A Tutorial on Optimization Techniques Applied to DSM Algorithms

Darlene Maciel Neves[a], Aldebaro Barreto da Rocha Klautau Júnior[a], Marcio Murilo Conte[a], Eduardo Lins de Medeiros[a], Jacklyn Dias Reis[a], Boris Dortschy[b]
[a]Signal Processing Laboratory, Pará, DEEC, Belém-PA, Brazil;
[b]Ericsson, Kista, Stockholm, Sweden

ABSTRACT

xDSL systems are widely used nowadays. Services such as VDSL2 can achieve high bitrates over copper wires. The usage of dynamic spectrum management techniques (DSM) can result in even better bitrates, through mitigation of crosstalk, the worst interference in such systems. This tutorial surveys the recent progress in DSM, covering the main algorithms and optimization concepts used by then.

Keywords: DSL, Optimization, Dynamic Spectrum Management (DSM), Power Allocation.

1. INTRODUCTION

One of the major impairments in DSL (Digital Subscriber Line) deployment is the crosstalk among users sharing the same binder. The static spectral masks currently adopted in DSL systems to define the power allocation for each user don't consider the rather dynamic nature of crosstalk interference, thus leading DSL systems to sub-optimal performance.

DSL modems should not transmit more power than necessary to achieve their target data rates with good quality of service and should not use more bandwidth than useful for communication. In order to accomplish such transmission efficiency, adaptive allocation techniques, known as dynamic spectrum management (DSM), can be applied for shaping the transmitted power spectrum density (PSD) in DSL lines. As a result, the PSD allocation in systems that support DSM depends not only on the channel characteristics, which are regarded nearly stationary, but also on the rather changing line crosstalk activity at the moment.

This paper provides a tutorial on the main DSM techniques. The purpose is to provide an integrated overview of the techniques using a consistent and uniform nomenclature. This paper is organized as follows. Section 2 introduces some basic optimization concepts. Section 3 makes some considerations about DMT and spectrum management. Section 4 describes the state-of-art DSM algorithms and their the solutions, some simulations are outlined in Section 5. The concluding remarks are presented in Section 6.

2. NON-LINEAR OPTIMIZATION OVERVIEW

This section provides a concise review of important optimization concepts and techniques, often used in DSM algorithms. It covers convexity, Karush-Kuhn-Tucker (KKT) optimality conditions and Lagrangian duality.

2.1 Convex Functions

A function $f(x) : \mathbb{R}^n \rightarrow \mathbb{R}$ is said to be convex if for any two points x, $y \in \mathbb{R}^n$

$$f(\theta x + (1-\theta)y) \leq \theta f(x) + (1-\theta)f(y), \ \forall \quad \theta \in [0,1] \tag{1}$$

Geometrically, this means that, the line segment joining $(x, f(x))$ and $(y, f(y))$ is above the graph of f. We say that some function $h(x)$ is concave if $-h(x)$ is convex. For that reason, we will refer to concave problems as convex. The most important property about convex function is the fact that they are closed under summation, positive scaling and the pointwise maximum operations.[1]

Further author information
E-mails: {dmn, aldebaro, marciomm, eduardom, jacklyn}@ufpa.br and boris.dortschy@ericsson.com

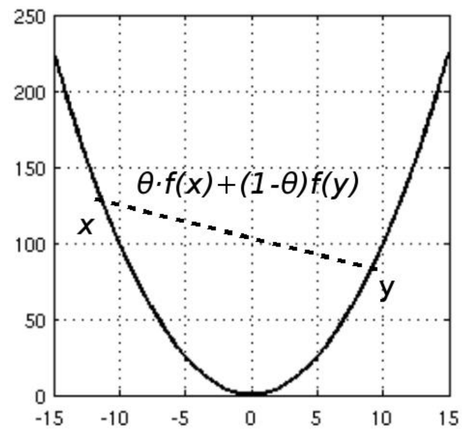

Figure 1. Example of a convex function x^2, it satisfies the condition (1). It presents just one descent direction and one global minimum.

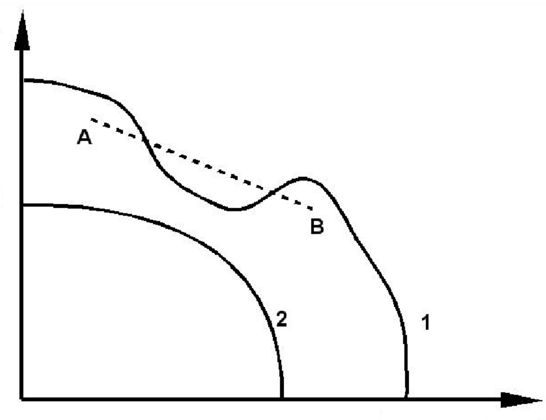

Figure 2. Nonconvex and convex curves. Curve 1 is said non-convex and curve 2 satisfies the convexity condition in (1).

2.2 Optimal Points

For a given function $f(x)$ a point x^* is called *global minimum* of $f(x)$ if

$$f(x^*) \leq f(x), \quad \forall x \tag{2}$$

The same analysis is done for a *global maximum*. A point x^* is called *global maximum* of $f(x)$ if

$$f(x^*) \geq f(x), \quad \forall x \tag{3}$$

If a point satisfies the condition in (2) not for all x value but only to a interval or a neighborhood then this point is a *local minimum*. The same analysis is done to the condition (3), then the point is said to be a *local maximum*.

If some function $h(x)$ is defined in the interval $a \leqslant x \leqslant b$, where $x = x^*$ is the local minimun or maximun (generally said as optimal) point, then this property must be satisfied:

$$\bigtriangledown f(x^*) = 0 \tag{4}$$

In example shown in figure 1 the point $x = 0$ is the local minimum and satisfies the property (4):

$$\frac{\partial x^2}{\partial x} = 2.x \rightarrow 2.(x = 0) = 0$$

Because of the convex nature of the function, this point is also the global minimum as well.

2.3 Mathematical Form of Optimization Problems

Consider the standard form of an optimization problem:

$$\min_{x \in S} \quad f_0(x) \tag{5}$$

$$\text{s.t.} \quad f_i(x) \leq 0, \text{for} \quad i = 1, ..., m \tag{6}$$

$$h_j(x) = 0, \text{for} \quad j = 1, ..., r, \tag{7}$$

where $f_0(x)$ is called the objective function (or cost function), $f_i(x) \leq 0$ and $h_j(x) = 0$ are called inequality and equality constraint functions, respectively, and S is called a constraint set. In practice, S can be implicitly defined if is said that $x \in S$ is feasible and satisfies all the problem's constraints.

The problem described in (5) is usually called *Primal*, because it is the original form of the problem, or in other words, it is still in the domain of the original variable x.

The optimization problem described in (5) is said to be convex if these three conditions are satisfied: "*I) the objective function must be convex; II) the inequality constraint functions must be convex; III) the equality constraint functions* $h_i(x) = a_i^T x - b_i$ *must be affine*".[2]

2.4 KKT Conditions

When including constraints in the optimization problem, the property (4) could not be valid anymore to test the optimality of points. To solve that, there are the *KKT Conditions.* The Karush-Kuhn-Tucker (KKT) conditions are necessary for a point x^* to be a optimal in a constrained problem. In other words, if some candidate point x' does not satisfy the KKT conditions, this point can not be considered optimal.

Consider that $f_0(x)$, $f_i(x)$ and $h_j(x)$ are differentiable. If x^* is a local minimun of the problem, then there are scalars $\lambda_i \geqslant 0$ and μ_j such that

$$-\bigtriangledown f_0(x^*) = \sum_{i=i}^{m} \lambda_i \bigtriangledown f_i(x^*) + \sum_{j=1}^{r} \mu_j \bigtriangledown h_j(x^*) \tag{8}$$

The KKT conditions are defined as bellow:

$$\bigtriangledown f_0(x^*) + \sum_{i=1}^{m} \lambda_i \bigtriangledown f_i(x^*) + \sum_{j=1}^{r} \mu_j \bigtriangledown h_j(x^*) = 0 \tag{9}$$

$$\lambda_i f_i(x^*) = 0, \quad \lambda_i \geqslant 0, \quad \mu \textit{ any} \tag{10}$$

where the λ and μ are called *Lagrange Multiplier* vectors.

2.5 Lagrangian Duality

The main objective of the Lagrangian Duality is to "*take the constraints (...) into account by augmenting the objective function with a weighted sum of the constraint functions*".[2] Several problems have been solved in the literature using this tecnhique, and it has led to several algorithms for solving large-scale optimization problems. More recently, it has proven to be useful in discrete optimization, where at least some of the variables of the problems are restricted to be integers numbers.

The *Lagrangian* for the problem (5) is defined as:

$$L(x, \lambda, \mu) = f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{j=1}^{r} \mu_j h_j(x) \tag{11}$$

where λ_i and μ_i are called Lagrange multipliers associated to the constraints $f_i(x)$ and $h_j(x)$ of the problem, respectively.

The *Lagrangian dual function* is defined as

$$g(\lambda, \mu) = \min_{x \in S} L(x, \lambda, \mu) \tag{12}$$

The new problem (dual) is defined as bellow and can be used to find the solution of the original problem (primal):

$$\begin{aligned} \max \quad & g(\lambda, \mu) & (13) \\ \text{s.t.} \quad & \lambda \geq 0 & (14) \end{aligned}$$

Now lets say that x^o and (λ^o, μ^o) are any feasible solutions of the primal and dual problem, respectively. So the *Weak Duallity* property is always assured:

$$f_0(x^o) \geqslant g(\lambda^o, \mu^o) \tag{15}$$

Numerically, the duallity gap is the difference between primal and dual objective function values. If this difference is zero, then we say that the problem has a *Strong Duallity*. This means that:

$$f_0(x^o) = g(\lambda^o, \mu^o) \tag{16}$$

and the primal can be solved by the dual problem without loss of accuracy.

It is possible to observe in (11) that the Lagrangian function includes the KKT conditions and the Lagrangian is a convex function, so a constraint problem (with objective function concave or not concave) can be efficiently solved by using Lagrangian function.[1]

2.5.1 Lagrangian Dual Decomposition

The dual decomposition methods decouple constraints in primal domain by using a dual variable. They can be used even if the objective function is nonconvex. Some kind of problems can be separated, as in the following example:

$$\begin{aligned} \min \quad & \sum_{k=1}^{q} f_k(x_k) & (17) \\ \text{s.t.} \quad & \sum_{k=1}^{q} g_k(x_k) \leq 0 & (18) \\ & \sum_{k=1}^{q} h_k(x_k) = 0 & (19) \end{aligned}$$

Both the constraints and objective function are decomposed into q. Applying the Lagrangian duality gives:

$$L(x, \lambda, \nu) = \sum_{k=1}^{q} \left[f_k(x_k) + \lambda^T h_k(x_k) + \nu^T g_k(x_k) \right] = \sum_{k=1}^{q} L_k(x, \lambda, \nu) \tag{20}$$

If the duality gap is zero, the primal problem in (17) is equivalent to $g(\lambda, \nu) = \min_x L(x, \lambda, \nu)$ and can be decomposed in:

$$\min_{\mathrm{x}_1, \ldots, \mathrm{x}_q} \quad \sum_{k=1}^{q} L_k(x, \lambda, \nu) \tag{21}$$

The optimization problem was then decomposed into q distinct and smaller subproblems, which can be solved more efficiently.

3. SPECTRUM MANAGEMENT

The use of optimization theory to develop DSM algorithms are facilited by the adoption of *Discrete Multitone* (DMT) by DSL systems standardization bodies. The ability to set the PSD level of each frequency carrier individually gives to DSM techniques the potential to improve the achievable rates.[3] The next section shows a brief description about DMT.

3.1 DMT-based DSL

On typical DSL deployments, N subscribers share a binder with 25 to 100 copper-pairs, with each one provoking *crosstalk* on neighbor pairs within the binder. The crosstalk between pairs in different binders is less significant and typically ignored.

Most DSL standards adopt discrete multitone modulation (DMT), where the channel is divided into K *subchannels* (or *tones*), spaced by $\Delta_f = 4.13125$ kHz. When transmitting data, each user n simultaneously sends K *symbols* (for example, complex numbers as QAM symbols) organized in a vector $\mathbf{x}_n = [x_n^1, \ldots, x_n^K]$. A fast Fourier transform (FFT) is used to convert these *frequency-domain* symbols into *time-domain* samples, which are sent through the channel. If some simplifying assumptions hold (see, e.g.,[4]), the K subchannels can be considered independent. In this case and assuming noise-free transmission, each element of the output vector $\mathbf{y}_n = [y_n^1, \ldots, y_n^K]$ is given by

$$y_n^k = h_{n,n}^k x_n^k, \tag{22}$$

where $h_{n,n}^k = H_n(f)|_{f=k\Delta_f}$ is the frequency response $H_n(f)$ of the n-th channel, which is the Fourier transform of the impulse response $h_n(t)$, calculated at the frequency corresponding to tone k. Hereafter, the subscripts will denote the user n while the superscripts denote the tone k.

In a DSL-enabled binder, a user n receives, at tone k, the signal $y_n^k = h_{n,n}^k x_n^k + \sum_{m \neq n} h_{n,m}^k x_m^k + z_n^k$, where $\sum_{m \neq n} h_{n,m}^k x_m^k$ is the crosstalk from the other $N-1$ users and z_n^k is the additive noise that models noise sources as radio-frequency interference (RFI), thermal noise and *alien* crosstalk. When studying the transmission of N DSL users of a given service (e.g., VDSL2), the alien crosstalk takes in account the interference of other users (other than the N users) that typically use other services (e.g., ISDN, ADSL, etc.).

Given that the subchannels of a DMT transmission can be treated independently, it is convenient to write

$$\mathbf{y}^k = \mathbf{x}^k \mathbf{H}^k + \mathbf{z}^k, \tag{23}$$

where $\mathbf{H}^k = \{h_{n,m}^k\}$ is a $N \times N$ matrix with direct channels in its main diagonal. An off-diagonal element at row n and column m is the crosstalk channel (in frequency-domain) between transmitter m and receiver n. The row vectors $\mathbf{y}^k = [y_1^k, \ldots, y_N^k]$, $\mathbf{x}^k = [x_1^k, \ldots, x_N^k]$ and $\mathbf{z}^k = [z_1^k, \ldots, z_N^k]$ correspond to the inputs, outputs and noises at tone k, respectively.

Due to the random characteristic of the transmitted signals and noise sources, instead of using (23), it is convenient to study the transmission as a stochastic process. The transmitted power spectral density (PSD) of user n at tone k is $s_n^k = \mathbb{E}\{|x_n^k|^2\}$, where $\mathbb{E} = \{\cdot\}$ denotes an expected value. Similarly, $\sigma_n^k = \mathbb{E}\{|z_n^k|^2\}$ is the noise PSD and $|h_{n,m}^k|^2 s_m^k$ is the PSD received by n from m.

$\overline{\mathbf{G}} = \{g_{n,m}^k\}$ define a matrix of dimension $N \times N \times K$, where $g_{n,m}^k = |h_{n,m}^k|^2$ and the elements of the vector $\mathbf{g}_{n,m} = [g_{n,m}^1, \ldots, g_{n,m}^K]$ are the squared magnitudes (*gains*) of the channel from m to n. Considering that each modem treats interference from the other modems as noise, the achievable bitloading of user n on tone k can be written as[5]

$$b_n^k \triangleq \log_2 \left(1 + \frac{1}{\Gamma} \frac{g_{n,n}^k s_n^k}{\sum_{m \neq n} g_{n,m}^k s_m^k + \sigma_n^k} \right), \tag{24}$$

bits per complex dimension, where Γ denotes the *gap* or SNR-gap to *capacity*. The gap is typically specified in $10\log_{10}(\Gamma)$ dB and converted to linear scale to be used in (24). The data rate on line n is thus $r_n = f_s \sum_{k=1}^K b_n^k$, where $f_s = 4$ kHz is the symbol rate adopted in current DSL standards such as ADSL.[6] In practical applications, it may be convenient to increase the *margin* γ [6,7] and, consequently, the robustness of the system. Instead of using (24), a margin $\gamma > 1$ can be obtained by using a smaller number of bits according to

$$\hat{b}_n^k \triangleq \log_2 \left(1 + \frac{\text{SNR}_n^k}{\Gamma \gamma} \right) \tag{25}$$

The margin is the amount by which the SNR on the channel may be lowered before performance degrades to a probability of error greater than the target error probability used when calculating the gap.[4] Dealing with both margin γ and gap Γ seems redundant because the margin could be incorporated into the gap by redefining it as $\hat{\Gamma} = \Gamma\gamma$.

For the sake of illustration, figure $(3a)$ shows curves generated with (24) for gaps of 0, 6 and 12 dB. Considering the curve for $\Gamma = 1$ (i.e, $\Gamma_{\text{dB}} = 0$), one can see that a $(\text{SNR}/\Gamma)_{\text{dB}}$=0 dB allows to load one bit, while 15 bits require approximately 45 dB. A popular rule-of-thumb in DSL is that an improvement of 3 dB in SNR is required to add an extra bit. From (25) and assuming that $\text{SNR}_n^k/(\Gamma\gamma) \gg 1$, one has $b_n^k \approx 0.1\ \text{SNR}_{\text{dB}} \log_2 10 \approx 0.32\ \text{SNR}_{\text{dB}}$.

This rule can also be seen from figure $(3a)$ or, alternatively, from figure $(3b)$ that shows, for example, that adding 5 bits (e.g., from 5 to 10 bits) requires approximately a 30 dB increase in SNR.

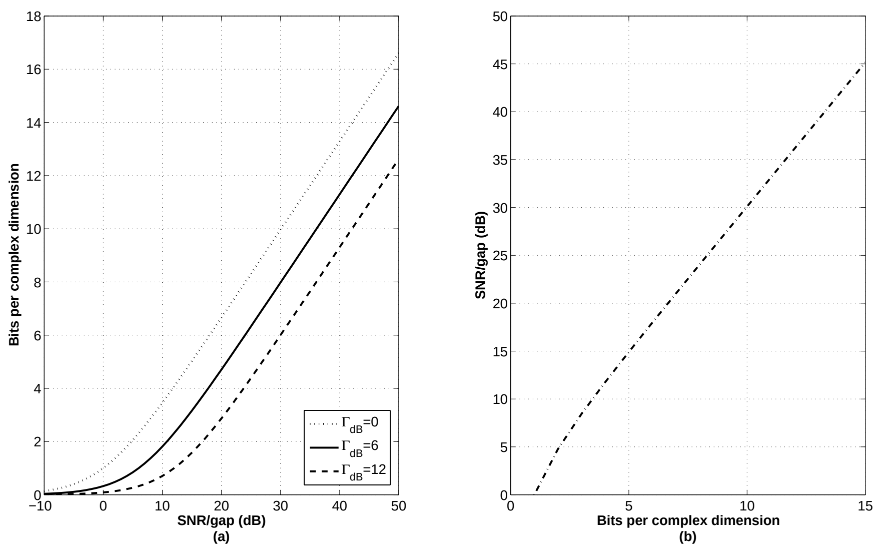

Figure 3. Curves generated with (24) for a gap of 0, 6 and 12 dB.

3.2 The Conventional Spectrum Management Problem in DSL

There are many ways to mathematically pose a DSM problem. The most common ones are rate maximization (also called rate adaptive or RA) or power minimization (also called fixed-margin or FM).[6] The RA problem is

$$\mathbf{P}^* = \arg\max_{\mathbf{P}} r_1 \quad \text{s.t.} \quad r_n \geq R_n^{\min},\ \forall n > 1 \quad \text{and} \quad p_n^{\text{tot}} \leq P_n^{\max},\ \forall n, \tag{26}$$

where r_n and $R_n^{\min}$ are the actual and minimum rates of the n-th user, respectively,

$$p_n^{\text{tot}} = \sum_{k=1}^{K} p_n^k, \tag{27}$$

is the total transmitted power of modem n and $P_n^{\max}$ is the maximum allowed total power (typically imposed by the modem's analog front end[6]).

Other way to solve the DSM problem is by usage of a function $\mathcal{F}$ that maps any operating point $(r_1, \ldots, r_N)$ of the rate region into a vector $\mathbf{w} = (w_1, \ldots, w_N)$ of *weights* for each user. Given $\mathbf{w}$, one can rewrite the original problems as a weighted rate-sum maximization and solve them using a dual formulation.[8,9] For the RA problem, one has:

$$\max \quad \sum_{n=1}^{N} w_n r_n \tag{28}$$

$$\text{s.t.} \quad p_n^{\text{tot}} \leq P^{\max},\ \forall n \tag{29}$$

The weights w can be interpreted as priorities given to each user. Along the optimization process, it is common to force the weights to sum up to a constant $C = \sum_{n=1}^{N} w_n$, such that w_n/C gives the priority of user n.

4. RELATED WORK ON SPECTRUM OPTIMIZATION PROBLEMS

4.1 Iterative Waterfilling - IWF

The iterative waterfilling (IWF) algorithm was proposed by Wei Yu.[8] The IWF solution is obtained as a result of a bit rate maximization of multiple users, where each modem optimizes its own PSD under the assumption that the noise coming from other modems is constant.

The IWF algorithm autonomously coordinates the DSL lines. In the algorithm the level of coordination is only implicit. Each modem can independently determine its own transmit spectra, by using measurements of its own background noise and direct channels. Each user will avoid noisy bands and focus on frequencies it can most efficiently utilize. So no spectrum management center (SMC) is required, as is the case when centralized algorithms are used.[10] This autonomy and its low complexity are the two main pros for this technique.

The problem in (26) is typically an optimization problem and can be solve using Lagrangian. On each k subchannel the corresponding power allocation is found by the waterfilling equation, where the modems will allocate their PSDs according to the noise-to-channel ratio on each tone, as stated next.

The Lagrangian of the optimization problem in (26) is given by

$$L(s_k^1, \ldots, s_n^k) = \sum_n \left(\sum_k R_n + \lambda_n \left(P_n - \sum_{k \epsilon K} s_n^k \right) \right), \tag{30}$$

where λ_n is the Lagrange multiplier of user n and can be interpreted as "price" of violating constraints. The correct values for λ_n are specified by the Karush-Kuhn-Tucker (KKT) conditions. After performing the derivative test, the solution is found and can be written as the waterfilling equation,

$$\left[s_n^k = \frac{log_2 e}{\lambda_n} - N_n^k \right]^+, \tag{31}$$

where $[x]^+ \triangleq max(0, x)$ and

$$N_n^k = \frac{\Gamma \sum_{m \neq n} |h_{n,m}^k|^2 s_m^k + \sigma_n^k}{|h_{n,n}^k|^2} \tag{32}$$

The waterfilling level (λ_n) is chosen such that the modem achieves its target data rate (R_n^{target}), which permits to adjust its PSD according to (31) in an interactive way with other modems.

4.2 Optimal Spectrum Balancing

For simplicity, we will describe the OSB on the two-users case, but it can be easily extended.

A fundamental step toward the OSB solution is the weighted rate-sum analysis. As mathematically proved in,[9] for some w, the spectrum management problem, for the two user case, is equivalent to

$$\max_{s_1, s_2} \quad wR_1 + (1 - w)R_2 \tag{33}$$

$$\text{s.t} \quad \sum_{k=1}^{K} s_k^n \leqslant P_n^{max}, \forall n \tag{34}$$

Varying w will make us find all optimal points in the rate region boundary.

The total power constraints can be incorporated into the optimization problem by defining the Lagrangian

$$L = wR_1 + (1-w)R_2 - \lambda_1(P_1 - \sum_k s_1^k) - \lambda_2(P_2 - \sum_k s_2^k) \ , \tag{35}$$

where λ_n is a Lagrangian multiplier for user n.

Now, defining the Lagrangian on tone k; $L_k = wb_1^k + (1-w)b_2^k - \lambda_1 s_1^k - \lambda_2 s_2^k$. The Lagrangian (35) can be decomposed into a sum across tones of L_k:

$$L = \sum_k L_k + \lambda_1 P_1 + \lambda_2 P_2 \tag{36}$$

It is now possible to split the optimization into the K tones, which are now coupled only through w, λ_1 and λ_2. The objective is to find the PSD pair $[s_1^k, s_2^k]$ (said to be optimal) that maximizes the Lagrangian and the corresponding λ_n's that enable the total power constraint to be respected. The interest is to find the maximum point of L_k and its corresponding PSD pair. With a larger λ_n, a smaller amount of power will be allocated for user n. On the other hand, a λ_n closer to zero leads to the allocation of a bigger amount of power. The convergence algorithm statement means that either $\lambda_n = 0$ or $\sum_k s_n^k = P_n$. The algorithm is not selfish as the IWF. As a result, an optimal balance between each user's data-rates is achieved in convergence.

This solution, nevertheless capable of finding all optimal operating points in the rate region boundary, is still too complex for application in commercial deployments. The algorithm has a complexity that scales exponentially with N, which makes it computationally intractable for more than 5-6 users.

The OSB for N users is simply an extension of OSB algorithm for 2 users. In the general case of N users, the first $N-1$ users will have a target rate constraint and the $N-th$ user will have the rate maximization.

The spectrum management problem of section 3.2 is then rewritten by using the same approach addressed so far for maximizing a weighted rate-sum

$$\max_{s_n, n\epsilon N} \sum_n \omega_n R_n \quad \text{s.t.} \quad \sum_{k\epsilon K} s_n^k \leq P_n, \ \forall n \text{ and } \quad 0 \leq s_n^k \leq s_{n,max}^k, \ \forall k \ , \tag{37}$$

where the weights $\omega_1, \ldots, \omega_{N-1}$ are chosen such that the target rate constraints on users $1, \ldots, N-1$ are met. To solve (37) Lagrangian is used:

$$\max_{s_1^k, \ldots, s_N^k} L_k \quad \text{s.t.} \quad 0 \leq s_n^k \leq s_{n,max}^k, \ \forall n \ , \tag{38}$$

where the Lagrangian on tone k is defined as

$$L_k(s_1^k, \ldots, s_N^k, \omega_1, \ldots, \omega_N, \lambda_1, \ldots, \lambda_{N-1}) = \sum_n \omega_n b_n^k(s_1^k, \ldots, s_N^k) - \lambda_n s_n^k \tag{39}$$

The exhaustive search is used to solve the (38) and to find λ_n and ω_n is used sub-gradient approach because bisection (used in 2-user case) becomes exponentially complex with N.

4.3 Iterative Spectrum Balancing

As an alternative approach to the computationally heavy proposal of the OSB, which is too complex for practical deployment, the Iterative Spectrum Balancing (ISB) algorithm[11] optimizes the PSDs in an iterative way through the users, such that each user is updated at a time, while the PSDs of all other users are fixed at their present values. The ISB method results in a lower complexity algorithm, (compared with the OSB), with near-optimal performance.

This optimization is then equivalent to the OSB maximization of the dual problem decomposed across the tones, but now for the PSD of each user in turn. The optimization procedure follows.

$$\max_{s_n^k} L_k(s_1^k, \ldots, s_N^k) \quad \text{s.t.} \quad 0 \leq s_n^k \leq s_{n,max}^k, \ \forall n \ , \tag{40}$$

The N-dimensional exhaustive search used in OSB can be replaced by a 1-dimensional exhaustive search, since the optimization of the PSD is done for a single user at a time. In a similar fashion to OSB, Lagrange multipliers λ_n and w_n are updated using sub-gradient approach. As a result, the complexity required to determine the PSD is reduced to a quadratic scale in N, rather than exponential.

Thus, the resulting ISB near-optimal algorithm, unlike OSB, is computationally tractable for large N, and feasible if a SMC is available, considering that the algorithm requires that each user must have knowledge of the noise PSDs and crosstalk channels of all modems within the binder. This requirement turns ISB a not fully autonomous DSM algorithm as IWF; however, the improvement of performance seems to be advantageous, since experiments shows ISB often doubles the data-rate of IWF.

4.4 Successive Convex Approximation for Low Complexity - SCALE

The successive convex approximation for low complexity (SCALE) algorithm was proposed in 2005 by.[12] It is an answer for the demand of a low complexity solution with near-optimal performance.

The first aproach is to consider the following relaxation of the problem to avoid the d.c structure.

$$\alpha \log(z) + \beta \leq \log(1+z), \quad \text{where} \begin{cases} \alpha = \frac{z_0}{1+z_0} \\ \beta = \log(1+z_0) - \frac{z_0}{1+z_0}\log(z_0) \end{cases} \tag{41}$$

This way (28) can be rewritten as

$$\max_{s_1,\ldots,s_n \geqslant 0} \quad \sum_{n=1}^{N} w_n \sum_{k=1}^{K} \alpha_n^k \log(\text{SNR}(s)) + \beta_n^k \tag{42}$$

$$\text{s.t.} \quad p_n^{\text{tot}} \leq P^{\max}, \ \forall n \tag{43}$$

However, with one little trick, a change of variable, we can turn this problem in a standard convex problem! The trick is to consider $x_n^k = \log(s_n^k)$ and solve the problem on the new variable x

$$\max_{s_1,\ldots,s_n \geqslant 0} \quad \sum_{n=1}^{N} w_n \sum_{k=1}^{K} \alpha_n^k \log(\text{SNR}(e^x)) + \beta_n^k \tag{44}$$

$$\text{s.t.} \quad \sum_{n=1}^{N} e^{x_n} \leq P^{\max}, \ \forall n \tag{45}$$

Knowing that this problem is convex, its Lagragian is defined as

$$L = \sum_{n=1}^{N} w_n \sum_{k=1}^{K} \alpha_n^k \log(\text{SNR}(e^x)) + \beta_n^k - \sum_{n=1}^{N} \lambda_n \{e^{x_n} - P_n^{max}\} \tag{46}$$

and we can find the optimal point using (4) and transforming back the problem to the original variable s.

$$\frac{\partial L}{\partial x_n^k} = 0 = w_n \alpha_n^k - s_n^k(\lambda_k + \sum_{i \neq j} w_i a_i^k h_{i,j}^k \frac{\text{SNR}_i^k(s)}{h_{i,i}^k s_i^k}) \tag{47}$$

This way we have a closed equation to update the PSDs:

$$s_n^k = \frac{w_n \alpha_n^k}{\lambda_k + \sum_{i \neq j} w_i a_i^k h_{i,j}^k \frac{\text{SNR}_i^k(s)}{h_{i,i}^k s_i^k}} \tag{48}$$

5. SIMULATIONS RESULTS

This section presents some simulation results that are useful to compare the performance of DSM algorithms assuming standard channel models. For all simulations, it was assumed VDSL2, band plan B8-3 ,[13] *ANSI VDSL Noise A* as background noise[5] and the upstream direction.

A two-user scenario was selected to make the comparison of achievable bitrate between the DSM algorithms when using standard channel models. The scenario is depicted in figure (4) and consists of two VDSL2 modems in a DSL provider or central office (CO). A maximum transmit power of 21.05 dBm was allowed to each modem. The results are presented in figure (5) as a rate region plot. A rate region is a plot containing all possible

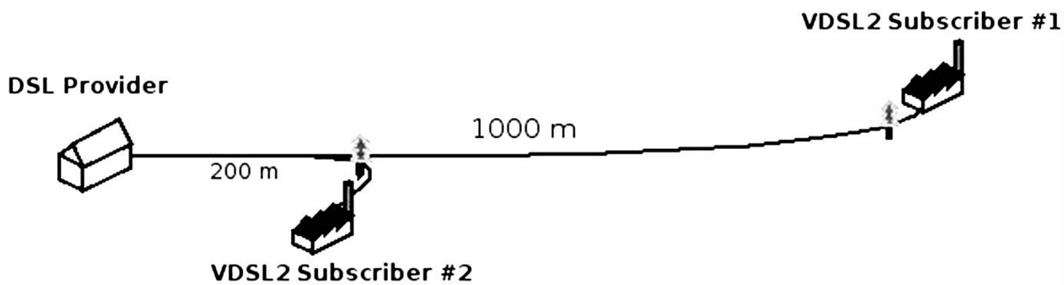

Figure 4. 2 non-equal length users using VDSL2 service.

operation points (a set of bitrates) that can be achieved for a certain algorithm with a certain maximum power. Algorithms solving the RA problem map the border of the rate region. Changing the maximum power moves the border, so that inner points can also be mapped.

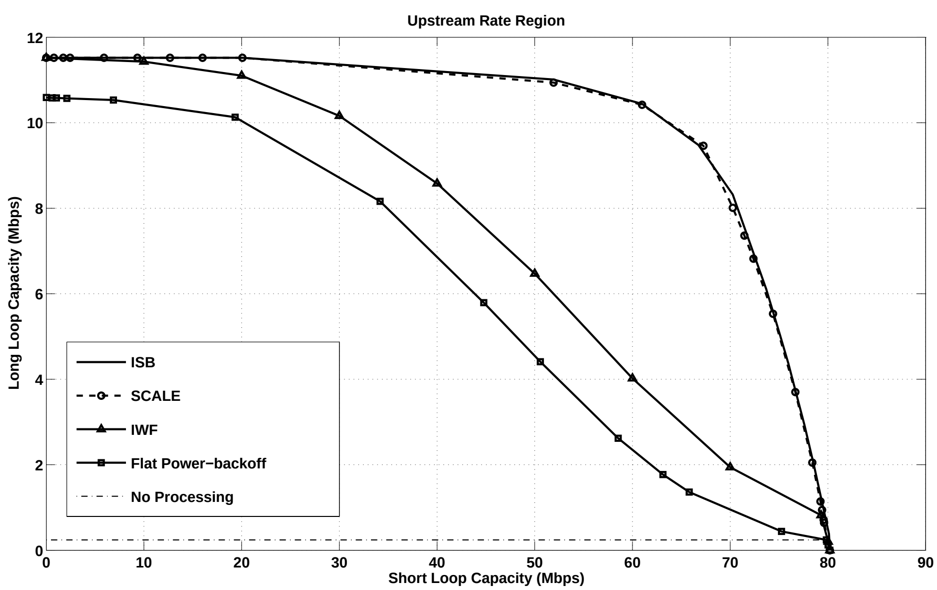

Figure 5. Rate regions obtained for DSM algorithms with 2 users.

It is possible to observe in figure (5) that the SCALE algorithm has a similar result to ISB, having much lower complexity. IWF achieves lower bitrates, but it doesn't require any crosstalk transfer function knowledge nor message passing.

6. FINAL REMARKS

Convex optimization concepts, nonconvex optimization problems and solutions for DSM algorithms were presented in this paper. It was shown the use of optimization concepts and techniques in the state-of-the-art DSM algorithms. Optimal and near-optimal solutions for the DSM problem were discussed, most of then based on different uses of the Lagrangian, but also on interesting relaxation techniques (SCALE).

Optimization provides a powerful set of tools for design and analysis of xDSL systems, specially in DSM algorithms. The concepts surveyed here can be used and extended to solve similar problems in other areas of telecommunications.

An open issue to be considered in future works is the development of new algorithms that can achieve optimal or near-optimal performance with lower computational cost, less message passing, and no previous knowledge about crosstalk transfer functions.

Acknowledgment

Authors would like to thank the Research and Development Center, Ericsson Telecommunications S.A., Brazil. Some of the authors want to acknowledge the financial support received from the European Commission IST 6th Framework and from the Swedish Agency for Innovation Systems, VINNOVA, through the IST - MUSE and the Eureka - Celtic BANITS projects respectively, which partially enabled this work.

REFERENCES

1. M. Bazaraa and C. M. Shetty, *Nonlinear Programming*. Jonh Wiley, 2006.
2. S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.
3. W. Yu, R. Lui, and R. Cendrillon, "Dual Optimization Methods for Multiuser Orthogonal Frequency Division Multiplex Systems," in *IEEE Global Communications Conference (Globecom)*, 2004, pp. 225–229.
4. J. M. Cioffi, "Class notes," 2002. [Online]. Available: http://www.stanford.edu/group/cioffi/
5. T. Starr, J. M. Cioffi, and P. J. Silverman, *Understanding Digital Subscriber Line Technology*. Prentice-Hall, 1999.
6. T. Starr, M. Sorbara, J. M. Cioffi, and P. J. Silverman, *DSL Advances*. Prentice-Hall, 2003.
7. S. Panigrahi and Y. Xu, "Enhanced multiuser resource allocation using near-continuous loading algorithms with adaptive coding for digital subscriber loops," *Electrical and Computer Engineering, 2005. Canadian Conference on*, vol. 1.
8. W. Yu, G. Ginis, and J. M. Cioffi, "Distributed Power Control for Digital Subscriber Lines," *IEEE J. Select. Areas Commun.*, vol. 20, no. 5, pp. 1105–15, June 2002.
9. R. Cendrillon, "Multi-user signal and spectra co-ordination for digital subscriber lines," Ph.D. dissertation, Katholieke Universiteit Leuven, Dec. 2004.
10. R. Cendrillon, M. Moonen, W. Yu, J. Verlinden, and T. Bostoen, "On the Optimality of Iterative Waterfilling," 2003.
11. R. Cendrillon and M. Moonen, "Iterative spectrum balancing for digital water-filling implemented on ADSL modems," in *Proc. IEEE Int. Conf. on Acoust., Speech and Sig. Processing (ICASSP)*, May 2005.
12. J. Papandriopoulos and J. S. Evans, "Low-Complexity Distributed Algorithms for Spectrum Balancing in Multi-User DSL Networks," 2005. [Online]. Available: http://www.cubinlab.ee.mu.oz.au/~jpap/
13. V. S. G.993.2, "International telecommunications union," 2005.