\xi), \xi)) \leq y,\ \forall \xi \\ & \big(y, x(\cdot), s(\cdot)\big) \in \mathbb{R} \times \mathcal{N}_1, \end{aligned} \tag{13}$$

a solution $\big(x^*(\cdot), s^*(\cdot)\big)$ to Problem (11) together with $y^* = \mathbb{E}\big(f(x^*(\xi), \xi)\big)$ serves a solution to Problem (13). Note that the objectives of Problem (12) and (13) are the same, and the feasible set of (12) is larger than the feasible set of (13). Thus, Problem (12), which is a convex problem, can be regarded as a convex relaxation of Problem (11), with the optimal value of Problem (12) being not greater than the optimal value of Problem (11). Moreover, if $\big(y^*, x^*(\cdot), s^*(\cdot)\big)$ is a solution to

Problem (12) and $y^* = \mathbb{E}\big(f(x^*(\xi), \xi)\big)$, then it is a feasible point of Problem (13). Since any feasible point $\big(y, x(\cdot), s(\cdot)\big)$ of Problem (13) is feasible to Problem (12), the following holds by the fact that $\big(y^*, x^*(\cdot), s^*(\cdot)\big)$ is a solution to (12):

$$\mathbb{E}\big(y^* + \beta s^*(\xi)^m\big) \le \mathbb{E}\big(y + \beta s(\xi)^m\big) \ \forall \text{ feasible } \big(y, x(\cdot), s(\cdot)\big) \text{ of problem (13)},$$

which infers that $\big(y^*, x^*(\cdot), s^*(\cdot)\big)$ is also a feasible solution to Problem (13) and is therefore a feasible solution to Problem (11).

If $f(x(\xi), \xi)$ is affine in $x(\xi)$ for all ξ, then equality $\mathbb{E}(f(x(\xi), \xi)) = y$ is a convex constraint, therefore, replacing the inequality $\mathbb{E}(f(x(\xi), \xi)) \le y$ in (12) with the equality $\mathbb{E}(f(x(\xi), \xi)) = y$ results in a convex program. It is easy to see that this convex program is equivalent to Problem (11). The proof is complete. $\square$

Remark 3.1. So far most applications we encounter have $f(x(\xi), \xi)$ to be affine in $x(\xi)$. Therefore the second part of Proposition 3.1 applies and the relaxation is exact. Otherwise, the error of the relaxation can be estimated by certain probability inequalities (e.g., Chebyshev inequality and Petrov exponential inequality), based on given moment information of ξ and other assumptions. This is an on-going work in robust optimization and is better left for a separate study.

It is often more convenient for computational purpose to replace the single variable y by a functional $y(\cdot) \in \mathscr{H}$ with an additional constraint $y(\xi) = y$ for all $\xi \in \Xi$, i.e., $y(\xi)$ independent with ξ and can be treated as an additional part of the first-stage variable. In this way, we gain some theoretical advantage at a minor computational cost. Denote

$$z(\cdot) = \big(y(\cdot), x(\cdot), s(\cdot)\big) \in \mathscr{H}_{n+2},$$

then problem (12) can be re-written as the following problem

$$\begin{aligned} \min_{z(\cdot)} \quad & H_2\big(z(\cdot)\big) = \mathbb{E}\big(h(z(\xi), \xi)\big) \\ \text{s.t.} \quad & z(\cdot) \in \mathcal{C}_2 \cap \mathcal{N}_2 \cap \mathcal{S}, \end{aligned} \tag{14}$$

where $h(z(\xi), \xi) := y(\xi) + \beta s(\xi)^m$, and

$$\mathcal{N}_2 = \{z(\cdot) \in \mathscr{H}_{n+2} \mid y(\xi) = y \ \forall \ \xi, x(\cdot) \in \mathcal{N}, \ s(\cdot) \text{ free}\}.$$

Note that the objective function is the expectation of $h(z(\xi), \xi)$, which infers that H_2 is continuously differentiable and its gradient has the decomposable structure in ξ as (7) with $F = \nabla h$. Besides, the first part of constraints is also decomposable in ξ, namely, $z(\xi) \in C_2(\xi) \ \forall \ \xi$, where

$$\mathcal{C}_2 = \{z(\cdot) \in \mathscr{H}_{n+2} : x(\xi) \in C(\xi), \ s(\xi) \ge 0, \ f(x(\xi), \xi) - s(\xi) - y(\xi) \le 0, \forall \xi\},$$

the second part of constraints is the nonanticipativity constraint $z(\cdot) \in \mathcal{N}_2$ and the last single constraint

$$z(\cdot) \in \mathcal{S} = \{z(\cdot) \in \mathscr{H}_{n+2} : \mathbb{E}\big(f(x(\xi), \xi) - y(\xi)\big) \le 0\}$$

is the so-called "linkage constraint" because it links all ξ together, or "cross constraint" as we named it in Sun et al (2019).

It is easy to show that, Problem (14) is a convex program. Thus, the sufficient and necessary condition for optimality of $z(\cdot)$ is

$$-\nabla H_2\big(z(\cdot)\big) \in N_{\mathcal{C}_2\cap\mathcal{N}_2\cap\mathcal{S}}\big(z(\cdot)\big), \tag{15}$$

which is similar to Problem (6) except the additional linkage constraint $\mathcal{S}$. Note that H_2 is differentiable, so we use ∇H_2 instead of ∂H_2.

3.2. *The Lagrangian PHA for MLPM minimization*

Now we focus on Problem (14). The idea to remove the linkage constraint is based on the Lagrangian function. First, we denote the Lagrangian function of Problem (14) by $L(z(\cdot),\lambda) : \mathscr{H}_{n+2}\times\mathbb{R}\to[-\infty,+\infty]$ and there hold

- if $z(\cdot)\in\mathcal{C}_2\cap\mathcal{N}_2,\ \lambda\geq 0$, then
$$L(z(\cdot),\lambda) = H_2(z(\cdot)) + \lambda\Big(\mathbb{E}\big(f(x(\xi),\xi)-y(\xi)\big)\Big) = \mathbb{E}\Big(\ell(z(\xi),\lambda)\Big),$$
where $\ell(z(\xi),\lambda) = h(z(\xi),\xi) + \lambda(f(x(\xi),\xi)-y(\xi))$;
- if $z(\cdot)\notin\mathcal{C}_2\cap\mathcal{N}_2,\ \lambda\geq 0$, then $L(z(\cdot),\lambda) = +\infty$;
- if $z(\cdot)\in\mathcal{C}_2\cap\mathcal{N}_2,\ \lambda< 0$, then $L(z(\cdot),\lambda) = -\infty$.

Due to the convexity of Problem (14), under the following CQ that

$$\operatorname{ri}\mathcal{C}_2\cap\mathcal{N}_2\cap\operatorname{ri}\mathcal{S}\neq\emptyset, \tag{16}$$

$z^*(\cdot)$ is the optimal solution of Problem (14) iff there exists $\lambda^*\in\mathbb{R}$ such that $(z^*(\cdot),\lambda^*)$ is a saddle point of the Lagrangian function $L(z(\cdot),\lambda)$ with respect to minimizing over $z(\cdot)\in\mathscr{H}_{n+2}$ and maximizing over $\lambda\in\mathbb{R}$, i.e.,

$$(z^*(\cdot),\lambda^*)\in\arg\left\{\min_{z(\cdot)\in\mathcal{C}_2\cap\mathcal{N}_2}\ \max_{\lambda\geq 0}\ L(z(\cdot),\lambda)\right\}, \tag{17}$$

which leads to

$$\begin{pmatrix} -\nabla_z L(z^*(\cdot),\lambda^*) \\ \nabla_\lambda L(z^*(\cdot),\lambda^*) \end{pmatrix} \in N_{(\mathcal{C}_2\cap\mathcal{N}_2)\times\mathbb{R}_+}(z^*(\cdot),\lambda^*). \tag{18}$$

In fact, $\nabla_z L(z^*(\cdot),\lambda^*)\in\mathscr{H}_{n+2}$ with

$$\nabla_z L(z^*(\cdot),\lambda^*) : \xi\mapsto\begin{pmatrix} 1-\lambda^* \\ \lambda^*\nabla f(x^*(\xi),\xi)) \\ \beta m s^*(\xi)^{m-1}\end{pmatrix},$$

and $\nabla_\lambda L(z^*(\cdot),\lambda^*)\in\mathbb{R}$ with

$$\nabla_\lambda L(z^*(\cdot),\lambda^*) = \mathbb{E}\big(f(x(\xi),\xi)-y(\xi)\big).$$

It can be seen that $\nabla_z L(z^*(\cdot),\lambda^*)$ has the decomposable structure as (7) while $\nabla_\lambda L(z^*(\cdot),\lambda^*)$ is not decomposable. In order to handle the nondecomposable part, we use the similar technique for y to introduce a variable $u(\cdot) \in \mathscr{H}$ and let $u(\cdot) \in \mathcal{C}_3 \cap \mathcal{N}_3$ with

$$\mathcal{C}_3 := \{u(\cdot) \in \mathscr{H} : u(\xi) \in \mathbb{R}_+ \ \forall\ \xi\},\ \mathcal{N}_3 := \{u(\cdot) \in \mathscr{H} : u(\xi) = \lambda \ \forall\ \xi\}.$$

Let $\bar{L}(z(\cdot), u(\cdot)) : (\mathcal{C}_2 \times \mathcal{C}_3) \cap (\mathcal{N}_2 \times \mathcal{N}_3) \to \mathbb{R}$ be

$$\bar{L}(z(\cdot), u(\cdot)) := \mathbb{E}\Big(\ell(z(\xi), u(\xi))\Big).$$

Then $(z^*(\cdot), \lambda^*)$ being a saddle point of $L(z(\cdot), \lambda)$ is equivalent to $(z^*(\cdot), u^*(\cdot))$ being a saddle point of $\bar{L}(z(\cdot), u(\cdot))$ with $u^*(\xi) = \lambda^* \ \forall\ \xi$, which further implies

$$-\begin{pmatrix} \nabla_z \bar{L}(z^*(\cdot), u^*(\cdot)) \\ -\nabla_u \bar{L}(z^*(\cdot), u^*(\cdot)) \end{pmatrix} \in N_{(\mathcal{C}_2\times\mathcal{C}_3)\cap(\mathcal{N}_2\times\mathcal{N}_3)}(z^*(\cdot), u^*(\cdot)). \tag{19}$$

Here $\nabla_z \bar{L}(z^*(\cdot), u^*(\cdot)) \in \mathscr{H}_{n+2}$ with decomposable structure:

$$\nabla_z \bar{L}(z^*(\cdot), u^*(\cdot)) : \xi \mapsto \begin{pmatrix} 1 - u^*(\xi) \\ u^*(\xi)\nabla f(x^*(\xi), \xi)) \\ \beta m s^*(\xi)^{m-1} \end{pmatrix},$$

and $\nabla_u \bar{L}(z^*(\cdot), u^*(\cdot)) \in \mathscr{H}$ has decomposable structure as well with

$$\nabla_u \bar{L}(z^*(\cdot), u^*(\cdot)) : \xi \mapsto f(x^*(\xi), \xi) - y^*(\xi).$$

So far, we come up with problem (19) in format (6) with

$$\mathcal{F} = \begin{pmatrix} \nabla_z \bar{L} \\ -\nabla_u \bar{L} \end{pmatrix},\ \mathcal{C} = \mathcal{C}_2 \times \mathcal{C}_3 \text{ and } \mathcal{N} = \mathcal{N}_2 \times \mathcal{N}_3,$$

therefore, under CQ that $\mathrm{ri}(\mathcal{C}_2 \times \mathcal{C}_3) \cap (\mathcal{N}_2 \times \mathcal{N}_3) \neq \emptyset$, (19) is equivalent to find $(z^*(\cdot), u^*(\cdot)) \in \mathcal{N}_2 \times \mathcal{N}_3$ and $(w^*(\cdot), v^*(\cdot)) \in \mathcal{M}_2 \times \mathcal{M}_3$ such that

$$-\begin{pmatrix} \nabla_z \bar{L}(z^*(\xi), u^*(\xi)) \\ -\nabla_u \bar{L}(z^*(\xi), u^*(\xi)) \end{pmatrix} - \begin{pmatrix} w^*(\xi) \\ v^*(\xi) \end{pmatrix} \in N_{\mathcal{C}_2(\xi)\times\mathcal{C}_3(\xi)}(z^*(\xi), u^*(\xi)),\ \forall\ \xi,$$

where $\mathcal{M}_i$ are the complementary subspace of $\mathcal{N}_i$ for $i = 2, 3$, then one can apply Algorithm 1 for the above equivalent problem and obtain a saddle point of $\bar{L}(z(\cdot), u(\cdot))$; furthermore, this saddle point can provide a solution to Problem (14).

More concretely, when applying Algorithm 1, it starts from $\big(z^\nu(\cdot), u^\nu(\cdot)\big) \in \mathcal{N}_2 \times \mathcal{N}_3$ and $\big(w^\nu(\cdot), v^\nu(\cdot)\big) \in \mathcal{M}_2 \times \mathcal{M}_3$. In Step 1, one obtains $(\hat{z}^\nu(\xi), \hat{u}^\nu(\xi))$ through the following variational inequality for every ξ,

$$-\begin{pmatrix} \nabla_z \bar{L}(z(\xi), u(\xi)) \\ -\nabla_u \bar{L}(z(\xi), u(\xi)) \end{pmatrix} - \begin{pmatrix} w^\nu(\xi) \\ v^\nu(\xi) \end{pmatrix} - r\begin{pmatrix} z(\xi) - z^\nu(\xi) \\ u(\xi) - u^\nu(\xi) \end{pmatrix} \in N_{\mathcal{C}_2(\xi)\times\mathcal{C}_3(\xi)}(z(\xi), u(\xi)),$$

which is actually the saddle point condition of the following small-size min-max problem for every ξ:

$$\min_{z\in C_2(\xi)} \max_{u\in\mathbb{R}_+} \phi_r\big(z, u; w^\nu(\xi), v^\nu(\xi), \xi\big), \tag{20}$$

where

$$\phi_r(z,u;w,v,\xi) = \ell(z,u) + z^T w - uv + \frac{r}{2}\|z - z^\nu(\xi)\|^2 - \frac{r}{2}(u - u^\nu(\xi))^2.$$

In fact, denote $g(z(\xi),\xi) = f(x(\xi),\xi) - y(\xi)$, then problem (20) is equivalent to the following problem

$$\min_{z\in C_2(\xi)} \left\{ h(z,\xi) + z^T w^\nu(\xi) + \frac{r}{2}\|z - z^\nu(\xi)\|^2 + \max_{u\in\mathbb{R}_+}\{ug(z,\xi) - uv^\nu(\xi) - \frac{r}{2}(u - u^\nu(\xi))^2\} \right\}. \tag{21}$$

Since the maximizer of problem $\max\limits_{u\in\mathbb{R}_+}\{ug(z,\xi) - uv^\nu(\xi) - \frac{r}{2}(u-u^\nu(\xi))^2\}$ is

$$\hat{u}^\nu(\xi) := \max\{0, u^\nu(\xi) + r^{-1}(g(z,\xi) - v^\nu(\xi))\},$$

we have (21) equal to

$$\min_{z\in C_2(\xi)} \quad \psi_r\big(z, u^\nu(\xi), w^\nu(\xi), v^\nu(\xi), \xi\big) \tag{22}$$

where

$$\begin{aligned}\psi_r(z,u,w,v,\xi) \triangleq h(z,\xi) + z^T w + \frac{r}{2}\|z - z^\nu(\xi)\|^2 + \frac{1}{2r}\big(g(z,\xi) - v\big)^2 \\ -\frac{r}{2}\mathrm{dist}^2_{\mathbb{R}_+}\Big(u^\nu(\xi) + r^{-1}\big(g(z,\xi) - v\big)\Big),\end{aligned}$$

with $\mathrm{dist}^2_{\mathbb{R}_+}(a) := \min_{x\in\mathbb{R}_+}(x-a)^2$. Thus, the solution of (20), $(\hat{z}^\nu(\xi), \hat{u}^\nu(\xi))$, can be calculated in terms of the augmented Lagrangians:

$$\begin{cases} \hat{z}^\nu(\xi) = \mathrm{argmin}_{\, z\in C_2(\xi)}\ \psi_r\big(z, u^\nu(\xi), w^\nu(\xi), v^\nu(\xi), \xi\big), \\ \hat{u}^\nu(\xi) = \max\{0, u^\nu(\xi) + r^{-1}(g(\hat{z}^\nu(\xi),\xi) - v^\nu(\xi))\}. \end{cases}$$

In Step 2, the projection $(z^{\nu+1}(\cdot), u^{\nu+1}(\cdot)) = P_{\mathcal{N}_2\times\mathcal{N}_3}(\hat{z}^\nu(\cdot), \hat{u}^\nu(\cdot))$ is equivalent to

$$\begin{cases} z^{\nu+1}(\cdot) = P_{\mathcal{N}_2}(\hat{z}^\nu(\cdot)), \\ u^{\nu+1}(\xi) = P_{\mathcal{N}_3}(\hat{u}^\nu(\cdot)), \end{cases} \Leftrightarrow \begin{cases} y^{\nu+1}(\xi) = \mathbb{E}(\hat{y}^\nu(\xi))\ \forall\ \xi, \\ x^{\nu+1}(\cdot) = P_{\mathcal{N}}(\hat{x}^\nu(\cdot)), \\ s^{\nu+1}(\cdot) = \hat{s}^\nu(\cdot), \\ u^{\nu+1}(\xi) = \mathbb{E}(\hat{u}^\nu(\xi))\ \forall\ \xi, \end{cases}$$

and the dual part is

$$\begin{cases} w^{\nu+1}(\cdot) = w^\nu(\cdot) + r(\hat{z}^\nu(\cdot) - z^{\nu+1}(\cdot)), \\ v^{\nu+1}(\cdot) = v^\nu(\cdot) + r(\hat{u}^\nu(\cdot) - u^{\nu+1}(\cdot)). \end{cases}$$

To summarize, the Lagrangian progressive hedging algorithm for Problem (14) and its convergence results are presented as Algorithm 2 and Theorem 3.1.

Theorem 3.1. *Suppose Problem (14) is solvable and CQ (16) holds. Then the iteration sequence $\{(z^\nu(\cdot), u^\nu(\cdot), w^\nu(\cdot), v^\nu(\cdot))\}$ generated by Algorithm 2 for any $r > 0$, starting from any $(z^0(\cdot), u^0(\cdot)) \in \mathcal{N}_2 \times \mathcal{N}_3$ and $(w^0(\cdot), v^0(\cdot)) \in \mathcal{M}_2 \times \mathcal{M}_3$, will converge (in the weak topology of $\mathscr{H}_{n+3} \times \mathscr{H}_{n+3}$) to a solution pair of problem (14),*

$(z^*(\cdot), u^*(\cdot), w^*(\cdot), v^*(\cdot))$. *Moreover, denoting* $Z = (z(\cdot), u(\cdot))$ *and* $W = (w(\cdot), v(\cdot))$, *and the norm in* $\mathscr{H}_{n+3} \times \mathscr{H}_{n+3}$ *as*

$$\|(Z,W)\|_r^2 = \|z(\cdot)\|^2 + \|u(\cdot)\|^2 + \frac{1}{r^2}\|w(\cdot)\|^2 + \frac{1}{r^2}\|v(\cdot)\|^2,$$

we have

$$\|(Z^{\nu+1}, W^{\nu+1}) - (Z^*, W^*)\|_r \leq \|(Z^\nu, W^\nu) - (Z^*, W^*)\|_r \text{ for all } \nu.$$

Proof. The proof is similar to the one in Rockafellar and Sun (2020). We omit the details. □

Algorithm 2. LPHA for MLPM minimization

Given $(z^\nu(\cdot), u^\nu(\cdot)) \in \mathcal{N}_2 \times \mathcal{N}_3$, $(w^\nu(\cdot), v^\nu(\cdot)) \in \mathcal{M}_2 \times \mathcal{M}_3$, and $r > 0$,

Step 1. For every ξ, obtain $\hat{z}^\nu(\xi)$ and $\hat{u}^\nu(\xi)$ via

$$\hat{z}^\nu(\xi) = \arg\min_{z \in C_2(\xi)} \psi_r(z, u^\nu(\xi), w^\nu(\xi), v^\nu(\xi)),$$
$$\hat{u}^\nu(\xi) = \max\big\{0, u^\nu(\xi) + r^{-1}(g(\hat{z}^\nu(\xi), \xi) - v^\nu(\xi))\big\};$$

Step 2. Update

$$z^{\nu+1}(\cdot) = P_{\mathcal{N}_2}(\hat{z}^\nu(\cdot)),\ u^{\nu+1}(\xi) = \mathbb{E}(\hat{u}^\nu(\xi))\ \forall\ \xi,$$
$$w^{\nu+1}(\cdot) = w^\nu(\cdot) + r(\hat{z}^\nu(\cdot) - z^{\nu+1}(\cdot)),$$
$$v^{\nu+1}(\cdot) = v^\nu(\cdot) + r(\hat{u}^\nu(\cdot) - u^{\nu+1}(\cdot));$$

$\nu := \nu + 1$, **repeat**.

Particularly, when $f(x(\xi), \xi)$ is affine in $x(\xi)$ for all ξ, Problem (12) is equivalent to the following convex problem based on Proposition 3.1.

$$\begin{aligned} \min_{z(\cdot)} \quad & H_2\big(z(\cdot)\big) \\ \text{s.t.} \quad & z(\cdot) \in \mathcal{C}_2 \cap \mathcal{N}_2 \cap \mathcal{S}', \end{aligned} \tag{23}$$

where $\mathcal{S}' = \{z(\cdot) \in \mathscr{H}_{n+2} : \mathbb{E}\big(f(x(\xi), \xi) - y(\xi)\big) = 0\}$. Utilizing the saddle point concept to achieve the decomposition, we have that the solution to the following SVI is a solution to Problem (23):

$$-\begin{pmatrix} \nabla_z \bar{L}(z^*(\cdot), u^*(\cdot)) \\ -\nabla_u \bar{L}(z^*(\cdot), u^*(\cdot)) \end{pmatrix} \in N_{(\mathcal{C}_2 \times \mathscr{H}) \cap (\mathcal{N}_2 \times \mathcal{N}_3)}(z^*(\cdot), u^*(\cdot)),$$

in which the only difference from (19) is the multiplier $u^*(\xi) \in \mathbb{R}$ rather than $u^*(\xi) \in \mathbb{R}_+$ for all ξ. Therefore, following the similar derivations to Algorithm 2, we can obtain the following Algorithm 3 for finding a solution to Problem (12) in case that $f(x(\xi), \xi)$ is affine.

Algorithm 3. LPHA for MLPM minimization with affine f

Given $(z^\nu(\cdot), u^\nu(\cdot)) \in \mathcal{N}_2 \times \mathcal{N}_3$, $(w^\nu(\cdot), v^\nu(\cdot)) \in \mathcal{M}_2 \times \mathcal{M}_3$, and $r > 0$,

Step 1. For every ξ, obtain $\hat{z}^\nu(\xi)$ and $\hat{u}^\nu(\xi)$ via

$$\hat{z}^\nu(\xi) = \arg\min_{z \in C_2(\xi)} \bar{\psi}_r(z, u^\nu(\xi), w^\nu(\xi), v^\nu(\xi)),$$

$$\hat{u}^\nu(\xi) = u^\nu(\xi) + r^{-1}(g(\hat{z}^\nu(\xi), \xi) - v^\nu(\xi)),$$

where $\bar{\psi}_r(z, u, w, v) = h(z, \xi) + z^T w + \frac{r}{2}\|z - z^\nu(\xi)\|^2 + \frac{1}{2r}\big(g(z, \xi) - v\big)^2$;

Step 2. Update

$$z^{\nu+1}(\cdot) = P_{\mathcal{N}_2}(\hat{z}^\nu(\cdot)),\ \ u^{\nu+1}(\xi) = \mathbb{E}(\hat{u}^\nu(\xi))\ \forall\ \xi,$$

$$w^{\nu+1}(\cdot) = w^\nu(\cdot) + r(\hat{z}^\nu(\cdot) - z^{\nu+1}(\cdot)),$$

$$v^{\nu+1}(\cdot) = v^\nu(\cdot) + r(\hat{u}^\nu(\cdot) - u^{\nu+1}(\cdot));$$

$\nu := \nu + 1$, **repeat**.

4. A Numerical Example

In this section, we consider an optimal irrigation water allocation problem over a multi-period planning horizon. Specifically, we adopt a real application problem from Dai and Li (2013), in which a water manager is charged with supplying the water resources to 15 irrigation subareas for 3 crops: wheat, maize and cotton, with uncertainty in the total amount of the future available water, and farmers are expanding their activities and investments with the amount of water which is promised by the water manager. If the amount of water is allocated as the manager promises, the local economy will achieve net benefits, while if insufficient water is available, the users will either curtail their expansion plans or obtain water from more expensive sources such as withdrawing groundwater which will leads to losses. In this reservoir, wheat is planted in rotation with maize and cotton. In detail, wheat is planted from winter this year to summer next year, while maize and cotton are planted in summer and autumn next year. Since water demand of each crop varies with its growth stage, and the same amount of water supplied or lacked in different growth stages may result in different amount of crop yield or reduction of yield, we consider the optimal irrigation water allocation problem over four periods, i.e., period $i = 1, 2, 3, 4$ stands for winter of this year, spring, summer and autumn of next year, respectively. Let ξ_i be the random variable representing the amount of available water in period i. For each i, ξ_i satisfies a discrete distribution with detailed data shown in Table 1. Before observing the random available water ξ_1, the water manager needs to make decisions on irrigation water target and let farmers know how much water they can expect so that their activities could be arranged, which results in the first-stage decision $x_1 \in \mathbb{R}^{45}$, consisted of the irrigation water target for different crops in different subareas. Here we assume the irrigation target remains the same during all periods. In fact, it is realistic since changes may not be

Table 1. Available water under different flow levels in different seasons

Flow level	Probability	Available water ξ_i (10^3m^3)			
		Period 1	Period 2	Period 3	Period 4
1 (Very low)	0.20	0	0	0	0
2 (Low)	0.25	8548.4	8149.9	3085.1	24522.9
3 (Medium)	0.40	28257.1	21505.0	17506.0	58083.0
4 (High)	0.10	34635.9	26631.5	26315.8	102523.4
5 (Very high)	0.05	44164.1	28308.7	28540.4	172815.7

cost-effective after the water pipes being set. When ξ_1 is realized, the manager may adjust his/her strategy and decide the actual amount of water to allocate in period 1, denoted by $x_2(\xi_1)$. And after observing ξ_2 in period 2, the manager is supposed to decide the actual amount of water to allocate in period 2, i.e., $x_3(\xi_1, \xi_2)$. Therefore, the whole decision process is as follows:

$$x_1, \xi_1, x_2(\xi_1), \xi_2, x_3(\xi_1, \xi_2), \xi_3, x_4(\xi_1, \xi_2, \xi_3), \xi_4, x_5(\xi_1, \xi_2, \xi_3, \xi_4),$$

where x_5 is the terminal response. Denote

$$\xi = (\xi_1, \xi_2, \xi_3, \xi_4)^T \in \mathbb{R}^4,$$

then the number of the possible values of ξ, i.e. the number of scenarios, is $5^4 = 625$. Besides, denote

$$x(\xi) = (x_1(\xi), x_2(\xi), x_3(\xi), x_4(\xi), x_5(\xi)),$$

then one can find that $x(\cdot)$ is in the following nonanticipativity set:

$$\mathcal{N} := \{x(\cdot) \ : \ x_i(\xi) \text{ dose not depend on } \xi_i, \cdots, \xi_4 \ \forall\ i = 1, \ldots, 4\}. \tag{24}$$

Decision variables and parameters involved in this model are listed in the following and the corresponding detailed data are presented in Tables 2-4 in the Appendix.

Decision variables:
$x_1(\xi) \in \mathbb{R}^{45}$: surface water irrigation area target of each crop in every subarea before random variable ξ is realized (ha);
$x_2(\xi), x_3(\xi) \in \mathbb{R}^{15}$: actual irrigation area of wheat in every subarea during period 1 and period 2 under scenario ξ (ha);
$x_4(\xi) \in \mathbb{R}^{45}$: actual irrigation area of each crop in every subarea during period 3 under scenario ξ (ha);
$x_5(\xi) \in \mathbb{R}^{30}$: actual irrigation area of maize and cotton in every subarea during period 4 under scenario ξ (ha);
Parameters:
b_i: net benefit per unit of area that water allocated in period i ($i = 1, 2, 3, 4$) (RMB/ha);
c_i: extra cost per unit of area not irrigated during period i ($i = 1, 2, 3, 4$) (RMB/ha);
a_i: irrigation quota for each crop during period i ($i = 1, 2, 3, 4$) (m^3/ha);
$q_i(\xi) \in \mathbb{R}$: random variable of total water availability for irrigation during period i

$(i = 1, 2, 3, 4)$ (m^3);
$l_b, u_b \in \mathbb{R}^{45}$: the lower and upper bounds of the irrigation area target (ha).

We can formulate this irrigation planning problem into the MLPM minimization (5), i.e.,

$$\begin{aligned} &\min\ \mathbb{E}\big(f(x(\xi),\xi)\big) + \beta\mathbb{E}\Big(f(x(\xi),\xi) - \mathbb{E}\big(f(x(\xi),\xi)\big)\Big)_+^m \\ &\text{s.t.}\ \ x(\cdot) \in \mathcal{C} \cap \mathcal{N}, \end{aligned} \tag{25}$$

with $\mathcal{N}$ defined as (24) and

$$f(x(\xi),\xi) = -\sum_{i=1}^{4} b_i^T x_{i+1}(\xi) + \sum_{i=1}^{4} c_i^T (x_1(\xi) - x_{i+1}(\xi)),$$

$$C(\xi) := \left\{ x \ : \ \begin{array}{c} \sum_{k=1}^{i} a_k^T x_{k+1} \le \sum_{k=1}^{i} q_k(\xi),\ \forall\ i = 1,2,3,4, \\ x_i \le x_1(1:15),\ i = 2,3, \\ x_4 \le x_1, \\ x_5 \le x_1(16:45), \\ l_b \le x_1 \le u_b, \end{array} \right\}.$$

where $x_1(j:k)$ means a subvector consisted from the j-th element to the k-th element of vector x_1. It should be pointed out that when reformulated this problem into a large-scale nonlinear programming, the dimension of variable is at least 24870, while adopting the proposed LPHA (Algorithm 3), one need to solve 625 subproblems with dimension being 152 at each iteration.

Next, we apply Algorithm 3 to solve the MLRM minimization problem (25) and in comparison with the expectation minimization problem, i.e. Problem (25) with $\beta = 0$, solving by Algorithm 1. The test code is written in Matlab R2015b and run on a PC with an Intel(R) Core(TM) i7-7500U 2.90 GHz CPU and 16 GB of RAM under WINDOWS 10 operating system. Parameters are set as $\beta = 1,\ m = 2,\ r = 1$. And the initial point is set by

$$\begin{aligned} \hat{z}^0(\xi) &= \text{argmin}_{\, z(\xi)\in C_2(\xi)} h(z(\xi),\xi) \quad \forall \xi, \\ z^0(\cdot) &= P_{\mathcal{N}_2}(\hat{z}^0(\cdot)), u^0(\cdot) = 0, w^0(\cdot) = 0, v^0(\cdot) = 0. \end{aligned}$$

The stopping criterion is taken as

$$\frac{\|(Z^{\nu+1}, W^{\nu+1}) - (Z^\nu, W^\nu)\|_r}{\|(Z^\nu, W^\nu)\|_r} \le 1.0e-5 \text{ or MaxIt} > 500,$$

where MaxIt presents the maximal number of iterations.

Table 4 lists the first-stage solutions and the optimal values of problem (25) with $\beta = 1, m = 2$ and $\beta = 0$, in which fval means the optimal value, iter means the number of convergence iterations, time(s) means the total time for solve the problems in seconds, and Avg-time(s) means the average time of each iteration in seconds. One can find that the decision made by problem (25) is more conservative than the decision made by expectation minimization, and the optimal value of problem (25) is greater than that of expectation minimization which verifies the risk-averseness of the risk measure LMR. Besides, for this problem, the number of iteration for converging by Algorithm 3 is 126 while the number of convergence iteration by original PHA for expectation minimization problem is 368. The time cost by solving MLRM minimization is 3220 seconds, which is longer than the time for solving expectation minimization problem, which is 2970 seconds.

Table 2. Numerical results

	MLRM ($\beta = 1, m = 2$)			Expectation Minimization		
Region	Wheat	Maize	Corn	Wheat	Maize	Corn
1	1730	1770	870	1730	1870	1000
2	2570	2000	1050	2570	2200	1150
3	1430	1050	550	1430	1200	650
4	2730	1716	1820	2730	1820	2080
5	2550	2280	144	2550	2490	174
6	1145	1145	95	1145	1250	117.5
7	4860	4400	600	4860	4740	734
8	2070	990	55.5	2070	1072.5	55.5
9	1980	2010	120	1980	2160	120
10	420	420	6	420	450	12
11	645	630	4.5	645	675	7.5
12	1950	1590	225	1950	1680	225
13	8000	9167.5	450	8000	10000	450
14	1354.7	437	89.7	1354.7	483	128.8
15	255	240	0	255	270	0
fval	-5.2×10^7			-9.1×10^7		
iter	126			368		
time(s)	3220			2970		
Avg-time(s)	25			8		

5. Conclusion

This paper is concerned with a popular risk-averse multistage stochastic optimization problem in financial engineering and business management. We develop monotone stochastic variational inequality models for this problem. Based on a delinkage-and-decomposition idea, a Lagrangian progressive hedging algorithm is designed to solve the stochastic variational inequalities. The proposed algorithm decomposes the original problem into hindside optimization problems for each scenario and then updates the iterative solution by a projection procedure to maintain its nonanticipativity. It is shown that this algorithm is convergent if the original model has a solution and satisfies a constraint qualification condition. A numerical real-world example is solved by the proposed algorithm. The computational results show that the Lagrangian progressive hedging algorithm can be effectively used in practice.

Acknowledgments

This work was partially supported by the Australian Research Council (Grant DP160102819), National Science Foundation of China (Grants 71631008, 11801130,11671029 and 11671116) and Chinese Academy of Sciences (Grant Y932231).

References

Bawa, VS (1975). Optimal rules for ordering uncertain prospects. *J. Financ. Econ.*, 2(1), 95–121.

Bawa, VS and Lindenberg, EB (1977). Capital market equilibrium in a mean-lower partial moment framework. *J. Financ. Econ.*, 5, 189–200.

Chen, L, He, S and Zhang, S (2011). Tight bounds for some risk measures, with applications to robust portfolio selection. *Oper. Res.*, 59(4), 847–865.

Dai, ZH and Li, YP (2013). *A multistage irrigation water allocation model for agriculturalland-use planning under uncertainty. Agr. Water Manag.*, 129, 69–79.

Fishburn, PC (1977). Mean-risk analysis with risk associated with below-target returns. *Am. Econ. Rev.*, 67(2), 116–126.

Grootveld, H and Hallerbach, W (1999). Variance vs downside risk: is there really that much difference. *Eur. J. Oper. Res.*, 114, 304–319.

Harlow, WV (1991). Asset allocation in a downside-risk framework. *Financ. Anal. J.*, 47(5), 28–40.

Ling, A, Sun, J and Wang, M (2019). Robust multi-period portfolio selection based on downside risk with asymmetrically distributed uncertainty set. *Eur. J. Oper. Res.* , Available online at https://doi.org/10.1016/j.ejor.2019.01.012

Markowitz, HM (1959). *Porfolio Selection: Efficient Diversification of Investment.* New York: John Wiley.

Rockafellar, RT (1970). *Convex Analysis.* New Jersey: Princeton University Press.

Rockafellar, RT and Sun, J (2019). Solving monotone stochastic variational inequality and stochastic complementarity problems by progressive hedging. *Math. Program.*, 174, 453–471.

Rockafellar, RT and Sun, J (2020). Solving Lagrangian variational inequalities with applications to stochastic programming, to appear in *Math. Program.* DOI: 10.1007/s10107-019-01458-0.

Rockafellar, RT and Wets, RJB (1991). Scenarios and policy aggregation in optimization under uncertainty. *Math. Oper. Res.*, 16(1), 119–147.

Sun, J, Xu, HL and Zhang, M (2019). A new interpretation of the progressive hedging algorithm for multistage stochastic minimization problems. *J. Ind. Manag. Optim.*, DOI: 10.3934/jimo.2019022

Yu, L, Zhang, S and Zhou, X (2006). A downside risk analysis based on financial index tracking models. *Stoch. Finan. Part I*, 213–236.

Appendix

Table 3. Unit net benefit and extra cost per area during each period (RMB/ha)

Unit net benefit b_i				Unit extra cost c_i			
b_1	b_2	b_3	b_4	c_1	c_2	c_3	c_4
194	291	1940	2550	42.5	85	722.5	520
211.2	316.8	2112	3200.25	42.5	85	722.5	520
207.6	311.4	2076	3106.75	42.5	85	722.5	520
227.2	340.8	2272	3259.75	42.5	85	722.5	520
199.2	298.8	1992	2541.5	42.5	85	722.5	520
195.6	293.4	1956	2584	42.5	85	722.5	520
226.4	339.6	2264	3285.25	42.5	85	722.5	520
206	309	2060	2779.5	42.5	85	722.5	520
209.6	314.4	2096	2528.75	42.5	85	722.5	520
219.2	328.8	2192	2911.25	42.5	85	722.5	520
209.6	314.4	2096	3043	42.5	85	722.5	520
122.4	183.6	1224	2159	42.5	85	722.5	520
216.8	325.2	2168	2707.25	42.5	85	722.5	520
203.6	305.4	2036	2834.75	42.5	85	722.5	520
229.2	343.8	2292	2800.75	42.5	85	722.5	520
		450	1869			130	357.5
		564.75	2432.5			130	357.5
		548.25	1827			130	357.5
		575.25	1904			130	357.5
		448.5	1760.5			130	357.5
		456	1841			130	357.5
		579.75	1893.5			130	357.5
		490.5	1389.5			130	357.5
		446.25	1022			130	357.5
		513.75	1981			130	357.5
		537	2366			130	357.5
		381	1270.5			130	357.5
		477.75	1200.5			130	357.5
		500.25	1956.5			130	357.5
		494.25	0			130	357.5
		801				292.5	
		1042.5				292.5	
		783				292.5	
		816				292.5	
		754.5				292.5	
		789				292.5	
		811.5				292.5	
		595.5				292.5	
		438				292.5	
		849				292.5	
		1014				292.5	
		544.5				292.5	
		514.5				292.5	
		838.5				292.5	
		0				292.5	

Table 4. Irrigation quota during each period (10^3m^3/ha), lower bound and upper bound (ha)

Irrigation quota a_i				lower bound	upper bound
a_1	a_2	a_3	a_4	l_b	u_b
0.165	0.33	2.805	1.76	1730	1830
0.165	0.33	2.805	1.76	2570	2680
0.165	0.33	2.805	1.76	1430	1570
0.165	0.33	2.805	1.76	2730	2925
0.165	0.33	2.805	1.76	2550	2670
0.165	0.33	2.805	1.76	1145	1207.5
0.165	0.33	2.805	1.76	4860	5140
0.165	0.33	2.805	1.76	2070	2175
0.165	0.33	2.805	1.76	1980	2025
0.165	0.33	2.805	1.76	420	450
0.165	0.33	2.805	1.76	645	675
0.165	0.33	2.805	1.76	1950	2070
0.165	0.33	2.805	1.76	8000	8667.5
0.165	0.33	2.805	1.76	1354.7	1430.6
0.165	0.33	2.805	1.76	255	270
		0.44	1.21	1770	1870
		0.44	1.21	2000	2200
		0.44	1.21	1050	1200
		0.44	1.21	1716	1820
		0.44	1.21	2280	2490
		0.44	1.21	1145	1250
		0.44	1.21	4400	4740
		0.44	1.21	990	1072.5
		0.44	1.21	2010	2160
		0.44	1.21	420	450
		0.44	1.21	630	675
		0.44	1.21	1590	1680
		0.44	1.21	9167.5	10000
		0.44	1.21	437	483
		0.44	1.21	240	270
		0.99		870	1000
		0.99		1050	1150
		0.99		550	650
		0.99		1820	2080
		0.99		144	174
		0.99		95	117.5
		0.99		600	734
		0.99		55.5	64.5
		0.99		120	150
		0.99		6	12
		0.99		4.5	7.5
		0.99		225	255
		0.99		450	492.5
		0.99		89.7	128.8
		0.99		0	0