Duality Gap Estimation and Polynomial Time Approximation for Optimal Spectrum Management

Zhi-Quan Luo*,† and Shuzhong Zhang‡

Abstract

Consider a communication system whereby multiple users share a common frequency band and must choose their transmit power spectra jointly in response to physical channel conditions including the effects of interference. The goal of the users is to maximize a system-wide utility function (e.g., weighted sum-rate of all users), subject to individual power constraints. A popular approach to solve the discretized version of this nonconvex problem is by Lagrangian dual relaxation. Unfortunately the discretized spectrum management problem is NP-hard and its Lagrangian dual is in general not equivalent to the primal formulation due to a positive duality gap. In this paper, we use a convexity result of Lyapunov to estimate the size of duality gap for the discretized spectrum management problem and show that the duality gap vanishes asymptotically at the rate $O(1/\sqrt{N})$, where N is the size of the uniform discretization of the shared spectrum. If the channels are frequency flat, the duality gap estimate improves to $O(1/N)$. Moreover, when restricted to the FDMA spectrum sharing strategies, we show that the Lagrangian dual relaxation, combined with a linear programming scheme, can generate an ϵ-optimal solution for the continuous formulation of the spectrum management problem in polynomial time for any $\epsilon > 0$.

Keywords: Spectrum management, cognitive radio, sum-rate maximization, complexity, duality, ϵ-approximation

I. Introduction

Crosstalk interference is a major obstacle to high speed reliable communication involving multiple users. A standard approach to eliminate multiuser interference is 'orthogonal channelization' whereby

Original manuscript submitted to *IEEE Trans. on Signal Processing*, August 22, 2007.

* Corresponding author, fax: 612-625-4583, phone: 612-625-0242, e-mail: luozq@ece.umn.edu.

† Department of Electrical and Computer Engineering, University of Minnesota, Minneapolis, MN Supported in part by the U.S. NSF, grant number DMS-0610037 and in part by the USDOD ARMY, grant number W911NF-05-1-0567.

‡ Department of Systems Engineering and Engineering Management, The Chinese University of Hong Kong, Shatin, Hong Kong. Email: zhang@se.cuhk.edu.hk. Research supported by Hong Kong RGC Earmarked Grants CUHK418505 and CUHK418406.

users share time, frequency or codes on a non-overlapping basis. Although this approach is easy to implement, it can lead to high system overhead and low system utilization. An important alternative to orthogonal channelization is to allow users' transmit power spectra to overlap over the shared spectrum, in a way that best exploits physical channel conditions. Such dynamic spectrum sharing strategy is flexible and can potentially achieve a substantially higher overall throughput. However, when users' transmit power spectra are allowed to overlap, crosstalk will appear, causing multiuser interference in the system. As a result, each user's performance will depend on not only the power spectral density of his own, but also those of other users in the system. Clearly, proper spectrum (and time) management is required for the maximization of the overall system performance. Dynamic spectrum management problem of this type arises in both wireless systems (e.g., cognitive radios [12]) and wireline communication (e.g., digital subscriber line (DSL) system [20]). In both cases, crosstalk is known to be the major source of signal distortion, and judicious management of spectrum among competing users can have a major impact on the overall system performance.

Existing work

The dynamic spectrum management (DSM) problem has attracted significant research interest in recent years, although a solution method that is both efficient and effective has yet to emerge. Theoretically, the DSM problem can be cast in the *continuous* frequency domain (i.e., without discretizing the frequency) by using power spectral density functions. However, in practice, the DSM problem is often considered in the *discretized* frequency domain (i.e., OFDMA scheme) using finite dimensional power allocation vectors. From the optimization perspective, the DSM problem can be formulated either as a Nash game [8], [10], [24] or as a nonconcave utility maximization problem [7], [25]. Several algorithms were proposed to compute a Nash equilibrium solution (Iterative Waterfilling method (IWFA) [8], [24]) or globally optimal power allocations (dual decomposition method [6], [14], [23]). However, IWFA is known to perform poorly when the interference is strong, while the dual decomposition algorithm is only known to deliver a dual optimal solution [6], [11], [14], [23] rather than the actual optimal transmit power spectra (i.e., primal optimal solution). Moreover, the existing analysis of these algorithms is quite unsatisfactory: the convergence of IWFA to a Nash equilibrium solution is established only when channels satisfy certain restrictive diagonal dominance conditions [15], [22], [24], while the dual decomposition algorithm can fail to converge to a feasible spectrum sharing solution. Is there an efficient (e.g., polynomial time) algorithm for solving the DSM problem, or the problem is intrinsically hard?

In a recent work [16], the authors have established the NP-hardness of the discretized DSM problem under various practical settings as well as for different choices of system utility functions, and have identified several subclasses of the problem which are polynomial time solvable. Among other things, it was shown [16] that the general sum-rate maximization problem is NP-hard, even for the case of two FDMA users. This apparent intractability can be attributed to the nonconvex nature of the discretized DSM problem, which usually gives rise to a positive duality gap between the primal and the dual formulations. In contrast, the *continuous* formulation of the DSM problem actually possesses a surprising hidden convexity. In particular, by using the Lyapunov convexity theorem [17], it was shown [16] that there is no duality gap for the *continuous* version of the DSM problem[1]. Furthermore, under a mild continuity assumption on channel parameters, the duality gap for the discretized DSM problem vanishes asymptotically as the size of the discretization becomes small, although no explicit estimates were given about the vanishing rate. Similar asymptotic strong duality results had previously been derived in mathematical economics for variational equilibrium problems defined by integrals, again relying on the Lyapunov convexity theorem; see [2], [13] for details.

The asymptotic strong duality result suggests that, while computing the exact optimal transmit power spectra for the discretized problem is difficult, it may still be possible to find a high quality approximate solution efficiently for the continuous DSM formulation using a Lagrangian dual relaxation approach. Specifically, we first solve the discretized dual formulation (which is convex and efficiently solvable when restricted to the FDMA strategies [25]), and then use its optimal solution to generate an approximately optimal solution for the continuous version of the DSM problem. To ensure and demonstrate the effectiveness of this Lagrangian dual relaxation approach, we must address two issues. First, given a dual optimal solution for the discretized DSM problem, we need an efficient method to derive a (primal) feasible and approximately optimal solution for the continuous DSM problem. This task is nontrivial because of the positive duality gap that exists in the discretized formulations. Second, we need to estimate the size of the duality gap for any finitely discretized DSM problem. Without such an estimate, we cannot determine the quality of the dual optimal solution, nor establish the near-optimality of the derived primal FDMA solution. In this paper, we positively resolve both of these issues.

To date, there has been little work in the optimization literature that explicitly estimates the size

[1]This strong duality property was first discovered in the DSM context by [23] using an intuitive *frequency-sharing* argument. But this argument was flawed since one cannot determine if their (limiting) frequency-sharing solution is feasible or not. Moreover, their argument requires an infinitesimal division of the shared spectrum over which one can assume constant noise spectral density and crosstalk coefficients and can apply a frequency-sharing solution result. A careful error analysis is needed to examine the effect of this approximation on feasibility and on the final achievable rate.

of duality gap for a nonconvex optimization problem. The notable exception was the work by Aubin and Ekeland [1] which provided explicit duality gap estimates when the objective function and the constraint functions have a separable structure (i.e., representable as a sum of N component functions each depending on a separate variable). For the problem of maximizing the achievable sum-rates in the system, Aubin and Ekeland's result implies that the corresponding duality gap is of order $O(1/N)$. However, for general system utility functions (e.g., maximizing the harmonic mean of users' rates), the duality gap estimate cannot be inferred from Aubin and Ekeland's result.

Summary of contributions

In this paper, we build on the previous work [1], [16] by focussing on two theoretical aspects of the dynamic spectrum management problem: duality gap estimation and polynomial time approximation algorithms. Our contributions are two fold.

1) Under a stronger condition of Lipschitz continuity, we strengthen the asymptotic strong duality result of [16] and the duality gap estimates of Aubin and Ekeland [1] by providing an explicit estimate of the duality gap for the discretized DSM problem with general concave system utility functions. In particular, we provide a unified analysis that shows that the duality gap vanishes asymptotically at the rate $O(1/\sqrt{N})$ under a Lipschitz continuity assumption on the channel parameters, where N is the size of the uniform discretization of the shared spectrum. Notice that the zero duality for the continuous formulation requires the use of Lebesgue integrals which cannot be well approximated by a Riemann partial sum. This creates technical difficulties in our analysis since the discretized DSM problem is obtained through the Riemann sum approximation of the corresponding Lebesgue integrals in the continuous formulation. Fortunately, even though the integrals can not be well approximated, the two optimization problems are still asymptotically equivalent, and the duality gap for the discretized problem approaches zero at the rate of $O(1/\sqrt{N})$. This rate improves to $O(1/N)$ when channels are frequency flat.
2) Our second contribution is a polynomial time approximation scheme to determine the optimal Frequency Division Multiple Access (FDMA) spectrum sharing strategy. The reason for focussing on FDMA solutions is two fold. First, the recent work [11] has shown that FDMA strategies are sum-rate optimal for scenarios with strong interference. Thus, there is no loss of generality by considering only FDMA solutions in this case. Secondly, when restricted to the class of FDMA strategies, the Lagrangian dual relaxation can be implemented in polynomial time [14]. Without imposing the FDMA structure, the Lagrangian dual is difficult to compute or optimize. Notice that, due to a

positive duality gap, the optimal dual solution does not lead to a feasible FDMA solution with equal utility value for the discretized DSM problem. In this paper, we devise a linear programming procedure which, when coupled with the Lagrangian dual relaxation scheme, can generate a near-optimal FDMA solution for the continuous formulation of the DSM problem. By the duality gap estimate $O(1/\sqrt{N})$, we show that this combined procedure constitutes a fully polynomial time approximation scheme for the continuous version of the DSM problem.

Discussions

According to the complexity analysis of [16], finding a good approximately optimal solution for the *discretized* DSM problem is NP-hard, even when restricted to the FDMA strategies. Therefore, it may seem as a surprise that finding an approximately optimal solution for the *continuous* DSM problem can actually be done in polynomial time (at least for the FDMA case), see the second claim above. However, there is no contradiction here, because the continuous formulation has a hidden convexity, whereas the discretized formulation does not. The continuous formulation allows a much richer set of solutions than the discretized formulation. The latter artificially restricts the power spectra to be piecewise constant over a fixed number of uniformly partitioned tones, which makes primal feasibility harder to achieve.

II. Primal and Dual Formulations

Consider a multi-user communication system consisting of K transmitter-receiver pairs sharing a common frequency band $f \in \Omega$. For simplicity, we will call each of such transmitter-receiver pair a "*user*". Upon normalization, we can assume Ω to be the unit interval in $\Re$, namely, $\Omega = [0, 1]$. Each user k has a fixed transmit power budget which he can allocate across Ω so as to maximize his own utility. Let $s_k(f) : \Omega \mapsto [0, \infty)$ denote the power spectral density (or power allocation) function of user k. The transmit power budget of user k can be represented as

$$\int_\Omega s_k(f)df \le P_k,$$

where $P_k > 0$ is a given constant. Due to multi-user interference, user k's utility depends on not only his own allocation function $s_k(f)$, but also those of others $\{s_\ell(f) : \ell \ne k\}$. Let user k's utility function be denoted by

$$u_k(s_1, s_2, ..., s_K) = \int_\Omega R_k(s_1(f), \dots, s_K(f), f)df, \tag{1}$$

where $R_k(\cdot) : \Re_+^K \times \Omega \mapsto [0, +\infty)$ is a Lebesgue integrable, possibly non-concave function.

Due to the complex coupling between users' utility functions, it is generally impossible to maximize u_1, u_2,..., u_K simultaneously. Instead, we seek to maximize a system-wide utility $H(u_1, ..., u_K)$ which balances the interests of all users in the system appropriately. This leads to the following spectrum management problem:

$$\begin{array}{ll} \max & H(u_1, \cdots, u_K) \\ \text{s.t.} & u_1 = \int_\Omega R_1(s_1(f), \ldots, s_K(f), f) df \\ & \quad \vdots \\ & u_K = \int_\Omega R_K(s_1(f), \ldots, s_K(f), f) df \\ & \int_\Omega s_k(f) df \le P_k,\, 0 \le s_k(f) \le S_{\max} \text{ and Lebesgue measurable, } k = 1, ..., K, \end{array} \qquad (P_c)$$

where $S_{\max}$ represents the maximum allowable magnitude of transmit power spectra. The subscript c in the notation "(P_c)" signifies the continuous domain of the formulation. The maximum value of (P_c) is called the *social optimum*.

There are four commonly used choices for the system utility function $H(u_1, \cdots, u_K)$:

i) Sum-rate utility: $H_1(u_1, ..., u_K) = \frac{1}{K} \sum_{k=1}^K u_k$;

ii) Proportional fairness utility: $H_2(u_1, ..., u_K) = \left(\prod_{k=1}^K u_k \right)^{1/K}$ (equivalent to maximizing $\sum_{k=1}^K \log u_k$);

iii) Harmonic-rate utility: $H_3(u_1, ..., u_K) = K \left(\sum_{k=1}^K u_k^{-1} \right)^{-1}$ (equivalent to maximizing $-\log(\sum_{k=1}^K u_k^{-1})$);

iv) Min-rate utility: $H_4(u_1, ..., u_K) = \min_{1 \le k \le K} u_k$.

In general, these utility functions can be ordered $H_1 \ge H_2 \ge H_3 \ge H_4,\ \forall\ (u_1, u_2, ..., u_K) \in \Re_+^K$. In terms user fairness, the order is reversed.

The DSM problem (P_c) is in general nonconvex. A powerful approach to analyze a nonconvex optimization is through its Lagrangian dual formulation. For the DSM problem (P_c), the Lagrangian function is given by

$$\begin{aligned} L(s, t; \lambda, \mu) &= H(t_1, \cdots, t_K) + \sum_{k=1}^K \lambda_k \left[\int_{f \in \Omega} R_k(s_1(f), s_2(f), ..., s_K(f), f) df - t_k \right] \\ &\quad + \sum_{k=1}^K \mu_k \left[P_k - \int_{f \in \Omega} s_k(f) df \right] \\ &= H(t_1, \cdots, t_K) - \sum_{k=1}^K t_k \lambda_k + \sum_{k=1}^K P_k \mu_k \\ &\quad + \int_{f \in \Omega} \sum_{k=1}^K \left[\lambda_k R_k(s_1(f), s_2(f), ..., s_K(f), f) - \mu_k s_k(f) \right] df. \end{aligned} \qquad (2)$$

Let

$$\begin{aligned} g(\lambda,\mu) \quad := \quad & \sup \quad L(s,t;\lambda,\mu) \\ & \text{s.t.} \quad t\in\Re^K \\ & \qquad 0\le s_k(f)\le S_{\max},\ f\in\Omega,\ s_k(\cdot) \text{ is Lebesgue measurable; } k=1,...,K. \end{aligned}$$

Since t is separated from s, we may simplify the expression for $g(\lambda,\mu)$ by using the conjugate dual function of H

$$H^*(\lambda) := \sup_{t\in\Re^K}\left(H(t)-\lambda^{\mathrm{T}}t\right) \tag{3}$$

which is convex. We have $g(\lambda,\mu)=H^*(\lambda)+P^{\mathrm{T}}\mu+\bar{g}(\lambda,\mu)$, where $P=(P_1,P_2,...,P_K)^{\mathrm{T}}$ and

$$\begin{aligned} \bar{g}(\lambda,\mu) \quad := \quad & \max \quad \int_{f\in\Omega}\sum_{k=1}^{K}\left[\lambda_k R_k(s_1(f),s_2(f),...,s_K(f),f)-\mu_k s_k(f)\right]df \\ & \text{s.t.} \quad 0\le s_k(f)\le S_{\max},\ f\in\Omega,\ s_k(\cdot) \text{ is Lebesgue measurable; } k=1,...,K. \end{aligned}$$

Clearly, $g(\lambda,\mu)$ is convex jointly in λ and μ. The Lagrangian dual problem of (P_c) is defined as

$$\boxed{\begin{array}{ll} \text{minimize} & g(\lambda,\mu) \\ \text{subject to} & \lambda,\mu\in\Re_+^K. \end{array} \qquad (D_c)}$$

The conjugate function H^* can be computed explicitly for various choices of system utility functions. It is relatively easy to verify the following.

1) For weighted sum-rate function $H(u)=w^{\mathrm{T}}u$, the conjugate function $H^*(\lambda)=0$ for $\lambda=w$ and $H^*(\lambda)=\infty$ when $\lambda\neq a$.
2) For the proportional fairness system utility function $H(u)=\sum_{k=1}^K \log u_k$, the conjugate function $H^*(\lambda)=K-\sum_{k=1}^K\log\lambda_k$.
3) For the min-rate utility function $H(u)=\min_{1\le k\le K}u_k$, the conjugate function $H^*(\lambda)=0$ if $\sum_{k=1}^K\lambda_k\ge 1$, and $H^*(\lambda)=+\infty$ otherwise.
4) For the harmonic-rate utility function $H(u)=(\sum_{k=1}^K u_k^{-1})^{-1}$, the conjugate function $H^*(\lambda)=0$ if $\sum_{k=1}^K\sqrt{\lambda_k}\ge 1$, and $H^*(\lambda)=+\infty$ otherwise.

The DSM problem (P_c) is in general nonconvex due to the nonconcavity of utility functions $u_1,...,u_K$. However, the dual problem (D_c) is always convex. Unfortunately, both problems are defined in continuous domain (infinite dimensional), with functions $s_1(f)$, $s_2(f)$,..., $s_K(f)$ as decision variables. As such, the spectrum management problem (P_c) is a difficult infinite dimensional nonlinear optimization problem.

To facilitate practical implementation (e.g., the OFDM scheme) and numerical solution, we typically discretize the frequency set so that $\Omega=\{0,\frac{1}{N},\frac{2}{N},...,1\}$. In this way, the continuous formulation of the resource management problem (P_c) can be discretized by replacing Lebesgue measure with a discrete uniform measure on $[0, 1]$. In particular, user k's resource allocation becomes

$$s_k^n \geq 0, \quad \frac{1}{N}\sum_{n=1}^{N} s_k^n \leq P_k$$

and the corresponding utility is

$$u_k = \frac{1}{N}\sum_{n=1}^{N} R_k(s_1^n, \ldots, s_K^n, n/N).$$

The corresponding social optimum is achieved by maximizing the total system utility $H(u_1, ..., u_K)$

$$\boxed{\begin{array}{ll} \max & H(u_1, \cdots, u_K) \\ \text{s.t.} & u_1 = \frac{1}{N}\sum_{n=1}^{N} R_1(s_1^n, \ldots, s_K^n, n/N) \\ & \quad \vdots \\ & u_K = \frac{1}{N}\sum_{n=1}^{N} R_K(s_1^n, \ldots, s_K^n, n/N) \\ & \frac{1}{N}\sum_{n=1}^{N} s_k^n \leq P_k,\ 0 \leq s_k^n \leq S_{\max},\ k = 1, ..., K. \end{array} \qquad (P_d^N)}$$

We use (P_d^N) to denote this discretized problem. Intuitively, $(P_d^N) \to (P_c)$ as $N \to \infty$, that is, as the discretization becomes infinitely fine, the discretized problem (P_d^N) should coincide with the continuous DSM problem (P_c). However, we caution that this intuition can be misleading due to the fact that, in general, a Lebesgue integral cannot be well approximated by a Riemann partial sum. Since (P_d^N) is obtained through the Riemann sum approximation of the corresponding Lebesgue integrals in the continuous formulation (P_c), it is unclear if the two optimization problems are asymptotically equivalent. A main contribution of this paper is to settle this issue in the affirmative; see Section III.

We can also develop a Lagrangian dual for the discrete DSM problem (P_d^N). Specifically, let

$$\begin{aligned} L_N(s, u; \lambda, \mu) &= H(u_1, \cdots, u_K) + \sum_{k=1}^{K} \lambda_k \left[\frac{1}{N}\sum_{n=1}^{N} R_k(s_1^n, \ldots, s_K^n, n/N) - u_k\right] \\ &\quad + \sum_{k=1}^{K} \mu_k \left[P_k - \frac{1}{N}\sum_{n=1}^{N} s_k^n\right] \\ &= H(u_1, \cdots, u_K) - \sum_{k=1}^{K} u_k\lambda_k + \sum_{k=1}^{K} P_k\mu_k + \frac{1}{N}\sum_{n=1}^{N}\sum_{k=1}^{K} \left[\lambda_k R_k(s_1^n, \ldots, s_K^n, n/N) - \mu_k s_k^n\right]. \end{aligned}$$

The dual function is then given by

$$\begin{aligned} g_N(\lambda, \mu) := \ & \sup \quad L_N(s, u; \lambda, \mu) \\ & \text{s.t.} \quad u \in \Re^K,\ 0 \leq s_k^n \leq S_{\max},\ k = 1, ..., K. \end{aligned}$$

Hence the Lagrangian dual of (P_d^N) can be written as

$$\boxed{\begin{array}{ll} \min & H^*(\lambda) + P^{\mathrm{T}}\mu + \bar{g}_N(\lambda,\mu) \\ \text{s.t.} & \lambda, \mu \in \Re^K_+, \end{array} \qquad (D_d^N)}$$

where H^* is defined by (3) and

$$\begin{aligned} \bar{g}_N(\lambda,\mu) \quad := \quad & \max \quad \frac{1}{N}\sum_{n=1}^{N}\sum_{k=1}^{K}[\lambda_k R_k(s_1^n,\ldots,s_K^n) - \mu_k s_k^n] \\ & \text{s.t.} \qquad 0 \le s_k^n \le S_{\max},\ n = 1,...,N;\ \ k = 1,...,K. \end{aligned}$$

Example: rate maximization

We can specify the choice of Lebesgue integrable functions $R_k(s_1(f),...,s_K(f))$ for the case where users do not have direct knowledge of each other's code/modulation schemes. In this situation, it is reasonable for each user to treat the signals from other users as Gaussian noise. As a result, $R_k(\cdot)$ can represent the data rate achievable by user k at frequency f (in the sense of Shannon [9]):

$$R_k(s_1(f),\ldots,s_K(f),f) = \log\left(1 + \frac{s_k(f)}{\sigma_k(f) + \sum_{j\neq k}\alpha_{kj}(f)s_j(f)}\right), \tag{4}$$

where $\sigma_k(f) > 0$ signifies the noise power at user k on frequency f, and $\alpha_{k\ell}(f) > 0$ denotes the normalized path loss coefficient for the channel between user ℓ and user k on frequency f. For a practical system (e.g., IEEE 802.11x standards), the available spectrum Ω is usually divided into multiple tones (or bands) and shared among the users. In this way and assuming $H(\cdot) = H_1(\cdot)$, the DSM problem (P_c) is discretized and becomes

$$\boxed{\begin{array}{ll} \text{maximize} & \displaystyle\frac{1}{NK}\sum_{k=1}^{K}\sum_{n=1}^{N}\log\left(1 + \frac{s_k^n}{\sigma_k^n + \sum_{j\neq k}\alpha_{kj}^n s_j^n}\right) \\ \text{subject to} & \displaystyle\frac{1}{N}\sum_{n=1}^{N}s_k^n \le P_k, \quad 0 \le s_k^n \le S_{\max}, \quad n = 1,2,...,N,\ k = 1,2,...,K. \end{array}}$$

III. Duality Analysis

Consider the discrete rate-maximization problem (P_d^N) in the previous section. Even though (P_d^N) is nonconvex, its dual (D_d^N) is always convex and easy to solve as long as $\bar{g}(\lambda,\mu)$ and its gradient are easily computable. Let P_N^* and D_N^* denote their respective optimal values. It follows from weak duality that $P_N^* \le D_N^*$. Simple examples exist [16] which show that the duality gap $D_N^* - P_N^*$ is typically positive. In this section we will estimate the size of duality gap $D_N^* - P_N^*$. The duality gap estimates derived in this section will be needed in the performance analysis of Lagrangian dual relaxation algorithms in Section IV.

Recall that the duality gap between the continuous primal-dual formulations (P_c) and (D_c) is zero under minimal assumptions; see [16], [23]. As noted before, there is one difference between the primal-dual formulations (P_c)–(D_c) and those given in [16]: we have added a bound constraint $0 \leq s_k(f) \leq S_{\max}$ in our formulations (P_c)–(D_c). This bound constraint is used to eliminate the possibility of delta functions in the limiting process, thus simplifying the duality gap analysis. Below we first generalize the strong duality result of [16], [23] to the bounded spectra case. The key property that enables this strong duality is a hidden convexity in the continuous problem formulation (P_c) which can be established using the Lyapunov convexity theorem [17].

Theorem 1 (Strong duality [16]) *Suppose that the system utility function $H(u_1, ..., u_K)$ is jointly concave in the variables $(u_1, u_2, ..., u_K)$ and is nondecreasing in each u_k. Assume that the rate functions $R_k(s, f)$ are Lebesgue integrable. Then, the optimal values of (P_c) and (D_c) are equal.*

Proof. The proof is similar to that of [16]. We only need to show the convexity of the set

$$W = \left\{ \left(\begin{array}{c} \int_0^1 R_1(s(f), f) df \\ \vdots \\ \int_0^1 R_K(s(f), f) df \end{array} \right) \middle| \begin{array}{l} s(f) \text{ Lebesgue integrable on } f \in [0,1] \\ \text{and } 0 \leq s_k(f) \leq S_{\max} \end{array} \right\}$$

By Lyapunov theorem [17], the following set

$$\bar{W} = \left\{ \left(\begin{array}{c} \int_0^1 R_1(s^2(f), f) df \\ \vdots \\ \int_0^1 R_K(s^2(f), f) df \\ \int_0^1 [S_{\max}^2 - \max_k s_k^2(f)]_- df \end{array} \right) \middle| \; s(f) \text{ Lebesgue integrable on } f \in [0,1] \right\}$$

is convex, where $s^2(f) \equiv (s_1^2(f), s_2^2(f), ..., s_K^2(f))$ and $[\cdot]_-$ denotes projection to the nonpositive real numbers. This is because $R_k(s^2(f), f)$ remains Lebesgue integrable and the function $[S_{\max}^2 - \max_k s_k^2(f)]_-$ is also Lebesgue integrable. Consider the set $\tilde{W} = \bar{W} \cap (\Re^K \times \{0\})$ which clearly remains convex. Due to the constraint

$$\int_0^1 [S_{\max}^2 - \max_k s_k^2(f)]_- df = 0$$

the set $\tilde{W}$ can be characterized by

$$\tilde{W} = \left\{ \left(\begin{array}{c} \int_0^1 R_1(s^2(f), f) df \\ \vdots \\ \int_0^1 R_K(s^2(f), f) df \\ 0 \end{array} \right) \middle| \begin{array}{l} s(f) \text{ Lebesgue integrable on } f \in [0,1] \\ \text{and } 0 \leq s_k^2(f) \leq S_{\max}^2 \end{array} \right\}.$$

Renaming $s_k^2(f)$ as $s_k(f)$ and projecting $\tilde{W}$ to $\Re^K$ gives W. Thus the set W is convex. This further implies the strong duality between the primal-dual pair (P_c)-(D_c). ∎

In what follows, we will use Theorem 1 to provide an upper bound on the duality gap between (P_d^N) and (D_d^N) under the additional assumption that the system utility function is Lipschitz continuous. The key step in the ensuing analysis is to establish an everywhere-dense property of uniformly spaced piecewise constant functions in the set of Lebesgue integrable functions; see Proposition 1.

Without loss of generality, let us assume $\Omega = [0, 1]$ and consider an uniform discretization of Ω $\{n/N \mid n = 0, 1, ..., N\}$. Corresponding to this discretization, the noise and power spectrum functions $\sigma_k(f)$, $s_k(f)$, as well as the interference coefficient functions $\alpha_{kj}(f)$ become finite dimensional vectors which we denote as follows:

$$\sigma_k(n/N) = \sigma_k^n,\ s_k(n/N) = s_k^n,\ \alpha_{kj}(n/N) = \alpha_{kj}^n,\quad 0 \le n \le N,\ 1 \le k, j \le K. \tag{5}$$

In this way, the primal and dual continuous spectrum management problem (P_c) and (D_c) are discretized and become (P_d^N) and (D_d^N) respectively.

Theorem 2 *Suppose the system utility function $H(u_1, u_2, ..., u_K)$ and the rate functions $R_k(s, f)$ (cf. see (1)) are Lipschitz continuous in the sense*

$$\begin{aligned} &|H(u_1, u_2, ..., u_K) - H(v_1, v_2, ..., v_K)| \le L\|u - v\|_\infty, \quad \forall\ u, v \ge 0, \\ &|R_k(s, f) - R_k(s', f')| \le L\left(|f - f'| + \|s - s'\|_\infty\right), \quad \forall\ f, f' \in [0, 1], \end{aligned} \tag{6}$$

where $L > 0$ is the Lipschitz constant. Let P_N^* and D_N^* denote the objective values of (P_d^N) and (D_d^N) respectively with σ_k^n, α_{kj}^n defined by (5). Then we have

$$0 \le D_N^* - P_N^* \le O\left(\frac{L}{\sqrt{N}}\right). \tag{7}$$

Moreover, if the rate functions $R_k(s, f)$ are independent of f (i.e., frequency flat), then the above duality gap estimate improves to $O(L/N)$.

Since the concavity and monotonicity and Lipschitz continuity assumptions in Theorem 2 are satisfied by the min-rate, harmonic-rate, proportional fairness rate and sum-rate functions, it follows that the duality gap estimate (7) holds for all of these choices of system utility functions.

The proof of Theorem 2 can be established through the following general result regarding Lebesgue integral and its approximation.

Proposition 1 *Suppose* $R(s(f),f) \equiv (R_1(s,f), R_2(s,f), ..., R_K(s,f))$ *satisfies the Lipschitz condition (6). Consider*

$$W = \left\{ \left(\begin{array}{c} \int_0^1 R_1(s(f),f)df \\ \vdots \\ \int_0^1 R_K(s(f),f)df \end{array} \right) \middle| \begin{array}{l} s(f) \textit{ Lebesgue integrable on } f \in [0,1] \\ \textit{and } 0 \le s_k(f) \le S_{\max} \end{array} \right\}$$

and

$$W^N = \left\{ \left(\begin{array}{c} \int_0^1 R_1(s(f),f)df \\ \vdots \\ \int_0^1 R_K(s(f),f)df \end{array} \right) \middle| \begin{array}{l} s(f) \textit{ step-function on } [0,1] \textit{ with breakpoints} \\ \textit{being } \frac{\ell}{N},\ \ell = 1, ..., N-1, \textit{ and } 0 \le s_k(f) \le S_{\max} \end{array} \right\}.$$

Then, for any $w \in W$*, there exists some* $\tilde{w} \in W^N$ *such that*

$$\|\tilde{w} - w\|_\infty \le (L(2+K) + \|R\|K + (K+1)\sqrt{K} + 1)N^{-1/2}, \tag{8}$$

where

$$\|R\|_\infty = \max_{k=1,2,...,K} \sup_{f\in[0,1]} R_k(s(f),f)$$

denotes the ∞*-norm of* $R(s(f),f)$*. Moreover, for the frequency flat case whereby the rate function* $R(s,f)$ *is not dependent on the frequency* f*, i.e.,* $R(s,f) = R(s)$*, the above estimate (8) can be strengthened to*

$$\|\tilde{w} - w\|_\infty \le \|R\|_\infty K N^{-1}.$$

Recall that a Lebesgue integral cannot be approximated simply by its Riemann partial sum. That is, for any given Lebesgue integrable function $s(f)$, we have in general

$$\frac{1}{N}\sum_{n=1}^{N} R(s(n/N), n/N) \not\to \int_0^1 R(s(f),f)df, \quad \text{as } N \to \infty.$$

Interestingly, Proposition 1 asserts that there *exists* some step function $\tilde{s}(n/N)$ (not necessarily equal to $s(n/N)$) such that

$$\frac{1}{N}\sum_{n=1}^{N} R(\tilde{s}(n/N), n/N) \to \int_0^1 R(s(f),f)df, \quad \text{as } N \to \infty.$$

Moreover, the convergence speed is at least $O(1/\sqrt{N})$.

The proof of Proposition 1 consists of an explicit construction of a step function $\tilde{s}(f)$ satisfying the given error estimates. The key is the use of Caratheodary theorem from convex analysis. The details of the proof are left to Appendix A. Theorem 2 follows from Proposition 1 by means of approximating the vector integrations using the step-functions on the uniform subintervals of $[0,1]$. The details of the proof are given in Appendix B.

A. Extension to FDMA formulations

The duality gap estimation can also be made for the FDMA version of the primal dual pair (P_d^N)-(D_d^N). In particular, since the power spectral densities $s_k(f)$ are nonnegative, the FDMA constraint

$$s_k(f)s_j(f) = 0 \;\; \text{for all } f \in \Omega \text{ and all } k \neq j$$

can be represented in the integral form

$$\int_\Omega s_k(f)s_j(f)df = 0, \quad \forall\, k \neq j.$$

Consider the continuous FDMA version of the DSM problem (P_c):

$$\boxed{\begin{array}{ll} \max & H(u_1, \cdots, u_K) \\ \text{s.t.} & u_1 = \int_\Omega R_1(s_1(f), \ldots, s_K(f), f)df \\ & \quad \vdots \\ & u_K = \int_\Omega R_K(s_1(f), \ldots, s_K(f), f)df \\ & \int_\Omega s_k(f)s_j(f)df = 0, \quad \forall\, k \neq j, \\ & \int_\Omega s_k(f)df \leq P_k,\ 0 \leq s_k(f) \leq S_{\max} \text{ and Lebesgue measurable, } k = 1, ..., K, \end{array} \qquad (P_c^{\text{fdma}})}$$

and its dual (D_c^{fdma}). Then Lyapunov theorem can be invoked to establish the strong duality between (P_c^{fdma}) and (D_c^{fdma}). Moreover, if we let $P^*_{N,\text{fdma}}$ and $D^*_{N,\text{fdma}}$ denote the optimal primal and dual objective value of the corresponding discretized FDMA spectrum management problems. Then we can use Proposition 1 to establish a duality gap estimate similar to that in Theorem 2:

$$0 \leq D^*_{N,\text{fdma}} - P^*_{N,\text{fdma}} \leq O(1/\sqrt{N}). \tag{9}$$

We omit the details of the proof (since they are mostly the same as the proof of Theorem 2).

IV. A Polynomial Time Approximation Scheme to Find FDMA Solutions

In this section, we propose a polynomial time approximation scheme to find an ϵ-optimal FDMA solution for the continuous formulation (P_c), for any $\epsilon > 0$. The key ingredient in our analysis is the duality gap estimates developed in the previous section. There are two reasons why we focus on FDMA solutions. First, the general dual optimization problem (D_d^N) is difficult to solve without FDMA structure. In fact, even evaluating the dual objective function is NP-hard in general. By focussing on FDMA solutions, we ensure the polynomial time solvability of the dual (D_d^N) [25]. Second, by the recent work [11], FDMA strategies are optimal for (P_d^N) when the crosstalk interference coefficients are sufficiently large (roughly $\alpha_{kj}^n \geq 0.5$ for all k, j, n). In other words, for scenarios with strong interference, there is no loss in the sum-rate performance by considering only FDMA solutions.

While restricting to FDMA strategies indeed simplifies the solution of the dual problem (D_d^N), it still does not ensure polynomial time solvability of the primal problem DSM (P_d^N). As shown in [16], computing a globally optimal FDMA solution for the discretized DSM problem (P_d^N) is NP-hard; in fact, even approximating the global optimum is intractable in this case. The reason is that, even if a dual optimal solution is available, it is still difficult to construct a primal feasible FDMA solution for (P_d^N) that can achieve the same dual objective value. In fact, this is in general impossible due to positive duality gap. The main contribution of this section is to show that we can use a dual optimal solution for (D_d^N) to generate, in polynomial time, a primal solution that is feasible for the continuous formulation (P_c) while achieving the same objective value. By the duality gap estimates of Theorem 2, the resulting algorithm constitutes a polynomial time approximation scheme for computing an ϵ-optimal solution of (P_c), for any $\epsilon > 0$.

A. *Polynomial time solvability of the dual FDMA DSM problem*

Notice that, under the FDMA constraint, the primal problem (P_d^N) is equivalent to the following:

$$\boxed{\begin{array}{lll} \max & \mathcal{P}_N(s) := H\left(\frac{1}{N}\sum_{n=1}^{N}\log\left(1+\frac{s_1^n}{\sigma_1^n}\right), ..., \frac{1}{N}\sum_{n=1}^{N}\log\left(1+\frac{s_K^n}{\sigma_K^n}\right)\right) & \\ \text{s.t.} & \frac{1}{N}\sum_{n=1}^{N} s_k^n \leq P_k,\ k=1,...,K, & (P_{d,\mathrm{fdma}}^N) \\ & 0 \leq s_k^n \leq S_{\max},\ \ n=1,...,N;\ k=1,...,K, & \\ & s_k^n s_\ell^n = 0,\ \ \forall\ 1 \leq k \neq \ell \leq K;\ n=1,...,N. & \end{array}}$$

with the corresponding dual given by

$$\boxed{\begin{array}{lll} \min & \mathcal{D}_N(\lambda, \mu) \equiv H^*(\lambda) + \mathbf{p}^T\mu + \bar{g}_N^{\mathrm{fdma}}(\lambda, \mu) & (D_{d,\mathrm{fdma}}^N) \\ \text{s.t.} & \lambda, \mu \in \Re_+^K, & \end{array}}$$

where $\bar{g}_N^{\mathrm{fdma}}(\lambda, \mu)$ is given by

$$\begin{array}{rcll} \bar{g}_N^{\mathrm{fdma}}(\lambda, \mu) & := & \max & \frac{1}{N}\sum_{n=1}^{N}\sum_{k=1}^{K}\left[\lambda_k \log\left(1+\frac{s_k^n}{\sigma_k^n}\right) - \mu_k s_k^n\right] \\ & & \text{s.t.} & 0 \leq s_k^n \leq S_{\max},\ n=1,...,N;\ \ k=1,...,K, \\ & & & s_k^n s_k^m = 0,\ \ \forall\ 1 \leq m \neq n \leq N. \end{array}$$

The above problem, though not convex due to the FDMA constraint (each frequency can only be assigned to at most one user), can be analytically solved by the following greedy procedure.

Procedure 1

For $n = 1, ..., N$ and $k = 1, ..., K$, let

$$\hat{s}_k^n \longleftarrow \text{argmax} \left\{ \lambda_k \log\left(1 + \frac{\bar{s}_k^n}{\sigma_k^n}\right) - \mu_k \bar{s}_k^n \,\middle|\, \bar{s}_k^n = \left(\frac{\lambda_k}{\mu_k} - \sigma_k^n\right)_+ \right\},$$

where $(\cdot)_+$ denotes projection to the interval $[0, S_{\max}]$, and let

$$I^n := \text{argmax} \left\{ \lambda_k \log \lambda_k - \lambda_k + \mu_k \sigma_k^n - \lambda_k \log\left(\mu_k \sigma_k^n\right) \mid 1 \le k \le K \right\}.$$

For $n = 1, ..., N$, take any $k_n \in I^n$, and return with the solution

$$s_j^n = \begin{cases} 0, & \text{if } j \neq k_n; \\ \hat{s}_{k_n}^n, & \text{if } j = k_n. \end{cases}$$

The above procedure has a time complexity of $O(NK)$. It finds a subgradient of the dual objective of $(D_{\text{d,fdma}}^N)$ as a byproduct, i.e.,

$$\begin{pmatrix} \mathbf{d} + \dfrac{1}{N} \displaystyle\sum_{\{n:\ k = k_n \in I^n\}}^{N} \log\left(1 + \frac{\hat{s}_k^n}{\sigma_k^n}\right) \mathbf{e}_k \\ \mathbf{p} - \dfrac{1}{N} \displaystyle\sum_{\{n:\ k = k_n \in I^n\}}^{N} \hat{s}_k^n \mathbf{e}_k \end{pmatrix}, \tag{10}$$

where $\mathbf{d} \in \partial H^*(\lambda)$, and $\mathbf{e}_k$ is the kth unit vector in $\Re^K$. The subdifferential of the dual objective function $\mathcal{D}_N(\lambda, \mu)$ can be characterized by

$$\partial \mathcal{D}_N(\lambda,\mu) = \begin{pmatrix} \partial H^*(\lambda) \\ 0 \end{pmatrix} + \text{Conv} \left\{ \begin{pmatrix} \dfrac{1}{N} \displaystyle\sum_{\{n:\ k = k_n\}}^{N} \log\left(1 + \frac{\hat{s}_k^n}{\sigma_k^n}\right) \mathbf{e}_k \\ \mathbf{p} - \dfrac{1}{N} \displaystyle\sum_{\{n:\ k = k_n\}}^{N} \hat{s}_k^n \mathbf{e}_k \end{pmatrix} : k_n \in I^n,\ 1 \le n \le N \right\}. \tag{11}$$

Notice that the convex hull is taken with respect to all the subgradients of the form (10) which are $|I^1| \times |I^2| \times \cdots \times |I^N|$ in total. As the convex hull is taken over these exponential number of vectors, it is inefficient to directly deal with (11). Fortunately, the subdifferential $\partial \mathcal{D}_N(\lambda, \mu)$ can be alternatively described in a compact way. In particular, due to the Cartesian product structure of the index set $I^1 \times I^2 \times \cdots \times I^N$, the convex hull in (11) can be computed as the Cartesian product of convex hulls with respect to individual index set I^n, $n = 1, 2, ..., N$. Thus, we have

$$\partial \mathcal{D}_N(\lambda, \mu) = \begin{pmatrix} \partial H^*(\lambda) \\ 0 \end{pmatrix} + \left\{ \left(\begin{array}{c} \frac{1}{N} \sum_{k=1}^{K} \sum_{\{n:\ I^n \ni k\}} \nu_{nk} \log\left(1 + \frac{\hat{s}_k^n}{\sigma_k^n}\right) \mathbf{e}_k \\ \mathbf{p} - \frac{1}{N} \sum_{n=1}^{N} \left(\sum_{k \in I^n} \nu_{nk} \hat{s}_k^n \mathbf{e}_k \right) \end{array} \right) \middle| \begin{array}{l} \sum_{k \in I^n} \nu_{nk} = 1, \\ \nu_{nk} \geq 0,\ k \in I^n \\ 1 \leq n \leq N \end{array} \right\}. \quad (12)$$

If the system utility function $H(u)$ is the sum-rate of all users, then $\lambda_k \equiv 1$ for all $1 \leq k \leq K$, so that

$$\hat{s}_k^n = \left(\frac{1}{\mu_k} - \sigma_k^n \right)_+,$$

and the index assignment is

$$k_n \quad \longleftarrow \quad \text{argmin}\ \{\mu_k \sigma_k^n - \log(\mu_k \sigma_k^n) - 1 \,|\, 1 \leq k \leq K\}.$$

In this case, the dual objective function $\mathcal{D}_N(\lambda, \mu)$ is constant with respect to λ and the subdifferential $\partial \mathcal{D}_N(\lambda, \mu)$ contains only the subgradient information with respect to the μ. Thus, the subdifferential of $\mathcal{D}_N(\lambda, \mu)$ (projected to the component μ) can be characterized by

$$\partial g_N^{\text{fdma}}(\mathbf{e}, \mu) = \mathbf{p} - \left\{ \frac{1}{N} \sum_{n=1}^{N} \left(\sum_{k \in I^n} \nu_{nk} \hat{s}_k^n \mathbf{e}_k \right) \middle| \sum_{k \in I^n} \nu_{nk} = 1,\ \nu_{nk} \geq 0,\ n = 1, 2, ..., N;\ k \in I^n \right\},$$

where $\mathbf{e}$ is the all-one vector. The same can be computed for a general system utility function H.

The dual problem $(D_{d,\text{fdma}}^N)$ can be solved in polynomial-time using a general purpose convex minimization algorithm (such as the ellipsoid method). More specifically, to find an ϵ-optimal dual solution, the ellipsoid method will need at most $O(K \log 1/\epsilon)$ iterations, with each iteration requiring $O(NK)$ arithmetic operations (to find the dual objective value and a subgradient; see **Procedure 1**). Thus, the ellipsoid method has an overall complexity of $O(NK^2 \log 1/\epsilon)$ to solve the dual problem $(D_{d,\text{fdma}}^N)$.

B. Generating a feasible time-shared FDMA solution for $(P_{d,\text{fdma}}^N)$ via linear programming

After obtaining an optimal solution for $(D_{d,\text{fdma}}^N)$, we must generate a primal feasible FDMA solution for $(P_{d,\text{fdma}}^N)$ while still achieving a near-optimal utility value. Notice that primal feasibility requires that each user satisfies a given instantaneous power constraint. Due to positive duality gap, achieving primal feasibility while maintaining near-optimality is highly nontrivial. Below we describe a linear programming procedure to find a near-optimal FDMA solution that can be made feasible through time-sharing.

Recall that at a dual optimal solution for $(D_{d,\text{fdma}}^N)$, say (λ^*, μ^*), we must have $\mathbf{0} \in \partial \mathcal{D}_N(\lambda^*, \mu^*)$. In light of the polyhedral representation (12), there exists some ν^* such that the following holds

$$\mathbf{p} - \frac{1}{N} \sum_{n=1}^{N} \left(\sum_{k \in I^n} \nu_{nk}^* \hat{s}_k^n \mathbf{e}_k \right) = 0,$$

or equivalently,

$$P_k = \frac{1}{N} \sum_{\{n:\ I^n \ni k\}} \nu^*_{nk} \hat{s}^n_k, \quad \forall\, k, \tag{13}$$

where $\sum_{k=1}^{K} \nu^*_{nk} = 1$, $\nu^*_{nk} \geq 0$, $\nu^*_{nk} = 0$, $k \notin I^n$. Such ν^* can be found in polynomial-time, e.g. by linear programming. One possible linear programming formulation to serve this purpose is

$$\boxed{\begin{array}{ll} \text{minimize} & \displaystyle\sum_{k=1}^{K} w_k \\ \text{subject to} & -w_k \leq P_k - \displaystyle\frac{1}{N}\sum_{n=1}^{N} \nu_{nk}\hat{s}^n_k \leq w_k,\ k = 1, ..., K, \\ & \displaystyle\sum_{k=1}^{K} \nu_{nk} = 1,\ n = 1, 2, ..., N, \\ & \nu_{nk} \geq 0,\ \forall\, k, n;\ \nu_{nk} = 0,\ k \notin I^n. \end{array}} \tag{14}$$

Since $\sum_{n=1}^{N} \sum_{k=1}^{K} \frac{1}{N}\nu^*_{nk} = 1$, the set of values $\{\frac{1}{N}\nu^*_{nk}\}$ forms a discrete probability distribution. This leads to the following (randomized) solution method to solve (P_c) using FDMA solutions.

Algorithm 1

Phase I.

Step 0 *Let $0 < \epsilon < 1$ be the required precision, and let $N := \lceil \frac{1}{\epsilon} \rceil$.*

Step 1 *Use the subgradient method or the ellipsoid method to solve $(D^N_{\text{d,fdma}})$, where at each iteration, apply* ***Procedure 1*** *to get a subgradient of $\mathcal{D}_N(\lambda, \mu)$ as needed.*

Step 2 *At optimal solution (λ^*, μ^*) for $(D^N_{\text{d,fdma}})$, compute the index sets I^n, $n = 1, 2, ..., N$, and solve (14) to get $\{\nu^*_{nk}\}$.*

Phase II.

For $k = 1, ..., K$ and $n = 1, 2, ..., N$, allocate power $\hat{s}^n_k$ to user k with probability $\frac{1}{N}\nu^*_{nk}$.

The overall complexity of **Algorithm 1** is $O(\frac{K^2}{\epsilon} \log \frac{1}{\epsilon})$ which is polynomial in terms of K and $\frac{1}{\epsilon}$. The Phase I of **Algorithm 1** generates a set of FDMA solutions, each of which by itself does not satisfy the prescribed user power budgets. The Phase II of **Algorithm 1** is a randomized procedure that generates a FDMA solution through time-sharing of tones. Randomized use of tones ensures that the average power budget P_k for each user k, $k = 1, ..., K$ can be satisfied. This power allocation policy is infeasible if the power budget is enforced instantaneously rather than on average over time. In other words, the instantaneous power may fluctuate each time, but the average power meets the prescribed user budgets.

C. A polynomial time ϵ-approximation method for the continuous DSM problem $(P_c^{\rm fdma})$

Unlike the previous subsection where the focus is on finding a high quality feasible FDMA solution for the discretized DSM problem $(D^N_{d,{\rm fdma}})$, we now develop a polynomial time strategy to find an approximately optimal FDMA solution for the continuous DSM problem $(P_c^{\rm fdma})$. The basic approach is by further partitioning of tones.

First, let us decompose the solution ν^* obtained from solving the linear program (14). Let $S_n := \{(\nu_{nk})_{k\in I^n} \mid \sum_{k\in I^n}\nu_{nk} = 1,\, \nu_{nk}\ge 0,\, k\in I^n\}$, (which is a simplex in $\Re^{|I^n|}$), with $n = 1, 2, ..., N$. Consider the Cartesian product of these simplices $S := S_1 \times S_2 \cdots \times S_N$. Clearly, a point in S is vertex of S if and only if it is a Cartesian product of vertices of simplices S_n. Also, each vertex of S corresponds to exactly one FDMA solution of $(P^N_{d,{\rm fdma}})$. Since $\nu^* \in S$, by Carathéodory's theorem, it can be expressed as a convex combination of no more than $\sum_{n=1}^N |I^n| + 1 \le NK + 1$ vertices in S. This can also be done by a simple sorting algorithm.

Procedure 2

Step 0 Initialize $\nu := \nu^*$ and let the index sets be $J^n := I^n$, $n = 1, 2, ..., N$; $\ell = 0$.

Step 1 Set $\ell := \ell + 1$. If $\nu = 0$ and J^n's are empty, stop; otherwise let $\delta := \min\{\nu_{nk} \mid n = 1, 2, ..., N;\, k \in J^n\}$. Also, for each $n = 1, 2, ..., N$, let $j_n^{\min} = \{j \in J^n \mid \nu_{nj} = \min_{k\in J^n}\nu_{nk}\}$. [If there is more than one $j \in J^n$ satisfying $\nu_{nj} = \min_{k\in J^n}\nu_{nk}$, then we pick any one of such j. In this way, $j_n^{\min}$ is always well defined.]

Step 2 Define $\omega_\ell := \delta$. For $n = 1, 2, ..., N$, let

$$J^n := \begin{cases} J^n \setminus \{j_n^{\min}\}, & \text{if } \delta = \nu_{nj_n^{\min}}, \\ J^n, & \text{otherwise} \end{cases} \quad \text{and} \quad \nu^\ell_{nk} := \begin{cases} 1, & \text{if } k = j_n^{\min}, \\ 0, & \text{else.} \end{cases}$$

[In this way, $\nu^\ell := \{\nu^\ell_{nk}\}$ is a vertex of S.]

Step 3 Update $\nu := \{\nu_{nk}\}$ by

$$\nu_{nk} := \begin{cases} \nu_{nk} - \delta, & \text{for } n = 1, 2, ..., N \text{ and } k = j_n^{\min}, \\ \nu_{nk}, & \text{for all other components.} \end{cases} \tag{15}$$

Go to **Step 1**.

In **Procedure 2**, the total cardinality of $\sum_{n=1}^N |J^n|$ is decreased by at least one whenever ℓ is increased by one. Thus, we can increase ℓ at most NK times. Moreover, it can be shown inductively that at the end of iteration m of **Procedure 2**, the following invariant relations hold

$$\nu^* = \sum_{\ell=1}^m \omega_\ell \nu^\ell + \nu, \quad \text{and} \quad \sum_{n,k}\nu^*_{nk} = N\sum_{\ell=1}^m \omega_\ell + \sum_{n,k}\nu_{nk} = N$$

where ν is given by (15). This implies that, upon termination of **Procedure 2**, we obtain the following vertex representation of ν^*

$$\nu^* = \sum_{\ell=1}^{L} \omega_\ell \nu^\ell \tag{16}$$

where ν^ℓ is a vertex in S, and $\sum_{\ell=1}^{L} \omega_\ell = 1$ with $\omega_\ell \geq 0$, $\ell = 1, ..., L$, and $L \leq NK + 1$. Moreover, by a pre-sort and a simple indexing data structure, **Procedure 2** can be implemented in $O((NK)^2 \log(NK))$ basic operations. This is because there are at most $L \leq NK + 1$ iterations, each requiring at most $O(NK \log(NK))$ arithmetic operations.

By (13), we have

$$\begin{aligned} P_k &= \frac{1}{N} \sum_{n=1}^{N} \nu^*_{nk} \hat{s}^n_k = \frac{1}{N} \sum_{n=1}^{N} \sum_{\ell=1}^{L} \omega_\ell \nu^\ell_{nk} \hat{s}^n_k \\ &= \frac{1}{N} \sum_{\ell=1}^{L} \sum_{n=1}^{N} \omega_\ell \nu^\ell_{nk} \hat{s}^n_k = \frac{1}{NL} \sum_{\ell=1}^{L} \sum_{n=1}^{N} s^{n,\ell}_k, \end{aligned} \tag{17}$$

where

$$s^{n,\ell}_k := L \omega_\ell \nu^\ell_{nk} \hat{s}^n_k. \tag{18}$$

Since ν^ℓ is a vertex of S, it follows that $\{s^{n,\ell}_k\}_{n,k}$ is a legitimate FDMA solution (over N tones) for each fixed ℓ. However, when L of these FDMA solutions are combined, they no longer satisfy the FDMA constraint since some tones may be shared across ℓ.

To accommodate these L FDMA solutions simultaneously in an non-overlapping manner, we can further partition each frequency interval $[(n-1)/N, n/N)$ into L equally-sized subintervals, each with length $1/(NL)$. We denote these subintervals subsequently as $F^n_1, ..., F^n_L$, i.e.,

$$F^n_\ell = [n/N + (\ell - 1)/(NL),\ n/N + \ell/(NL)),$$

where $\ell = 1, 2, ..., L$.

Now for each $\ell = 1, 2, ..., L$, we implement the ℓ-th FDMA solution $\{s^{n,\ell}_k\}_{n,k}$ on frequency slots $F^1_\ell, F^2_\ell, ..., F^N_\ell$. In this way, all L FDMA solutions are simultaneously implemented in an non-overlapping fashion. This gives an overall FDMA solution over NL frequency tones. According to (17), the total power allocated to user k satisfies the prescribed power budget. This leads to the following polynomial time approximation algorithm for solving (P^{fdma}_c), which can be considered as interchanging the time dimension (by the repetition of the assignments in time) in **Algorithm 1** with the space dimension (by further splitting frequency slots).

Algorithm 2

Phase I.

Step 0 *Let* $0 < \epsilon < 1$ *be the required precision, and let* $N > 0$ *be given.*

Step 1 *Use the ellipsoid method to solve* $(D^N_{d,\text{fdma}})$ *whereby at each iteration a subgradient of* $\mathcal{D}_N(\lambda, \mu)$ *is computed according to* ***Procedure 1****.*

Step 2 *At optimal solution* (λ^*, μ^*) *for* $(D^N_{d,\text{fdma}})$*, compute the index sets* I^n*,* $n = 1, 2, ..., N$*, and solve (14) to get* $v^* = \{\nu^*_{kn}\}$*.*

Phase II.

Step 0 *Use* ***Procedure 2*** *to find a convex combination for* ν^**:*

$$\nu^* = \sum_{\ell=1}^{L} \omega_\ell \nu^\ell$$

such that ν^ℓ *is a vertex in* S*, and* $\sum_{\ell=1}^{L} \omega_\ell = 1$ *with* $\omega_\ell \geq 0$*,* $\ell = 1, ..., L$*, and* $L \leq NK + 1$*.*

Step 1 *For each* $1 \leq \ell \leq L$*, create a set of subintervals* $F^n_\ell = [n/N + (\ell - 1)/(NL), n/N + \ell/(NL))$*, where* $n = 1, 2, ..., N$*.*

Step 2 *For* $\ell = 1, 2, ..., L$*, implement the* ℓ*-th FDMA solution* $\{s^{n,\ell}_k\}_{n,k}$ *on frequency slots* $F^1_\ell, F^2_\ell, ..., F^N_\ell$*. This gives an overall FDMA solution over* NL *frequency tones satisfying the prescribed power budget constraint (17).*

Theorem 3 *Suppose that noise power spectral density functions* $\sigma_k(\cdot)$ *and the crosstalk coefficients* $\alpha_{kj}(\cdot)$ *are Lipschitz continuous and are bounded*

$$0 \leq \alpha_{kj}(f) \leq \alpha, \quad \sigma_k(f) \geq \sigma > 0, \quad \forall\, f \in \Omega,\ k, j.$$

Then, for any given $\epsilon > 0$*,* ***Algorithm 2*** *finds an* ϵ*-optimal FDMA solution for* (P^{fdma}_c)*. Moreover,* ***Algorithm 2*** *runs in at most* $O\left(\frac{K^{3.5}}{\epsilon^7} \log \frac{1}{\epsilon}\right)$ *basic operations.*

Proof. Let (λ^*, μ^*) be an optimal dual solution for the dual problem $(D^N_{d,\text{fdma}})$. As shown by (17), the power allocations (18) generated by **Algorithm 2** satisfy the individual power constraint for each user. These allocations $\{s^{n,\ell}_k\}_{n,k,\ell}$ are FDMA and require NL frequency tones. Hence it is a feasible FDMA solution for the primal problem $(P^{NL}_{d,\text{fdma}})$. Moreover, this FDMA solution achieves a primal objective value $\mathcal{P}_{NL}(\{s^{n,\ell}_k\}_{n,k,\ell})$ which is equal to the dual optimal value of $(D^N_{d,\text{fdma}})$:

$$\mathcal{P}_{NL}(\{s^{n,\ell}_k\}_{n,k,\ell}) = \mathcal{D}_N(\lambda^*, \mu^*) = D^*_{N,\text{fdma}}.$$

Under the assumptions of Theorem 3, it can be checked that the rate functions

$$R_k(s_1(f),\ldots,s_K(f),f)=\log\left(1+\frac{s_k(f)}{\sigma_k(f)+\sum_{j\neq k}\alpha_{kj}(f)s_j(f)}\right),\quad k=1,2,...,K,$$

satisfy the Lipschitz condition (6). By the duality gap estimate (9) (which is an extension of Theorem 2), the duality gap $D^*_{N,\mathrm{fdma}}-P^*_{N,\mathrm{fdma}}$ is bounded by $O(1/\sqrt{N})$, where the constant in $O(\cdot)$ depends on σ, α, and $S_{\max}$. This further implies that $\mathcal{P}_{NL}(\{s_k^{n,\ell}\}_{n,k,\ell})$ is within $O(1/\sqrt{N})$ from the optimal value of (P_c^{fdma}). Thus, by selecting a large enough $N=O(1/\epsilon^2)$, we can ensure the gap between $\mathcal{P}_{NL}(\{s_k^{n,\ell}\}_{n,k,\ell})$ and the optimal value of (P_c^{fdma}) is no more than ϵ. In other words, $\{s_k^{n,\ell}\}_{n,k,\ell}$ is an ϵ-optimal solution of the continuous FDMA spectrum management problem (P_c^{fdma}).

The running time of **Algorithm 2** can be analyzed as follows: there are a total of $O(K\log(1/\epsilon))$ iterations in the ellipsoid method. Each of ellipsoid iteration requires a call of **Procedure 1** which has a complexity of $O(NK)$. Thus, computing (λ^*,μ^*) requires $O(NK^2\log(1/\epsilon)$ arithmetic operations. Solving the linear program (14) requires $O((NK)^{3.5}\log(1/\epsilon))$ operations. Phase II of **Algorithm 2** requires a call of **Procedure 2** which has a complexity of $O((NK)^2\log(NK)$. Thus, the overall complexity of **Algorithm 2** is

$$O(NK^2\log(1/\epsilon)+O((NK)^{3.5}\log(1/\epsilon))+O((NK)^2\log(NK))=O\left(\frac{K^{3.5}}{\epsilon^7}\log\frac{1}{\epsilon}\right),$$

since $N=O(1/\epsilon^2)$. [Notice that the constant in $O(\cdot)$ notation depends on the values of σ, α and $S_{\max}$.] Thus, for any fixed precision $\epsilon>0$, **Algorithm 2** finds an ϵ-optimal solution of (P_c^{fdma}) in time that is polynomial in K.

If the channels are frequency flat, then the duality gap shrinks to zero at the rate of $O(1/N)$. In this case, we can choose $N=O(1/\epsilon)$ (instead of $O(1/\epsilon^2)$ in the frequency selective case), and the resulting complexity bound of **Algorithm 2** reduces to $O\left(K^{3.5}\epsilon^{-3.5}\log 1/\epsilon\right)$. ■

Though we have provided an efficient scheme to approximately solve the continuous DSM problem (P_c^{fdma}), in practice the discretized DSM problem $(P^N_{d,\mathrm{fdma}})$ is of more interest since practical OFDMA systems have discrete sub-carriers. Furthermore, the idea of dividing a frequency tone into smaller sized frequency slots may be infeasible in practice where the tone width is usually predefined and cannot be changed. Hence, the theoretical claim of Theorem 3 may have a limited practical value. Nonetheless, the polynomial time procedure (consisting of **Algorithm 2** and **Procedure 2**) used to establish Theorem 3 is still quite valuable. First, it gives an efficient way to generate a small set (polynomial size) of candidate FDMA solutions which can be implemented, on a time-sharing basis and without tone-partitioning, to satisfy the prescribed power budget constraints *on average*. [Notice that the number of candidate FDMA solutions generated by **Algorithm 1** may be exponential.] To do so we only need to use $\{\omega_\ell\}_{\ell=1}^{L}$ (c.f. (16)) as probabilities to randomly choose the candidate FDMA solutions over time. Second, the procedure can be used to efficiently estimate the size of duality gap; the latter can be used to benchmark the performance of heuristically generated FDMA solutions.

Our final comment is on numerical stability. Recall that all the existing dual relaxation based algorithms try to generate a primal optimum from a dual optimum (either exact optimum or approximate). This process is not only numerically unstable but also theoretically unfounded (due to positive duality). After all, finding a primal optimal FDMA solution is NP-hard anyways. The way we have dealt with this problem is to use the sub-differential at the dual optimal point to construct a set of candidate FDMA solutions, and then use a linear programming procedure to combine these candidate solutions into a single high quality FDMA solution. Due to the use of sub-differential and the linear programming scheme, our approach does not suffer from the numerical sensitivity problem.

Simulation of duality gap

We have implemented the linear programming based polynomial time approximation scheme for (P_c) and tested it in a VDSL scenario with 2 upstream users and 2 downstream users, see Figure 1(a). In the simulation, the number of VDSL loops for each user is 3, while the lengths of the loops for the 4 users are 1000-ft, 1500-ft, 1200-ft and 1000-ft respectively. The frequency range is 25KHz to 12MHz. The noise variance is set to be -140dBm/Hz and the system gap is 12dBm. For each user, the power constraint over the whole spectrum and the per-tone spectrum mask are -3dBm and -1.8dBm/tone, respectively.

We discretized the available spectrum into n tones, where n varies from 2 to 200 with an increment of 2. For each n, we generated the VDSL communication channel parameters under the above mentioned setup by using a VDSL system simulation program (courtesy of Professor Wei Yu, University of Toronto). Then, we apply Phase I of **Algorithm 2** to each of the n-tone sum-rate spectrum management problem with FDMA constraint. This gives us the dual optimal value $D^*_{n,\mathrm{fdma}}$. We then solve the linear program (14) using the SeDuMi software [21] to obtain a set of L_n FDMA solutions (each by itself may not satisfy the primal power constraint). In Phase II of **Algorithm 2**, every one of the n tones is further divided to L_n sub-tones. Over the newly created $N = nL_n$ tones, the L_n n-tone FDMA solutions can be combined to obtain a feasible N-tone FDMA solution, achieving a certain sum-rate value which we denote by $P_{N,\mathrm{fdma}}$. Let $P^*_{N,\mathrm{fdma}}$ denote the maximum achievable sum-rate for a N-tone FDMA solution. Then, we have $P^*_{N,\mathrm{fdma}} \geq P_{N,\mathrm{fdma}}$. Thus, the duality gap for the N-tone FDMA sum-rate maximization problem can be upper bounded by

$$D^*_{N,\mathrm{fdma}} - P^*_{N,\mathrm{fdma}} \le D^*_{N,\mathrm{fdma}} - P_{N,\mathrm{fdma}}.$$

Given the primal-dual values $\{D^*_{N,\mathrm{fdma}}, P_{N,\mathrm{fdma}} : N = nL_n,\ n = 2, 4, 6, ..., 200.\}$, this upper bound is computable. In Figure 1(b), we plot the duality gap upper bound $D^*_{N,\mathrm{fdma}} - P_{N,\mathrm{fdma}}$ versus N. It is clear from this plot that the duality gap does decrease to zero when N increases, as claimed by Theorem 2.

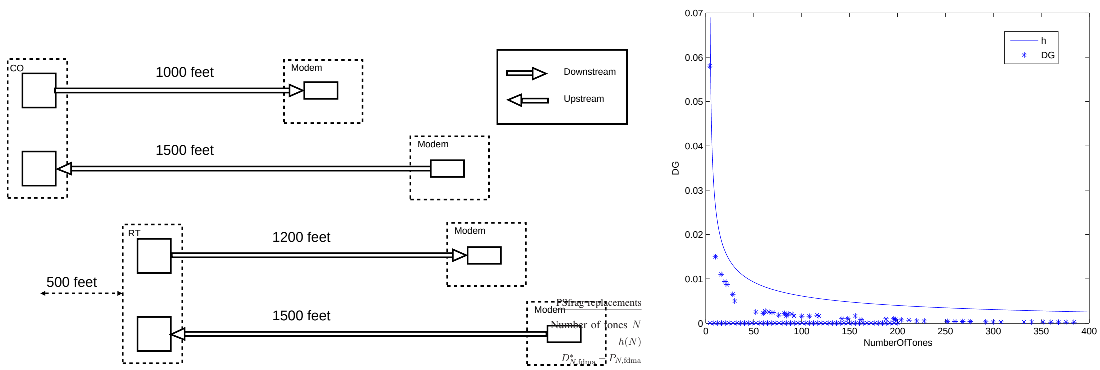

Fig. 1. (a) VDSL system setup (b) Duality gap v.s. N

Moreover, in Figure 1(b), the curve $h(N) = \frac{0.1}{\sqrt{N}} + 0.04$ upper bounds the duality estimate $D^*_{N,\mathrm{fdma}} - P_{N,\mathrm{fdma}}$ for all N, illustrating that the duality gap vanishes asymptotically at the rate of $\mathcal{O}(1/\sqrt{N})$. Notice that there are some values of N for which the duality gap is zero. This happens exactly when a feasible FDMA solution is found in the Phase I of **Algorithm 2**, that is, $L_n = 1$.

V. Discussions

In a multiuser environment, dynamic spectrum management (DSM) is an important technique for achieving high spectral efficiency. Central to the success of DSM is the ability to efficiently compute optimal spectrum sharing strategies in response to physical channel conditions. For the continuous DSM problem (P_c), it is known that the strong duality (i.e., zero duality gap) holds, even though the problem lacks apparent convexity. This surprising result was first discovered by Yu and Lui [23] and later rigorously established [16] under a Lebesgue integral framework and using a classical result of Lyapunov [17]. For practical implementations (using, say, OFDM), it is necessary to discretize the corresponding Lebesgue integrals using Riemann partial sums and consider the resulting discretized DSM problem (P_d^N).

However, the discretized DSM problem (P_d^N) is very difficult to optimize or to approximately solve in general; in fact, both tasks are NP-hard, even when restricted to the set of FDMA strategies [16]. In this paper, we instead focus on the original continuous DSM problem (P_c), since the latter possesses a hidden convexity. We have shown that it is possible to find an ϵ-optimal FDMA solution for (P_c) in polynomial time, for any $\epsilon > 0$. A major step in this work is to use the hidden convexity to analyze the effect of discretization on the DSM problem (P_c) as the size of discretization decreases to zero. Our analysis shows that the duality gap for the discretized DSM problem (P_d^N) converges to zero at the rate of $O(1/\sqrt{N})$ if channels are frequency selective, and at the rate of $O(1/N)$ when channels are frequency flat. At present, it is not clear if the slower rate of $O(1/\sqrt{N})$ is a genuine consequence of frequency selectivity or merely a weakness of our analysis, although we suspect the latter. Another key step in the design of our polynomial time approximation algorithm is to avoid the exponential size of a linear program that must be solved in order to identify a small number of FDMA solution candidates. Jointly (but not individually), these candidates satisfy the power constraints of all users, and they can be easily implemented in an non-overlapping fashion once we further partition each frequency tone into a few smaller sized frequency slots. In this way, we derive a feasible FDMA solution for the original continuous DSM problem (P_c) while still achieving a provably near-optimal system utility value.

It is important to point out that the duality gap estimates $O(1/\sqrt{N})$ hold for general spectrum sharing strategies, whether they are FDMA or non-FDMA. However, the linear programming based polynomial time approximation scheme for the DSM problem (P_c^{fdma}) does require FDMA assumption. This is a major limitation of current work. Notice that the (only) part in our approach that requires FDMA assumption is in the solution of the dual problem (D_d^N). FDMA assumption ensures polynomial time solvability of (D_d^N). Does there exist any other classes of spectrum sharing strategies for which the dual DSM problem (D_d^N) is polynomial time solvable? Any progress in this direction, when combined with the analysis in this paper, will lead to an overall polynomial time approximation scheme for a broader class of DSM problems. This is an interesting direction for future research. Finally, we believe the full impact of Lyapunov theorem is still not well understood. It may be possible to use the hidden convexity associated with the DSM problem in a game theoretic framework or in a more network oriented formulation involving multiple wireless users.

Acknowledgement: The authors wish to thank Professor Jan Brinkhuis, Erasmus University, for the discussions leading to the proof of Proposition 1. They are also grateful to Professor Wei Yu, University of Toronto, for providing his VDSL system simulation program and to Yao Huang, University of Minnesota, for her help in generating Figure 1.

Appendix A Proof of Proposition 1

Throughout this proof, we use $\|\cdot\|$ to denote the infinity norm. The Lyapunov theorem ([5], [17]) asserts that W is a closed and convex set. Take an arbitrary $w \in W$ and let $s(f)$ be the corresponding Lebesgue integrable function such that $w_k = \int_0^1 R_k(s(f), f)df$, $k = 1, ..., K$. For any given $N > 0$, by definition of the Lebesgue integration (as the limit of the integration of a series of simple functions; see e.g. Chapter 4 of [19]) there exist mutually disjoint Lebesgue measurable sets $A_1, ..., A_M$, such that $\bigcup_{j=1}^M A_j = [0,1]$ and

$$\left\| \int_0^1 R(s(f), f)df - \sum_{j=1}^M R(s(f_j), f_j)\mu(A_j) \right\| < 1/\sqrt{N}, \tag{19}$$

where $f_j \in A_j$, $j = 1, ..., M$. In the above approximation, A_j's are level sets associated with some appropriately partitioned range set $\{R(s(f), f) \mid f \in [0,1]\}$. By further refining each A_j if necessary, we can assume that the range of $R(s(f), f)$ over A_j is contained in some interval of size $1/n$, where $n := \lceil \sqrt{N} \rceil$. Define two new simple functions:

$$\begin{aligned} \bar{s}(f) &:= s(f_j) && \text{for } f \in A_j, \\ \bar{R}(f) &:= R(s(f_j), f_j) && \text{for } f \in A_j. \end{aligned} \tag{20}$$

Let

$$\bar{w} := \int_0^1 \bar{R}(f)df = \sum_{j=1}^M R(s(f_j), f_j)\mu(A_j). \tag{21}$$

By (19) it follows that $\|w - \bar{w}\|_\infty < 1/\sqrt{N}$. Now let $A_j^\ell := [\frac{\ell}{n}, \frac{\ell+1}{n}) \cap A_j$, and $t_j^\ell := \mu(A_j^\ell)$, $j = 1, ..., M; \ell = 1, ..., n$. Clearly, $t_j^\ell \geq 0$ and $\sum_{j=1}^M t_j^\ell = \frac{1}{n}$, for all $\ell = 1, ..., n$. Let

$$\bar{w}^\ell := \int_{\frac{\ell}{n}}^{\frac{\ell+1}{n}} \bar{R}(f)df = \sum_{j=1}^M \bar{R}(f_j)t_j^\ell.$$

It follows from Carathéodory's theorem that there is a subset $\mathcal{J}^\ell \subseteq \{1, ..., M\}$ and $\bar{t}_j^\ell > 0$ with $\sum_{j \in \mathcal{J}^\ell} \bar{t}_j^\ell = \frac{1}{n}$, such that $K_\ell := |\mathcal{J}^\ell| \leq K + 1$, and

$$\bar{w}^\ell = \sum_{j \in \mathcal{J}^\ell} \bar{R}(f_j)\bar{t}_j^\ell.$$

Since $\bar{t}_j^\ell > 0$, it follows that $A_j^\ell \neq \emptyset$. Let $f_j^\ell \in A_j^\ell \subset A_j$. Recall that the range of $R(s(f), f)$ over A_j is contained in an interval of size $1/n$ and $f_j, f_j^\ell \in A_j$. It follows that

$$\|R(s(f_j), f_j) - R(s(f_j^\ell), f_j^\ell)\| \leq \sqrt{K}/n. \tag{22}$$

Also, since $R(s, f)$ is Lipschitz continuous (see (6) and $|f_j^\ell - \ell/n| \leq 1/n$, it follows that

$$\|R(s(f_j^\ell), f_j^\ell) - R(s(f_j^\ell), \ell/n)\| \leq L/n, \text{ for any } \ell = 1, ..., n, \tag{23}$$

where $L > 0$ is the Lipschitz constant.

By rearranging the indices in $\mathcal{J}^\ell$ if necessary, we can assume $\mathcal{J}^\ell = \{1, ..., K_\ell\}$. Furthermore, denote

$$T_0^\ell := \frac{\ell}{n},\, T_{j+1}^\ell := T_j^\ell + \bar{t}_{j+1}^\ell,\ \text{for } j = 0, 1, ..., K_\ell - 1,$$

and introduce two new step-functions on the interval $[\frac{\ell}{n}, \frac{\ell+1}{n})$:

$$\begin{aligned} \hat{s}^\ell(f) &:= s(f_j^\ell) && \text{for } f \in [T_{j-1}^\ell, T_j^\ell),\ j = 1, 2, ..., K_\ell \\ \hat{R}^\ell(f) &:= \bar{R}(f_j) = R(s(f_j), f_j) && \text{for } f \in [T_{j-1}^\ell, T_j^\ell),\ j = 1, 2, ..., K_\ell. \end{aligned} \tag{24}$$

By this construction, both $\hat{s}^\ell(f)$ and $\hat{R}^\ell(f)$ have at most $K+1$ breakpoints on $[\frac{\ell}{n}, \frac{\ell+1}{n})$. Moreover, we have

$$\int_{\frac{\ell}{n}}^{\frac{\ell+1}{n}} \hat{R}^\ell(f) df = \sum_{j \in \mathcal{J}^\ell} \bar{R}(f_j) \bar{t}_j^\ell = \bar{w}^\ell = \int_{\frac{\ell}{n}}^{\frac{\ell+1}{n}} \bar{R}^\ell(f) df. \tag{25}$$

Now we can use (22)–(23) to obtain

$$\begin{aligned} \left\| \int_{\frac{\ell}{n}}^{\frac{\ell+1}{n}} \hat{R}^\ell(f) df - \int_{\frac{\ell}{n}}^{\frac{\ell+1}{n}} R(\hat{s}^\ell(f), \ell/n) df \right\| &= \left\| \sum_{\ell=1}^{K_\ell} \left(\int_{T_{j-1}^\ell}^{T_j^\ell} R(s(f_j), f_j) df - \int_{T_{j-1}^\ell}^{T_j^\ell} R(s(f_j^\ell), \ell/n) df \right) \right\| \\ &\le \left\| \sum_{\ell=1}^{K_\ell} \left(\int_{T_{j-1}^\ell}^{T_j^\ell} R(s(f_j), f_j) df - \int_{T_{j-1}^\ell}^{T_j^\ell} R(s(f_j^\ell), f_j^\ell) df \right) \right\| \\ &\quad + \left\| \sum_{\ell=1}^{K_\ell} \left(\int_{T_{j-1}^\ell}^{T_j^\ell} R(s(f_j^\ell), f_j^\ell) df - \int_{T_{j-1}^\ell}^{T_j^\ell} R(s(f_j^\ell), \ell/n) df \right) \right\| \\ &\le \sum_{\ell=1}^{K_\ell} \left\| \int_{T_{j-1}^\ell}^{T_j^\ell} R(s(f_j), f_j) df - \int_{T_{j-1}^\ell}^{T_j^\ell} R(s(f_j^\ell), f_j^\ell) df \right\| \\ &\quad + \sum_{\ell=1}^{K_\ell} \left\| \int_{T_{j-1}^\ell}^{T_j^\ell} R(s(f_j^\ell), f_j^\ell) df - \int_{T_{j-1}^\ell}^{T_j^\ell} R(s(f_j^\ell), \ell/n) df \right\| \\ &\le K_\ell(\sqrt{K}/n^2 + L/n^2) \\ &\le (K+1)(\sqrt{K} + L)/n^2. \end{aligned} \tag{26}$$

On each interval $[\frac{\ell}{n}, \frac{\ell+1}{n})$, the function $\hat{R}^\ell(f)$ is a step function with at most $K+1$ breakpoints. However, the breakpoints are not necessarily on the uniformly spaced grid points in $[0, 1]$. For this reason, we need to further approximate $\hat{R}^\ell(f)$ by further subdividing each interval $[\frac{\ell}{n}, \frac{\ell+1}{n})$. In particular, let us consider step-functions defined over $[0, 1)$ with n^2 uniformly spaced breakpoints. In this way, there are n uniform subintervals in each interval $[\frac{\ell}{n}, \frac{\ell+1}{n})$. Since $R(\hat{s}^\ell(f), \ell/n)$ is a step-function in $[\frac{\ell}{n}, \frac{\ell+1}{n})$ with at most $K+1$ breakpoints, these $K+1$ breakpoints can be approximated by uniformly spaced grid points in $[\frac{\ell}{n}, \frac{\ell+1}{n})$ with $1/n^2$ accuracy. Thus, there must be a step function $\tilde{s}^\ell(f)$ defined on the n uniform subintervals of

$[\frac{\ell}{n}, \frac{\ell+1}{n})$ such that

$$\left\| \int_{\frac{\ell}{n}}^{\frac{\ell+1}{n}} R(\hat{s}^{\ell}(f), \ell/n) df - \int_{\frac{\ell}{n}}^{\frac{\ell+1}{n}} R(\tilde{s}^{\ell}(f), \ell/n) df \right\| \leq \|R\| K/n^2, \tag{27}$$

where $\|R\|$ is the maximum absolute value of $R(s(f), f)$ over $[0, 1]$. Again, by Lipschitz continuity, we have

$$\left\| \int_{\frac{\ell}{n}}^{\frac{\ell+1}{n}} R(\tilde{s}^{\ell}(f), \ell/n) df - \int_{\frac{\ell}{n}}^{\frac{\ell+1}{n}} R(\tilde{s}^{\ell}(f), f) df \right\| \leq L/n^2. \tag{28}$$

Combining the inequalities (26), (27) and (28) yields

$$\left\| \int_{\frac{\ell}{n}}^{\frac{\ell+1}{n}} R(\tilde{s}^{\ell}(f), f) df - \int_{\frac{\ell}{n}}^{\frac{\ell+1}{n}} \hat{R}^{\ell}(f) df \right\| \leq c/n^2,$$

where $c = (L(2+K) + \|R\|K + (K+1)\sqrt{K})$. Due to (25), the above inequality yields

$$\left\| \int_{\frac{\ell}{n}}^{\frac{\ell+1}{n}} R(\tilde{s}^{\ell}(f), f) df - \bar{w}^{\ell} \right\| \leq c/n^2. \tag{29}$$

Extending $\tilde{s}^{\ell}(f)$ to the entire interval $[0, 1)$, i.e., let $\tilde{s}(f) := \tilde{s}^{\ell}(f)$ for $f \in [\frac{\ell}{n}, \frac{\ell+1}{n})$ yields

$$\left\| \int_0^1 R(\tilde{s}(f), f) df - \sum_{\ell=1}^{n} \bar{w}^{\ell} \right\| = \left\| \int_0^1 R(\tilde{s}(f), f) df - \bar{w} \right\| \leq c/n. \tag{30}$$

Combining (19), (21), (30), and noting that $n \leq \sqrt{N} + 1$, it follows that

$$\left\| \int_0^1 R(\tilde{s}(f), f) df - w \right\| < (c+1)/\sqrt{N}.$$

Since $\int_0^1 R(\tilde{s}(f), f) df \in W^N$, we have established (8).

For a frequency flat environment, the rate function $R(s(f), f)$ depends on f only implicitly through $s(f)$. In other words, $R(s(f), f)$ takes the form of $R(s(f))$. In this case, the error estimation in (8) can be further improved. In particular, for any $N > 0$, there exists some $M > 0$ such that

$$\left\| \int_0^1 R(s(f)) df - \sum_{j=1}^{M} R(s(f_j)) \mu(A_j) \right\| < 1/N, \tag{31}$$

where A_j, $j = 1, 2, ..., M$ are some disjoint measurable subsets of $[0, 1]$. We can also define $\bar{R}(f)$, $\bar{s}(f)$ and $\bar{w}$ just like (20)–(21). Moreover, let $t_j = \mu(A_j)$, $j = 1, 2, ..., M$, so that $\sum_j t_j = 1$. We can construct step functions $\hat{s}(f)$, $\hat{R}(f)$ similar to the construction of $\hat{s}^{\ell}(f)$ and $\hat{R}^{\ell}(f)$ in (24) except we replace the subinterval $[\ell/n, (\ell+1)/n]$ by the interval $[0, 1]$. The step functions $\hat{s}^{\ell}(f)$ and $\hat{R}^{\ell}(f)$ have at most $K+1$ breakpoints. These breakpoints can be approximated to $1/N$ accuracy by a uniformly spaced grids $\{0, 1/N, 2/N, ..., (N-1)/N, 1\}$. Thus, we can construct step functions $\tilde{s}(f)$ and $\tilde{R}(f)$ over the N uniformly partitioned subintervals of $[0, 1]$ so that

$$\left\| \int_0^1 R(\hat{s}(f)) df - \int_0^1 R(\tilde{s}(f)) df \right\| \leq \|R\| K/N.$$

■

Appendix B Proof of Theorem 2

Fix an $N > 0$. Let $s^*(f)$ denote the optimal transmit power spectra for the continuous spectrum management problem (P_c). Define

$$R_k(s(f), f) = s_{k-K}(f), \quad k = K+1, ..., 2K.$$

Denote $R(s(f), f) = (R_1(s(f), f), ..., R_{2K}(s(f), f))$ and let

$$\|R\|_\infty = \max_{k=1,\ldots,2K} \sup_{f\in[0,1]} |R_k(s^*(f), f)|.$$

By Proposition 1, there exists some step function $\tilde{s}(f)$ with breakpoints $\{0, 1/N, ..., (N-1)/N, 1\}$ such that $0 \le \tilde{s}_k(f) \le S_{\max}$ and

$$\left\| \int_0^1 R(\tilde{s}(f), f) df - \int_0^1 R(s^*(f), f) df \right\|_\infty < c_1 N^{-1/2}, \tag{32}$$

where $c_1 \equiv (L(2+K) + \|R\|K + (K+1)\sqrt{K}) + 1$. Let $v_k^* \equiv u_k(s^*(f))$ and $v_k \equiv u_k(\tilde{s})$ for all $k = 1, 2, ..., K$. Then the above estimate (32) implies

$$|v_k - v_k^*| \le c_1 N^{-1/2}, \quad \forall\, k = 1, 2, ..., K. \tag{33}$$

Furthermore, the estimate (32) also implies

$$\int_0^1 R_k(\tilde{s}(f), f) df < \int_0^1 R_k(s^*(f), f) df + c_1 N^{-1/2}, \quad k = K+1, ..., 2K,$$

which can be written as

$$\int_0^1 \tilde{s}_k(f) df \le P_k + c_1 N^{-1/2}, \quad k = 1, ..., K.$$

Define

$$\bar{s}_k(f) = \frac{P_k}{P_k + c_1 N^{-1/2}} \tilde{s}_k(f), \quad k = 1, 2, ..., K.$$

Then, $\bar{s}_k(f)$ is a step function with breakpoints being $\{0, 1/N, ..., (N-1)/N, 1\}$ and there holds

$$\int_0^1 \bar{s}_k(f) df = \frac{1}{N} \sum_{n=1}^{N} \bar{s}_k(n/N) \le P_k, \quad k = 1, 2, ..., K,$$

which means that $\{\bar{s}_k(f)\}$ satisfy all the constraints of the discretized spectrum management problem (P_d^N). Moreover, we have

$$|\bar{s}_k(f) - \tilde{s}_k(f)| \le \frac{c_1 N^{-1/2}}{(P_k + c_1 N^{-1/2})} \le c_2 N^{-1/2}, \quad \forall\, k = 1, 2, ..., K,$$

where the constant c_2 is a positive constant defined by

$$c_2 \equiv \max_{N,k\geq 1} \frac{c_1}{(P_k + c_1 N^{-1/2})} = c_1(\max_k P_k^{-1}).$$

Let $u_k \equiv u_k(\bar{s})$. This further implies

$$\begin{aligned} |u_k - v_k| &\leq \left| \int_0^1 R_k(\bar{s}, f)df - \int_0^1 R_k(\tilde{s}, f)df \right| \leq \int_0^1 |R_k(\bar{s}, f) - R_k(\tilde{s}, f)|df \\ &\leq \int_0^1 L\|\bar{s} - \tilde{s}\|_\infty df \leq c_2 L N^{-1/2}, \end{aligned}$$

where we have used the Lipschitz continuity of the R_k functions. Combining this with (33) yields

$$|u_k - v_k^*| \leq (c_1 + c_2 L)N^{-1/2}, \quad \forall\, k.$$

Finally, since the system utility function $H(u_1, u_2, ..., u_K)$ is Lipschitz continuous, we have

$$|H(u_1, u_2, ..., u_K) - H(v_1^*, v_2^*, ..., v_K^*)| \leq L\|u - v^*\|_\infty$$

implying

$$H(u_1, u_2, ..., u_K) \geq H(v_1^*, v_2^*, ..., v_K^*) - (c_1 + c_2 L)LN^{-1/2},$$

or

$$P_N^* \geq P_c^* - (c_1 + c_2 L)LN^{-1/2}.$$

By a similar argument, we can bound the dual objective values as

$$D_N^* \leq D_c^* + c_3 L N^{-1/2},$$

for some positive constant c_3. Since $P_c^* = D_c^*$ (strong duality for the continuous formulation), we conclude

$$D_N^* - P_N^* \leq (c_1 + c_3 + c_2 L)LN^{-1/2},$$

which establishes Theorem 2. ■

References

[1] Aubin, J.P. and Ekeland, I., "Estimates of the Duality Gap in Nonconvex Optimization," *Mathematics of Operations Research,* Vol. 1, No. 3, pp. 225–245, 1976.

[2] Aumann, R. and Perles, M., "A Variational Problem Arising in Economics," *Journal of Mathematical Analysis and Applications*, Vol. 11, pp. 488–503, 1965.

[3] Avriel, M., W.E. Diewert, S. Schaible, and I. Zang, *Generalized Concavity*. Plenum Press, New York, 1988.

[4] Barvinok, A., *A Course in Convexity*, Graduate Studies in Mathematics, Vol. 54, American Mathematical Society, 2002.

[5] Blackwell, D., "On a Theorem of Lyapunov," *The Annals of Mathematical Statistics*, 22, pp. 112 – 114, 1951.

[6] Chan, V.M.K. and Yu, W., "Joint Multi-User Detection And Optimal Spectrum Balancing For Digital Subscriber Lines," *European Journal on Applied Signal Processing (EURASIP), special issue on Digital Subscriber Lines,* 2006.

[7] Cendrillon, R., Yu, W., Moonen, M., Verliden, J. and Bostoen, T., "Optimal Multi-User Spectrum Management for Digital Subscriber Lines," *IEEE Transactions on Communications*, Vol. 54, No. 5, pp. 922–933, May 2006.

[8] Chung, S.T., Kim, S. J., Lee, J. and Cioffi, J.M., "A Game-Theoretic Approach to Power Allocation in Frequency-Selective Gaussian Interference Channels," *Proceeding in 2003 IEEE International Symposium on Information Theory,* Yokohama, Japan, 2003.

[9] Cover, T.M. and Thomas, J.A., *Elements of Information Theory*, John Wiley & Sons, Inc., 1991.

[10] Etkin, R., Parekh, A. and Tse, D., "Spectrum Sharing for Unlicensed Bands," *Proceedings of First IEEE Symposium on New Frontiers in Dynamic Spectrum Access Networks*, pp. 251–258, 2005.

[11] Hayashi, S. and Luo, Z.-Q., "Spectrum Management for Interference-limited Multiuser Comminication Systems," Accepted for publication in *IEEE Transactions on Information Theory*, 2008.

[12] Haykin, S.,"Cognitive Radio: Brain-Empowered Wireless Communications," *IEEE Journal Selected Areas in Communications,* Vol. 23, pp. 201–220, 2005.

[13] Lasry, J.M. and Berliocchi, H., "Integrandes Nomales et Mesures Parametrees en Calcul des Variations," *Bulletin of the Society of Mathematics of France*, Vol. 101, pp. 129–184, 1973.

[14] Lui, R. and Yu, W., "Low Complexity Near Optimal Spectrum Balancing For Digital Subscriber Lines," *IEEE International Conference on Communications (ICC)*, Seoul, Korea, 2005.

[15] Luo, Z.-Q. and Pang, J.-S., "Analysis of Iterative Water-Filling Algorithm For Multi-User Power Control In Digital Subscriber Lines," Special issue of *EURASIP Journal on Applied Signal Processing on Advanced Signal Processing Techniques for Digital Subscriber Lines,* Vol. 2006, Article ID 24012, 10 pages, 2006.

[16] Luo, Z.-Q. and Zhang, S., "Dynamic Spectrum Management: Complexity and Duality," *IEEE Journal of Selected Topics in Signal Processing*, Special Issue on Signal Processing and Networking for Dynamic Spectrum Access, Vol. 2, No. 1, pp. 57–73, February 2008.

[17] Lyapunov, A.M., "Sur les Fonctions-vecteur Complètement Adďitives," *Bull. Acad. Sci. URSS. Sèr. Math.* Vol. 4, pp. 465–478, 1940.

[18] Rockafellar, R.T., *Convex Analysis*, Princeton University Press, 1970.

[19] Royden, H.L., *Real Analysis*, 3rd Edition, Prentice Hall, 1988.

[20] Song, K.B., Chung, S.T., Ginis, G. and Cioffi, J.M., "Dynamic Spectrum Management for Next-generation DSL Systems," *IEEE Communications Magazine,* Vol. 40, pp. 101–10, 2002.

[21] Sturm, J. F., "Using SeDuMi 1.02, a MATLAB Toolbox for Optimization over Symmetric Cones," *Optimization Methods and Software*, vol. 11-12, pp. 625–653, 1999; also see the website http://sedumi.mcmaster.ca/.

[22] Yamashita, N. and Luo, Z.-Q., "A Nonlinear Complementarity Approach To Multi-User Power Control For Digital Subscriber Lines," *Optimization Methods and Software,* Vol. 19, pp. 633–652, 2004.

[23] Yu, W. and R. Lui. "Dual Methods for Nonconvex Spectrum Optimization of Multicarrier Systems," *IEEE Transactions on Communications,* Vol. 54, pp. 1310–1322, 2006.

[24] Yu, W., Ginis, G., and Cioffi, J. M., "Distributed Multi-User Power Control For Digital Subscriber Lines," *IEEE Journal on Selected Areas in Communications*, Vol. 20, pp. 1105–1115, 2002.

[25] Yu, W., Lui, R. and Cendrillon, R. , "Dual Optimization Methods For Multi-User Orthogonal Frequency Division Multiplex Systems," *IEEE Global Communications Conference (Globecom)*, Vol. 1, pp. 225–229, Dallas, USA, 2004.