# AN IMPROVED DUAL DECOMPOSITION APPROACH TO DSL DYNAMIC SPECTRUM MANAGEMENT

*Paschalis Tsiaflakis, Ion Necoara, Johan A. K. Suykens, and Marc Moonen*

Katholieke Universiteit Leuven, Electrical Engineering, ESAT-SCD
Kasteelpark Arenberg 10, B-3001, Leuven-Heverlee, Belgium
email:{paschalis.tsiaflakis,ion.necoara,johan.suykens,marc.moonen} @esat.kuleuven.be

## ABSTRACT

Modern DSL networks suffer from crosstalk among different lines in the same cable bundle. By carefully choosing the modems' transmit power spectra, the impact of crosstalk can be minimized leading to spectacular data rate performance gains. This is also referred to as dynamic spectrum management (DSM). DSM algorithms based on an iterative convex approximation approach are recognized as being very effective in tackling the corresponding non-convex optimization problems. One crucial ingredient of this type of algorithms, is a subgradient-based dual decomposition approach to solve the corresponding convex approximations. Although a dual decomposition approach decouples the problem into manageable subproblems, the subgradient-based updates are known to exhibit a slow convergence, with a difficult but crucial stepsize selection. This paper presents an improved dual decomposition approach that improves on the convergence of existing subgradient-based approaches by one order of magnitude. It uses a smoothing technique for the Lagrangian combined with an optimal gradient-based scheme for updating the Lagrange multipliers. Furthermore, the optimal stepsize parameters are selected automatically. The proposed approach makes an important step towards obtaining numerically fast and effective DSM algorithms.

## 1. INTRODUCTION

Digital subscriber line (DSL) technology remains by far the most popular broadband access technology. The increasing demand for higher data rates forces DSL systems to use higher frequencies. At these high frequencies, electromagnetic coupling becomes particularly harmful and causes interference, also called crosstalk, among lines operating in the same cable bundle. This crosstalk is a major obstacle for modern DSL systems towards reaching higher data rates.

Dynamic spectrum management (DSM) [1] refers to a set of solutions to the crosstalk problem. These solutions consist of signal level and/or spectrum level coordination amongst the different modems. In this paper the focus is on spectrum level coordination, which is also referred to as spectrum balancing. Here the modems' transmit power spectra are designed so as to mitigate the impact of crosstalk interference, leading to spectacular performance gains. The problem of optimally choosing the transmit power spectra to maximize the data rates of the network can be formulated as an optimization problem [2], and is referred to as the spectrum management problem. Unfortunately this optimization problem is a very difficult, NP-hard, nonconvex optimization problem. State-of-the-art DSM algorithms (CA-DSB [3], SCALE [4]) use an iterative convex approximation approach to tackle this nonconvex problem. This approach consists of iteratively executing the following two steps: (i) approximating the nonconvex problem by a convex optimization problem, and (ii) solving the convex approximation using a standard (sub)gradient-based dual decomposition approach. We will focus on the second step, which requires the major part of the computational complexity. The standard subgradient-based updates used in this step can lead to very slow convergence. This is mainly because of two reasons: (i) subgradient methods are generally known to exhibit a slow convergence, i.e. a worst case convergence of order $\mathcal{O}(\frac{1}{\varepsilon^2})$ with $\varepsilon$ referring to the required accuracy of the approximation of the optimum [5], and (ii) the stepsizes used by subgradient methods are very difficult to tune so as to guarantee fast convergence.

In this paper we propose a novel improved dual decomposition approach for iterative convex approximation based DSM algorithms inspired by recent advances in mathematical programming [6]. More specifically the novel approach improves on the convergence of existing subgradient-based approaches by one order of magnitude with the same computational complexity. The proposed method uses (i) a smoothing technique for the Lagrangian that preserves the separability of the problem, (ii) an optimal gradient-based scheme, and (iii) optimal stepsizes, which leads to straightforward tuning.

This paper is organized as follows. In Section 2 the system model for the crosstalk environment is described. In Section 3 the spectrum management problem is reviewed. In Section 4 the iterative convex approximation approach for DSL DSM is briefly reviewed. In Section 5 the improved dual decomposition approach is proposed with corresponding proofs on the convergence speed-up. Finally in Section 6 simulation results are given.

## 2. SYSTEM MODEL

Most current DSL systems use discrete multi-tone (DMT) modulation. For the standardly assumed case of perfect tone synchronisation, the transmission for a binder of $N$ modems, using a frequency range of $K$ tones, can be modeled on each tone $k$ by

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

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 $N$ modems. $[\mathbf{H}_k]_{n,m} = h_k^{n,m}$ is an $N \times N$ matrix containing the channel transfer functions from transmitter $m$ to receiver $n$ on tone $k$. 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.

 

The transmit power is denoted by $s_k^n \triangleq \Delta_f E\{|x_k^n|^2\}$, the noise power by $\sigma_k^n \triangleq \Delta_f E\{|z_k^n|^2\}$. The vector containing the transmit power of modem $n$ 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 by $f_s$, the tone spacing by $\Delta_f$. The set of users is denoted by $\mathscr{N} = \{1, \ldots, N\}$

In our model, no signal coordination is assumed among transmitting and receiving modems. Each modem views the signals from the other modems as noise. When the number of interfering modems is large, the interference is well approximated by a Gaussian distribution. Under this standard assumption the achievable bit loading for modem $n$ on tone $k$, given the transmit spectra $\mathbf{s}_k \triangleq [s_k^1, s_k^2, \ldots, s_k^N]^T$ of all modems in the system, is

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

where $\Gamma$ denotes the SNR-gap to capacity, which is a function of the desired BER, the coding gain and noise margin. The total bit rate for modem $n$ and the total power used by modem $n$ are $R^n = f_s \sum_{k \in \mathscr{K}} b_k^n$ and $P^n = \sum_{k \in \mathscr{K}} s_k^n$ respectively.

## 3. SPECTRUM MANAGEMENT PROBLEM

The problem of optimally balancing the transmit power spectra $s_k^n, k \in \mathscr{K}, n \in \mathscr{N}$, to maximize the data rates of the DSL network is referred to as the rate adaptive spectrum management problem. The objective is to find the optimal transmit spectra for a bundle of interfering DSL modems, maximizing a weighted bit rate, subject to per-modem total power constraints and spectral mask constraints. This can be formulated as the following nonconvex optimization problem [7]:

$$\begin{array}{lll} \max\limits_{\mathbf{s}^n, n \in \mathscr{N}} & \sum\limits_{n \in \mathscr{N}} w_n R^n & \\ \text{s.t.} & \sum\limits_{k \in \mathscr{K}} s_k^n \leq P^{n,\text{tot}} \quad , \; n \in \mathscr{N}, & (\mathbf{F}) \\ & 0 \leq s_k^n \leq s_k^{n,\text{mask}}, \; n \in \mathscr{N}, k \in \mathscr{K}, & \end{array} \qquad (2)$$

where $P^{n,\text{tot}}$ denotes the total power budget for modem $n$ and $s_k^{n,\text{mask}}$ denotes the spectral mask for modem $n$ on tone $k$. The weights $w_n$ are used to put more emphasis on some modems. Let us also define $\mathbf{P}^{\text{tot}} = [P^{1,\text{tot}}, \ldots, P^{N,\text{tot}}]^T$.

## 4. DSM BASED ON ITERATIVE CONVEX APPROXIMATIONS

DSM algorithms based on iterative convex approximations, such as CA-DSB and SCALE, are known to be very effective in tackling the nonconvex optimization problem $\mathbf{F}$ (2). Their basic approach is summarized in Algorithm 1. It starts with an initial convex approximation $\mathbf{F}_{\text{cvx}}$ of the nonconvex problem $\mathbf{F}$ in line 1. In line 3 the obtained convex approximation is solved using a subgradient-based dual decomposition approach, which will be discussed in more detail later in this section. In line 4 the approximation is improved based on the solution $\mathbf{s}_{k,\text{cvx}}, k \in \mathscr{K}$, obtained in line 3. This iterative scheme converges to a locally optimal solution of (2) under certain conditions [8] on the chosen convex approximations, which are indeed satisfied for both CA-DSB and SCALE. In the remaining of this text we will elaborate the proposed schemes for CA-DSB. This can similarly be done for SCALE but requires more complicated notations because of the inherent exponential transformation of variables.

**Algorithm 1** DSM based on iterative convex approximations

1: Approximate $\mathbf{F}$ by a convex approximation $\mathbf{F}_{\text{cvx}}$
2: **repeat**
3: Solve $\mathbf{F}_{\text{cvx}}$ using a subgradient-based dual decomposition approach, to obtain $\mathbf{s}_{k,\text{cvx}}, k \in \mathscr{K}$
4: Tighten convex approximation $\mathbf{F}_{\text{cvx}}$ in $\mathbf{s}_{k,\text{cvx}}, k \in \mathscr{K}$
5: **until** convergence

Note that line 3 of Algorithm 1 requires the major part of the computational cost. It involves solving a high-dimensional convex optimization problem, i.e. with dimension $NK$, where the number of users $N$ typically ranges between 2-100 and the number of tones $K$ can go up to 4000. For CA-DSB, this convex problem is as follows:

$$\max_{\mathbf{s}_k \in \mathscr{S}_k, k \in \mathscr{K}} \sum_{k \in \mathscr{K}} b_{k,\text{cvx}}(\mathbf{s}_k) \;\; \text{s.t.} \sum_{k \in \mathscr{K}} s_k^n \leq P^{n,\text{tot}}, \; n \in \mathscr{N} \;\; (\mathbf{F}_{\text{cvx}}) \qquad (3)$$

where $\mathscr{S}_k = \{\mathbf{s}_k \in \mathscr{R}^n : 0 \leq s_k^n \leq s_k^{n,\max}, n \in \mathscr{N}\}$ is a compact convex set with $s_k^{n,\max} := \min\{s_k^{n,\text{mask}}, P^{n,\text{tot}}\}$ and $P^{n,\text{tot}} < \infty$, and where $b_{k,\text{cvx}}(\mathbf{s}_k)$ is concave and given as:

$$\begin{aligned} b_{k,\text{cvx}}(\mathbf{s}_k) = & \sum_{n \in \mathscr{N}} w_n f_s \log_2\Big(\sum_{m \in \mathscr{N}} |\tilde{h}_k^{n,m}|^2 s_k^m + \Gamma \sigma_k^n\Big) \\ & - \sum_{n \in \mathscr{N}} w_n f_s \Big(\sum_{m \neq n} a_k^{m,n} s_k^m + c_k^n\Big), \end{aligned} \qquad (4)$$

and where $a_k^{n,m}, c_k^n, \forall n, m, k$ are constant approximation parameters, obtained by a closed-form formula during the approximation step (line 1 and 4 of Algorithm 1), and

$$|\tilde{h}_k^{n,m}|^2 \begin{cases} = \Gamma |h_k^{n,m}|^2, & n \neq m \\ = |h_k^{n,m}|^2, & n = m. \end{cases} \qquad (5)$$

The standard way of solving $\mathbf{F}_{\text{cvx}}$ (3) is via its dual problem formulation $\mathbf{F}_{\text{cvx,dual}}$, as shown in (6), which leads to the same solution because the duality gap is zero. The advantage of the dual formulation is that the dual objective function $g_{\text{cvx}}(\lambda)$ can be decomposed into independent subproblems $g_{k,\text{cvx}}(\lambda)$ for each tone $k$, which are much more simple to solve. The dual problem $\mathbf{F}_{\text{cvx,dual}}$, dual function $g_{\text{cvx}}(\lambda)$ and Lagrangian $\mathscr{L}_{k,\text{cvx}}(\mathbf{s}_k, \lambda)$, for tone $k$, are given as follows:

$$\min_{\lambda \geq 0} g_{\text{cvx}}(\lambda) \qquad (\mathbf{F}_{\text{cvx,dual}}) \qquad (6)$$

$$\text{with} \quad g_{\text{cvx}}(\lambda) = \sum_{k \in \mathscr{K}} g_{k,\text{cvx}}(\lambda) = \sum_{k \in \mathscr{K}} \max_{\mathbf{s}_k \in \mathscr{S}_k} \mathscr{L}_{k,\text{cvx}}(\mathbf{s}_k, \lambda)$$

$$\mathscr{L}_{k,\text{cvx}}(\mathbf{s}_k, \lambda) = b_{k,\text{cvx}}(\mathbf{s}_k) - \sum_{n \in \mathscr{N}} \lambda_n s_k^n + \sum_{n \in \mathscr{N}} \lambda_n P^{n,\text{tot}}/K$$

The dual problem $\mathbf{F}_{\text{cvx,dual}}$ is solved using the standard subgradient-based dual decomposition approach, as shown in Algorithm 2, so as to find the solution of the convex problem (line 3 of Algorithm 1). Note that $[x]^+$ denotes the projection of $x \in \mathscr{R}^N$ onto $\mathscr{R}_+^N$, and that the stepsize $\delta$ can be chosen using different procedures [2] [7], e.g. $\delta = q/t$ where $q$ is the initial stepsize and $t$ is the iteration counter. Note that line 4 of Algorithm 2 corresponds to solving $K$ independent convex subproblems of dimension $N$. This can be done by using state-of-the-art iterative methods (e.g. Newton's-method) or by using iterative fixed point updates [3] [4].

**Algorithm 2** Subgradient-based dual decomposition approach to solve $\mathbf{F}_{\mathrm{cvx}}$ for CA-DSB

1: $t := 1$
2: **repeat**
3: $\quad \lambda_n^{t+1} = \left[\lambda_n^t + \delta(\sum_{k\in\mathscr{K}} s_k^n - P^{n,\mathrm{tot}})\right]^+, \forall n \in \mathscr{N}$
4: $\quad \forall k : \tilde{\mathbf{s}}_k = \operatorname*{argmax}_{\mathbf{s}_k\in\mathscr{S}_k} \mathscr{L}_{k,\mathrm{cvx}}(\mathbf{s}_k, \lambda^{t+1})$
5: $\quad t := t+1$
6: **until** convergence (accuracy $\varepsilon$)
7: $\mathbf{s}_{k,\mathrm{cvx}} = \tilde{\mathbf{s}}_k,\ k \in \mathscr{K}$

## 5. IMPROVED DUAL DECOMPOSITION FOR DSL DSM

The subgradient update approach (Algorithm 2) to solve $\mathbf{F}_{\mathrm{cvx}}$, often exhibits a very slow convergence. Therefore we propose an improved approach that is inspired by recent advances in mathematical programming, in particular the proximal center method of [6]. Our improved scheme uses an optimal gradient-based scheme and automatically selects optimal stepsizes. Furthermore we prove that the proposed scheme improves on the convergence of subgradient based schemes by one order of magnitude, i.e. from $\mathscr{O}(\frac{1}{\varepsilon^2})$ to $\mathscr{O}(\frac{1}{\varepsilon})$, with the same computational complexity, resulting in much faster DSM algorithms.

The basic steps in this approach are as follows. First an approximated (smoothed) dual function $\bar{g}_{\mathrm{cvx}}(\lambda)$ is defined that can be chosen to be arbitrarily close to the original dual function $g_{\mathrm{cvx}}(\lambda)$. Then it is proven that this smoothed dual function $\bar{g}_{\mathrm{cvx}}$ is differentiable and has a Lipschitz continuous gradient, contrary to $g_{\mathrm{cvx}}(\lambda)$ that is non-differentiable and has no Lipschitz continuous gradient. Finally an optimal gradient scheme [5] is applied to the smoothed dual function $\bar{g}_{\mathrm{cvx}}(\lambda)$ leading to an efficiency estimate of the order $\mathscr{O}(\frac{1}{\varepsilon})$, i.e. one order of magnitude better than the subgradient based dual decomposition approach (Algorithm 2).

We introduce the following functions $d_k(\mathbf{s}_k)$ which are called prox-functions in [6] and are defined as follows:

**Definition 5.1.** *A prox-function $d_k(\mathbf{s}_k)$ has the following properties:*

- *$d_k(\mathbf{s}_k)$ is a non-negative continuous and strongly convex function with convexity parameter $\sigma_{\mathscr{S}_k}$*
- *$d_k(\mathbf{s}_k)$ is defined for the compact convex set $\mathscr{S}_k$*

An example of a valid prox-function is $d_k(\mathbf{s}_k) = \frac{1}{2}\|\mathbf{s}_k\|^2$. Since $\mathscr{S}_k, \forall k$, are compact and $d_k(\mathbf{s}_k)$ are continuous, we can choose finite and positive constants such that

$$D_{\mathscr{S}_k} \geq \max_{\mathbf{s}_k\in\mathscr{S}_k} d_k(\mathbf{s}_k), \forall k. \tag{7}$$

The prox-functions are used to smoothen the dual function $g_{\mathrm{cvx}}(\lambda)$ to obtain a smoothed dual function $\bar{g}_{\mathrm{cvx}}(\lambda)$ as follows:

$$\bar{g}_{\mathrm{cvx}}(\lambda) = \max_{\mathbf{s}_k\in\mathscr{S}_k, k\in\mathscr{K}} \sum_{k\in\mathscr{K}} \left\{ b_{k,\mathrm{cvx}}(\mathbf{s}_k) - \sum_{n\in\mathscr{N}} \lambda_n (s_k^n - \frac{P^{n,\mathrm{tot}}}{K}) - c d_k(\mathbf{s}_k) \right\}, \tag{8}$$

where $c$ is a positive smoothness parameter that will be defined later in this section. By using a sufficiently small value for $c$, the smoothed dual function can be arbitrarily close to the original dual function. Note that the particular choice of the prox-functions does not destroy the tone-separability of the objective function in (8).

Denote by $\bar{\mathbf{s}}_{k,\mathrm{cvx}}(\lambda), k \in \mathscr{K}$, the optimal solution of the maximization problem in (8). The following theorem describes the properties of the smoothed dual function $\bar{g}_{\mathrm{cvx}}(\lambda)$:

**Theorem 5.1** ([6])**.** *The function $\bar{g}_{\mathrm{cvx}}(\lambda)$ is convex and continuous differentiable at any $\lambda \in \mathscr{R}^n$. Moreover, its gradient $\nabla\bar{g}_{\mathrm{cvx}}(\lambda) = \sum_{k\in\mathscr{K}} \bar{\mathbf{s}}_{k,\mathrm{cvx}}(\lambda) - \mathbf{P}^{\mathrm{tot}}$ is Lipschitz continuous with Lipschitz constant $L_c = \sum_{k\in\mathscr{K}} \frac{1}{c\sigma_{\mathscr{S}_k}}$. The following inequalities also hold:*

$$\bar{g}_{\mathrm{cvx}}(\lambda) \leq g_{\mathrm{cvx}}(\lambda) \leq \bar{g}_{\mathrm{cvx}}(\lambda) + c \sum_{k\in\mathscr{K}} D_{\mathscr{S}_k} \qquad \forall\lambda \in \mathscr{R}^n. \tag{9}$$

The addition of the prox-functions thus leads to a convex differentiable dual function $\bar{g}_{\mathrm{cvx}}(\lambda)$ with Lipschitz continuous gradient. Now instead of solving the original dual problem (6), we will focus on the following problem:

$$\min_{\lambda\geq 0} \bar{g}_{\mathrm{cvx}}(\lambda). \tag{10}$$

Note that, by making $c$ sufficiently small in (8), the solution of (10) can be brought arbitrarily close to the solution of (6). This means that the solution of (6) can be found by solving (10), up to a certain accuracy determined by the choice of $c$. Taking the particular structure of (10) into account, i.e. a differentiable objective function with Lipschitz continuous gradient, we propose Algorithm 3, which is an optimal gradient-based scheme derived from [6] to solve (10).

**Algorithm 3** Improved dual decomposition scheme for (10)

1: $i := 0$, tmp $:= 0$
2: initialize $i_{\max}$, $\lambda^i$
3: **for** $i = 0 \ldots i_{\max}$ **do**
4: $\quad \forall k : \mathbf{s}_k^{i+1} = \operatorname*{argmax}_{\mathbf{s}_k\in\mathscr{S}_k} b_{k,\mathrm{cvx}}(\mathbf{s}_k) - \sum_{n\in\mathscr{N}} \lambda_n^i s_k^n - c d_k(\mathbf{s}_k)$
5: $\quad d\bar{g}_c^{i+1} = \sum_{k\in\mathscr{K}} \mathbf{s}_k^{i+1} - \mathbf{P}^{\mathrm{tot}}$
6: $\quad \mathbf{u}^{i+1} = [\frac{d\bar{g}_c^{i+1}}{L_c} + \lambda^i]^+$
7: $\quad \mathrm{tmp} := \mathrm{tmp} + \frac{i+1}{2} d\bar{g}_c^{i+1}$
8: $\quad \mathbf{v}^{i+1} = [\frac{\mathrm{tmp}}{L_c}]^+$
9: $\quad \lambda^{i+1} = \frac{i+1}{i+3}\mathbf{u}^{i+1} + \frac{2}{i+3}\mathbf{v}^{i+1}$
10: **end for**
11: Build $\hat{\lambda} = \lambda^{i_{\max}+1}$ and $\hat{\mathbf{s}}_k = \sum_{i=0}^{i_{\max}} \frac{2(i+1)}{(i_{\max}+1)(i_{\max}+2)} \mathbf{s}_k^{i+1}$

In Algorithm 3, the specific value for $L_c$ depends on the chosen prox-function $d_k(\mathbf{s}_k)$, as given in Theorem 5.1. The specific value for $c$ will be defined later in Theorem 5.2. The index $i$ refers to the iteration counter. Note that lines 5-9 of Algorithm 3 correspond to the improved Lagrange multiplier updates. By comparing this with the standard subgradient Lagrange multiplier update (line 3 of Algorithm 2), one can observe that the standard and improved updates require a similar complexity.

The remaining issue is to prove that $\hat{\mathbf{s}}_k, k \in \mathscr{K}$, in Algorithm 3, after $i_{\max}$ iterations has converged to an $\varepsilon$-optimal solution where $i_{\max}$ is of the order $\mathcal{O}(\frac{1}{\varepsilon})$. For this we define the following lemmas that will be used in the sequel.

**Lemma 5.1.** *For any $\mathbf{y} \in \mathscr{R}^n$ and $\mathbf{z} \geq 0$, the following inequality holds*[1]*:*

$$\mathbf{y}^T\mathbf{z} \leq \|[\mathbf{y}]^+\|\|\mathbf{z}\|. \tag{11}$$

*Proof.* Let us define the following index sets: $I^- = \{i \in \{1\ldots n\} : y_i < 0\}$ and $I^+ = \{i \in \{1\ldots n\} : y_i \geq 0\}$. Then,

$$\mathbf{y}^T\mathbf{z} = \sum_{i\in I^-} y_i z_i + \sum_{i\in I^+} y_i z_i \leq \sum_{i\in I^+} y_i z_i = ([\mathbf{y}]^+)^T\mathbf{z} \leq \|[\mathbf{y}]^+\|\|\mathbf{z}\|.$$

The last inequality follows from the Cauchy-Schwartz inequality. □

The following lemma gives a lower bound for the primal gap, $f^*_{\text{cvx}} - \sum_{k\in\mathscr{K}} b_{k,\text{cvx}}(\hat{\mathbf{s}}_k)$, of (3), where $f^*_{\text{cvx}}$ is the optimal objective value of (3),

**Lemma 5.2.** *Let $\lambda^*$ be any optimal Lagrange multiplier, then for any $\hat{\mathbf{s}}_k \in \mathscr{S}_k, \forall k$, the following lower bound on the primal gap holds:*

$$f^*_{\text{cvx}} - \sum_{k\in\mathscr{K}} b_{k,\text{cvx}}(\hat{\mathbf{s}}_k) \geq -\|\lambda^*\|\|[\sum_{k\in\mathscr{K}} \hat{\mathbf{s}}_k - \mathbf{P}^{\text{tot}}]^+\|. \tag{12}$$

*Proof.* From the assumptions of the lemma we have

$$\begin{aligned} f^*_{\text{cvx}} &= \max_{\mathbf{s}_k\in\mathscr{S}_k, k\in\mathscr{K}} \sum_{k\in\mathscr{K}} b_{k,\text{cvx}}(\mathbf{s}_k) - \lambda^{*T}(\sum_{k\in\mathscr{K}} \mathbf{s}_k - \mathbf{P}^{\text{tot}}) \\ &\geq \sum_{k\in\mathscr{K}} b_{k,\text{cvx}}(\hat{\mathbf{s}}_k) - \lambda^{*T}(\sum_{k\in\mathscr{K}} \hat{\mathbf{s}}_k - \mathbf{P}^{\text{tot}}). \end{aligned} \tag{13}$$

Formula (12) is then obtained by applying Lemma 5.1. □

A consequence of Lemma 5.2 is that if $\|[\sum_{k\in\mathscr{K}} \hat{\mathbf{s}}_k - \mathbf{P}^{\text{tot}}]^+\| \leq \varepsilon_c$, then the primal gap is bounded: for all $\hat{\lambda} \in \mathscr{R}^N_+$

$$-\varepsilon_c\|\lambda^*\| \leq f^*_{\text{cvx}} - \sum_{k\in\mathscr{K}} b_{k,\text{cvx}}(\hat{\mathbf{s}}_k) \leq g_{\text{cvx}}(\hat{\lambda}) - \sum_{k\in\mathscr{K}} b_{k,\text{cvx}}(\hat{\mathbf{s}}_k). \tag{14}$$

Therefore, if we are able to derive an upper bound $\varepsilon$ for the dual gap, $g_{\text{cvx}}(\hat{\lambda}) - \sum_{k\in\mathscr{K}} b_{k,\text{cvx}}(\hat{\mathbf{s}}_k)$, and an upper bound $\varepsilon_c$ for the coupling constraints for some given $\hat{\lambda}$ ($\geq 0$) and $\hat{\mathbf{s}}_k \in \mathscr{S}_k, \forall k$, then we conclude that $\hat{\mathbf{s}}_k$ is an $(\varepsilon, \varepsilon_c)$-solution for $\mathbf{F}_{\text{cvx}}$ (since in this case $-\varepsilon_c\|\lambda^*\| \leq f^*_{\text{cvx}} - \sum_{k\in\mathscr{K}} b_{k,\text{cvx}}(\hat{\mathbf{s}}_k) \leq \varepsilon$). The next theorem derives these upper bounds for Algorithm 3 and provides a specific value for $c$.

**Theorem 5.2.** *Let $\lambda^*$ be an optimal Lagrange multiplier, taking $c = \frac{\varepsilon}{\sum_{k\in\mathscr{K}} D_{\mathscr{S}_k}}$ and $i_{\max} + 1 = 2\sqrt{(\sum_k \frac{1}{\sigma_{\mathscr{S}_k}})(\sum_k D_{\mathscr{S}_k})}\frac{1}{\varepsilon}$, then after $i_{\max}$ iterations, Algorithm 3 obtains an approximate solution $\hat{\mathbf{s}}_k, \forall k \in \mathscr{K}$, to the convex approximation (3) with a duality gap less than $\varepsilon$, i.e.*

$$g_{\text{cvx}}(\hat{\lambda}) - \sum_{k\in\mathscr{K}} b_{k,\text{cvx}}(\hat{\mathbf{s}}_k) \leq \varepsilon, \tag{15}$$

*and the constraints satisfy:*

$$\|[\sum_k \hat{\mathbf{s}}_k - \mathbf{P}^{\text{tot}}]^+\| \leq \varepsilon(\|\lambda^*\| + \sqrt{\|\lambda^*\|^2 + 2}). \tag{16}$$

[1]For the sake of an easy exposition we consider only the Euclidian norm $\|\cdot\|$. Other norms can also be used (see [6] for a detailed exposition).

*Proof.* Using a similar reasoning as in the proof of Theorem 3.4 in [6] we can show that for any $c$ the following inequality holds:

$$\begin{aligned} \bar{g}_{\text{cvx}}(\hat{\lambda}) \leq \min_{\lambda\geq 0}\Big\{ &\tfrac{2L_c}{(i_{\max}+1)^2}\|\lambda\|^2 \\ &+ \sum_{i=0}^{i_{\max}} \tfrac{2(i+1)}{(i_{\max}+1)(i_{\max}+2)}[\bar{g}_{\text{cvx}}(\lambda^i) + (\nabla\bar{g}_{\text{cvx}}(\lambda^i))^T(\lambda - \lambda^i)]\Big\} \end{aligned}$$

By replacing $\bar{g}_{\text{cvx}}(\lambda^i)$ and $\nabla\bar{g}_{\text{cvx}}(\lambda^i)$ with their expressions given in (8) and Theorem 5.1, respectively and taking into account the functions $b_{k,\text{cvx}}$ are concave, we obtain the following inequality:

$$\begin{aligned} &g_{\text{cvx}}(\hat{\lambda}) - \sum_{k\in\mathscr{K}} b_{k,\text{cvx}}(\hat{\mathbf{s}}_k) \\ &\leq c(\sum_{k\in\mathscr{K}} D_{\mathscr{S}_k}) + \min_{\lambda\geq 0}\Big\{\tfrac{2L_c}{(i_{\max}+1)^2}\|\lambda\|^2 - \langle\lambda, \sum_k \hat{\mathbf{s}}_k - \mathbf{P}^{\text{tot}}\rangle\Big\} \\ &= c(\sum_{k\in\mathscr{K}} D_{\mathscr{S}_k}) - \tfrac{(i_{\max}+1)^2}{8L_c}\|[\sum_k \hat{\mathbf{s}}_k - \mathbf{P}^{\text{tot}}]^+\|^2 \leq c(\sum_{k\in\mathscr{K}} D_{\mathscr{S}_k}). \end{aligned}$$

Therefore taking $c$ as in the theorem we obtain (15). For the constraints using Lemma 5.2 and the previous inequality we get that $\|[\sum_k \hat{\mathbf{s}}_k - \mathbf{P}^{\text{tot}}]^+\|$ satisfies the second order inequality in $y$: $\frac{(i_{\max}+1)^2}{8L_c}y^2 - \|\lambda^*\|y - \varepsilon \leq 0$. Therefore, $\|[\sum_k \hat{\mathbf{s}}_k - \mathbf{P}^{\text{tot}}]^+\|$ must be less than the largest root of the corresponding second-order equation, i.e.

$$\|[\sum_k \hat{\mathbf{s}}_k - \mathbf{P}^{\text{tot}}]^+\| \leq \big(\|\lambda^*\| + \sqrt{\|\lambda^*\|^2 + \tfrac{\varepsilon(i_{\max}+1)^2}{2L_c}}\big)\tfrac{4L_c}{(i_{\max}+1)^2}.$$

Taking $i_{\max}$ as defined in the theorem, we also get (16). □

From Theorem 5.2 we can conclude that by taking $c = \frac{\varepsilon}{\sum_{k\in\mathscr{K}} D_{\mathscr{S}_k}}$, Algorithm 3 converges to a solution of $\mathbf{F}_{\text{cvx}}$ with duality gap less than $\varepsilon$ and the constraints violation satisfy $\|[\sum_k \hat{\mathbf{s}}_k - \mathbf{P}^{\text{tot}}]^+\| \leq \varepsilon(\|\lambda^*\| + \sqrt{\|\lambda^*\|^2 + 2})$ after $i_{\max} = 2\sqrt{(\sum_k \frac{1}{\sigma_{\mathscr{S}_k}})(\sum_k D_{\mathscr{S}_k})}\frac{1}{\varepsilon}$ iterations, i.e. the efficiency estimate of our scheme is of the order $\mathcal{O}(\frac{1}{\varepsilon})$, one order of magnitude better than the standard subgradient based method that has an efficiency estimate of the order $\mathcal{O}(\frac{1}{\varepsilon^2})$.

Note that this approach is fully automatic, i.e. it does not require any stepsize tuning as in the subgradient approach, which is known to be a very difficult and crucial process. For our proposed scheme, one decides on the required accuracy $\varepsilon$ and then simply executes the algorithm. The extension of CA-DSB with the improved dual decomposition approach will be referred to as I-CA-DSB, i.e. improved CA-DSB.

A final remark on Algorithm 3 is that the independent convex 'per-tone' problems (line 4) are slightly modified with respect to the standard 'per-tone' problems $g_{k,\text{cvx}}$. This is a consequence of the addition of the extra prox-function term. One can use state-of-the-art iterative methods (e.g. Newton's-method) to solve these convex subproblems with guaranteed convergence. Another popular method consists in using an iterative fixed point update approach, as this is shown to work well with very small complexity and can easily be extended to a distributed implementation [4] [3]. The fixed point update formula for the transmit powers $s_k^n$ used by CA-DSB can be adapted to take the extra prox-term into account. Following the same procedure as explained in [3], we obtain the following transmit power update formula, that only differs in the presence of the term PROX:

$$s_k^n = \Bigg[\Bigg(\frac{w_n f_s/\log(2)}{\lambda_n + \underbrace{2cs_k^n}_{\text{PROX}} + \sum_{m\neq n} \omega_m f_s a_k^{n,m} - \sum_{m\neq n} \frac{w_m f_s \Gamma |h_k^{m,n}|^2/\log(2)}{\sum_p |\tilde{h}_k^{m,p}|^2 s_k^p + \Gamma\sigma_k^m}}\Bigg) - \frac{\sum_{m\neq n} \Gamma |h_k^{n,m}|^2 s_k^m + \Gamma\sigma_k^n}{|h_k^{n,n}|^2}\Bigg]^+ . \quad (17)$$

By iteratively updating the transmit powers $s_k^n$ using (17) over all users and tones, a fast convergence to the solution of the convex subproblem is achieved (line 4 of Algorithm 3). Providing convergence conditions for these types of iterative fixed point updates is outside the scope of this paper. In [3, 9, 10], limited convergence results are given for these types of iterative updates. Although convergence is only proven under certain conditions, one always observes convergence for realistic DSL scenarios as shown in [3, 9, 10].

Finally note that the improved dual decomposition approach can straightforwardly be extended to more general DSM problem formulations that incorporate energy-awareness, such as those presented in [11] (green DSL).

## 6. SIMULATION RESULTS

Simulations are performed for a near-far CO-RT (central office - remote terminal) scenario, consisting of a CO-line with length 5000m, a RT-line with length 3000m, and a CO-RT distance of 4000m. In Figure 1 the convergence behaviour is compared for the improved scheme (Algorithm 3) and the subgradient scheme (Algorithm 2), where convergence is defined as achieving the optimal dual value within accuracy 0.05%. For the subgradient scheme we used the stepsize update rule $\delta = q/t$, where $q$ is the initial stepsize and $t$ is the iteration counter. Different initial stepsizes $q$ lead to a different convergence behaviour and this is generally difficult to tune. Note that for all initial stepsizes, the subgradient scheme is still far from convergence after 500 iterations. The improved scheme, on the other hand, automatically tunes its stepsize and converges very rapidly in only 40 iterations.

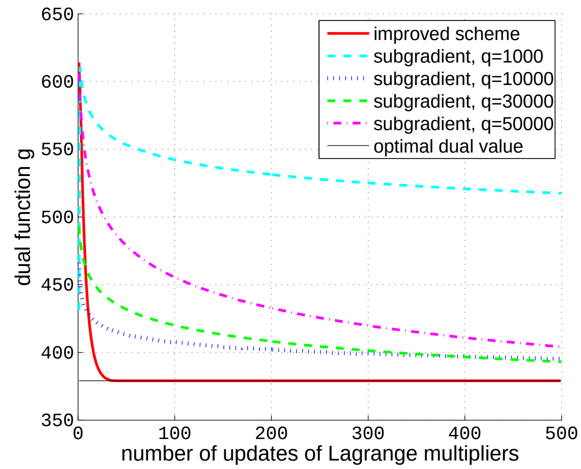


Figure 1: Comparison of convergence behaviour between subgradient schemes, with different initial stepsizes $q$, and the improved scheme.

## 7. CONCLUSION

Dynamic spectrum management has the potential to dramatically increase the data rates in current DSL broadband access networks. State-of-the-art DSM algorithms use an iterative convex approximation approach to tackle the corresponding nonconvex optimization problems, but rely on a subgradient-based dual decomposition approach that is known to exhibit slow convergence. This paper leverages on recent advances in mathematical programming to obtain a novel dual decomposition approach that improves on the convergence of the standard subgradient approach by one order of magnitude. The proposed approach makes an important step towards obtaining numerically fast and effective DSM algorithms.

**Acknowledgment.** This research work was carried out at the ESAT laboratory of Katholieke Universiteit Leuven, in the frame of K.U.Leuven Research Council: GOA AMBioRICS, CoE EF/05/006, OT/03/12, FWO projects: G.0235.07, G.0499.04, G.0211.05, G.0226.06, G.0302.07; Research communities (ICCoS, ANMMM, MLDM); AWI: BIL/05/43; Belgian Federal Science Policy Office: IUAP DYSCO.

## REFERENCES


[1] K.B. Song, S.T. Chung, G. Ginis, J.M. Cioffi, "Dynamic spectrum management for next-generation DSL systems," *IEEE Comm. Magazine*, vol. 40, no. 10, pp. 101–109, Oct. 2002.

[2] P. Tsiaflakis, J. Vangorp, M. Moonen, J. Verlinden, "A low complexity optimal spectrum balancing algorithm for DSL," *Sign. Proc.*, vol. 87, no. 7, pp. 1735–1753, July 2007.

[3] P. Tsiaflakis, M. Diehl, M. Moonen, "Distributed Spectrum Management Algorithms for Multiuser DSL Networks," *IEEE Trans. on Sign. Proc.*, vol. 56, no. 10, pp. 4825–4843, Oct 2008.

[4] J. Papandriopoulos, J.S. Evans, "Low-complexity distributed algorithms for spectrum balancing in multi-user DSL networks," in *IEEE Int. Conf. on Comm.*, vol. 7, June 2006, pp. 3270–3275.

[5] Y. Nesterov, *Introductory Lectures on Convex Optimization: A Basic Course*. Kluwer Boston, 2004.

[6] I. Necoara, J.A.K. Suykens, "Application of a smoothing technique to decomposition in convex optimization," *IEEE Trans. on Automatic Control*, vol. 11, no. 53, pp. 2674–2679, 2008.

[7] W. Yu and R. Lui, "Dual methods for nonconvex spectrum optimization of multicarrier systems," *IEEE Trans. on Comm.*, vol. 54, no. 7, July 2006.

[8] M. Chiang, C.W. Tan, D.P. Palomar, D. O'Neill, D. Julian, "Power control by geometric programming," *IEEE Trans. on Wireless Comm.*, vol. 6, no. 7, pp. 2640–2651, July 2007.

[9] R. Cendrillon, J. Huang, M. Chiang, M. Moonen, "Autonomous spectrum balancing for digital subscriber lines," *IEEE Trans. on Sign. Proc.*, vol. 55, no. 8, pp. 4241–4257, August 2007.

[10] W. Yu, G. Ginis, J. Cioffi, "Distributed multiuser power control for digital subscriber lines," *IEEE J. Sel. Area. Comm.*, vol. 20, no. 5, pp. 1105–1115, Jun. 2002.

[11] P. Tsiaflakis, Y. Yi, M. Chiang, and M. Moonen, "Green DSL: Energy-efficient DSM," in *Proceedings of IEEE ICC*, 2009.