Dual Methods for Nonconvex Spectrum Optimization of Multicarrier Systems

Wei Yu, *Member, IEEE*, and Raymond Lui

***Abstract*—The design and optimization of multicarrier communications systems often involve a maximization of the total throughput subject to system resource constraints. The optimization problem is numerically difficult to solve when the problem does not have a convexity structure. This paper makes progress toward solving optimization problems of this type by showing that under a certain condition called the *time-sharing* condition, the duality gap of the optimization problem is always zero, regardless of the convexity of the objective function. Further, we show that the time-sharing condition is satisfied for practical multiuser spectrum optimization problems in multicarrier systems in the limit as the number of carriers goes to infinity. This result leads to efficient numerical algorithms that solve the nonconvex problem in the dual domain. We show that the recently proposed *optimal spectrum balancing algorithm* for digital subscriber lines can be interpreted as a dual algorithm. This new interpretation gives rise to more efficient dual update methods. It also suggests ways in which the dual objective may be evaluated approximately, further improving the numerical efficiency of the algorithm. We propose a low-complexity iterative spectrum balancing algorithm based on these ideas, and show that the new algorithm achieves near-optimal performance in many practical situations.**

***Index Terms*—Digital subscriber lines (DSLs), discrete multitone (DMT), duality theory, dynamic spectrum management (DSM), iterative spectrum balancing (ISB), nonconvex optimization, optimal spectrum balancing (OSB), orthogonal frequency-division multiplex (OFDM).**

I. Introduction

THE design of communication systems often involves the optimization of a design objective subject to various resource constraints. The optimization problem becomes numerically difficult to solve when either the objective function or the constraint lacks a convexity structure. This paper deals with efficient numerical solutions for nonconvex optimization problems for a particular class of optimization problems that often arise in multicarrier communication systems. In a multicarrier system, the transmission frequency spectrum is partitioned into a large number of frequency bins on which parallel data transmissions take place. The most common examples of multicarrier systems include wireless orthogonal frequency-division multiplex (OFDM) systems, such as the 802.11 and digital audio broadcasting systems, and wireline discrete multitone (DMT) systems, such as the digital subscriber line (DSL) systems. In both DMT and OFDM systems, a pair of discrete Fourier transform (DFT) and inverse discrete Fourier transform (IDFT) is used to partition the frequency band into independent subchannels. Adaptive spectrum shaping and bit allocation may be easily implemented on a carrier-by-carrier basis.

A central issue in the design of adaptive multicarrier systems is that of optimal spectrum and bit allocation across the frequency domain. The issue is well understood for single-user systems in which the optimal solution resembles an information-theoretically optimal "waterfilling" solution. However, the problem is nontrivial when multiple users are present at the same time. In the latter case, the design objective function and the constraints are often nonconvex, and the optimization problem becomes computationally difficult to solve.

This paper makes progress toward numerical solution of nonconvex optimization problems for multicarrier systems by studying their fundamental properties. In particular, we focus on the characterization of the Lagrangian dual of these nonconvex problems. Our main result is that a nonconvex optimization problem in multicarrier systems has a *zero* duality gap whenever a so-called "time-sharing" condition is satisfied. Further, the time-sharing condition is always satisfied for practical multiuser spectrum optimization problems in multicarrier systems when the number of frequency carriers goes to infinity. This result is surprising at a first glance, as nonconvex optimization problems generally have a nonzero duality gap. Yet, the result is very useful, as it opens up the possibilities of rigorously solving for the global optimum of nonconvex problems in the dual domain.

Lagrangian duality theory for general convex optimization problems is well known [1], [2]. For nonconvex multiuser spectrum optimization problems, existing approaches in the literature typically focus on either the convex relaxation of the problem [3]–[5] or heuristic methods that approximate the global solution of the problem [6]–[11]. In both cases, the global optimality of the solution is difficult to prove. Recently, an *optimal spectrum balancing* (OSB) method for the DSL multiuser spectrum optimization problem is proposed in [12], where a first proof of global optimality of a bit-loading algorithm for a nonconvex problem is provided. This paper generalizes the result of [12] and reinterprets the OSB algorithm in a dual optimization framework. Our main contribution is a theoretical

Paper approved by G.-H. Im, the Editor for Transmission Systems of the IEEE Communications Society. Manuscript received July 14, 2005; revised December 12, 2005. This work was supported in part by Bell Canada University Laboratories, in part by Communications and Information Technology Ontario (CITO), in part by the Natural Sciences and Engineering Council (NSERC) of Canada, and in part by the Canada Research Chairs program. This paper was presented in part at the IEEE Global Telecommunications Conference, Dallas, TX, December 2004, and in part at the IEEE International Conference on Communications, Paris, France, May 2005.

The authors are with the Edward S. Rogers Sr. Electrical and Computer Engineering Department, University of Toronto, Toronto, ON M5S 3G4, Canada (e-mail: weiyu@comm.utoronto.ca; rwmlui@gmail.com).

Digital Object Identifier 10.1109/TCOMM.2006.877962

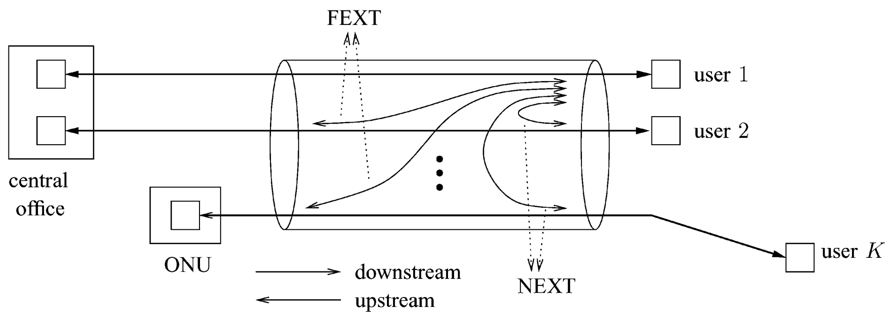


Fig. 1. Multiuser DSLs.

treatment of nonconvex optimization problems for multicarrier systems, and a precise condition for zero duality gap. Our general theory is inspired by the earlier work of Aubin and Ekeland [13] and Bertsekas *et al.* [14], [15], where an integer programming problem is considered in a similar context.

The general theory presented in this paper also gives rise to several algorithmic improvements to the OSB algorithm. First, by reinterpreting the OSB algorithm as a dual algorithm, we show that the dual updates may be done efficiently using subgradient and ellipsoid algorithms. This significantly speeds up the convergence of the implementation. Second, we propose a low-complexity method to approximate the evaluation of the Lagrangian dual objective, which allows the computational complexity to be further reduced at a small cost to performance. Both are significant steps toward making the algorithm practical. This paper aims to provide an optimization-theoretical viewpoint to the multiuser spectrum optimization problem. Theoretical approaches to this problem have been proved to be fruitful in the past. For example, in a recent work [16], a nonlinear complementarity approach to the spectrum optimization problem is used to provide very interesting insights to the DSL problem.

The rest of the paper is organized as follows. Section II contains the system model and the problem formulation. In Section III, a general theory of nonconvex optimization problems for multicarrier systems is presented and the main duality gap result is proved. Section IV provides efficient ways to update dual variables for OSB. In Section V, a new low-complexity algorithm is proposed for the evaluation of dual objective, which further reduces the computational complexity of the algorithm. Section VI contains simulation results. Concluding remarks are made in Section VII.

II. Spectrum Optimization Problems

This paper is primarily motivated by the recent surge of interest in dynamic spectrum management (DSM) for digital subscriber lines. In a DSL system, multiple copper twisted pairs are bundled together. The electromagnetic coupling between the copper pairs causes crosstalk interference, which has long been identified as the primary source of line impairment in DSL deployments. Current DSL systems use a static spectrum management (SSM) approach where a fixed transmit power spectral density (PSD) is applied to each line regardless of the loop topology or user service requirements. The performance projection under SSM is based on the levels of worst-case crosstalk interference.

Future generations of DSL services are envisioned to implement DSM [17], [18], in which, each line is given an ability to adapt to its loop environment and service requirements individually across the spectrum. DSM is facilitated by the adoption of multicarrier modulation by DSL standardization bodies. The ability to set the PSD level of each frequency carrier individually gives DSM techniques the potential to greatly improve the achievable rates and service ranges of current DSL systems. On the other hand, the large number of design variables also present a research challenge from an optimization point of view. DSM is an active area of research, both within the research community and within the standardization bodies [19].

Under the SSM scheme, where the line interference is assumed to be fixed (or assumed to have a worst-case PSD level), the spectrum optimization problem simplifies to the following. The design objective is to maximize the overall system throughput, which is the sum of individual rates in each frequency carrier. The design constraint is a power constraint coupled across all the carriers. Let S^n denote the transmit PSD at the nth carrier. The optimization problem is

$$\begin{aligned} \text{maximize} \quad & \sum_{n=1}^{N} \log\left(1+\frac{S^n}{\sigma^n}\right) \\ \text{subject to} \quad & \sum_{n=1}^{N} S^n \le P \\ & S^n \ge 0 \end{aligned} \tag{1}$$

where σ^n is the combined noise and interference PSD at the nth carrier normalized by $\Gamma/|H^n|^2$. Here, Γ is the gap to capacity, H^n is the channel transfer function in the nth carrier, P is the total power constraint, and N is the number of frequency carriers. The above problem has a well-known waterfilling solution. Efficient solution readily exists in this case, because the objective function is concave in the optimizing variable S^n.

The spectrum optimization problem becomes much more challenging when the PSDs of multiple users need to be optimized at the same time. The need for such a joint optimization is most clearly illustrated in the situation depicted in Fig. 1, where the channel transfer functions are heavily unbalanced. As shown in the figure, when an optical network unit (ONU) is deployed remotely, it may emit excessive interference to a neighboring customer-premise modem served from the central office. DSM enables the joint optimization of the transmit

PSDs by both the central office modem and the ONU modem, allowing both to operate at the same time.

Mathematically, the multiuser spectrum optimization problem may be formulated as follows:

$$\begin{aligned}\text{maximize} \quad & \sum_{k=1}^{K} \omega_k \sum_{n=1}^{N} \log\left(1 + \frac{S_k^n}{\sigma_k^n + \sum_{j\neq k} \alpha_{jk}^n S_j^n}\right) \\ \text{subject to} \quad & \sum_{n=1}^{N} S_k^n \leq P_k \quad \forall k \\ & S_k^n \geq 0 \quad \forall k, n\end{aligned} \tag{2}$$

where the effective noise PSD for the kth user at the nth carrier σ_k^n is again normalized by $\Gamma/|H_{kk}^n|^2$, and the effective interference coefficient α_{jk}^n is defined as $\Gamma|H_{jk}^n|^2/|H_{kk}^n|^2$. Here, K is the number of users, N is the number of frequency carriers, H_{jk}^n is the channel transfer function from the jth transmitter to the kth receiver in carrier n, ω_k is the relative weight given to the kth user in the optimization problem, S_k^n is the optimization variable denoting the power allocation for user k in the nth carrier, and P_k is the total power constraint for the user k. Throughout the paper, the sidelobe effect between adjacent carriers is neglected.[1] This is realistic for frame-synchronous DSL systems implementing a zipper-like modulation [20], or where a sufficient amount of transmit windowing is included. By solving the above optimization problem with varying ω_k, the entire achievable rate region can be generated. Because the objective function is not concave in S_k^n, numerical optimization is difficult. Clearly, an exhaustive search is not feasible, as the complexity would be exponential in the total number of variables, which is KN, where N can be as large as 4096.

Iterative waterfilling (IWF) [7] is one of the early multiuser spectrum optimization techniques that take advantage of the ability for DSL modems to perform spectral shaping. In this algorithm, each user iteratively maximizes its own achievable rate by performing a single-user waterfilling with the crosstalk interference from all other users treated as noise. However, the IWF process does not seek to find the global optimum for the entire DSL bundle. Instead, each user participates in a noncooperative game, and the convergence point of the IWF process corresponds to a competitive equilibrium. Although not optimum, the IWF algorithm has been shown to outperform SSM schemes.

Recently, an exact OSB algorithm to solve this problem globally and optimally was proposed in [12]. The basic strategy is to transform the spectrum optimization problem (2) into the dual domain by forming its Lagrangian dual

$$\begin{aligned}\text{maximize} \quad & \sum_{k=1}^{K} \omega_k \sum_{n=1}^{N} \log\left(1 + \frac{S_k^n}{\sigma^n + \sum_{j\neq k} \alpha_{jk}^n S_j^n}\right) \\ & + \sum_{k=1}^{K} \lambda_k \left(P_k - \sum_{n=1}^{N} S_k^n\right) \\ \text{subject to} \quad & S_k^n \geq 0 \quad \forall k.\end{aligned} \tag{3}$$

[1]When the desired DMT symbol and the interfering DMT symbols are not frame-synchronized, the frequency tones of the desired and interfering signals are not orthogonal to each other, creating sidelobes for each tone. In this case, S_j^n may interfere not only with S_k^n, but also with S_k^{n+1}, S_k^{n-1}, etc.

The idea is to solve the Lagrangian for each set of nonnegative and fixed $(\lambda_1, \ldots, \lambda_K)$. Then the solution to the original problem may be found by a nested bisection search in the λ-space. It can be shown that the OSB algorithm has a computational complexity that is linear in the number of frequency carriers N. As illustrated in [12], the OSB algorithm can provide a significant performance improvement, as compared with IWF.

However, the computational complexity of the OSB algorithm, although linear in N, is still exponential in the number of users K. This is so for two reasons. First, with K users, K nested loops of bisections are needed, one for each λ_k. Thus, the λ search is exponentially complex. Second, the maximization of the Lagrangian for a fixed set of $(\lambda_1, \ldots, \lambda_K)$ involves an exhaustive search over $(S_1^n, \ldots, S_K^n)$ in each tone n, which has a computational complexity that is also exponential in K. When the number of users is large, the complexity of OSB becomes prohibitive.

The purpose of this paper is to refine the OSB algorithm with an aim of eliminating its exponential complexity. Toward this end, we establish a general theory of dual optimization for multicarrier systems, and show that contrary to general nonconvex problems, the duality gap for multiuser spectrum optimization always tends to zero as the number of frequency carriers goes to infinity, regardless of whether the optimization problem is convex. This key observation leads to efficient λ search methods that optimize the dual objective function directly.

Second, to overcome the exponential complexity of an exhaustive search over $(S_1^n, \ldots, S_K^n)$, we propose iterative and low-complexity ways to evaluate the dual objective approximately. The resulting algorithm is a middle ground between IWF and OSB. We show by simulation that such an iterative spectrum balancing (ISB) technique achieves most of the gain of OSB in many cases of practical importance, while having a much lower computational complexity.

The computational methods proposed in this paper have a wider implication beyond that of DSL applications. The DSL spectrum balancing problem is very similar to the optimal power allocation and bit-loading problem for OFDM systems in wireless applications [3], [5], [21], [22]. A low-complexity solution to the DSL problem is likely to be applicable to wireless systems, as well.

III. Duality Gap of Nonconvex Optimization

In this section, we present a general duality theory for nonconvex optimization problems in multicarrier systems. In a multicarrier system, the optimization objective and the constraints typically consist of a large number of individual functions, each corresponding to one of the N frequency carriers. So, the optimization problem has the following general form:

$$\begin{aligned}\text{maximize} \quad & \sum_{n=1}^{N} f_n(\mathbf{x}_n) \\ \text{subject to} \quad & \sum_{n=1}^{N} \mathbf{h}_n(\mathbf{x}_n) \leq P\end{aligned} \tag{4}$$

where $\mathbf{x}_n \in \mathbb{R}^K$ are vectors of optimization variables, $f_n(\cdot)$ are $\mathbb{R}^K \rightarrow \mathbb{R}$ functions that are not necessarily concave, and $\mathbf{h}_n(\cdot)$ are $\mathbb{R}^K \rightarrow \mathbb{R}^L$ functions that are not necessarily convex. Power constraints are denoted by an L-vector $\mathbf{P}$. Here, "$\leq$" is used to denote a component-wise inequality. For the multiuser spectrum optimization considered before, $K = L, \mathbf{x}_n = (S_1^n, \ldots, S_K^n), f_n = \sum_k \omega_k \log(1+S_k^n/(\sigma^n+\sum_{j\neq k} S_j^n)), \mathbf{h}_n(\mathbf{x}_n) = [S_1^n \cdots S_K^n]^T$, and $\mathbf{P} = [P_1 \cdots P_K]^T$.

The idea of dual optimization is to solve (4) by forming its Lagrangian dual

$$L(\mathbf{x}_n, \lambda) = \sum_{n=1}^{N} f_n(\mathbf{x}_n) + \lambda^T \left(\mathbf{P} - \sum_{n=1}^{N} \mathbf{h}_n(\mathbf{x}_n)\right) \tag{5}$$

where λ is a vector of Lagrangian dual variables. Define the dual objective $g(\lambda)$ as an unconstrained maximization of the Lagrangian

$$g(\lambda) = \max_{\mathbf{x}_n} L(\mathbf{x}_n, \lambda). \tag{6}$$

The dual optimization problem is

$$\begin{aligned} \text{minimize} \quad & g(\lambda) \\ \text{subject to} \quad & \lambda \geq 0. \end{aligned} \tag{7}$$

When $f_n(\mathbf{x}_n)$'s are concave and $\mathbf{h}_n(\mathbf{x}_n)$'s are convex, standard convex optimization results guarantee that the primal problem (4) and the dual problem (7) have the same solution. When convexity does not hold, the dual problem provides a solution, which is an upper bound to the solution of (4). The upper bound is not always tight, and the difference between the upper bound and the true optimum is called the "duality gap."

The main objective of this section is to characterize a condition under which the duality gap is zero even when the optimization problem is not convex. Toward this end, we define the following *time-sharing* condition.

Definition 1: Let $\mathbf{x}_n^*$ and $\mathbf{y}_n^*$ be optimal solutions to the optimization problem (4) with $\mathbf{P} = \mathbf{P_x}$ and $\mathbf{P} = \mathbf{P_y}$, respectively. An optimization problem of the form (4) is said to satisfy the time-sharing condition if for any $\mathbf{P_x}, \mathbf{P_y}$ and for any $0 \leq \nu \leq 1$, there always exists a feasible solution $\mathbf{z}_n$, such that $\sum_{n=1}^N \mathbf{h}_n(\mathbf{z}_n) \leq \nu\mathbf{P_x} + (1-\nu)\mathbf{P_y}$, and $\sum_{n=1}^N f_n(\mathbf{z}_n) \geq \nu \sum_{n=1}^N f_n(\mathbf{x}_n^*) + (1-\nu)\sum_{n=1}^N f_n(\mathbf{y}_n^*)$.

The time-sharing condition has the following intuitive interpretation. Consider the maximum value of the optimization problem (4) as a function of the constraint $\mathbf{P}$. Clearly, a larger $\mathbf{P}$ implies a more relaxed constraint. So, roughly speaking, the maximum value is an increasing function of $\mathbf{P}$. The time-sharing condition implies that the maximum value of the optimization problem is a concave function of $\mathbf{P}$.

Note that if f_n's are concave and $\mathbf{h}_n$'s are convex, then the time-sharing condition is always satisfied. This can be easily verified by setting $\mathbf{z}_n = \nu\mathbf{x}_n^* + (1-\nu)\mathbf{y}_n^*$, in which case the concavity of f_n implies $\sum f_n(\mathbf{z}_n) \geq \nu \sum f_n(\mathbf{x}_n^*) + (1-\nu)\sum f_n(\mathbf{y}_n^*)$, and the convexity of $\mathbf{h}_n$ implies $\sum_n \mathbf{h}_n(\mathbf{z}_n) \leq \nu\mathbf{h}_n(\mathbf{x}_n^*) + (1-\nu)\mathbf{h}_n(\mathbf{y}_n^*) \leq \nu\mathbf{P_x} + (1-\nu)\mathbf{P_y}$. However, the converse is not necessarily true. As is shown later in the paper, for many multicarrier systems of practical interest, the time-sharing condition holds even when f_n's are not concave and $\mathbf{h}_n$'s are not convex.

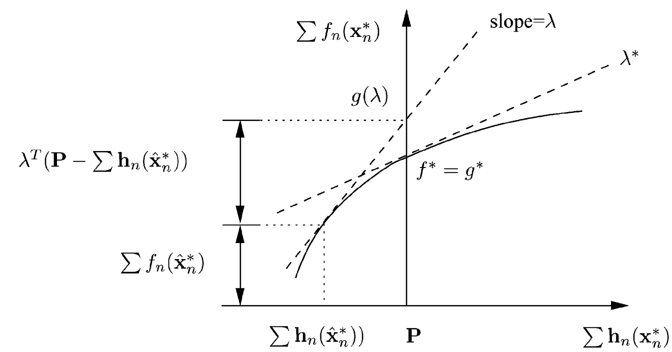


Fig. 2. Time-sharing property implies zero duality gap.

The main result of this section is that the time-sharing property implies zero duality gap. Further, for many practical optimization problems in the multicarrier context, the time-sharing condition is satisfied.

Theorem 1: Consider an optimization problem of the form (4). If the optimization problem satisfies the time-sharing property, then it has a zero duality gap, i.e., the primal problem (4) and the dual problem (7) have the same optimal value.

Proof: The theorem is a standard result in convex optimization if f_n's are concave and $\mathbf{h}_n$'s are convex functions. In this case, the optimization problem (4) is a convex programming problem, which has a zero duality gap under general constraint qualification conditions. The main novelty of the theorem resides in cases where (4) is not convex, but for which the time-sharing condition nevertheless holds. The proof is divided into two parts.

Let $\mathbf{P_x}, \mathbf{P_y}$ and $\mathbf{P_z}$ be vectors of power constraints with $\mathbf{P_z} = \nu\mathbf{P_x} + (1-\nu)\mathbf{P_y}$ for some $0 \leq \nu \leq 1$. Let $\mathbf{x}_n^*$, $\mathbf{y}_n^*$, and $\mathbf{z}_n^*$ be the optimal solutions to the optimization problem (4) with constraints $\mathbf{P_x}, \mathbf{P_y}$ and $\mathbf{P_z}$, respectively. In the first part of the proof, we show that time sharing implies that $\sum_n f_n(\mathbf{x}_n^*)$ is a concave function of $\mathbf{P_x}$. The concavity follows from the definition of the time-sharing property. Since $\mathbf{P_z} = \nu\mathbf{P_x} + (1-\nu)\mathbf{P_y}$, the time-sharing property implies that there exists a $\mathbf{z}_n$ such that $\sum_n \mathbf{h}_n(\mathbf{z}_n) \leq \nu\mathbf{P_x} + (1-\nu)\mathbf{P_y}$ and $\sum f_n(\mathbf{z}_n) \geq \nu \sum f_n(\mathbf{x}_n^*) + (1-\nu)\sum f_n(\mathbf{y}_n^*)$. Since $\mathbf{z}_n$ is a feasible solution for the optimization problem, this means that $\sum_n f_n(\mathbf{z}_n^*) \geq \sum_n f_n(\mathbf{z}_n) \geq \nu \sum f_n(\mathbf{x}_n^*) + (1-\nu)\sum f_n(\mathbf{y}_n^*)$, thus proving the concavity.

Second, we show that the concavity of the optimal $\sum_n f_n$ in $\mathbf{P}$ implies zero duality gap. Fig. 2 is a graphical illustration of the proof. Consider a sequence of the optimization problem parameterized by the constraint $\mathbf{P}$. The solid line in Fig. 2 is a plot of optimal $(\sum \mathbf{h}_n(\mathbf{x}_n^*), \sum f_n(\mathbf{x}_n^*))$ as the constraint varies. The curve is plotted with $\sum \mathbf{h}_n(\mathbf{x}_n^*)$ on the x-axis. The y-axis is located at the point where $\sum \mathbf{h}_n(\mathbf{x}_n^*) = \mathbf{P}$. Thus, the intersection of the curve with the y-axis is exactly the primal optimal solution to (4), which is denoted as f^* on the plot.

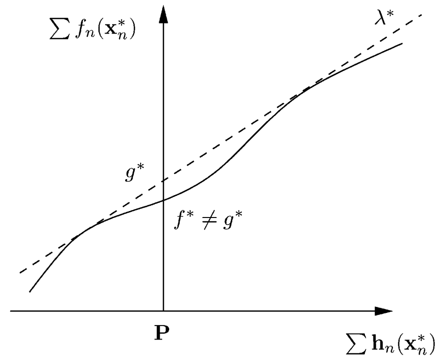


Fig. 3. Duality gap is nonzero without the time-sharing condition.

Now, consider the dual objective function $g(\lambda)$ for a fixed λ

$$g(\lambda) = \max_{\mathbf{x}_n} \left\{ \sum_n f_n(\mathbf{x}_n) + \lambda^T \left(\mathbf{P} - \sum_{n=1}^{N} \mathbf{h}_n(\mathbf{x}_n) \right) \right\}. \quad (8)$$

Let $\hat{\mathbf{x}}_n^*$ be the optimal solution to the above optimization problem. The value of $g(\lambda)$ can be graphically obtained by drawing a tangent line to the $(\sum \mathbf{h}_n(\mathbf{x}_n^*), \sum f_n(\mathbf{x}_n^*))$ curve through the point $(\sum \mathbf{h}_n(\hat{\mathbf{x}}_n^*), \sum f_n(\hat{\mathbf{x}}_n^*))$. By the definition of $g(\lambda)$, it is not difficult to see that the tangent line has a slope λ. Further, the intersection of the tangent line with the y-axis is $\sum_n f_n(\hat{\mathbf{x}}_n^*) + \lambda^T(\mathbf{P} - \sum_{n=1}^{N} \mathbf{h}_n(\hat{\mathbf{x}}_n^*))$, which is exactly the value of $g(\lambda)$, as illustrated in Fig. 2. This allows the minimization of $g(\lambda)$ to be visualized. The dual optimum, denoted as g^*, is the minimum $g(\lambda)$ over all nonnegative λ's. Clearly, when the optimal $(\sum \mathbf{h}_n(\mathbf{x}_n^*), \sum f_n(\mathbf{x}_n^*))$ curve is concave, among all tangent lines with various slopes λ, the λ^* that minimizes the y-axis intersection is precisely the one that intersects the y-axis at f^*. Thus, $f^* = g^*$ and the duality gap is zero. ■

To illustrate the importance of the time-sharing condition, Fig. 3 depicts a situation in which time sharing does not hold. In this case, the curve $(\sum \mathbf{h}_n(\mathbf{x}_n^*), \sum f_n(\mathbf{x}_n^*))$ is not concave, and the minimum $g(\lambda)$ is strictly larger than the maximum $\sum f_n(\mathbf{x}_n)$.

It turns out that the time-sharing condition is always satisfied for practical spectrum optimization problems in multicarrier applications in the limit, as the number of carriers N goes to infinity. The intuitive reason is as follows. The time-sharing condition is clearly satisfied if an actual time-division multiplexing may be implemented. Let $\mathbf{x}_n$ and $\mathbf{y}_n$ be two spectrum allocations. In this case, the entire frequency band can be assigned to $\mathbf{x}_n$ for ν percentage of the time, and $\mathbf{y}_n$ for $(1-\nu)$ percentage of the time. The overall $\sum_n f_n$ then becomes a linear combination $\sum_n \nu f_n(\mathbf{x}_n) + (1-\nu) f_n(\mathbf{y}_n)$. The constraint becomes a linear combination also, thus satisfying the time-sharing condition.

In practical multicarrier systems with a large number of frequency carriers, channel conditions in adjacent carriers are often similar. Then, time sharing may be approximately implemented with frequency sharing. By interleaving $\mathbf{x}_n$ and $\mathbf{y}_n$ in the frequency domain with a proportionality ν, the overall $\sum_n f_n$ becomes approximately $\sum_n \nu f_n(\mathbf{x}_n) + (1-\nu) f_n(\mathbf{y}_n)$, and the same applies to the constraints. The approximation is exact when $N \to \infty$.

To make the intuition precise, consider the spectrum optimization problem (2) with continuous frequency variables

$$\begin{aligned} \text{maximize} \quad & \sum_{k=1}^{K} \omega_k \\ & \cdot \int \log \left(1 + \frac{S_k(f)}{\sigma_k(f) + \sum_{j \neq k} \alpha_{jk}(f) S_j(f)} \right) df \\ \text{subject to} \quad & \int S_k(f) \leq P_k \quad \forall k \\ & S_k(f) \geq 0 \quad \forall k, f \end{aligned} \quad (9)$$

where the effective noise PSD $\sigma_k(f)$ is again normalized by $\Gamma/|H_{kk}(f)|^2$, and the effective interference coefficient $\alpha_{jk}(f)$ is defined as $\Gamma|H_{jk}(f)|^2/|H_{kk}(f)|^2$. Again, $H_{jk}(f)$ is the channel transfer function from the jth transmitter to the kth receiver.

Theorem 2: Consider an optimization problem (9) in which $\sigma_k(f)$ and $\alpha_{jk}(f)$ are both continuous functions of f. Then, the time-sharing condition is always satisfied. In addition, its discretized version (2) also satisfies the time-sharing condition in the limit as $N \to \infty$.

Proof: Let $S_{k,1}^*(f)$ and $S_{k,2}^*(f)$ be the optimal solutions to the spectrum optimization problem (9) with power constraints $\mathbf{P}_1 = [P_{1,1} \cdots P_{K,1}]^T$ and $\mathbf{P}_2 = [P_{1,2} \cdots P_{K,2}]^T$, respectively. Let R_1^* and R_2^* be their respective optimal values. To prove the time-sharing property, we need to construct an $S_k(f)$ such that it satisfies a power constraint $\nu\mathbf{P}_1 + (1-\nu)\mathbf{P}_2$ and achieves a rate equal to or higher than $\nu R_1^* + (1-\nu) R_2^*$ for all ν between zero and one.

We first prove the theorem for the case in which $\sigma_k(f)$ and $\alpha_{jk}(f)$ are constant functions of f for all j and k. First, observe that the optimal solution to (9) is always a frequency-division multiplex (FDM) of at most 2^K frequency bands, where each frequency band corresponds to a transmission strategy for which a subset of K users transmit. Further, within each frequency band, the optimal $S_1^*(f)$ and $S_2^*(f)$ must be constant. This is because within each band, the same Karush–Kuhn–Tucker (KKT) condition (which is a necessary condition even for nonconvex problems) must be satisfied for each frequency f, and the optimal set of spectra is the KKT point that maximizes the weighted sum rate. As the same condition applies to all frequencies, the power allocation within each band must be flat.

Now, let $(S_{1,x}^*(f), \ldots, S_{K,x}^*(f))$ be the optimal solution of (9) with a power constraint $\mathbf{P}_x$, and $(S_{1,y}^*(f), \ldots, S_{K,y}^*(f))$ be the optimal solution of (9) with a power constraint $\mathbf{P}_y$. Let the achievable rate in the two cases be R_x^* and R_y^*, respectively. To prove the time-sharing property, we need to construct an $(S_{1,z}(f), \ldots, S_{K,z}(f))$ that achieves at least $\nu R_x^* + (1-\nu) R_y^*$, with a power that is at most $\nu\mathbf{P}_x + (1-\nu)\mathbf{P}_y$ for all ν between zero and one. Such a $S_{k,z}(f)$ may be constructed by taking the union of the two frequency partitions corresponding to $\mathbf{P}_x$ and $\mathbf{P}_y$, then further subdividing each frequency band in the union into two, ν proportion of which has $S_{k,z}(f) = S_{k,x}^*(f)$, and $(1-\nu)$ proportion of which has $S_{k,z}(f) = S_{k,y}^*(f)$. Clearly, the resulting $S_{k,z}(f)$ satisfies the power constraint $\int S_{k,z}(f) \leq \nu P_{k,x} + (1-\nu) P_{k,y}$. Further, it also achieves a rate $\nu R_x^* +$

$(1-\nu)R_y^*$. Therefore, for an optimization problem with constant channel gain and noise power spectrum, the time-sharing property holds.

To show that the time-sharing property holds for all optimization problems with continuous channel gains and noise spectra, a limiting argument is needed. The idea is to divide the total frequency into a set of infinitesimal frequency bands. By continuity, the channel gains and noise spectra within each band approaches a constant value as the subdivision becomes finer and finer. Then, the constant channel result proved earlier applies in each infinitesimal band. Therefore, the time-sharing property applies to the entire optimization problem.

Finally, the above argument also shows that in a discretized version of the optimization problem (4), as $N \to \infty$, the time-sharing property holds in the limit. ■

For the ease of presentation, *Theorem 2* has been stated and proved for the continuous bit-loading problem. However, the theorem continues to hold even when an additional integer bit constraint is imposed. The steps of the proof follow exactly the same way. First, the theorem can be shown to hold when $\sigma_k(f)$ and $\alpha_{jk}(f)$ are constants. With integer bit loading, there are clearly only a finite number of frequency bands in which optimal $S_1^*(f)$ and $S_2^*(f)$ are constants. Then, FDM of the two gives the desirable time-sharing points.

Note that for practical systems with a large but finite N, although the duality gap is not strictly zero, it is nearly so. This is because the carrier width in multicarrier systems is always chosen so that the channel response in adjacent subchannels are approximately the same. In this case, $f_n, \ldots, f_{n+m}$ are sufficiently similar (and likewise for $\mathbf{h}_n, \ldots, \mathbf{h}_{n+m}$) so that time-sharing may be implemented via FDM.

The main consequence of *Theorems 1* and *2* is that as long as N is sufficiently large, even nonconvex spectrum optimization problems can be solved by solving its dual. For optimization problems of the form (4), solving the dual problem can be much easier. This is so because of the following two observations. First, the dual objective function $g(\lambda)$ decouples into N independent problems

$$g(\lambda) = \sum_{n=1}^{N} \left\{ \max_{\mathbf{x}_n} \left\{ f_n(\mathbf{x}_n) - \lambda^T \mathbf{h}_n(\mathbf{x}_n) \right\} \right\} + \lambda^T \mathbf{P}. \quad (10)$$

Thus, the evaluation of $g(\lambda)$ has a complexity which is linear in N. Note that the per-carrier optimization problem does not have a convexity structure. So, solving the per-carrier problem globally may still involve an exhaustive search. However, as the optimization problem is unconstrained, it is more manageable.

Second, the function $g(\lambda)$ is convex even if $f_n(\mathbf{x}_n)$ is not concave and $\mathbf{h}_n(\mathbf{x}_n)$ is not convex. (This is because $L(\mathbf{x}_n, \lambda)$ is linear in λ for each fixed x_n, and $g(\lambda)$ is the maximum of linear functions, and is, therefore, convex.) The complexity of optimizing $g(\lambda)$ depends on the dimension of λ, which is the number of constraints in the original problem. For the spectrum optimization problem, the number of constraints is K, which is independent of N.

Putting these two facts together, the entire dual optimization process has an $O(N)$ complexity. As N can be large in multicarrier spectrum optimization problems, the reduction from an exponential complexity to a linear complexity in N is significant.

We note here that *Theorems 1* and *2* guarantee that the optimal value of the original optimization problem is exactly the minimal value of $g(\lambda)$ over $\lambda \geq 0$. However, in some cases, extra care must be taken when recovering the optimal primal solution of the original optimization problem $\mathbf{x}_n^*$ from the optimal dual solution λ^*. In particular, there are cases in which the optimal $\mathbf{x}_n^*$ that solves the maximization problem (10) is not unique. In this case, a set of $\mathbf{x}_n^*$ that satisfies the constraints of the original problem must be chosen. Such a set of feasible $\mathbf{x}_n^*$ always exists when the time-sharing condition is satisfied. For the spectrum optimization problem, this corresponds to the case in which many equivalent FDM solutions exist.

As mentioned earlier, our treatment of duality gap for the nonconvex optimization problem is inspired by Aubin and Ekeland [13] and Bertsekas *et al.* [14], [15], who derived estimates of the duality gap for nonconvex integer programming problems. Although the main ideas of the proofs are similar, the problem setup of [13]–[15] is somewhat different, and the earlier results are not directly applicable to this setting.

IV. Dual Update Methods

The OSB algorithm developed in [12] is one of the first dual optimization algorithms for nonconvex spectrum optimization problems. In [12], the implementation of OSB is based on a bisection search in each component of λ. It is proved in [12] that the optimal power allocation in a multicarrier system is a continuous and monotonic function of λ. Thus, bisection search is guaranteed to converge to the optimum. However, as the multiuser spectrum optimization problem (2) consists of K constraints, successive bisection on each component of λ has a complexity that is exponential in K. One of the main motivations for developing a general duality theory for nonconvex problems, as presented in the previous section, is that such a general result allows a direct optimization of $g(\lambda)$. This gives rise to efficient dual-update methods that have a polynomial complexity in K.

The main idea is to minimize $g(\lambda)$ directly by updating all components of λ at the same time along some search direction (instead of successively updating one component at a time). Because $g(\lambda)$ is convex, a gradient-type search is guaranteed to converge to the global optimum. However, the main difficulty is that $g(\lambda)$, although convex, is not necessarily differentiable. Thus, it does not always have a gradient. Nevertheless, it is possible to find a search direction based on what is known as a subgradient. A vector $\mathbf{d}$ is a subgradient of $g(\lambda)$ at λ, if for all λ'

$$g(\lambda') \geq g(\lambda) + \mathbf{d}^T(\lambda' - \lambda). \quad (11)$$

Subgradient is a generalization of gradient for (possibly) nondifferentiable functions. Intuitively, d is a subgradient if the linear function with slope $\mathbf{d}$ passing through $(\lambda, g(\lambda))$ lies entirely below $g(\lambda)$. Fortunately, for the $g(\lambda)$ defined in (6), a subgradient is easy to find.

Proposition 1: For the optimization problem (2) with a dual objective $g(\lambda)$ defined in (6), the following choice of $\mathbf{d}$ is a subgradient for $g(\lambda)$:

$$\mathbf{d} = \mathbf{P} - \sum_{n=1}^{N} \mathbf{h}_n(\mathbf{x}_n^*) \tag{12}$$

where $\mathbf{x}_n^*$ is the optimizing variable in the maximization problem in the definition of $g(\lambda)$.

Proof: By definition, $g(\lambda) = \max_{\mathbf{x}_n}\{\sum_n f_n(\mathbf{x}_n) + \lambda^T(\mathbf{P} - \sum_{n=1}^{N}\mathbf{h}_n(\mathbf{x}_n))\}$. Let $\mathbf{x}_n^*$ be the optimizing variable in the definition of $g(\lambda)$. Then

$$\begin{aligned} g(\lambda') &= \max_{\mathbf{x}_n}\left\{\sum_n f_n(\mathbf{x}_n) + \lambda'^T\left(\mathbf{P} - \sum_{n=1}^{N}\mathbf{h}_n(\mathbf{x}_n)\right)\right\} \\ &\geq \sum_n f_n(\mathbf{x}_n^*) + \lambda'^T\left(\mathbf{P} - \sum_{n=1}^{N}\mathbf{h}_n(\mathbf{x}_n^*)\right) \\ &= g(\lambda) + \left(\mathbf{P} - \sum_{n=1}^{N}\mathbf{h}_n(\mathbf{x}_n^*)\right)^T(\lambda' - \lambda) \end{aligned}$$

thus verifying the definition of subgradient (11). ■

The subgradient search direction suggests that the kth component of λ should increase if the corresponding constraint is exceeded, i.e., the kth component of $\sum_{n=1}^{N}\mathbf{h}_n(\mathbf{x}_n^*)$ exceeds P_k, and decrease otherwise. This is intuitive, as λ represents a price for power. Price should increase if the power constraint is exceeded. Price should decrease, otherwise. In fact, λ updates can be done systematically. In the following, we propose two λ update methods based on the well-known subgradient and ellipsoid methods. These methods, for example, were used recently in [23] for a joint routing and resource allocation problem.

A. Subgradient Method

The idea of the subgradient update method is to design a step-size sequence s^l to update λ in the subgradient direction. More specifically, the update may be performed as follows:

$$\lambda^{l+1} = \left[\lambda^l - s^l\left(\mathbf{P} - \sum_n \mathbf{h}_n(\mathbf{x}_n)\right)\right]^+ \tag{13}$$

where l is the iteration number, s^l is a sequence of scalar step sizes, and $[\cdot]^+$ is defined as $[\cdot]^+ = \max(\cdot, 0)$. The above subgradient update is guaranteed to converge to the optimal λ as long as s^l is chosen to be sufficiently small [24]. A common criterion for choosing the step sizes is that s^l must be square summable, but not absolute summable [25], [26]. When the norm of the subgradient is bounded, the following choice:

$$s^l = \frac{\beta}{l} \tag{14}$$

for some constant β is guaranteed to converge to the optimal $g(\lambda)$. Other update rules include $s^l = \beta$ and $s^l = (\beta)/(\sqrt{l})$.

B. Ellipsoid Method

The update of the dual variables may also be done using the so-called cutting-plane methods. The idea is to localize the set of candidate λ's within some closed and bounded set. Then, by evaluating the subgradient of $g(\lambda)$ at an appropriately chosen center of such a region, roughly half of the region may be eliminated from the candidate set. The iterations continue as the size of the candidate set diminishes until it converges to an optimal λ. More precisely, the definition of subgradient (11) guarantees that $g(\lambda') \geq g(\lambda) + \mathbf{d}^T(\lambda' - \lambda)$. Thus, for all λ' that satisfies

$$\mathbf{d}^T(\lambda' - \lambda) \geq 0 \tag{15}$$

we must have $g(\lambda') \geq g(\lambda)$. So, all λ''s that are in the half-plane defined by (15) can be eliminated in each step. The cutting-plane method is a generalization of the one-dimensional (1-D) bisection method to higher dimensions.

A common choice of the candidate region is the minimal-sized ellipsoid containing all candidate λ's. An ellipsoid with a center $\mathbf{z}$ and a shape defined by positive semidefinite matrix $\mathbf{A}$ is defined to be

$$E(\mathbf{A}, \mathbf{z}) \triangleq \{\mathbf{x} | (\mathbf{x} - \mathbf{z})^T\mathbf{A}(\mathbf{x} - \mathbf{z}) \leq 1\}. \tag{16}$$

Let $\mathbf{d_i}$ be the subgradient of $g(\lambda)$ at the center of the ellipsoid $\mathbf{z_i}$. In each iteration, half of the ellipsoid is eliminated based on $\mathbf{d_i}$. A new ellipsoid, which is the minimal-volume ellipsoid containing the other half, is formed. Mathematically, the update algorithm is as follows [26]:

$$1)\quad \tilde{\mathbf{d}}_\mathbf{i} = \frac{\mathbf{d_i}}{\sqrt{\mathbf{d_i}^T\mathbf{A_i}^{-1}\mathbf{d_i}}} \tag{17}$$

$$2)\quad \mathbf{z_{i+1}} = \mathbf{z_i} - \frac{1}{K+1}\mathbf{A_i}^{-1}\tilde{\mathbf{d}}_\mathbf{i} \tag{18}$$

$$3)\quad \mathbf{A_{i+1}}^{-1} = \frac{K^2}{K^2 - 1} \times \left(\mathbf{A_i}^{-1} - \frac{2}{K+1}\mathbf{A_i}^{-1}\tilde{\mathbf{d}}_\mathbf{i}\tilde{\mathbf{d}}_\mathbf{i}^T\mathbf{A_i}^{-1}\right) \tag{19}$$

where K is the dimension of λ, i.e., the number of users in the problem. Fig. 4 illustrates the update process.

To choose an initial ellipsoid, we need to bound all candidate λ's in a closed and bounded set. The following result gives a suitable choice of the initial set.

Proposition 2: For the K-user spectrum optimization problem (2), the optimal set of dual variables λ^* must satisfy

$$0 \leq \lambda_k^*/\omega_k \leq \lambda_k^{\text{single}} \quad \forall k \tag{20}$$

where ω_k is the relative weight for the user k, and $\lambda_k^{\text{single}}$ is the optimal dual variable in the single-user spectrum optimization problem (1) with $\sigma^n = \sigma_k^n$ and $P = P_k$.

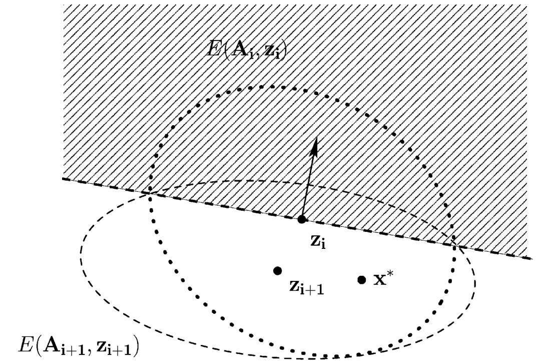


Fig. 4. Graphical illustration of ellipsoid update.

Proof: To be the optimal λ_k^* for the problem (2), λ_k^* must satisfy a set of KKT conditions. Differentiating the Lagrangian of (2) with respect to S_k^n, we obtain

$$\omega_k \cdot \frac{1}{\ln(2)} \cdot \frac{1}{\sigma_k^n + S_k^n + \sum_{j \neq k} \alpha_{jk}^n S_j^n} + \text{negative terms} = \lambda_k^*. \tag{21}$$

Intuitively, only user k's rate is an increasing function of S_k^n, so only one term in the derivative is positive. Now, in the single-user problem, $\lambda_k^{\text{single}}$ may be computed directly as

$$\frac{1}{\ln(2)} \cdot \frac{1}{\sigma_k^n + S_k^n} = \lambda_k^{\text{single}}. \tag{22}$$

From the above two equations, it follows that

$$\begin{aligned} \lambda_k^* &\le \omega_k \cdot \frac{1}{\ln(2)} \cdot \frac{1}{\sigma_k^n + S_k^n + \sum_{j \neq k} \alpha_{jk}^n S_j^n} \\ &\le \omega_k \frac{1}{\ln(2)} \cdot \frac{1}{\sigma_k^n + S_k^n} = \omega_k \lambda_k^{\text{single}}. \end{aligned} \tag{23}$$

■

Using the above result, the following initial ellipsoid may be chosen to enclose a rectangular region in which the optimal λ must reside:

$$\mathbf{A}_0^{-1} = \begin{bmatrix} K\left(\frac{\omega_1 \lambda_1^{\text{single}}}{2}\right)^2 & 0 & \cdots & 0 \\ 0 & K\left(\frac{\omega_2 \lambda_2^{\text{single}}}{2}\right)^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & K\left(\frac{\omega_K \lambda_K^{\text{single}}}{2}\right)^2 \end{bmatrix}$$

$$\mathbf{z}_0 = \left[\frac{\omega_1 \lambda_1^{\text{single}}}{2} \quad \cdots \quad \frac{\omega_K \lambda_K^{\text{single}}}{2}\right]^T. \tag{24}$$

The ellipsoid updates (17)–(19) can then be carried out from this starting point.[2] When the subgradient of the dual objective function $g(\lambda)$ has a bounded norm [26], the ellipsoid update is guaranteed to converge to the optimal λ. The bounded norm condition is satisfied for this problem, because λ is constrained to be in a closed and bounded set.

Fig. 5 compares the convergence behaviors of the subgradient update and the ellipsoid update. As seen in the figure, the convergence speed of subgradient methods depends heavily on the choice of step size, while the ellipsoid method appears to converge faster and is more stable across a wide variety of situations. However, the subgradient algorithm is more suitable for distributed implementation, where each user may update its own dual variable autonomously. This is not possible with the ellipsoid method. We also note that the computational costs per iteration for the two methods are similar.

[2]If at any given point in the iteration, the center of the ellipsoid moves out of the feasibility region, i.e., some components of λ become negative, they can be simply set to zero.

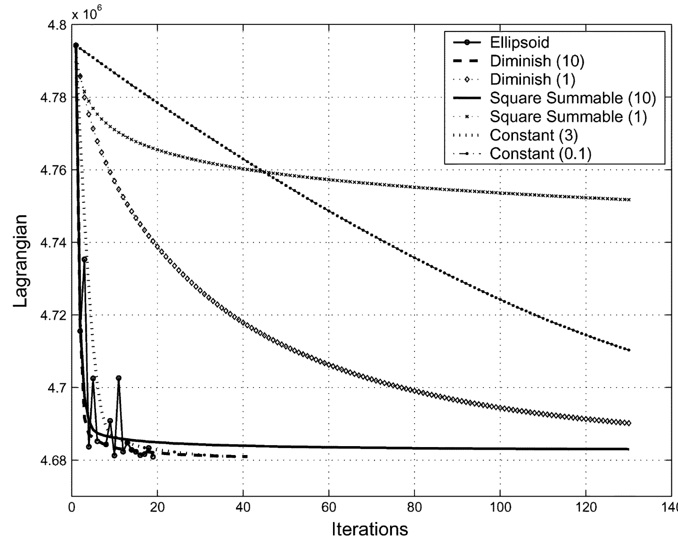


Fig. 5. Comparison between the ellipsoid method and the subgradient method. Various step-size sequences are used for the subgradient method. Constant (β) refers to the update rule $s^l = \beta$. Square Summable (β) refers to the update rule $s^l = \beta/l$. Diminish (β) refers to the update rule $s^l = \beta/\sqrt{l}$.

TABLE I
THE ISB ALGORITHM

Algorithm: Iterative Spectrum Balancing
initialize $(\lambda_1, \cdots, \lambda_K)$.
repeat
initialize $(S_1^n, \cdots, S_K^n)$.
repeat
for $k = 1$ to K,
set $S_k^n = \text{argmax}_{S_k^n} \sum_{l=1}^K (w_l b_l^n - \lambda_l S_l^n)$.
end
until $(S_1^n, \cdots, S_K^n)$ converges
update $(\lambda_1, \cdots, \lambda_K)$ using ellipsoid or subgradient method.
until $(\lambda_1, \cdots, \lambda_K)$ converges

TABLE II
COMPUTATIONAL COMPLEXITY COMPARISON

Algorithm	**Computational Complexity**
Exhaustive Search	$O(B^{NK})$
Optimal Spectrum Balancing	$O(T_2 N B^K)$
Iterative Spectrum Balancing	$O(T_1 T_2 B N K)$
Iterative Water-Filling	$O(T_3 K N \log(N))$

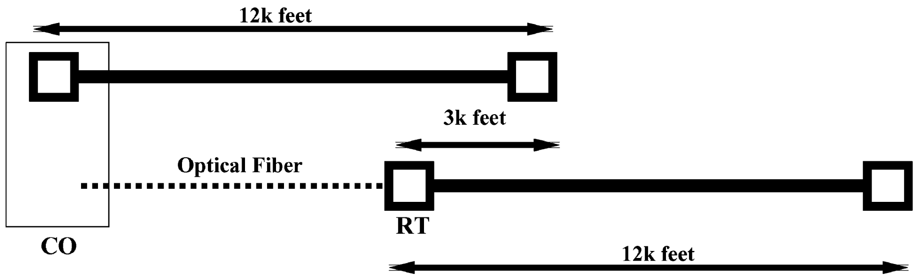


Fig. 6. Loop topology for two downstream ADSL users. The OSB and ISB algorithms are compared with the IWF and SSM approaches.

V. Iterative Spectrum Balancing

The results of the previous section show that the exponential complexity of the λ search can be avoided by using subgradient or ellipsoid updates. However, the complexity of evaluating $g(\lambda)$, although linear in N, is still exponential in K, as it involves solving N nonconvex optimization problems, corresponding to the N tones, each with K variables. Nevertheless, for practical problems, suboptimal low-complexity methods often exist. In this section, we propose an iterative method that eliminates the exponential complexity in evaluating $g(\lambda)$.

We begin the discussion by reviewing the approach taken in OSB [12]. Our prior discussion has focused on the spectrum optimization problem (2) with continuous bit loading. In practice, the bit allocation on each frequency carrier must be integer-valued. With discrete bit loading, the solution to the per-tone nonconvex optimization problem simplifies into an exhaustive search. More specifically, for a spectrum optimization problem (2) with discrete bit constraints, define

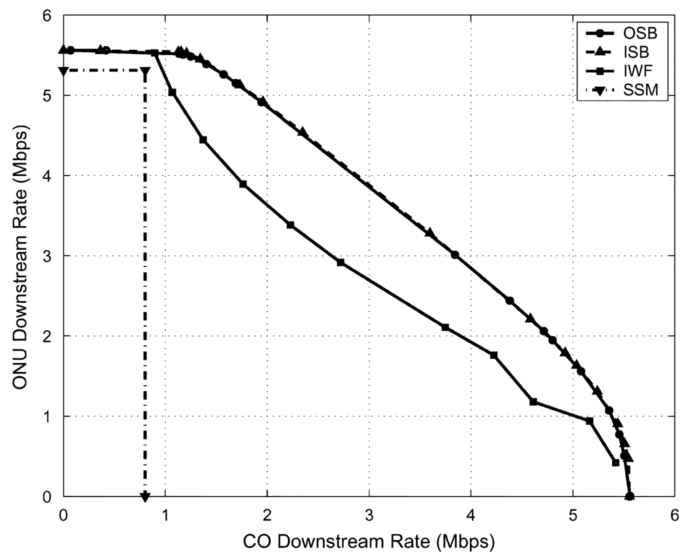


Fig. 7. Rate region for two downstream ADSL users.

$$b_k^n = \left\lfloor \log_2 \left(1 + \frac{S_k^n}{\sigma_k^n + \sum_{i \neq k} \alpha_{i,k}^n S_i^n} \right) \right\rfloor . \quad (25)$$

The evaluation of $g(\lambda_1, \ldots, \lambda_K)$ is exactly that of exhaustively searching through the set of discrete b_k^n's

$$\begin{aligned} g(\lambda_1, \ldots, \lambda_K) &= \max_{\{S_1^n, \ldots, S_K^n\}_{n=1}^N} \sum_{k=1}^K \left(w_k \sum_{n=1}^N b_k^n \right) - \sum_{k=1}^K \lambda_k \left(\sum_{n=1}^N S_k^n - P_k \right) \\ &= \left(\sum_{n=1}^N \max_{S_1^n, \ldots, S_K^n} \sum_{k=1}^K (w_k b_k^n - \lambda_k S_k^n) \right) + \sum_{k=1}^K \lambda_k P_k. \quad (26) \end{aligned}$$

The strategy of [12] is to map each discrete $(b_1^n, b_2^n, \ldots, b_K^n)$ into a set of $(S_1^n, S_2^n, \ldots, S_K^n)$ via (25), then to select the $\{b_k^n\}$ that maximizes

$$\tilde{g}(S_1^n, \ldots, S_K^n) \triangleq \sum_{k=1}^K (w_k b_k^n - \lambda_k S_k^n) . \quad (27)$$

Clearly, such an exhaustive search has a computational complexity $O(B^K)$ in each frequency carrier, where B is the maximum allowable bit cap.

The main contribution of this section is an efficient algorithm that approximately evaluates $g(\lambda_1, \ldots, \lambda_K)$. The main idea is to locally optimize $\tilde{g}(S_1^n, \ldots, S_K^n)$ via coordinate descent. For each fixed set of $(\lambda_1, \ldots, \lambda_K)$, our proposed approach first finds the optimal S_1^n while keeping $S_2^n, \ldots, S_K^n$ fixed, then finds the optimal S_2^n keeping all other S_k^n fixed, then $S_3^n, \ldots, S_K^n$, then $S_1^n, S_2^n, \ldots$, and so on. Note that when optimizing each S_k^n, only a small finite number of power levels (corresponding to a finite number of integer bits) need to be searched over. Further, such an iterative process is guaranteed to converge, because each iteration strictly increases the objective function. The convergence point must have integer bit values for all users, and it is guaranteed to be at least a local maximum for $\tilde{g}(S_1^n, \ldots, S_K^n)$.

This new approach bears some resemblance to the IWF algorithm [7]. However, it differs from IWF in the following two key aspects. First, unlike the IWF algorithm, where each user maximizes its own rate in each step of the iteration, the above algorithm optimizes an objective function that includes the joint rates of all users. Thus, the new algorithm has the potential to reach a joint optimum. Second, the power constraint in the IWF process is handled on an *ad hoc* basis, while the new algorithm proposed in this paper dualizes the power constraint in an optimal fashion. The correct values of the dual variables are then used in a subgradient or ellipsoid search. We call this approach ISB. Table I summarizes the algorithm.

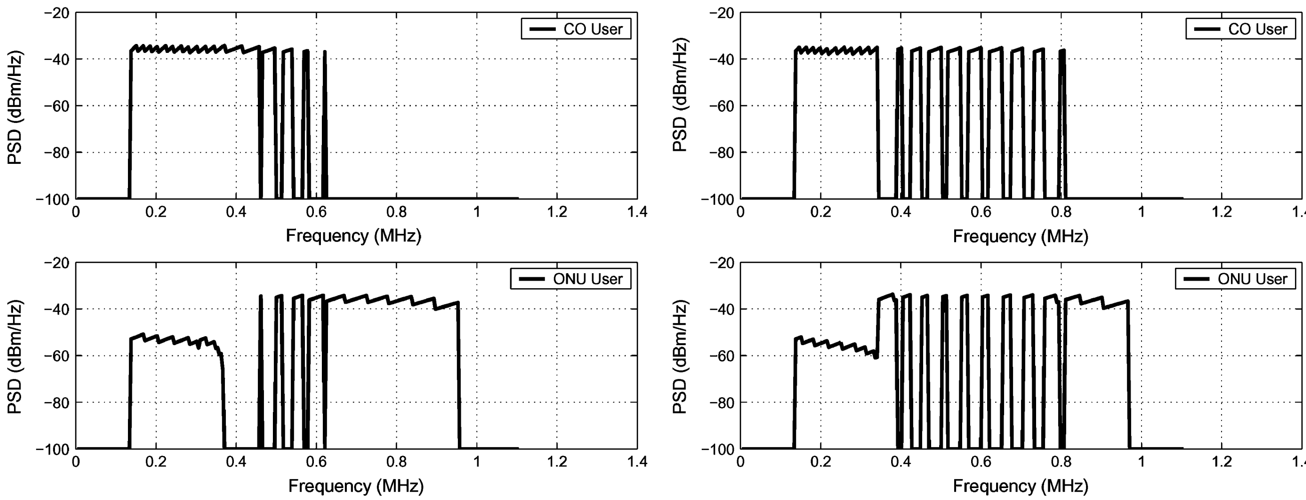


Fig. 8. PSDs of OSB (left) and ISB (right) for the two distributed ADSL users at equal rate. PSDs for both CO-based (top) and RT-based (bottom) lines are plotted.

It should be noted that the ISB algorithm is a suboptimal algorithm. In particular, the local optimum depends on the initial starting point and the ordering of iterations. Further, with an approximate evaluation of $g(\lambda)$, the subgradient property (i.e., *Proposition 1*) can no longer be guaranteed, and the proof of convergence becomes an issue. Nonetheless, the ISB algorithm has been observed to converge in all simulation settings that we have tried. In addition, as the simulation results in the next section show, its performance can be near-optimal in many practical situations. Further, the ISB algorithm may be implemented autonomously in a distributed environment, provided that suitable bit-allocation information is shared by the neighboring lines.

The computational complexity of this new iterative approach is significantly lower than that of the OSB algorithm proposed in [12]. In the evaluation of $\tilde{g}(S_1^n, \ldots, S_k^n)$, each iteration has a computational complexity that is linear in K. Let T_1 be the number of iterations needed in the evaluation of each $\tilde{g}(S_1^n, \ldots, S_k^n)$. Let T_2 be the number of subgradient or ellipsoid updates needed. The total computational complexity of ISB is $O(T_1 T_2 BNK)$. Computational experience suggests both T_1 and T_2 are polynomial functions of K. This is significant, as K is usually large in realistic DSL deployment scenarios. Table II summarizes the computational complexity comparison. Here, T_3 is the number of iterations needed in IWF. In actual implementation, T_3 is comparable with T_1. Both are relatively small. Fig. 5 shows the values of T_2 for different dual-update methods.

Finally, we note that a slightly different ISB algorithm has been independently and simultaneously proposed by Cendrillon and Moonen [27]. As compared with the algorithm proposed in this paper, the iteration and the dual updates are interchanged in this alternative approach.

VI. Simulations

In this section, we present an extensive set of simulation results to evaluate the performance of the proposed low-complexity ISB algorithm for the DSL spectrum optimization problem. In the simulation, all DSL lines are 26-gauge twisted pairs with a background noise level of -140 dBm/Hz. Users

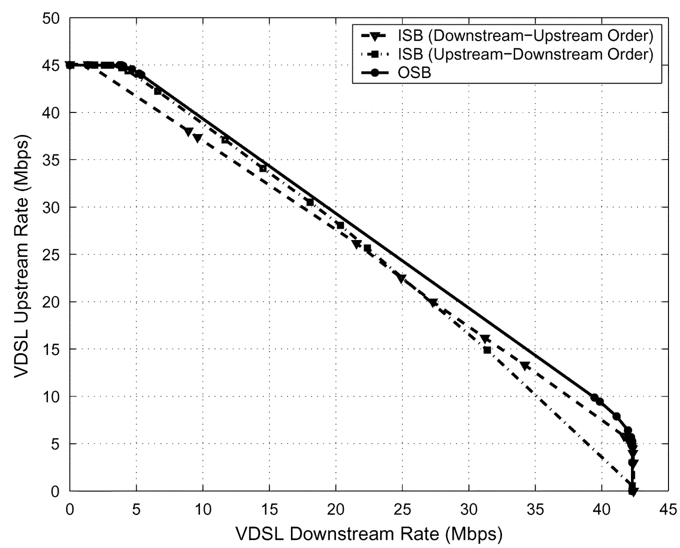


Fig. 9. Rate region for five-user full duplex VDSL.

are assumed to be symbol-synchronized so that the sidelobe interference is not in effect. No spectral masks are enforced.

A. Two-User ADSL Downstream

The first set of simulations examines a two-user asymmetric DSL (ADSL) downstream distributed environment with both users having a loop length of 12 k ft and with a crosstalk distance of 3 k ft. No other crosstalk sources are assumed to exist in the binder. The loop topology is shown in Fig. 6. Such a distributed environment is expected to benefit significantly from DSM because of its highly unbalanced crosstalk channels. The power constraint for each user is set to 20.4 dBm, as defined in [28].

Fig. 7 shows the achievable rate regions of OSB, ISB, IWF, and SSM algorithms. As can be seen in the figure, the rate regions for OSB and ISB are almost identical to each other. Both outperform IWF significantly. Interestingly, although the achievable rates of OSB and ISB are identical, the optimal spectra obtained from the two algorithms can be different. Fig. 8 shows the downstream spectra obtained from the two algorithms. The main difference between the spectra of OSB

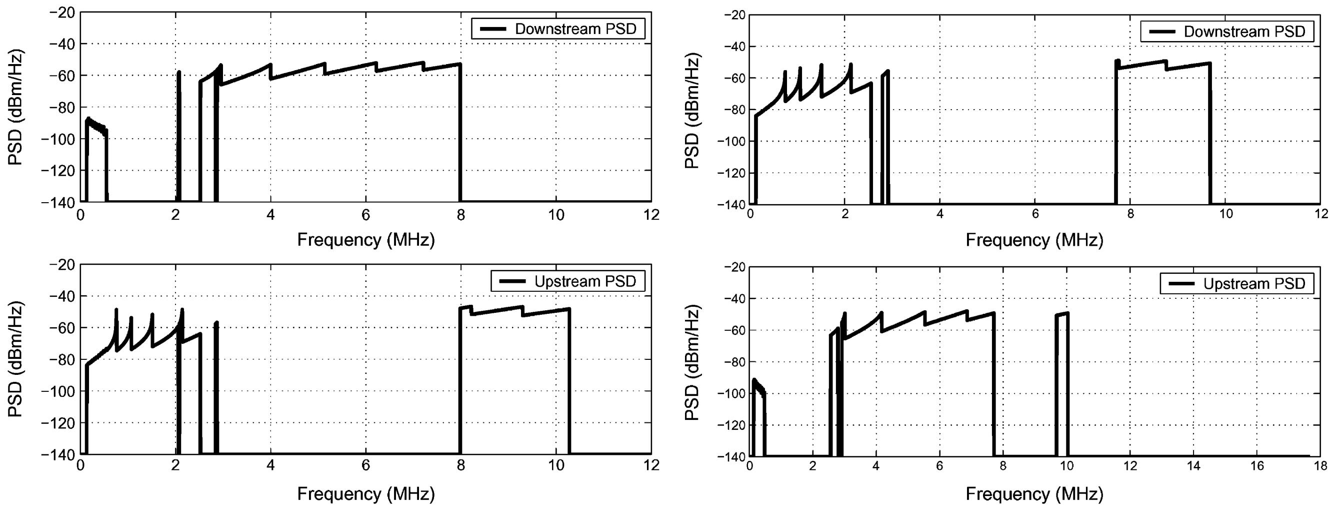


Fig. 10. Downstream (top) and upstream (bottom) PSDs for five-user full duplex VDSL at equal rate. The power spectra depend on the ordering in iteration in ISB. Downstream–upstream order is on the left, and the upstream–downstream order is on the right.

and ISB is in the FDM region (frequency beyond 380 kHz). Both PSDs essentially achieve the same rates because there are many equivalent permutations of FDM possible. We note that in the simulation for ISB, the iteration order of CO-modem followed by RT-modem is found to have a superior performance, as compared with the RT-CO order. This is because the CO-modem is more susceptible to crosstalk interference in this configuration. Therefore, it is advantageous to let the CO-modem occupy the best spectrum.

B. Five-User VDSL Full Duplex

The current very-high-speed digital subscriber line (VDSL) standard [29] uses a fixed frequency bandplan (the so-called 998 bandplan [30]) to separate upstream and downstream. This is not optimal, because no overlapping of upstream and downstream transmissions is allowed. In this set of simulations, we explore the achievable rate region and the optimal power allocations with overlap spectra for full duplex transmission in a VDSL environment. The simulation setup consists of five users with the same loop length (3-k-ft long) in the same binder. As the loop characteristics for the five users are identical, this is essentially a two-user scenario between upstream and downstream. Perfect echo cancellation is assumed. The near-end crosstalk (NEXT) is modeled in addition to the far-end crosstalk (FEXT). The downstream transmission has a power constraint of 11.5 dBm, and the upstream transmission has a power constraint of 14.5 dBm, in accordance with [30].

Fig. 9 shows the achievable rate regions obtained from the OSB and ISB algorithms. As can be seen, the performance of ISB is very close to that of OSB, although ISB is clearly a suboptimal algorithm. Furthermore, it is observed that the solution provided by ISB is not unique. The nonuniqueness of this algorithm is exposed by choosing a different order of users during the iteration procedure in ISB. ISB gives slightly different rate regions for different iteration orders. Interestingly, no particular order has a rate region that is completely superior to the rate regions of all other orders. In addition, as seen in PSD plots, ordering affects the power spectral densities as well. Fig. 10 shows the PSD pairs corresponding to the downstream–upstream ordering and the upstream–downstream ordering. As can be seen, a narrow low-frequency spectrum is always shared by both directions. In the high-frequency range, frequency-division duplexing (FDD) separates the upstream and the downstream. FDD is optimal in the high-frequency range because of the strong NEXT interference. Interestingly, if the downstream–upstream ordering is used in ISB, the resulting frequency division follows an up–down–up pattern. The situation is reversed when the upstream–downstream ordering is used. The upstream–downstream ordering produces a FDD solution that follows a down–up–down pattern.

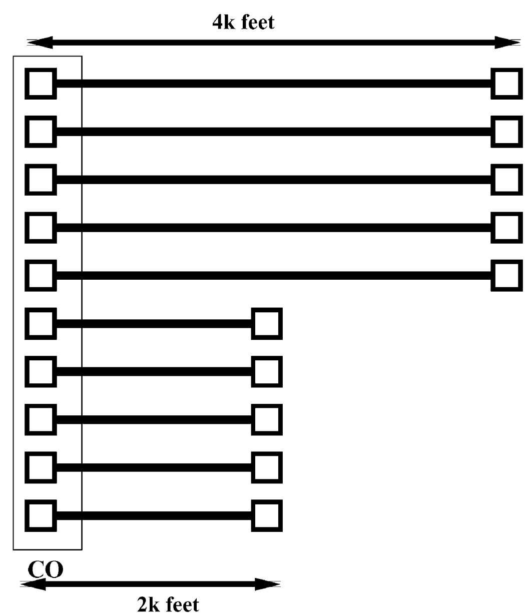


Fig. 11. Loop topology for 10-user VDSL.

C. 10-User VDSL Full-Duplex

In this final set of simulations, we explore the full duplex transmission of a 10-user VDSL scenario with the topology shown in Fig. 11. Again, overlapping spectra is allowed and perfect echo cancellation is assumed. The OSB algorithm as proposed in [12] is not computationally practical in this case.

TABLE III
MAXIMUM MINIMUM RATE FOR 10-USER FULL DUPLEX VDSL

Transmitter	ISB	IWF
4k ft Downstream	14.3Mbps	11.3Mbps
4k ft Upstream	17.4Mbps	11.7Mbps
2k ft Downstream	21.8Mbps	13.6Mbps
2k ft Upstream	29.9Mbps	11.7Mbps

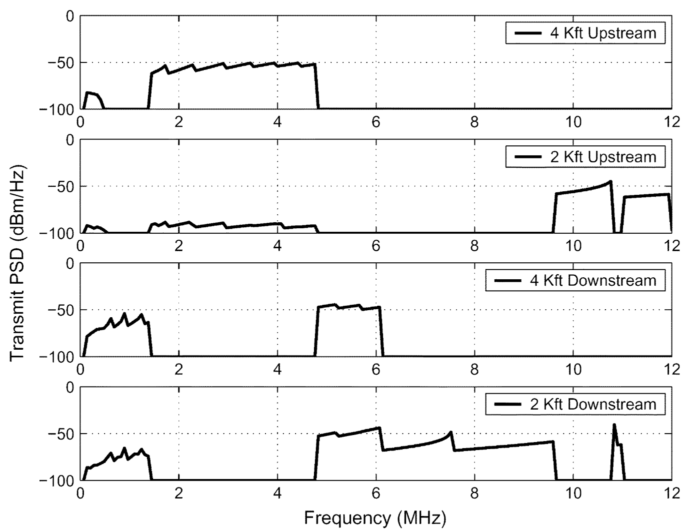


Fig. 12. PSDs for 10-user full duplex VDSL at maximum minimal rate.

With four types of transmitters, the rate region is 4-D, which is difficult to visualize. Instead, we use the maximum minimal rate as the performance metric. The maximum minimum rate is defined as the maximum $\min\{R_1, R_2, R_3, R_4\}$ subject to the power constraint on the four types of transmitters. The dual algorithm developed in this paper is applicable to general nonconvex optimization problems. So, it can be applied to the maximum minimal-rate problem as well. Table III compares the maximum minimum rate computed using the proposed ISB algorithm with that computed from IWF. As seen in Table III, IWF is able to support a minimal data rate of 11.3 Mb/s, while ISB is able to achieve at least 14.3 Mb/s. A minimum gain of 3 Mb/s is possible.

The PSDs obtained from the ISB algorithm are shown in Fig. 12. Interestingly, a small low-frequency band is shared by all four transmitters with full duplex operation. In the middle frequency band, FDM separates upstream and downstream transmissions of the 2 k ft and 4 k ft users. The high-frequency band is used exclusively by the 2 k ft lines. Again, FDD is used there. This type of optimal spectrum usage is nonobvious, and is channel- and user-data-rates-dependent. In this example, the choice of initial ordering is also found to be important. The best results are obtained with the following order: five 4 k ft upstream transmitters first, then five 2 k ft upstream transmitters, then five 4 k ft downstream transmitters, and finally, five 2 k ft downstream transmitters.

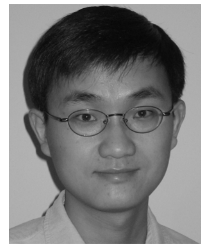
VII. CONCLUSION

A duality theory for nonconvex optimization problems in multicarrier communication systems is presented in this paper. It is shown that the duality gap for a nonconvex optimization problem is zero if the optimization problem satisfies a time-sharing condition. Further, the time-sharing condition is always satisfied for the multiuser spectrum optimization problem in multicarrier systems when the number of frequency carriers goes to infinity. This observation leads to two improvements to the OSB algorithm for DSL. First, the reinterpretation of the OSB algorithm as a dual algorithm leads to effective dual-update methods, such as the subgradient method and the ellipsoid method. Second, we propose a low-complexity and iterative algorithm to approximately evaluate the dual objective. When compared with previous OSB methods, this new ISB algorithm offers a significant complexity reduction with a small loss of optimality in many practical situations. The proposed iterative algorithm is a significant step forward in making OSB practical.

REFERENCES

[1] R. T. Rockafellar, *Convex Analysis*. Princeton, NJ: Princeton Univ. Press, 1970.
[2] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.
[3] C. Y. Wong, R. S. Cheng, K. B. Letaief, and R. D. Murch, "Multiuser OFDM with adaptive subcarrier, bit, and power allocation," *IEEE J. Sel. Areas Commun.*, vol. 17, no. 10, pp. 1747–1758, Oct. 1999.
[4] W. Yu and J. M. Cioffi, "FDMA capacity of Gaussian multiple-access channels with ISI," *IEEE Trans. Commun.*, vol. 50, no. 1, pp. 102–111, Jan. 2002.
[5] L. M. C. Hoo, B. Halder, J. Tellado, and J. M. Cioffi, "Multiuser transmit optimization for multicarrier broadcast channels: Asymptotic FDMA capacity region and algorithms," *IEEE Trans. Commun.*, vol. 52, no. 6, pp. 922–930, Jun. 2004.
[6] K. S. Jacobsen, "Methods of upstream power backoff on very high-speed digital subscriber lines," *IEEE Commun. Mag.*, pp. 210–216, Mar. 2001.
[7] W. Yu, G. Ginis, and J. M. Cioffi, "Distributed multiuser power control for digital subscriber lines," *IEEE J. Sel. Areas Commun.*, vol. 20, no. 5, pp. 1105–1115, Jun. 2002.
[8] *On the Optimality of Power Back-Off Methods*, ANSI-T1E1.4 Contrib. 2000/235, Aug. 2000, G. Cherubini, E. Eleftheriou, and S. Ölçer.
[9] S. T. Chung and J. M. Cioffi, "Methods of upstream power backoff on very high-speed digital subscriber lines," *IEEE Trans. Commun.*, vol. 51, no. 11, pp. 1768–1772, Nov. 2003.
[10] J. Lee, R. V. Sonalkar, and J. M. Cioffi, "A multi-user power control algorithm for digital subscriber lines," *IEEE Commun. Lett.*, vol. 9, no. 3, pp. 193–195, Mar. 2005.
[11] D. Statovci, T. Nordstrom, and R. Nilsson, "The normalized-rate iterative algorithm: A practical dynamic spectrum management method for DSL," *EURASIP J. Appl. Signal Process.*, vol. 2006, pp. 1–17, 2006, Paper 95175.
[12] R. Cendrillon, W. Yu, M. Moonen, J. Verlinden, and T. Bostoen, "Optimal multiuser spectrum balancing for digital subscriber lines," *IEEE Trans. Commun.*, vol. 54, no. 5, pp. 922–933, May 2006.
[13] J. P. Aubin and I. Ekeland, "Estimates of the duality gap in nonconvex optimization," *Math. Oper. Res.*, vol. 1, no. 3, pp. 225–245, 1976.
[14] D. Bertsekas, G. Lauer, N. Sandell, Jr., and T. Posbergh, "Optimal short-term scheduling of large-scale power systems," *IEEE Trans. Autom. Control*, vol. AC-28, no. 1, pp. 1–11, Jan. 1983.
[15] D. Bertsekas, *Constrained Optimization and Lagrange Multiplier Methods*. New York: Academic, 1982.
[16] N. Yamashita and Z.-Q. Luo, "A nonlinear complementarity approach to multiuser power control for digital subscriber lines," *Optim. Methods Softw.*, vol. 19, no. 5, Oct. 2004.
[17] K. B. Song, S. T. Chung, G. Ginis, and J. M. Cioffi, "Dynamic spectrum management for next-generation DSL systems," *IEEE Commun. Mag.*, vol. 17, no. 2, pp. 101–109, Oct. 2002.
[18] K. S. Jacobsen, "Methods of upstream power backoff on very high-speed digital subscriber lines," *IEEE Commun. Mag.*, vol. 39, no. 3, pp. 210–216, Mar. 2001.
[19] *Dynamic Spectrum Management Report*, ANSI-T1E1.4 Contrib. 2003-018R9, 2003, J. M. Cioffi, Ed.

[20] F. Sjöberg, M. Isaksson, R. Nilsson, P. Ödling, S. K. Wilson, and P. O. Börjesson, "Zipper: A duplex method for VDSL based on DMT," *IEEE Trans. Commun.*, vol. 47, no. 8, pp. 1245–1252, Aug. 1999.
[21] Z. Shen, J. G. Andrews, and B. L. Evans, "Optimal power allocation in multiuser OFDM systems," in *Proc. IEEE Global Telecommun. Conf.*, San Francisco, CA, 2003, vol. 1, pp. 337–341.
[22] G. Song and Y. Li, "Cross-layer optimization for OFDM wireless networks—Part I: Theoretical framework; Part II: Algorithm development," *IEEE Trans. Wireless Commun.*, vol. 4, no. 2, pp. 614–634, Mar. 2005.
[23] L. Xiao, M. Johansson, and S. P. Boyd, "Simultaneous routing and resource allocation via dual decomposition," *IEEE Trans. Commun.*, vol. 52, no. 7, pp. 1136–1144, Jul. 2004.
[24] D. Bertsekas, *Nonlinear Programming*. Belmont, MA: Athena Scientific, 1999.
[25] N. Z. Shor, *Minimization Methods for Non-Differentiable Functions*. New York: Springer, 1985.
[26] S. Boyd, EE392o Course Notes Stanford Univ., Stanford, CA, 2004 [Online]. Available: http://www.stanford.edu/class/ee392o
[27] R. Cendrillon and M. Moonen, "Iterative spectrum balancing for digital subscriber lines," in *IEEE Int. Conf. Commun.*, Seoul, Korea, 2005, pp. 1937–1941.
[28] *Asymmetric Digital Subscriber Line (ADSL) Transceivers*, ITU Rec. G.992.1, Jul. 1999.
[29] *Very High Speed Digital Subscriber Line (VDSL)*, ETSI TS 101 270-1 v1.2.1, 1999.
[30] *Very-High-Bit-Rate Digital Subscriber Line (VDSL) Metallic Interface*, ANSI-T1E1.4 Contrib. 2003-210R1, Aug. 2003, S. Schelstraete.



Wei Yu (S'97–M'02) received the B.A.Sc. degree in computer engineering and mathematics from the University of Waterloo, Waterloo, ON, Canada in 1997, and the M.S. and Ph.D. degrees in electrical engineering from Stanford University, Stanford, CA, in 1998 and 2002, respectively.

Since 2002, he has been an Assistant Professor with the Electrical and Computer Engineering Department at the University of Toronto, Toronto, ON, Canada, where he also holds a Canada Research Chair. His main research interests include multiuser information theory, optimization, wireless communications, and broadband access networks.

Dr. Yu is currently an Editor of the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS. He was a Guest Editor of a special issue of the IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS on "Nonlinear Optimization of Communications Systems," and a Guest Editor of a special issue of the *EURASIP Journal on Applied Signal Processing* on "Advanced Signal Processing for Digital Subscriber Lines."

Raymond Lui received the B.A.Sc. and M.A.Sc. degrees in electrical engineering from the University of Toronto, Toronto, ON, Canada, in 2003 and 2006, respectively.

His main research interests are the optimization of multiuser DMT-based xDSL systems and OFDM systems.