Convex Separable Problems with Linear and Box Constraints in Signal Processing and Communications

Antonio A. D'Amico, Luca Sanguinetti *Member, IEEE*, Daniel P. Palomar *Fellow, IEEE*.

Abstract—In this work, we focus on separable convex optimization problems with box constraints and a set of triangular linear constraints. The solution is given in closed-form as a function of some Lagrange multipliers that can be computed through an iterative procedure in a finite number of steps. Graphical interpretations are given casting valuable insights into the proposed algorithm and allowing to retain some of the intuition spelled out by the water-filling policy. It turns out that it is not only general enough to compute the solution to different instances of the problem at hand but also remarkably simple in the way it operates. We also show how some power allocation problems in signal processing and communications can be solved with the proposed algorithm.

Index Terms—Convex problems, separable functions, linear constraints, box constraints, power allocation, water-filling, cavefilling, multi-level water-filling, multi-level cave-filling.

I. INTRODUCTION

CONSIDER the following problem:

$$(\mathcal{P}): \quad \min_{\{x_n\}} \quad \sum_{n=1}^{N} f_n(x_n) \tag{1}$$

$$\text{subject to} \quad \sum_{n=1}^{j} x_n \leq \rho_j \quad j = 1, 2, \ldots, N$$

$$l_n \leq x_n \leq u_n \quad n = 1, 2, \ldots, N$$

where $\{x_n\}$ are the optimization variables, the coefficients $\{\rho_j\}$ are real-valued parameters, and the constraints $l_n \leq x_n \leq u_n$ are called variable bounds or box constraints with $-\infty \leq l_n < u_n \leq +\infty$. The functions f_n are real-valued, continuous and strictly convex in $[l_n, u_n]$, and continuously differentiable in (l_n, u_n). If f_n is not defined in l_n and/or in u_n, then it is extended by continuity as $f_n(l_n) = \lim_{x_n \to l_n^+} f_n(x_n)$ and $f_n(u_n) = \lim_{x_n \to u_n^-} f_n(x_n)$. Possible extensions of $(\mathcal{P})$ will be discussed in Section II.C.

A. A. D'Amico and L. Sanguinetti are with the University of Pisa, Department of Information Engineering, Via Caruso 56126 Pisa, Italy (e-mail: {a.damico, luca.sanguinetti}@iet.unipi.it). L. Sanguinetti is also with the Alcatel-Lucent Chair on Flexible Radio, Supélec, Gif-sur-Yvette, France (e-mail: luca.sanguinetti@supelec.fr.)

D. P. Palomar is with the Hong Kong University of Science and Technology, Department of Electronic and Computer Engineering, Clear Water Bay, Kowloon, Hong Kong (e-mail: palomar@ust.hk).

A. Motivation and contributions

Constrained optimization problems of the form (1) arise in connection with a wide range of power allocation problems in different applications and settings in signal processing and communications. For example, they arise in connection with the design of multiple-input multiple-output (MIMO) systems dealing with the minimization of the power consumption while meeting the quality-of-service (QoS) requirements over each data stream (see for example [1], [2], [3], [4], [5] for point-to-point communications and [6], [7], [8], [9], [10] for amplify-and-forward relay networks). A survey of some of these problems for point-to-point MIMO communications can be found in [11]. It also appears in the design of optimal training sequences for channel estimation in multi-hop transmissions using decode-and-forward protocols [12] and in the optimal power allocation for the maximization of the instantaneous received signal-to-noise ratio in amplify-and-forward multi-hop transmissions under short-term power constraints [13]. Other instances of (1) are shown to be the rate-constrained power minimization problem over a code division multiple-access channel with correlated noise [14] and the power allocation problem in amplify-and-forward relaying scheme for multiuser cooperative networks under frequency-selective block-fading [15]. Formulations as in (1) arise also in wireless communications with energy harvesting constraints. For example, they appear in [16] wherein the authors look for the optimal energy management scheme that maximizes the throughput in a point-to-point link with an energy harvesting transmitter operating over a fading channel. They can also be found in the design of the precoding strategy that maximizes the mutual information along independent channel accesses under non-causal knowledge of the channel state and harvested energy [17].

Clearly, the optimization problem in (1) can always be solved using standard convex solvers. Although possible, this in general does not provide any insights into its solution and does not exploit the particular structure of the problem itself. In this respect, all the aforementioned works go a step further and provide ad-hoc algorithms for specific instances of (1) in the attempt of giving some intuition on the solutions. However, this is achieved at the price of a loss of generality in the sense that most of them can only be used for the specific problem at hand. On the contrary, the main contribution of this work is to develop a general framework that allows one to compute the solution (and its structure) for any problem in the form of (1). In other words, whenever a problem can be put in the form of (1), then its solution can be efficiently obtained by particularizing the proposed algorithm to the problem at hand without the need of developing specific solutions.

B. Related work

The main related literature to this paper is represented by [18] and [19] in which the authors focus on solving problems of the form:

$$\min_{\{x_n\}} \quad \sum_{n=1}^{N} f_n(x_n) \tag{2}$$

$$\text{subject to} \quad \sum_{n=1}^{j} x_n \leq \sum_{n=1}^{j} \alpha_n \quad j = 1, 2, \ldots, N$$

$$0 \leq x_n \leq u_n \quad n = 1, 2, \ldots, N$$

with $\alpha_n \geq 0$ for any n. The above problems are known as separable convex optimization problems with linear *ascending inequality* constraints and box constraints. In particular, in [18] the authors propose a dual method to numerically evaluate the solution of the above problem in no more than $N - 1$ iterations under an ordering condition on the slopes of the functions at the origin. An alternative solution improving the worst case complexity of [18] is illustrated in [19]. Differently from [18] and [19], we consider more general problems in which the inequality constraints are *not necessarily* in ascending order since the box constraint values l_n and u_n may possibly be equal to $-\infty$ and $+\infty$, respectively. All this makes (1) more general than problems of the form given in (2). Observe, however, that if the lower bounds l_n are all finite, then problem (1) boils down to (2) (as it can be easily shown using simple mathematical arguments). Compared to [18] and [19], however, we also follow a different approach that allows us (simply exploiting the inherent structure of (1)) to focus only on functions f_n that are continuous, strictly convex and monotonically decreasing in the intervals $[l_n, u_n]$. Furthermore, differently from [18] we do not impose any constraints on the slopes of f_n.

It is also worth mentioning that at the time of submission we became aware (through a private correspondence with the authors) of [20] in which the problem originally solved in [18] has been revisited in light of the theory of polymatroids. In particular, in [20] the authors have removed some of the restrictions on functions f_n that were present in [18]. This allows them to come up with a solution similar to the one we propose in this work.

C. Organization

The remainder of the paper is structured as follows. Some preliminary results are discussed in the next section together with some possible extensions of the problem at hand. Section III provides the main result of the paper: an algorithm to evaluate the solution to $(\mathcal{P})$. Section IV presents some graphical interpretations of the way the proposed solution operates. This leads to an interesting water-filling inspired policy. Section V shows how some power allocation problems of practical interest in signal processing and communications can be solved with the proposed algorithm. Finally, some conclusions are drawn in Section VI.

II. PRELIMINARY RESULTS AND DISCUSSIONS

Some preliminary results are discussed in the sequel. In particular, we first study the feasibility (admissibility) of (1) and then we show that the optimization in (1) reduces to solve an equivalent problem in which all the functions f_n are continuous, strictly convex and monotonically decreasing in the intervals $[l_n, u_n]$. In addition, we also discuss some possible extensions of (1).

A. Feasibility

The feasibility of (1) simply amounts to verifying that for given values of $\{l_n\}$, $\{u_n\}$ and $\{\rho_n\}$, the *feasible set* (or *constraint set*) is not empty [21]. A necessary and sufficient condition for (1) to be feasible is provided in the following proposition.

Proposition 1. *The solution to* (1) *exists if and only if*

$$\sum_{n=1}^{j} l_n \leq \rho_j \quad j = 1, 2, \ldots, N. \tag{3}$$

Proof: The proof easily follows from $(\mathcal{P})$ since the point $(l_1, l_2, \ldots, l_N)$ is feasible. ∎

In all subsequent discussions, we assume that (3) is satisfied and denote $x_n^\star$, for $n = 1, 2, \ldots, N$, the solutions of (1).

B. Monotonic properties of f_n

Observe that since f_n is by definition strictly convex in $[l_n, u_n]$ and continuously differentiable in (l_n, u_n), then the three following cases may occur.

a) The function f_n is monotonically increasing in $[l_n, u_n]$ or, equivalently, $f_n'(x_n) > 0$ for any $x_n \in (l_n, u_n)$.

b) There exists a point z_n in (l_n, u_n) such that $f_n'(z_n) = 0$ with $f_n'(x_n) < 0$ and $f_n'(x_n) > 0$ for any x_n in (l_n, z_n) and (z_n, u_n), respectively.

c) The function f_n is monotonically decreasing in $[l_n, u_n]$ or, equivalently, $f_n'(x_n) < 0$ for any $x_n \in (l_n, u_n)$.

Lemma 1. *If f_n is monotonically increasing in $[l_n, u_n]$ and $l_n \neq -\infty$, then $x_n^\star$ is given by*

$$x_n^\star = l_n. \tag{4}$$

Proof: The proof is given in Appendix A. ∎

The above result can be used to find an equivalent form of (1). Denote by $\mathcal{A} \subseteq \{1, 2, \ldots, N\}$ the set of indices n in (1) for which case **a)** holds true and assume (without loss of generality) that $\mathcal{A} = \{1, 2, \ldots, |\mathcal{A}|\}$. Using the results of Lemma 1, it follows that $x_n^\star = l_n$ for any $n \in \mathcal{A}$ while the computation of the remaining variables with indices $n \notin \mathcal{A}$ requires to solve the following reduced problem:

$$\min_{\{x_n\}} \quad \sum_{n=|\mathcal{A}|+1}^{N} f_n(x_n) \tag{5}$$

$$\text{subject to} \quad \sum_{n=|\mathcal{A}|+1}^{j} x_n \leq \rho_j' \quad j = |\mathcal{A}| + 1, \ldots, N$$

$$l_n \leq x_n \leq u_n \quad n = |\mathcal{A}| + 1, \ldots, N$$

with

$$\rho'_j = \rho_j - \sum_{n=1}^{|\mathcal{A}|} l_n \qquad (6)$$

for $j = |\mathcal{A}| + 1, \ldots, N$[1]. The above optimization problem is exactly in the same form of (1) except for the fact that all its functions f_n fall into cases **b)** or **c)**. To proceed further, we make use of the following result.

Lemma 2. *If there exists a point z_n in (l_n, u_n) such that $f'_n(z_n) = 0$ with $f'_n(x_n) < 0 \; \forall x_n \in (l_n, z_n)$ and $f'_n(x_n) > 0 \; \forall x_n \in (z_n, u_n)$, then it is always*

$$l_n \leq x_n^{\star} \leq z_n. \qquad (7)$$

Proof: The proof is given in Appendix A. ∎

Using the above result, it follows that solving (5) amounts to looking for the solution of the following equivalent problem:

$$\min_{\{x_n\}} \quad \sum_{n=|\mathcal{A}|+1}^{N} f_n(x_n) \qquad (8)$$

$$\text{subject to} \quad \sum_{n=|\mathcal{A}|+1}^{j} x_n \leq \rho'_j \quad j = |\mathcal{A}| + 1, \ldots, N$$

$$l_n \leq x_n \leq u'_n \quad n = |\mathcal{A}| + 1, \ldots, N$$

where

$$u'_n = z_n \quad \text{if } n \in \mathcal{B} \qquad (9)$$
$$u'_n = u_n \quad \text{otherwise} \qquad (10)$$

with $\mathcal{B}$ being the set of indices n in (5) for which case **b)** holds true. The above problem is in the same form as (1) with the only difference that all functions f_n are monotonically decreasing in (l_n, u'_n) and thus fall into case **c)**.

The results of Lemmas 1 and 2 can be summarized as follows. Once the optimal values of the variables associated with functions f_n that are monotonically increasing have been trivially computed through (4), it remains to solve the optimization problem (5) in which the functions f_n belong to either case **b)** or **c)**. In turn, problem (5) is equivalent to problem (8) with only class **c)** functions. This means that we can simply consider optimization problems of the form in (1) in which all functions f_n fall into case **c)**. Accordingly, in the following we assume that (3) is satisfied and only focus on functions f_n that are continuous, strictly convex and monotonically decreasing in the intervals $[l_n, u_n]$. For notational simplicity, however, in all subsequent derivations we maintain the notation given in (1), though we assume that the results of Lemmas 1 and 2 have been already applied.

[1]Notice that in order for problem in (5) and thus for the original problem in (1) to be well-defined it must be $l_n \neq -\infty \; \forall n \in \mathcal{A}$.

C. Possible extensions

An equivalent form of $(\mathcal{P})$, which is sometimes encountered in literature, is given by:

$$\min_{\{x_n\}} \quad \sum_{n=1}^{N} f_n(x_n) \qquad (11)$$

$$\text{subject to} \quad \sum_{n=1}^{j} x_n \geq \rho_j \quad j = 1, 2, \ldots, N$$

$$l_n \leq x_n \leq u_n \quad n = 1, 2, \ldots, N.$$

The above problem can be rewritten in the same form as in (1) simply replacing x_n with $y_n = -x_n$ in (11). In doing this, we obtain

$$\min_{\{y_n\}} \quad \sum_{n=1}^{N} f_n(-y_n) \qquad (12)$$

$$\text{subject to} \quad \sum_{n=1}^{j} y_n \leq -\rho_j \quad j = 1, 2, \ldots, N$$

$$-u_n \leq y_n \leq -l_n \quad n = 1, 2, \ldots, N$$

which is exactly in the same form of $(\mathcal{P})$.

Consider also the following problem

$$\min_{\{x_n\}} \quad \sum_{n=1}^{N} f_n(x_n) \qquad (13)$$

$$\text{subject to} \quad \sum_{n=1}^{j} g_n(x_n) \leq \rho_j \quad j = 1, 2, \ldots, N$$

$$l_n \leq x_n \leq u_n \quad n = 1, 2, \ldots, N$$

in which g_n is a continuos and strictly increasing function. Setting $y_n = g_n(x_n)$ yields

$$\min_{\{x_n\}} \quad \sum_{n=1}^{N} p_n(y_n) \qquad (14)$$

$$\text{subject to} \quad \sum_{n=1}^{j} y_n \leq \rho_j \quad j = 1, 2, \ldots, N$$

$$l'_n \leq y_n \leq u'_n \quad n = 1, 2, \ldots, N.$$

where $p_n = f_n \circ g_n^{-1}$, $l'_n = g_n(l_n)$ and $u'_n = g_n(u_n)$ with $l'_n < u'_n$ since g_n is strictly increasing. Clearly, (14) is in the same form of $(\mathcal{P})$ in (1) provided that p_n is continuous and strictly convex in $[l'_n, u'_n]$, and continuously differentiable in (l'_n, u'_n). This happens for example when: $i)$ f_n is a strictly convex decreasing function and g_n^{-1} is a concave function (or, equivalently, g_n is a convex function); $ii)$ f_n is a strictly convex increasing function and g_n^{-1} is a convex function (or, equivalently, g_n is a concave function).

Similar arguments can be used when g_n in (13) is a strictly decreasing function. This means that the results of this work can also be applied to the case in which the constraints have the following form:

$$\sum_{n=1}^{j} g_n(x_n) \leq \rho_j \qquad (15)$$

with g_n being continuously differentiable and invertible in $[l_n, u_n]$.

III. THE MAIN RESULT

This section proposes an iterative algorithm that computes the solutions $x_n^\star$ for $n = 1, 2, \ldots, N$ in a finite number of steps $L < N$. We begin by denoting

$$h_n(x_n) = -f_n'(x_n) \qquad (16)$$

which is a positive and strictly decreasing function since f_n is by definition monotonically decreasing, strictly convex in $[l_n, u_n]$ and continuously differentiable in (l_n, u_n). We take $h_n(l_n) = \lim_{x_n \to l_n^+} h_n(x_n)$ and $h_n(u_n) = \lim_{x_n \to u_n^-} h_n(x_n)$. We also define the functions $\xi_n(\varsigma)$ for $n = 1, 2, \ldots, N$ as follows

$$\xi_n(\varsigma) = \begin{cases} u_n & 0 \le \varsigma < h_n(u_n) \\ h_n^{-1}(\varsigma) & h_n(u_n) \le \varsigma < h_n(l_n) \\ l_n & h_n(l_n) \le \varsigma \end{cases} \qquad (17)$$

where $0 \le \varsigma < +\infty$ and h_n^{-1} denotes the inverse function of h_n within the interval $[l_n, u_n]$. Since h_n is a continuous and strictly decreasing function, then h_n^{-1} is continuous and strictly decreasing whereas ξ_n is continuous and non-increasing. Functions $\xi_n(\varsigma)$ in (17) can be easily rewritten in the following compact form:

$$\xi_n(\varsigma) = \min\left\{\max\left\{h_n^{-1}(\varsigma), l_n\right\}, u_n\right\} \qquad (18)$$

from which it is seen that each $\xi_n(\varsigma)$ projects $h_n^{-1}(\varsigma)$ onto the interval $[l_n, u_n]$.

Theorem 1. *The solutions of* $(\mathcal{P})$ *are given by*

$$x_n^\star = \xi_n(\sigma_n^\star) \qquad (19)$$

where the quantities $\sigma_n^\star \ge 0$ *for* $n = 1, 2, \ldots, N$ *are some Lagrange multipliers satisfying the following conditions[2]:*

$$0 \le (\sigma_n^\star - \sigma_{n+1}^\star) \perp \left(\sum_{j=1}^n x_j^\star - \rho_n\right) \le 0 \qquad (20)$$

with $\sigma_{N+1}^\star = 0$.

Proof: The proof is given in Appendix B. ∎

From (18) and (19), it easily follows that $x_n^\star$ can be compactly represented as

$$x_n^\star = \min\left\{\max\left\{h_n^{-1}(\sigma_n^\star), l_n\right\}, u_n\right\}. \qquad (21)$$

Lemma 3. *The Lagrange multipliers* $\sigma_n^\star$ *satisfying* (20) *can be computed by means of the iterative procedure illustrated in* **Algorithm 1**.

Proof: The proof is given in Appendix C. ∎

As seen, Algorithm 1 proceeds as follows (see also Section IV for a more intuitive graphical illustration). At the first iteration it sets $j = 0$ and $\gamma_n = \rho_n$, $\forall n$, and for those values of $n \in \{1, 2, \ldots, N\}$ such that

$$\gamma_n < \sum_{i=1}^n u_i \qquad (26)$$

[2]We use $0 \le x \perp y \le 0$ to denote $0 \le x$, $y \le 0$ and $xy = 0$.

Algorithm 1 Iterative procedure for solving $(\mathcal{P})$ in (1).

1) Set $j = 0$ and $\gamma_n = \rho_n$ for every n.
2) **While** $j < N$
   a) Set $\mathcal{N}_j = \{j+1, \ldots, N\}$
   b) For every $n \in \mathcal{N}_j$.
      i) If $\gamma_n < \sum_{i=j+1}^n u_i$ then compute $\varsigma_n^\star$ as the solution of
      $$c_n(\varsigma) = \sum_{i=j+1}^n \xi_i(\varsigma) = \gamma_n \qquad (22)$$
      for ς.
      ii) If $\gamma_n \ge \sum_{i=j+1}^n u_i$ then set
      $$\varsigma_n^\star = 0. \qquad (23)$$
   c) Evaluate
      $$\mu^\star = \max_{n \in \mathcal{N}_j} \varsigma_n^\star \qquad (24)$$
      and
      $$k^\star = \max_{n \in \mathcal{N}_j}\left\{n \mid \varsigma_n^\star = \mu^\star\right\}. \qquad (25)$$
   d) Set $\sigma_n^\star \leftarrow \mu^\star$ for $n = j+1, \ldots, k^\star$.
   e) Use $\sigma_n^\star$ in (19) to obtain $x_n^\star$ for $n = j+1, \ldots, k^\star$.
   f) Set $\gamma_n \leftarrow \gamma_n - \gamma_{k^\star}$ for $n = k^\star + 1, \ldots, N$.
   g) Set $j \leftarrow k^\star$.

---

it computes the unique solution $\varsigma_n^\star$ (see Appendix D for a detailed proof on the existence and uniqueness of $\varsigma_n^\star$) of the following equation

$$c_n(\varsigma) = \sum_{i=1}^n \xi_i(\varsigma) = \gamma_n. \qquad (27)$$

On the other hand, for those values of $n \in \{1, 2, \ldots, N\}$ such that

$$\gamma_n \ge \sum_{i=1}^n u_i \qquad (28)$$

it sets $\varsigma_n^\star = 0$. The values $\varsigma_n^\star$ computed as described above, for $n = 1, 2, \ldots, N$, are used in (24) and (25) to obtain $\mu^\star$ and $k^\star$, respectively. As it follows from (24) and (25), $\mu^\star$ is set equal to the maximum value of $\{\varsigma_n^\star\}$ with $n \in \{1, 2, \ldots, N\}$ while $k^\star$ stands for its corresponding index. Both are then used to replace $\sigma_n^\star$ with $\mu^\star$ for $n = 1, 2, \ldots, k^\star$. Note that if two or more indices can be associated with $\mu^\star$ (meaning that $\varsigma_n^\star = \mu^\star$ for all such indices), then according to (25) the maximum one is selected.

Once $\{\sigma_1^\star, \sigma_2^\star, \ldots, \sigma_{k^\star}^\star\}$ have been computed, Algorithm 1 moves to the second step, which essentially consists in solving

the following reduced problem:

$$\min_{\{x_n\}} \sum_{n=k^\star+1}^{N} f_n(x_n) \tag{29}$$

$$\text{subject to } \sum_{n=k^\star+1}^{j} x_n \le \gamma_j - \gamma_{k^\star} \quad j = k^\star+1, k^\star+2, \dots, N$$

$$l_n \le x_n \le u_n \quad n = k^\star+1, k^\star+2, \dots, N$$

using the same procedure as before. The iterative procedure terminates in a finite number of steps when all quantities $\sigma_n^\star$ are computed. According to Theorem 1, the solutions of $(\mathcal{P})$ for $n = 1, 2, \dots, N$ are eventually obtained as $x_n^\star = \xi_n(\sigma_n^\star)$.

A. Remarks

The following remarks are of interest.

Remark 1. It is worth observing than in deriving Algorithm 1 we have implicitly assumed that the number of linear constraints in (1) is exactly N. When this does not hold true, Algorithm 1 can be slightly modified in an intuitive and straightforward manner. Specifically, let $\mathcal{L} \subset \{1, 2, \dots, N\}$ denote the subset of indices associated to the linear constraints of the optimization problem at hand. In these circumstances, we have that (1) reduces to:

$$\min_{\{x_n\}} \sum_{n=1}^{N} f_n(x_n) \tag{30}$$

$$\text{subject to } \sum_{n=1}^{j} x_n \le \rho_j \quad j \in \mathcal{L}$$

$$l_n \le x_n \le u_n \quad n = 1, 2, \dots, N.$$

The solution of (30) can still be computed through the iterative procedure illustrated in Algorithm 1 once the two following changes are made:

- Step a) – Replace $\mathcal{N}_j$ with $\mathcal{N}_j \cap \mathcal{L}$.
- Step f) – Replace the statement "Set $\gamma_n \leftarrow \gamma_n - \gamma_{k^\star}$ for $n = k^\star+1, k^\star+2, \dots, N$" with "Set $\gamma_n \leftarrow \gamma_n - \gamma_{k^\star}$ for $n \in \{k^\star+1, k^\star+2, \dots, N\} \cap \mathcal{L}$".

As seen, when only a subset $\mathcal{L}$ of constraints must be satisfied, then Algorithm 1 proceeds computing the quantities $\varsigma_n^\star$ only for the indices $n \in \mathcal{L}$.

Remark 2. The number of iterations L required by Algorithm 1 to compute all the Lagrange multipliers $\sigma_n^\star$ (and, hence, to compute all the solutions $x_n^\star$) depends on the cardinality $|\mathcal{L}|$ of $\mathcal{L}$ (or, equivalently, on the number of linear constraints). In general, L is less than or equal to $|\mathcal{L}|$. However, if $|\mathcal{L}| = 1$ only one iteration is required and thus $L = 1$. Also, if there is no linear constraint (which means $\mathcal{L} = \emptyset$ and $|\mathcal{L}| = 0$) the solutions of $(\mathcal{P})$ can be computed without running Algorithm 1 since they are trivially given by $x_n^\star = u_n$. On the other hand, if $|\mathcal{L}| = N$ the maximum number of iterations required is $N-1$. Indeed, assume that at each iteration Algorithm 1 provides only one $x_n^\star$ (which amounts to saying that at the first iteration Algorithm 1 computes $x_1^\star$, at the second $x_2^\star$, and so forth). Accordingly, at the end of the $(N-1)$th iteration the values of $x_1^\star, x_2^\star, \dots, x_{N-1}^\star$ are available, and $x_N^\star$ can be directly

computed as $x_N^\star = \min\{(\rho_N - \sum_{n=1}^{N-1} x_n^\star), u_n\}$ without the need of performing the Nth iteration. For simplicity, in the sequel we assume that the Nth iteration is always performed so that it is assured that the last value of $k^\star$ computed through (25) is always equal to N.

Remark 3. Observe that if there exists one or more values of $j \in \mathcal{L}$ in (30) for which the following condition holds true

$$\rho_j = \sum_{i=1}^{j} l_i \tag{31}$$

then it easily follows that $x_n^\star = l_n$ for $n = 1, 2, \dots, j_{\max}$, with $j_{\max}$ being the maximum value of $j \in \mathcal{L}$ such that the above condition is satisfied. This means that solving (30) basically reduces to find the solution of the following problem:

$$\min_{\{x_n\}} \sum_{n=j_{\max}+1}^{N} f_n(x_n) \tag{32}$$

$$\text{subject to } \sum_{n=j_{\max}+1}^{j} x_n \le \rho_j' \quad j \in \mathcal{L} \setminus \mathcal{C}$$

$$l_n \le x_n \le u_n \quad n = j_{\max}+1, \dots, N$$

in which $\mathcal{C} = \{1, 2, \dots, j_{\max}\}$ and $\rho_j' = \rho_j - \sum_{n=1}^{j_{\max}} l_n$.

Remark 4. For later convenience, we concentrate on the computation of $\mu^\star$ in the last step of Algorithm 1. For this purpose, denote by $\{\mu_1^\star, \mu_2^\star, \dots, \mu_L^\star\}$ and $k_1^\star < k_2^\star < \cdots < k_L^\star$ (with $k_1^\star \ge 1$ and $k_L^\star = N$) the values of $\mu^\star$ and $k^\star$ provided by (24) and (25), respectively, at the end of the L iterations required to solve $(\mathcal{P})$. Setting $k_0^\star = 0$, we may write

$$\sigma_n^\star = \mu_j^\star \quad k_{j-1}^\star+1 \le n \le k_j^\star \text{ and } j = 1, 2, \dots, L \tag{33}$$

with $\{\mu_1^\star, \mu_2^\star, \dots, \mu_{L-1}^\star\}$ such that

$$\sum_{n=k_{j-1}^\star+1}^{k_j^\star} x_n^\star = \rho_{k_j^\star} - \rho_{k_{j-1}^\star} \quad j = 1, 2, \dots, L-1. \tag{34}$$

For $j = L$ two cases may occur, namely $\mu_L^\star > 0$ or $\mu_L^\star = 0$. In the former, $\mu_L^\star$ is such that

$$\sum_{n=k_{L-1}^\star+1}^{k_L^\star} x_n^\star = \rho_{k_L^\star} - \rho_{k_{L-1}^\star} \tag{35}$$

while in the latter we simply have that

$$x_n^\star = u_n \quad n = k_{L-1}^\star+1, \dots, k_L^\star. \tag{36}$$

Remark 5. At any given iteration, Algorithm 1 requires to solve at most $N - k^\star$ non-linear equations (where $k^\star$ is the value obtained from (25) at the previous iteration):

$$c_n(\varsigma) = \sum_{i=k^\star+1}^{n} \xi_i(\varsigma) = \gamma_n \quad n = k^\star+1, k^\star+2, \dots, N. \tag{37}$$

When the solutions $\{\varsigma_n^\star\}$ of the above equations can be computed in closed form, the computational complexity required by each iteration is nearly negligible. On the other hand, when a closed-form does not exist, this may result in excessive computation. In this latter case, a possible means of reducing
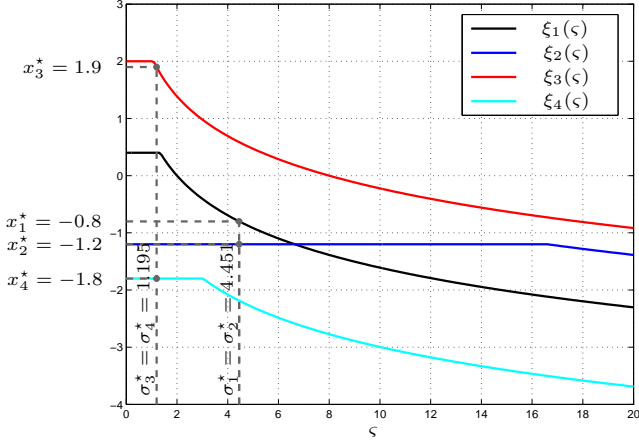
Fig. 1. Graphical illustration of the solutions $x_n^\star$. The intersection of $\xi_n(\varsigma)$ with the vertical dashed line at $\varsigma = \sigma_n^\star$ yields $x_n^\star$.

the computational complexity relies on the fact that $c_n(\varsigma)$ is a non-increasing function as it is the sum of non-increasing functions. Now, assume that the solution of (37) has been computed for $n = n'$. Since we are interested in the maximum between the solutions of (37), as indicated in (24), then for $n'' > n'$ $c_{n''}(\varsigma) = \gamma_{n''}$ must be solved only if $c_{n''}(\varsigma_{n'}^\star) > \gamma_{n''}$. Indeed, only in this case $\varsigma_{n''}^\star$ would be greater than $\varsigma_{n'}^\star$. Accordingly, we may proceed as follows. We start by solving (37) for $n = k^\star + 1$. Then, we look for the first index $n > k^\star + 1$ for which $c_n(\varsigma_{k^\star+1}^\star) > \gamma_n$ and solve the equation associated to such an index. We proceed in this way until $n = N$. In this way, the number of non-linear equations solved at each iteration is smaller than or equal to that required by Algorithm 1.

Remark 6. From the above remark, it follows that the proposed algorithm can be basically seen as composed of two layers. The outer layer computes the Lagrange multipliers $\{\sigma_n^\star\}$ whereas the inner layer evaluates the solution to (37). If the latter can be solved in closed form, then the complexity required by the inner layer is negligible and thus the number of iterations required to solve the problem is essentially given by the number of iterations of the outer layer, which is at most $N - 1$ with N being the number of linear constraints. On the other hand, if the solution to (37) cannot be computed in closed form, then the total number of iterations should also take into account the complexity of the inner layer. However, this cannot be easily quantified as it largely depends on the particular structure of (37) and the specific iterative procedure used to solve it.

IV. GRAPHICAL INTERPRETATIONS

In the next, we provide graphical interpretations of the general policy spelled out by Theorem 1 and Lemma 3.

A. Charts

A direct depiction of Theorem 1 and Lemma 3 can be easily obtained by plotting $c_n(\varsigma)$ and $\xi_n(\varsigma)$ for $n = 1, 2, \ldots, N$ as a

function of $\varsigma \geq 0$. From (22), it follows that the intersections of curves $c_n(\varsigma)$ with the horizontal lines at γ_n yield $\varsigma_n^\star$ from which $\mu^\star$ and $k^\star$ are computed as indicated in (24) and (25). According to (19), the solutions $x_n^\star$ for $n = 1, 2, \ldots, k^\star$ correspond to the interception of the corresponding functions $\xi_n(\varsigma)$ with the vertical line at $\varsigma = \sigma_n^\star = \mu^\star$. Once $x_n^\star$ for $n = 1, 2, \ldots, k^\star$ are computed, the algorithm proceeds with the computation of the remaining solutions by solving the corresponding reduced problem.

For illustration purposes, we assume $N = 4$, $l_n = -\infty$ for any n, $\mathbf{u} = [0.4, -1.2, 2, -1.8]$ and $\boldsymbol{\rho} = [0.2, -2, 1.1, -1.9]$. In addition, we set

$$f_n(x_n) = w_n e^{-x_n} \tag{38}$$

with $[w_1, w_2, w_3, w_4] = [2, 5, 8, 0.5]$. Then, it follows that $h_n(x_n) = w_n e^{-x_n}$ and $h_n^{-1}(\varsigma) = \ln w_n - \ln \varsigma$. Then, from (17) we obtain

$$\xi_n(\varsigma) = \begin{cases} u_n & 0 \leq \varsigma < w_n e^{-u_n} \\ \ln w_n - \ln \varsigma & w_n e^{-u_n} \leq \varsigma \end{cases} \tag{39}$$

or, more compactly,

$$\xi_n(\varsigma) = \min\{\max\{\ln w_n - \ln \varsigma, 0\}, u_n\} \tag{40}$$

whose graph is shown in Fig. 1.

As seen, the first operation of Algorithm 1 is to compute the quantities $\varsigma_n^\star$ for $n = 1, \ldots, 4$ according to step b). Since the condition $\gamma_n \leq \sum_{i=1}^n u_i$ is satisfied for $n = 1, 2, \ldots, 4$, the computation of $\varsigma_n^\star$ requires to solve (22) for $n = 1, 2, \ldots, 4$. Using (40), we easily obtain:

$$\varsigma_1^\star = e^{\ln w_1 - \gamma_1} = 1.637 \tag{41}$$

$$\varsigma_2^\star = e^{\ln w_1 + u_2 - \gamma_2} = 4.451 \tag{42}$$

$$\varsigma_3^\star = e^{\ln w_3 + u_1 + u_2 - \gamma_3} = 1.196 \tag{43}$$

$$\varsigma_4^\star = e^{\frac{\ln w_1 + \ln w_3 + u_2 + u_4 - \gamma_4}{2}} = 2.307. \tag{44}$$

A direct depiction of the above results can be easily obtained by plotting $c_n(\varsigma)$ for $n = 1, 2, \ldots, 4$ as a function of $\varsigma \geq 0$. As shown in Fig. 2, the intersections of curves $c_n(\varsigma)$ with the horizontal lines at $\varsigma = \gamma_n$ yield $\varsigma_n^\star$.

Using the above results into (24) and (25) of step c) yields $\mu^\star = 4.451$ and $k^\star = 2$ from which (according to step d)) we obtain

$$\sigma_1^\star = \sigma_2^\star = \mu^\star = 4.451. \tag{45}$$

Once the optimal $\sigma_1^\star$ and $\sigma_2^\star$ are computed, Algorithm 1 proceeds solving the following reduced problem:

$$\min_{\{x_3, x_4\}} \sum_{n=3}^4 w_n e^{-x_n} \tag{46}$$

$$\text{subject to} \quad \sum_{n=3}^j x_n \leq \gamma_j \quad j = 3, 4$$

$$x_n \leq u_n \quad n = 3, 4$$

with $\gamma_3 = 3.1$ and $\gamma_4 = 0.1$ as obtained from $\gamma_j \leftarrow \gamma_j - \gamma_{k^\star}$ observing that $\gamma_{k^\star} = \gamma_2 = -2$. Since $\gamma_3 > u_3$, from step b)
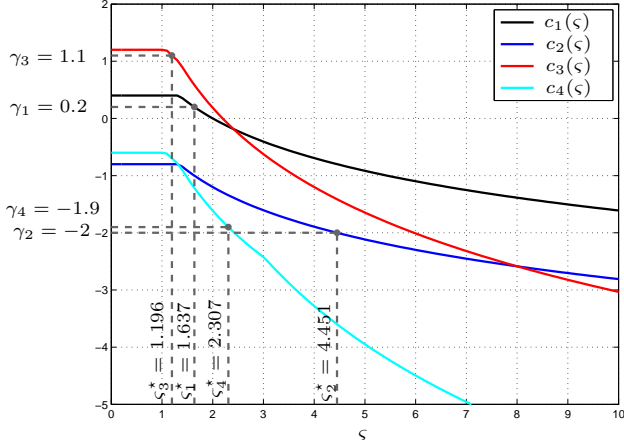
Fig. 2. Graphical illustration of $c_n(\varsigma)$. Their intersection with the horizontal dashed lines at $\gamma_1 = 0.2$, $\gamma_2 = -2$, $\gamma_3 = 1.1$ and $\gamma_4 = -1.9$ yields respectively $\varsigma_1^\star = 1.637$, $\varsigma_2^\star = 4.451$, $\varsigma_3^\star = 1.196$ and $\varsigma_4^\star = 2.307$.

we have that $\varsigma_3^\star = 0$ while $\varsigma_4^\star$ turns out to be given by

$$\varsigma_4^\star = e^{\ln w_3 + u_4 - \gamma_4} = 1.195. \tag{47}$$

As before, $\varsigma_4^\star$ can be obtained as the intersection of new function

$$c_4(\varsigma) = \sum_{n=3}^{4} \xi_n(\varsigma) \tag{48}$$

with the horizontal line at $\varsigma = \gamma_4 = 0.1$. Then, from (24) and (25), we have that

$$\mu^\star = \max_{n=3,4} \varsigma_n^\star = 1.195 \tag{49}$$

and thus $k^\star = 4$. This means that $\sigma_3^\star = \sigma_4^\star = 1.195$.

The optimal $x_n^\star$ are eventually obtained as $x_n^\star = \xi_n(\sigma_n^\star)$. This yields $x_1^\star = -0.8$, $x_2^\star = -1.2$, $x_3^\star = 1.9$ and $x_4^\star = -1.8$. As depicted in Fig. 1, the solution $x_n^\star$ corresponds to the interception of $\xi_n(\varsigma)$ with the vertical line at $\varsigma = \sigma_n^\star$.

B. Water-filling inspired policy

While the charts used in the foregoing example are quite useful, we put forth an alternative interpretation that allows retaining some of the intuition of the water-filling policy. This interpretation is valid for cases in which the optimization variables $\{x_n; n = 1, 2, \ldots, N\}$ can only take non-negative values, which amounts to setting $l_n = 0$ for $n = 1, 2, \ldots, N$.

We start considering the simple case in which a single linear constraint is imposed:

$$\min_{\{x_n\}} \quad \sum_{n=1}^{N} f_n(x_n) \tag{50}$$

$$\text{subject to} \quad \sum_{n=1}^{N} x_n \leq \rho_N$$

$$0 \leq x_n \leq u_n \quad n = 1, 2, \ldots, N.$$

Using the results of Theorem 1 and Lemma 3, the solution to (50) is found to be

$$x_n^\star = \xi_n(\sigma^\star) = \min\left\{\max\left\{h_n^{-1}(\sigma_n^\star), 0\right\}, u_n\right\} \tag{51}$$
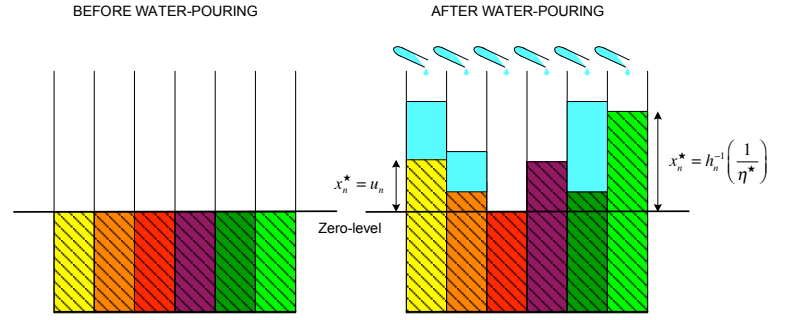

Fig. 3. Water-filling inspired interpretation of the solutions $x_n^\star$.

where the values of $\sigma_n^\star$ are obtained through Algorithm 1. Since a single constraint is present in (54), then a single iteration is required to compute all the values of $\sigma_n^\star$ for $n = 1, 2, \ldots, N$. In particular, it turns out that $\sigma_n^\star = \sigma^\star$ for any n, with $\sigma^\star$ such that the following condition is satisfied:

$$\sum_{n=1}^{N} x_n^\star = \sum_{n=1}^{N} \xi_n(\sigma^\star) = \rho_N. \tag{52}$$

Consider now N vessels, which are filled with a proper material (different from vessel to vessel) up to a given level. Think of it as the *zero-level* and assume that it is the same for all vessels, as illustrated in Fig. 3 for $N = 6$. Assume that a certain quantity η of water (measured in proper units) is poured into each vessel and let each material be able to first absorb it and then to expand accordingly up to a certain level. In particular, assume that the behaviour of material n is regulated by $\xi_n(\varsigma)$ with $\varsigma = 1/\eta$. More precisely, $\xi_n(\varsigma)$ is the difference between the new level of material n and the zero-level. From (17), it easily follows that the expansion starts only when η reaches the level $\eta = 1/h_n(0)$ while it stops when $\eta = 1/h_n(u_n)$, corresponding to a maximum expansion of $\xi_n(\varsigma) = u_n$. This means that additional water beyond the quantity $1/h_n(u_n)$ does not produce any further expansion - it is simply accumulated in vessel n above the level u_n as depicted in Fig. 3.

Using (51) and (52), the solutions $\{x_n^\star\}$ to (50) can thus be interpreted as obtained trough the following procedure, which is reminiscent of the water-filling policy.

1) Consider N vessels;
2) Assume vessel n is filled with a proper material up to a certain zero-level (the same for each vessel);
3) Let the behaviour of material n be regulated by ξ_n;
4) Compute $\sigma^\star$ through (52);
5) Poor the same quantity $\eta^\star = 1/\sigma^\star$ of water into each vessel;
6) The material height over the zero-level in vessel n gives $x_n^\star$.

The extension of the above water-filling interpretation to the general form in (1) is straightforward. Assume that the jth iteration is considered. Then, Algorithm 1 proceeds as follows.

1) Consider N_j vessels with indices $n = j + 1, \ldots, N$;
2) Assume vessel n is filled with a proper material up to a certain zero-level (the same for each vessel);
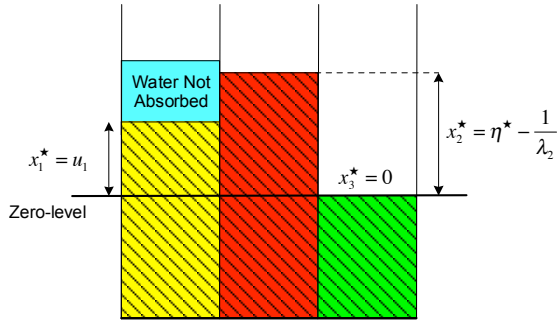3) Let the behaviour of material n be regulated by ξ_n;

Fig. 4. Illustration of the water-filling inspired policy for problem (54) when $N = 3$.

Fig. 5. Illustration of the cave-filling policy for problem (54) when $N = 3$.

4) Compute $\mu^\star$ and $k^\star$ through (24) and (25);
5) Poor the same quantity $\eta^\star = 1/\mu^\star$ of water into vessels $n = j + 1, \ldots, k^\star$;
6) The material height over the zero-level gives $x_n^\star$ for $n = j + 1, \ldots, k^\star$.

Remark 7. Observe that the speed by which material n expands itself depends on ξ_n' defined as the first derivative of ξ_n with respect to $\eta = 1/\varsigma$. It can be easily shown that

$$\xi_n' = \frac{1}{\eta^2 f_n'' \left(h_n^{-1}(1/\eta) \right)} \tag{53}$$

from which it follows that the rate of growth is inversely proportional to the second derivative of f_n evaluated at $h_n^{-1}(1/\eta)$.

V. PARTICULARIZATION TO POWER ALLOCATION PROBLEMS

In the following, we show how some power allocation problems in signal processing and communications can be put in the form of (1), and thus can be solved with the generalized algorithm illustrated above[3].

A. Classical water-filling and cave-filling policies

Consider the classical problem of allocating a certain amount of power P among a bank of non-interfering channels to maximize the capacity. This problem can be mathematically formulated as follows:

$$\max_{\{x_n\}} \quad \sum_{n=1}^{N} \log(1 + \lambda_n x_n) \tag{54}$$

$$\text{subject to} \quad \sum_{n=1}^{N} x_n \leq P$$

$$0 \leq x_n \leq u_n \quad n = 1, 2, \ldots, N$$

where x_n represents the transmit power allocated over the nth channel of gain λ_n whereas $\log(1 + \lambda_n x_n)$ gives the capacity of the nth channel. Clearly, we assume that $\sum_{n=1}^{N} u_n > P$, otherwise (54) has the trivial solution $x_n^\star = u_n$.

[3]Due to the considerable amount of works in this field, our exposition will be necessarily incomplete and will reflect the subjective tastes and interests of the authors. To compensate for this partiality, we refer the interested reader to the list of references for an entree into the extensive literature on this subject.
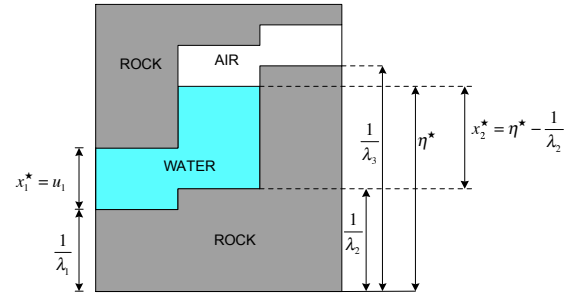
The above problem can be put in the same form of (30) setting $f_n(x_n) = -\log(1 + \lambda_n x_n)$, $l_n = 0 \; \forall n$, $\mathcal{L} = \{N\}$ and $\rho_N = P$. Observing that

$$h_n^{-1}(\varsigma) = \frac{1}{\varsigma} - \frac{1}{\lambda_n} \tag{55}$$

from (21) one gets

$$x_n^\star = \min \left\{ \max \left\{ \frac{1}{\sigma^\star} - \frac{1}{\lambda_n}, 0 \right\}, u_n \right\} \tag{56}$$

with $\sigma^\star$ such that

$$\sum_{n=1}^{N} x_n^\star = \sum_{n=1}^{N} \xi_n(\sigma^\star) = P. \tag{57}$$

Using the water-filling policy illustrated in Section IV, the solutions in (56) have the visual interpretation shown in Fig. 4, where we have assumed $N = 3$ and set $\eta^\star = 1/\sigma^\star$. The material inside the nth vessel starts expanding when the quantity of water η poured in the vessel equals $1/\lambda_n$. Due to the particular form of f_n, the expansion follows the linear law $\xi_n(1/\eta) = \eta - 1/\lambda_n$ as long as $\eta \leq u_n + 1/\lambda_n$. After that, water is no more absorbed and the expansion stops. The additional water is accumulated in the vessel above the maximum level of the material. As shown in Fig. 4, this is precisely what happens with the yellow material in vessel 1. On the other hand, we have that $\eta^\star - 1/\lambda_2 < u_2$ and thus no water is accumulated on the top of the red material in vessel 2. Finally, the green material in vessel 3 is such that no expansion occurs since $\eta^\star < 1/\lambda_3$.

An alternative visual interpretation of (56) (commonly used in the literature) is given in Fig. 5, where $1/\lambda_n$ and $u_n + 1/\lambda_n$ are viewed as the ground and the ceiling levels of patch n, respectively. In this case, the solution is computed as follows. We start by flooding the region with water to a level η. The total amount of water used is then given by

$$\sum_{n=1}^{N} \min \left\{ \max \left\{ \eta - \frac{1}{\lambda_n}, 0 \right\}, u_n \right\}. \tag{58}$$

The flood level is increased until a total amount of water equal to P is used. The depth of water inside patch n gives $x_n^\star$. This solution method is known as cave-filling due to its specific physical meaning. Clearly, if $u_n = +\infty$ for any n in (54) then $x_n^\star$ reduces to

$$x_n^\star = \max \left\{ \frac{1}{\sigma^\star} - \frac{1}{\lambda_n}, 0 \right\}. \tag{59}$$
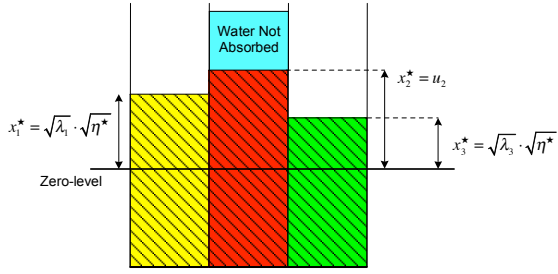
Fig. 6. Illustration of the water-filling inspired policy for problem (60) .

which is the well-known and classical water-filling solution.

A problem whose solution has the same visual interpretation of Fig. 5 is considered also in [12] (see problem (21)) in which the authors design the optimal training sequences for channel estimation in multi-hop transmissions using decode-and-forward protocols.

B. General water-filling policies

Consider now the following problem:

$$\min_{\{x_n\}} \quad \sum_{n=1}^{N} \frac{\lambda_n}{x_n} \tag{60}$$

$$\text{subject to} \quad \sum_{n=1}^{j} x_n \leq \rho_j \quad j = 1, 2, \ldots, N$$

$$0 \leq x_n \leq 1 \quad n = 1, 2, \ldots, N$$

where $\{\lambda_n > 0\}$ are positive parameters. This problem is considered in [8] in the context of linear transceiver design architectures for MIMO networks with a single non-regenerative relay. It also appears in [1] where the authors deal with the linear transceiver design problem in MIMO point-to-point networks to minimize the power consumption while satisfying specific QoS constraints on the mean-square-errors (MSEs). A similar instance can also be found in [22] and corresponds to the minimization of the weighted arithmetic mean of the MSEs in a multicarrier MIMO system with a total power constraint. All the above examples could in principle be solved with (specifically designed) multi-level water-filling algorithms [11]. Easy reformulations allow to use the more general Algorithm 1 as shown next for problem (60).

Setting $f_n(x_n) = \lambda_n / x_n$ and letting $l_n = 0$ and $u_n = 1, \forall n$, it is easily seen that (60) has the same form as (1). Then, one gets $h_n(x_n) = \lambda_n / x_n^2$ and $h_n^{-1}(\varsigma) = \sqrt{\lambda_n / \varsigma}$. The solution to (60) is given by

$$x_n^\star = \min\left\{\max\left\{\sqrt{\frac{\lambda_n}{\sigma_n^\star}}, 0\right\}, 1\right\} \tag{61}$$

where $\{\sigma_1^\star, \sigma_2^\star, \ldots, \sigma_N^\star\}$ are computed through Algorithm 1 and take the form (33) with $\{\mu_1^\star, \mu_2^\star, \ldots, \mu_{L-1}^\star\}$ such that

$$\sum_{n=k_{j-1}^\star+1}^{k_j^\star} x_n^\star = \rho_{k_j^\star} - \rho_{k_{j-1}^\star} \quad j = 1, 2, \ldots, L-1. \tag{62}$$

According to Remark 4, if $\mu_L^\star$ is greater than 0 then

$$\sum_{n=k_{L-1}^\star+1}^{N} x_n^\star = \rho_N - \rho_{k_{L-1}^\star} \tag{63}$$

otherwise when $\mu_L^\star = 0$ one gets

$$x_n^\star = 1 \quad n = k_{L-1}^\star + 1, \ldots, N. \tag{64}$$

The solutions $x_n^\star$ in (61) can be thought as obtained through the water-filling policy illustrated in Section IV in which the expansion of material n is regulated by the square-root law $\xi_n(1/\eta) = \sqrt{\lambda_n \eta}$ with rate of growth given by

$$\xi_n(1/\eta) = \sqrt{\frac{\lambda_n}{\eta}}, \tag{65}$$

according to (53). This is illustrated in Fig. 6 wherein we consider the first iteration of Algorithm 1 under the assumption that $k_1^\star = 3$ and $\lambda_3 < \lambda_1 < \lambda_2$. As expected, the level of the red material in the 2nd vessel is higher than the others.

C. Some other examples

Consider now the following problem:

$$\min_{\{x_n\}} \quad \sum_{n=1}^{N} \lambda_n e^{-x_n} \tag{66}$$

$$\text{subject to} \quad \sum_{n=1}^{j} x_n \leq \rho_j \quad j = 1, 2, \ldots, N$$

$$x_n \leq 0 \quad n = 1, 2, \ldots, N.$$

The above problem arises in [4] where the authors deal with the power minimization in MIMO point-to-point networks with non-linear architectures at the transmitter or at the receiver. A similar problem arises when two-hop MIMO networks with a single amplify-and-forward relay are considered [8]. The solution of (66) has the form

$$x_n^\star = \min\left\{\max\left\{\log\left(\frac{\lambda_n}{\sigma_n^\star}\right), 0\right\}, u_n\right\} \tag{67}$$

where the quantities $\sigma_n^\star$ are given by (33).

Another instance of (1) arises in connection with the computation of the optimal power allocation for the maximization of the instantaneous received signal-to-noise ratio in amplify-and-forward multi-hop transmissions under short-term power constraints [13]. Denoting by N the total number of hops, the problem can be mathematically formalized as follows [13]

$$\max_{\{x_n\}} \quad \left(\prod_{n=1}^{N}\left(1 + \frac{1}{x_n \lambda_n}\right) - 1\right)^{-1} \tag{68}$$

$$\text{subject to} \quad \sum_{n=1}^{N} x_n \leq P$$

$$0 \leq x_n \leq u_n \quad n = 1, 2, \ldots, N.$$

where x_n represents the power allocated over the nth hop and P denotes the available power. In addition, λ_n is the channel

gain over the nth hop. The above problem can be equivalently reformulated as follows

$$\min_{\{x_n\}} \quad \sum_{n=1}^{N} \log\left(1 + \frac{1}{x_n \lambda_n}\right) \tag{69}$$

$$\text{subject to} \quad \sum_{n=1}^{N} x_n \leq P$$

$$0 \leq x_n \leq u_n \quad n = 1, 2, \ldots, N$$

from which it is clear that it is in the same form as (30) with

$$f_n(x_n) = \log\left(1 + \frac{1}{x_n \lambda_n}\right) \tag{70}$$

$\mathcal{L} = \{N\}$ and $\rho_N = P$. Then,

$$h_n^{-1}(\varsigma) = \frac{\sqrt{1 + \frac{4\lambda_n}{\varsigma}} - 1}{2\lambda_n}. \tag{71}$$

It is assumed $\sum_{n=1}^{N} u_n > P$, otherwise (69) has the trivial solution $x_n^\star = p_n$. Using (71) into (19) yields

$$x_n^\star = \min\left\{\max\left\{\frac{1}{2\lambda_n}\left(\sqrt{1 + \frac{4\lambda_n}{\sigma^\star}} - 1\right), 0\right\}, u_n\right\} \tag{72}$$

with $\sigma^\star$ such that $\sum_{n=1}^{N} x_n^\star = P$.

VI. CONCLUSIONS

An iterative algorithm has been proposed to compute the solution of separable convex optimization problems with a set of linear and box constraints. The proposed solution operates through a two layer architecture, which has a simple graphical water-filling inspired interpretation. The outer layer requires at most $N-1$ steps with N being the number of linear constraints whereas the number of iterations of the inner layer depends on the complexity of solving a set of (possibly) non-linear equations. If solvable in closed form, then the computational burden of the inner layer is negligible. The problem under investigation is particularly interesting since a large number of existing (and likely future) power allocation problems in signal processing and communications can be reformulated as instances of its general form, and thus can be solved with the proposed algorithm without the need of developing specific solutions for each of them.

APPENDIX A
PROOF OF LEMMAS 1 AND 2

We start considering case **a**). Without loss of generality, we concentrate on f_1, which is assumed monotonically increasing in $[l_1, u_1]$, and aim at proving that $x_1^\star = l_1$. We start denoting by $\mathcal{S}(x_1)$ the feasible set of $x_2, x_3, \ldots, x_N$ for a given $x_1 \in [l_1, u_1]$. Mathematically, $\mathcal{S}(x_1)$ is such that

$$\sum_{n=2}^{j} x_n \leq \rho_n - x_1 \quad j = 2, \ldots, N \tag{73}$$

$$l_n \leq x_n \leq u_n \quad n = 2, \ldots, N.$$

Clearly, we have that $\mathcal{S}(x_1) \subseteq \mathcal{S}(l_1)$ for any $x_1 \in (l_1, u_1]$. For notational convenience, we also define $F(x_1)$ as

$$F(x_1) = \min_{\{x_2, x_3, \ldots, x_N\} \in \mathcal{S}(x_1)} \sum_{n=2}^{N} f_n(x_n). \tag{74}$$

Observe now that the optimal value $x_1^\star$ is such that $f_1(x_1) + F(x_1)$ is minimized. To this end, we recall that: **i)** $f_1(l_1) < f_1(x_1)$ since f_1 is strictly increasing in $[l_1, u_1]$; **ii)** $F_1(l_1) \leq F_1(x_1)$ since $\mathcal{S}(x_1) \subseteq \mathcal{S}(l_1)$ for any $x_1 \in (l_1, u_1]$. Therefore, it easily follows that $f_1(l_1) + F(l_1) < f_1(x_1) + F(x_1)$ for any $x_1 \in (l_1, u_1]$, which proves that $x_1^\star = l_1$. The same result can easily be extended to a generic x_n with $n \neq 1$ using similar arguments. This proves Lemma 1.

Consider now case **b)** and assume that there exists a point z_n in (l_n, u_n) such that $f_n'(z_n) = 0$ with $f_n'(x_n) < 0 \,\forall x_n \in (l_n, z_n)$ and $f_n'(x_n) > 0 \,\forall x_n \in (z_n, u_n)$. We aim at proving that $x_n^\star \in [l_n, z_n]$

Since $f_n'(x_n) > 0 \,\forall x_n \in (z_n, u_n)$, then $f_n(x_n)$ is monotonically increasing in $[z_n, u_n]$. Consequently, by Lemma 1 it follows that $x_n^\star$ cannot be greater than z_n. This amounts to saying that $x_n^\star$ must belong to the interval $[l_n, z_n]$, as stated in (7).

Finally, for case **c)** nothing can be said a priori apart for that the solution $x_n^\star$ lies in interval $[l_n, u_n]$ as required by the box constraints in (1).

APPENDIX B
PROOF OF THEOREM 1

We begin by writing the Karush-Kuhn-Tucker (KKT) conditions for the convex problem $(\mathcal{P})$:

$$-h_n(x_n) + \sum_{j=n}^{N} \lambda_j + \nu_n - \kappa_n = 0 \qquad n = 1, \ldots, N \tag{75}$$

$$0 \leq \lambda_n \perp \left(\sum_{j=1}^{n} x_j - \rho_n\right) \leq 0 \qquad n = 1, \ldots, N \tag{76}$$

$$0 \leq \nu_n \perp (x_n - u_n) \leq 0 \qquad n = 1, \ldots, N \tag{77}$$

$$0 \leq \kappa_n \perp (x_n - l_n) \geq 0 \qquad n = 1, \ldots, N \tag{78}$$

where $h_n(x) = -f_n'(x)$. Letting $\sigma_n = \sum_{j=n}^{N} \lambda_j$ and $\sigma_{N+1} = 0$, we may rewrite (75) – (78) in the following equivalent form:

$$-h_n(x_n) + \sigma_n + \nu_n - \kappa_n = 0 \qquad n = 1, \ldots, N \tag{79}$$

$$\sigma_n \geq 0 \qquad \sigma_{N+1} = 0 \tag{80}$$

$$0 \leq (\sigma_n - \sigma_{n+1}) \perp \left(\sum_{j=1}^{n} x_j - \rho_n\right) \leq 0 \quad n = 1, \ldots, N \tag{81}$$

$$0 \leq \nu_n \perp (x_n - u_n) \leq 0 \qquad n = 1, \ldots, N \tag{82}$$

$$0 \leq \kappa_n \perp (x_n - l_n) \geq 0 \qquad n = 1, \ldots, N. \tag{83}$$

Since $(\mathcal{P})$ is convex, solving the KKT conditions is equivalent to solving $(\mathcal{P})$. Accordingly, we let $x_n^\star$, $\nu_n^\star$, $\kappa_n^\star$ and $\sigma_n^\star$ to denote the solution of (79) – (83) for $n = 1, \ldots, N$. In the next, it is shown that $x_n^\star$, $\nu_n^\star$ and $\kappa_n^\star$ are given by

$$x_n^\star = \xi_n(\sigma_n^\star) \tag{84}$$

$$\nu_n^\star = \max\{h_n(x_n^\star) - \sigma_n^\star, 0\} \qquad (85)$$

$$\kappa_n^\star = \max\{\sigma_n^\star - h_n(x_n^\star), 0\} \qquad (86)$$

where $\xi_n(\sigma_n^\star)$ is computed as in (17) with $\varsigma = \sigma_n^\star$:

$$\xi_n(\sigma_n^\star) = \begin{cases} u_n & 0 \le \sigma_n^\star < h_n(u_n) \\ h_n^{-1}(\sigma_n^\star) & h_n(u_n) \le \sigma_n^\star < h_n(l_n) \\ l_n & h_n(l_n) \le \sigma_n^\star. \end{cases} \qquad (87)$$

The following three cases are considered separately: **a)** $x_n^\star = l_n$; **b)** $l_n < x_n^\star < u_n$; **c)** $x_n^\star = u_n$.

Case **a)** If $x_n^\star = l_n$ then from (82) it immediately follows $\nu_n^\star = 0$ whereas (79) reduces to $-h_n(x_n^\star) + \sigma_n^\star = \kappa_n^\star$, from which using (83) we get

$$-h_n(x_n^\star) + \sigma_n^\star \ge 0 \qquad (88)$$

or, equivalently, $\sigma_n^\star \ge h_n(x_n^\star) = h_n(l_n)$. Using the above result into (87) yields

$$x_n^\star = l_n = \xi_n(\sigma_n^\star) \qquad (89)$$

as stated in (84). From the above results, it also follows that:

$$\nu_n^\star = 0 = \max\{h_n(x_n^\star) - \sigma_n^\star, 0\} \qquad (90)$$

$$\kappa_n^\star = \sigma_n^\star - h_n(x_n^\star) = \max\{\sigma_n^\star - h_n(x_n^\star), 0\} \qquad (91)$$

as given in (85) and (86), respectively.

Case **b)** From (82) and (83) we obtain $\nu_n^\star = 0$ and $\kappa_n^\star = 0$ so that (79) reduces to

$$-h_n(x_n^\star) + \sigma_n^\star = 0 \qquad (92)$$

from which we have that $x_n^\star = -h_n^{-1}(\sigma_n^\star)$. Since in this case $l_n < x_n^\star < u_n$, then

$$h_n(u_n) < \sigma_n^\star < h_n(l_n) \qquad (93)$$

so that we obtain

$$x_n^\star = h_n^{-1}(\sigma_n^\star) = \xi_n(\sigma_n^\star). \qquad (94)$$

Also, taking (92) into account yields

$$\nu_n^\star = 0 = \max\{h_n(x_n^\star) - \sigma_n^\star, 0\} \qquad (95)$$

$$\kappa_n^\star = 0 = \max\{\sigma_n^\star - h_n(x_n^\star), 0\}. \qquad (96)$$

Case **c)** In this case, from (83) one gets $\kappa_n^\star = 0$ whereas (79) reduces to $\nu_n^\star = h_n(x_n^\star) - \sigma_n^\star$. Since (82) is satisfied for $\nu_n^\star \ge 0$, then $h_n(x_n^\star) - \sigma_n^\star \ge 0$ or, equivalently,

$$\sigma_n^\star \le h_n(x_n^\star) = h_n(u_n). \qquad (97)$$

Accordingly, we can write

$$x_n^\star = u_n = \xi_n(\sigma_n^\star) \qquad (98)$$

$$\nu_n^\star = h_n(x_n^\star) - \sigma_n^\star = \max\{h_n(x_n^\star) - \sigma_n^\star, 0\} \qquad (99)$$

and

$$\kappa_n^\star = 0 = \max\{\sigma_n^\star - h_n(x_n^\star), 0\}. \qquad (100)$$

Using all the above results together, (84) – (86) easily follow from which it is seen that $x_n^\star$ depends solely on $\sigma_n^\star$. The latter must be chosen so as to satisfy (80) and (81).

Algorithm 2 Equivalent form of **Algorithm 1**.

1) Set $j = 1$, $k_0^\star = 0$ and $\gamma_n = \rho_n$ for every n.
2) **While** $k_{j-1}^\star < N$
   a) Set $\mathcal{N}_{k_j^\star} = \{k_{j-1}^\star + 1, k_{j-1}^\star + 2, \ldots, N\}$.
   b) For every n in $\mathcal{N}_{k_j^\star}$.
      i) If $\gamma_n < \sum_{i=k_{j-1}^\star+1}^{n} u_i$ then compute $\varsigma_{n,j}^\star$ as the solution of

$$c_n(\varsigma; j) = \sum_{i=k_{j-1}^\star+1}^{n} \xi_i(\varsigma) = \gamma_n \qquad (101)$$

      for ς.
      ii) If $\gamma_n \ge \sum_{i=k_{j-1}^\star+1}^{n} u_i$ then set

$$\varsigma_{n,j}^\star = 0. \qquad (102)$$

   c) Evaluate

$$\mu_j^\star = \max_{n \in \mathcal{N}_{k_j^\star}} \varsigma_{n,j}^\star \qquad (103)$$

   and

$$k_j^\star = \max_{n \in \mathcal{N}_{k_j^\star}} \{n | \varsigma_{n,j}^\star = \mu_j^\star\}. \qquad (104)$$

   d) Set

$$\sigma_n^\star \leftarrow \mu_j^\star \quad \text{for } k_{j-1}^\star < n \le k_j^\star. \qquad (105)$$

   e) Set $\gamma_n \leftarrow \rho_n - \rho_{k_{j-1}^\star}$ for $k_{j-1}^\star < n \le N$.
   f) Set $j \leftarrow j + 1$.

APPENDIX C
PROOF OF LEMMA 3

For the sake of clarity, the steps of Algorithm 1 are put in the equivalent forms illustrated in **Algorithm 2** in which basically some indices and equations are introduced or reformulated in order to ease understanding of the mathematical arguments and steps reported below.

As seen, the jth iteration of Algorithm 2 computes the real parameter $\mu_j^\star$ and the integer $k_j^\star \in \{k_{j-1}^\star + 1, \ldots, N\}$ through (103) and (104), respectively. The latter are then used in (105) to obtain $\sigma_n^\star$ for $n = k_{j-1}^\star + 1, \ldots, k_j^\star$:

$$\sigma_n^\star = \mu_j^\star \quad \text{for } k_{j-1}^\star < n \le k_j^\star. \qquad (106)$$

In the next, it is shown that the quantities $\sigma_n^\star$ given by (106) satisfy (80) and (81).

We start proving that $\sigma_n^\star \ge 0$ as required in (80). Since the domain of the function $c_n(\varsigma; j)$ in (101) is the interval $[0, +\infty)$ then the solution $\varsigma_{n,j}^\star$ of $c_n(\varsigma; j) = \rho_n - \rho_{k_{j-1}^\star}$ is non-negative or it does not exist. In the latter case, Algorithm 2 sets $\varsigma_{n,j}^\star = 0$ (according to (102)) and hence $\varsigma_{n,j}^\star \ge 0$ in any case. This means that $\mu_j^\star$, as computed through (103), is non-negative and, consequently, $\sigma_n^\star$ is non-negative as well.

To proceed further, we now show that

$$\sigma_n^\star - \sigma_{n+1}^\star \ge 0 \qquad (107)$$

for $n = 1, \ldots, N$ as required in (81). To this end, we start

observing that $\forall j$

$$\sigma_n^\star - \sigma_{n+1}^\star = 0 \quad \text{for } k_{j-1}^\star < n < k_j^\star \tag{108}$$

as immediately follows from (106). On the other hand, for $n = k_j^\star$ one has

$$\sigma_{k_j^\star}^\star - \sigma_{k_j^\star+1}^\star = \mu_j^\star - \mu_{j+1}^\star. \tag{109}$$

From (108) and (109), it clearly follows that to prove (107) it suffices to show that $\mu_j^\star \geq \mu_{j+1}^\star$. To see how this comes about, we start observing that since each $\xi_n(\varsigma)$ in (17) is non-increasing then $c_n(\varsigma; j)$ in (101) is non-increasing as well, so that we may write

$$c_n(\mu_j^\star; j) \leq c_n(\varsigma_{n,j}^\star; j) = \rho_n - \rho_{k_{j-1}^\star} \quad n = k_{j-1}^\star + 1, \ldots, N \tag{110}$$

where we have taken into account that by definition $\mu_j^\star \geq \varsigma_{n,j}^\star$ (as it follows from (103)). In particular, (110) is satisfied with equality for $n = k_j^\star$ whereas it is a strict inequality for $n > k_j^\star$. Indeed, it cannot exist an index $\bar{k} > k_j^\star$ such that $c_{\bar{k}}(\mu_j^\star; j) = \rho_{\bar{k}} - \rho_{k_{j-1}^\star}$ because this would mean that $\mu_j^\star$ is solution of both $c_{\bar{k}}(\varsigma; j) = \rho_{\bar{k}} - \rho_{k_{j-1}^\star}$ and $c_{k_j^\star}(\varsigma; j) = \rho_{k_j^\star} - \rho_{k_{j-1}^\star}$. If that is the case, in applying (104) at the jth step $\bar{k}$ would have been chosen instead of $k_j^\star$.

Based on the above results, setting $n = k_{j+1}^\star > k_j^\star$ into (110) yields

$$c_{k_{j+1}^\star}(\mu_j^\star; j) < \rho_{k_{j+1}^\star} - \rho_{k_{j-1}^\star}. \tag{111}$$

Also, a close inspection of (101) reveals that for $n > k_j^\star$ $c_n(\varsigma; j)$ can be rewritten as follows

$$c_n(\varsigma; j) = c_{k_j^\star}(\varsigma; j) + c_n(\varsigma; j+1) \tag{112}$$

from which setting $n = k_{j+1}^\star$ we obtain

$$c_{k_{j+1}^\star}(\varsigma; j) = c_{k_j^\star}(\varsigma; j) + c_{k_{j+1}^\star}(\varsigma; j+1). \tag{113}$$

Replacing ς with $\mu_j^\star$ in (113) yields

$$c_{k_{j+1}^\star}(\mu_j^\star; j) = \rho_{k_j^\star} - \rho_{k_{j-1}^\star} + c_{k_{j+1}^\star}(\mu_j^\star; j+1) \tag{114}$$

where we have taken into account that

$$c_{k_j^\star}(\mu_j^\star; j) = \rho_{k_j^\star} - \rho_{k_{j-1}^\star} \tag{115}$$

as it easily follows from the definition of $c_n(\varsigma; j)$ in (101) and from those of $\mu_j^\star$ and $k_j^\star$ in (103) and (104).

Using (111) with (114) leads to

$$c_{k_{j+1}^\star}(\mu_j^\star; j+1) < \rho_{k_{j+1}^\star} - \rho_{k_j^\star} \tag{116}$$

from which recalling that

$$c_{k_{j+1}^\star}(\mu_{j+1}^\star; j+1) = \rho_{k_{j+1}^\star} - \rho_{k_j^\star} \tag{117}$$

we obtain

$$c_{k_{j+1}^\star}(\mu_j^\star; j+1) < c_{k_{j+1}^\star}(\mu_{j+1}^\star; j+1). \tag{118}$$

Since the functions $c_n(\varsigma; j)$ are non-increasing, from the above inequality we eventually obtain $\mu_j^\star > \mu_{j+1}^\star$ from which using (109) we have that $\sigma_{k_j^\star}^\star - \sigma_{k_j^\star+1}^\star > 0$. Accordingly, from (108) it follows that $\sigma_n^\star - \sigma_{n+1}^\star \geq 0$ as required by (81).

We proceed showing that

$$\sum_{i=1}^n x_i^\star - \rho_n \leq 0 \quad n = 1, \ldots, N. \tag{119}$$

For this purpose, observe that for $k_{j-1}^\star < n \leq k_j^\star$ we have that

$$\sum_{i=1}^n x_i^\star = \rho_{k_{j-1}^\star} + \sum_{i=k_{j-1}^\star+1}^n \xi_i(\sigma_i^\star) =$$

$$= \rho_{k_{j-1}^\star} + \sum_{i=k_{j-1}^\star+1}^n \xi_i(\mu_j^\star) = \rho_{k_{j-1}^\star} + c_n(\mu_j^\star; j)$$

from which using (110) it easily follows that the inequality in (119) is always satisfied.

We are now left with proving that

$$(\sigma_n^\star - \sigma_{n+1}^\star)\Big(\sum_{i=1}^n x_i^\star - \rho_n\Big) = 0 \quad \text{for } k_{j-1}^\star < n \leq k_j^\star. \tag{120}$$

For this purpose, we start observing that (120) is trivially satisfied for $k_{j-1}^\star < n < k_j^\star$ due to (108). On the other hand, setting $n = k_j^\star$ into (120) yields

$$(\sigma_{k_j}^\star - \sigma_{k_j+1}^\star)\Big(\sum_{i=1}^{k_j^\star} x_i^\star - \rho_{k_j^\star}\Big) = 0 \tag{121}$$

which holds true if and only if $\sum_{i=1}^{k_j^\star} x_i^\star - \rho_{k_j^\star} = 0$ since $\sigma_{k_j}^\star - \sigma_{k_j+1}^\star > 0$. To this end, we observe that

$$\sum_{n=1}^{k_j^\star} x_n^\star = \sum_{\ell=0}^{j-1} \sum_{i=k_\ell^\star+1}^{k_{\ell+1}^\star} x_i^\star = \sum_{\ell=0}^{j-1} \sum_{i=k_\ell^\star+1}^{k_{\ell+1}^\star} \xi_i(\mu_{\ell+1}^\star)$$

$$= \rho_{k_1^\star} + \sum_{\ell=1}^{j-1}\Big(\rho_{k_{\ell+1}^\star} - \rho_{k_\ell^\star}\Big) = \rho_{k_j^\star} \tag{122}$$

which shows that also (121) is satisfied.

Collecting all the above results together, it follows that the quantities $\{\sigma_n^\star\}$ computed by means of Algorithm 1 satisfy the KKT conditions.

APPENDIX D
EXISTENCE AND UNIQUENESS OF $\varsigma_n^\star$

The purpose of this Appendix is to show that the quantities $\varsigma_n^\star$ required by (24) and (25) for the computation of $\mu^\star$ and $k^\star$, respectively, are always well-defined. This amounts to proving that at each iteration either (22) or (23) provide a *unique* $\varsigma_n^\star \geq 0$ for any $n \in \mathcal{N}_j$.

As done in Appendix C, we refer to the equivalent form illustrated in Algorithm 2 and start considering the first iteration for which $j = 1$, $\gamma_n = \rho_n$ and $k_0^\star = 0$. Under the assumption that the problem $(\mathcal{P})$ is feasible (see Proposition 1 in Section II.A), and recalling Remark 4 (see Section III.A), the following two cases are of interest: **a)** $\sum_{i=1}^n l_i < \rho_n < \sum_{i=1}^n u_i$; **b)** $\rho_n \geq \sum_{i=1}^n u_i$. In particular, case **a)** can be easily handled observing that $c_n(\varsigma; 1)$ is strictly decreasing in the interval (ω_n, Ω_n) with

$$\omega_n = \min\{h_i(u_i), i = 1, \ldots, n\} \tag{123}$$

$$\Omega_n = \max\{h_i(l_i), i = 1, \ldots, n\} \tag{124}$$

whereas $c_n(\varsigma;1) = \sum_{i=1}^{n} u_i$ for $\varsigma \in [0, \omega_n]$, and $c_n(\varsigma;1) = \sum_{i=1}^{n} l_i$ for $\varsigma \in [\Omega_n, +\infty)$. Accordingly, if case **a)** holds true, then the solution of $c_n(\varsigma;1) = \gamma_n$ exists and is unique, it belongs to the interval (ω_n, Ω_n), and coincides with the quantity $\varsigma_{n,1}^\star$ as computed through (101). On the other hand, if case **b)** holds true, then $\varsigma_{n,1}^\star = 0$ as given by (102). In both cases, Algorithm 2 produces a single value of $\varsigma_{n,1}^\star \geq 0$ for any n.

Consider now the $(j+1)$th step of Algorithm 2. Assume that at the jth step the value of $\varsigma_{n,j}^\star$ is well-defined (in the sense specified above) for any $n > k_{j-1}^\star$. This means that $\varsigma_{n,j}^\star = 0$ if

$$\gamma_n = \rho_n - \rho_{k_{j-1}^\star} \geq \sum_{i=k_{j-1}^\star+1}^{n} u_i. \tag{125}$$

On the other hand, $\varsigma_{n,j}^\star > 0$ is the unique solution of $c_n(\varsigma;j) = \rho_n - \rho_{k_{j-1}^\star}$, when

$$\sum_{i=k_{j-1}^\star+1}^{n} l_i < \rho_n - \rho_{k_{j-1}^\star} < \sum_{i=k_{j-1}^\star+1}^{n} u_i. \tag{126}$$

In the sequel, it is shown that if the above assumptions hold true then the value of $\varsigma_{n,j+1}^\star$ for $n > k_j^\star$ is also well-defined at the $(j+1)$th step. This amounts to saying that

$$\sum_{i=k_j^\star+1}^{n} l_i < \rho_n - \rho_{k_j^\star} \tag{127}$$

for $n > k_j^\star$. By contradiction, assume that there exists an index $\bar{k} > k_j^\star$ such that

$$\sum_{i=k_j^\star+1}^{\bar{k}} l_i \geq \rho_{\bar{k}} - \rho_{k_j^\star}. \tag{128}$$

This would mean that $\forall \varsigma > 0$

$$c_{\bar{k}}(\varsigma;j+1) \geq \sum_{i=k_j^\star+1}^{\bar{k}} l_i \geq \rho_{\bar{k}} - \rho_{k_j^\star} \tag{129}$$

from which, recalling that $c_{\bar{k}}(\varsigma;j+1) = c_{\bar{k}}(\varsigma;j) - c_{k_j^\star}(\varsigma;j)$ and setting $\varsigma = \mu_j^\star$, one would get

$$c_{\bar{k}}(\mu_j^\star;j) \geq c_{k_j^\star}(\mu_j^\star;j) + \rho_{\bar{k}} - \rho_{k_j^\star} = \rho_{\bar{k}} - \rho_{k_{j-1}^\star}. \tag{130}$$

This would contradict the fact that $c_n(\mu_j^\star;j) < \rho_n - \rho_{k_{j-1}^\star}$ for any $n > k_j^\star$, as already shown in (110). Accordingly, we must conclude that it cannot exist an index $\bar{k} > k_j^\star$ for which (128) is satisfied, and hence that (127) holds $\forall n > k_j^\star$. In turn, this amounts to saying that if the values of $\varsigma_{n,j}^\star$ for $n > k_{j-1}^\star$ are well-defined at the jth step, allowing the computation of $\mu_j^\star$ and $k_j^\star$, then the values of $\varsigma_{n,j+1}^\star$ for $n > k_j^\star$, computed at the $(j+1)$th step, are well-defined as well, allowing the computation of $\mu_{j+1}^\star$ and $k_{j+1}^\star$. Since the values of $\varsigma_{n,j}^\star$ are well-defined at the first iteration, then they are always well-defined. This concludes the proof.

References

[1] D. Palomar, M.-A. Lagunas, and J. Cioffi, "Optimum linear joint transmit-receive processing for MIMO channels with QoS constraints," *IEEE Trans. Signal Process.*, vol. 52, no. 5, pp. 1179 – 1197, May 2004.

[2] D. Palomar, M. Bengtsson, and B. Ottersten, "Minimum ber linear transceivers for MIMO channels via primal decomposition," *IEEE Trans. Signal Process.*, vol. 53, no. 8, pp. 2866 – 2882, Aug 2005.

[3] D. Palomar, "Convex primal decomposition for multicarrier linear MIMO transceivers," *IEEE Trans. Signal Process.*, vol. 53, no. 12, pp. 4661 – 4674, Dec 2005.

[4] Y. Jiang, W. W. Hager, and J. Li, "Tunable channel decomposition for MIMO communications using channel state information," *IEEE Trans. Signal Process.*, vol. 54, no. 11, pp. 4405 – 4418, Nov 2006.

[5] S. Bergman, D. Palomar, and B. Ottersten, "Joint bit allocation and precoding for MIMO systems with decision feedback detection," *IEEE Trans. Signal Process.*, vol. 57, no. 11, pp. 4509 – 4521, Nov 2009.

[6] Y. Fu, L. Yang, W.-P. Zhu, and C. Liu, "Optimum linear design of two-hop MIMO relay networks with QoS requirements," *IEEE Trans. Signal Process.*, vol. 59, no. 5, pp. 2257 – 2269, May 2011.

[7] S. Ren and M. van der Schaar, "Distributed power allocation in multi-user multi-channel cellular relay networks," *IEEE Trans. Wireless Commun.*, vol. 9, no. 6, pp. 1952 – 1964, June 2010.

[8] L. Sanguinetti and A. D'Amico, "Power allocation in two-hop amplify-and-forward MIMO relay systems with QoS requirements," *IEEE Trans. Signal Process.*, vol. 60, no. 5, pp. 2494 – 2507, May 2012.

[9] A. Liu, V. Lau, and Y. Liu, "Duality and optimization for generalized multi-hop MIMO amplify-and-forward relay networks with linear constraints," *IEEE Trans. Signal Process.*, vol. 61, no. 9, pp. 2356 – 2365, May 2013.

[10] L. Sanguinetti, A. D'Amico, and Y. Rong, "On the design of amplify-and-forward MIMO-OFDM relay systems with QoS requirements specified as Schur-convex functions of the MSEs," *IEEE Trans. Veh. Technol.*, vol. 62, no. 4, pp. 1871 – 1877, May 2013.

[11] D. Palomar and J. Fonollosa, "Practical algorithms for a family of waterfilling solutions," *IEEE Trans. Signal Process.*, vol. 53, no. 2, pp. 686 – 695, Feb. 2005.

[12] F. Gao, T. Cui, and A. Nallanathan, "Optimal training design for channel estimation in decode-and-forward relay networks with individual and total power constraints," *IEEE Trans. Signal Process.*, vol. 56, no. 12, pp. 5937 – 5949, Dec 2008.

[13] G. Farhadi and N. Beaulieu, "Power-optimized amplify-and-forward multi-hop relaying systems," *IEEE Trans. Wireless Commun.*, vol. 8, no. 9, pp. 4634 – 4643, 2009.

[14] A. Padakandla and R. Sundaresan, "Power minimization for CDMA under colored noise," *IEEE Trans. Commun.*, vol. 57, no. 10, pp. 3103 – 3112, Oct. 2009.

[15] T. Pham, H. Nguyen, and H. Tuan, "Power allocation in MMSE relaying over frequency-selective rayleigh fading channels," *IEEE Trans. Commun.*, vol. 58, no. 11, pp. 3330 – 3343, Nov. 2010.

[16] O. Ozel, K. Tutuncuoglu, J. Yang, S. Ulukus, and A. Yener, "Transmission with energy harvesting nodes in fading wireless channels: Optimal policies," *IEEE J. Sel. Areas Commun.*, vol. 29, no. 8, pp. 1732 – 1743, September 2011.

[17] M. Gregori and M. Payaro, "On the precoder design of a wireless energy harvesting node in linear vector gaussian channels with arbitrary input distribution," *IEEE Trans. Commun.*, vol. 61, no. 5, pp. 1868 – 1879, May 2013.

[18] A. Padakandla and R. Sundaresan, "Separable convex optimization problems with linear ascending constraints," *SIAM J. Opt.*, vol. 20, no. 3, pp. 1185 – 1204, July 2007.

[19] Z. Wang, "On solving convex optimization problems with linear ascending constraints," *CoRR*, vol. abs/1212.4701v3, 2014.

[20] P. Akhil, R. Singh, and R. Sundaresan, "A polymatroid approach to separable convex optimization with linear ascending constraints," in *Communications (NCC), 2014 Twentieth National Conference on*, Feb 2014, pp. 1–5.

[21] S. Boyd and L. Vandenberghe, *Convex Optimization*. New York, NY, USA: Cambridge University Press, 2004.

[22] D. Palomar, J. Cioffi, and M.-A. Lagunas, "Joint Tx-Rx beamforming design for multicarrier MIMO channels: a unified framework for convex optimization," *IEEE Trans. Signal Process.*, vol. 51, no. 9, pp. 2381 – 2401, Sept 2003.

Antonio A. D'Amico received the Dr. Ing. Degree in Electronic Engineering in 1992 and the Ph. D. degree in 1997, both from the University of Pisa, Italy. He is currently a Research Fellow at the Department of Information Engineering of the University of Pisa. His research interests are in digital communication theory, with emphasis on synchronization algorithms, channel estimation and detection techniques.

Luca Sanguinetti (S'04, M'06) received the Laurea Telecommunications Engineer degree (cum laude) and the Ph.D. degree in information engineering from the University of Pisa, Italy, in 2002 and 2005, respectively. Since 2005 he has been with the Dipartimento di Ingegneria dell'Informazione of the University of Pisa. In 2004, he was a visiting Ph.D. student at the German Aerospace Center (DLR), Oberpfaffenhofen, Germany. During the period June 2007 - June 2008, he was a postdoctoral associate in the Dept. Electrical Engineering at Princeton. During the period June 2010 - Sept. 2010, he was selected for a research assistantship at the Technische Universitat Munchen. From July 2013 to July 2015 he was with the Alcatel-Lucent Chair on Flexible Radio, Supélec, Gif-sur-Yvette, France. He is an Assistant Professor at the Dipartimento di Ingegneria dell'Informazione of the University of Pisa. He is currently serving an Associate Editor for the IEEE Transactions on Wireless Communications. His expertise and general interests span the areas of communications and signal processing, estimation and game theory. He was the co-recipient of the Best Paper Award at WCNC13 and WCNC14. He is also the recipient of the FP7 Marie Curie IEF 2013 "Dense deployments for green cellular networks".

Daniel P. Palomar (S'99-M'03-SM'08-F'12) received the Electrical Engineering and Ph.D. degrees from the Technical University of Catalonia (UPC), Barcelona, Spain, in 1998 and 2003, respectively. He is an Associate Professor in the Department of Electronic and Computer Engineering at the Hong Kong University of Science and Technology (HKUST), Hong Kong, which he joined in 2006. He is also consulting for the research center CTTC-HK. He had previously held several research appointments, namely, at King's College London (KCL), London, UK; Technical University of Catalonia (UPC), Barcelona; Stanford University, Stanford, CA; Telecommunications Technological Center of Catalonia (CTTC), Barcelona; Royal Institute of Technology (KTH), Stockholm, Sweden; University of Rome "La Sapienza", Rome, Italy; and Princeton University, Princeton, NJ. His current research interests include applications of convex optimization theory, game theory, and variational inequality theory to financial systems and communication systems.

Dr. Palomar is an IEEE Fellow, a recipient of a 2004/06 Fulbright Research Fellowship, the 2004 Young Author Best Paper Award by the IEEE Signal Processing Society, the 2002/03 best Ph.D. prize in Information Technologies and Communications by the Technical University of Catalonia (UPC), the 2002/03 Rosina Ribalta first prize for the Best Doctoral Thesis in Information Technologies and Communications by the Epson Foundation, and the 2004 prize for the best Doctoral Thesis in Advanced Mobile Communications by the Vodafone Foundation and COIT.

He serves as an Associate Editor of IEEE Transactions on Information Theory, and has been an Associate Editor of IEEE Transactions on Signal Processing, a Guest Editor of the IEEE Signal Processing Magazine 2010 Special Issue on "Convex Optimization for Signal Processing," the IEEE Journal on Selected Areas in Communications 2008 Special Issue on "Game Theory in Communication Systems," and the IEEE Journal on Selected Areas in Communications 2007 Special Issue on "Optimization of MIMO Transceivers for Realistic Communication Networks." He serves on the IEEE Signal Processing Society Technical Committee on Signal Processing for Communications (SPCOM).