



**HAL**  
open science

# A multiplicative weights update algorithm for MINLP

Luca Mencarelli, Youcef Sahraoui, Leo Liberti

► **To cite this version:**

Luca Mencarelli, Youcef Sahraoui, Leo Liberti. A multiplicative weights update algorithm for MINLP. EURO Journal on Computational Optimization, 2017, 5 (1-2), pp.31-86. 10.1007/s13675-016-0069-8 . hal-02105301

**HAL Id: hal-02105301**

**<https://hal.science/hal-02105301v1>**

Submitted on 20 Apr 2019

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# A multiplicative weights update algorithm for MINLP

LUCA MENCARELLI<sup>1</sup>, YUCEF SAHRAOUI<sup>1,2</sup>, LEO LIBERTI<sup>1</sup>

<sup>1</sup> *CNRS LIX, École Polytechnique, F-91128 Palaiseau, France*  
Email:{liberti,mencarelli,sahraoui}@lix.polytechnique.fr

<sup>2</sup> *OSIRIS, EDF R&D, F-92141 Clamart, France*  
Email:youcef.sahraoui@edf.fr

May 4, 2016

## Abstract

We discuss an application of the well-known Multiplicative Weights Update (MWU) algorithm to non-convex and mixed-integer nonlinear programming. We present applications to: (a) the distance geometry problem, which arises in the positioning of mobile sensors and in protein conformation; (b) a hydro unit commitment problem arising in the energy industry, and (c) a class of Markowitz' portfolio selection problems. The interest of the MWU with respect to one of its closest competitors (classic Multi-Start) is that it provides a relative approximation guarantee on a certain quality measure of the solution.

## 1 Introduction

The Multiplicative Weights Update (MWU) algorithm [2] is a stochastic heuristic with a relative performance guarantee on a weighted average of the errors. Its typical application is to decide the best way to take advice from a set of advisors which repeatedly express different opinions. At each iteration, the MWU performs the following tasks: (i) it updates a probability distribution on the advisors' performance, based on gains/costs at the preceding iteration; (ii) it samples new decisions from the current distribution; and (iii) it updates the gains/costs vector based on the decisions. The MWU yields an approximation guarantee on the weighted average gains/costs relative to the performance of the best advisor.

The MWU was used in the past to derive approximation algorithms for Linear Programming (LP) [33] and Semidefinite Programming (SDP) [1]. In this paper, we describe the computational application of the MWU to two novel settings: nonconvex Nonlinear Programming (NLP) and Mixed-Integer Nonlinear Programming (MINLP). Although we do not derive an approximation algorithm, we demonstrate the applicability of the MWU as a heuristic for NLP and MINLP. We showcase its application to three well known problems: the Distance Geometry Problem (DGP) [29], a variant of the Hydro Unit Commitment (HUC) problem from the energy industry [40], and a variant of Markowitz' Mean-Variance Portfolio Selection Problem (MVPS) [32]. We show that the MWU still retains its (relative) approximation guarantee in these new settings, whilst generally performing as well as, or better than, a classic Multi-Start (MS) approach (see Alg 1). Note that some work is necessary to adapt MS to any given optimization problem.

---

### Algorithm 1 Multi-Start

---

- 1: **while** termination condition is not met **do**
  - 2:   sample a starting point  $x'$
  - 3:   perform a local descent from  $x'$ , yielding  $\tilde{x}$
  - 4:   if  $\tilde{x}$  improves the best optimum  $x^*$  so far, update  $x^*$  with  $\tilde{x}$
  - 5: **end while**
- 

We must determine the best distribution to sample from, what kind of local descent algorithm to employ,

and which termination conditions are appropriate. We shall see that the MWU heuristic applies generally to optimization problems in a similar way: a nontrivial amount of work is necessary to adapt it to the given problem.

Our main motivation for singling out the MWU as a good heuristic is that, unlike the vast majority of heuristics, the MWU comes with a relative approximation guarantee on some user-defined cumulative (over all MWU iterations) mean costs (or errors)  $\psi_1, \dots, \psi_q$ , which turn out to be bounded above by a piecewise linear function of the cumulative (again, all over MWU iterations) *lowest* cost  $\min_{i \leq q} \psi_i$ . Although, in general, this is a rather weak guarantee, it is surprising that it should exist at all, considering the generality of the method. The MWU readily turns into an approximation algorithm whenever the structure of the problem allows one to provide a guaranteed upper bound to the lowest cost. Although we do not derive such approximations in this paper, we mean to study this issue in further works, and hope that this paper will spark interest in the matter.

## 1.1 The MWU algorithm

We believe that Arora et al.'s excellent survey [2] provides an introduction to the MWU that we cannot hope to improve as regards clarity: we therefore borrow from it here to introduce the MWU by means of an application example.

An investor employs  $q$  advisors to help her decide every day whether the price of a given stock will increase or decrease. We let  $\theta_i$  be the prediction of the  $i$ -th advisor, for each  $i \leq q$ . The investor maintains a weight vector  $\omega \in [0, 1]^q$ , where the component  $\omega_i$  is used to weigh the reliability of the  $i$ -th advisor, for each  $i \leq q$ . At the outset, all weights are initialized to 1. As the days pass, each wrong prediction is penalized by a unit cost  $\psi_i$  to be born by the  $i$ -th advisor, whereas correct predictions have zero cost. Note that we can also assign gains, represented as negative costs; and, more generally, we consider a cost vector  $\psi \in [-1, 1]^q$ , where  $\psi_i$  denotes the cost, or gain, to be attributed to advisor  $i$  (for  $i \leq q$ ). At the end of the day, the weights are updated as follows:

$$\forall i \leq q \quad \omega_i \leftarrow \omega_i(1 - \eta\psi_i), \quad (1.1)$$

where  $\eta$  is a given positive constant  $\leq \frac{1}{2}$ . The next day, each prediction  $\theta_i$  is weighed by a random number sampled in  $[0, \omega_i]$  (for  $i \leq q$ ), and the investor makes her decision based on the weighted average of the predictions. The weights  $\omega$  should be thought of as a discrete distribution over  $\mathcal{Q} = \{1, \dots, q\}$  which is updated every day. As such, we define their normalized version as  $p = \omega / \|\omega\|_1$ .

More formally, given positive constants  $\eta \leq \frac{1}{2}$  and  $T \in \mathbb{N}$ , the MWU maintains a list of weights  $\omega = (\omega_i \mid i \leq q) \in [0, 1]^q$  which are used to randomly update the values of a vector  $\theta \in \Theta \subseteq \mathbb{R}^q$  of decision variables related to an associated vector  $\psi \in [-1, 1]^q$  of costs in a componentwise fashion; these costs are then used to update the weights  $\omega$  according to Eq. (1.1) before starting the next iteration. This procedure is repeated for  $T$  iterations (see Alg. 2). The idea is that the weights  $\omega$  iteratively adapt

---

### Algorithm 2 (Didactical) Multiplicative Weights Update

---

- 1: **while** termination condition is not met **do**
  - 2:   weigh each prediction  $\theta_i$  by  $p_i$
  - 3:   compute the cost vector  $\psi$  associated to  $\theta$
  - 4:   update  $\omega$  using  $\psi$  as in Eq. (1.1)
  - 5: **end while**
- 

$\theta$  to being a good solution of an optimization problem which aims at minimizing a weighted average of the costs  $\psi$ . As the  $i$ -th cost  $\psi_i$  gets smaller (and perhaps negative, becoming a gain), the associated weight  $\omega_i$  increases (because of Eq. (1.1)). One (heuristically) hopes that this will yield an even smaller cost at the next iteration. As the MWU proceeds for  $T$  iterations, we index  $\omega$ ,  $p$  and  $\psi$  by  $t \leq T$ . Thus,

the expected cost of the process on day  $t \leq T$  is  $p^t \psi^t$ , and the cumulative expected cost over the time horizon is  $\sum_{t \leq T} p^t \psi^t$ .

As already mentioned, it takes a nontrivial amount of work to adapt the MWU to the NLP and MINLP settings: how do we relate  $\theta$  to a given (MI)NLP, and how do we compute  $\psi$ ? This paper provides some answers to these questions, based on three very different applications.

## 1.2 The MWU approximation guarantee

We recall here the precise statement of the MWU approximation guarantee. Given  $q, T, \eta, \omega, \psi, p$  as above, it can be shown that [2]:

$$E_{\text{MWU}} \triangleq \sum_{t \leq T} \psi^t p^t \leq \min_{i \leq q} \left( \sum_{t \leq T} \psi_i^t + \eta \sum_{t \leq T} |\psi_i^t| \right) + \frac{\ln q}{\eta}. \quad (1.2)$$

We remark that Eq. (1.2) is actually an immediate consequence of the more general bound:

$$\forall i \leq q \quad E_{\text{MWU}} \leq \sum_{t \leq T} \psi_i^t + \eta \sum_{t \leq T} |\psi_i^t| + \frac{\ln q}{\eta},$$

and that different weight update rules yield slightly different bounds.

## 1.3 The MWU in mixed-integer nonlinear programming

The objective of this paper is to propose an adaptation of the MWU to optimization problems in the following very general Mixed-Integer Nonlinear Programming (MINLP) form:

$$\left. \begin{array}{l} \min_{x \in \mathbb{R}^n} f(x) \\ \forall \ell \leq m \quad g_\ell(x) \leq 0 \\ \forall j \in Z \quad x_j \in \mathbb{Z}, \end{array} \right\} [P] \quad (1.3)$$

where  $Z \subseteq \{1, \dots, n\}$  is given. If  $Z = \emptyset$ , the problem is called a Nonlinear Programming (NLP) problem.

There are three main issues in abstracting Alg. 2 to Eq. (1.3).

1. As presented in Alg. 2, the MWU algorithm appears to make a single decision (the investor's) based on a sequence of random predictions, whereas Eq. (1.3) makes  $n$  decisions.
2. The objective function  $f(x)$  is given, and will generally be different than the cumulative weighted cost average that the MWU attempts to minimize.
3. Eq. (1.3) is hard to solve also because of feasibility, not just optimality. Nothing in Alg. 2 is about feasibility.

We shall address all three issues by means of a particular reformulation of Eq. (1.3) which partitions the decisions into two vectors of decision variables. One of the two vectors is simply  $x$ , the decision variables in Eq. (1.3). The other vector  $\theta$  encodes (directly or indirectly) the values of some problematic terms in Eq. (1.3). The values for  $\theta$  are decided using the MWU algorithm, whereas those for  $x$  are decided by solving the reformulation (with  $\theta$  fixed at the values assigned by the MWU) at each iteration of the. Naturally, the replacement of problematic terms by easier terms in function  $\theta$  will be chosen so as to make the resulting reformulation easier to solve than Eq. (1.3). We remark that the reformulation is parametrized by  $\theta$ .

This strategy addresses the second and third issues: the current solution found by solving the reformulation at each MWU iteration can be “plugged into” Eq. (1.3): this yields an objective function value for Eq. (1.3) at each iteration, as well as a measure of infeasibility of the current solution. The cost vector  $\psi$  can be defined in terms of the relative progress of the objective function value, as well as the infeasibilities.

Let us see how this strategy addresses the first issue. Specifically, the vector  $\theta$  in Alg. 2 is a vector of “random predictions” rather than decisions. This strategy, however, requires  $\theta$  to match the value of the terms they replace, or at least get sufficiently close. We achieve this by modifying Step 2 in Alg. 3 so that  $\theta$  is *updated* at each iteration by a factor sampled randomly in  $[0, \omega_i]$  (see Alg. 3). This will help  $\theta$  be “decided” by the MWU rather than “randomly sampled”. Notice that the relative approximation guarantee of Sect. 1.2 still holds, since it only depends on the weights update rule.

---

**Algorithm 3** Multiplicative Weights Update
 

---

- 1: **while** termination condition is not met **do**
  - 2:   for each  $i \leq q$ , multiply  $\theta_i$  by a factor sampled randomly from  $[0, \omega_i]$
  - 3:   compute the cost vector  $\psi$  associated to  $\theta$
  - 4:   update  $\omega$  using  $\psi$  as in Eq. (1.1)
  - 5: **end while**
- 

In the strategy that we will propose in Section 3,  $\theta$  could have a different dimension with respect to  $\omega$  and  $\psi$ . This is one of the issues for which adapting Alg. 3 to applications requires nontrivial work.

The rest of this paper is organized as follows. In Sect. 2, we propose a reformulation-based methodology in order to relate  $\theta$  to Eq. (1.3), and discuss the application of this methodology to three real-world applications. We present the adaptation of the MWU for (MI)NLP in Sect. 3, and the computational results in Sect. 4.

## 2 Pointwise reformulations

We introduce the reformulation referred to in Sect. 1.3. Broadly speaking, we reformulate  $P$  (see Eq. (1.3)) by replacing  $r$  “problematic” terms (e.g., the nonconvex terms) with simpler terms parametrized by  $\theta$ . This yields a simplified formulation  $R$  in the original decision variables  $x$ , which varies in function of  $\theta$ . We shall then iteratively solve  $R$  with  $\theta$  fixed to values determined by the MWU framework. We remark that, with this approach, our MWU-based algorithm belongs to the category of *mat-heuristics* [31], meaning heuristics based on Mathematical Programming (MP).

Notationally, for a MP formulation  $P$ , we write  $\text{val}(P)$  to denote the objective function value of a global optimum of  $P$ , and  $\text{feas}(P)$  to denote the feasible set of  $P$ .

### 2.1 Definition

Given a MINLP  $P$  as in Eq. (1.3), a *pointwise reformulation*  $R^\theta = \underset{\mathbf{t} \leftarrow \mathbf{t}'(\theta)}{\text{ptw}}(P)$  is a family of MINLP formulations, parametrized by  $\theta = (\theta_s \mid s \leq r)$ , which are obtained by replacing given occurrences  $\mathbf{t}_1, \dots, \mathbf{t}_r$  of terms appearing in  $P$  by corresponding parametrized terms  $\mathbf{t}'_s(\theta_s)$  (for  $s \leq r$ ).  $\square$

Notationally,  $\mathbf{t}$  and  $\mathbf{t}'$  denote “terms” in the MINLP formulation  $P$ . We define a *term* formally as a symbolic expression represented by its parsing tree [15]. The parsing tree has leaf nodes labelled by the values of the constants or the index of the variables appearing in the term, and all other nodes labelled by the operators. The terms  $\mathbf{t}'$  are expressed in function of the parameter vector  $\theta$ , so we denote them  $\mathbf{t}'(\theta)$ . Both  $\mathbf{t}$  and  $\mathbf{t}'$  also depend on the decision variable vector  $x$ . These terms can be evaluated by means of a very simple recursive algorithm which starts at the root of the parsing tree: at each non-leaf

node  $v$  of the tree it calls itself on all the subnodes  $v_1, \dots, v_h$ , and then returns the value  $\otimes(v_1, \dots, v_h)$  where  $\otimes$  is the operator represented by the label of  $v$ ; if  $v$  is a leaf node, the value of the constant or of the variable represented by the leaf node is returned. Through the algorithm, each term corresponds to a function  $\mathbb{R}^n \rightarrow \mathbb{R}$ , which we denote by  $\mathbf{t}(x)$  and  $\mathbf{t}'(\theta, x)$ .

For every replaced term  $\mathbf{t}_s$  (for  $s \leq r$ ) in Defn. 2.1, let  $D_s$  be the interval range of  $\mathbf{t}_s(x)$ . For every replacement term  $\mathbf{t}'_s$  (for  $s \leq r$ ) let  $D'_s(\theta_s)$  be the interval range of  $\mathbf{t}'_s(\theta_s, x)$ . For  $s \leq r$ , let  $\Theta_s$  be the interval range of the corresponding parameter  $\theta_s$ , and let  $\Theta = (\Theta_s \mid s \leq r)$ .

Given a parameter vector  $\theta \in \Theta$  and a function  $\phi$ , we denote by  $\phi^\theta$  the function obtained by replacing the terms  $\mathbf{t}$  by the terms  $\mathbf{t}'(\theta)$ . Thus, for example, the objective and constraints of  $R$  are denoted as  $f^\theta, g^\theta$ , respectively.

We can therefore write a pointwise reformulation  $R^\theta$  of  $P$  as follows:

$$\left. \begin{array}{l} \min_{x \in \mathbb{R}^{n'}} f^\theta(x) \\ \forall \ell \leq m \quad g_\ell^\theta(x) \leq 0 \\ \forall j \in Z' \quad x_j \in \mathbb{Z}, \end{array} \right\} [R^\theta] \quad (2.1)$$

where  $Z' \subseteq \{1, \dots, n'\}$ .

Note that Eq. (2.1) is actually a family of formulations, parametrized by  $\theta$ . Note also that, whereas the number of variables  $n'$  may be different from  $n$  (since many variables occurring in replaced terms might be replaced terms involving fewer or more parameter symbols),  $P$  and  $R^\theta$  have the same number of constraints  $m$ . If a replacement yields a trivial constraint, we stipulate that it is still formally part of Eq. (2.1). In practice, when solving pointwise reformulations, trivially satisfied constraints may be dropped, of course.

## 2.1 Theoretical properties

### 2.2 Definition

Given a MINLP  $P$  and  $R^\theta = \underset{\mathbf{t} \leftarrow \mathbf{t}'(\theta)}{\text{ptw}}(P)$ , both defined on a vector  $x$  of decision variables in  $\mathbb{R}^n$ :

- (a)  $R^\theta$  is *spanning* if, for any  $x \in \mathbb{R}^n$ , there are values of  $\theta$  such that evaluating the functions of  $P$  and of  $R^\theta$  at  $x$  yields the same results — more precisely,  $\exists \bar{\theta} \in \Theta_s$  such that

$$\forall s \leq r \quad D_s \subset \bigcup_{\theta_s \in \Theta_s} D'_s(\theta_s) \quad \wedge \quad [\mathbf{t}'_s(\bar{\theta}_s)](x) = \mathbf{t}_s(x)$$

(note that the first condition in the conjunction is a form of consistency);

- (b)  $R^\theta$  is *exact* if, for each globally optimal solution  $x^*$  of  $P$ , there is at least one vector  $\theta' \in \Theta$  such that  $x^*$  is also an optimal solution of  $R^{\theta'}$ ;
- (c)  $R^\theta$  is *efficient* if there is a polynomial-time algorithm for approximately solving  $R^\theta$  (for  $\theta \in \Theta$ ) to within a given  $\varepsilon > 0$  approximation factor.  $\square$

Note that the exactness property of a pointwise reformulation only makes sense for feasible problems. We arbitrarily define any pointwise reformulation of a class of infeasible problems to be exact.

More informally, we say that a pointwise reformulation is *good* if there is an established, practically efficient technology for solving  $R^\theta$  either optimally or approximately. For example, if  $R^\theta$  turns out to be a Linear Program (LP) or convex NLP (cNLP), then  $R^\theta$  is efficient; if it turns out to be a Mixed Integer Linear Program (MILP),  $R^\theta$  is good, since current MILP solution technology can routinely solve fairly large-scale MILPs, even though MILP itself is **NP**-hard [25]. Obviously, every efficient pointwise reformulation is also good.

### 2.3 Example

Consider the following formulation  $P$ :

$$\min x^2(1 - y) + y \tag{2.2}$$

$$x + 2y \geq 2 \tag{2.3}$$

$$y \in \{0, 1\}. \tag{2.4}$$

If we set  $y = 0$ , Eq. (2.3) and the objective function direction force  $x = 2$ , whereas if  $y = 1$  we can let  $x = 0$ ; therefore the global optimum is  $(x^*, y^*) = (0, 1)$ . We replace the term  $x^2$  in the objective function by the term consisting of the scalar parameter  $\theta$ , obtaining a pointwise reformulation  $R^\theta$ :

$$\min (1 - \theta)y + \theta$$

$$x + 2y \geq 2$$

$$y \in \{0, 1\}.$$

It is easy to see that  $R^\theta$  is spanning whenever  $\theta \in \mathbb{R}_+$ . If we set  $y = 0$  we obtain  $x \geq 2$ , whereas  $y = 1$  yields no constraints on  $x$ . The objective function value if  $y = 0$  is  $\theta$ ;  $y = 1$  yields  $1 - \theta + \theta = 1$ . So for  $\theta < 1$  the set of global optima is  $[2, \infty) \times \{0\}$ ;  $\theta > 1$  yields the global optimal set  $\mathbb{R}_+ \times \{1\}$ , and if  $\theta = 1$  every feasible solution is optimal, with optimal objective function value equal to 1. Hence  $(x, y) = (x^*, y^*) = (0, 1)$  yields an optimum as long as  $\theta \geq 1$ , which means that this pointwise reformulation is exact. This pointwise reformulation is not efficient, since it is a MILP, but it is good.  $\square$

### 2.4 Lemma

Given  $P$  and a spanning reformulation  $R^\theta = \underset{\mathbf{t} \leftarrow \mathbf{t}'}{\text{ptw}}(P)$ , we have:

$$\text{feas}(P) \subseteq \bigcup_{\theta \in \Theta} \text{feas}(R^\theta). \tag{2.5}$$

*Proof.* Let  $x' \in \text{feas}(P)$ . Since  $R^\theta$  is spanning, there is  $\xi \in \Theta$  such that  $\mathbf{t}_s(x') = [\mathbf{t}_s(\xi_s)](x')$  for each  $s \leq r$ , which implies  $g_\ell^\xi(x^*) = g_\ell(x^*)$  for all  $\ell \leq m$ . Since  $x' \in \text{feas}(P)$ ,  $g_\ell(x') \leq 0$  for all  $\ell \leq m$ , hence  $x'$  is also feasible in  $R^\xi$ . Since  $\text{feas}(R^\xi)$  is a subset of the right hand side of Eq. (2.5) for each possible  $\xi \in \Theta$ , the result follows.  $\square$

The following example shows that Eq. (2.5) cannot be tightened to an equality.

### 2.5 Example

Consider the pure feasibility NLP formulation  $F$ :

$$x \geq \frac{1}{2} \tag{2.6}$$

$$x^2 = x, \tag{2.7}$$

where  $x$  is a continuous decision variable. The constraint in Eq. (2.7) is equivalent to  $x \in \{0, 1\}$ , so Eq. (2.6) forces  $x = 1$ , hence  $\text{feas}(F) = \{1\}$ . We rewrite  $x^2 = x$  and replace the first occurrence of  $x$  by  $\theta$ , which yields the pointwise reformulation  $\underset{\mathbf{t} \leftarrow \mathbf{t}'}{\text{ptw}}(F) = R^\theta$ :

$$x \geq \frac{1}{2} \tag{2.8}$$

$$\theta x = x. \tag{2.9}$$

Any value of  $\theta \neq 1$  requires  $x = 0$  in order for Eq. (2.9) to hold, but  $x = 0$  is infeasible by Eq. (2.8). Setting  $\theta = 1$  yields  $\text{feas}(R^1) = [\frac{1}{2}, \infty)$ . In particular, we have

$$\{1\} = \text{feas}(F) \subsetneq \bigcup_{\theta \in \mathbb{R}} \text{feas}(R^\theta) = \text{feas}(R^1) = [\frac{1}{2}, \infty). \tag{2.10}$$

Since  $F$  and  $R^\theta$  have no objective function, every feasible solution is optimal by definition. Verifying exactness reduces to checking that, for every feasible solution of  $F$ , there are values of  $\theta$  such that the same solution is feasible in  $R^\theta$ , which is established by Eq. (2.10). So this reformulation is exact. Since  $F$  is a nonconvex NLP but  $R^\theta$  is an LP, this reformulation is efficient.  $\square$

A relaxation of a MP formulation  $Q$  provides a guaranteed bound (in the optimization direction) at every feasible point of  $Q$ . Since relaxations must be efficiently solvable, and since one usually looks for a bound to the *optimal* objective function value of  $Q$ , rather than to any objective function value achieved by a feasible point in  $Q$ , it makes sense to generalize a relaxation so it is about optimal rather than feasible points. A *bounding reformulation* of  $Q$  as a reformulation which, when solved to optimality, provides a bound in the optimization direction to the optimal objective function value of  $Q$  (and, moreover, its feasible set contains the feasible set of  $Q$ ). Obviously, all relaxations are bounding reformulations.

### 2.6 Lemma

For any formulation  $P$  and spanning pointwise reformulation  $R^\theta$ , there exists  $\xi \in \Theta$  such that  $R^\xi$  is a bounding reformulation of  $P$ .

*Proof.* The proof of Lemma 2.4 implies that, if  $\xi \in \Theta$  is such that  $\mathbf{t}_s(x^*) = [\mathbf{t}_s(\xi_s)](x^*)$  for all  $s \leq r$ ,  $x^* \in \text{feas}(R^\xi)$ . Similarly, we show  $f^\xi(x^*) = f(x^*)$ , and therefore:

$$\text{val}(R^\xi) \leq f^\xi(x^*) = f(x^*) = \text{val}(P),$$

which establishes the result.  $\square$

We remark that the bounding reformulation guaranteed by Lemma 2.6 need not be a relaxation in the traditional sense.

### 2.7 Remark

Often, the replacement process  $\mathbf{t} \leftarrow \mathbf{t}'(\theta)$  raises a cardinality issue: we replace  $r$  terms of the original formulation, and we have to use its solution to compute  $q$  costs, where  $r$  might in general be different from  $q$ . We shall discuss this issue in Sect. 3.1 below.  $\square$

Last but not least, note that pointwise reformulations can be used in more general settings than just the MWU algorithm. Although they have been devised with the MWU in mind, what they really achieve is a general mechanism for automatically decomposing the solution process of a MINLP in two phases: one for deciding values of  $\theta$ , and the other for deciding values of  $x$ , based on solving the corresponding pointwise reformulation.

## 2.2 Applications

In this section, we present pointwise reformulations for three real-world MINLP problems, known to be difficult to solve both from a theoretical and a practical viewpoint, namely:

- the Distance Geometry Problem (DGP) with Euclidean distances;
- the Hydro-power (short-term) Unit Commitment problem (HUC);
- a subclass of nonconvex variants of the Markowitz' Mean-Variance Portfolio Selection Problem (MVPS).

The structure of each of those problems is exploited to construct a pointwise reformulation, in view of solving the problem using the MWU algorithm.



### 2.2.1 Distance Geometry Problem

The Distance Geometry Problem (DGP) with Euclidean distances is defined formally as follows [29]: given an integer  $K > 0$ , a simple undirected graph  $G = (V, E)$ , and an edge weight function  $d : E \rightarrow \mathbb{R}_+$ , establish or deny the existence of a vertex realization function  $x : V \rightarrow \mathbb{R}^K$  such that:

$$\forall \{u, v\} \in E \quad \|x_u - x_v\|_2 = d_{uv}. \quad (2.11)$$

The DGP arises in many important applications: determination of protein conformation from distance data [30], localization of mobile sensors in communication networks [18], synchronization of clocks from phase information [38], control of unmanned submarine fleets [3], spatial logic [19], and more [29]. It is **NP**-complete when  $K = 1$  and **NP**-hard for larger values of  $K$  [34]. Notationwise, we let  $n = |V|$  and  $m = |E|$ .

The most common MP formulation for the DGP is:

$$\min_x \sum_{\{u,v\} \in E} (\|x_u - x_v\|_2^2 - d_{uv}^2)^2, \quad (2.12)$$

It is obvious that the given DGP instance is YES if and only if the globally optimal value of Eq. (2.12) is zero. Note, however, that the DGP is not currently known to be in **NP** for  $K > 1$  [6]. Therefore, if we wish to employ floating point based solution methods, we will not be able to determine precisely whether a given objective function value is exactly zero. We therefore take a more practical standpoint, and postulate that the quality of a solution is proportional to the value of a given error function.

Consider now the following formulation:

$$\left. \begin{array}{l} \max_x \sum_{\{u,v\} \in E} \|x_u - x_v\|_2^2 \\ \forall \{u, v\} \in E \quad \|x_u - x_v\|_2^2 \leq d_{uv}^2. \end{array} \right\} \quad (2.13)$$

### 2.8 Proposition

Eq. (2.13) is an exact reformulation<sup>1</sup> of Eq. (2.12).

*Proof.* Replacing the outer square in Eq. (2.12) by an absolute value trivially yields an exact reformulation:

$$\min_x \sum_{\{u,v\} \in E} |\|x_u - x_v\|_2^2 - d_{uv}^2|, \quad (2.14)$$

since Eq. (2.14) is zero iff Eq. (2.12) is zero. We reformulate Eq. (2.14) as:

$$\left. \begin{array}{l} \min_{x, \hat{t} \geq 0} \sum_{\{u,v\} \in E} \hat{t}_{uv} \\ \forall \{u, v\} \in E \quad -\hat{t}_{uv} \leq \|x_u - x_v\|_2^2 - d_{uv}^2 \leq \hat{t}_{uv}. \end{array} \right\} \quad (2.15)$$

The concave constraints  $-\hat{t}_{uv} \leq \|x_u - x_v\|_2^2 - d_{uv}^2$  can be moved back to the objective, yielding:

$$\left. \begin{array}{l} \min_{x, \hat{t} \geq 0} \sum_{\{u,v\} \in E} ((d_{uv}^2 - \|x_u - x_v\|_2^2) + \hat{t}_{uv}) \\ \forall \{u, v\} \in E \quad \|x_u - x_v\|_2^2 - d_{uv}^2 \leq \hat{t}_{uv}. \end{array} \right\} \quad (2.16)$$

Assume first that the given DGP is YES; then Eq. (2.15) must have globally optimal objective function value zero, which implies that  $\hat{t}_{uv} = 0$  (for all  $\{u, v\} \in E$ ) at the optimum, and so the following

<sup>1</sup>An *exact reformulation* (formally defined in [28] as a surjective mapping from the optima of the exact reformulation to the optima of the original problem) is not the same as an *exact pointwise reformulation*. Intuitively speaking, solving an exact reformulation of a problem directly yields a solution of the problem itself, which is not generally the case for an exact pointwise reformulation.

reformulation of Eq. (2.16),

$$\left. \begin{array}{l} \min_x \sum_{\{u,v\} \in E} (d_{uv}^2 - \|x_u - x_v\|_2^2) \\ \forall \{u, v\} \in E \quad \|x_u - x_v\|_2^2 - d_{uv}^2 \leq 0, \end{array} \right\} \quad (2.17)$$

preserves a mapping between solutions of the DGP and global optima of Eq. (2.15). If the DGP is NO then the globally optimal objective function value of Eq. (2.15) must be strictly greater than zero, which implies that either Eq. (2.17) is infeasible, or it has a strictly positive globally optimal objective function value, whence Eq. (2.17) is an exact reformulation of Eq. (2.12). Finally, we eliminate the objective constant (which does not change global optima) and write  $\min -f$  as  $\max f$ , obtaining the exact reformulation:

$$\left. \begin{array}{l} \max_x \sum_{\{u,v\} \in E} \|x_u - x_v\|_2^2 \\ \forall \{u, v\} \in E \quad \|x_u - x_v\|_2^2 \leq d_{uv}^2, \end{array} \right\}$$

as claimed.  $\square$

### 2.9 Corollary

If the given DGP instance is YES, at any global optimum  $x^*$  of Eq. (2.13) all constraints  $(\forall \{u, v\} \in E \|x_u - x_v\|_2^2 \leq d_{uv}^2)$  of Eq. (2.13) are active.

*Proof.* By Lemma 2.8, Eq. (2.13) correctly solves a DGP instance where the given partial Euclidean Distance Matrix (pEDM) is  $(d_{uv})$ . The global optimum  $x^*$  must therefore satisfy Eq. (2.11), which implies that  $\|x_u^* - x_v^*\|_2 = d_{uv}$  for each  $\{u, v\} \in E$ , as claimed.  $\square$

We can easily derive a pointwise reformulation of Eq. (2.13) by replacing the quadratic term  $(x_{uk} - x_{vk})^2 = (x_{uk} - x_{vk})(x_{uk} - x_{vk})$  occurring in the objective function with a linear term  $\theta_{uvk}(x_{uk} - x_{vk})$ :

$$\left. \begin{array}{l} \max_x \sum_{\{u,v\} \in E} \sum_{k \leq K} \theta_{uvk}(x_{uk} - x_{vk}) \\ \forall \{u, v\} \in E \quad \|x_u - x_v\|_2^2 \leq d_{uv}^2. \end{array} \right\} \quad (2.18)$$

It is easy to see that Eq. (2.18) is spanning, since, given a solution  $x'$ , it suffices to set  $\theta'_{uvk} = (x'_{uk} - x'_{vk})$  for the objective and constraints of the pointwise reformulation to take identical values to the objective and constraints of the original formulation.

### 2.10 Proposition

If the given DGP instance is YES, for each globally optimal solution  $x^*$  of Eq. (2.13) there is a parameter matrix  $\theta \in \mathbb{R}^{mK}$  such that  $x^*$  is a globally optimal solution of Eq. (2.18).

*Proof.* By Lemma 2.6, Eq. (2.18) is a bounding reformulation of Eq. (2.13) for  $\theta^* = (x_{uk}^* - x_{vk}^* | \{u, v\} \in E \wedge k \in K)$ , where  $x^*$  is a global optimum of Eq. (2.13). Since Eq. (2.18) is a maximization problem, we have to show that there is no optimum  $x'$  of Eq. (2.18) such that  $f^{\theta^*}(x') > f^{\theta^*}(x^*)$ . To aim at a contradiction, we suppose the existence of such an  $x'$ . Then there must be at least one edge  $\{u, v\} \in E$  such that

$$\begin{aligned} \sum_{k \leq K} \theta_{uvk}^* (x'_{uk} - x'_{vk}) &> \|x_u^* - x_v^*\|_2^2 \\ \Rightarrow \sum_{k \leq K} (x_{uk}^* - x_{vk}^*) (x'_{uk} - x'_{vk}) &> \|x_u^* - x_v^*\|_2^2 \\ \Rightarrow (x_u^* - x_v^*) (x'_u - x'_v) &> \|x_u^* - x_v^*\|_2^2. \end{aligned}$$

The last equation can be re-written as  $y \cdot z > \|z\|^2$ , for  $y = x'_u - x'_v$  and  $z = x_u^* - x_v^*$ , and hence as  $\|y\| \|z\| \cos(\mu) > \|z\|^2$ , where  $\mu$  is the angle between  $y$  and  $z$ . Since  $\|z\| = 0$  yields  $0 > 0$ , we assume

$\|z\| > 0$ , and we can hence divide through by  $\|z\|$ , yielding  $\|y\| \cos(\mu) > \|z\|$ . Since  $\cos(\mu) \leq 1$ , the only way the latter equation can hold is if  $\|y\| > \|z\|$ . Hence:

$$\begin{aligned} \|x'_u - x'_v\|_2 &> \|x_u^* - x_v^*\|_2 \\ \Rightarrow \|x'_u - x'_v\|_2^2 &> \|x_u^* - x_v^*\|_2^2 = d_{uv}^2 \end{aligned}$$

by Cor. 2.9. So  $\|x'_u - x'_v\|_2^2 > d_{uv}^2$  which means that  $x'$  is infeasible in Eq. (2.18). The result follows.  $\square$

### 2.11 Theorem

The formulation in Eq. (2.18) is an exact and efficient pointwise reformulation of Eq. (2.11).

*Proof.* The fact that Eq. (2.18) is a pointwise reformulation of Eq. (2.11) follows by Prop. 2.8. The fact that it is exact follows by Prop. 2.10. The fact that it is efficient follows by the existence of polynomial-time algorithms for cNLP (see e.g. [12]).  $\square$

### 2.2.2 Hydro-power unit commitment

The short-term HUC problem is essentially a scheduling problem, defined as follows: over a uniformly discretized time horizon  $H = \{1, \dots, \bar{h}\}$  (ranging from one day to a week), given an initial water volume  $V_0$  in a water reservoir, the goal is to find the optimal schedule of released water flow ( $x_h \mid h \in H$ ) (expressed in  $\text{m}^3$ ) which maximizes the revenue obtained by providing the generated power ( $y_h \mid h \in H$ ) (expressed in MW) to the grid, such that the final volume of water remaining in the reservoir reaches a desired target  $V_{\bar{h}}$  (again expressed in  $\text{m}^3$ ).

Additional features of the model include:

- time period duration  $\tau$  (expressed in h)
- volume bounds  $\underline{V}$  and  $\bar{V}$  (expressed in  $\text{m}^3$ ) on the volume  $v_h$ , for each  $h \in H$
- maximum flow bound  $\bar{Q}$  (expressed in  $\text{m}^3/\text{s}$ ) on the released flow  $x_h$ , for each  $h \in H$
- maximum ramp-down  $Q^-$  and ramp-up  $Q^+$  (expressed in  $\text{m}^3/\text{s/h}$ ) for the flow  $x_h$ , for each  $h \in H$
- forecasted inflows  $I_h$  (expressed in  $\text{m}^3/\text{s}$ ), for each  $h \in H$
- forecasted power selling prices  $\Pi_h$  (expressed in currency/MWh), for each  $h \in H$
- parameters and coefficients  $K_1, \dots, K_6, L_1, \dots, L_6, \underline{L}$  and  $R_0$  of a polynomial function which models the generated power  $y_h$ , for each  $h \in H$ .

The simplified version can be formulated as follows, where  $v$  is the volume of water in the reservoir:

$$\left. \begin{aligned} \max_{x,y,v} \quad & \sum_{h \in H} \tau \Pi_h y_h \\ & v_0 = V_0 \\ & v_{\bar{h}} = V_{\bar{h}} \\ \forall h \in H \quad & v_{h+1} - v_h = 3600\tau(I_h - x_h) \\ \forall h \in H \quad & x_h - x_{h+1} \leq Q^- \\ \forall h \in H \quad & x_{h+1} - x_h \leq Q^+ \\ \forall h \in H \quad & y_h = \varphi(x_h, v_h) \\ \forall h \in H \quad & v_h \in [\underline{V}, \bar{V}] \\ \forall h \in H \quad & x_h \in [0, \bar{Q}], \end{aligned} \right\} \quad (2.19)$$

where  $\varphi(x, v) = 9.81x(\sum_{l=0}^6 L_l x^l)(\sum_{k=0}^6 K_k v^k - \underline{L} - R_0 x^2)$  is the function expressing the power generated depending on the water flow released  $x$  and the water volume  $v$  in the reservoir.

By replacing each nonconvex multivariate function  $\varphi(x_h, v_h)$  (for  $h \in H$ ) with an affine approximation (i.e. a first-order approximation at a given point  $(\tilde{x}_h, \tilde{v}_h)$ ), we derive a pointwise linear reformulation of Eq. (2.19):

$$\left. \begin{aligned} \max_{x,y,v} \quad & \sum_{h \in H} \tau \Pi_h y_h \\ & v_0 = V_0 \\ & v_{\bar{h}} = V_{\bar{h}} \\ \forall h \in H \quad & v_{h+1} - v_h = 3600\tau(I_h - x_h) \\ \forall h \in H \quad & x_h - x_{h+1} \leq Q^- \\ \forall h \in H \quad & x_{h+1} - x_h \leq Q^+ \\ \forall h \in H \quad & y_h = \nu_{0h} + \nu_{1h}(x_h - \tilde{x}_h) + \nu_{2h}(v_h - \tilde{v}_h) \\ \forall h \in H \quad & v_h \in [\underline{V}, \bar{V}] \\ \forall h \in H \quad & x_h \in [0, \bar{Q}], \end{aligned} \right\} \quad (2.20)$$

We remark that, in the pointwise reformulation Eq. (2.20), the parameter vector  $\theta$  is structured as the  $h \times 5$  matrix  $(\tilde{x}, \tilde{v}, \nu_0, \nu_1, \nu_2)$ .

### 2.12 Lemma

The pointwise reformulation Eq. (2.20) is spanning.

*Proof.* For each  $h \in H$ , the following holds: if  $x'_h \in [0, \bar{Q}]$  and  $v'_h \in [\underline{V}, \bar{V}]$ , the replacement term  $\nu_{0h} + \nu_{1h}(x_h - \tilde{x}_h) + \nu_{2h}(v_h - \tilde{v}_h)$  matches the replaced term  $\varphi(x_h, v_h)$  for all values of  $\theta$  satisfying  $\tilde{x}_h = x'_h$ ,  $\tilde{v}_h = v'_h$  and  $\nu_{0h} = \varphi(x'_h, v'_h)$ .  $\square$

### 2.13 Proposition

For each globally optimal solution  $X^* = (x^*, v^*, y^*)$  of Eq. (2.19), there is a  $\theta^* \in \mathbb{R}^{5\bar{h}}$  such that  $X^*$  is a globally optimal solution of Eq. (2.20).

*Proof.* Let us show that  $X^*$  is a globally optimal solution of the pointwise problem Eq. (2.20) for

$$\theta_h^* = (x_h^*, v_h^*, \varphi(x_h^*, v_h^*), \frac{\partial \varphi}{\partial x_h}(x_h^*, v_h^*), \frac{\partial \varphi}{\partial v_h}(x_h^*, v_h^*)),$$

where  $h \in H$ . Since this definition of  $\theta^*$  satisfies the spanning property in Prop. 2.12, we can invoke Lemma 2.4 to conclude that  $X^*$  is feasible for Eq. (2.20). Since  $X^*$  is a global optimum of Eq. (2.19), in particular it is also a local optimum, and therefore it satisfies the Karush-Kuhn-Tucker (KKT) conditions. By the choice of  $\theta^*$ , the gradients of the objective function and the constraints of Eq. (2.20) at  $X^*$  are identical to gradients of objective and constraints of Eq. (2.19) at  $X^*$ . Therefore,  $X^*$  also satisfies first-order optimality conditions of Eq. (2.20) when parametrized by  $\theta^*$ . Since Eq. (2.20) is an LP, any KKT point is also a global optimum, which concludes the proof.  $\square$

### 2.14 Theorem

The formulation in Eq. (2.20) is an exact and efficient pointwise reformulation of Eq. (2.19).

*Proof.* Exactness of Eq. (2.20) follows by Prop. 2.13. Efficiency follows because Eq. (2.20) is an LP, which can be solved in polynomial time.  $\square$

### 2.2.3 Mean-variance portfolio selection

Our third application of pointwise reformulations is an uncountable class of problems, parametrized by symbols which denote abstract univariate functions (any assignment of concrete functions to the symbols

represents a different problem). Using this very general setting, we mean to provide an insight of how much better or worse the MWU performs, compared to MS, depending on how far or close such functions are to being linear. In Sect. 4 below, we shall show that the MWU performs better than the MS on functions which are “very nonlinear”, whereas the MS wins out on functions which are “close to linear”. Since this is only a qualitative statement derived empirically on a few univariate functions, we leave the interpretation of these informally expressed notions to the inspection of the function graphs (see Fig. 4).

The MVPS problem [32] is defined as follows: given a set of  $n$  possibly risky stocks, characterized by a mean return vector  $\rho \in \mathbb{R}^n$  and a covariance matrix  $Q \in \mathbb{R}^{n \times n}$ , determine a composition of the portfolio, i.e., the fraction  $x_i$  of value invested in stock  $i \leq n$ , by simultaneously considering two conflicting targets: maximizing expected return of the portfolio and minimizing a measure of its risk. For an in-depth analysis on mean-variance approaches to portfolio selection problems, we refer the reader to the survey [17].

We consider here the MVPS variant with transaction costs and maximum number of held assets (also called the *sparsity* of the portfolio). Transaction costs are usually defined by means of a separable function  $C : \mathbb{R}^n \rightarrow \mathbb{R}_+$  of the assets held in the portfolio:

$$C(x) = \sum_{i=1}^n C_i(x_i), \quad (2.21)$$

where, for all  $i \leq n$ ,  $C_i : \mathbb{R} \rightarrow \mathbb{R}_+$  is a univariate possibly nonconvex and nonconcave function. Note that this setting is much more general than the one usually considered in literature, involving concave non-decreasing transaction cost (see [26, 41]).

Let  $K \in \mathbb{R}_+$  and  $\sigma \in \mathbb{R}_+$  be respectively the maximum sparsity level and the maximum risk desired for the portfolio. The MVPS variant we are interested in can be formally stated as follows:

$$\left. \begin{array}{l} \max_x \quad \rho^\top x - C(x) \\ \mathbf{1} x = 1 \\ x^\top Q x \leq \sigma \\ \|x\|_0 \leq K \\ x \in [\underline{x}, \bar{x}] \cup \{0\}, \end{array} \right\} \quad (2.22)$$

where  $\mathbf{1} \in \mathbb{R}^n$  is the all-one  $n$ -dimensional vector,  $\|x\|_0$  is the so-called  $\ell_0$ -norm of  $x$ , i.e.,  $\|x\|_0 = |\{x_i : x_i \neq 0\}|$ ,  $\underline{x}$  and  $\bar{x}$  are respectively minimum and maximum buy-in thresholds, and  $x$  is a vector of  $n$  semi-continuous decision variables [21, 39].

In Eq. (2.22), the first constraint ensures the whole available capital is invested in the portfolio. Portfolio sparsity is used for modeling the will of investors to limit e.g. brokerage fees, bid-ask spreads and monitoring costs [35]. Moreover, we assume  $0 \leq \underline{x} \leq \bar{x}$ , so that short selling, i.e., the possibility for the investor to sell financial assets he/she does not own, is avoided. Eq. (2.22) is an **NP**-hard problem, even if  $n = 3$  [8, 37], and can be exactly reformulated (in the sense of [28]) as the following non-convex MINLP:

$$\left. \begin{array}{l} \max_{x,y} \quad \rho^\top x - C(x) \\ \mathbf{1} x = 1 \\ x^\top Q x \leq \sigma \\ \mathbf{1} y \leq K \\ y^\top \underline{x} \leq x \leq y^\top \bar{x} \\ y \in \{0, 1\}^n. \end{array} \right\} \quad (2.23)$$

We can obtain a pointwise reformulation by replacing terms  $C_i(x_i)$  with terms  $(1 + x_i)\theta_i$ , for every  $i \leq n$ . This yields the pointwise reformulated objective function:

$$\rho^\top x - \sum_{i \leq n} (\theta_i x_i + \theta_i) = (\rho - \theta)^\top x - \mathbf{1} \theta,$$

which is an affine function of  $x$ . Thus, the pointwise reformulation is:

$$\left. \begin{aligned} \max_{x,y} \quad & (\rho - \theta)^\top x - \mathbf{1} \theta \\ & \mathbf{1} x = 1 \\ & x^\top Q x \leq \sigma \\ & \mathbf{1} y \leq K \\ & y^\top \underline{x} \leq x \leq y^\top \bar{x} \\ & y \in \{0, 1\}^n. \end{aligned} \right\} \quad (2.24)$$

### 2.15 Remark

The pointwise reformulation in Eq. (2.24) is spanning, since the replacement terms  $\theta_i$  match the replaced terms  $\frac{C_i(x_i)}{x_i+1}$  at each feasible point ( $x_i \mid i \leq n$ ): it suffices to divide  $C_i(x_i) = (1+x_i)\theta_i$  through by  $1+x_i$ . By Lemma 2.6, there exist values of  $\theta$  which make Eq. (2.24) a bounding reformulation for the original problem Eq. (2.22).  $\square$

### 2.16 Example

The pointwise reformulation in Eq. (2.24) is not exact. It suffices to consider a portfolio with  $n = 2$  stocks,  $Q = I_2$ ,  $K = 2$ ,  $\underline{x} = 0$ ,  $\bar{x} = 0.55$ , i.e. essentially unconstrained apart from the total budget constraint  $x_1 + x_2 = 1$ , with  $\rho = (1, 0.40)^\top$ , and  $C(x) = (C_1(x_1), C_2(x_2))^\top = (x_1, 0)^\top$ . Note that, for all  $i \in \{1, 2\}$ ,  $\theta_i = \frac{C_i(x_i)}{1+x_i}$ , and hence  $\Theta_i = [C_i(\underline{x})/(1+\bar{x}), C_i(\bar{x})/(1+\underline{x})]$ . Therefore, in this example,  $\Theta_1 = [0, 0.55]$  and  $\Theta_2 = \{0\}$ . The objective function of the original formulation Eq. (2.22) is simply  $\max(0.40x_2)$ , which implies a global optimum  $x^* = (0.45, 0.55)^\top$ . The objective function of the pointwise reformulation is  $\max((1-\theta_1)x_1 + 0.40x_2 - \theta_1)$ . Now,  $x^* = (0.45, 0.55)^\top$  is a global optimum of the pointwise reformulation if and only if  $(1-\theta_1) < 0.40$ , i.e., iff  $\theta_1 > 0.60$ : however, these values of  $\theta_1$  do not belong to the set  $\Theta_1$ .

### 2.17 Remark

The pointwise reformulation in Eq. (2.24) is not efficient, since the MVPS is **NP-hard** [8]. However, the pointwise reformulation in Eq. (2.24) is good, since the solver technology for finding solutions of convex MIQPs is more advanced and reliable than that for solving nonconvex nonconcave MINLP.  $\square$

The previous models can be easily extended in order to take into account the presence of fixed transaction costs. Furthermore, the simplicity of the previous proposed approach makes it a promising pointwise reformulation candidate for the extension to other mixed integer nonlinear problems with separable nonconvexities.

## 3 The MWU for MINLPs

The MWU (Alg. 3) samples some values from an iteratively updated distribution in order to optimize a given loss or gain criterion. We adapt this framework to (MI)NLP by exploiting good pointwise reformulations, which can be solved more efficiently than the original formulation, and hence can be solved repeatedly at a relatively low computational cost. We rely on a set of parameters  $\theta$  which can be guessed using the MWU framework. This yields a loop which updates the values of  $\theta$ , uses them to solve a pointwise reformulation, and then estimates the error over each variable in order to update  $\theta$  again at the next iteration.

The pseudocode of the MWU algorithm for MINLP is shown in Alg. 4. It takes a MINLP formulation  $P$  and a pointwise reformulation  $\text{ptw}(P)$  as inputs, and hopefully produces a good solution as output.

Step 5 in Alg. 4 is not actually needed to ensure that the relative approximation guarantee Eq. (1.2) holds. Without it, however, the MWU is computationally much less efficient and effective.

**Algorithm 4** MWU( $P$ )

- 
- 1: Initially set the iteration index  $t = 1$ , weights  $\omega^t = \mathbf{1}$ , parameters  $\theta^{t-1}$  chosen uniformly at random in  $\Theta$ , an incumbent  $x^* = \infty$  and  $t = 1$
  - 2: **while**  $t \leq T$  **do**
  - 3:   assign  $\theta^t \leftarrow \theta^{t-1} \bar{\omega}^t$  where  $\bar{\omega}^t$  is vector chosen uniformly at random in  $[0, \omega^t]$
  - 4:   solve  $\underset{t \leftarrow t'(\theta^t)}{\text{ptw}}(P)$ , get solution  $x^t$
  - 5:   refine  $x^t$  (e.g. using local descent on  $P$ )
  - 6:   if  $x^t$  is better than the incumbent, replace  $x^* \leftarrow x^t$
  - 7:   compute costs  $\psi^t \in [-1, 1]^q$  from  $x^t$
  - 8:   update weights for the next iteration:  $\forall i \leq q \ \omega_i^{t+1} \leftarrow \omega_i^t (1 - \frac{\psi_i^t}{2})$
  - 9:   normalize weights such that  $\sum_{i \leq q} \omega_i^{t+1} = 1$
  - 10:   increase  $t$
  - 11: **end while**
- 

We now clarify the dependencies between symbols occurring in the MWU algorithm, symbols occurring in the pointwise reformulation, and the optima of the original formulation.  $E_{\text{MWU}}$ , the function that the MWU aims at minimizing, depends on  $p$  and  $\psi$ ;  $p$  depends on  $\omega$ , which is updated using  $\psi$ , and  $\psi$  depends on the local solution  $x$  of the pointwise reformulation, which replaces the incumbent  $x^*$  whenever it improves it (with respect to the objective function of the original formulation). Note also that  $x$  depends on  $\theta$  through the pointwise reformulation, and that  $\theta$  is randomly chosen from the discrete distribution  $p$ , proportional to  $\omega$ .

Below, we discuss Steps 3, 4-5 and 7 of Alg. 4 in more detail.

### 3.1 Sampling

As mentioned in Remark 2.7, the dimension  $r$  of  $\theta$  and the dimension  $q$  of  $\omega$  and  $\psi$  might be different. The question is how to apply Step 2 in Alg. 3, since it implicitly assumes that  $r = q$ . We deal with this issue by defining  $\theta$  using aggregations (if  $r > q$ ) or disaggregations (if  $q > r$ ) of the values  $\bar{\omega}$  sampled from  $[0, \omega]$  (see Step 3 in Alg. 4). Aggregations and disaggregations are obtained by applying any number of operators, such as products or sums, to  $\bar{\omega}$ . Since there are many ways to split products and sums in a prescribed number of different parts, the precise details of this step are heuristic in nature. We note that the MWU performance guarantee on  $E_{\text{MWU}}$  (see Eq. (1.2)) depends on  $q$  but not on  $r$ , and so it is not impacted by such details.

In Sect. 3.3 we sketch a general methodology which eschews this issue by ensuring that  $r = q$  (so that sampling  $\theta$  from the distribution  $p$  poses no problem). This methodology is then applied to the HUC and MVPS problems. In the DGP application, on the other hand, we have  $mK = r > q = n$ , where  $n = |V|$ ,  $m = |E|$  and  $(K, G = (V, E, d))$  is the DGP input.

### 3.2 Solution and refinement

#### 3.2.1 Solving pointwise reformulations

If  $P$  has no integer variable and  $\underset{t \leftarrow t'}{\text{ptw}}(P)$  is efficient, it is likely to be an LP or a cNLP, both of which can be either solved or accurately approximated in polynomial time. If the pointwise reformulation is not efficient but at least good, it may be a type of nonconvex NLP for which we have a practically fast solver which scales reasonably well.

If  $P$  has some integer variables and  $\underset{t \leftarrow t'}{\text{ptw}}(P)$  is good,  $P$  might be a MILP or a convex MINLP (cMINLP). This complicates matters, since solving a MILP or a cMINLP to optimality at each iteration is usually computationally costly, even if good solver technologies exist. Note, however, that all the MWU algorithm needs in order for its guarantee to hold is simply an error vector  $\psi^t$  at each iteration  $t \leq T$  and the weight update rule Eq. (1.1). So in fact we can run *any heuristic we like* on the pointwise reformulation. In practice, we run Branch-and-Bound methods with a bound on computation time or optimality gap.

### 3.2.2 Refining the pointwise optimum

The refinement step is optional in theory, but computational experience shows it is necessary in practice for the MWU to perform well. If  $x'$  is the solution of the pointwise reformulation, any solver which is designed to improve  $x'$  with respect to the *original* formulation  $P$  (at least locally) can be used to refine  $x'$ .

### 3.3 Computing the MWU costs

This is the most critical step of the MWU algorithm, since it influences the performance guarantee. It has two requirements:

- (a)  $\psi^t \in [-1, 1]^q$  for all  $t \leq T$ ;
- (b) for any  $t \leq T$ ,  $\psi^t$  measures the error of the current local solution  $x^t$  of the pointwise reformulation with respect to optimality and feasibility of the original formulation.

We need (a) to prove the MWU relative approximation guarantee, and (b) in order to relate  $E_{\text{MWU}}$  to the solution quality of the incumbent  $x^*$  (see issues 2-3 in Sect. 1.3).

Based on (a) and (b), we compute a scalar  $\alpha^t$  related to optimality, and a set  $\{\beta_\ell^t \mid \ell \leq m\}$  of vectors related to feasibility. Since we penalize infeasibilities but we generally do not award “better feasibility”, the components of  $\beta_\ell^t$  are usually required to be in  $[0, 1]$  rather than  $[-1, 1]$ .

More specifically, let

$$R^{\theta^t} = \underset{t \leftarrow t'(\theta^t)}{\text{ptw}}(P)$$

be the pointwise reformulation at iteration  $t \leq T$ , let  $f(x), g_\ell(x) \leq 0$  (for  $\ell \leq m$ ) be the objective function and constraints of  $P$  as per Eq. (1.3), and let  $f^{\theta^t}(x), g_\ell^{\theta^t}(x) \leq 0$  be those of  $R^t$  (for  $\ell \leq m$ ). After Steps 4-5 of Alg. 4, we can evaluate the current solution  $x^t$  in the pointwise reformulation by computing  $f^{\theta^t}(x^t)$  and  $g_\ell^{\theta^t}(x^t)$  for each  $\ell \leq m$ . We define arrays of values,  $\alpha^t, \beta^t$  at each iteration  $t \leq T$ :

- let  $\alpha^t$  be proportional to  $f^{\theta^t}(x^t) - f(x^t)$ , so as to favor a pointwise reformulation with a lower objective value (for a MINLP in the general minimization form Eq. (1.3); if considering a maximization problem,  $\alpha^t$  should be replaced by  $-\alpha^t$ );
- for all  $\ell \leq m$ , let  $\beta_\ell^t$  be proportional to  $\max(g_\ell^{\theta^t}(x^t), 0)$ , so as to penalize a pointwise reformulation which makes a feasible solution infeasible.

The arrays  $\alpha, \beta$  can be scaled in any way which makes them satisfy requirement (a) above. We assume this scaling is application-dependent. We can now define  $\psi^t$  in a very simple way as the concatenation of  $\alpha^t$  and  $\beta_\ell^t$  for all  $\ell \leq m$ , which also fixes  $q = 1 + m$ . Other application-dependent ways of defining  $\psi$  by means of  $\alpha, \beta$  are also possible (see Sect. 3.4.1).



### 3.4 Adapting the MWU for MINLP to applications

In this section we discuss the adaptations of Alg. 4 to the different application settings we considered.

#### 3.4.1 Distance Geometry Problem

In this section, we discuss the adaptation of Alg. 4 to the DGP application setting.

The pointwise reformulation Eq. (2.18) we employ for the DGP relies on  $\theta^t = (\theta_{uvk}^t \mid \{u, v\} \in E \wedge k \leq K)$  having dimension  $r = K|E|$ ,  $x^t = (x_v^t \mid v \in V)$  having dimension  $n = |V|$ , and  $\psi^t = (\psi_{uv}^t \mid \{u, v\} \in E)$  having dimension  $q = |E|$ , at any iteration  $t \leq T$ . Also, the original formulation Eq. (2.11) is a pure feasibility problem consisting of  $|E|$  quadratic equations, without objective function or integrality constraints. We therefore define:

$$\forall \{u, v\} \in E \quad \psi_{uv}^t = \frac{|\|x_u^t - x_v^t\|_2 - d_{uv}|}{\max(\|x_u^t - x_v^t\|_2, d_{uv})}. \quad (3.1)$$

This definition of  $\psi$  is close to the interpretation of  $\beta$  given in Sect. 3.3. Since  $\beta_\ell^t$  is defined for the  $\ell$ -th inequality constraint, and equality constraints correspond to pairs of opposing inequality constraints, we simply observe that  $\max(\tilde{g}_\ell(x^t), 0)$  turns into

$$\max(g_\ell^{\theta^t}(x^t), -g_\ell^{\theta^t}(x^t)) = |g_\ell^{\theta^t}(x^t)| = |\|x_u^t - x_v^t\|_2 - d_{uv}|,$$

which is the numerator in the definition of  $\psi^t$  in Eq. (3.1).

As mentioned above and in Sect. 3.1, in the DGP application we have a discrepancy between  $r$  and  $q$ ; accordingly, we sample a random vector  $\bar{\omega}^t$  uniformly at random from  $[0, \omega^t]$ , and define  $\theta^{t+1}$  as a disaggregation of  $\bar{\omega}^t$  over  $k \in K$ , as follows:

$$\forall k \leq K, \{u, v\} \in E \quad \theta_{uvk}^{t+1} = \bar{\omega}_{uv}^t (x_{uk}^t - x_{vk}^t),$$

i.e., making a reasonable guess, we postulate that a good parameter  $\theta_{uvk}^{t+1}$  will be proportional to the  $k$ -th component of the vector  $x_u^t - x_v^t$ .

The performance guarantee of the MWU applied to the DGP turns out to be slightly better than the general one in Eq. (1.2), due to the fact that infeasibility is penalized but feasibility is not rewarded:

$$\min_{t \leq T} \sum_{\{u, v\} \in E} \psi_{uv}^t D_{uv}^t \leq \frac{1}{T} \left( \frac{\ln |E|}{\eta} + (1 + \eta) \min_{\{u, v\} \in E} \sum_{t \leq T} \psi_{uv}^t \right), \quad (3.2)$$

which is actually a statement on a weighted feasibility error of the MWU solution  $x^*$ , since it concerns the minimum over all iterations. For a proof of this statement, see [16, Prop. 4.1].

#### 3.4.2 Hydro-power unit commitment

In this section, we discuss the adaptation of Alg. 4 to the HUC application setting.

For each time period of the scheduling horizon, an upper bound on the contribution to the overall revenue is given by  $\bar{\Pi} \bar{\varphi} \triangleq \max_h \Pi_h \max_{x, v} \varphi(x, v)$ . This allows us to define costs:

$$\forall h \in H \quad \psi_h^t = \frac{\bar{\Pi} \bar{\varphi} - \Pi_h y_h}{\bar{\Pi} \bar{\varphi}}. \quad (3.3)$$

This definition of  $\psi^t$  is close to the interpretation of  $\alpha^t$  given in Sect. 3.3. The overall gain/cost reward is scaled appropriately for each time period.

The pointwise reformulation we employ for the HUC (Eq. (2.20)) relies on a parameter matrix  $\theta^t = (\tilde{x}^t, \tilde{v}^t, \nu_0^t, \nu_1^t, \nu_2^t)$ , and  $\bar{h}$ -dimensional cost vectors  $\psi^t$ . Note that  $\theta^t$  is  $\bar{h} \times 5$ , and that the first two vectors are actually points in the  $(x, v)$ -space. We therefore set  $\tilde{x}^t = x^{t-1}$ ,  $\tilde{v}^t = v^{t-1}$ , and sample  $\nu_a^t$  (for  $a \in \{0, 1, 2\}$ ) from  $\omega^t$ : it suffices to draw three different samples from the same distribution  $p$  proportional to  $\omega^t$ . More precisely,

$$\begin{aligned} \forall h \in H, \quad \nu_{0h}^t &= \bar{\omega}_h^t \varphi(\tilde{x}_h^t, \tilde{v}_h^t), \\ \nu_{1h}^t &= \bar{\omega}_h^t \frac{\partial \varphi}{\partial x_h}(\tilde{x}_h^t, \tilde{v}_h^t), \\ \nu_{2h}^t &= \bar{\omega}_h^t \frac{\partial \varphi}{\partial v_h}(\tilde{x}_h^t, \tilde{v}_h^t). \end{aligned}$$

Since the distance to  $\bar{\Pi}\bar{\varphi}$  is always penalized with nonnegative  $\psi$ , the performance guarantee of the MWU applied to the HUC is similar to Eq. (3.2):

$$\min_{t \leq T} \sum_{h \in H} \psi_h^t p_h^t \leq \frac{1}{T} \left( \frac{\ln \bar{h}}{\eta} + (1 + \eta) \min_{h \in H} \sum_{t \leq T} \psi_h^t \right). \quad (3.4)$$

### 3.4.3 Mean-variance portfolio selection

Note that the only complicating terms of the MVPS formulation Eq. (2.23) are the transaction costs in the objective function. Since feasibility is not an issue, we only need to define  $\psi$  for optimality purposes, i.e. we only consider the  $\alpha$  arrays mentioned in Sect. 3.3. Specifically, at each iteration  $t \leq T$ , we define a vector  $\alpha^t \in [0, 1]^n$ , whose components are:

$$\alpha_i^t = \frac{C_i(x_i^t) - (x_i^t + 1)\theta_i^t}{\max(|(x_i^t + 1)\theta_i^t|, |C_i(x_i^t)|)}. \quad (3.5)$$

Note that we define  $\alpha^t$  to be a vector of  $n$  components, instead of a scalar as in Sect. 3.3. This adaptation fits the MVPS well, since its objective function is separable, and since satisfying its constraints is not hard. Each component of  $\alpha^t$  defines a cost/gain relative to the contribution of each asset quantity towards the value of the objective function. We then simply set:

$$\forall i \leq n \quad \psi_i^t = \alpha_i^t. \quad (3.6)$$

In Alg. 4, the  $\theta^t$  vector is updated in terms of  $\omega$ , which are themselves updated as in Eq. (1.1). For this application, we decided to add a step with the purpose of explicitly scale each  $\theta_i^t$  (for  $i \leq n$ ) by  $\mathbf{t}_i = \frac{C_i(x_i^t)}{(1+x_i^t)}$  first, in order to make  $\theta_i^t$  a better replacement candidate for  $\mathbf{t}_i$ :

$$\forall i \leq n \quad \theta_i^t = \psi_i^t \frac{C_i(x_i^t)}{(1+x_i^t)}. \quad (3.7)$$

After this modification, we resume Alg. 4: we proceed to update  $\omega$ , then compute  $\theta^{t+1}$  as an update to  $\theta^t$  multiplied by random sample from the distribution  $p$ , proportional to  $\omega$ .

## 4 Computational experiments

In this section we present comparative computational experiments to validate the behaviour of the MWU algorithm for (MI)NLP in practice. We always compare the MWU with a classic MS heuristic, which is possibly the most similar existing algorithm to the MWU for (MI)NLP, on a fixed number  $T$  of iterations, and with initial points sampled uniformly within the instance ranges. We chose  $\eta = 0.5$  and  $T = 20$  for most of the experiments (unless stated otherwise).

## 4.1 DGP results

The test set consists of a set of 46 Protein Data Bank (PDB) [7] files, to be realized in  $\mathbb{R}^3$ . The protein graph has an edge  $\{i, j\}$ , weighted by the inter-atomic distance between atoms  $i$  and  $j$ , whenever this distance is smaller than 5Å (compatibly with Nuclear Magnetic Resonance [36] data). Note that, since all the proteins we tested actually exist, all of these instances are YES instances of the DGP. Based on the computational results, we partition these instances in two groups: EASY (27 instances) and HARD (19 instances). We configured both MWU and MS solvers with  $T = 20$  iterations, and use the IPOPT solver [14] as the local NLP solver in both methods.

Instance	$n$	$m$	$d:\text{min}$	max	avg	MWU			MS		
						err:avg	max	CPU	avg	max	CPU
C0700odd.1	8	28	0.93	4.06	2.39	0.00	0.00	0.61	0.00	0.00	<b>0.46</b>
C0700odd.2	8	28	0.93	4.06	2.39	0.00	0.00	0.61	0.00	0.00	<b>0.46</b>
C0700odd.3	8	28	0.93	4.06	2.39	0.00	0.00	0.62	0.00	0.00	<b>0.46</b>
C0700odd.5	8	28	0.93	4.06	2.39	0.00	0.00	0.61	0.00	0.00	<b>0.39</b>
C0700odd.6	8	28	0.93	4.06	2.39	0.00	0.00	0.60	0.00	0.00	<b>0.49</b>
C0700odd.7	8	28	0.93	4.06	2.39	0.00	0.00	0.64	0.00	0.00	<b>0.39</b>
C0700odd.8	8	28	0.93	4.06	2.39	0.00	0.00	0.61	0.00	0.00	<b>0.45</b>
C0700odd.9	8	28	0.93	4.06	2.39	0.00	0.00	0.61	0.00	0.00	<b>0.39</b>
C0700odd.A	8	28	0.93	4.06	2.39	0.00	0.00	0.61	0.00	0.00	<b>0.45</b>
C0700odd.B	8	28	0.93	4.06	2.39	0.00	0.00	0.62	0.00	0.00	<b>0.44</b>
lavor11	11	40	1.48	4.87	2.77	0.00	0.00	0.77	0.00	0.00	<b>0.51</b>
lavor11.7-2	11	47	1.49	4.97	2.95	0.00	0.00	<b>0.72</b>	0.00	0.00	0.83
lavor11.7	11	47	1.44	4.98	2.96	0.00	0.00	<b>0.68</b>	0.00	0.00	0.81
lavor11.7-1	11	47	1.44	4.98	2.96	0.00	0.00	<b>0.73</b>	0.00	0.00	0.80
lavor11.7-b	11	47	1.44	4.98	2.96	0.00	0.00	<b>0.78</b>	0.00	0.00	0.87
C0150alter.1	26	191	1.02	4.99	2.99	0.00	0.00	<b>2.01</b>	0.00	0.00	5.24
lavor30.6-4	30	191	0.64	4.96	2.88	0.00	0.00	<b>1.74</b>	0.00	0.00	4.00
lavor30.6-1	30	192	0.64	4.98	2.90	0.00	0.00	<b>1.85</b>	0.00	0.00	3.29
lavor30.6-8	30	193	0.64	4.99	2.90	0.00	0.00	<b>1.91</b>	0.00	0.00	3.85
lavor30.6-6	30	195	0.64	4.99	2.96	0.01	0.15	<b>1.71</b>	0.00	0.00	3.47
lavor30.6-7	30	195	0.64	4.98	2.93	0.00	0.00	<b>1.87</b>	0.00	0.00	4.11
lavor30.6-3	30	195	0.64	4.98	2.92	0.00	0.00	<b>1.73</b>	0.00	0.00	3.65
lavor30.6-5	30	195	0.64	4.99	2.93	0.00	0.00	<b>1.79</b>	0.00	0.00	3.37
lavor30.6-2	30	202	0.64	4.99	3.03	0.00	0.00	<b>1.81</b>	0.00	0.00	4.45
tiny	27	252	0.93	4.99	3.10	0.00	0.00	<b>1.56</b>	0.00	0.00	8.75
C0700.odd.G	36	308	0.98	4.99	3.04	0.00	0.00	<b>2.58</b>	0.00	0.00	8.29
C0700.odd.H	36	308	0.98	4.99	3.04	0.00	0.00	<b>2.55</b>	0.00	0.00	10.01
avg (CPU)								<b>1.22</b>			2.62
geo (CPU)								<b>1.05</b>			1.40

Table 1: Comparative results on the EASY instances of the DGP. The last line reports the arithmetic and geometric means of the CPU times.

The result tables 1 and 2 report, for each of the instance sets EASY and HARD, the instance PDB name, the number  $n$  of atoms, the number  $m$  of given distances, the minimum, maximum and average distance values in the instance, and the solver performance statistics. These are average and maximum edge errors, defined respectively as:

$$\frac{1}{m} \sum_{\{i,j\} \in E} |d_{ij} - \|x_i - x_j\|_2|$$

$$\max_{\{i,j\} \in E} |d_{ij} - \|x_i - x_j\|_2|,$$

and seconds of user CPU time of an i7 at 2GHz with 8GB RAM running Darwin 13.4.0.

Both methods are iterative, both have been run for the same number of iterations, MWU calls the local NLP solver twice per iteration (solution of the pointwise reformulation and refinement) and MS calls it only once. However, the MWU takes much less time (Fig. 1), and yields better results, as shown in Tables 1-2.

For the HARD instances only, we also carried out a comparison to a very well-known and efficient metaheuristic called Variable Neighbourhood Search (VNS) [23], applied to the DGP as described in [27], and using IPOPT as its local NLP solver. The range of applications where VNS excels borders on unbelievable. As is clear from the results in Table 3 and Fig. 2, it excels in the DGP case too, in

Instance	$n$	$m$	$d$ :min	max	avg	MWU			MS		
						$err$ :avg	max	CPU	avg	max	CPU
2er1-frag-bp1	39	406	0.81	4.99	3.11	0.00	0.00	<b>2.63</b>	0.00	0.00	15.06
C0080create.1	60	681	0.98	5.00	3.33	<b>0.04</b>	<b>0.41</b>	<b>6.17</b>	0.26	1.27	37.79
C0080create.2	60	681	0.98	5.00	3.33	<b>0.04</b>	<b>0.41</b>	<b>6.05</b>	0.06	0.99	36.50
names	82	840	1.21	5.00	3.48	0.16	1.28	<b>10.91</b>	<b>0.15</b>	<b>1.17</b>	79.85
pept	107	999	1.16	5.00	3.55	<b>0.16</b>	<b>1.38</b>	<b>15.53</b>	0.42	1.45	108.80
C0020pdb	107	999	1.16	5.00	3.55	<b>0.16</b>	<b>1.45</b>	<b>14.69</b>	0.44	1.92	101.72
1guu-1	150	959	1.20	4.99	3.46	<b>0.09</b>	<b>0.79</b>	<b>15.09</b>	0.10	1.06	30.80
1guu-4000	150	968	0.33	4.99	3.51	<b>0.11</b>	<b>0.68</b>	<b>19.31</b>	0.16	1.00	42.88
1guu	150	955	1.30	5.00	3.46	<b>0.09</b>	<b>0.70</b>	<b>15.33</b>	0.12	1.07	31.58
res_5000	108	1392	0.94	5.00	3.42	<b>0.11</b>	<b>1.75</b>	<b>22.79</b>	0.43	2.15	162.53
res_2000	108	1404	0.95	5.00	3.42	<b>0.17</b>	<b>1.80</b>	<b>21.46</b>	0.42	2.24	169.70
res_0	108	1410	0.94	5.00	3.43	<b>0.12</b>	2.07	<b>22.11</b>	0.38	<b>1.98</b>	154.44
res_3000	108	1487	0.97	5.00	3.48	<b>0.20</b>	<b>1.99</b>	<b>25.29</b>	0.41	2.38	209.24
res_1000	108	1506	0.94	5.00	3.49	<b>0.16</b>	<b>1.71</b>	<b>25.19</b>	0.40	2.23	213.81
res_2kxa	177	2627	0.93	5.00	3.52	<b>0.18</b>	<b>2.16</b>	<b>74.32</b>	0.39	2.75	621.78
2kxa	177	2711	0.93	5.00	3.53	<b>0.26</b>	<b>2.79</b>	<b>78.66</b>	0.44	3.07	834.17
C0030pk1	198	3247	0.94	5.00	3.56	<b>0.22</b>	<b>2.86</b>	<b>115.21</b>	0.45	3.43	1249.54
cass...-130731	281	4871	0.94	5.00	3.50	<b>0.21</b>	<b>3.03</b>	<b>256.58</b>	0.46	3.46	2029.28
helix_amber	392	6265	0.96	5.00	3.52	<b>0.26</b>	<b>3.30</b>	<b>388.29</b>	0.46	3.63	2600.84
avg						<b>0.14</b>	<b>1.61</b>	<b>59.76</b>	0.31	1.96	459.49
geo (CPU)								<b>25.43</b>			156.54

Table 2: Comparative results (MWU vs. MS) on the HARD instances of the DGP. The last line reports the average edge errors and the arithmetic and geometric mean of the CPU times.

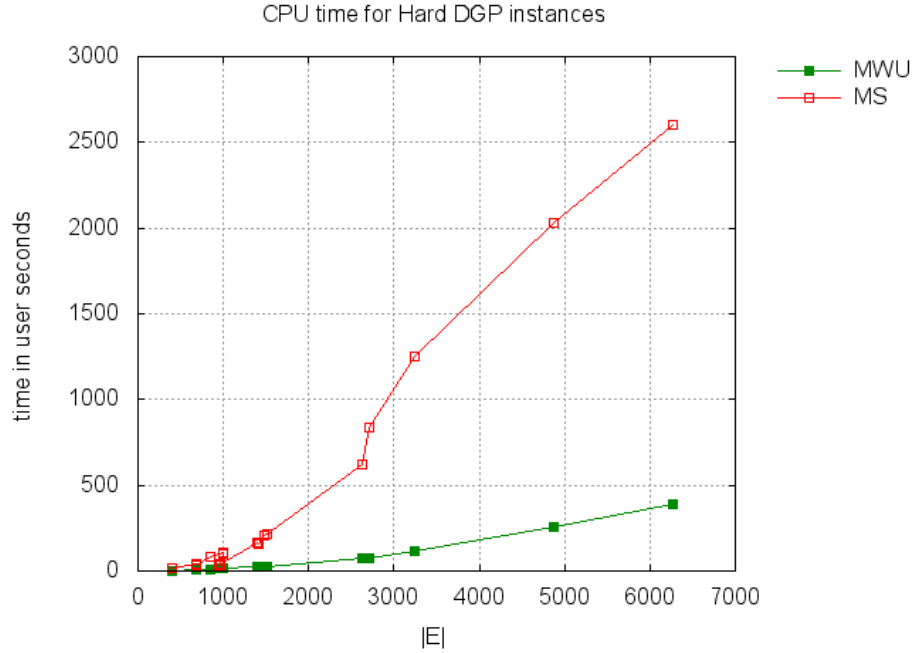


Figure 1: CPU time vs.  $m$ : MWU vs. MS.

both quality and CPU time. What is surprising, however, is that the MWU is often able to find a better maximum distance error. This measure is important in the context of proteins, since a single large maximum distance error might well mean a different conformation altogether (possibly closer to a different isomer), and therefore a protein with a different function. On the other hand, the average distance error is somewhat less important (as long as it remains low enough), as it might simply be due to experimental errors. Unlike the MWU, and similarly to MS, VNS has no approximation guarantee whatsoever, even a relative one.

Instance	$n$	$m$	$d$ :min	max	avg	MWU			VNS		
						err:avg	max	CPU	avg	max	CPU
2er1-frag-bp1	39	406	0.81	4.99	3.11	0.00	0.00	<b>2.63</b>	0.00	0.00	4.00
C0080create.1	60	681	0.98	5.00	3.33	0.04	0.41	<b>6.17</b>	<b>0.00</b>	<b>0.00</b>	9.19
C0080create.2	60	681	0.98	5.00	3.33	0.04	0.41	6.05	<b>0.00</b>	<b>0.00</b>	<b>0.7</b>
names	82	840	1.21	5.00	3.48	0.16	1.28	<b>10.91</b>	<b>0.01</b>	<b>0.63</b>	16.91
pept	107	999	1.16	5.00	3.55	0.16	<b>1.38</b>	15.53	<b>0.15</b>	1.91	<b>11.62</b>
C0020pdb	107	999	1.16	5.00	3.55	0.16	<b>1.45</b>	14.69	<b>0.12</b>	2.78	<b>6.44</b>
1guu-1	150	959	1.20	4.99	3.46	0.09	<b>0.79</b>	15.09	<b>0.04</b>	1.11	<b>1.46</b>
1guu-4000	150	968	0.33	4.99	3.51	<b>0.11</b>	19.31	<b>0.68</b>	<b>0.11</b>	1.08	<b>10.57</b>
1guu	150	955	1.30	5.00	3.46	0.09	<b>0.70</b>	15.33	<b>0.06</b>	1.00	<b>9.64</b>
res_5000	108	1392	0.94	5.00	3.42	0.11	<b>1.75</b>	22.79	<b>0.06</b>	2.57	<b>10.24</b>
res_2000	108	1404	0.95	5.00	3.42	0.17	<b>1.80</b>	21.46	<b>0.15</b>	1.87	<b>2.50</b>
res_0	108	1410	0.94	5.00	3.43	<b>0.12</b>	<b>2.07</b>	22.11	0.2	2.60	<b>2.32</b>
res_3000	108	1487	0.97	5.00	3.48	0.20	1.99	25.29	<b>0.01</b>	<b>0.86</b>	<b>20.11</b>
res_1000	108	1506	0.94	5.00	3.49	0.16	<b>1.71</b>	25.19	<b>0.14</b>	2.89	<b>10.49</b>
res_2kxa	177	2627	0.93	5.00	3.52	0.18	2.16	74.32	<b>0.00</b>	<b>0.00</b>	<b>20.11</b>
2kxa	177	2711	0.93	5.00	3.53	0.26	2.79	78.66	<b>0.01</b>	<b>1.26</b>	<b>20.86</b>
C0030pk1	198	3247	0.94	5.00	3.56	0.22	2.86	115.21	<b>0.00</b>	<b>0.01</b>	<b>7.80</b>
cass...-130731	281	4871	0.94	5.00	3.50	0.21	<b>3.03</b>	256.58	<b>0.12</b>	3.54	<b>21.09</b>
helix_amber	392	6265	0.96	5.00	3.52	0.26	<b>3.30</b>	388.29	<b>0.11</b>	4.52	<b>23.12</b>
avg						0.14	1.61	59.76	<b>0.07</b>	<b>1.51</b>	<b>11.02</b>
geo (CPU)								25.43			<b>7.85</b>

Table 3: Comparative results (MWU vs. VNS) on the HARD instances of the DGP. The last line reports the average edge errors and the arithmetic and geometric mean of the CPU times.

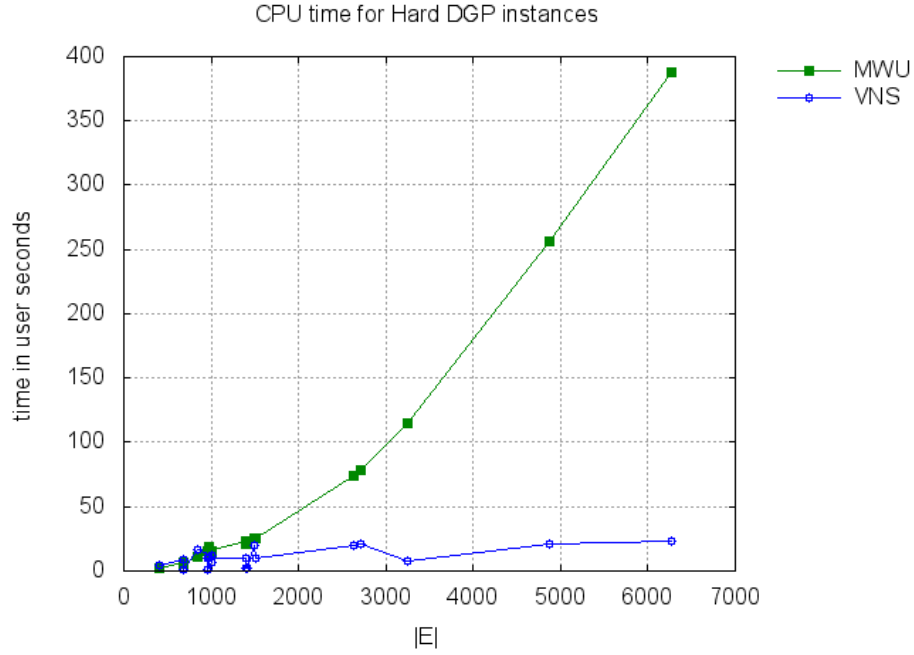


Figure 2: CPU time vs.  $m$ : MWU vs. VNS.

## 4.2 HUC results

In this section we present the results on the HUC application.

### 4.2.1 Test configuration

We again use the IPOPT solver [14] as the local NLP solver in both the MWU (Step 5 of Alg. 4) and MS (Step 3 of Alg. 1) methods; and CPLEX [24] as the LP solver for the pointwise reformulation of the MWU (Step 4 of Alg. 4). External solvers are invoked without time limits. Tests were executed on a machine configured with eight 64-bit Intel Xeon CPU E5504 running at 2.00 GHz and 11.7 GB of RAM, running the Linux operating system.

The authors of [11] provided us with three simplified instances, sharing the following common characteristics:

- $\bar{h} = 128, \tau = 1\text{h}$
- $V_0 = V_{\bar{h}} = 21078580, \underline{V} = 15000000, \bar{V} = 33000000\text{m}^3$
- $\bar{Q} = 42\text{m}^3/\text{s}$
- $Q^- = Q^+ = 70\text{m}^3/\text{s}/\text{hour}$
- $R_0 = 0.01$
- $\underline{L} = 385$
- $L = (L_\ell \mid \ell \in \{0, \dots, 6\}) =$   
 $= (4.0986, -1.2554, 0.1605, -9.762 \times 10^{-3}, 3.0943 \times 10^{-4}, -4.9293 \times 10^{-6}, 3.1152 \times 10^{-8})$
- $K = (K_k \mid k \in \{0, \dots, 6\}) =$   
 $= (3.074 \times 10^2, 3.88 \times 10^1, -4.37, 0.265, -8.87 \times 10^{-3}, 1.55 \times 10^{-4}, -1.11 \times 10^{-6})$

All of the the values of the inflows  $I$  ( $\text{m}^3/\text{s}$ ) and prices  $\Pi$  (currency/MWh) per time period are reported in Tab. 8 in the appendix.

In order to work with a larger test set, nine further instances were generated from each of the three original instances by uniformly sampling price vectors  $\Pi = (\Pi_h \mid h \in \{1, \dots, 168\})$  in  $[\min \Pi, \max \Pi]$ . The generated instances are noted A2 to A10, B2 to B10, C2 to C10. For example, instance C2 features the same data as C1 except for the prices, which, however, lay in the same range.

### 4.2.2 Comparative results on solution quality and CPU time

Both MWU and MS are configured with  $T = 20$  iterations. Tests are run on the (A1-C10) 30-instance set introduced in Sect. 4.2.1. The comparative computational results are reported in Tab. 4 as follows:

- the first column shows the instance name
- second and third columns show objective value and CPU time (in seconds of user time) for the MS algorithm
- the fourth and fifth columns show objective value and CPU time (in seconds of user time) for the MWU algorithm
- the fifth column shows relative objective value improvement from MS to MWU computed as

$$\Delta = \frac{\text{val}(\text{MWU}) - \text{val}(\text{MS})}{\text{val}(\text{MS})} \quad (4.1)$$

- the sixth column shows time improvement ratio from MS to MWU computed as

$$\Lambda = \frac{\text{cpu}(\text{MS})}{\text{cpu}(\text{MWU})}. \quad (4.2)$$

For the last two columns, the comparison metrics are summarized in the last line with the average (avg) and the standard deviation (std) across all 30 instances.

Instance	MS		MWU		MS vs. MWU	
	objective	CPU	objective	CPU	$\Delta$	$\Lambda$
A1	4.12E+4	30.2	<b>4.17E+4</b>	<b>5.7</b>	1.27%	5.35
A2	5.18E+4	33.4	<b>5.32E+4</b>	<b>5.7</b>	2.59%	5.89
A3	4.88E+4	33.4	<b>4.97E+4</b>	<b>5.6</b>	1.80%	5.99
A4	4.86E+4	31.8	<b>4.98E+4</b>	<b>5.9</b>	2.56%	5.42
A5	5.01E+4	34.4	<b>5.19E+4</b>	<b>5.5</b>	3.63%	6.29
A6	5.00E+4	29.8	<b>5.10E+4</b>	<b>6.1</b>	2.11%	4.91
A7	5.10E+4	33.0	<b>5.20E+4</b>	<b>5.8</b>	1.97%	5.68
A8	5.07E+4	33.2	<b>5.24E+4</b>	<b>5.8</b>	3.37%	5.69
A9	5.08E+4	32.7	<b>5.21E+4</b>	<b>5.7</b>	2.48%	5.71
A10	5.04E+4	32.0	<b>5.13E+4</b>	<b>6.0</b>	1.61%	5.32
B1	1.98E+4	29.5	1.98E+4	<b>4.4</b>	0.00%	6.70
B2	2.67E+4	33.3	2.67E+4	<b>5.0</b>	0.00%	6.64
B3	2.53E+4	30.7	2.53E+4	<b>4.4</b>	0.00%	7.03
B4	2.53E+4	29.2	2.53E+4	<b>4.7</b>	0.00%	6.22
B5	2.60E+4	30.2	2.60E+4	<b>5.0</b>	0.00%	6.05
B6	2.60E+4	28.2	2.60E+4	<b>4.6</b>	0.00%	6.11
B7	2.62E+4	28.6	2.62E+4	<b>4.4</b>	0.00%	6.54
B8	2.65E+4	29.6	2.65E+4	<b>4.5</b>	0.00%	6.54
B9	2.62E+4	29.5	2.62E+4	<b>4.8</b>	0.00%	6.20
B10	2.61E+4	31.8	2.61E+4	<b>4.5</b>	0.00%	7.08
C1	5.09E+4	26.8	<b>5.17E+4</b>	<b>5.0</b>	1.60%	5.38
C2	6.82E+4	33.0	<b>6.99E+4</b>	<b>5.7</b>	2.53%	5.82
C3	6.51E+4	33.6	<b>6.65E+4</b>	<b>5.4</b>	2.06%	6.29
C4	6.58E+4	31.9	<b>6.70E+4</b>	<b>5.5</b>	1.85%	5.77
C5	6.70E+4	33.1	<b>6.87E+4</b>	<b>5.0</b>	2.43%	6.62
C6	6.64E+4	30.2	<b>6.82E+4</b>	<b>5.0</b>	2.77%	6.07
C7	6.83E+4	31.3	<b>6.87E+4</b>	<b>4.8</b>	0.55%	6.51
C8	6.83E+4	33.2	<b>6.86E+4</b>	<b>5.0</b>	0.44%	6.69
C9	6.72E+4	30.9	<b>6.88E+4</b>	<b>4.7</b>	2.27%	6.62
C10	6.77E+4	32.5	<b>6.84E+4</b>	<b>5.3</b>	1.03%	6.14
avg	4.67E+4	31.4	<b>4.75E+4</b>	<b>5.2</b>	<b>1.36%</b>	<b>6.11</b>
std	1.71E+4	1.9	1.76E+4	0.5	1.19%	0.54
geom	4.33E+4	31.3	<b>4.39E+4</b>	<b>5.1</b>	NA	<b>6.09</b>

Table 4: Objective and CPU time of MS and MWU, relative objective improvement  $\Delta$  and CPU time improvement  $\Lambda$  from MS to MWU.

The MWU algorithm is always better than the MS algorithm objective-wise but the improvement is relatively small. It must be emphasized, however, that a few percentage points in the objective functions of energy-related optimization problems often translate in consistent savings in absolute terms. As for the time performance, the MWU algorithm is six times faster than the MS algorithm on average, with low variability: this is an extremely desirable feature in short-term unit commitment problems such as this.

### 4.2.3 Sensitivity to instance size

In order to study in the relative performance sensitivity of the MWU algorithm on instance size, we artificially vary the time horizon  $\bar{h}$  of the instances from one day to two weeks.

The instances introduced in Sect. 4.2.1 are one week long with hourly time steps ( $\bar{h} = 168$ ); we now consider instances with  $\bar{h}$  in  $\{24, 48, \dots, 168, \dots, 336\}$ , defined as follows:

- when  $\bar{h} \leq 168$ , data from the 168-long instances is cropped to  $h \in [1, \bar{h}]$ ;
- when  $\bar{h} > 168$ , data for  $h \in [1, 168]$  is identical to the 168-long instances, and data for  $h \in [169, \bar{h}]$  is duplicated from the first interval and taken from time period  $h - 168 \in [1, 168]$ .

Both MWU and MS are configured with  $T = 20$ . For each instance subset A (A1 to A10), B (B1 to B10) and C (C1 to C10), the comparative computational results for the 14 different sizes are summarized by average and standard deviation in Tab. 5, as follows:

- the first column shows the instance subset
- the second column shows instance size  $\bar{h}$

- the third and fourth columns show average (avg) and standard deviation (std) of the CPU time improvement  $\Lambda$  taken over the 10 instances in the first column having size specified in the second
- the fifth and sixth columns show average (avg) and standard deviation (std) of the relative objective improvement  $\Delta$  taken over the 10 instances in the first column having size specified in the second

For the last three columns, the last line shows average, standard deviation and geometric mean of the corresponding metrics taken over the whole 420 ( $= 3 \times 10 \times 14$ ) tested instances.

Instance subset	Instance size $\bar{h}$	$\Delta$		$\Lambda$		
		avg	std	avg	std	geo
A	24	<b>0.38%</b>	7.73%	<b>2.76</b>	0.17	<b>2.75</b>
	48	<b>3.10%</b>	1.05%	<b>3.82</b>	0.17	<b>3.82</b>
	72	<b>3.19%</b>	1.97%	<b>4.46</b>	0.21	<b>4.46</b>
	96	<b>2.30%</b>	0.88%	<b>5.12</b>	0.23	<b>5.11</b>
	120	<b>3.67%</b>	1.32%	<b>5.11</b>	0.23	<b>5.11</b>
	144	<b>1.81%</b>	1.22%	<b>5.62</b>	0.35	<b>5.61</b>
	168	<b>2.23%</b>	0.66%	<b>5.73</b>	0.29	<b>5.73</b>
	192	<b>2.73%</b>	0.98%	<b>6.18</b>	0.33	<b>6.17</b>
	216	<b>2.40%</b>	1.11%	<b>6.39</b>	0.37	<b>6.38</b>
	240	<b>1.86%</b>	1.25%	<b>6.48</b>	0.44	<b>6.47</b>
	264	<b>2.12%</b>	0.54%	<b>7.18</b>	0.40	<b>7.17</b>
	288	<b>2.30%</b>	0.74%	<b>6.77</b>	0.45	<b>6.76</b>
	312	<b>1.84%</b>	0.73%	<b>6.65</b>	0.39	<b>6.64</b>
	336	<b>1.78%</b>	0.91%	<b>7.01</b>	0.50	<b>7.00</b>
B	24	0.00%	0.00%	<b>3.19</b>	0.24	<b>3.18</b>
	48	0.00%	0.00%	<b>4.87</b>	0.27	<b>4.86</b>
	72	0.00%	0.00%	<b>5.92</b>	0.38	<b>5.91</b>
	96	0.00%	0.00%	<b>6.51</b>	0.38	<b>6.50</b>
	120	0.00%	0.00%	<b>6.31</b>	0.29	<b>6.31</b>
	144	0.00%	0.00%	<b>6.75</b>	0.46	<b>6.74</b>
	168	0.00%	0.00%	<b>7.01</b>	0.46	<b>6.99</b>
	192	0.00%	0.00%	<b>7.41</b>	0.72	<b>7.38</b>
	216	0.00%	0.00%	<b>7.36</b>	0.71	<b>7.33</b>
	240	0.00%	0.00%	<b>7.40</b>	0.31	<b>7.40</b>
	264	0.00%	0.00%	<b>7.68</b>	0.52	<b>7.67</b>
	288	0.00%	0.00%	<b>7.60</b>	0.98	<b>7.54</b>
	312	0.00%	0.00%	<b>7.92</b>	0.75	<b>7.88</b>
	336	0.00%	0.00%	<b>8.45</b>	0.82	<b>8.41</b>
C	24	0.00%	0.00%	<b>3.06</b>	0.15	<b>3.05</b>
	48	<b>2.08%</b>	1.87%	<b>4.31</b>	0.21	<b>4.31</b>
	72	<b>2.59%</b>	2.11%	<b>4.77</b>	0.33	<b>4.76</b>
	96	<b>2.99%</b>	1.01%	<b>5.17</b>	0.19	<b>5.16</b>
	120	<b>3.30%</b>	1.86%	<b>5.47</b>	0.23	<b>5.47</b>
	144	<b>2.75%</b>	0.99%	<b>5.98</b>	0.30	<b>5.97</b>
	168	<b>1.89%</b>	0.72%	<b>6.59</b>	0.39	<b>6.58</b>
	192	<b>1.81%</b>	0.96%	<b>6.74</b>	0.60	<b>6.72</b>
	216	<b>1.08%</b>	0.76%	<b>6.53</b>	0.44	<b>6.52</b>
	240	<b>1.70%</b>	1.34%	<b>7.41</b>	0.46	<b>7.40</b>
	264	<b>1.20%</b>	0.82%	<b>7.32</b>	0.85	<b>7.28</b>
	288	<b>1.22%</b>	0.81%	<b>7.29</b>	0.54	<b>7.27</b>
	312	<b>1.46%</b>	0.85%	<b>7.31</b>	0.45	<b>7.30</b>
	336	<b>0.62%</b>	1.03%	<b>7.15</b>	0.46	<b>7.14</b>
	avg	<b>1.34%</b>	0.86%	<b>6.16</b>	0.41	<b>6.15</b>
	std	1.20%	1.25%	1.38	0.20	1.37
	geo	NA	NA	<b>5.98</b>	0.37	<b>5.97</b>

Table 5: Relative objective improvement  $\Delta$  and CPU time improvement  $\Lambda$  from MS to MWU, according to instance subset and instance size  $\bar{h}$ .

The results from Tab. 5 corroborate those from Tab. 4: the MWU algorithm almost always outputs a slightly better objective than the MS algorithm while consistently outperforming the MS algorithm in CPU time.

We graphically compare CPU times averaged over instance sizes in Fig. 3. Each point corresponds to (size, average CPU time) over all the instances A1-C10.



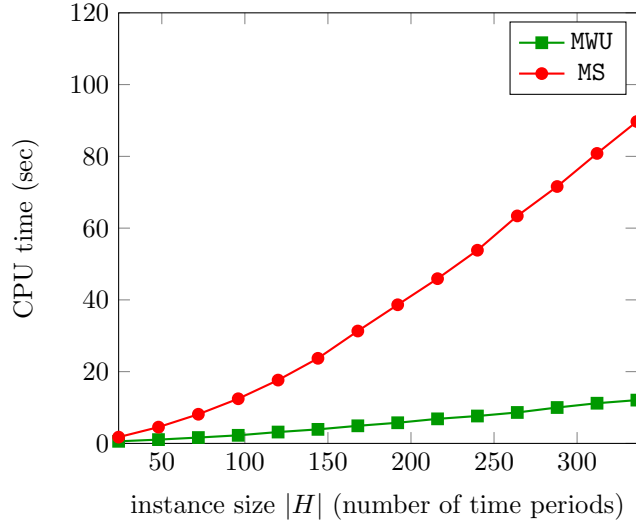


Figure 3: Average of CPU time (sec) vs. instance size  $|H| = \bar{h}$ . Each point represents a CPU time average over three equally-sized instances (in groups A, B, C).

#### 4.2.4 Influence of initialization

In this section we empirically show that the MWU method is robust to varying initial conditions. We define the *objective dispersion*  $\Xi$  as the standard deviation over the average of a sample of objective function values collected from a sequence of runs with randomly chosen starting vectors  $\theta$ .

The MWU search is somewhat diversified due to the random sampling of  $\theta$  (Step 3 of Alg. 4), which is used to define the pointwise reformulation. Whenever the  $\psi$  costs are only defined through feasibility (i.e.  $\psi$  is proportional to  $\beta$ , see Sect. 3.3), it is easy to show that the weights  $\omega$  can only decrease during MWU execution (Step 8 of Alg. 4), which means the sampling domain is increasingly small. In the extreme case where all weights  $\omega$  are set to zero for all iterations, the search is deterministic and only depends the values initially set for parameters  $\theta$ . We therefore test whether MWU results are conditioned by initialization.

To this end, we look at the dispersion of 20 MWU runs (all of them configured with  $T = 20$ ), started from with 20 different randomly sampled vectors  $\theta$ . In addition, for each sampled initial point, a local descent is independently performed to solve problem Eq. (2.19); the best of these descents is equivalent to the result of a 20-iteration MS run.

Tests are run on the (A1-C10) 30-instance set introduced in Sect. 4.2.1 (with  $\bar{h} = 168$  time periods). Computational results are reported in Tab. 6, as follows:

- the first column shows the instance name
- the second column shows the objective dispersion  $\Xi$
- the third column shows average objective function value improvements from MS to each of the 20 MWU runs ( $\Delta_{\text{avg}}$ )
- the fourth column indicates whether the worst objective function value obtained over the 20 MWU runs is as good as the MS result.

The comparison metrics are summarized in the last line with the average across all 30 instances.

On average, the dispersion of the MWU results is low ( $\text{avg}(\Xi) = 0.38\%$ ), which shows low sensitivity to variability in initialization. In addition, the dispersion of the MWU results is lower on average than the average relative improvement ( $\text{avg}(\Delta_{\text{avg}}) = 1.17\%$ ), thus showing that the MWU is consistently better

Instance	$\Xi$	$\Delta_{\text{avg}}$	MWU $\geq$ MS
A1	0.40%	0.08%	false
A2	0.63%	2.50%	true
A3	0.56%	2.76%	true
A4	0.83%	1.87%	true
A5	0.67%	2.02%	true
A6	0.33%	0.20%	false
A7	0.69%	3.93%	true
A8	0.57%	1.53%	true
A9	0.41%	2.17%	true
A10	0.61%	2.02%	true
B1	0.00%	0.00%	true
B2	0.00%	0.00%	true
B3	0.00%	0.00%	true
B4	0.00%	0.00%	true
B5	0.00%	0.00%	true
B6	0.00%	0.00%	true
B7	0.00%	0.00%	true
B8	0.00%	0.00%	true
B9	0.00%	0.00%	true
B10	0.00%	0.00%	true
C1	0.41%	1.43%	true
C2	0.63%	3.44%	true
C3	0.45%	1.70%	true
C4	0.70%	3.96%	true
C5	0.46%	1.40%	true
C6	0.52%	0.48%	false
C7	0.57%	2.47%	true
C8	0.70%	0.59%	false
C9	0.56%	0.93%	false
C10	0.58%	0.54%	false
avg	<b>0.38%</b>	<b>1.20%</b>	<b>80%</b>

Table 6: Objective dispersion  $\Xi$ , average relative objective value improvement from MS  $\Delta_{\text{avg}}$ , and worst-case objective value comparison to MS (last column) for the 20 MWU runs.

than MS; this comparison is all the more relevant as the 20 MWU runs and the 20 MS iterations are started with the same values. Similarly, four out of five times (80%) even the worst MWU run yields a result as good as the MS.

Therefore, the MWU method does not suffer from a mitigated diversification compared to the MS method, and is robust to varying initial conditions.

#### 4.2.5 Importance of the pointwise step

In this section we show that solving the pointwise problem within the MWU method (Step 4 of Alg. 4) contributes to its effectiveness compared to the MS.

Both MWU and MS feature a randomly started local descent. On the one hand, the initialization phase for MS is based on sampling from a uniform distribution. On the other hand, the MWU has a specific routine to sample parameters, which are then used to solve the pointwise problem Eq. (2.20), the solution of which is employed as a starting point to a local descent (refinement step) for the original problem Eq. (2.19). In the MWU loop applied to the HUC, a solution of the pointwise problem is feasible for the original problem (except for the dependent variables  $y$ ), whereas the initial point randomly sampled in MS may be infeasible. To test if the latter characteristic is responsible for the very good relative time performances of the MWU, we compare results against an enhanced MS with a pointwise feasibility recovery step, as described in Alg. 5.

Both MWU and MS<sub>ptw</sub> are configured with  $T = 20$ . Tests are run on the (A1-C10) 30-instance set introduced in Sect. 4.2.1, (with  $\bar{h} = 168$  time periods). The comparative computational results are reported in Tab. 7 as follows:

- the first column shows instance name
- the second and third columns show objective value and CPU time (in seconds) for the MS<sub>PTW</sub>

**Algorithm 5** MSptw( $P$ )

---

```

1: while  $t \leq T$  do
2:   sample  $\theta^t$  uniformly at random
3:   solve ptw ( $P$ ), get solution  $x^t$ 
       $t \leftarrow t'(\theta^t)$ 
4:   refine  $x^t$  (e.g. using local descent)
5:   if  $x^t$  is better than the incumbent, replace  $x^* \leftarrow x^t$ 
6:   increase  $t$ 
7: end while

```

---

algorithm

- the fourth and fifth columns show objective value and CPU time (in seconds) for the MWU algorithm
- the fifth column shows relative objective value improvement  $\Delta$  from MSptw to MWU
- the sixth column shows time improvement ratio  $\Lambda$  from MSptw to MWU.

For the last two columns, the comparison metrics are summarized in the last line with the average (avg) and the standard deviation (std) across all 30 instances.

Instance	MWU		MSptw		MWU vs. $\Delta$	MSptw $\Lambda$
	objective	CPU	objective	CPU		
A1	4.17E+4	<b>5.48</b>	<b>4.18E+4</b>	7.93	-0.27%	1.45
A2	<b>5.32E+4</b>	<b>5.22</b>	5.19E+4	8.02	2.40%	1.54
A3	<b>4.97E+4</b>	<b>5.29</b>	4.95E+4	8.51	0.49%	1.61
A4	<b>4.98E+4</b>	<b>5.51</b>	4.90E+4	8.3	1.74%	1.51
A5	<b>5.19E+4</b>	<b>5.35</b>	5.06E+4	8.86	2.56%	1.66
A6	<b>5.10E+4</b>	<b>5.49</b>	5.06E+4	8.36	0.81%	1.52
A7	<b>5.20E+4</b>	<b>5.3</b>	5.11E+4	8.3	1.84%	1.57
A8	<b>5.24E+4</b>	<b>5.38</b>	5.17E+4	7.83	1.36%	1.46
A9	<b>5.21E+4</b>	<b>5.33</b>	5.14E+4	8.18	1.28%	1.53
A10	<b>5.13E+4</b>	<b>5.57</b>	5.04E+4	8.54	1.64%	1.53
B1	1.98E+4	<b>4.2</b>	1.98E+4	7.33	0.00%	1.75
B2	2.67E+4	<b>4.57</b>	2.67E+4	7.52	0.00%	1.65
B3	2.53E+4	<b>4.11</b>	2.53E+4	7.64	0.00%	1.86
B4	2.53E+4	<b>4.37</b>	2.53E+4	7.21	0.00%	1.65
B5	2.60E+4	<b>4.8</b>	2.60E+4	8.28	0.00%	1.73
B6	2.60E+4	<b>4.35</b>	2.60E+4	7.8	0.00%	1.79
B7	2.62E+4	<b>4.1</b>	2.62E+4	7.8	0.00%	1.90
B8	2.65E+4	<b>4.28</b>	2.65E+4	7.97	0.00%	1.86
B9	2.62E+4	<b>4.65</b>	2.62E+4	8.04	0.00%	1.73
B10	2.61E+4	<b>4.33</b>	2.61E+4	7.25	0.00%	1.67
C1	<b>5.17E+4</b>	4.57	5.16E+4	6.7	0.09%	1.47
C2	<b>6.99E+4</b>	<b>5.3</b>	6.84E+4	8.26	2.11%	1.56
C3	<b>6.65E+4</b>	<b>5.26</b>	6.57E+4	9.05	1.12%	1.72
C4	<b>6.70E+4</b>	<b>5.64</b>	6.54E+4	7.97	2.51%	1.41
C5	6.87E+4	<b>5</b>	<b>6.88E+4</b>	7.9	-0.11%	1.58
C6	<b>6.82E+4</b>	<b>4.9</b>	6.71E+4	8.25	1.77%	1.68
C7	<b>6.87E+4</b>	<b>5.08</b>	6.67E+4	8.54	3.02%	1.68
C8	<b>6.86E+4</b>	<b>5.03</b>	6.79E+4	8.02	1.06%	1.59
C9	<b>6.88E+4</b>	<b>4.7</b>	6.73E+4	8.18	2.25%	1.74
C10	<b>6.84E+4</b>	<b>5.39</b>	6.78E+4	8.82	0.96%	1.64
avg	<b>4.75E+4</b>	<b>4.95</b>	4.70E+4	8.05	<b>0.95%</b>	<b>1.63</b>
std	1.76E+4	0.49	1.72E+4	0.52	1.02%	0.13
geo	<b>4.39E+4</b>	<b>4.93</b>	4.35E+4	<b>8.03</b>	NA	<b>1.63</b>

Table 7: Objective and CPU time of MWU and MS with ptw initialization, relative objective improvement  $\Delta$  and CPU time improvement  $\Lambda$  from MSptw to MWU.

The MWU method still outperforms the enhanced MSptw method both as regards the objective function value and the CPU time. However, the improvement margin is less marked as with plain MS (see tab 4). Solving a pointwise reformulation to quickly obtain a feasible solution for local descent serves the effectiveness of the MWU method (this is all the more so if the pointwise reformulation is efficient).

### 4.3 MVPS results

The test set consists in the 20 real-world instances described in [13], publicly available from OR-Library [4, 5] on the web site <http://www.brunel.ac.uk/~mastjjb/jeb/info.html>, differing for the number of assets and for value for the risk level  $\sigma$ . We imposed, as in [13],  $\underline{x} = 0.01$ ,  $\bar{x} = 1$  and  $K = 10$ .

#### 4.3.1 Transaction cost functions, ranked by nonlinearity

We consider different transaction cost functions, which represent different categories of “nonlinearity”. In particular, we analyze the performance of the MWU algorithm with respect to the following five univariate functions (see Fig. 4):

- (a)  $C_i(x_i) = -\rho_i \ln\left(\frac{20-0.06(1+x_i)}{1+x_i}\right)$  for all  $i \leq n$ : this function is increasing, concave and “almost linear”.
- (b)  $C_i(x_i) = -\rho_i \ln\left(\frac{0.2-0.01(0.00001+x_i)}{0.00001+x_i}\right)$  for all  $i \leq n$ : this function is concave and replicates the behavior of the transaction cost function described in [26].
- (c)  $C_i(x_i) = \rho_i(4x_i + 0.12 \sin(40x_i))$  for all  $i \leq n$ : this function has a sinusoidal behavior similar to a step function.
- (d)  $C_i(x_i) = \rho_i(4x_i + 0.3 \sin(40x_i))$  for all  $i \leq n$ : this function is similar to the one in (c) but with a “stronger nonlinear behavior”.
- (e)  $C_i(x_i) = \rho_i(0.5x_i + \sin(50x_i))$  for all  $i \leq n$ : this is the “most nonlinear” transaction cost function among which we tested the methods.

We were careful to use quotes around this informally held quantitative view of the nonlinearity of a function. Since these functions are univariate, we trust most readers will agree with our categorization, by inspection of Fig. 4.

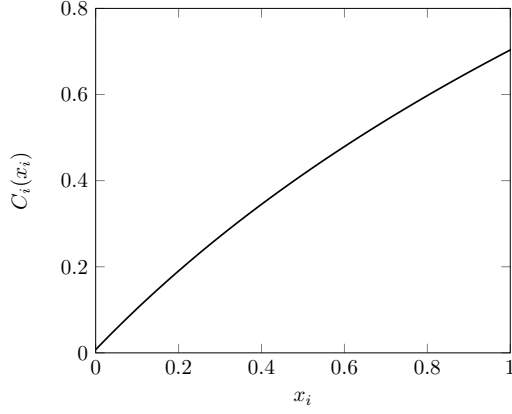
#### 4.3.2 Configuration of the computational platform

We use  $T = 20$  iterations of both MWU and MS. We use BONMIN [9] as the local MINLP solver for both methods, with time limit equal to 600 seconds. Specifically, we employ BONMIN’s native Branch-and-Bound (B-BB) algorithm [22, 10], since it is generally more stable for nonconvex MINLPs. We use CPLEX [24] as the convex MIQP solver for the pointwise reformulation Eq. (2.24), with a 600s time limit, using only one thread. All of the computational experiments were performed on an Intel Xeon CPU E5649, 2.53GHz, using only one processor.

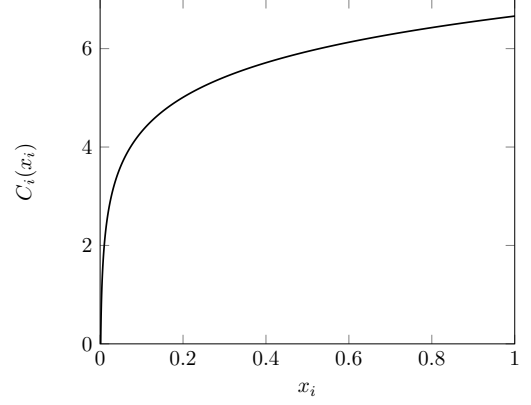
#### 4.3.3 Localization of the MS subsolver

Since the MVPS problem is the only application involving integer variables, a further discussion is in order.

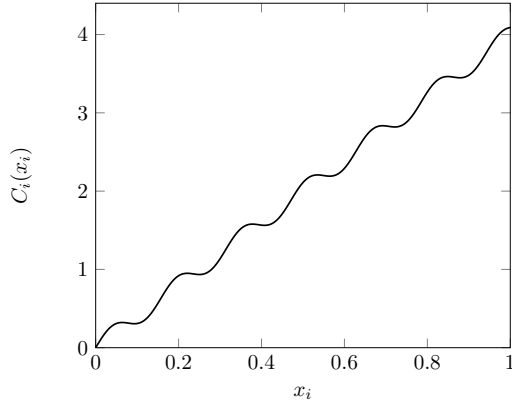
The MS algorithm is based on randomly sampling a starting point, and performing a local optimization starting from that point. On the other hand, since B-BB is a local optimization procedure when deployed on nonconvex MINLP, the only influence of the starting point is that it sets a cut-off value for the optimum. From some preliminary tests, BONMIN behaves more like a global solver on our test-set than a local one, and essentially defies the purpose of comparing the MWU against the MS.



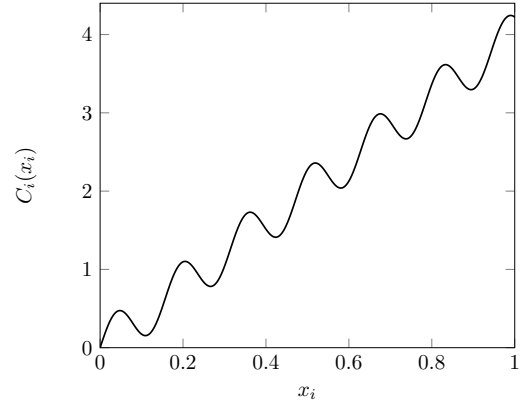
(a) Easy concave transaction costs.



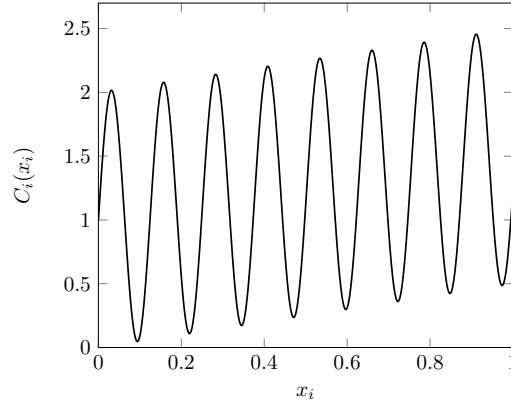
(b) Hard concave transaction costs.



(c) Easy trigonometric transaction costs.



(d) Medium trigonometric transaction costs.



(e) Hard trigonometric transaction costs.

Figure 4: Examples of transaction cost functions.

In order to turn the B-BB into a truly local solver, we added a local branching constraint [20] to the formulation, which limits the amount of flips of the binary variables to a fixed constant  $\lfloor \nu n \rfloor$ , where  $\nu \in [0, 1]$ :

$$\sum_{\substack{i \leq n \\ y'_i = 0}} y_i + \sum_{\substack{i \leq n \\ y'_i = 1}} (1 - y_i) \leq \lfloor \nu n \rfloor, \quad (4.3)$$

where  $y'$  is the starting point. This enforces a local exploration in combinatorial neighbourhood of the

starting point  $y'$ . Some trial and error yielded a very high threshold  $\nu = 0.96$  (lower values made the instance infeasible excessively often).

#### 4.3.4 Summary of results

The results (see Fig. 5) seem to confirm a preliminary observation that MWU finds better solutions than MS for problems involving “very nonlinear” functions. With transaction costs (a)-(b), MS performs better than MWU; with respect to the transaction costs (c) and (e), however, MWU performs better than MS. In particular, for transaction costs (c)-(d) the average value of relative objective value improvement from MS to MWU is considerably high.

Another interesting (and unexpected) observation concerning solution quality is that, on average, the number of assets in the best portfolio produced by the MWU (a secondary goal when solving such portfolio problems in real life) is smaller with regards to the MS. At the moment we are unable to explain why, so we leave this as an open issue.

As for the geometric mean for the CPU time, MS shows an advantage only in case (a), all the other cases being dominated by MWU.

## 5 Conclusion

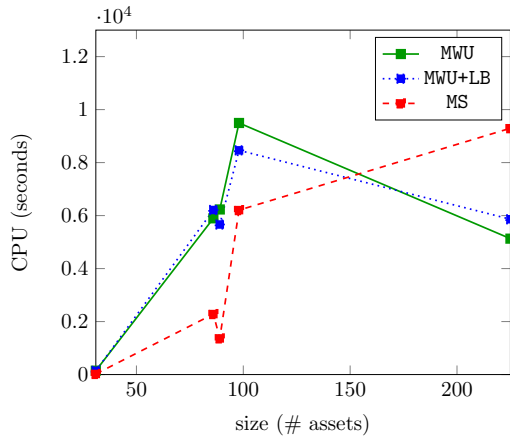
This paper is about the adaptation of the multiplicative weights algorithm to the (nonconvex) nonlinear and mixed integer nonlinear programming setting, based on a particular parametrized reformulation of the problem. Though this algorithm has been previously employed as a theoretical tool to derive approximation algorithms, we decided to benchmark it computationally to solve three classes of hard problems from different application fields. We found it compares quite favorably to the well-known multi-start method, which, unlike the MWU, offers no approximation guarantee.

## Acknowledgments

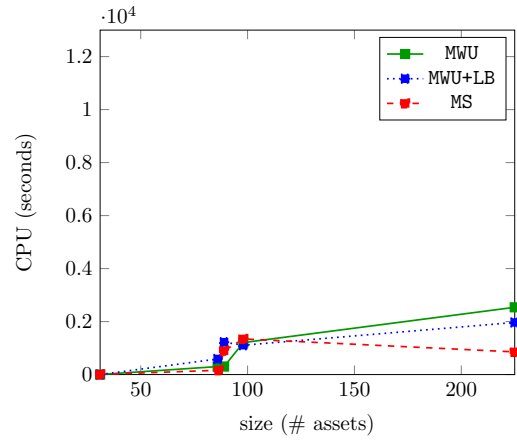
We are very grateful to Dr. Pascale Bendotti (EDF) for useful suggestions about the HUC problem. Luca Mencarelli is sponsored by a Ph.D. Fellowship from the FP7 Marie Curie ITN “MINO” project. Youcef Sahraoui is sponsored by a CIFRE Ph.D. Fellowship with Électricité De France (EDF). Leo Liberti was partly sponsored by the ANR Bip:Bip project under contract ANR-10-BINF-0003, and completed this work during a visiting term at IMECC, University of Campinas (SP), Brazil, sponsored by the Chaires Françaises dans l'état de São Paulo (CFSP) program, a collaboration between the French Consulate in São Paulo and the three main universities in São Paulo: UNICAMP, USP and UNESP.

## References

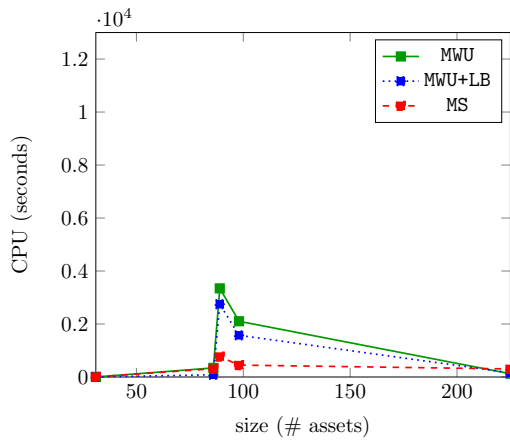
- [1] Arora, S., Hazan, E., Kale, S.: Fast algorithms for approximate semidefinite programming using the multiplicative weights update method. In: Foundations of Computer Science, *FOCS*, vol. 46, pp. 339–348. IEEE (2005)
- [2] Arora, S., Hazan, E., Kale, S.: The multiplicative weights update method: a meta-algorithm and applications. *Theory of Computing* **8**, 121–164 (2012)
- [3] Bahr, A., Leonard, J., Fallon, M.: Cooperative localization for autonomous underwater vehicles. *International Journal of Robotics Research* **28**(6), 714–728 (2009)



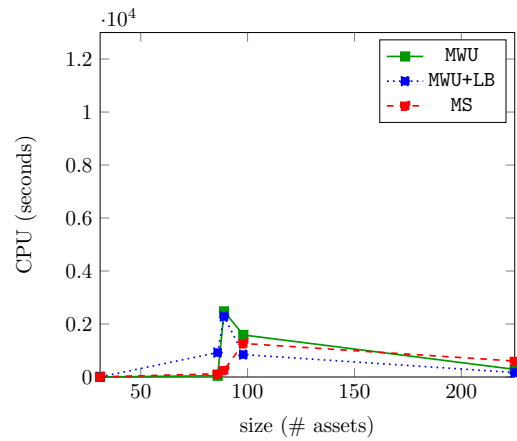
(a) "Easy" concave transaction costs.



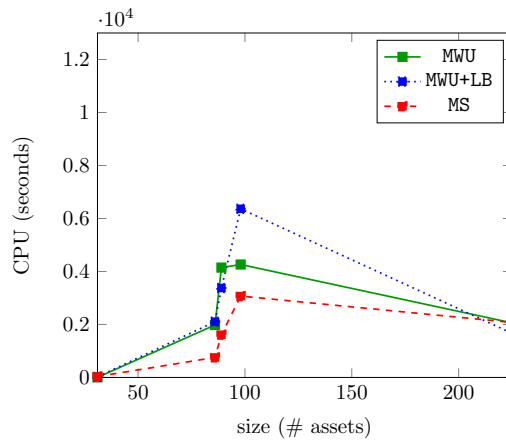
(b) "Hard" concave transaction costs.



(c) "Easy" trigonometric transaction costs.



(d) "Medium" trigonometric transaction costs.



(e) "Hard" trigonometric transaction costs.

Figure 5: CPU time vs. size of the problem  $n$  (# assets).

[4] Beasley, J.: OR-Library: Distributing test problems by electronic mail. *Journal of the Operational Research Society* **41**(11), 1069–1072 (1990)

[5] Beasley, J.: Obtaining test problems via internet. *Journal of Global Optimization* **8**(4), 429–433

- (1996)
- [6] Beeker, N., Gaubert, S., Glusa, C., Liberti, L.: Is the distance geometry problem in **NP**? In: A. Mucherino, C. Lavor, L. Liberti, N. Maculan (eds.) *Distance Geometry: Theory, Methods, and Applications*. Springer, New York (2013)
  - [7] Berman, H., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T., Weissig, H., Shindyalov, I., Bourne, P.: The protein data bank. *Nucleic Acid Research* **28**, 235–242 (2000)
  - [8] Bienstock, D.: Computational study of a family of mixed-integer quadratic programming problems. *Mathematical Programming* **74**(2), 121–140 (1996)
  - [9] Bonami, P., Lee, J.: *BONMIN user’s manual*. Tech. rep., IBM Corporation (2007)
  - [10] Bonami, P., Lee, J., Leyffer, S., Waechter, A.: More Branch-and-Bound experiments in convex nonlinear integer programming (2011). Preprint ANL/MCS-P1949-0911, Argonne National Laboratory, Mathematics and Computer Science Division
  - [11] Borghetti, A., D’Ambrosio, C., Lodi, A., Martello, S.: Optimal scheduling of a multiunit hydro power station in a short-term planning horizon. In: K.G. Murty (ed.) *Case Studies in Operations Research, International Series in Operations Research & Management Science*, vol. 212, pp. 167–181. Springer New York (2015). DOI 10.1007/978-1-4939-1007-6\_8. URL [http://dx.doi.org/10.1007/978-1-4939-1007-6\\_8](http://dx.doi.org/10.1007/978-1-4939-1007-6_8)
  - [12] Boyd, S., Vandenberghe, L.: *Convex Optimization*. Cambridge University Press, Cambridge (2004)
  - [13] Chang, T.J., Meade, N., Beasley, J., Sharaiha, Y.: Heuristics for cardinality constrained portfolio optimization. *Computers and Operations Research* **27**(13), 1271–1302 (2000)
  - [14] COIN-OR: Introduction to IPOPT: A tutorial for downloading, installing, and using IPOPT (2006)
  - [15] Costa, A., Hansen, P., Liberti, L.: Formulation symmetries in circle packing. In: R. Mahjoub (ed.) *Proceedings of the International Symposium on Combinatorial Optimization, Electronic Notes in Discrete Mathematics*, vol. 36, pp. 1303–1310. Elsevier, Amsterdam (2010)
  - [16] D’Ambrosio, C., Ky, V., Lavor, C., Liberti, L., Maculan, N.: Solving distance geometry problems with interval data using formulation-based methods. Tech. rep., LIX Ecole Polytechnique (working paper) (2014)
  - [17] D’Ambrosio, C., Mencarelli, L.: Complex portfolio selection via convex mixed-integer quadratic approaches: A survey. Tech. rep., LIX, École Polytechnique (working paper) (2014)
  - [18] Ding, Y., Krislock, N., Qian, J., Wolkowicz, H.: Sensor network localization, Euclidean distance matrix completions, and graph realization. *Optimization and Engineering* **11**, 45–66 (2010)
  - [19] Du, H., Alechina, N., Stock, K., Jackson, M.: The logic of NEAR and FAR. In: T.T. et al. (ed.) *COSIT, LNCS*, vol. 8116, pp. 475–494. Springer, Switzerland (2013)
  - [20] Fischetti, M., Lodi, A.: Local branching. *Mathematical Programming, Series B* **98**(1–3), 23–47 (2003)
  - [21] Frangioni, A., Gentile, C.: Perspective cuts for a class of convex 0-1 mixed integer programs. *Mathematical Programming, Series A* **106**(2), 225–236 (2006)
  - [22] Gupta, O., Ravindran, A.: Branch-and-Bound experiments in convex nonlinear integer programming. *Management Science* **31**(12), 1533–1546 (1985)
  - [23] Hansen, P., Mladenović, N.: Variable neighbourhood search: Principles and applications. *European Journal of Operations Research* **130**, 449–467 (2001)
  - [24] IBM: *ILOG CPLEX 12.2 User’s Manual*. IBM (2010)



- [25] Kannan, R., Monma, C.: On the computational complexity of integer programming problems. In: R. Henn, B. Korte, W. Oettli (eds.) *Optimization and Operations Research, Lecture Notes in Economics and Mathematical Systems*, vol. 157, pp. 161–172. Springer–Verlag (1978)
- [26] Konno, H., Wiyayanayake, A.: Portfolio optimization problem under concave transaction costs and minimal transaction unit constraints. *Mathematical Programming, Series B* **89**(2), 233–250 (2001)
- [27] Lavor, C., Liberti, L., Maculan, N.: Computational experience with the molecular distance geometry problem. In: J. Pintér (ed.) *Global Optimization: Scientific and Engineering Case Studies*, pp. 213–225. Springer, Berlin (2006)
- [28] Liberti, L.: Reformulations in mathematical programming: Definitions and systematics. *RAIRO-RO* **43**(1), 55–86 (2009)
- [29] Liberti, L., Lavor, C., Maculan, N., Mucherino, A.: Euclidean distance geometry and applications. *SIAM Review* **56**(1), 3–69 (2014)
- [30] Malliavin, T., Mucherino, A., Nilges, M.: Distance geometry in structural biology. In: A. Mucherino, C. Lavor, L. Liberti, N. Maculan (eds.) *Distance Geometry: Theory, Methods, and Applications*. Springer, New York (2013)
- [31] Maniezzo, V., Stützle, T., Voß, S. (eds.): *Hybridizing metaheuristics and mathematical programming, Annals of Information Systems*, vol. 10. Springer, New York (2009)
- [32] Markowitz, H.: Portfolio selection. *The Journal of Finance* **7**(1), 77–91 (1952)
- [33] Plotkin, S., Shmoys, D., Tardos, E.: Fast approximation algorithm for fractional packing and covering problems. *Mathematics of Operations Research* **20**, 257–301 (1995)
- [34] Saxe, J.: Embeddability of weighted graphs in  $k$ -space is strongly **NP**-hard. *Proceedings of 17th Allerton Conference in Communications, Control and Computing* pp. 480–489 (1979)
- [35] Scherer, B., Martin, D.: *Introduction to Modern Portfolio Optimization*. Springer–Verlag (2005)
- [36] Schlick, T.: *Molecular modelling and simulation: an interdisciplinary guide*. Springer, New York (2002)
- [37] Shaw, D., Liu, S., Kopman, L.: Lagrangian relaxation procedure for cardinality-constrained portfolio optimization. *Optimization Methods and Software* **23**(3), 411–420 (2008)
- [38] Singer, A.: Angular synchronization by eigenvectors and semidefinite programming. *Applied and Computational Harmonic Analysis* **30**, 20–36 (2011)
- [39] Sun, X., Zheng, X., Li, D.: Recent advances in mathematical programming with semi-continuous variables and cardinality constraint. *Journal of the Operations Research Society of China* **1**(1), 55–77 (2013)
- [40] Tahanan, M., van Ackooij, W., Frangioni, A., Lacalandra, F.: Large-scale unit commitment under uncertainty: a literature survey. *4OR* **13**, 115–171 (2015)
- [41] Xue, H.G., Xu, G.X., Feng, Z.X.: Mean-variance portfolio optimal problem under concave transaction cost. *Applied Mathematics and Computation* **174**(1), 1–12 (2006)

## A Parameter values for the HUC instances

	instance A1		instance B1		instance C1			instance A1		instance B1		instance C1	
$h$	$I_h$	$\Pi_h$	$I_h$	$\Pi_h$	$I_h$	$\Pi_h$	$h$	$I_h$	$\Pi_h$	$I_h$	$\Pi_h$	$I_h$	$\Pi_h$
1	2.66	50.17	0.53	42.37	1.53	48.06	85	2.19	57.99	0.66	84.77	3.06	71.69
2	2.66	40.17	0.53	29.75	1.53	48.06	86	2.19	61.75	0.66	92.72	3.06	57.64
3	2.66	35.17	0.53	30.09	1.53	47.56	87	2.19	61.69	0.66	99.75	3.06	55.62
4	2.66	35.17	0.53	30.12	1.53	47.00	88	2.19	56.24	0.66	103.72	3.06	55.62

$h$	instance A1		instance B1		instance C1		$h$	instance A1		instance B1		instance C1	
	$I_h$	$\Pi_h$	$I_h$	$\Pi_h$	$I_h$	$\Pi_h$		$I_h$	$\Pi_h$	$I_h$	$\Pi_h$	$I_h$	$\Pi_h$
5	2.66	35.15	0.53	30.20	1.53	47.55	89	2.19	56.24	0.66	99.75	3.06	71.62
6	2.66	40.05	0.53	30.29	1.53	47.73	90	2.19	51.70	0.66	93.68	3.06	96.89
7	2.66	57.17	0.53	52.40	1.53	54.20	91	2.19	55.63	0.66	76.65	3.06	96.92
8	2.66	75.17	0.53	60.24	1.53	75.00	92	2.19	64.64	0.66	63.71	3.06	97.00
9	2.66	90.00	0.53	93.60	1.53	105.00	93	2.19	70.00	0.66	68.75	3.06	96.97
10	2.66	147.00	0.53	140.11	1.53	110.61	94	2.19	61.00	0.66	76.75	3.06	74.92
11	2.66	146.94	0.53	148.11	1.53	110.63	95	2.34	57.12	0.67	68.62	2.84	67.67
12	2.66	139.95	0.53	141.11	1.53	100.61	96	2.48	56.02	0.67	55.67	2.62	50.62
13	2.66	95.00	0.53	82.61	1.53	78.61	97	2.63	40.50	0.68	60.61	2.39	48.62
14	2.66	90.19	0.53	77.61	1.53	77.71	98	2.78	40.50	0.68	50.61	2.17	47.63
15	2.66	95.00	0.53	90.61	1.53	94.91	99	2.93	35.30	0.69	45.61	1.95	47.54
16	2.66	95.36	0.53	103.61	1.53	95.63	100	2.93	35.31	0.69	30.61	1.95	47.00
17	2.66	90.61	0.53	107.60	1.53	188.11	101	2.93	30.25	0.69	30.61	1.95	47.00
18	2.66	75.49	0.53	97.60	1.53	199.13	102	2.93	30.26	0.69	30.61	1.95	47.43
19	2.66	60.39	0.53	73.61	1.53	199.13	103	2.93	40.25	0.69	41.61	1.95	49.62
20	2.66	80.50	0.53	62.61	1.53	110.63	104	2.93	48.25	0.69	56.34	1.95	75.82
21	2.66	95.62	0.53	61.61	1.53	79.63	105	2.93	48.25	0.69	76.61	1.95	92.85
22	2.66	65.25	0.53	75.45	1.53	61.62	106	2.93	60.56	0.69	92.50	1.95	109.51
23	2.51	60.25	0.52	61.25	1.53	58.49	107	2.93	60.71	0.69	92.41	1.95	109.02
24	2.36	55.05	0.51	59.15	1.52	52.85	108	2.93	60.86	0.69	79.41	1.95	93.93
25	2.21	50.24	0.50	41.75	1.52	48.60	109	2.93	60.50	0.69	61.49	1.95	76.00
26	2.06	40.16	0.49	30.24	1.52	48.49	110	2.93	47.45	0.69	58.17	1.95	75.92
27	1.91	35.15	0.48	30.14	1.52	48.49	111	2.93	47.40	0.69	58.00	1.95	76.45
28	1.91	35.07	0.48	30.12	1.52	48.38	112	2.93	47.45	0.69	58.75	1.95	99.92
29	1.91	35.09	0.48	29.75	1.52	48.38	113	2.93	51.45	0.69	59.75	1.95	184.03
30	1.91	40.16	0.48	29.75	1.52	48.62	114	2.93	60.40	0.69	59.75	1.95	199.14
31	1.91	57.06	0.48	52.40	1.52	54.62	115	2.93	60.40	0.69	59.75	1.95	199.14
32	1.91	65.23	0.48	60.13	1.52	75.62	116	2.93	60.45	0.69	59.75	1.95	120.44
33	1.91	100.61	0.48	93.00	1.52	105.63	117	2.93	79.71	0.69	70.11	1.95	80.47
34	1.91	147.61	0.48	132.00	1.52	110.63	118	2.93	69.89	0.69	70.11	1.95	72.43
35	1.91	147.62	0.48	140.00	1.52	105.63	119	2.63	60.45	0.71	56.75	1.92	59.64
36	1.91	130.61	0.48	133.00	1.52	100.62	120	2.34	57.45	0.73	50.00	1.88	49.70
37	1.91	95.61	0.48	81.93	1.52	75.63	121	2.05	40.25	0.75	55.75	1.85	47.70
38	1.91	80.59	0.48	77.14	1.52	75.63	122	1.76	40.11	0.77	30.48	1.82	47.55
39	1.91	95.61	0.48	90.22	1.52	95.62	123	1.47	35.22	0.80	30.50	1.79	47.00
40	1.91	95.60	0.48	110.00	1.52	95.63	124	1.47	30.25	0.80	30.49	1.79	47.00
41	1.91	90.60	0.48	110.00	1.52	185.15	125	1.47	30.25	0.80	30.50	1.79	47.00
42	1.91	85.60	0.48	96.77	1.52	199.27	126	1.47	35.29	0.80	30.60	1.79	47.58
43	1.91	70.60	0.48	72.75	1.52	199.27	127	1.47	40.25	0.80	30.50	1.79	53.69
44	1.91	70.61	0.48	61.75	1.52	110.64	128	1.47	47.45	0.80	30.50	1.79	75.25
45	1.91	95.61	0.48	74.75	1.52	75.65	129	1.47	48.26	0.80	50.61	1.79	118.00
46	1.91	80.59	0.48	60.93	1.52	75.63	130	1.47	60.08	0.80	60.60	1.79	138.10
47	1.82	60.48	0.48	60.88	1.74	70.62	131	1.47	60.62	0.80	60.50	1.79	118.50
48	1.73	55.29	0.48	58.69	1.96	53.63	132	1.47	60.78	0.80	60.50	1.79	100.29
49	1.64	55.73	0.48	50.75	2.18	48.63	133	1.47	60.61	0.80	60.49	1.79	78.25
50	1.56	44.75	0.48	30.60	2.40	48.61	134	1.47	47.87	0.80	50.61	1.79	78.07
51	1.47	39.75	0.48	30.33	2.63	48.61	135	1.47	47.45	0.80	50.40	1.79	95.50
52	1.47	34.75	0.48	30.25	2.63	48.61	136	1.47	48.05	0.80	50.23	1.79	95.63
53	1.47	39.68	0.48	37.25	2.63	48.60	137	1.47	51.62	0.80	50.20	1.79	184.94
54	1.47	50.59	0.48	40.25	2.63	47.79	138	1.47	60.08	0.80	50.20	1.79	199.13
55	1.47	52.73	0.48	52.04	2.63	59.55	139	1.47	60.40	0.80	50.25	1.79	198.50
56	1.47	62.00	0.48	61.10	2.63	75.68	140	1.47	60.66	0.80	50.40	1.79	100.63
57	1.47	96.00	0.48	88.63	2.63	95.91	141	1.47	79.86	0.80	60.25	1.79	79.35
58	1.47	145.50	0.48	101.64	2.63	105.97	142	1.47	70.00	0.80	70.21	1.79	74.89
59	1.47	145.49	0.48	140.60	2.63	105.90	143	1.58	60.89	0.74	60.00	1.67	69.63
60	1.47	145.00	0.48	121.64	2.63	95.87	144	1.70	61.00	0.68	59.84	1.56	52.75
61	1.47	94.49	0.48	84.40	2.63	71.82	145	1.81	40.40	0.62	53.00	1.44	48.00
62	1.47	85.29	0.48	89.40	2.63	59.60	146	1.93	40.07	0.57	30.25	1.33	47.68
63	1.47	89.34	0.48	90.61	2.63	59.57	147	2.04	35.07	0.51	30.25	1.22	47.50
64	1.47	85.49	0.48	104.63	2.63	64.50	148	2.04	29.63	0.51	29.75	1.22	45.51
65	1.47	88.89	0.48	100.64	2.63	74.78	149	2.04	29.68	0.51	29.40	1.22	46.87
66	1.47	79.70	0.48	94.62	2.63	95.80	150	2.04	40.07	0.51	29.4	1.22	48.02
67	1.47	61.61	0.48	76.40	2.63	99.89	151	2.04	56.75	0.51	30.24	1.22	54.61
68	1.47	83.75	0.48	68.40	2.63	95.80	152	2.04	70.16	0.51	59.40	1.22	75.49
69	1.47	92.75	0.48	64.60	2.63	77.00	153	2.04	90.47	0.51	94.75	1.22	118.62
70	1.47	80.07	0.48	76.40	2.63	69.64	154	2.04	147.43	0.51	110.00	1.22	118.62
71	1.61	63.95	0.52	68.40	2.71	59.67	155	2.04	147.28	0.51	110.06	1.22	118.62
72	1.76	61.24	0.55	56.61	2.80	48.61	156	2.04	140.21	0.51	100.45	1.22	100.62
73	1.90	62.60	0.59	50.62	2.89	52.62	157	2.04	75.20	0.51	87.77	1.22	78.02
74	2.05	52.65	0.62	29.61	2.97	50.53	158	2.04	80.00	0.51	87.88	1.22	78.62
75	2.19	36.65	0.66	29.62	3.06	47.64	159	2.04	90.28	0.51	92.00	1.22	95.62

$h$	instance A1		instance B1		instance C1	
	$I_h$	$\Pi_h$	$I_h$	$\Pi_h$	$I_h$	$\Pi_h$
76	2.19	33.63	0.66	29.62	3.06	47.62
77	2.19	30.64	0.66	36.51	3.06	47.56
78	2.19	32.61	0.66	39.62	3.06	47.56
79	2.19	52.68	0.66	51.65	3.06	47.67
80	2.19	57.75	0.66	64.61	3.06	50.09
81	2.19	59.39	0.66	87.68	3.06	49.44
82	2.19	69.99	0.66	100.81	3.06	71.66
83	2.19	70.00	0.66	139.82	3.06	74.00
84	2.19	65.00	0.66	109.88	3.06	72.59

$h$	instance A1		instance B1		instance C1	
	$I_h$	$\Pi_h$	$I_h$	$\Pi_h$	$I_h$	$\Pi_h$
160	2.04	95.27	0.51	92.11	1.22	95.62
161	2.04	90.25	0.51	92.18	1.22	185.12
162	2.04	75.10	0.51	82.00	1.22	199.12
163	2.04	74.79	0.51	81.75	1.22	199.12
164	2.04	84.89	0.51	60.75	1.22	100.62
165	2.04	109.89	0.51	74.77	1.22	79.62
166	2.04	70.04	0.51	65.45	1.22	75.62
167	2.04	55.19	0.51	55.07	1.22	70.00
168	2.04	54.67	0.51	51.00	1.22	53.07

Table 8: Inflows  $I_h$  ( $\text{m}^3/\text{s}$ ) and prices  $\Pi_h$  (currency/MWh) for time periods  $h \in \{1, \dots, \bar{h} = 164\}$  for the 3 provided instances.

## B Detailed results for the MVPS problem

Tables 10-13 report the numeric results for each transaction cost function. Their columns are as follows:

- instance name;
- maximum risk level  $\sigma$ ;
- number  $n$  of assets quoted on the financial market;
- objective value for the MWU algorithm;
- CPU time (in seconds) for the MWU algorithm;
- objective value for the MWU algorithm with the local branching constraint;
- CPU time (in seconds) for the MWU algorithm with the local branching constraint;
- objective value for the MS algorithm with the local branching constraint;
- CPU time (in seconds) for the MS algorithm with the local branching constraint;
- relative objective value improvement from MS to MWU computed as

$$\Gamma = \frac{\text{val}(\text{MWU}) - \text{val}(\text{MS})}{|\text{val}(\text{MS})|}; \tag{B.1}$$

- time improvement ratio  $\Lambda$  from MS to MWU (see Eq. (4.2));
- relative objective value improvement from MS to MWU with the local branching constraint (see Eq. (B.1));
- time improvement ratio  $\Lambda$  from MS to MWU with the local branching constraint (see Eq. (4.2)).

The comparison metrics are summarized in the last three lines with the sum ( $\sum$ ), average (avg) and the standard deviation (std) across all 20 instances.

Instance	$\sigma$	n	MWU			CPU			MS			MWU vs. MS			MWU+LB vs. MS		
			objective	$\ x^*\ _0$	CPU	objective	$\ x^*\ _0$	CPU	objective	$\ x^*\ _0$	CPU	$\Gamma$	$\Delta$	$\Gamma$	$\Delta$		
port-orlib1_000	0.0047755010	31	1.0742596	292.536	314.918	1	1.0742596	31.203	1	1.0742596	0.000%	0.107	0.000%	0.099			
port-orlib1_050	0.0021522075	31	1.0742596	268.278	212.915	1	1.0742596	31.469	1	1.0742596	0.000%	0.117	0.000%	0.148			
port-orlib1_100	0.0010585969	31	1.0592082	13.587	12.915	3	1.0592082	13.180	3	1.0592082	0.000%	0.970	0.000%	1.021			
port-orlib1_150	0.0007158421	31	1.0180977	11.998	12.391	5	1.0180977	13.999	5	1.0180977	0.000%	0.980	0.000%	0.949			
port-orlib2_000	0.0028352430	85	1.8750415	16000.371	18516.395	1	1.8750415	1339.651	1	1.8750415	0.000%	0.084	0.000%	0.072			
port-orlib2_050	0.0004953237	85	1.8741070	30.941	196.865	3	1.8741070	182.681	3	1.8741070	0.000%	0.552	0.000%	0.928			
port-orlib2_100	0.0002704062	85	1.8587348	7038.599	5745.031	3	1.8587348	7183.440	3	1.8587348	0.000%	1.021	0.000%	1.250			
port-orlib2_150	0.0001663200	85	1.7828905	214.691	384.026	9	1.7828905	425.750	9	1.7828905	0.000%	1.983	0.000%	1.109			
port-orlib3_000	0.0015166351	89	2.3898574	18019.824	17862.785	1	2.3898574	1396.067	1	2.3898574	0.000%	0.077	0.000%	0.078			
port-orlib3_050	0.0005849758	89	2.3892174	2831.204	213.402	2	2.3892174	333.062	2	2.3892174	0.000%	0.118	0.000%	1.561			
port-orlib3_100	0.0003215941	89	2.2652696	743.189	4405.874	8	2.2652696	490.434	8	2.2652696	-0.098%	0.660	-0.098%	2.824			
port-orlib3_150	0.002239117	89	2.9234597	3329.717	26538.931	10	2.9234597	4777.167	10	2.9234597	-0.141%	0.973	-0.141%	0.735			
port-orlib4_000	0.0002987241	98	2.9200045	20875.539	29265.584	3	2.9200045	742.838	3	2.9200045	0.000%	0.160	0.000%	0.180			
port-orlib4_050	0.0003052450	98	2.9067690	210.979	192.266	4	2.9067690	11884.665	4	2.9067690	0.000%	3.521	0.000%	3.864			
port-orlib4_100	0.0001613979	98	2.8526192	895.863	6364.984	3	2.8526192	6461.036	3	2.8526192	-0.019%	1.698	-0.026%	1.867			
port-orlib4_150	0.0016485224	225	4.6677119	3478.872	754.278	10	4.6677119	5736.267	10	4.6677119	0.007%	7.212	0.007%	8.566			
port-orlib5_000	0.0005150277	225	4.6668554	5652.420	5288.969	2	4.6668554	10932.209	2	4.6668554	0.000%	1.934	0.000%	2.209			
port-orlib5_050	0.0003918260	225	4.6577793	3451.980	3640.422	4	4.6577793	9833.002	4	4.6577793	0.000%	2.849	0.000%	2.701			
port-orlib5_100	0.0003272876	225	4.6264568	7934.132	9564.905	9	4.6264568	10684.629	9	4.6264568	-0.078%	1.347	-0.078%	1.117			
avg			51.2448641	107592.611	105345.493	85	51.2448641	75729.075	86	51.2448641	-0.218%	25.011	-0.335%	32.362			
std			2.5622432	5379.631	5267.275	4.250	2.5622432	3786.454	4.300	2.5622432	-0.011%	1.401	-0.017%	1.618			
geo (CPU)			(1.2414709)	(7721.167)	(7478.113)	(3.210)	(1.2413996)	(4287.608)	(3.045)	(1.2413996)	(0.032%)	(1.672)	(0.040%)	(1.928)			
			1324.256		1090.151		909.83										

Table 9: Comparative results of MS and MWU for the transaction cost function (a).

Instance	$\sigma$	n	MWU+LB			CPU			MS			MWU vs. MS			MWU+LB vs. MS		
			objective	$\ x^*\ _0$	CPU	objective	$\ x^*\ _0$	CPU	objective	$\ x^*\ _0$	CPU	$\Gamma$	$\Delta$	$\Gamma$	$\Delta$		
port-orlib1_000	0.0047755010	31	0.3283893	1.502	1.295	2	0.3283893	4.896	1	0.3283893	0.000%	3.260	0.000%	3.781			
port-orlib1_050	0.0021522075	31	0.3268002	1.698	1.488	2	0.3268002	5.310	1	0.3268002	-0.017%	3.127	-0.017%	3.569			
port-orlib1_100	0.0010585969	31	0.3264242	1.807	1.977	3	0.3264242	5.493	3	0.3264242	0.000%	3.040	0.000%	2.778			
port-orlib1_150	0.0007158421	31	0.3255758	2.890	3.002	6	0.3255758	5.649	6	0.3255758	0.000%	1.955	0.000%	1.882			
port-orlib2_000	0.0028352430	85	0.5696578	7.404	7.059	1	0.5696578	43.570	1	0.5696578	0.000%	5.885	0.000%	6.172			
port-orlib2_050	0.0004953237	85	0.5677075	19.638	11.335	8	0.5677075	52.401	5	0.5677075	0.000%	2.688	0.000%	4.623			
port-orlib2_100	0.0002704062	85	0.5671808	611.424	776.712	10	0.5671808	85.439	10	0.5671808	0.000%	0.058	0.000%	0.046			
port-orlib2_150	0.0001663200	85	0.5668838	568.841	1560.205	10	0.5668838	1737.240	10	0.5668838	0.000%	3.054	0.000%	1.113			
port-orlib3_000	0.0015166351	89	0.7248280	19.936	13.463	1	0.7248280	37.717	1	0.7248280	0.000%	3.686	0.000%	2.802			
port-orlib3_050	0.0005849758	89	0.7234556	19.936	17.177	2	0.7234556	46.714	2	0.7234556	-0.005%	2.343	-0.005%	2.436			
port-orlib3_100	0.0003215941	89	0.7229391	17.926	1210.131	7	0.7229391	651.578	7	0.7229391	0.000%	36.348	0.000%	0.538			
port-orlib3_150	0.002239117	89	0.7225929	1165.350	3643.605	10	0.7225929	2864.165	10	0.7225929	0.000%	4.458	0.000%	0.756			
port-orlib4_000	0.002987241	98	0.8862800	11.850	29.022	1	0.8862800	50.988	1	0.8862800	0.000%	2.117	0.000%	1.214			
port-orlib4_050	0.0003052450	98	0.8846300	28.928	44.461	7	0.8846300	61.320	2	0.8846300	0.000%	0.678	0.000%	1.203			
port-orlib4_100	0.0001613979	98	0.8830698	1005.380	44.461	7	0.8830698	681.456	7	0.8830698	0.000%	1.126	0.000%	15.203			
port-orlib4_150	0.0016485224	225	1.4120658	3778.496	4306.084	10	1.4120658	4925.459	10	1.4120658	0.000%	1.801	0.000%	0.988			
port-orlib5_000	0.0005150277	225	1.4112907	4805.147	3055.888	5	1.4112907	581.806	5	1.4112907	0.000%	0.182	0.000%	3.283			
port-orlib5_050	0.0003918260	225	1.4109834	4291.447	2455.898	7	1.4109834	892.922	7	1.4109834	0.000%	0.181	0.000%	0.292			
port-orlib5_100	0.0003272876	225	1.4108196	755.727	2242.568	7	1.4108196	1368.145	7	1.4108196	0.000%	1.810	0.000%	0.317			
avg			15.6561636	17407.895	19946.639	102	15.6561636	13958.057	101	15.6561636	-0.024%	80.060	-0.024%	54.240			
std			0.7828082	870.395	977.352	5.100	0.7828082	697.903	5.090	0.7828082	-0.001%	4.003	-0.001%	2.712			
geo (CPU)			(0.3736084)	(1342.238)	(1406.937)	(3.243)	(0.3736081)	(1121.600)	(3.517)	(0.3736083)	(0.004%)	(7.751)	(0.004%)	3.380			
			71.630		87.557		127.168										

Table 10: Comparative results of MS and MWU for the transaction cost function (b).



Instance	$\sigma$	n	MWU			MWHLB			MS			MWU vs. MS			MWHLB vs. MS			$\Lambda$
			objective	CPU	$\ x^*\ _0$	objective	CPU	$\ x^*\ _0$	objective	CPU	$\ x^*\ _0$	$\Gamma$	$\Lambda$	$\Gamma$	$\Lambda$	$\Gamma$	$\Lambda$	
port-orlib1_000	0.0047755010	31	0.0217170	1.877	6	0.0395563	14.952	10	0.0487432	43.180	10	-124.447%	23.005	-23.225%	2.888			
port-orlib1_050	0.0021522075	31	0.0469337	3.138	10	0.0476659	56.161	10	0.0488171	37.351	10	-4.013%	11.903	-2.415%	0.665			
port-orlib1_100	0.0010585969	31	0.0432059	2.984	10	0.0413710	4.962	10	0.0498939	58.515	10	-15.479%	19.610	-20.601%	11.793			
port-orlib1_150	0.0007158421	31	0.0399172	2.588	9	0.0376060	6.172	9	0.0388779	6.757	9	2.604%	2.611	-3.382%	1.095			
port-orlib2_000	0.0028352430	85	0.0136601	10.510	4	0.0133721	8.726	3	0.0184047	375.738	10	-84.734%	35.751	-37.635%	43.060			
port-orlib2_050	0.0004953237	85	0.0383876	39.267	7	0.0387789	33.522	6	0.0186861	522.376	10	51.323%	13.303	51.814%	15.445			
port-orlib2_100	0.0002704062	85	0.0283756	4226.211	10	0.0166890	4382.785	10	0.0207083	524.517	10	27.021%	0.124	-24.083%	0.120			
port-orlib2_150	0.0001663200	85	0.0315010	3642.877	10	0.0307374	3989.962	10	0.0268662	1599.415	10	14.713%	0.439	12.594%	0.401			
port-orlib3_000	0.0015166351	89	0.0116564	9.548	2	0.0029338	22.035	10	0.0234077	240.592	10	-100.814%	25.198	-697.857%	10.919			
port-orlib3_050	0.0005849758	89	0.0347923	19.026	8	0.0342867	102.802	8	0.0239912	441.404	10	31.044%	0.796	31.204%	4.294			
port-orlib3_100	0.0003215941	89	0.0342867	903.586	10	0.0269692	2142.790	10	0.0188460	719.657	10	45.034%	0.796	52.018%	0.336			
port-orlib3_150	0.0002239117	89	0.0284570	15650.882	10	0.0392775	11231.379	10	0.0255798	5050.864	10	-78.545%	0.323	-89.058%	6.556			
port-orlib4_000	0.0029387241	98	0.0130890	4.641	2	0.0123611	66.034	10	0.0233697	432.908	10	10.111%	93.279	5.152%	0.450			
port-orlib4_050	0.0006828450	98	0.0583659	32.832	10	0.0541541	91.192	10	0.0239543	272.405	10	58.958%	8.297	55.766%	2.987			
port-orlib4_100	0.0003059553	98	0.0135301	1082.307	10	0.0078843	696.938	10	0.0248876	513.718	10	-83.943%	0.475	-215.660%	0.757			
port-orlib4_150	0.0001613979	98	0.0236738	15929.002	10	0.0260514	24604.167	10	0.0252765	5982.823	10	-6.770%	0.376	2.974%	0.243			
port-orlib5_000	0.0016485224	225	0.0326302	1448.934	6	0.0328336	788.687	6	0.0017349	5513.344	10	94.683%	3.805	94.717%	6.991			
port-orlib5_050	0.0005150277	225	0.0031309	302.041	9	0.0073133	370.425	6	0.0018351	350.660	9	41.386%	1.161	74.907%	0.947			
port-orlib5_100	0.0003918260	225	0.0078564	5516.043	7	0.0035668	4774.516	8	0.0103946	1525.448	10	-32.308%	0.277	-191.430%	0.319			
port-orlib5_150	0.0003272876	225	0.0143170	968.684	10	0.0143169	730.436	10	0.0143170	913.920	10	0.000%	0.943	0.000%	1.251			
$\sum$			0.5394835	49796.978	159	0.5283151	54118.943	178	0.4885919	25125.592	198	-104.177%	264.875	-924.200%	111.494			
avg			0.0269742	2489.849	7.950	0.0264158	2705.947	8.900	0.0244296	1256.280	9.900	-5.209%	13.244	-46.210%	5.875			
std			(0.0145719)	(4819.775)	(2.685)	(0.0155312)	(5846.227)	(1.917)	(0.0135588)	(1891.435)	(0.308)	(56.991%)	(21.729)	(172.079%)	9.900			
geo (CPU)			135.816	247.786					426.464									

Table 13: Comparative results of MS and MWU for the transaction cost function (e).