

RESEARCH

Open Access

Coupled-decompositions: exploiting primal–dual interactions in convex optimization problems

Antoni Morell*, José López Vicario and Gonzalo Seco-Granados

Abstract

Decomposition techniques implement the so-called “divide and conquer” in convex optimization problems, being primal and dual decompositions the two classical approaches. Although both solutions achieve the goal of splitting the original program into several smaller problems (called the subproblems), these techniques exhibit in general slow speed of convergence. This is a limiting factor in practice and in order to circumvent this drawback, we develop in this article the coupled-decompositions method. As a result, the number of iterations can be reduced by more than one order of magnitude. Furthermore, the new technique is self-adjustable, i.e., it does not depend on user-defined parameters, as opposite to what happens with classical strategies. Given that in signal processing applied to communications and networking we usually deal with a variety of problems that exhibit certain coupling structures, our method is useful to design decentralized as well as centralized optimization schemes with advantages over the existing techniques in the literature. In this article, we expose there different resource allocation problems where the proposed method is successfully applied.

1 Introduction

Convex optimization theory [1,2] has provided in the last decades a powerful framework to solve optimization problems in many distinct areas. Besides the numerous applications existing in the signal processing literature, it is also possible to find examples in topics such as filter design, machine learning, or finance among others. This great success has been motivated by (i) convex optimization provides relevant insights into each specific problem, thanks to a mature theoretical framework, (ii) some problems can be solved analytically or semi-analytically applying the so-called Karush–Kuhn–Tucker (KKT) optimality conditions, and (iii) efficient numerical methods, e.g., interior point methods, have been developed to solve generic convex problems in polynomial time.

In many engineering areas, optimization problems with a partially coupled structure arise. In particular, we consider programs where the objective can be expressed as a sum of functions that depend on disjoint sets of variables, which are additionally coupled by the problem constraints

(e.g., [3–5]). The optimization of such programs is the topic addressed by decomposition methods [6] and a common strategy is to split the original problem into several smaller subproblems that are somehow coordinated until they reach the optimal solution. Additionally and as a by-product, the resulting methods can deal more naturally with decentralized implementations [7,8].

However, existing decomposition methods exhibit some drawbacks in practice. Roughly speaking, the speed of convergence of the algorithms is in general slow (this can be appreciated, for instance, in the numerical examples of [6]) and furthermore, it is necessary to manually adjust the step-size used in the successive updates of the algorithms. Since there is no universal rule to do that optimally, the performance of the methods is compromised [9]. In order to overcome these drawbacks, we introduce a novel technique, the coupled-decompositions method (CDM). It can be applied to decentralized implementations and furthermore, due to its superior computational performance in terms of convergence speed, the new technique is also competitive when compared to well-established centralized methods.

*Correspondence: antoni.morell@uab.cat

Telecommunications and System Engineering Department (TES), Universitat Autònoma de Barcelona (UAB), Q Building, 08193 Cerdanyola del Valles, Spain

In the following, we synthesize the main contributions of this article: (i) development of new interactions between the primal and dual domains in convex decomposition problems, (ii) development of a new method based on these novel interactions for problems with a single coupling constraint, (iii) convergence proof of the proposed method, (iv) further analysis of the method when it is applied to a subset of the problems of interest, and (v) presentation of numerical examples that show the benefits of having an unsupervised and efficient solution (in terms of both computational cost and convergence speed).

The remainder of the article is organized as follows. Section 2 formulates the type of problems that we deal with and it also reviews the classical decomposition techniques. Section 3 describes the proposed CDM and proves its convergence to the optimal solution whereas Section 4 provides further analysis on the proposed method when the problem is particularized. Finally, Section 5 presents numerical examples of the proposed method and Section 6 concludes the article.

2 Problem formulation and existing solutions

In this section, we first define the type of problems that we deal with throughout the text. Thereafter, the existing decomposition techniques in the literature are reviewed.

2.1 Problem formulation

Let us consider the following optimization problem,

$$\begin{aligned} \min_{\{\mathbf{x}_j\}} \quad & \sum_{j=1}^J f_j(\mathbf{x}_j) \\ \text{s.t.} \quad & \mathbf{x}_j \in \mathcal{X}_j, \quad j = 1, \dots, J \\ & \sum_{j=1}^J h_j(\mathbf{x}_j) \leq C \end{aligned} \quad (1)$$

with variables $\mathbf{x}_j \in \mathbb{R}^{n_j}$. The functions $f_j, h_j : \mathbb{R}^{n_j} \rightarrow \mathbb{R}$ are assumed convex and differentiable in the sets \mathcal{X}_j , also convex and compact too. These sets are defined as $\mathcal{X}_j = \{\mathbf{x}_j \mid \mathbf{g}_j(\mathbf{x}_j) \leq \mathbf{0}\}^a$ with $\mathbf{g}_j(\mathbf{x}_j) = [g_j^1(\mathbf{x}_j), \dots, g_j^{G_j}(\mathbf{x}_j)]^T$, where the functions $g_j^k : \mathbb{R}^{n_j} \rightarrow \mathbb{R}$ are convex and differentiable. Therefore, Equation (1) defines a convex problem and if we further assume that its feasible region has non-empty relative interior, then strong duality holds.

Note that we may interpret (1) as the distribution of a quantity C of resources among J entities where the j th entity aims to set the values of the variables in \mathbf{x}_j (constrained to lie in \mathcal{X}_j) in order to minimize the global cost function $\sum_{j=1}^J f_j(\mathbf{x}_j)$ without exceeding the coupling constraint $\sum_{j=1}^J h_j(\mathbf{x}_j) \leq C$. The presented formulation applies, among others, to fair dynamic bandwidth allocation (DBA) in point-to-multipoint networks [4], to problems related with multiple-input multiple-output design [3,10,11] or problems related to OFDM system design [12,13].

The problem in (1) is suitable for a dual decomposition approach and also for a primal decomposition if it is adequately reformulated. In the next sections, those classical solutions are reviewed.

2.2 Primal decomposition

Let us consider the following modified version of (1),

$$\begin{aligned} \min_{\{\mathbf{x}_j\}, \mathbf{y}} \quad & \sum_{j=1}^J f_j(\mathbf{x}_j) \\ \text{s.t.} \quad & \mathbf{x}_j \in \mathcal{X}_j, \quad j = 1, \dots, J \\ & h_j(\mathbf{x}_j) \leq y_j, \quad j = 1, \dots, J \\ & \sum_{j=1}^J y_j \leq C \\ & \mathbf{y} \in \mathcal{Y}, \quad \mathcal{Y} = \mathcal{Y}_1 \times \dots \times \mathcal{Y}_J \end{aligned} \quad (2)$$

where we have introduced the coupling variables $\mathbf{y} = [y_1, \dots, y_J]^T$. The subsets $\mathcal{Y}_j \in \mathbb{R}$ are defined as the images of \mathcal{X}_j through the functions h_j , i.e., $h_j : \mathcal{X}_j \rightarrow \mathcal{Y}_j$. Since the functions h_j are convex over \mathcal{X}_j and so continuous, the subsets \mathcal{Y}_j are guaranteed to be compact ([14], Th. 5.2.2). Therefore, each \mathcal{Y}_j has both a minimum and a maximum.

In primal decomposition, we assume that the coupling variables are fixed to a given value $\mathbf{y} \in \mathcal{Y}$ (more details can be found in [15], Sec. 6.4.2). Then, the problem in (2) is solved as J independent problems in the variables \mathbf{x}_j . They are called the subproblems and they are expressed as

$$p_j(y_j) = \begin{cases} \min_{\mathbf{x}_j} & f_j(\mathbf{x}_j) \\ \text{s.t.} & h_j(\mathbf{x}_j) \leq y_j \\ & \mathbf{x}_j \in \mathcal{X}_j \end{cases} \quad (3)$$

Interestingly, we know from ([15], Sec. 5.4.4) that $-\lambda_j$, i.e., minus the Lagrange multiplier associated to the constraint $h_j(\mathbf{x}_j) \leq y_j$, is in fact a subgradient^b of p_j at y_j .

Having defined the primal subproblems, we can rewrite (2) as

$$\begin{aligned} \min_{\{\mathbf{y}_j\}} \quad & \sum_{j=1}^J p_j(y_j) \\ \text{s.t.} \quad & \sum_{j=1}^J y_j \leq C \\ & \mathbf{y} \in \mathcal{Y} \end{aligned} \quad (4)$$

and (4) is referred to as the primal master problem. Note that since the subgradients of the primal subproblems are obtained at no cost, we can use a projected gradient approach ([15], Sec. 2.3) to solve the problem. In other words, the following recursion (k indexes iterations)

$$\mathbf{y}^{k+1} = \left[\mathbf{y}^k - \alpha^k \mathbf{s}^k \right]^\dagger \quad (5)$$

with $\mathbf{s}^k = -[\lambda_1^*(y_1^k), \dots, \lambda_J^*(y_J^k)]^T$ and where $[\cdot]^\dagger$ is the projection onto the feasible set (i.e., $\mathbf{y} \in \{\mathbf{y} \mid \mathbf{y} \in \mathcal{Y}, \sum_{j=1}^J y_j \leq C\}$) converges to \mathbf{y}^* . The interested reader can find more details about primal decomposition in ([15], Sec. 6.4.2) and also in [6]. However, note that it is necessary to appropriately adjust the value of α^k in order to guarantee the convergence to the desired solution [6,9].

2.3 Dual decomposition

Dual decomposition is the dual-domain alternative to primal decomposition. Let us compute the partial Lagrangian of (1) by means of relaxing only the coupling constraint

$$q(\mu) = \sum_{j=1}^J \min_{\mathbf{x}_j \in \mathcal{X}_j} \{f_j(\mathbf{x}_j) + \mu h_j(\mathbf{x}_j)\} - \mu C \quad (6)$$

Clearly, the problem in (6) decouples into J independent problems, called the dual subproblems and defined as

$$q_j(\mu) = \begin{cases} \min_{\mathbf{x}_j} f_j(\mathbf{x}_j) + \mu h_j(\mathbf{x}_j) \\ \text{s.t.} \quad \mathbf{x}_j \in \mathcal{X}_j \end{cases} \quad (7)$$

Note that the dual subproblems are convex programs for $\mu \geq 0$ and that given a value of μ , the values of the variables in \mathbf{x}_j are found after solving the subproblems in (7) for $j = 1, \dots, J$, which can be computed in parallel. In particular, the optimal values of the primal variables, i.e., $\{\mathbf{x}_j\}$, are obtained from an optimal value of the dual variable, i.e., μ^* .

Using the dual subproblems, the dual master problem is written as

$$\max_{\mu} \sum_{j=1}^J q_j(\mu) - \mu C \quad (8) \\ \text{s.t.} \quad \mu \geq 0$$

and, as in primal decomposition, a projected gradient approach can be applied ([15], Sec. 6.4.1) to finally get μ^* . The recursion is

$$\mu^{k+1} = [\mu^k + \alpha^k s^k]^+ \quad (9)$$

where k indexes iterations and $[a]^+ = \max\{0, a\}$. As well as in primal decomposition, it can be shown that a subgradient of q_j at μ^k is readily found as^c $h_j(\mathbf{x}_j^*(\mu^k))$ once the dual subproblems are solved ([15], Sec. 6.1) and therefore, a subgradient of q at μ^k is given by $s^k = \sum_{j=1}^J h_j(\mathbf{x}_j^*(\mu^k)) - C$. Finally, note that a user-defined step-size is also necessary in dual decomposition and, as we discuss later, this is a serious drawback of classical decomposition methods in practice.

2.4 Primal–dual techniques

There is a huge list of methods in the literature that are termed primal–dual but, to the best of our knowledge, the essentials in our proposed CDM have not been established previously. In general, all the reviewed methods suffer from (i) slow speed of convergence to the optimal solution (this restricts the number of practical applications), (ii) no consideration for the separated nature of the problem (i.e., the techniques are not decomposition-based approaches), and/or (iii) the decentralized implementation of the methods is not taken into account. On the contrary, all these aspects are addressed in the proposed CDM.

A first group of existing primal–dual techniques focus on iteratively finding a saddle-point of the Lagrangian, which is a convex and concave function of the primal and dual variables, respectively. Although these methods were not originally conceived from a decomposition perspective, they can be applied to the problems of interest in this article (and also implemented in a decentralized manner). Among these techniques, we find the classical work of Arrow et al. [16] or the more recent Mean Value Cross (MVC) decompositions method [17,18]. However, both techniques need to fix an step-size (explicit or implicit as in the MVC decompositions method) and, as a consequence, they penalize in terms of convergence speed in practice.

In a second group of techniques we include all the possible combinations of classical primal and dual decompositions, as described in [6]. The idea in this case is to solve some parts of the problem with a primal decomposition approach while other parts are tackled by means of a dual decomposition. Therefore, these solutions do not consider full primal–dual interactions as in the proposed CDM, where each part considers both domains simultaneously. Furthermore, they also suffer from slow convergence speeds due, in part, to the manually adjusted step-sizes. However, it is important to remark that in the last decade a significant progress has been made in dual-decomposition-based solutions using smoothing [19] or path-following [20] strategies, improving the number of iterations of the classical dual decomposition by an order of magnitude. Notwithstanding, these methods tackle problems with linear constraints and are not designed under a decentralized implementation perspective.

Finally, let us mention the primal–dual interior point methods ([2], Sec. 11.7) and its variants [21,22] as the third group of primal–dual approaches. In this case, the basic idea is to iteratively solve the KKT conditions of the problem using numerical methods typically applied to the resolution of systems of nonlinear equations such as the Newton method. These techniques have received great attention during the past years due to their good performance in terms of convergence speed when used in generic convex problems. However, since they were not conceived to exploit the separability of the problem (if it exists), it is not straightforward to derive decentralized solutions from this third group of techniques (one of the goals in this article).

3 The CDM

In order to overcome the detected drawbacks, we design our CDM with the aim to (i) exploit the primal and dual domains in convex optimization problems and (ii) *simultaneously* benefit from the separability of the problem in order to derive decentralized solutions. Although there are solutions in the literature that exploit both solution

domains as discussed, the development of a fast technique satisfying (i) and (ii) is still pending. In the following, we first describe our proposed CDM and thereafter, we prove that the iterates of the method converge to the optimal solution.

3.1 Description of the method

The proposed method has four building blocks: the primal subproblems, the dual subproblems, the primal projection, and the dual projection. These blocks are connected as depicted in Figure 1 and in what follows, we describe the actions taken at each step of the method and we provide a summary of the technique in algorithmic form. Thereafter, the convergence of the successive updates of the CDM, i.e., μ^k , towards an optimal value of the dual variable, i.e., μ^* , is proved (and the same is valid for the rest of variables, primal, and dual).

3.1.1 Step 1: dual subproblems

From μ^k , the primal value y_j^k is obtained after solving the following convex optimization problem in d_j

$$d_j(\mu^k) = \begin{cases} \min_{x_j, y_j^k} f_j(x_j) + \mu^k y_j^k \\ \text{s.t. } h_j(x_j) \leq y_j^k \\ x_j \in \mathcal{X}_j \end{cases} \quad (10)$$

Note that $d_j(\mu^k)$ coincides with (7) if we substitute y_j^k by $h_j(x_j)$. Note also that λ_j^k , i.e., the dual variable associated to the constraint $h_j(x_j) \leq y_j^k$, always takes the value of μ^k . This can be checked using one of the KKT optimality conditions of the problem as follows

$$\frac{\partial L(x_j, y_j^k, \lambda_j^k, \dots)}{\partial y_j^k} = \mu^k - \lambda_j^k = 0 \quad \rightarrow \quad \lambda_j^k = \mu^k \quad (11)$$

where $L(x_j, y_j^k, \lambda_j^k, \dots)$ stands for the Lagrangian function of the problem. The interested reader can find more details on the Lagrangian function as well as on the KKT optimality conditions of convex problems in ([2], Sec. 5.1, Sec. 5.5).

3.1.2 Step 2: primal projection

In the second step of the method, the values in y_j^k from all the subproblems are grouped in $\mathbf{y}^k = [y_1^k, \dots, y_J^k]^T$ and projected to the subset $\mathcal{Y} \cap \{\mathbf{y} \mid \sum_i y_i = C\}$ if $\mu^k > 0$ and to the subset $\mathcal{Y} \cap \{\mathbf{y} \mid \sum_i y_i \leq C\}$ otherwise. Note that both projections force the values in $\hat{\mathbf{y}}^k$ to be feasible and that the choice of the projection subset depending on μ is in accordance with the complementary slackness constraint $\mu(\sum_{j=1}^J y_j - C) = 0$ of the problem.

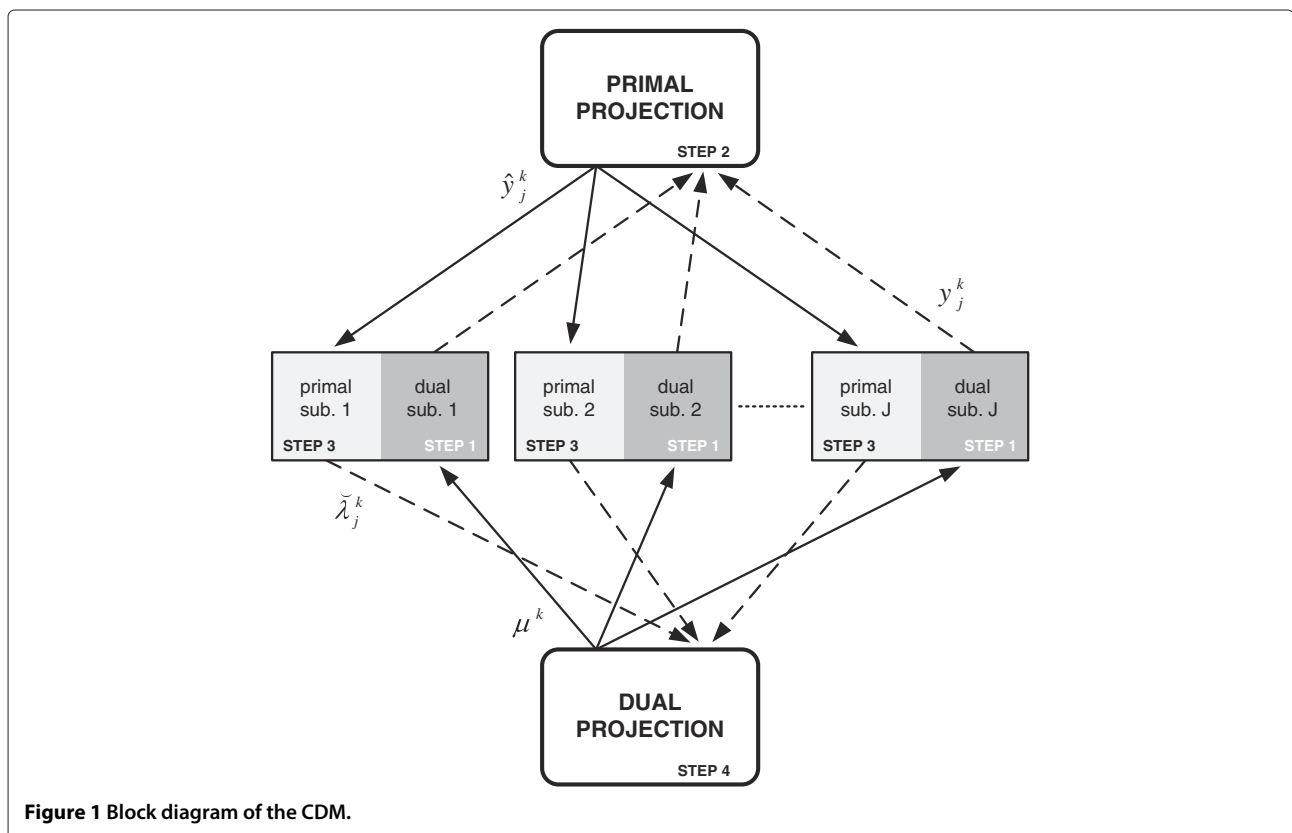


Figure 1 Block diagram of the CDM.

In the most usual case, that is, for $\mu^k > 0$, the following convex problem has to be solved

$$\begin{aligned} \min_{\hat{\mathbf{y}}^k} \quad & \|\mathbf{y}^k - \hat{\mathbf{y}}^k\|^2 \\ \text{s.t.} \quad & \sum_{j=1}^J \hat{y}_j^k = C \\ & \hat{\mathbf{y}}^k \in \mathcal{Y} \end{aligned} \quad (12)$$

which can be done semi-analytically as discussed in “Proof of Proposition 2” in Appendix.

3.1.3 Step 3: primal subproblems

The j th primal subproblem is defined as

$$p_j(\hat{y}_j^k) = \begin{cases} \min_{\mathbf{x}_j} & f_j(\mathbf{x}_j) \\ \text{s.t.} & h_j(\mathbf{x}_j) \leq \hat{y}_j^k \\ & \mathbf{x}_j \in \mathcal{X}_j \end{cases} \quad (13)$$

and it can be solved once \hat{y}_j^k is available. In this case, we are interested in the optimal value of the Lagrange multiplier associated to $h_j(\mathbf{x}_j) \leq \hat{y}_j^k$, that is, $\hat{\lambda}_j^k$. As later discussed in Section 3.4, the step 4 of the method uses only the values of $\hat{\lambda}_j^k$ that result from $\hat{y}_j^k \notin \mathbf{bd} \mathcal{Y}_j$, where $\mathbf{bd} \mathcal{A}$ stands for the boundary of the subset \mathcal{A} . The selected values are then grouped in the list $\{\check{\lambda}_j^k\}$, which is the input of the dual projection. Note that if $\sum_{j=1}^J \hat{y}_j^k = C$, then the list is guaranteed to be non-empty as shown in Proposition 3 (Section 3.4). Besides, it is important to solve the primal subproblem in (13) according to its dual version in (10). In other words, if \hat{y}_j^k is fixed in the j th primal subproblem then $\hat{\lambda}_j^k$ (not necessarily unique) is accepted as valid only if $d_j(\hat{\lambda}_j^k)$ gives $y_j^k = \hat{y}_j^k$.

3.1.4 Step 4: dual projection

If $\sum_{j=1}^J \hat{y}_j^k = C$, a new update of μ , i.e., μ^{k+1} , is obtained as the solution of the following optimization problem

$$\begin{aligned} \min_{\mu^{k+1}} \quad & \|\mu^{k+1} - \mu^k\|^2 \\ \text{s.t.} \quad & \mu^{k+1} \in \{\check{\lambda}_j^k\} \end{aligned} \quad (14)$$

In other words, μ^{k+1} takes the value in $\{\check{\lambda}_j^k\}$ that is the closest to μ^k . As discussed in Section 3.3, this is equivalent to set $\mu^{k+1} = \min\{\check{\lambda}_j^k\}$ if $\mu^k < \mu^*$ and $\mu^{k+1} = \max\{\check{\lambda}_j^k\}$ if $\mu^k > \mu^*$.

If $\sum_{j=1}^J \hat{y}_j^k < C$ then μ^{k+1} is fixed to 0, which is in accordance with the complementary slackness constraint $\mu(\sum_{j=1}^J y_j - C) = 0$.

3.2 The CDM in algorithmic form

Let us consider without loss of generality a decentralized implementation of the proposed method with a controller

and J independent participants. Each participant is able to solve the corresponding primal and dual subproblems whereas the task of the controller is to compute the primal and dual projections. Note that both operations involve simple computations as discussed in the steps 2 and 4 above. The proposed CDM is then summarized in the following algorithm,

Choose an initial value for μ^0 and repeat

1. The controller sends μ^k to the participants, which compute $d_j(\mu^k)$ in (10) and return y_j^k .
2. With $\mathbf{y}^k = [y_1^k, y_2^k, \dots, y_J^k]$, the controller computes $\hat{\mathbf{y}}^k$ using the primal projection (step 1 above) and sends \hat{y}_j^k to the participants if $\hat{y}_j^k \notin \mathbf{bd} \mathcal{Y}_j$.
3. The participants compute $p_j(\hat{y}_j^k)$ in (13) and return $\hat{\lambda}_j^k$ to the controller.
4. The controller fixes μ^{k+1} to the received value that is closer to μ^k .

Until convergence.

3.3 Resource–price interpretation

Often in convex optimization, primal variables are interpreted as resources and dual variables as prices to be paid for them. In the sequel, we revisit the proposed technique under this resource–price perspective. Initially, a global price μ^k is fixed and sent to the parts. Given that price, the parts estimate the amount of resources they want to buy. Intuitively, there will be a deficit of resources (a total request over C) if the price is too low and an excess if it is too high. In both cases, the primal projection corrects the allocation in order to distribute all the available resources among the parts. However, there is no guarantee that the distribution follows a common market law. In order to correct the situation, the primal subproblems estimate the price to be paid for the new resource allocation and, in case the individual prices differ, the dual projection fixes a new common price μ^{k+1} in order to advance towards a consensus price μ^* .

3.4 Proof of the method

Before proving that the successive updates of the proposed method converge to the optimal solution, let us establish the relationship between primal and dual variables in the subproblems with the following proposition.

Proposition 1. *Take the j th primal subproblem p_j in (13) and the j th dual subproblem d_j in (10) of the CDM. Then, the following two statements hold: (i) $\hat{\lambda}_j^k(\hat{y}_j^k)$ is non-increasing on \hat{y}_j^k in (13) and (ii) $y_j^k(\mu^k)$ is non-increasing on μ^k in (10).*

Proof. See “Proof of Proposition 1” in Appendix. \square

Next, the goal is to verify that the primal and dual projections effectively coordinate the subproblems towards the optimal solution. Let us assume, without loss of generality, that the initial guess is $\mu^0 = 0$ so that $\mu^0 \leq \mu^*$. From that value, the CDM starts by solving the dual subproblems in (10) in order to obtain \mathbf{y}^0 . As a result, there are two possibilities, namely, (i) $\sum_j y_j^0 \leq C$ and (ii) $\sum_j y_j^0 > C$. In the first situation, μ^0 as well as \mathbf{y}^0 and the corresponding values in $\{\mathbf{x}_j\}$ are optimal. Note that the subproblems are in this case decoupled and therefore the individual optimization carried out in the dual subproblems is globally optimal, too. For the sake of brevity, we do not discuss here what are the outputs of the following steps and iterations of the method, but it can be checked that the solution remains unaltered as expected. In the second case, $\mu^0 = 0$ is clearly non-optimal and in the sequel we show how the successive updates of μ^k converge to an optimal value of the dual variable, that is, $\mu^* > 0$.

Let us revisit then a complete iteration of the method starting at the dual subproblems in (10) with $\mu^k < \mu^*$, which holds at least for $k = 0$. Since y_j^k is a non-increasing function of μ^k in the j th dual subproblem (see Proposition 1), $\mu^k < \mu^*$ and $y_j^k(\mu^*) = y_j^*$, it is true that $y_j^k \geq y_j^*$. Moreover, if we take into account that $\sum_{j=1}^J y_j^* = C$ (we are considering the case where the optimal solution is coupled), we can establish that $\sum_{j=1}^J y_j^k > C$ unless $\mathbf{y}^k = \mathbf{y}^*$.

Thereafter, it is verified in the second step of the method (primal projection) that $\hat{\mathbf{y}}^k \leq \mathbf{y}^k$ ($\hat{y}_j^k < y_j^k$ for some j) according to Proposition 2 next.

Proposition 2. *Given the optimization problem in (12) and $\mathbf{y}^k \geq \mathbf{y}^*$ ($\mathbf{y}^k \neq \mathbf{y}^*$), its optimal solution can be expressed as $\hat{\mathbf{y}}^k = \mathbf{y}^k - \mathbf{r}$ with $\mathbf{r} \geq \mathbf{0}$ ($r_j > 0$ for some j).*

Proof. See “Proof of Proposition 2” in Appendix. \square

In the third step of the method, the j th primal subproblem defined in (13) computes the individual price $\hat{\lambda}_j^k$ and the list of individual prices $\{\hat{\lambda}_j^k\}$ is constructed with the values obtained from the J independent subproblems, indexed by $j = 1, \dots, J$. Note, however, that our main interest is not in the prices $\hat{\lambda}_j^k$ but in finding a global consensus price μ^* . Fortunately, if we come back to the problem definition in (2), we notice that there is a dependence between the dual variable associated to the constraint $h_j(\mathbf{x}_j) \leq y_j$, i.e., λ_j , and the dual variable associated to the constraint $\sum_{j=1}^J y_j \leq C$, i.e., μ (in terms of the proposed algorithm, \hat{y}_j^k , $\hat{\lambda}_j^k$ and μ^k play the role of y_j , λ_j and μ , respectively). This dependence motivates in our algorithm the selection of some of the values in the list $\{\hat{\lambda}_j^k\}$. To be more specific, the value $\hat{\lambda}_j^k$ is chosen if the

corresponding primal variable \hat{y}_j^k satisfies $\hat{y}_j^k \notin \mathbf{bd} \mathcal{Y}_j$ as discussed next.

Let us first write the Lagrangian of the problem in (2), that is

$$L(\{\mathbf{x}_j\}, \mathbf{y}, \lambda, \mu, \{\xi_j\}, \{\psi_j\}) = \sum_{j=1}^J f_j(\mathbf{x}_j) + \sum_{j=1}^J \xi_j^T \mathbf{g}_j(\mathbf{x}_j) + \sum_{j=1}^J \psi_j^T \mathbf{q}_j(y_j) + \sum_{j=1}^J \lambda_j (h_j(\mathbf{x}_j) - y_j) + \mu (\sum_{j=1}^J y_j - C) \quad (15)$$

where the set of convex functions $\mathbf{q}_j(y_j)$ with associated Lagrange multipliers ψ_j define the subset \mathcal{Y}_j . From the Lagrangian function we derive some of the KKT optimality conditions of the convex optimization problem as far as the optimal values of the variables form a saddle-point in the function plot. In particular, let us consider the following condition

$$\frac{\partial L}{\partial y_j} = \mu - \lambda_j + \psi_j^T \frac{\partial}{\partial y_j} \mathbf{q}_j(y_j) = 0 \quad (16)$$

that reveals

$$\mu = \lambda_j - \psi_j^T \frac{\partial}{\partial y_j} \mathbf{q}_j(y_j) \quad (17)$$

This equality is not very useful in general and neither from an algorithmic point of view because the values of the multipliers in ψ_j are unknown. However, we can make use of the following complementary slackness conditions ([2], Sec. 5.5.2) of the problem, compactly written as $\psi_j \odot \mathbf{q}_j(y_j) = \mathbf{0}$,^d and observe that if $y_j \notin \mathbf{bd} \mathcal{Y}_j$ then $\mathbf{q}_j(y_j) < \mathbf{0}$ and consequently $\psi_j = \mathbf{0}$. In that case, the link between μ and λ_j is clear,

$$\mu = \lambda_j \quad \text{if } y_j \notin \mathbf{bd} \mathcal{Y}_j \quad (18)$$

Back to the algorithm, this result motivates the use of $\hat{\lambda}_j^k$ only if it is derived from $\hat{y}_j^k \notin \mathbf{bd} \mathcal{Y}_j$ and so a new list $\{\check{\lambda}_j^k\}$ that contains all these suitable dual values is constructed. Besides, it is necessary to guarantee that the new list $\{\check{\lambda}_j^k\}$ is non-empty or, equivalently, that after the dual projection at least one value in $\hat{\mathbf{y}}^k$ satisfies $\hat{y}_j^k \notin \mathbf{bd} \mathcal{Y}_j$. This is the result of Proposition 3 next.

Proposition 3. *Let $\hat{\mathbf{y}}^k = \mathbf{y}^k - \mathbf{r}$ ($\hat{\mathbf{y}}^k \neq \mathbf{y}^*$) be a primal point resulting from the primal projection of the CDM with the value of $\mathbf{r} \geq \mathbf{0}$ suitable to fulfill $\sum_{j=1}^J \hat{y}_j^k = C$, $\hat{\mathbf{y}}^k \in \mathcal{Y}$. Then, at least one value in $\{\check{y}_j^k\}$ verifies $\hat{y}_j^k \notin \mathbf{bd} \mathcal{Y}_j$ and also $\hat{y}_j^k > y_j^*$.*

Proof. See “Proof of Proposition 3” in Appendix. \square

Finally, we need to prove that the last step of the method, i.e., the dual projection in (14), is able to find an update of the global price μ from the list $\{\check{\lambda}_j^k\}$ such that $\mu^k \xrightarrow{k \rightarrow \infty} \mu^*$.

Since we have assumed that primal and dual subproblems are reciprocal in the sense that they agree on the values of the dual variables $\hat{\lambda}_j^k$ and λ_j^k when $y_j^k = \hat{y}_j^k$ (see Section 3.1, step 3), a consequence is that $\check{\lambda}_j^k(y_j^*)$ computed in (13) equals $\lambda_j^* = \mu^*$ as well as $y_j^k(\mu^*) = y_j^*$ in (10). Note that we have intentionally written $\check{\lambda}_j^k$ instead of $\hat{\lambda}_j^k$ because our focus is only on the primal subproblems with $\hat{y}_j^k \notin \mathbf{bd} \mathcal{Y}_j$, which ensures $\lambda_j = \mu$ according to (18). Additionally, the following two claims can be made: (i) all the values in $\{\check{\lambda}_j^k\}$ satisfy $\check{\lambda}_j^k \geq \mu^k$ and (ii) at least one value in the list verifies $\check{\lambda}_j^k \leq \mu^*$. The first statement uses Proposition 1 and in particular that $\hat{\lambda}_j^k$ (or $\check{\lambda}_j^k$ equivalently) is a non-increasing function of y_j^k in the primal subproblems. Recalling that $p_j(y_j^k)$ in (13) would produce λ_j^k as inner Lagrange multiplier and that $\lambda_j^k = \mu^k$ according to (11), it is true that $\check{\lambda}_j^k \geq \mu^k$ since $\hat{y}_j^k \leq y_j^k$ (as a result of the primal projection). The second statement is verified in a similar manner taking into account that at least one value in $\{\hat{y}_j^k\}$ verifies $\hat{y}_j^k \notin \mathbf{bd} \mathcal{Y}_j$ and also $\hat{y}_j^k > y_j^*$ (see Proposition 3). Since $\check{\lambda}_j^k(y_j^*) = \lambda_j^* = \mu^*$ in the j th primal subproblem, Proposition 1 establishes that $\check{\lambda}_j^k \leq \mu^*$.

Figure 2a explains the effects of the three steps of the CDM graphically from the dual domain point of view. Each bar represents an entity (J in total) and a point in that bar indicates the value of the dual variable λ_j^k or $\hat{\lambda}_j^k$. The highest the point the highest the value. At the beginning of the k th iteration, the dual subproblems enforce $\lambda_j^k = \mu^k \forall j$ and translate these dual values to the primal variables in \mathcal{Y}^k . Immediately after the primal projection, the corrected values in $\hat{\mathcal{Y}}^k$ are converted again to dual variables, i.e., $\{\hat{\lambda}_j^k\}$. In the figure, we appreciate the effect of the primal projection on the Lagrange multipliers of interest. In short, we notice that (i) all values increase and (ii) there is at least one value below μ^* .

The role of the dual projection in (14) is then to update to μ^{k+1} by selecting the closest value to μ^k from the list

$\{\check{\lambda}_j^k\}$, that is, $\mu^{k+1} = \min\{\check{\lambda}_j^k\}$ if $\mu^k < \mu^*$, as depicted in Figure 2b. Together with the previous results, i.e., $\check{\lambda}_j^k \geq \mu^k$ and $\check{\lambda}_j^k \leq \mu^*$, the new update verifies $\mu^{k+1} \in [\mu^k, \mu^*]$ and thus our initial hypothesis ($\mu^k < \mu^*$) is also satisfied for the next iteration unless $\mu^{k+1} = \mu^*$. Therefore, successive iterations confirm $\mu^k \xrightarrow{k \rightarrow \infty} \mu^*$ and, accordingly, $\hat{y}_j^k \xrightarrow{k \rightarrow \infty} y_j^* \forall j$. This concludes the proof of the proposed method.

4 Convergence rate analysis and stopping criterion

This section provides additional insights into the proposed CDM by means of the following particularization of (1),

$$\begin{aligned} \min_{\{x_j\}, y} \quad & \sum_{j=1}^J f_j(x_j) \\ \text{s.t.} \quad & x_j \in \mathcal{X}_j, \quad j = 1, \dots, J \\ & h_j(x_j) \leq y_j, \quad j = 1, \dots, J \\ & \sum_{j=1}^J y_j \leq C \\ & y \in \mathcal{Y}, \quad \mathcal{Y} = \mathcal{Y}_1 \times \dots \times \mathcal{Y}_J \end{aligned} \quad (19)$$

where the variables in $\{x_j\}$ as well as the subsets in $\{\mathcal{X}_j\}$ are uni-dimensional. To be precise, not all the problems that can be formulated as in (19) are considered in the following convergence analysis but only those with the following dependance between the primal variable y_j and the dual variable λ_j in the subproblems of the CDM, still interesting as far as usual problems in the literature exhibit that relationship (see some examples in Section 5),

$$\begin{aligned} y_j = a_j (\lambda_j)^{-\alpha} + b_j, \quad & \text{for a certain value of } \alpha > 0, \\ & a_j > 0 \text{ and } b_j \in \mathbb{R} \end{aligned} \quad (20)$$

In the general case, the relationship between y_j and λ_j can be established again, thanks to the KKT optimality conditions of the problem. Therefore, let us construct the Lagrangian of (19), that is,

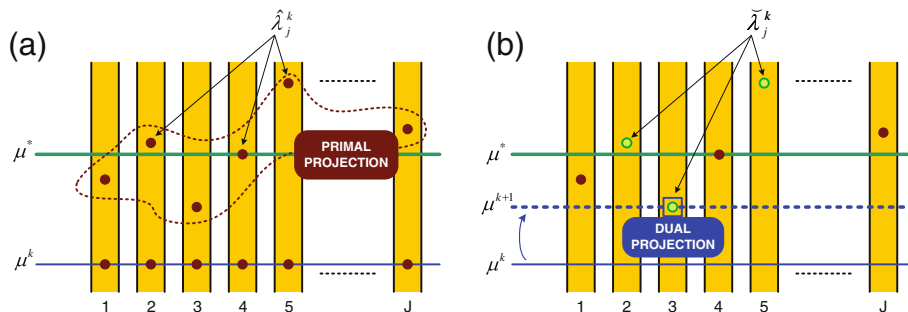


Figure 2 Primal and dual projections lead the updates of μ towards μ^* .

$$L(\{x_j\}, \mathbf{y}, \{\lambda_j\}, \mu, \{\xi_j\}) = \sum_{j=1}^J f_j(x_j) + \mu (\sum_{j=1}^J y_j - C) + \sum_{j=1}^J \lambda_j (h_j(x_j) - y_j) + \sum_{j=1}^J \xi_j^T (\mathbf{g}_j(x_j)) + \sum_{j=1}^J \psi_j^T \mathbf{q}_j(y_j) \quad (21)$$

and consider the following optimality condition,

$$\frac{\partial L}{\partial x_j} = \dot{f}_j(x_j) + \lambda_j \dot{h}_j(x_j) + \xi_j^T \frac{\partial}{\partial x_j} \mathbf{g}_j(x_j) = 0 \quad (22)$$

where \dot{f} and \dot{h} stand for the first derivatives of the functions f and h , respectively. Note that if $x_j \notin \mathbf{bd} \mathcal{X}_j$ then $\xi_j = \mathbf{0}$ due to complementary slackness and

$$\lambda_j = -\frac{\dot{f}_j(x_j)}{\dot{h}_j(x_j)} \quad (23)$$

Moreover, if the constraint $h(x_j) \leq y_j$ is satisfied with equality (the usual case as we consider coupled problems) then $x_j = h_j^{-1}(y_j)$ and the relationship between λ_j and y_j is established.

Finally, as we show in Section 5, (20) is found for common functions f and h appearing in usual problems. Furthermore, the convergence rate of the proposed method can be derived assuming (20) and a stopping criterion that enhances the performance of the CDM can be designed. These two issues are developed in the following subsections.

4.1 Convergence rate analysis

In order to find out the convergence rate of the proposed method, let us compare the value of $|(\mu^k)^{-\alpha} - (\mu^{k+1})^{-\alpha}|$ in two successive iterations, i.e., k and $k+1$. First, let us classify the optimal primal variables $\{y_j^*\}$ into three groups: \mathcal{I}^* includes the indexes j corresponding to the variables that satisfy $y_j^* = \inf \mathcal{Y}_j$, \mathcal{S}^* embraces the indexes where $y_j^* = \sup \mathcal{Y}_j$ and finally, \mathcal{A}^* contains the remaining indexes, i.e., those associated to $y_j \notin \mathbf{bd} \mathcal{Y}_j$. Using (20) and recalling the optimality condition $\lambda_j^* = \mu^*$ seen in (11), it is true that

$$y_j^* = \begin{cases} a_j (\mu^*)^{-\alpha} + b_j & j \in \mathcal{A}^* \\ m_j & j \in \mathcal{I}^* \\ d_j & j \in \mathcal{S}^* \end{cases} \quad (24)$$

where $m_j = \inf \mathcal{Y}_j$ and $d_j = \sup \mathcal{Y}_j$. Assuming that $\sum_{j=1}^J y_j^* = C$ is fulfilled, we get

$$(\mu^*)^{-\alpha} = \frac{C - \sum_{j \in \mathcal{A}^*} b_j - \sum_{j \in \mathcal{I}^*} m_j - \sum_{j \in \mathcal{S}^*} d_j}{\sum_{j \in \mathcal{A}^*} a_j} \quad (25)$$

For any other value $\mu^k \neq \mu^*$ we define

$$y_j^k = \begin{cases} a_j (\mu^k)^{-\alpha} + b_j & j \in \mathcal{A}^k \\ m_j & j \in \mathcal{I}^k \\ d_j & j \in \mathcal{S}^k \end{cases} \quad (26)$$

where the subsets \mathcal{A}^k , \mathcal{I}^k , and \mathcal{S}^k are defined likewise \mathcal{A}^* , \mathcal{I}^* , and \mathcal{S}^* but refer to the indexes of the variables in $\{y_j^k\}$.

Let us assume $\mu^k < \mu^*$ and let us obtain $\{y_j^k\}$ from (26). Clearly, since $(\mu^k)^{-\alpha} > (\mu^*)^{-\alpha}$, it holds that $y_j^k \geq y_j^* \forall j$. As a result of the primal projection in (12), now with the objective value modified by the weighting matrix $\mathbf{W} = [1/a_1, \dots, 1/a_J]^T$, i.e., $\mathbf{W}^{1/2} \|\mathbf{y}^k - \hat{\mathbf{y}}^k\|^2$, the corrected \hat{y}_j^k values can be expressed as

$$\hat{y}_j^k = \begin{cases} a_j (\mu^k)^{-\alpha} + b_j - a_j K & j \in \mathcal{A}^k \\ m_j & j \in \mathcal{I}^k \\ d_j - a_j K & j \in \mathcal{S}^k \end{cases} \quad (27)$$

for the value of $K > 0$ to be determined. The proof is very similar to the case $\mathbf{W} = \mathbf{I}$ in "Proof of Proposition 1" in Appendix and the convergence of the method is not affected. We use this projection in this particularized version simply because it offers better performance and we did not use it before just because we had no means to find a better weighting matrix than the identity matrix.

At the third step of the method, i.e., the dual subproblems, the reduced list $\{\check{\lambda}_j^k\}$ is obtained from the values \hat{y}_j^k in (27) with $j \in \mathcal{A}^k \cup \mathcal{S}^k$. In other words, reversing (20) we find

$$(\check{\lambda}_j^k)^{-\alpha} = \frac{\hat{y}_j^k - b_j}{a_j} = \begin{cases} (\mu^k)^{-\alpha} - K, & \forall j \in \mathcal{A}^k \\ \frac{d_j - b_j}{a_j} - K, & \forall j \in \mathcal{S}^k \end{cases} \quad (28)$$

Finally, in the dual projection we select the minimum value in $\{\check{\lambda}_j^k\}$, which is in this case the closest to μ^k given $\mu^k < \mu^*$ because $\mu^{k+1} \in [\mu^k, \mu^*]$ (see Section 3.3),

$$\mu^{k+1} = \min \{\check{\lambda}_j^k\} \quad (29)$$

or equivalently,

$$(\mu^{k+1})^{-\alpha} = \max \{\check{\lambda}_j^k\} = (\mu^k)^{-\alpha} - K \quad (30)$$

since $\frac{d_j - b_j}{a_j}$ is always lower than $(\mu^k)^{-\alpha}$ or otherwise $\frac{d_j - b_j}{a_j} - K$ would belong to \mathcal{A}^k . Note in (27) that the definition of the subsets \mathcal{A}^k and \mathcal{S}^k implies $d_j < a_j (\mu^k)^{-\alpha} + b_j$.

Using the previous results, we can state that

$$|(\mu^{k+1})^{-\alpha} - (\mu^*)^{-\alpha}| = |(\mu^k)^{-\alpha} - (\mu^*)^{-\alpha} - K| \quad (31)$$

This can be further refined if K is developed using (27) and $\sum_{j=1}^J \hat{y}_j^k = C$,

$$\begin{aligned} K &= \frac{(\mu^k)^{-\alpha} \sum_{j \in \mathcal{A}^k} a_j + \sum_{j \in \mathcal{A}^k} b_j + \sum_{j \in \mathcal{I}^k} m_j + \sum_{j \in \mathcal{S}^k} d_j - C}{\sum_{j \in \mathcal{A}^k \cup \mathcal{S}^k} a_j} \\ &= \frac{\sum_{j \in \mathcal{A}^k} a_j}{\sum_{j \in \mathcal{A}^k \cup \mathcal{S}^k} a_j} (\mu^k)^{-\alpha} - \frac{\sum_{j \in \mathcal{A}^k} a_j}{\sum_{j \in \mathcal{A}^k \cup \mathcal{S}^k} a_j} \\ &\quad \times \left[\frac{C - \sum_{j \in \mathcal{A}^k} b_j - \sum_{j \in \mathcal{I}^k} m_j - \sum_{j \in \mathcal{S}^k} d_j}{\sum_{j \in \mathcal{A}^k} a_j} \right] \end{aligned} \quad (32)$$

Particularly, note that the expression within brackets in (32) is exactly $(\mu^*)^{-\alpha}$ when the subsets \mathcal{A}^k , \mathcal{I}^k and \mathcal{S}^k coincide with the optimal ones. We say that the algorithm is in the *optimal zone* when the sets $(\mathcal{A}^k, \mathcal{I}^k, \mathcal{S}^k)$ coincide with $(\mathcal{A}^*, \mathcal{I}^*, \mathcal{S}^*)$.

Finally, we can conclude that the speed of convergence within the optimal zone obeys the following rule, which is obtained by plugging (32) into (31),

$$\begin{aligned} |(\mu^{k+1})^{-\alpha} - (\mu^*)^{-\alpha}| &= |(\mu^k)^{-\alpha} - (\mu^*)^{-\alpha}| \\ &\quad \times \left(1 - \frac{\sum_{j \in \mathcal{A}^*} a_j}{\sum_{j \in \mathcal{A}^* \cup \mathcal{S}^*} a_j} \right) \end{aligned} \quad (33)$$

In other words, $(\mu^k)^{-\alpha}$ converges linearly to $(\mu^*)^{-\alpha}$ except when $\mathcal{S}^* = \{\emptyset\}$, showing superlinear convergence. Alternatively, if the initial hypothesis is $\mu^0 > \mu^*$, the convergence is also linear except for $\mathcal{I}^* = \{\emptyset\}$, in which case it is superlinear. Note in both cases that since (1) and (19) are assumed coupled problems, $\mathcal{A}^* \neq \{\emptyset\}$.

4.2 Stopping criterion

The previous convergence rule in (33) can be used to define a stopping criterion for the CDM. It is based on the particular evolution followed by μ^k inside the optimal zone. For that purpose, let us take three consecutive values of μ , i.e., μ^k , μ^{k+1} , and μ^{k+2} , all of them in the optimal zone. The successive application of (33) leads to

$$\begin{aligned} |(\mu^{k+l})^{-\alpha} - (\mu^*)^{-\alpha}| &= |(\mu^k)^{-\alpha} - (\mu^*)^{-\alpha}| \\ &\quad \times \left(1 - \frac{\sum_{j \in \mathcal{A}^*} a_j}{\sum_{j \in \mathcal{A}^* \cup \mathcal{S}^*} a_j} \right)^l \end{aligned} \quad (34)$$

$l = \{0, 1, 2\}$

From (34) it is verified that

$$\frac{(\mu^{k+2})^{-\alpha} - (\mu^{k+1})^{-\alpha}}{(\mu^{k+1})^{-\alpha} - (\mu^k)^{-\alpha}} = 1 - \frac{\sum_{j \in \mathcal{A}^*} a_j}{\sum_{j \in \mathcal{A}^* \cup \mathcal{S}^*} a_j} \quad (35)$$

and therefore, in the optimal zone, the left side of (35) is a constant number regardless of k . From the practical point of view and thanks to this result, we can monitor the evolution of

$$SC^k = \frac{(\mu^{k+2})^{-\alpha} - (\mu^{k+1})^{-\alpha}}{(\mu^{k+1})^{-\alpha} - (\mu^k)^{-\alpha}}, \quad \forall k \quad (36)$$

and stop the iterations when SC^k stabilizes to a constant value. Afterwards, the optimal solution is readily obtained since at that point we know which allocations saturate to either m_j or d_j and the exact value of μ^* can be computed by means of (25).

4.3 Graphical comparison among decomposition techniques

In the sequel, we include a graphical comparison among decomposition techniques and the goal is to highlight the manner in which the different methods operate in essence. We do this with the support of the following toy optimization problem

$$\begin{aligned} \min_{x_1, x_2, y_1, y_2} \quad & a_1(x_1 - c_1)^2 + a_2(x_2 - c_2)^2 \\ \text{s.t.} \quad & x_i \leq y_i, \quad i = 1, 2 \\ & y_1 + y_2 \leq C \\ & 0 \leq x_i \leq x_i^{\max}, \quad i = 1, 2 \\ & 0 \leq y_i \leq x_i^{\max}, \quad i = 1, 2 \end{aligned} \quad (37)$$

where we have included the variables y_i to match the formulation of the proposed CDM and a primal decomposition as well. In Figure 3, we compare our proposed method to the classical decomposition techniques. In all the cases, the feasibility region of the problem in terms of the variables y_1, y_2 is marked in grey. Also, the contour lines of the objective function (centered at $\mathbf{c} = [c_1, c_2]^T$) are represented in the plots (even we know that the dependence of the objective function is on x_1, x_2 instead of y_1, y_2).

As depicted in the figure, a primal decomposition approach updates \mathbf{y}^k by adding the subgradient to the point and, if the result is not feasible, a projection corrects the situation by finding the closest point in the feasible set. In the figure, note that arrows represent subgradients and dashed lines projections. In this way, the successive projections tend to the optimal solution, i.e., \mathbf{y}^* . Next, let us consider a dual decomposition approach. In order to analyze it from the perspective of the primal variables, we need to establish first the relationship between y_j and λ_j

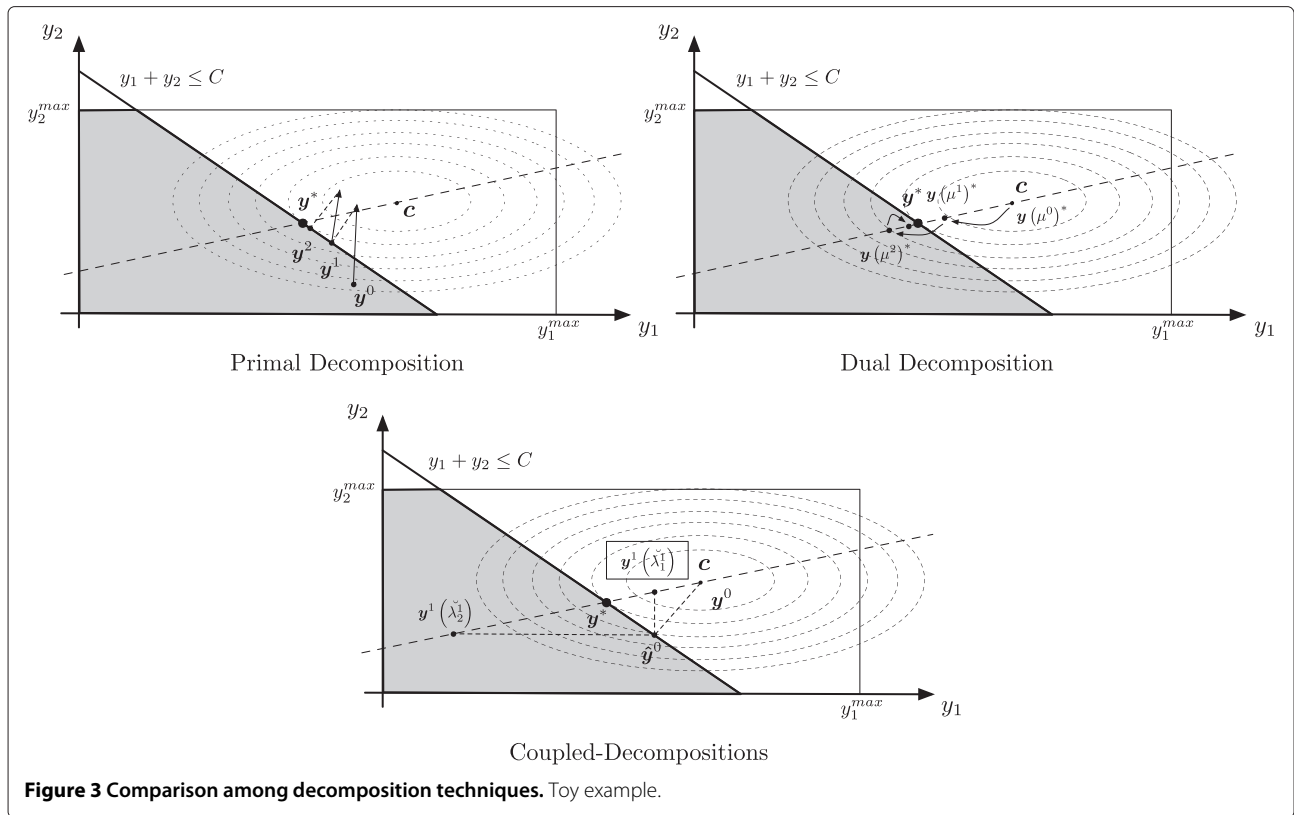


Figure 3 Comparison among decomposition techniques. Toy example.

and also between λ_j and μ . For that purpose, we consider again the Lagrangian of the problem, that is

$$\begin{aligned}
 L(x, y, \lambda, \mu, \xi_1, \xi_2, \psi_1, \psi_2) &= a_1(x_1 - c_1)^2 + a_2(x_2 - c_2)^2 + \lambda^T(x - y) \\
 &+ \mu(y_1 + y_2 - C) \\
 &- \sum_{i=1}^2 \xi_{1,i}x_i + \sum_{i=1}^2 \xi_{1,i}(x_i - x_i^{max}) \\
 &- \sum_{i=1}^2 \psi_{1,i}y_i + \sum_{i=1}^2 \psi_{1,i}(y_i - y_i^{max})
 \end{aligned} \tag{38}$$

For the case $x_i \in (0, x_i^{max}), i = 1, 2$ and $y_i = x_i$, the dual variables in ψ_i and ξ_i satisfy $\psi_i = \xi_i = \mathbf{0}, i = 1, 2$ due to slackness (note that $y_i^{max} = x_i^{max}, i = 1, 2$). In that conditions, the KKT optimality condition $\partial L / \partial x_i = 0$ forces

$$x_i = c_i - \frac{\lambda_i}{a_i} \tag{39}$$

Furthermore, if we take into account that $y_i = x_i$ in the case of interest and $\lambda_i = \mu$ due to the KKT optimality condition $\partial L / \partial y_i = 0$, then we verify that

$$y_i(\mu) = c_i - \frac{\mu}{a_i} \tag{40}$$

The dashed line in Figure 3 shows all the points that can be obtained by changing the value of μ in (40). Note in

particular that for $\mu = 0$ the optimum of the unconstrained version of (38) is achieved. Using the dual decomposition technique (in the figure we start with $\mu^0 = 0$), the successive updates move along this line (the orientation and module defined by the subgradient) until the optimal solution is achieved.

In the two classical decomposition approaches, primal and dual decomposition, a good election of the step-size that modifies the length of the subgradient plays a central role and it is recommended to choose a value that diminishes with the iteration number in order to prevent the successive updates from moving indefinitely around the optimal solution without reaching it. This issue is in fact an important practical impairment of both solutions. Notwithstanding, the method we propose in this article avoids the usage of a user-defined step-size. As the reader can appreciate in the figure, once the initial guess y^0 derived from $\mu^0 = 0$ is projected to the feasible subset, the proposed method finds out several candidates to update μ^0 to μ^1 (two in this case, i.e., $\check{\lambda}_1^1$ and $\check{\lambda}_2^1$). These two candidates provide the possible updates $y^1(\check{\lambda}_1^1)$ and $y^1(\check{\lambda}_2^1)$ and the method always chooses the dual candidate that provides the smallest possible update. This operation guarantees that the primal update, i.e., y^1 in this case, remains in the same half-space (with respect to the frontier $y_1 + y_2 = C$). Note with this simple example the

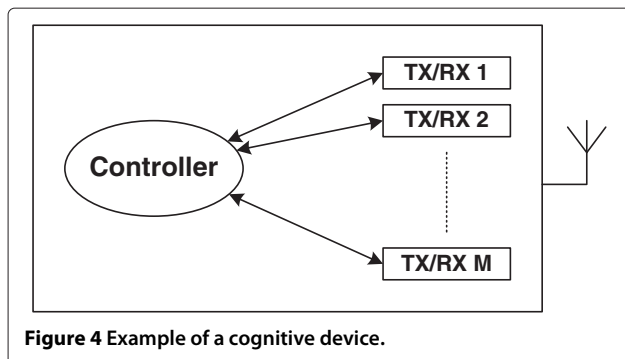
interesting feature of the proposed method in comparison with the other techniques, that is, the step is automatically controlled.

5 Applications and numerical results

In the sequel, we present three different applications of the proposed method, the first two are related to power allocation problems and the third one deals with DBA in satellite networks. In the first problem, a decentralized solution is required to reduce the amount of signaling information. In the second one, a centralized implementation of the CDM is used to solve a time-varying water-filling problem and the aim is to show the benefits of having an unsupervised method in that changing conditions. Finally, the third example shows the advantages of the proposed technique when a small allocation time is required in order to accommodate a large number of users.

5.1 Decentralized power allocation for cognitive radios

Let us consider a communication device that is able to establish simultaneous communication links by joining several networks or using multiple channels within the same system (e.g., this is possible in IEEE 802.11n). To do so, the device integrates multiple radio transceivers [23,24] which, at their turn, operate over multiple subchannels or subcarriers in order to combat the multipath fading (see Figure 4). We assume that the device can sense the wireless channel and determine the non-used subcarriers in each subsystem, as it is usual in cognitive scenarios. Furthermore, each transceiver is able to optimally allocate the available power among its subcarriers using the water-filling solution. This is advantageous from the system design point of view because we can employ off-the-shelf radio transceivers and simply balance the device power among them. Finally, there is a central controller that performs the distribution task, being the global objective to maximize the total sum rate capacity. Note that, depending on the signal strength and capacity in each subsystem, some of the transceivers may remain temporarily idle.



The problem can be formulated for M radio transceivers as

$$\begin{aligned} \max_{\{P_i\}} \quad & \sum_{i=1}^M r_i(P_i) \\ \text{s.t.} \quad & \sum_{i=1}^M P_i \leq P_T \\ & P_i \geq 0, \quad i = 1, \dots, M \end{aligned} \quad (41)$$

where $r_i(P_i)$ is the transmission rate of the i th transceiver when power P_i is allocated to it and P_T is the total available power. Each of the transmission rates is actually the result of another optimization problem, that is,

$$r_i(P_i) = \begin{cases} \max_{\{p_j^i\}} \sum_{j=1}^{N_s^i} BW_i \log \left(1 + \frac{p_j^i H_j^i}{N_0 BW_i} \right) \\ \text{s.t.} \quad p_j^i \geq 0 \\ \sum_{j=1}^{N_s^i} p_j^i \leq P_i \end{cases} \quad (42)$$

where N_s^i is the number of subcarriers of the i th radio, BW_i stands for subcarrier bandwidth, N_0 is the noise power spectral density, and H_j^i is the channel gain at the j th subcarrier of the i th transceiver.

Note that given the separability of the problem, i.e., there are many independent transceivers coupled by a total power constraint, a decentralized optimization method is adequate both from a mathematical and a practical point of view. In this approach, the controller decides the total power per transceiver and each subsystem computes its own optimal allocation. The application of the CDM to solve (41) is briefly detailed next.

Given $\mu^k < \mu^*$, each transceiver computes the following dual subproblem,

$$\begin{aligned} \max_{\{p_{i,j}\}, P_i^k} \quad & \sum_{j=1}^{N_s^i} BW_i \log \left(1 + \frac{p_{i,j} H_{i,j}}{N_0 BW_i} \right) - \mu^k P_i^k \\ \text{s.t.} \quad & \sum_{j=1}^{N_s^i} p_{i,j} \leq P_i^k \\ & p_{i,j} \geq 0 \end{aligned} \quad (43)$$

and the application of the KKT conditions gives the solution

$$\begin{aligned} p_{i,j} &= BW_i \left(\frac{1}{\mu^k} - \frac{N_0}{H_{i,j}} \right)^+ \\ P_i^k &= \sum_{j=1}^{N_s^i} p_{i,j} \end{aligned} \quad (44)$$

As a result of the dual subproblems we obtain $\mathbf{P}^k = [P_1^k, \dots, P_M^k]^T$ and we use it as the input for the primal projection, that is,

$$\begin{aligned} \min_{\{\hat{P}_i^k\}} \quad & \|\hat{\mathbf{P}}^k - \mathbf{P}^k\|^2 \\ \text{s.t.} \quad & \sum_{i=1}^M \hat{P}_i^k = P_T \\ & \hat{P}_i^k \geq 0 \end{aligned} \quad (45)$$

where $\hat{\mathbf{P}}^k = [\hat{P}_1^k, \dots, \hat{P}_M^k]^T$.

The corrected values \hat{P}_i^k are then used in the primal subproblems defined in (42) and each transceiver computes

the optimal allocation by its own. As a result of the primal subproblems, the Lagrange multipliers associated to the constraints $\sum_{j=1}^{N_s^i} p_{i,j} \leq \hat{P}_i^k$ at the k th iteration, i.e., $\hat{\lambda}_i^k$, are obtained. Discarding the values that result from $\hat{P}_i^k = 0$, we obtain the reduced list $\{\check{\lambda}_i^k\}$ and finally, the dual projection updates μ^k from $\{\check{\lambda}_i^k\}$ by doing

$$\mu^{k+1} = \min\{\check{\lambda}_i^k\} \quad (46)$$

A completely different approach is to gather all the information at the controller and to compute there the optimal power allocation. Afterwards, the result is sent back to the transceivers. Note that this centralized solution has an important drawback in terms of signaling because the powers in all the subcarriers and all the transceivers need to be exchanged. On the contrary, decentralized solutions benefit from transceiver-level signaling. In the numerical results below, we compare the CDM to other approaches.

5.1.1 Numerical results

We consider a device with three different OFDM transmitters. The first transmitter employs 256 subcarriers spanning a total bandwidth of 1.536 MHz (6 kHz per subchannel), the second one has 256 subcarriers as well and 3.072 MHz of bandwidth (12 kHz per subchannel) and the third one manages 128 subcarriers in 1.28 MHz of bandwidth (10 kHz per subchannel), so that a total of 640 subcarriers and 5.888 MHz have to be controlled (see Table 1). We assume frequency selective Rayleigh-fading channels in all three systems with a channel length of 20 taps and an exponential power delay profile where the delay spread is 1 ms. Mean channel gain is 0 dB in system 1, -10 dB in system 2, and -5 dB in system 3. Moreover, we assume that the noise power spectral density is flat over frequency with $N_0 = \sigma_n^2/BW_1$, being BW_1 the subcarrier bandwidth in system 1. Initially, we set up a uniform power allocation in all the methods and the total available power is always $P_T(\text{dB}) = \sigma_n^2(\text{dB}) + 10 \log_{10}(640) + 5$.

Figure 5 shows a multi-system water-filling allocation example. The plot at the top depicts one channel realization for the three systems whereas the plot at the bottom shows the optimal power allocation. As expected, most

Table 1 Description of the subsystems

Subsystem number	Number of subcarriers	Subcarrier bandwidth (BW_i)	Subsystem bandwidth
1	256	6 kHz	1.536 MHz
2	256	12 kHz	3.072 MHz
3	128	10 kHz	1.28 MHz

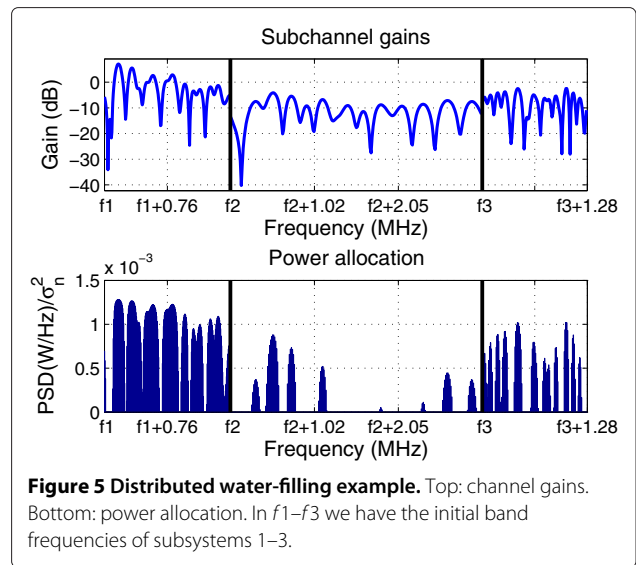


Figure 5 Distributed water-filling example. Top: channel gains. Bottom: power allocation. In f_1 – f_3 we have the initial band frequencies of subsystems 1–3.

of the power is allocated to transceivers 1 and 3, which are the ones that have the best channel condition. On the contrary, subsystem 2 only allocates power to a few subcarriers that have the highest channel gains. Notwithstanding, in absolute terms, transceiver 2 receives quite a large allocation in order to exploit the higher subcarrier bandwidth.

Figure 6 shows the evolution of the Normalized Mean Squared Error (NMSE) in the power allocation with respect to the number of messages exchanged between the transmission subsystems and the central controller. The optimal power allocation is computed using the bisection method (relative error below 10^{-5}). We compare the proposed CDM to the classical primal and dual decomposition techniques, the classical primal–dual algorithm of Arrow et al. [16] and also to a centralized approach. The classical decomposition techniques use $\alpha^k = 1/\sqrt{k}$

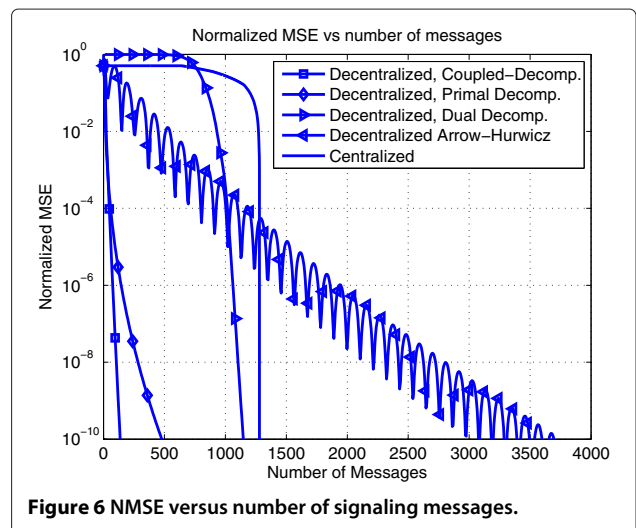


Figure 6 NMSE versus number of signaling messages.

as step-size and the Arrow–Hurwicz method initializes the value in the dual variable μ to 0, the primal variables with a uniform power allocation and the step-size is fixed to 0.1. On the one hand, results show that the proposed CDM is the best option, whereas the remaining alternatives require at least to double the amount of signaling in order to achieve the same allocation error. On the other hand, note that the classical decomposition techniques as well as the primal–dual approach are penalized in terms of convergence speed even taking into account that we have manually adjusted the step-size of each method in order to achieve the best possible result. Finally, note also that a centralized approach is not efficient at all as far as the allocation error becomes small enough only when the entire allocation has been transmitted. This requires 640 messages in our case to send the channel gains to the controller and 640 messages more to return the optimal power allocation values to the radios.

5.2 Power allocation in a conventional OFDM transmission

In the following, we apply our method to a classical water-filling problem where a decentralized solution is not necessary. In this occasion, we are interested in the adaptability of the method in time-varying scenarios.

Let us consider the well-known single-user water-filling solution over parallel Gaussian channels ([25], Sec. 10.4), which provides the optimal power allocation to the subcarriers of an OFDM-based system in order to maximize the mutual information given a total power constraint. Mathematically,

$$\begin{aligned} \max_{\{p_i\}} \quad & \sum_{i=1}^{N_s} \log \left(1 + \frac{p_i}{\sigma_{n_i}^2} \right) \\ \text{s.t.} \quad & p_i \geq 0 \\ & \sum_{i=1}^{N_s} p_i \leq P \end{aligned} \quad (47)$$

where N_s is the total number of subcarriers or parallel channels in the system, P is the total transmission power, $\sigma_{n_i}^2$ is the noise variance in the i th subcarrier and p_i stands for the allocated power. The application of the KKT optimality conditions to (47) leads to the solution

$$p_i = \left(\frac{1}{\mu} - \sigma_{n_i}^2 \right)^+ \quad (48)$$

where $(a)^+ = \max\{0, a\}$ and $\frac{1}{\mu}$ is denoted as the water-level and shall be chosen in order to satisfy the total power constraint. Typically, the bisection method is employed to find μ^* . However, note that (47) can be rewritten in the form of (2) and also (19). Therefore, we can apply the proposed CDM as well. Indeed, (48) and the relationship in (20) match if we identify p_i with y_i and μ with λ_i (remember that the required relationship applies only to $y_i \notin \text{bd } \mathcal{Y}_i$, that is, $y_i = p_i > 0$).

5.2.1 Numerical results

Let us assume $N_s = 512$ subcarriers. The channel is time-varying and frequency selective; it has 20 taps. The power delay profile is assumed exponential with a delay spread of 1 ms and the baseband sampling time is $1 \mu\text{s}$. We compare now the proposed CDM to the bisection method and also to the classical primal–dual algorithm in [16]. It is remarkable that the CDM requires no modification at all (it is completely unsupervised) and the same holds for the primal–dual algorithm. On the contrary, the bisection method requires a slight modification to be able to track the time-varying scenario. For that purpose, we introduce the updating factor α_u . Initially, the method is applied as usual, that is, having the initial hypothesis on μ_l^0 and μ_u^0 (two values that are below and above μ^* , respectively), we compute $\mu^1 = 1/2(\mu_l^0 + \mu_u^0)$ and we update μ_l^1 to μ^1 if $\sum_{i=1}^{N_s} p_i(\mu^1) > P$ or μ_h^1 to μ^1 otherwise. In the subsequent iterations, given that the channel is time-varying, we need to check first if μ_l^k and μ_h^k are still valid. If $\sum_{i=1}^{N_s} p_i(\mu_l^k) > P$ is not accomplished, we update μ_l^k to $\frac{\mu_l^k}{\alpha_u}$ and we repeat this while $\sum_{i=1}^{N_s} p_i(\mu_l^k) > P$. Similarly, if $\sum_{i=1}^{N_s} p_i(\mu_u^k) < P$ is not attained, we modify μ_u^k to $\alpha_u \cdot \mu_u^k$ and we repeat this while $\sum_{i=1}^{N_s} p_i(\mu_u^k) < P$. Then, we compute $\mu^{k+1} = 1/2(\mu_l^k + \mu_u^k)$ and we update the hypothesis accordingly, as in the normal version of the technique.

Figure 7 plots the NMSE of the power allocation for both methods as a function of the mean SNR. As in the previous application example, we compute the optimal power allocation using the bisection method (relative error below 10^{-5}). Moreover, all the algorithms are initialized to the optimal solution for the current channel condition, $\alpha_u = 1.05$ in the bisection technique, the step-size is fixed to 0.001 in the Arrow–Hurwicz method and we have considered three different channel velocities,

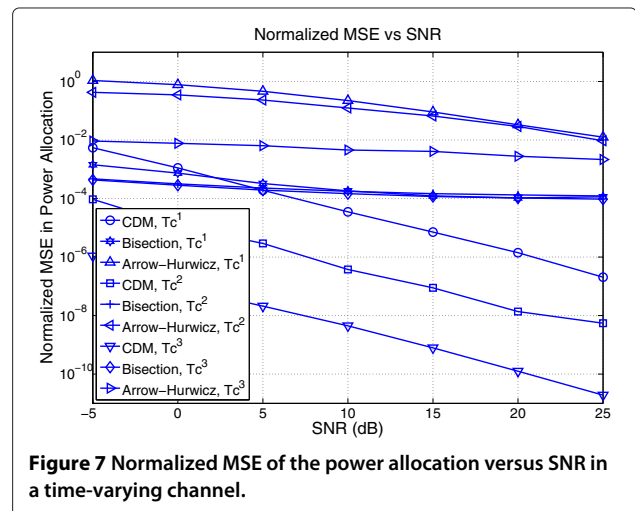


Figure 7 Normalized MSE of the power allocation versus SNR in a time-varying channel.

namely, (i) $T_c^1 = 10 \cdot T_{\text{CDM}}$, (ii) $T_c^2 = 100 \cdot T_{\text{CDM}}$, and (iii) $T_c^3 = 1000 \cdot T_{\text{CDM}}$, where T_{CDM} is the time taken by one complete iteration of the CDM and T_c^i is the coherence time of the channel at the i th scenario. Note that we have manually adjusted α_u in the bisection method and the step-size in the primal-dual algorithm in order to achieve the best possible performance at the worst channel condition, that is, when the channel coherence time is the smallest one, i.e., T_c^1 .

Results show that the CDM usually outperforms the bisection method and it is far better than the primal-dual algorithm. Indeed, it performs worse than the bisection only for T_c^1 and at low SNR. Note that since the CDM has no user-defined parameter, it automatically adapts to the different channel velocities. On the contrary, this adaptation does not occur in the other two methods. This is reflected in Figure 7, where, for example, the bisection method saturates to an NMSE around 10^{-4} for T_c^1 , T_c^2 , and T_c^3 as the SNR grows.

5.3 Fair DBA

The fair DBA problem arises in many-to-one communication systems [26,27] and the goal is to fairly distribute the available bandwidth. In many cases and specially in systems with a huge number of users [28], the computational cost of the techniques plays an important role. Additionally, let us remark that recent works on the topic aim at providing mechanisms for QoS differentiation [4,29] to modify a plain fair allocation. Therefore, we consider the following network utility maximization (NUM) formulation to solve a fair DBA problem,

$$\begin{aligned} \max_{\{r_j\}} \quad & \sum_{j=1}^N U_j(r_j; p_j) \\ \text{s.t.} \quad & m_j \leq r_j \leq d_j, \quad j = 1, \dots, N, \\ & \sum_{j=1}^N r_j \leq B \end{aligned} \quad (49)$$

where B is the available bandwidth, r_j is the rate allocated to the j th flow, and U_j is the j th utility function (the terms bandwidth and rate are used interchangeably). The parameters m_j , d_j (with $0 \leq m_j < d_j$), and $p_j > 0$ are used to define the QoS requirements for each ongoing connection and they represent the minimum necessary rate, the required (maximum) bandwidth and the priority of the j th flow, respectively. Furthermore, we assume that $\sum_j m_j < B < \sum_j d_j$, i.e., the problem is coupled. As argued before, the utility functions can adequately be chosen in order to achieve a fair distribution of resources in different degrees. The following family of functions parameterized by γ

$$U_j(r_j; p_j, \gamma) = \begin{cases} p_j \log(r_j), & \gamma = 1 \\ p_j \frac{r_j^{1-\gamma}}{1-\gamma}, & \gamma \neq 1 \end{cases} \quad (50)$$

define different types of fairness, being $\gamma \rightarrow \infty$ (max-min fairness) and $\gamma = 1$ (proportional fairness) the most relevant ones [29].

Note that (49) can be rewritten in the form of (19) and in particular, the problem is strictly convex and we assume that strong duality holds, i.e., there is at least one strictly feasible point. Therefore, we can apply the KKT optimality conditions to solve (49) semi-analytically. In this case, the optimal rates must verify^e

$$r_j^*(\mu) = \left[\left(\frac{p_j}{\mu} \right)^{\frac{1}{\gamma}} \right]_{m_j}^{d_j} \quad (51)$$

and the optimal value of μ is such that $\sum_{j=1}^N r_j^*(\mu^*) = B$. The bisection method is a classical technique widely used in the literature in order to approximate μ^* but, alternatively, we can also apply the enhanced version of the CDM. Specifically, by adding the new variables $\{y_j\}$ and identifying $f_j(r_j)$ with $-U_j(r_j)$ and $h_j(r_j) = r_j$, (23) together with $r_j = h_j^{-1}(y_j)$, (51) turns into

$$y_j = \left(\frac{p_j}{\lambda_j} \right)^{\frac{1}{\gamma}} \quad (52)$$

when $m_j < y_j < d_j$ and has the required form in (20). Therefore, once the subsets \mathcal{S}^* , \mathcal{T}^* , and \mathcal{A}^* are known, the optimal value of μ is readily found according to (25) as

$$\mu^* = \left[\frac{\sum_{i \in \mathcal{A}^*} \lambda_i \sqrt[p_i]{p_i}}{B - \sum_{i \in \mathcal{T}^*} m_i - \sum_{i \in \mathcal{I}^*} d_i} \right]^\gamma. \quad (53)$$

5.3.1 Numerical results

Let us draw the values of m_j from an integer uniform distribution between 0 and 10. Each request d_j is obtained summing m_j and an integer random number between 0 and 100. The priority values p_j are drawn from a uniform distribution that takes values between 0.25 and 5 in steps of 0.25 and $\gamma = 1$. Figure 8 plots the mean allocation time, i.e., execution time, of the CDM when centrally computed in combination with the stopping criterion defined in Section 4.2. The algorithm has been executed in a Intel[®] Core 2 Duo CPU running at 2.2 GHz and programmed in Matlab[®]. We have considered three different values for the total available bandwidth, namely $B1 = \sum_j m_j + 0.25 \sum_j d_j$, $B2 = \sum_j m_j + 0.5 \sum_j d_j$, and $B3 = \sum_j m_j + 0.75 \sum_j d_j$. The results of the CDM have been compared to the classical bisection method and to the hypothesis testing method [30]. Since the allocation time is not sensitive to the available capacity for the latter methods, in Figure 8 we distinguish among $B1$, $B2$, and $B3$ only for the CDM.

In order to provide a fair comparison among the methods, it is necessary to take into account the accuracy

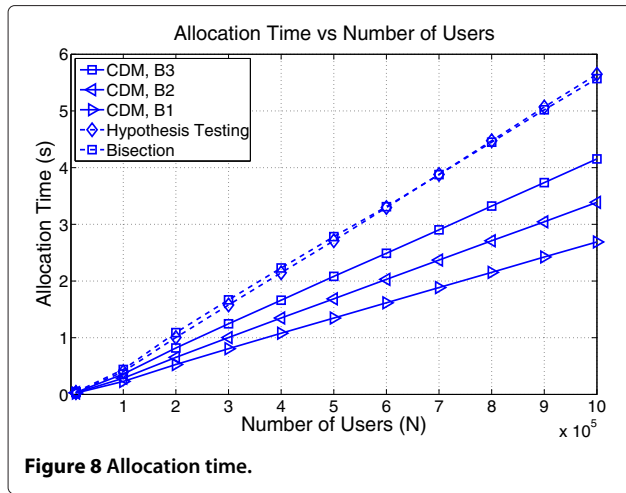


Figure 8 Allocation time.

with respect to the optimal solution. The hypothesis testing strategy always achieves the exact optimal solution (see the details in [30]). The bisection method has been adjusted to achieve a relative error in the allocation lower than 10^{-6} , or in other words,

$$\frac{\|r^{BI} - r^*\|}{\|r^*\|} \leq 10^{-6}, \quad (54)$$

where r^* is the optimal allocation (which can be obtained with the hypothesis testing method) and r^{BI} is the allocation achieved by the bisection method. Initially, the two hypothesis for the values of μ are 0 and 10. In the CDM we stop the iterations when

$$\frac{|SC^{k+1} - SC^k|}{|SC^{k+1}|} \leq 10^{-2}. \quad (55)$$

Note that as the number of users grows, the difference in time between the proposed method and the others also grows, specially when the system is more restricted in terms of capacity, i.e., for B1. In this case, the CDM is able to compute the allocation in half the time required by classical methods. In terms of accuracy in the solution, (55) gives the exact optimal solution for B1, B2 and a relative allocation error lower than 10^{-4} for B3. Overall, the solution is good in practice; it is optimal in capacity-constrained scenarios and nearly optimal in less critical situations. In order to illustrate the selection of the threshold for the stopping criterion, we plot in Figure 9 the evolution of the relative error and allocation time as a function of the accuracy in the stopping criterion. Note that a threshold of 10^{-2} provides a good trade-off between both performance metrics for the worst scenario, i.e., for B3. Finally, if we consider a higher available bandwidth, e.g., 99% of the whole system demand, this threshold value keeps the allocation time small as in B3 at the expenses of a higher allocation error (around 5%). However, the accuracy degradation appears only in this extreme case and it

is not critical in practice as far as all the users nearly reach their demands.

6 Conclusions and future work

This article has contributed with novel decomposition ideas that efficiently intertwine the classical primal and dual decomposition approaches in a single iteration of a new technique, called the CDM. It solves generic convex optimization problems that have one coupling constraint with the known advantages of decomposition-based approaches, that is, the implementation of decentralized solutions. However, it reduces the number of iterations by more than one order of magnitude with respect to the classical primal and dual decomposition solutions and furthermore, it is completely unsupervised, that is, there is no parameter that requires a manual adjustment. Moreover, when the problem is particularized (but still of interest), additional results regarding the convergence rate of the proposed technique are achieved and a stopping criterion that enhances the performance of the method (in terms of the number of iterations required to achieve the optimal solution) is derived.

The proposed method has been tested in three different problems, two dealing with power allocation in OFDM-based systems and a third one dealing with DBA. In the first two cases, the goal is to find the well-known water-filling solution in power. In one case, we benefit from a decentralized approach that suits the system architecture whereas in the other case, the proposed method is applied to a conventional OFDM transmission that deals with a time-varying channel. In both examples, we have compared our solution to other decomposition strategies and our approach performs significantly better than the available alternatives when a decentralized solution is required. In particular, our results show that the signaling requirements can be reduced at least by a factor 2.

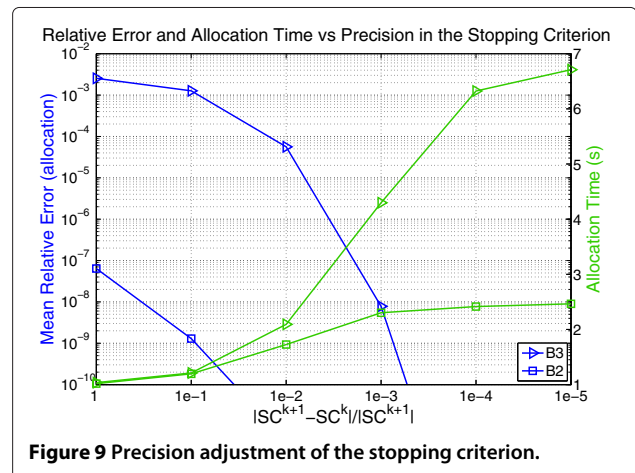


Figure 9 Precision adjustment of the stopping criterion.

Moreover, in the centralized application of the method, that is, the conventional OFDM transmission, the proposed method benefits from being unsupervised and the channel variability does not compromise the performance as classical methods do. In particular, and at the high SNR regime, the difference in the NMSE of the power allocation between the proposed method and the bisection is more than two orders of magnitude in all the explored scenarios.

Finally, when applied to a NUM problem and thanks to the enhanced version of the method, the proposed technique reduces the computational time by a factor 2 with respect to well-established techniques such as the bisection method. This reduction is very important in systems having a large number of users (as it happens in satellite communication networks), where the bandwidth allocation has to be computed in a short-time interval.

Appendix

Proof of Proposition 1

For the sake of simplicity, in what follows we obviate the iteration index k as well as the modifier $(\hat{\cdot})$ in \hat{y}_j^k , $\hat{\lambda}_j^k$, and μ^k . Let us consider first the dependance of λ_j on y_j in the primal subproblems. Interestingly, the function p_j in (13) has already been studied in the convex literature and it is known as the primal function ([15], Sec. 5.4.4). We recall here two main results related to the primal function: (i) p_j defines a convex function over the set \mathcal{P}_j defined as $\mathcal{P}_j = \{y_j \mid p_j(y_j) < \infty\}$ and (ii) the optimal value of the dual variable associated to the constraint $h_j(\mathbf{x}_j) \leq y_j$ and with opposite sign, say $-\lambda_j^*(y_j)$, is a subgradient of p_j at y_j . In our case, note that $\mathcal{Y}_j \subseteq \mathcal{P}_j \forall j$ and thus, these results can be applied to (13). Furthermore, since p_j is convex in the region of interest, it is guaranteed to be continuous although not necessarily differentiable. However, given that the objective function as well as all the constraints in the problem (finite number of them) are differentiable, then for each singular value in \mathcal{Y}_j we can find an open interval that includes it where p_j is differentiable (except of course at the singular point). Then, taking into account that the first derivative of a convex function is non-decreasing by definition and noting that the subgradient equals to the gradient where the function is differentiable, we obtain the desired result. In other words, $-\lambda_j^*(y_j)$ is non-decreasing in the intervals where p_j is differentiable just because the subgradient and the first derivative coincide whereas it takes a value in-between the right and left derivatives at the singular points, thus preserving the non-decreasing property. Finally, by removing the minus sign we can state that $\lambda_j^*(y_j)$ is non-increasing on y_j .

Second, we prove that $y_j(\mu)$ is non-increasing on μ in (10). Let us first rewrite (10) as

$$d_j(\mu) = \min_{y_j} \{p_j(y_j) + \mu y_j\} \quad (56)$$

Then take any two values of μ , say μ_1 and μ_2 , that accomplish: (i) $0 \leq \mu_1 < \mu_2$ and (ii) there exist two values in \mathcal{Y}_j , $y_{j,1}^*$ and $y_{j,2}^*$, such that $d_j(\mu_1) = p_j(y_{j,1}^*) + \mu_1 y_{j,1}^*$ and $d_j(\mu_2) = p_j(y_{j,2}^*) + \mu_2 y_{j,2}^*$. In other words, $y_{j,1}^*$ and $y_{j,2}^*$ are minimizers of $d_j(\mu_1)$ and $d_j(\mu_2)$, respectively. Now, since $y_{j,1}^*$ is not necessarily a minimizer of $d_j(\mu_2)$, we can establish the following inequality,

$$\begin{aligned} d_j(\mu_2) &= p_j(y_{j,2}^*) + \mu_1 y_{j,2}^* + (\mu_2 - \mu_1) y_{j,2}^* \\ &\leq p_j(y_{j,1}^*) + \mu_1 y_{j,1}^* + (\mu_2 - \mu_1) y_{j,1}^* \end{aligned} \quad (57)$$

Next, we prove by contradiction assuming $y_{j,2}^* > y_{j,1}^*$. In this case, it is true that (i) $(\mu_2 - \mu_1) y_{j,2}^* > (\mu_2 - \mu_1) y_{j,1}^*$ since $\mu_2 - \mu_1 > 0$ and (ii) $p_j(y_{j,2}^*) + \mu_1 y_{j,2}^* \geq p_j(y_{j,1}^*) + \mu_1 y_{j,1}^*$ since $y_{j,1}^*$ is a minimizer of $d_j(\mu_1)$. Finally, observations (i) and (ii) together contradict the inequality in (57) and therefore $y_{j,2}^*$ must necessarily satisfy $y_{j,2}^* \leq y_{j,1}^*$. This proves our second statement.

Proof of Proposition 2

Let us apply the KKT optimality conditions corresponding to (12). The Lagrangian is

$$\begin{aligned} L(\hat{\mathbf{y}}^k, \mu, \boldsymbol{\alpha}, \boldsymbol{\beta}) &= \sum_{j=1}^J (y_j^k - \hat{y}_j^k)^2 + \mu \left(\sum_{j=1}^J \hat{y}_j^k - C \right) \\ &\quad + \boldsymbol{\alpha}^T (\mathbf{y}_{min} - \hat{\mathbf{y}}) + \boldsymbol{\beta}^T (\hat{\mathbf{y}} - \mathbf{y}_{max}) \end{aligned} \quad (58)$$

where $\mathbf{y}_{min} = [\min \mathcal{Y}_1, \dots, \min \mathcal{Y}_J]$ and $\mathbf{y}_{max} = [\max \mathcal{Y}_1, \dots, \max \mathcal{Y}_J]$ and, if we look at the optimality condition $\partial L / \partial \hat{y}_j^k = 0$, we get

$$\hat{y}_j^k = y_j^k - \mu - \beta_j + \alpha_j, \quad j = 1, \dots, J \quad (59)$$

Therefore, assuming that the j th optimal primal value, i.e., $\hat{y}_j^{k,*}$, lies inside the interval $(\min \mathcal{Y}_j, \max \mathcal{Y}_j)$, then $\beta_j = \alpha_j = 0$ due to slackness and $\hat{y}_j^k = y_j^k - \mu$ (with $\mu \in \mathbb{R}$). If this is not the case, then either $\hat{y}_j^{k,*} = \min \mathcal{Y}_j$ or $\hat{y}_j^{k,*} = \max \mathcal{Y}_j$. In both cases, note that we can choose an adequate value of the free dual multipliers α_j or β_j , respectively, in order to satisfy (59). Finally, taking these results into account, we can conclude that the solution to the primal projection is

$$\hat{y}_j^k = \left[y_j^k - \mu \right]_{\min \mathcal{Y}_j}^{\max \mathcal{Y}_j} \quad (60)$$

where μ is adjusted to accomplish $\sum_{j=1}^J \hat{y}_j^k = C$. Since $y^k \geq y^*$ ($\sum y_j^k > C$ unless $y_j^k = y_j^* \forall j$) and $\sum_{j=1}^J y_j^* = C$, it is necessary that $\mu > 0$.

Proof of Proposition 3

From Proposition 2 it is verified that non-optimal values y that attain $\sum_{j=1}^J y_j > C$ diminish its value in the primal projection in order to achieve $\sum_{j=1}^J \hat{y}_j = C$ unless $y_j = \min \mathcal{Y}_j$, in which case $\hat{y}_j = y_j$. Let us distinguish two subsets of variables: \mathcal{I} includes the indexes of the values y_j that attain $y_j = \min \mathcal{Y}_j$ and $\bar{\mathcal{I}}$ the rest. Note that $\bar{\mathcal{I}}$ exactly contains the indexes of the $\{\lambda_j\}$ candidates used in the dual projection. Then it is true that

$$\hat{y}_j \leq y_j^*, \forall j \in \mathcal{I} \implies \sum_{j \in \mathcal{I}} \hat{y}_j \leq \sum_{j \in \mathcal{I}} y_j^*, \quad (61)$$

and therefore,

$$\sum_{j \in \bar{\mathcal{I}}} \hat{y}_j > \sum_{j \in \bar{\mathcal{I}}} y_j^* \quad (62)$$

since the equality constraints $\sum_{j=1}^J \hat{y}_j = \sum_{j=1}^J y_j^* = C$ are always fulfilled. This fact assures that there is at least one value \hat{y}_j with $j \in \bar{\mathcal{I}}$ that attains $\hat{y}_j > y_j^*$ unless all the values are optimal yet.

Endnotes

^aNotation: \leq , $<$, \geq , and $>$ stand for component-wise inequalities.

^bThe vector s is a subgradient of the function $f: \mathbb{R}^n \rightarrow \mathbb{R}$ at $x \in \mathbb{R}^n$ if $f(y) \geq f(x) + (y - x)^T s$, $\forall y \in \mathbb{R}^n$. If f is differentiable at x , the subgradient s and the gradient $\nabla f(x)$ coincide. Otherwise, there exist many subgradients.

^cThe notation x_j^* (μ^k) stands for the optimal solution of the j th subproblem given μ^k .

^dNotation: \odot stands for vector element-wise product. If $a = [a_1, a_2, \dots, a_N]^T$ and $b = [b_1, b_2, \dots, b_N]^T$, then $a \odot b = [a_1 \cdot b_1, a_2 \cdot b_2, \dots, a_N \cdot b_N]^T$.

^eNotation: $[a]_{m_j}^{d_j}$ equals a if $m_j < a < d_j$, m_j if $a \leq m_j$ and d_j if $a \geq d_j$.

Competing interests

The authors declare that they have no competing interests.

Acknowledgements

This work is supported by the Spanish Government under project TEC2011-28219 and the Catalan Government under grant 2009 SGR 298.

Received: 24 October 2012 Accepted: 23 January 2013

Published: 5 March 2013

References

1. D Bertsekas, A Nedić, A Ozdaglar, *Convex Analysis and Optimization* (Athena Scientific, Belmont, MA, USA, 2003)
2. L Boyd, S Vandenberghe, *Convex Optimization* (Cambridge University Press, Cambridge, MA, 2003)
3. D Palomar, J Cioffi, M Lagunas, Joint Tx-Rx beamforming design for multicarrier MIMO channels: a unified framework for convex optimization. *IEEE Trans. Signal Process.* **51**(9), 2381–2401 (2003)
4. H Yache, R Mazumdar, C Rosenberg, A game theoretic framework for bandwidth allocation and pricing in broadband networks. *IEEE/ACM Trans. Netw.* **8**(5), 667–678 (2000)
5. L Xiao, M Johansson, S Boyd, Simultaneous routing and resource allocation via dual decomposition. *IEEE Trans. Commun.* **52**(7), 1136–1144 (2004)
6. D Palomar, M Chiang, Alternative decompositions for distributed maximization of network utility: framework and applications. *IEEE Trans. Autom. Control.* **52**(12), 2254–2269 (2007)
7. CW Tan, D Palomar, M Chiang, in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, vol. 5. Distributed optimization of coupled systems with applications to network utility maximization, Toulouse, May 2006, pp. 981–984
8. S Liao, W Cheng, W Liu, Z Yang, Y Ding, in *IEEE International Conference on Communications (ICC)*. Distributed optimization for utility-energy tradeoff in wireless sensor networks, Glasgow, June 2007, pp. 3190–3194
9. L Xiao, S Boyd, Optimal scaling of a gradient method for distributed resource allocation. *J. Optim. Theory Appl. (JOTA)*. **129**(3), 469–488 (2006)
10. D Palomar, J Fonollosa, Practical algorithms for a family of waterfilling solutions. *IEEE Trans. Signal Process.* **53**(2), 686–695 (2005)
11. DP Palomar, M Bengtsson, B Ottersten, Minimum BER linear transceivers for MIMO channels via primal decomposition. *IEEE Trans. Signal Process.* **53**(8), 2866–2882 (2005)
12. A Scaglione, S Barbarossa, GB Giannakis, in *IEEE 52nd Vehicular Technology Conference (VTC Fall 2000)*, vol. 3. Optimal adaptive precoding for frequency-selective Nakagami-m fading channels, Boston, September 2000, pp. 1291–1295
13. AG Marqués, FF Digham, GB Giannakis, Optimizing power efficiency of OFDM using quantized channel state information. *IEEE J. Sel. Areas Commun.* **24**(8), 1581–1592 (2006)
14. A Arkhangel'skii, V Fedorchuk, in *General Topology, I, Encyclopedia of the Mathematical Sciences. The Basic Concepts and Constructions of General Topology* (Springer, New York, 1990)
15. D Bertsekas, *Nonlinear Programming* (Athena, Scientific, Belmont, MA, USA, 1999)
16. KJ Arrow, L Hurwicz, H Uzawa, in *Studies in Linear and Nonlinear Programming. Iterative Methods in Concave Programming* (Stanford University Press, Palo Alto, 1958), pp. 154–165
17. K Holmberg, K Kiwiel, Mean value cross decomposition for nonlinear convex problems. *Optim. Methods Softw.* **21**(3), 401–417 (2006)
18. K Holmberg, in *Design Models for Hierarchical Organizations: Computation, Information, and Decentralization. Primal and Dual Decomposition as Organizational Design: Price and/or Resource Directive Decomposition* (Kluwer Academic Publishers, Dordrecht, 1995), pp. 61–92
19. I Necoara, JAK Suykens, Application of a smoothing technique to decomposition in convex optimization. *IEEE Trans. Autom. Control.* **53**(11), 2674–2679 (2008)
20. QT Dinh, I Necoara, C Savorgnan, M Diehl, An inexact perturbed path-following method for lagrangian decomposition in large-scale separable convex optimization. *SIAM J. Optim.* **23**(1), 95–125 (2013)
21. P Tseng, D Bertsekas, On the convergence of the exponential multipliers method for convex programming. *Math. Program.* **60**(1-3), 1–19 (1993)
22. R Polyak, Primal–dual exterior point method for convex optimization. *Optim. Methods Softw.* **23**(1), 141–160 (2008)
23. I Akyildiz, S Mohanty, J Xie, A ubiquitous mobile communication architecture for next-generation heterogeneous wireless systems. *IEEE Commun. Mag.* **43**(6), S29–S36 (2005)
24. D Wang, K Miao, V John, S Rungta, W Chan, in *IEEE WICoM. Considering wireless mesh network with heterogeneous multiple radios*, Shanghai, September 2007, pp. 1681–1684
25. T Cover, J Thomas, *Elements of Information Theory* (Wiley, New York, 1991)
26. IEEE, Air Interface for Fixed and Mobile Broadband Wireless Access Systems; Amendment 2: Physical and Medium Access Control Layers for Combined Fixed and Mobile Operation in Licensed Band and Corrigendum 1, IEEE Standards, (2006)
27. ETSI: Digital Video Broadcasting (DVB); Interaction Channel for Satellite Distribution Systems. ETSI EN 301 790, 2005

28. G Acar, C Rosenberg, Weighted fair bandwidth-on-demand (WFBoD) for geostationary satellite networks with on-board processing. *Comput. Netw.* **39**(1), 5–20 (2002)
29. J Mo, J Walrand, Fair end-to-end window-based congestion control. *IEEE/ACM Trans. Netw.* **8**(5), 556–567 (2000)
30. G Seco-Granados, M Vazquez-Castro, A Morell, F Vieira, in *Proceedings of the IEEE Global Telecommunication Conference (GLOBECOM)*. Algorithm for fair bandwidth allocation with QoS constraints in DVB-S2/RCS, San Francisco, USA, November 2006), pp. 1–5

doi:10.1186/1687-6180-2013-41

Cite this article as: Morell *et al.*: Coupled-decompositions: exploiting primal–dual interactions in convex optimization problems. *EURASIP Journal on Advances in Signal Processing* 2013 **2013**:41.

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- ▶ Convenient online submission
- ▶ Rigorous peer review
- ▶ Immediate publication on acceptance
- ▶ Open access: articles freely available online
- ▶ High visibility within the field
- ▶ Retaining the copyright to your article

Submit your next manuscript at ▶ springeropen.com
