Katholieke Universiteit Leuven

Departement Elektrotechniek

ESAT-SISTA/TR 05-173

An efficient Lagrange multiplier search algorithm for Optimal Spectrum Balancing in crosstalk dominated xDSL systems[1]

Paschalis Tsiaflakis, Jan Vangorp and Marc Moonen[2]
Jan Verlinden and Katleen Van Acker
Raphael Cendrillon

August 2005
Submitted for publication in
IEEE Journal on Selected Areas in Communications

[1]This report is available by anonymous ftp from *ftp.esat.kuleuven.ac.be* in the directory *pub/sista/ptsiafla/reports/05-173_LagrangeSearch.pdf*

[2]K.U.Leuven, Dept. of Electrical Engineering (ESAT), Research group SISTA, Kasteelpark Arenberg 10, B-3001 Leuven, Belgium, Tel. 32/16/32 17 09, Fax 32/16/32 19 70, WWW: *http://www.esat.kuleuven.ac.be/sista*. E-mail: *Paschalis.Tsiaflakis@esat.kuleuven.ac.be*. This work was supported in part by the Belgian Programme on Interuniversity Attraction Poles, initiated by the Belgian Federal Science Policy Office IUAP P5/22 ('Dynamical Systems and Control: Computation, Identification and Modelling') and P5/11 ('Mobile multimedia communication systems and networks'), the Concerted Research Action GOA-AMBioRICS, Research Project FWO nr.G.0196.02 ('Design of efficient communication techniques for wireless time-dispersivemulti-user MIMO systems'), IWT project 030054: 'SOLIDT: Solutiuons for xDSL interoperability, deployment and new technologies' and CELTIC/IWT project 040049: 'BANITS 'Broadband Access Networks Integrated Telecommunications' and was partially sponsored by Alcatel-Bell.

Abstract

In modern DSL systems, multi-user crosstalk is the major source of performance degradation. Optimal Spectrum Balancing (OSB) is a centralized algorithm that optimally allocates the available transmit power over frequencies, thereby mitigating the effect of crosstalk. OSB uses Lagrange multipliers to enforce constraints that are coupled over frequencies. However, finding the optimal Lagrange multipliers can become complex when more than two users are considered. Starting from the single user case, this paper presents a number of properties, which are then extended to the multi-user case and lead to an efficient search algorithm for the Lagrange multipliers. Simulations show that the number of Lagrange multiplier evaluations is as small as 40, independent of the number of users, which is much faster than currently known search algorithms.

An efficient Lagrange multiplier search algorithm for Optimal Spectrum Balancing in crosstalk dominated xDSL systems

Paschalis Tsiaflakis, Jan Vangorp and Marc Moonen
Department of Electrical Engineering
Katholieke Universiteit Leuven, Belgium
{Paschalis.Tsiaflakis, Jan.Vangorp, Marc.Moonen}@esat.kuleuven.be

Jan Verlinden and Katleen Van Acker
DSL Research and Innovation
Alcatel Bell, Belgium
{Jan.VJ.Verlinden, Katleen.Van_Acker}@alcatel.be

Raphael Cendrillon
School of Information Technology and Electrical Engineering
University of Queensland, Australia
cendrillon@ieee.org

Abstract

In modern DSL systems, multi-user crosstalk is the major source of performance degradation. Optimal Spectrum Balancing (OSB) is a centralized algorithm that optimally allocates the available transmit power over frequencies, thereby mitigating the effect of crosstalk. OSB uses Lagrange multipliers to enforce constraints that are coupled over frequencies. However, finding the optimal Lagrange multipliers can become complex when more than two users are considered. Starting from the single user case, this paper presents a number of properties, which are then extended to the multi-user case and lead to an efficient search algorithm for the Lagrange multipliers. Simulations show that the number of Lagrange multiplier evaluations is as small as 40, independent of the number of users, which is much faster than currently known search algorithms.

Index Terms

Optimal Spectrum Balancing, Lagrange multiplier, subgradient, dual decomposition, multi-user power loading, multi-user bit loading, spectrum management, crosstalk, xDSL, waterfilling

I. Introduction

TO remain competitive with other emerging broadband access technologies, DSL operators must improve their techniques for data transmission over the existing telephone network. These advanced techniques aim at maximizing the available capacity. The ever increasing demand for higher data rates then forces DSL systems to use higher frequencies, up to 30 MHz for VDSL2 [1]. At these frequencies, electromagnetic coupling becomes particularly harmfull and causes crosstalk between systems operating in the same bundle. This crosstalk, typically 10-15dB larger than the background noise, is the dominant source of performance degradation in DSL systems currently under development [2].

There are two strategies for dealing with this crosstalk: crosstalk cancellation and spectrum management. Crosstalk cancellation can remove the crosstalk completely with minimal noise enhancement [3] [4] [5], but requires signal level cooperation between receivers or transmitters. Unfortunately, in an unbundled scenario this cooperation is not available. In this case crosstalk can be mitigated through the use of spectrum management.

Current DSL systems use a Static Spectrum Management (SSM) approach where fixed spectral masks ensure that crosstalk levels remain within an acceptable range [6] [7]. Because these spectral masks are designed for worst case loop characteristics,

Paschalis Tsiaflakis is a Research Assistant with the F.W.O. Vlaanderen.
Jan Vangorp is a research assistant with the SISTA laboratory.
This research work was carried out at the ESAT laboratory of the Katholieke Universiteit Leuven, in the frame of Belgian Programme on Interuniversity Attraction Poles, initiated by the Belgian Federal Science Policy Office IUAP P5/22 (‘Dynamical Systems and Control: Computation, Identification and Modelling’) and P5/11 (‘Mobile multimedia communication systems and networks’), the Concerted Research Action GOA-AMBioRICS, Research Project FWO nr.G.0196.02 (‘Design of efficient communication techniques for wireless time-dispersivemulti-user MIMO systems’), IWT project 030054: ’SOLIDT: Solutiuons for xDSL interoperability, deployment and new technologies’ and CELTIC/IWT project 040049: ’BANITS ’Broadband Access Networks Integrated Telecommunications’ and was partially sponsored by Alcatel-Bell. The scientific responsibility is assumed by its authors.

this approach can be extremely suboptimal. Dynamic Spectrum Management (DSM) overcomes this problem by designing the transmit spectrum of each modem according to the topology of the network. In this way spectra take into account the current requirements of all users, causing as little disturbance as possible.

One of the first DSM algorithms proposed is Iterative Waterfilling (IW) [8]. In this algorithm, each user waterfills its spectrum against the noise and interference. By repeating this in an iterative fashion over the different users, this converges to a selfish optimum. IW is a low complexity distributed algorithm, meaning it does not need any form of centralized control. Although IW significantly outperforms SSM, it is not optimal. This is especially so in heavily unbalanced scenarios, where some lines cause much more crosstalk than others (e.g. near-far scenario).

The Optimal Spectrum Balancing (OSB) algorithm [9] [10] provides a computationally tractable way to calculate optimal transmit spectra. By optimizing a weighted rate sum, this algorithm can make every possible trade off between the rates of different users. The damage done to other modems in the network is taken into account explicitly, avoiding the selfish optimum and thereby improving on the performance of IW. However, this can only be done when complete information about the channel is available (direct channels as well as crosstalk channels), making OSB only suitable with centralized control in a Spectrum Management Center (SMC).

OSB uses Lagrange multipliers to enforce constraints that are coupled over frequencies. However, finding the optimal Lagrange multipliers can become complex when more than two users are considered. Starting from the single user case, this paper presents a number of properties which are then extended to the multi-user case and lead to an efficient search algorithm for the Lagrange multipliers. Simulations show that the number of Lagrange multiplier evaluations is as small as 40, independent of the number of users, which is much faster than currently known search algorithms.

The paper is organized as follows. In section II, different viewpoints on the spectrum management problem are discussed. It is shown that a dual decomposition can decouple the spectrum management problem into many per-tone problems, leading to a dual problem formulation. Finally, currently available methods to solve this dual problem are discussed.

Section III discusses the single user spectrum management problem from a dual decomposition viewpoint. Some interesting properties and relations are derived. These relations are then extended to the multi-user case in section IV. Based on these relations, an efficient search algorithm for the Lagrange multipliers is proposed.

Section V gives some simulation results of the algorithm for a 2, 3 and 4-user scenario.

II. Optimal Spectrum Balancing

A. System Model

Most current DSL systems use Discrete Multi-Tone (DMT) modulation. The available frequency band is divided in a number of parallel subchannels or tones. Each tone is capable of transmitting data independently from other tones, and so the transmit power and the number of bits can be assigned individually for each tone. This gives a large flexibility in optimally shaping the transmit spectrum.

Transmission for a binder of N users can be modelled on each tone k by

$$\mathbf{y}_k = \mathbf{H}_k \mathbf{x}_k + \mathbf{z}_k \qquad k = 1 \ldots K.$$

The vector $\mathbf{x}_k = [x_k^1, x_k^2, \ldots, x_k^N]^T$ contains the transmitted signals on tone k for all N users. $[\mathbf{H}_k]_{n,m} = h_k^{n,m}$ is an $N \times N$ matrix containing the channel transfer functions from transmitter m to receiver n. The diagonal elements are the direct channels, the off-diagonal elements are the crosstalk channels. $\mathbf{z}_k$ is the vector of additive noise on tone k, containing thermal noise, alien crosstalk, RFI,... The vector $\mathbf{y}_k$ contains the received symbols.

We denote the transmit power as $s_k^n \triangleq \Delta_f E\{|x_k^n|^2\}$, the noise power as $\sigma_k^n \triangleq \Delta_f E\{|z_k^n|^2\}$. The vector containing the transmit power of user n on all tones is $\mathbf{s}^n \triangleq [s_1^n, s_2^n, \ldots, s_K^n]^T$. The DMT symbol rate is denoted as f_s, the tone spacing as Δ_f.

It is assumed that each modem treats interference from other modems as noise. When the number of interfering modems is large, the interference is well approximated by a Gaussian distribution. Under this assumption the achievable bit loading of user n on tone k, given the transmit spectra of all modems in the system, is

$$b_k^n \triangleq \log_2\left(1 + \frac{1}{\Gamma}\frac{|h_k^{n,n}|^2 s_k^n}{\sum_{m \neq n} |h_k^{n,m}|^2 s_k^m + \sigma_k^n}\right), \tag{1}$$

where Γ denotes the SNR-gap to capacity, which is a function of the desired BER, the coding gain and noise margin. The data rate for user n is

$$R^n = f_s \sum_k b_k^n.$$

When the bit loading b_k^n of the users is given for a specific tone k, the required transmit power $\mathbf{s}_k = [s_k^1, s_k^2, \ldots, s_k^N]^T$ for the modems in the system can be calculated by [10]

$$\mathbf{s}_k = \left(\mathbf{D}_k - \mathbf{\Lambda}_k \mathbf{A}_k\right)^{-1} \mathbf{\Lambda}_k \sigma_k \tag{2}$$

$$
\begin{aligned}
\mathbf{D}_k &\triangleq \operatorname{diag}\{|h_k^{1,1}|^2, |h_k^{2,2}|^2, \ldots, |h_k^{N,N}|^2\} \\
\mathbf{\Lambda}_k &\triangleq \operatorname{diag}\{2^{b_k^1} - 1, 2^{b_k^2} - 1, \ldots, 2^{b_k^N} - 1\} \\
[\mathbf{A}_k]_{n,m} &\triangleq a_k^{n,m} \quad \text{with} \quad a_k^{n,m} \triangleq \begin{cases} 0 & n = m \\ \Gamma |h_k^{n,m}|^2 & n \neq m \end{cases} \\
\sigma_k &\triangleq \Gamma[\sigma_k^1, \sigma_k^2, \ldots, \sigma_k^N]^T
\end{aligned}
$$

The total power used by user n is then

$$P^n = \sum_k s_k^n.$$

B. The Spectrum Management Problem

The spectrum management problem amounts to finding optimal transmit spectra for a bundle of interfering DSL lines, following a certain criterion and subject to a number of constraints.

First of all, there is a *total power constraint* $P^{n,tot}$ for each user. This constraint ensures the user's total power does not exceed the maximum allowed total transmit power. On top of this constraint there can be a *spectral mask constraint* $s_k^{n,mask}$ for each tone to guarantee electromagnetic compatibility with other systems.

Note that when using a spectrum management model where both these constraints are present, one of them can be made inactive by merely setting a proper constraint value. The total power constraint can be made inactive by setting its value large enough so that the spectral mask constraint is the most restrictive one:

$$P^{n,tot} \geq \sum_k s_k^{n,mask}.$$

Similarly, by setting the total power constraint as most restrictive, the spectral mask constraint can be made inactive:

$$s_k^{n,mask} \geq P^{n,tot} \qquad k = 1 \ldots K.$$

A second type of constraint is a *rate constraint* for each user. Typically service providers offer a number of profiles and guarantee a certain Quality of Service. The rate constraint then indicates a minimum data rate required by the user.

The spectrum management problem can be viewed from different angles, each time leading to a different criterion that is then to be optimized by the spectrum management algorithm. Either rate, margin or power of the users can be optimized. An optimal solution has to be found within the domain set out by the various constraints.

- In *rate adaptive mode*, the spectrum management problem is to maximize the sum of the data rates of the users. This will be done by using all available power to load a maximum number of bits on tones. The rate is thus limited by the total power and spectral mask constraints. It is possible that there are many solutions to this problem. Several trade-offs between the rates of individual users may exist, all resulting in the same total sum rate. Rate constraints can be used to select a solution with reasonable rates for all users.

$$
\begin{array}{lll}
\text{maximize}_{\mathbf{s}^n} & \sum_{n=1}^N R^n & \\
\text{subject to} & P^n \leq P^{n,tot} & n = 1 \ldots N \\
& 0 \leq s_k^n \leq s_k^{n,mask} & n = 1 \ldots N, k = 1 \ldots K \\
& R^n \geq R^{n,target} & n = 1 \ldots N
\end{array}
\tag{3}
$$

  In this equation, the first set of constraints indicate total power constraints per user. The second set of constraints are spectral mask constraints and the third set are rate constraints per user.

- In *power adaptive mode*, the spectrum management problem is to minimize the total power needed by all users and still meet the rate constraints. This has to be done without violating the spectral mask constraints and total power constraints.

$$
\begin{array}{lll}
\text{minimize}_{\mathbf{s}^n} & \sum_{n=1}^N P^n & \\
\text{subject to} & P^n \leq P^{n,tot} & n = 1 \ldots N \\
& 0 \leq s_k^n \leq s_k^{n,mask} & n = 1 \ldots N, k = 1 \ldots K \\
& R^n \geq R^{n,target} & n = 1 \ldots N
\end{array}
\tag{4}
$$

  Total power constraints are represented in the first set of constraints, the second set of constraints are spectral mask constraints and the third set are rate constraints.

- In *margin adaptive mode*, the spectrum management algorithm will use the available power to tightly satisfy the rate constraints. In this mode, the noise margin is maximized, thus minimizing the bit error rate for the requested data rate. In single-user mode, this can be achieved by using the spectrum management algorithm in power adaptive mode and assigning all power left per user over the used tones by scaling with the same factor, while not violating spectral mask constraints. In the multi-user case however, this strategy should be revised because adding power to a user causes more crosstalk to other users.

C. Dual Decomposition

We will now focus on the *rate adaptive* formulation of the spectrum management problem. The power adaptive form can be treated in a similar way.

The rate adaptive optimization problem (3) is a non-convex problem and therefore difficult to solve. To find the global optimum one must exhaustively search through all possible transmit spectra. This leads to an exponential complexity in both the number of users and tones, namely $\mathcal{O}(B^{NK})$ where B is the number of possibilities for the bit or power loading for each tone and each user in case of discrete or continuous loading respectively.

With $K = 256$ in ADSL and $K = 4096$ in VDSL, exhaustively searching all possible transmit spectra is seen to be computationally intractable. The reason behind this exponential complexity in the number of tones is that the total power constraints and rate constraints are coupled across tones. Therefore transmit spectra have to be searched jointly across tones.

In [9] [10] it was shown that this complexity can be reduced by using the method of dual decomposition to decouple the optimization problem across tones. By using Lagrange multipliers to move constraints coupled over tones into the unconstrained part of the optimization problem, the spectrum management problem can be solved in a per-tone fashion. By choosing appropriate values for the Lagrange multipliers, the constraints can still be enforced.

Following the approach of [11] we first formulate the rate adaptive spectrum management problem (3) in a slightly different way. Instead of optimizing the sum rate of the users in a general fashion, while trying to satisfy individual rate constraints, we now enforce a fixed ratio between the rates for all users. Solutions to this problem are more restricted than solutions to (3) because now rates in excess of the individual rate constraints cannot be assigned randomly but are divided over the users in proportion to their rate constraints. The resulting spectrum management problem is the maximization of the so-called base rate R while satisfying total power and spectral mask constraints. This set of solutions is further restricted by a rate constraint for the users expressed as a fixed proportion of the base rate.

$$\begin{array}{lll} \text{maximize}_{\mathbf{s}^n,R} & R & \\ \text{subject to} & P^n \leq P^{n,tot} & n = 1 \ldots N \\ & 0 \leq s_k^n \leq s_k^{n,mask} & n = 1 \ldots N, k = 1 \ldots K \\ & R^n \geq \beta^n R & n = 1 \ldots N \end{array} \tag{5}$$

The third set of constraints replaces the original rate constraints, now enforcing a fixed ratio β^n of the base rate to be assigned to user n.

In this optimization problem (5), the total power constraints and the rate constraints are coupled across the tones. This results in an exponential complexity in the number of tones. By using Lagrange multipliers, these coupled constraints can be moved into the unconstrained part of the optimization problem. The result is an optimization problem with only per-tone constraints.

$$\mathbf{s}^{1,opt}, \ldots, \mathbf{s}^{N,opt}, R^{opt} = \underset{\mathbf{s}^1,\mathbf{s}^2,\ldots,\mathbf{s}^N,R}{\text{argmax}} \; J(\omega_1, \ldots, \omega_N, \lambda_1, \ldots, \lambda_N, \mathbf{s}^1, \ldots, \mathbf{s}^N, R) \tag{6}$$

$$\begin{array}{lll} \text{with} \quad J & = & R + \sum_{n=1}^N \omega_n \big(R^n - \beta^n R\big) + \sum_{n=1}^N \lambda_n \big(P^{n,tot} - \sum_{k=1}^K s_k^n\big) \\ & = & \big(1 - \sum_{n=1}^N \omega_n \beta^n\big) R + \sum_{n=1}^N \omega_n R^n + \sum_{n=1}^N \lambda_n \big(P^{n,tot} - \sum_{k=1}^K s_k^n\big) \\ \text{and} & & \lambda_n \geq 0, \omega_n \geq 0, \quad n = 1 \ldots N \end{array}$$

If $\big(1 - \sum_n \omega_n \beta^n\big) > 0$ the maximization over R results in $R = \infty$. If $\big(1 - \sum_n \omega_n \beta^n\big) < 0$ the maximization over R results in $R = 0$. There only exists a non-trivial solution if

$$\big(1 - \sum_n \omega_n \beta^n\big) = 0. \tag{7}$$

This results in the following Lagrangian:

$$J = \sum_{n=1}^N \omega_n R^n + \sum_{n=1}^N \lambda_n \big(P^{n,tot} - \sum_{k=1}^K s_k^n\big). \tag{8}$$

This Lagrangian is decoupled across the tones:

$$\begin{aligned} J &= \sum_{k=1}^{K} \underbrace{\left(\sum_{n=1}^{N} \omega_n f_s b_k^n - \sum_{n=1}^{N} \lambda_n s_k^n \right)}_{J_k} + \underbrace{\sum_{n=1}^{N} \lambda_n P^{n,tot}}_{\text{constant}} \\ &= \sum_{k=1}^{K} J_k + \text{constant}. \end{aligned}$$

The constant has no influence on the maximization and can be discarded. Then (for a particular choice of $\lambda_n, \omega_n, n = 1 \dots N$) the problem is reduced to a maximization of a sum across tones, which is equal to the sum of independent maximizations. The original complexity of $\mathcal{O}(B^{NK})$, exponential in K, is now reduced to a linear complexity in K: $\mathcal{O}(KB^N)$.

In a 3 user VDSL system with 14 possible bit loadings per tone, the complexity is reduced from one maximization over $14^{3\times 4096}$ possibilities to 4096 maximizations over 14^3 possibilities. This is a spectacular reduction in complexity.

Solving the dual problem (see also section II-D) does not necessarily correspond to solving the original constrained problem. Optimization theory states that the solution to the dual problem provides only an upper bound to the solution of the original problem. The difference between the upper bound and the optimum of the primal problem is called the duality gap. In [11] it is shown that if multi-carrier systems like DMT satisfy certain conditions, the duality gap is zero. This means the solution to the dual problem is also the solution to the original problem. In practice, it is found that, while the necessary conditions might not be satisfied, the dual problem formulation leads to adequate solutions, and so is currently the method of choice.

D. Solving the Dual Problem

In the previous section the spectrum management problem defined by the following primal problem

$$\begin{array}{lll} \text{maximize}_{s^n,R} & R & \\ \text{subject to} & P^n \le P^{n,tot} & n = 1 \dots N \\ & 0 \le s_k^n \le s_k^{n,mask} & n = 1 \dots N, k = 1 \dots K \\ & R^n \ge \beta^n R & n = 1 \dots N \end{array}$$

was transformed by dual decomposition into the dual problem, decoupled across the tones:

$$\text{for } k = 1 \dots K, \qquad \mathbf{s}_k^{1,opt}, \dots, \mathbf{s}_k^{N,opt} = \underset{\mathbf{s}_k^1, \dots, \mathbf{s}_k^N}{\text{argmax}} \sum_{n=1}^{N} \omega_n f_s b_k^n - \sum_{n=1}^{N} \lambda_n s_k^n \tag{9}$$

$$\begin{array}{lll} \text{subject to} & 0 \le s_k^n \le s_k^{n,mask} & n = 1 \dots N \\ & \lambda_n \ge 0, \omega_n \ge 0 & n = 1 \dots N \\ & \left(1 - \sum_n \omega_n \beta^n\right) = 0 & \end{array}$$

Given $\omega_n, \lambda_n, n = 1 \dots N$ this maximization problem can be easily solved by performing an exhaustive search on each tone over all possible bit or power loading combinations for the users. This results in transmit spectra for all users. For random $\omega_n{}'$s and $\lambda_n{}'$s, the power and rate constraints are generally not satisfied. By choosing appropriate values for the Lagrange multipliers, these constraints can be enforced.

Looking at (9), it can be seen that the $\lambda_n{}'$s influence the resulting spectra. A larger λ_n for user n results in a larger penalty in the cost function when power is allocated to s_k^n. Therefore the $\lambda_n{}'$s can be viewed as setting a cost for power.

The $\omega_n{}'$s have a similar intuitive interpretation. A larger ω_n for user n results in an increased importance attached to its rate. The larger ω_n, the higher the rate allocated to user n compared to other users.

While searching for the $\omega_n{}'$s and $\lambda_n{}'$s corresponding to the constraints at hand, the constraints can be checked at all times by performing an exhaustive search on each tone over all possible bit or power loading combinations for the users. If the constraints are not met with the current $\omega_n{}'$s and $\lambda_n{}'$s, the obtained solution however corresponds to an optimization problem with other constraints. For the $\lambda_n{}'$s there is no choice but to make sure the total power constraints are met. The $\omega_n{}'$s can be treated more loosly. If they turn out not to satisfy the rate constraints of (5), they are still the solution to an optimization problem with some other form of rate constraints. Therefore (9) can be used to solve a spectrum management problem with generic rate constraints like (3). In this case the third condition in (9) dissapears. For any combination of $\omega_n{}'$s there exist $\beta_n{}'$s corresponding to the resulting rate ratio for the users.

E. Searching ω_n's and λ_n's

In [9] [10] a bisection method is proposed to find the λ_n's. It is shown that the total power $\sum_k s_k^n(\lambda_n)$ is monotonically decreasing in λ_n. This monotonicity ensures that bisection can be used to find the λ_n that satisfies the total power constraint of user n. This leads to **algorithm 1** in case there are two users.

Algorithm 1 Two-user spectrum management with bisection on λ's

for given $\omega_1, \omega_2, \lambda_1^{max}, \lambda_1^{min}, \lambda_2^{max}, \lambda_2^{min}$
repeat
  $\lambda_1 = \frac{\lambda_1^{max}+\lambda_1^{min}}{2}$
  repeat
    $\lambda_2 = \frac{\lambda_2^{max}+\lambda_2^{min}}{2}$
    $[\mathbf{s}^1, \mathbf{s}^2]$ = exhaustive search to find optimal spectra
    if $\sum_k s_k^2 > P^{2,tot}$ **then** $\lambda_2^{min} = \lambda_2$ **else** $\lambda_2^{max} = \lambda_2$
  until total power constraint user 2 satisfied
  if $\sum_k s_k^1 > P^{1,tot}$ **then** $\lambda_1^{min} = \lambda_1$ **else** $\lambda_1^{max} = \lambda_1$
until total power constraint user 1 satisfied

For a given value of λ_1 from the outer loop, the inner loop searches the correct λ_2 to satisfy the total power constraint for user 2. When this λ_2 is found, the outer loop updates λ_1 in the appropriate direction, to satisfy the total power constraint of user 1. This will have an influence on λ_2, requiring a new search for λ_2. Because of this nesting of the loops, this search algorithm has an exponential complexity in the number of users.

A second method to find the ω_n's and λ_n's is proposed in [11]. This method relies on the convexity of the dual problem. Therefore, any hill-climbing algorithm is guaranteed to converge. Because the dual problem is not necessarily differentiable, one has to resort to a subgradient method, leading to an update formula for the λ_n's as follows:

$$\lambda_n^{t+1} = \left[\lambda_n^t + \varepsilon\Big(\sum_k s_k^n - P^{n,tot}\Big)\right]^+, \tag{10}$$

where $[x]^+$ means $max(0, x)$. The next λ_n at time $t+1$ is derived from the current λ_n at time t and the distance from satisfying the constraint. ε is a step size parameter which has to be chosen small enough to ensure convergence. In practice a value smaller than 1 is claimed to work well [11].

The same method can be used to search for ω_n's such that the rate constraints are satisfied:

$$\omega_n^{t+1} = \left[\omega_n^t + \varepsilon\Big(R^{n,target} - f_s \sum_k b_k^n\Big)\right]^+. \tag{11}$$

III. Single-user Spectrum Management

In this section, the single user spectrum management problem will be discussed within the framework provided by the dual decomposition method. For the single user case, the spectrum management problem can be formulated as follows:

$$\begin{array}{ll} \text{maximize}_{\mathbf{s}} & R \\ \text{subject to} & \sum_k s_k \le P^{tot} \\ & 0 \le s_k \le s_k^{mask} \quad k = 1 \ldots K \end{array} \tag{12}$$

where now superscript '1'(for user 1) has been omitted. Using a Lagrange multiplier λ to incorporate the total power constraint, the dual problem is formulated as

$$\begin{array}{ll} \text{maximize}_{\mathbf{s}} & R + \lambda(P^{tot} - \sum_k s_k) \\ \text{subject to} & 0 \le s_k \le s_k^{mask} \qquad k = 1 \ldots K \end{array} \tag{13}$$

Recalling that λ represents a cost for power, larger λ's will result in less power being used, e.g. $\lambda = \infty$ leads to $P = \sum_k s_k = 0$ and thus $R = 0$. For any other λ, the optimal spectrum can be found through a per-tone exhaustive search over all possible bit or power loadings.

This relation can be represented on the power axis shown in **figure 1**. This axis shows the total power obtained when optimizing the dual problem for a particular λ. If for a particular λ a total power and rate (P^λ, R^λ) is obtained, then optimality of this solution implies that a total power P less than P^λ, i.e. $P < P^\lambda$ must correspond to a rate R smaller than R^λ, i.e. $R < R^\lambda$. This then ensures that if a λ is found that makes the power constraint tight, the primal problem is solved.



Fig. 1. Power axis

Secondly, it is easily proven that if λ is decreased, then both the achieved bit rate and the consumed power for the optimal solution do not decrease, i.e

$$\begin{aligned} P^{\lambda_B} &\geq P^{\lambda_A} \quad \text{if} \quad \lambda_B < \lambda_A \\ R^{\lambda_B} &\geq R^{\lambda_A} \quad \text{if} \quad \lambda_B < \lambda_A \end{aligned}$$

where $P^{\lambda_A}, P^{\lambda_B}$ and $R^{\lambda_A}, R^{\lambda_B}$ are the optimal total power and rate when $\lambda = \lambda_A$ and $\lambda = \lambda_B$ respectively. This can be proven as follows. Optimality of $(P^{\lambda_A}, R^{\lambda_A})$ for λ_A implies that

$$R^{\lambda_B} - \lambda_A P^{\lambda_B} \leq R^{\lambda_A} - \lambda_A P^{\lambda_A}, \tag{14}$$

because the right-hand side indeed uses the spectra that maximize the Lagrangian for λ_A. Any other spectra, e.g. those corresponding to $(P^{\lambda_B}, R^{\lambda_B})$, result in a smaller value for the Lagrangian. A similar statement can be made on the optimality of $(P^{\lambda_B}, R^{\lambda_B})$ for λ_B:

$$R^{\lambda_A} - \lambda_B P^{\lambda_A} \leq R^{\lambda_B} - \lambda_B P^{\lambda_B}. \tag{15}$$

Taking the sum of (14) and (15) results in

$$\underbrace{(\lambda_B - \lambda_A)}_{\Delta\lambda}\underbrace{(P^{\lambda_B} - P^{\lambda_A})}_{\Delta P} \leq 0. \tag{16}$$

Assuming $\lambda_B < \lambda_A$, (16) implies

$$P^{\lambda_B} \geq P^{\lambda_A}. \tag{17}$$

Then by using this in (15) we obtain

$$R^{\lambda_B} \geq R^{\lambda_A}. \tag{18}$$

Relation (16) can be used to construct a simple procedure to find the λ that makes the total power constraint tight. By starting with a random λ, e.g. $\lambda = 0$, we maximize the Lagrangian to obtain a bit and power loading. If the total power then exceeds the total power constraint, λ has to increase ($\Delta\lambda > 0$ so that $\Delta \mathrm{P} \leq 0$). When the total power is below the total power constraint, λ has to decrease ($\Delta\lambda < 0$ so that $\Delta \mathrm{P} \geq 0$). Note that λ has to remain positive. If by decreasing λ we end up with $\lambda = 0$, the total power constaint is made inactive by some other per-tone power constraint, e.g. spectral mask constraints. An update formula for λ is then as follows

$$\Delta\lambda = \mu\Big(\sum_k s_k - P^{tot}\Big) \quad \Rightarrow \quad \lambda^{t+1} = \left[\lambda^t + \mu\Big(\sum_k s_k - P^{tot}\Big)\right]^+. \tag{19}$$

By varying the step size μ according to **algorithm 2** the λ can be found which makes the power constraint tight. This is shown in **figure 2**. Starting from some initial λ (e.g. $\lambda = 0$), μ is always doubled, creating a trajectory of points on the power axis towards the target power. If for a new point the distance to the target is larger then the distance of the currently best known point, a new trajectory is started by reinitializing (19) with the λ and $\sum_k s_k$ of the best known point.

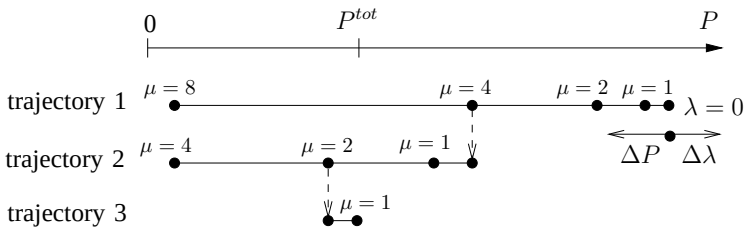


Fig. 2. Single-user search procedure

Formula (19) is the same as the subgradient approach (10) adopted from [11]. Note however that in our derivation we did not have to choose a definition for a subgradient. Moreover, due to the intuitive interpretation of the search algorithm, larger steps can be taken towards the target. This leads to a faster convergence.

Algorithm 2 Single-user λ search algorithm

```
while distance > tolerance do
  λ = best λ so far
  μ = 1
  while distance ≤ previousDistance do
    previousDistance = distance
    μ = μ × 2
    Δλ = −μ(P^tot − P^λ)
    [P^{λ+Δλ}, s_k] = calculateLoading(λ + Δλ)      (per-tone exhaustive search)
    distance = |P^tot − P^{λ+Δλ}|
  end while
end while
```

Using the theory provided by [11], a formula similar to (16) can also be derived. One can plot the maximum achievable rate as a function of the total power as in **figure 3**. This plot can be obtained by performing the maximization of the Lagrangian $max_{\mathbf{s}} R - \lambda P$ for all possible λ's. For each λ this results in an optimal total power usage $P = P^{opt}$ and a corresponding optimal rate $R = R^{opt}$. It is argued in [11] that this function is concave, based on the assumption that the number of tones is *infinite*. Solving $max_{\mathbf{s}} R - \lambda P$ for some given λ corresponds to finding the point on the curve where the difference between R^{λ} and λP^{λ} is largest. From figure 3 it can be seen that this corresponds to finding the point on the curve where the tangent is equal to λ. Hence for the optimal solution

$$\frac{\partial R(P = P^{\lambda})}{\partial P} = \lambda. \tag{20}$$

Mathematically, if the maximum achievable rate as a function of the optimal power budget is twice differentiable, the concavity of this curve can be expressed as

$$\frac{\partial^2 R}{\partial^2 P} \leq 0. \tag{21}$$

Based on (20) and (21), for a *small* ΔP the same relation as (16) is derived as follows:

$$\frac{\partial^2 R}{\partial^2 P} \leq 0 \Rightarrow \frac{\partial}{\partial P}\Big(\underbrace{\frac{\partial R}{\partial P}}_{\lambda}\Big) \leq 0 \Rightarrow \frac{\partial \lambda}{\partial P} \leq 0 \overset{\Delta P \Delta P \geq 0}{\Rightarrow} \Big(\underbrace{\frac{\partial \lambda}{\partial P}\Delta P}_{\Delta\lambda}\Big)\Delta P \leq 0.$$

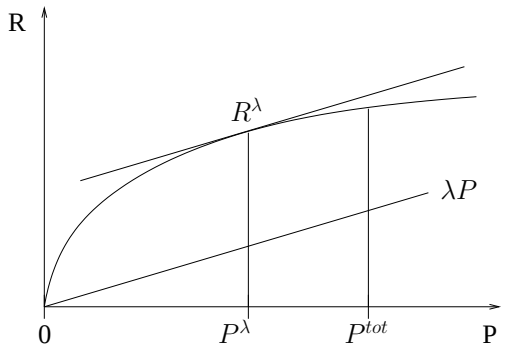


Fig. 3. Optimal rate vs power

Finally, note that the λ of the dual problem formulation (13) without the spectral mask constraint is related to the well known waterfilling solution [12] in the continuous bit loading case. Finding this solution to the optimization problem can be done by differentiating $R + \lambda(P^{tot} - \sum_k s_k)$ with respect to s_k. This leads to

$$s_k + \frac{\Gamma \sigma_k}{|h_k|^2} = \frac{1}{\lambda \ln(2)}.$$

The constant at the right-hand side is the waterfilling level, which is seen to be inversely related to the Lagrange multiplier. A higher waterfilling level corresponds to more power allocated, which can be obtained by setting a small λ.

IV. Multi-user Spectrum Management

In the previous section, interesting relations (16), (17) and (18) were derived between λ, the total power and the data rate. These relations gave rise to a simple procedure for searching the λ and leading to a total power tightly satisfying the total power constraint. In this section these relations will be investigated for the multi-user case. We will focus on the 2-user case first, and then extend the results to the general N-user case.

For the N-user case, the spectrum management problem can be formulated as in formula (3). Formula (9) specified the per-tone dual formulation of the generic spectrum management problem. For a two-user case the dual problem (Lagrangian) can be formulated as

$$\begin{array}{l} \text{argmax}_{\mathbf{s}^1,\mathbf{s}^2}\, \omega_1 R_1 + \omega_2 R_2 + \lambda_1\big(P^{1,tot} - \sum_k s_k^1\big) + \lambda_2\big(P^{2,tot} - \sum_k s_k^2\big) \\ \text{subject to} \quad 0 \le s_k^n \le s_k^{mask} \qquad n = 1,2; k = 1 \ldots K \end{array} \tag{22}$$

Given $\boldsymbol{\omega} = (\omega_1, \omega_2)$ and $\boldsymbol{\lambda} = (\lambda_1, \lambda_2)$, this dual problem, decoupled over the tones, can be solved easily by performing an exhaustive search for each tone over all possible bit or power loading combinations for the users.

$$\begin{array}{l} \text{for } k = 1 \ldots K, \qquad \mathbf{s}_k^{1,opt}, \mathbf{s}_k^{2,opt} = \underset{\mathbf{s}_k^1, \mathbf{s}_k^2}{\text{argmax}} \sum_{n=1}^{2} \omega_n f_s b_k^n - \sum_{n=1}^{2} \lambda_n s_k^n \\ \text{subject to} \quad 0 \le s_k^n \le s_k^{n,mask} \qquad n = 1,2 \end{array} \tag{23}$$

The optimal solution is then a bit and power loading corresponding to total powers and data rates

$$\big(P^{1,\boldsymbol{\omega},\boldsymbol{\lambda}}, R^{1,\boldsymbol{\omega},\boldsymbol{\lambda}}, P^{2,\boldsymbol{\omega},\boldsymbol{\lambda}}, R^{2,\boldsymbol{\omega},\boldsymbol{\lambda}}\big) \tag{24}$$

The optimality of this solution implies that for this $\boldsymbol{\lambda}$ and $\boldsymbol{\omega}$ there exists no other bit or power loading giving a larger value to the Lagrangian. This then implies that for a weighted total power budget $\lambda_1 P^1 + \lambda_2 P^2$ smaller than $P^{\boldsymbol{\omega},\boldsymbol{\lambda}} \triangleq \lambda_1 P^{1,\boldsymbol{\omega},\boldsymbol{\lambda}} + \lambda_2 P^{2,\boldsymbol{\omega},\boldsymbol{\lambda}}$, it is impossible to achieve a weighted rate sum (with weights ω_1, ω_2) that is larger than $R^{\boldsymbol{\omega},\boldsymbol{\lambda}} \triangleq \omega_1 R^{1,\boldsymbol{\omega},\boldsymbol{\lambda}} + \omega_2 R^{2,\boldsymbol{\omega},\boldsymbol{\lambda}}$.

This is shown graphically in the power plane of **figure 4** (which is the 2-user version of figure 1). $\big(P^{1,\boldsymbol{\omega},\boldsymbol{\lambda}}, P^{2,\boldsymbol{\omega},\boldsymbol{\lambda}}\big)$ is the optimal power allocation for a given $\boldsymbol{\lambda}$ and $\boldsymbol{\omega}$. Every loading corresponding to total powers in the marked triangle then has a smaller weighted rate sum.

If a $\boldsymbol{\lambda}$ is found such that $\big(P^{1,\boldsymbol{\omega},\boldsymbol{\lambda}} = P^{1,tot}, P^{2,\boldsymbol{\omega},\boldsymbol{\lambda}} = P^{2,tot}\big)$, this solution satisfies the total power constraints and so has a weighted rate sum larger than every other possible loading in the marked rectangle (i.e. a subset of the marked triangle). Thus the primal problem would be solved now if the rate constraints are satisfied.

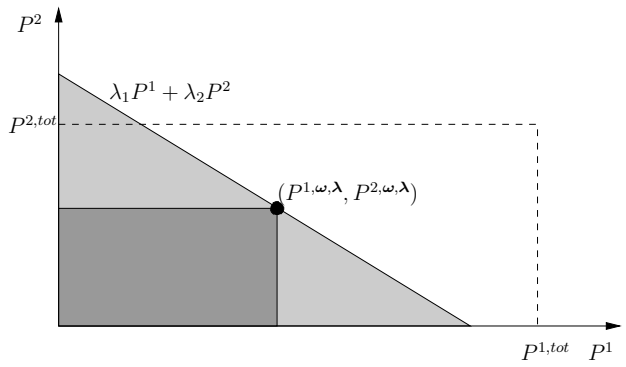


Fig. 4. 2-user power plane

Hence we would like to tune $\boldsymbol{\lambda}$ and $\boldsymbol{\omega}$ such that the total power constraints and rate constraints are satisfied: $P^{1,\boldsymbol{\omega},\boldsymbol{\lambda}} = P^{1,tot}$, $P^{2,\boldsymbol{\omega},\boldsymbol{\lambda}} = P^{2,tot}$, $R^{1,\boldsymbol{\omega},\boldsymbol{\lambda}} \ge R^{1,target}$ and $R^{2,\boldsymbol{\omega},\boldsymbol{\lambda}} \ge R^{2,target}$. Relations between $\boldsymbol{\lambda}$'s, $\boldsymbol{\omega}$'s, total powers and data rates can be explored in order to find strategies to tune the Lagrange multipliers. The goal of this tuning is to enforce the total power constraints and data rate constraints.

The same strategy as in the single-user case can be followed. Starting from two optimal solutions $(R^{1,\boldsymbol{\omega}_A,\boldsymbol{\lambda}_A}, P^{1,\boldsymbol{\omega}_A,\boldsymbol{\lambda}_A}, R^{2,\boldsymbol{\omega}_A,\boldsymbol{\lambda}_A}, P^{2,\boldsymbol{\omega}_A,\boldsymbol{\lambda}_A})$ and $(R^{1,\boldsymbol{\omega}_B,\boldsymbol{\lambda}_B}, P^{1,\boldsymbol{\omega}_B,\boldsymbol{\lambda}_B}, R^{2,\boldsymbol{\omega}_B,\boldsymbol{\lambda}_B}, P^{2,\boldsymbol{\omega}_B,\boldsymbol{\lambda}_B})$ corresponding to $(\boldsymbol{\omega}_A, \boldsymbol{\lambda}_A)$ and $(\boldsymbol{\omega}_B, \boldsymbol{\lambda}_B)$ respectively, optimality for $(\boldsymbol{\omega}_A, \boldsymbol{\lambda}_A)$ implies

$$\begin{array}{rl} & \omega_{1,A} R^{1,\boldsymbol{\omega}_B,\boldsymbol{\lambda}_B} + \omega_{2,A} R^{2,\boldsymbol{\omega}_B,\boldsymbol{\lambda}_B} - \lambda_{1,A} P^{1,\boldsymbol{\omega}_B,\boldsymbol{\lambda}_B} - \lambda_{2,A} P^{2,\boldsymbol{\omega}_B,\boldsymbol{\lambda}_B} \\ \le & \omega_{1,A} R^{1,\boldsymbol{\omega}_A,\boldsymbol{\lambda}_A} + \omega_{2,A} R^{2,\boldsymbol{\omega}_A,\boldsymbol{\lambda}_A} - \lambda_{1,A} P^{1,\boldsymbol{\omega}_A,\boldsymbol{\lambda}_A} - \lambda_{2,A} P^{2,\boldsymbol{\omega}_A,\boldsymbol{\lambda}_A} \end{array} \tag{25}$$

Optimality for $(\boldsymbol{\omega}_B, \boldsymbol{\lambda}_B)$ implies

$$\begin{aligned} & \omega_{1,B} R^{1,\boldsymbol{\omega}_A,\boldsymbol{\lambda}_A} + \omega_{2,B} R^{2,\boldsymbol{\omega}_A,\boldsymbol{\lambda}_A} - \lambda_{1,B} P^{1,\boldsymbol{\omega}_A,\boldsymbol{\lambda}_A} - \lambda_{2,B} P^{2,\boldsymbol{\omega}_A,\boldsymbol{\lambda}_A} \\ \leq \; & \omega_{1,B} R^{1,\boldsymbol{\omega}_B,\boldsymbol{\lambda}_B} + \omega_{2,B} R^{2,\boldsymbol{\omega}_B,\boldsymbol{\lambda}_B} - \lambda_{1,B} P^{1,\boldsymbol{\omega}_B,\boldsymbol{\lambda}_B} - \lambda_{2,B} P^{2,\boldsymbol{\omega}_B,\boldsymbol{\lambda}_B} \end{aligned} \tag{26}$$

Taking the sum of (25) and (26) results in

$$\begin{aligned} & -\underbrace{\left(\omega_{1,B} - \omega_{1,A}\right)}_{\Delta\omega_1} \underbrace{\left(R^{1,\boldsymbol{\omega}_B,\boldsymbol{\lambda}_B} - R^{1,\boldsymbol{\omega}_A,\boldsymbol{\lambda}_A}\right)}_{\Delta R^1} - \underbrace{\left(\omega_{2,B} - \omega_{2,A}\right)}_{\Delta\omega_2} \underbrace{\left(R^{2,\boldsymbol{\omega}_B,\boldsymbol{\lambda}_B} - R^{2,\boldsymbol{\omega}_A,\boldsymbol{\lambda}_A}\right)}_{\Delta R^2} \\ & + \underbrace{\left(\lambda_{1,B} - \lambda_{1,A}\right)}_{\Delta\lambda_1} \underbrace{\left(P^{1,\boldsymbol{\omega}_B,\boldsymbol{\lambda}_B} - P^{1,\boldsymbol{\omega}_A,\boldsymbol{\lambda}_A}\right)}_{\Delta P^1} + \underbrace{\left(\lambda_{2,B} - \lambda_{2,A}\right)}_{\Delta\lambda_2} \underbrace{\left(P^{2,\boldsymbol{\omega}_B,\boldsymbol{\lambda}_B} - P^{2,\boldsymbol{\omega}_A,\boldsymbol{\lambda}_A}\right)}_{\Delta P^2} \leq 0 \end{aligned} \tag{27}$$

Relation (27) for two users can be extended straightforward to the multi-user case:

$$\begin{bmatrix} -(\Delta\boldsymbol{\omega})^T & (\Delta\boldsymbol{\lambda})^T \end{bmatrix} \begin{bmatrix} \Delta\mathbf{R} \\ \Delta\mathbf{P} \end{bmatrix} \leq 0. \tag{28}$$

$\boldsymbol{\lambda} = [\lambda_1, \ldots, \lambda_N]^T$ and $\boldsymbol{\omega} = [\omega_1, \ldots, \omega_N]^T$ are vectors containing the λ's and ω's for the N users, $\mathbf{P} = [P^1, \ldots, P^N]^T$ and $\mathbf{R} = [R^1, \ldots, R^N]^T$ are vectors with the corresponding total powers and data rates.

Two special cases can be derived from formula (28).

$$\text{fixed } \boldsymbol{\omega} \quad (\Delta\boldsymbol{\omega} = 0) \quad \Rightarrow \quad (\Delta\boldsymbol{\lambda})^T \Delta\mathbf{P} \leq 0 \tag{29}$$

$$\text{fixed } \boldsymbol{\lambda} \quad (\Delta\boldsymbol{\lambda} = 0) \quad \Rightarrow \quad (\Delta\boldsymbol{\omega})^T \Delta\mathbf{R} \geq 0 \tag{30}$$

Note that (16) is the single-user version of (29).

Relation (28) can be used to construct a simple procedure to find the $\boldsymbol{\omega}$ and $\boldsymbol{\lambda}$ that make the rate and total power constraints tight. To simplify the graphical illustration of the procedure, we limit ourselves to a procedure that only updates the $\boldsymbol{\lambda}$ Lagrange multipliers. However, this procedure can be straightforwardly extended to also include the update of the $\boldsymbol{\omega}$ Lagrange multipliers by extending the vectors as in formula (28).

In the power plane, formula (29) can be represented graphically as two vectors with a non-positive inner product, as in **figure 5**. When searching for the Lagrange multipliers that make the total power constraint tight, we need to make changes to

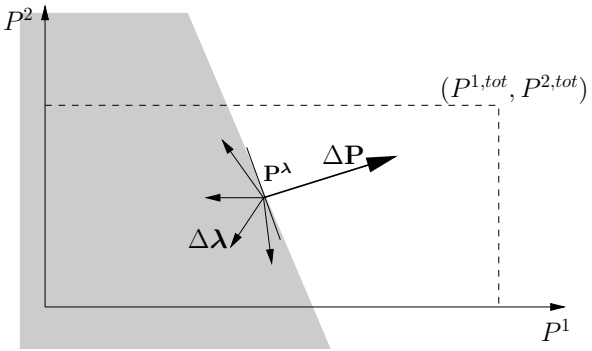


Fig. 5. 2-user power plane

the $\boldsymbol{\lambda}$ such that in the power plane we end up in the point where every user is at maximum power. Because of the non-positive inner product of $\Delta\boldsymbol{\lambda}$ and $\Delta\mathbf{P}$, $\Delta\boldsymbol{\lambda}$ for a desired $\Delta\mathbf{P}$ must be somewhere in the gray half plane opposite to the $\Delta\mathbf{P}$ vector.

This relation between $\Delta\boldsymbol{\lambda}$ and $\Delta\mathbf{P}$ can be used to steer the used power towards the total power constraint. By changing the current $\boldsymbol{\lambda}$ vector by a $\Delta\boldsymbol{\lambda}$ in the opposite direction of the desired $\Delta\mathbf{P}$, formula (29) guarantees that the step taken with $\Delta\mathbf{P}$ will get the used power closer to the power constraint $(P^{1,tot}, P^{2,tot})$, as long as $\Delta\boldsymbol{\lambda}$ is not too large. This is shown in **figure 6**, where the $\Delta\boldsymbol{\lambda}$ brings $\mathbf{P}^{\boldsymbol{\lambda}}$ to the next point inside the shaded circle.

Mathematically this procedure can be captured in the following update formula:

$$\Delta\boldsymbol{\lambda} = -\mu\big(\mathbf{P}^{tot} - \mathbf{P}^{\boldsymbol{\lambda}}\big) \quad \Rightarrow \quad \boldsymbol{\lambda}^{t+1} = \left[\boldsymbol{\lambda}^t - \mu\Big(\mathbf{P}^{tot} - \sum_k \mathbf{s}_k\Big)\right]^+. \tag{31}$$

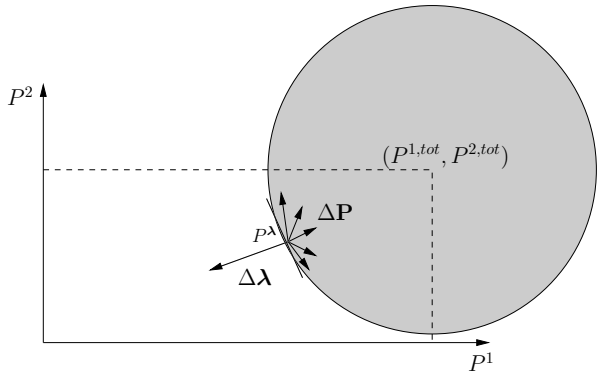


Fig. 6. 2-user power plane

By starting with a small μ, e.g. $\mu = 1$, the used power makes a small step closer to the desired total power. As long as the used power keeps getting closer to the desired total power, μ can be increased, e.g. doubled. A trajectory of points is then followed, each point with a total power closer to the power constraint. At some point, a $\Delta\boldsymbol{\lambda}$ will be selected taking the used power further from the desired total power than the previous point found along the trajectory. Then this last step has to be discarded and a new trajectory is started using a new direction $\left(\mathbf{P}^{tot} - \mathbf{P}^{\boldsymbol{\lambda}}\right)$. This procedure is formally represented in **algorithm 3**. The outer loop of this algorithm iterates over the trajectories while the inner loop follows one of the trajectories. A possible evolution of the total power using this strategy is shown in **figure 7**.

Formula (31) is the same as the subgradient approach (10) adopted from [11]. Note however that in our derivation we again did not have to choose a definition for a subgradient. The derivation for the multi-user case still allows for an intuitive interpretation of the search algorithm. Larger steps can be taken towards the target, which leads to a faster convergence.

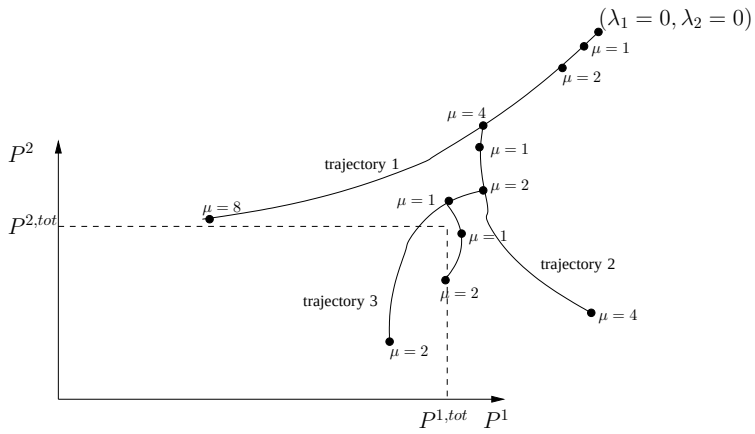


Fig. 7. Trajectories of total power in the power plane

Algorithm 3 Multi-user $\boldsymbol{\lambda}$ search algorithm

while distance $>$ tolerance **do**
  $\boldsymbol{\lambda}$ = best $\boldsymbol{\lambda}$ so far
  $\mu = 1$
  while distance $\leq$ previousDistance **do**
    previousDistance = distance
    $\mu = \mu \times 2$
    $\Delta\boldsymbol{\lambda} = -\mu\left(\mathbf{P}^{tot} - \mathbf{P}^{\boldsymbol{\lambda}}\right)$
    $[\mathbf{P}^{\boldsymbol{\lambda}+\Delta\boldsymbol{\lambda}}, \mathbf{s}^n]$ = calculateLoading($\boldsymbol{\lambda} + \Delta\boldsymbol{\lambda}$) (per-tone exhaustive search)
    distance $= \|\mathbf{P}^{tot} - \mathbf{P}^{\boldsymbol{\lambda}+\Delta\boldsymbol{\lambda}}\|$
  end while
end while

In the framework of [11], a similar relation to (29) results from the concavity of the function that gives the optimal weighted rate $R = \omega_1 R^1 + \omega_2 R^2$ in function of the used total power $(\mathrm{P}^1, \mathrm{P}^2)$. Because of this concavity, for *any* x and y the following relation must hold:

$$\begin{bmatrix} x & y \end{bmatrix} \underbrace{\begin{bmatrix} \frac{\partial^2 R}{\partial^2 P^1} & \frac{\partial^2 R}{\partial P^1 \partial P^2} \\ \frac{\partial^2 R}{\partial P^2 \partial P^1} & \frac{\partial^2 R}{\partial^2 P^2} \end{bmatrix}}_{\begin{bmatrix} \frac{\partial \lambda_1}{\partial P^1} & \frac{\partial \lambda_2}{\partial P^1} \\ \frac{\partial \lambda_1}{\partial P^2} & \frac{\partial \lambda_2}{\partial P^2} \end{bmatrix}} \begin{bmatrix} x \\ y \end{bmatrix} \leq 0$$

By choosing $x = \Delta P^1$ and $y = \Delta P^2$, for small ΔP^1, ΔP^2 relation (29) is obtained.

V. Simulation results

In this section some simulation results will be presented based on the algorithm of the previous section. Its performance will be compared to the other search algorithms presented earlier in section II-E.

All scenarios use a line diameter of 0.5 mm (24 AWG), the maximum transmit power is 20.4 dBm. The SNR gap Γ is set to 12.9 dB, corresponding to a target symbol error probability of 10^{-7}, coding gain of 3 dB and a noise margin of 6 dB. The tone spacing $\Delta_f = 4.3125$ kHz and the DMT symbol rate $f_s = 4$ kHz [13].

All simulations start from initial λ Lagrange multipliers set to zero, ω's are fixed.

A. 2-user scenario

As a first result, the performance of algorithm 3 is compared to the bisection [9] [10] and subgradient [11] search methods. The scenario for this simulation is a 2-user downstream ADSL system with mixed CO and RT deployment. The CO deployed line has a length of 5 km, the RT deployed line is 3 km. The distance between the CO and RT is 4 km as shown in **figure 8**. **Figure 9** shows part of the resulting rate region. The point in the rate region with $R^{\mathrm{CO}} = 1.6$ Mbps and $R^{\mathrm{RT}} = 5.4$ Mbps has spectra as shown in **figure 10**.

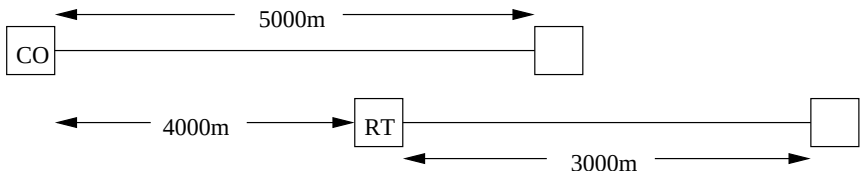


Fig. 8. 2-user downstream ADSL scenario

Depending on the point in the rate region, about 100 to 150 $\boldsymbol{\lambda}$-evaluations are performed with algorithm 3. Because each $\boldsymbol{\lambda}$-evaluation requires a per-tone exhaustive search to determine the loading on all tones, it is important to keep this number as small as possible. For the bisection method of [9] [10] 400 to 600 $\boldsymbol{\lambda}$-evaluations are needed for this 2-user case. Because of its exponential complexity in the number of users, this will get worse as the number of users increases. For the subgradient method with $\varepsilon = 1$ as suggested in [11], more than 20000 $\boldsymbol{\lambda}$-evaluations are required. However, when the number of users increases, the number of evaluations does not increase exponentially for the subgradient method.

The intuitive interpretation of the algorithm allows for the number of $\boldsymbol{\lambda}$-evaluations to be further reduced. By starting a trajectory with step size $\mu = 1$, a number of $\boldsymbol{\lambda}$-evaluations is wasted to increase the μ to a magnitude for which the total power starts converging towards the total power constraint. Instead, one could start a trajectory with a μ inspired by the best μ of the previous trajectory, avoiding unnecessary $\boldsymbol{\lambda}$-evaluations. In this way the algorithm converges in less than 40 $\boldsymbol{\lambda}$-evaluations.

B. 3-user and 4-user scenarios

The performance of algorithm 3 was also tested on the 3 and 4-user scenarios shown in **figure 11 and 13** respectively. The 3-user scenario required 115 $\boldsymbol{\lambda}$-evaluations in 7 trajectories to converge to the optimal spectra in **figure 12**. In this case a point on the rate region was found where the users had a bit rate of respectively 1.5 Mbps, 6.2 Mbps and 4.1 Mbps. The calculation of these spectra took about 10 hours on a Pentium IV. For the bisection method, we estimate convergence would take on the order of 8000 $\boldsymbol{\lambda}$-evaluations. The subgradient method of [11] with step size $\varepsilon = 1$ is expected to have similar performance as in the 2-user case (more than 20000 $\boldsymbol{\lambda}$-evaluations, which is estimated to be over 10 weeks of simulation on a Pentium IV).

For the 4-user case, 160 $\boldsymbol{\lambda}$-evaluations in 10 trajectories where needed to converge. The resulting spectra are shown in **figure 14**. A point on the rate region was found where the data rates of the users were respectively 1.5 Mbps, 5.7 Mbps,

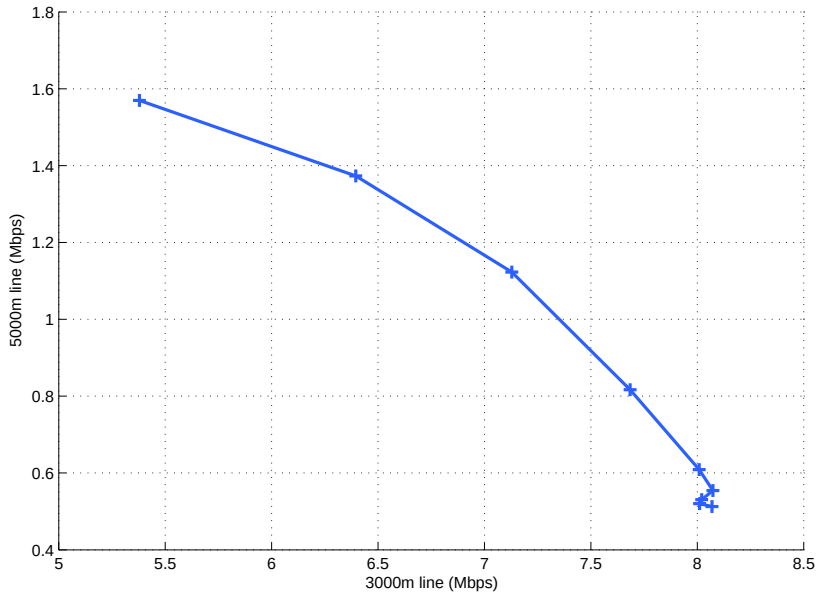


Fig. 9. Rate region for 2-user downstream ADSL scenario

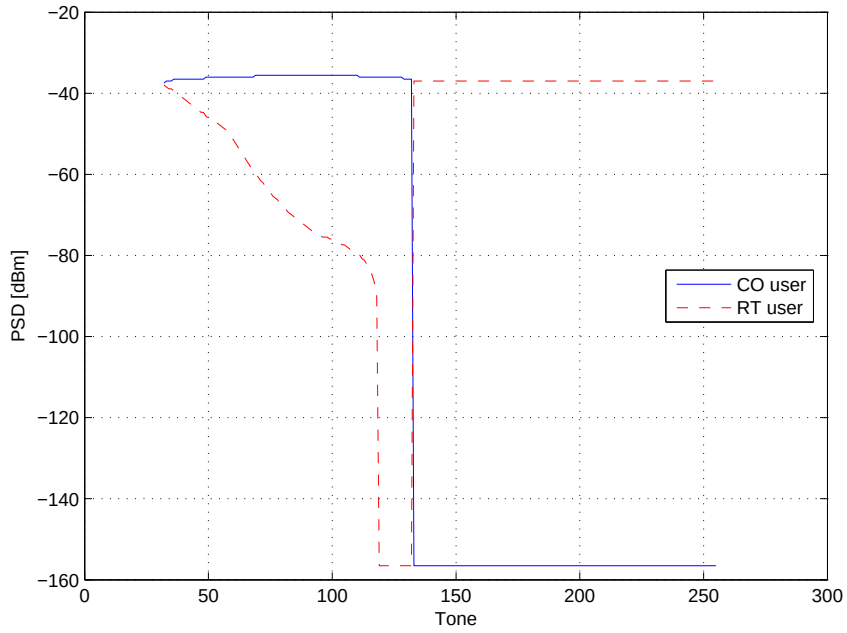


Fig. 10. Spectra for 2-user downstream ADSL scenario

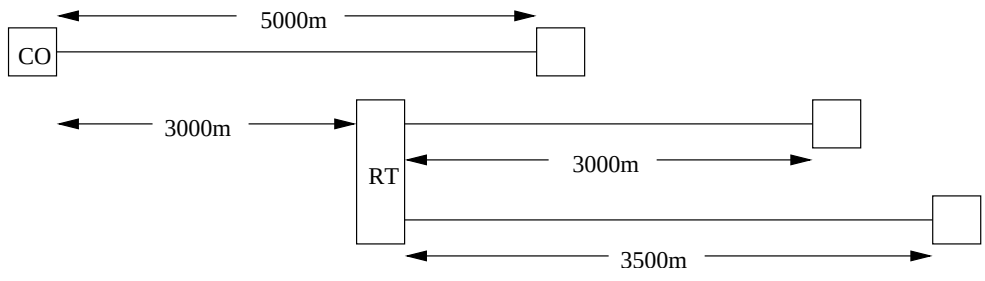


Fig. 11. 3-user downstream ADSL scenario

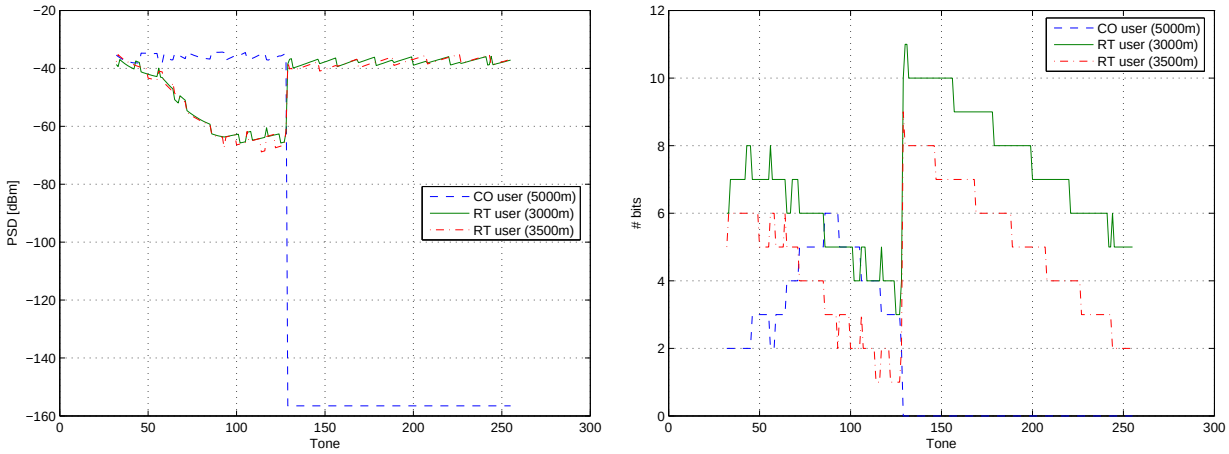


Fig. 12. Spectra and corresponding bitloading for 3-user downstream ADSL scenario

3.8 Mbps and 7.1 Mbps. The calculation of these spectra took about 8 days on a Pentium IV. For the bisection method, we estimate convergence would take on the order of 160000 $\boldsymbol{\lambda}$-evaluations. The subgradient method of [11] again will not increase in complexity.

It is seen that in all scenarios, the number of $\boldsymbol{\lambda}$-evaluations for algorithm 3 is roughly independent of the number of users. The increase in simulation time is due to the complexity of the per-tone exhaustive search, which is exponential in the number of users. This complexity has to be combated with more efficient search methods on the per-tone level if scenarios of more than 4 users have to be simulated. This in particular is a topic of current research. Currently proposed algorithms [14], [15] use an iterative approach giving near-optimal performance.

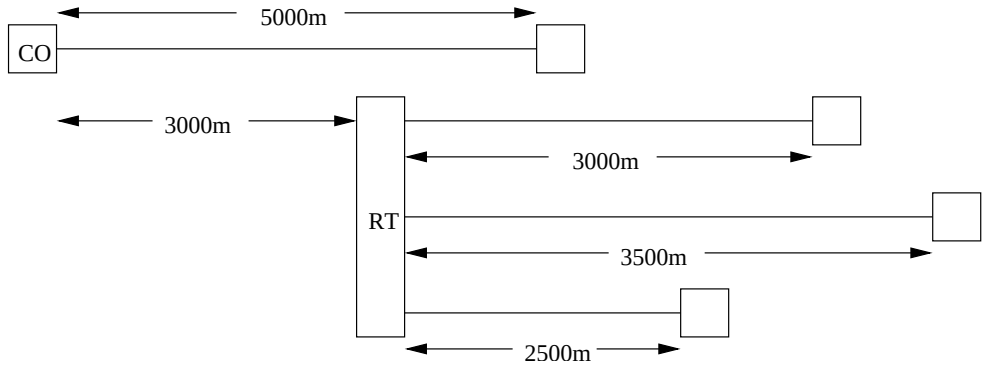


Fig. 13. 4-user downstream ADSL scenario

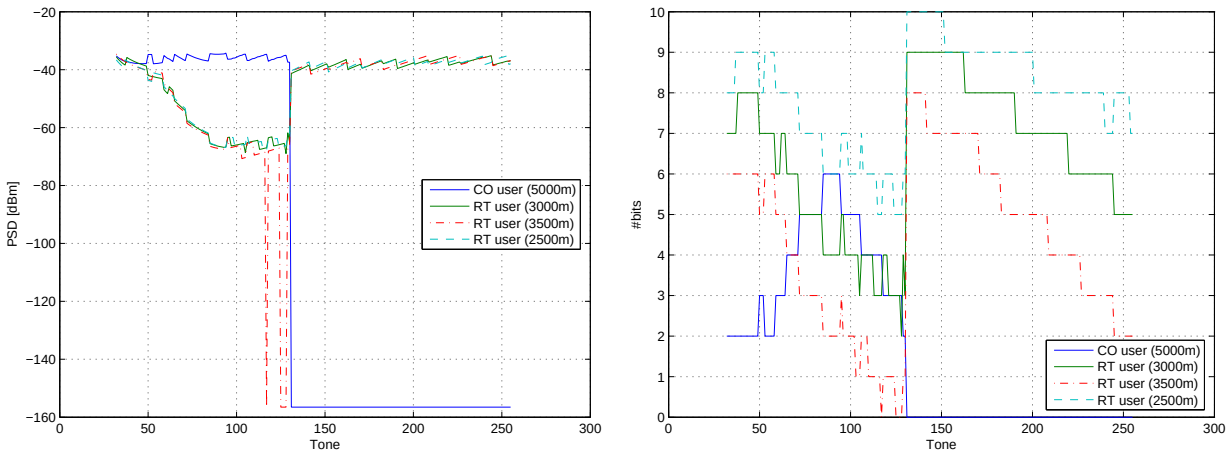


Fig. 14. Spectra and corresponding bitloading for 4-user downstream ADSL scenario

VI. Conclusion

In modern DSL systems, multi-user crosstalk is the major source of performance degradation. Optimal Spectrum Balancing (OSB) is a centralized algorithm that optimally allocates the available transmit power over frequencies, thereby mitigating the effect of crosstalk. OSB uses Lagrange multipliers to enforce constraints that are coupled over frequencies. However, finding the optimal Lagrange multipliers can become complex when more than two users are considered. In this paper, the problem of finding the Lagrange multipliers is analyzed. In the single-user case, this leads to an intuitive procedure to search the Lagrange multiplier. Insights from the single-user case are extended to the multi-user case, leading to basic relations between the Lagrange multipliers and their resulting constrained variables. These relations are used in a search procedure to find the Lagrange multipliers that enforce the constraints. Here *all* Lagrange multipliers can be updated in parallel, leading to a complexity which is found to be roughly independent of the number of users.

Simulations show that for 2-, 3- and 4-user scenarios, 100-150 evaluations of the Lagrange multipliers are sufficient to enforce the constraints. Moreover, the intuitive interpretation of the search algorithm allows for more efficient updates of the step size. This results in an even faster convergence, typically under 40 evaluations of the Lagrange multipliers. Because of the efficient search procedure, the remaining complexity of OSB is situated at the level of the per-tone exhaustive search. This is an area of current research.

References

[1] *Transmission and Multiplexing (TM); Access transmission systems on metallic access cables; Very high speed Digital Subscriber Line (VDSL); Functional Requirements*, ETSI Std. TS 101 270-1, Rev. V.1.3.1, 2003.

[2] Thomas Starr, John M. Cioffi, Peter J. Silverman, *Understanding Digital Subscriber Lines*. Prentice Hall, 1999.

[3] Raphael Cendrillon, Marc Moonen, Etienne Van den Bogaert and George Ginis, "The Linear Zero-Forcing Crosstalk Canceller is Near-optimal in DSL Channels," in *IEEE Global Communications Conference (Globecom)*, vol. 4, December 2004, pp. 2334–2338.

[4] Raphael Cendrillon, Marc Moonen, George Ginis, Katleen Van Acker, Tom Bostoen, Piet Vandaele, "Partial Crosstalk Cancellation for Upstream VDSL," *EURASIP Journal on Applied Signal Processing*, vol. 2004, no. 10, pp. 1520–1535, August 2004.

[5] George Ginis and John M. Cioffi, "Vectored Transmission for Digital Subscriber Line Systems," *IEEE Journal on Selected Areas in Communications*, vol. 20, no. 5, pp. 1085–1104, June 2002.

[6] *Spectrum Management for Loop Transmission Systems*, ANSI Std. T1.417, Issue 2, 2003.

[7] Thomas Starr, Massimo Sorbara, John M. Cioffi, Peter J. Silverman, *DSL Advances*. Prentice Hall, 2003.

[8] Wei Yu, George Ginis and John Cioffi, "Distributed Multiuser Power Control for Digital Subscriber Lines," *IEEE Journal on Selected Areas in Communications*, vol. 20, no. 5, pp. 1105–1115, June 2002.

[9] Raphael Cendrillon, Marc Moonen, Jan Verlinden, Tom Bostoen and Wei Yu, "Optimal Multi-user Spectrum Management for Digital Subscriber Lines," in *IEEE International Conference on Communications (ICC)*, vol. 1, June 2004, pp. 1–5.

[10] Raphael Cendrillon, Wei Yu, Marc Moonen, Jan Verlinden and Tom Bostoen, "Optimal Multi-user Spectrum Management for Digital Subscriber Lines," *accepted for IEEE Transactions on Communications*.

[11] Wei Yu, Raymond Lui and Raphael Cendrillon, "Dual Optimization Methods for Multiuser Orthogonal Frequency-Division Multiplex Systems," in *IEEE Global Telecommunications Conference (Globecom)*, vol. 1, December 2004, pp. 225–229.

[12] Brian S. Krongold, Kannan Ramchandran and Douglas L. Jones, "Computationally Efficient Optimal Power Allocation Algorithms for Multicarrier Communication Systems," *IEEE Transactions on Communications*, vol. 48, no. 1, pp. 23–27, January 2000.

[13] *Asymmetric digital subscriber line transceivers2 (ADSL2)*, ITU-T Std. G.992.3, 2002.

[14] Raphael Cendrillon, Marc Moonen, "Iterative Spectrum Balancing for Digital Subscriber Lines," in *International Communications Conference (ICC)*, May 2005.

[15] Raymond Lui and Wei Yu, "Low-Complexity Near-Optimal Spectrum Balancing for Digital Subscriber Lines," in *International Communications Conference (ICC)*, May 2005.