KU Leuven

Departement Elektrotechniek

ESAT-STADIUS/TR 13-211

Dynamic Spectrum Management with Spherical Coordinates [1]

Rodrigo B. Moraes, Martin Wolkerstorfer, Paschalis Tsiaflakis and Marc Moonen[2]

Submitted for publication, IEEE Transactions on Signal Processing
December 2013

[1]This report is available by anonymous ftp from *ftp.esat.kuleuven.ac.be* in the directory *pub/sista/rmoraes/reports/13-211.pdf*

[2]K.U.Leuven, Dept. of Electrical Engineering (ESAT), Research group STADIUS, Kastelpark Arenberg 10, 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: *rodrigo.moraes@esat.kuleuven.ac.be*. This research work was carried out at the ESAT Laboratory of the KU Leuven, in the frame of the KU Leuven Research Council CoE EF/05/006 Optimization in Engineering (OPTEC) and PFV/10/002 (OPTEC); Concerted Research Action GOA-MaNet and the Belgian Programme on Interuniversity Attraction Poles initiated by the Belgian Federal Science Policy Office IUAP P7/23 (Belgian network on Stochastic modeling, analysis, design and optimization of communication systems, BESTCOM, 2012-2017). The scientific responsibility is assumed by its authors.

Dynamic Spectrum Management with Spherical Coordinates

Rodrigo B. Moraes, Martin Wolkerstorfer, Paschalis Tsiaflakis and Marc Moonen

***Abstract*— Multiuser interference, i.e. crosstalk, is the main bottleneck for digital subscriber lines (DSL) technology. Dynamic spectrum management (DSM) mitigates crosstalk by focusing on the multi-user power/frequency resource allocation problem, and it can provide formidable gains in performance. In this paper, we look at the DSM problem from a different perspective. We formulate the problem with the power allocation vectors defined with spherical coordinates, i.e. as a function of a radius and angles. We see that this reformulation permits us to exploit structure in the problem. We propose two algorithms. In the first of them, we use the fact that the DSM problem is concave in the radial dimension and perform an exhaustive search for the angles. The second algorithm uses a block coordinate descent approach, i.e. a sequence of line searches. We show that there is structure to be found in the radial dimension (it is always concave) and in the angle dimensions. For the latter, we provide conditions for the line searches to be concave or convex for each of the angles. The fact that we use structure leads to large savings in computational complexity. For example, we see that our first algorithm can be up to 60 times faster than a corresponding previously proposed algorithm. Our second algorithm is 2-15 times faster than a relevant previously proposed algorithm.**

***Index Terms*—SPC-DSLP, SPC-INTF, SPC-MULT.**

I. Introduction

In a communications system where multiple users have competing utilities, the intelligent allocation of the system resources offers the system designer means to significantly improve the network performance. With proper resource allocation, the competing users can be coordinated such that the transmission of each user is designed so as to maximize its own utility while being as little detrimental as possible to the transmission of all other users. The system designer can count on a wide range of options so as to perform this coordination, such as the dimensions of power, code, space, frequency, time and waveform.

In this paper, we focus on the dimensions of frequency and power. More specifically, we treat a multi-user, multitone interference channel (IC) where the goal is to judiciously allocate power throughout frequency such that the weighted rate sum (WRS) of the users is maximized. Each user is subject to a power constraint (PC), which complicates the problem further.

The applications of this problem are numerous, and include both wireless and wireline systems. For the latter, the continued research activities to find efficient and high performance solutions to the power/frequency resource allocation problem is referred to as dynamic spectrum management (DSM), and its focus is usually on digital subscriber line (DSL) networks.[1]

DSL counts with a share of more than 70% of the broadband access market worldwide, with a total of more than 450 million subscribers [2]. DSL has been coping well with the increasing demand for higher data rates and with the competition from optical fiber. Today, it is generally accepted that DSL will be around for decades to come [3]. In this technology, transmission is done over twisted copper pairs (i.e. a DSL line). Typically, one dedicated DSL line serves one user, and several such lines are collected in quantities of up to 60 in a cable binder. Due to electromagnetic coupling, a signal transmitted in a given line leaks to the neighboring lines. Therein lies the effect of multi-user interference, more commonly known as crosstalk. Crosstalk has been repeatedly identified as the main bottleneck in DSL transmission.

In the past ten to fifteen years extensive theoretical research [4]–[19] has shown that managing crosstalk with DSM leads to formidable gains in performance. In this paper, we propose two algorithms for the solution of the WRS maximization problem in DSL. Both algorithms exploit structure in order to save on computational complexity. Our approach is based on a change of variables. For a network of N users, the classical way to represent the decision variables of the problem is to define a vector $\mathbf{p}^k = \begin{bmatrix} p_1^k & \cdots & p_N^k \end{bmatrix}^{\mathrm{T}} \in \mathbb{R}_+^N$ for each tone (i.e. sub-channel) k, where p_n^k is the power allocated for user n on tone k. The problem then consists in optimizing $\mathbf{p}^k$ for every tone while respecting the per-user PCs. For the first algorithm, we change the decision variables as follows: we re-write $\mathbf{p}^k$ in spherical coordinates, i.e. $\mathbf{p}^k = \rho^k \mathbf{d}^k$, with

A preliminary version of this paper was presented at the IEEE International Conference on Communications (ICC), Budapest, Hungary, in June 2013.

This research work was carried out at the ESAT Laboratory of the KU Leuven, in the frame of the KU Leuven Research Council CoE EF/05/006 Optimization in Engineering (OPTEC) and PFV/10/002 (OPTEC); Concerted Research Action GOA-MaNet and the Belgian Programme on Interuniversity Attraction Poles initiated by the Belgian Federal Science Policy Office IUAP P7/23 (Belgian network on Stochastic modeling, analysis, design and optimization of communication systems, BESTCOM, 2012-2017). The scientific responsibility is assumed by its authors.

The Competence Center FTW Forschungszentrum Telekommunikation Wien GmbH is funded within the program COMET—Competence Centers for Excellent Technologies by BMVIT, BMWA and the City of Vienna. The COMET program is managed by the FFG.

R. B. Moraes, P. Tsiaflakis and M. Moonen are with the Department of Electrical Engineering (ESAT, STADIUS Center for Dynamical Systems, Signal Processing and Data Analytics), KU Leuven, Kasteelpark Arenberg 10, B-3001 Leuven, Belgium (e-mail: rodrigo.moraes@esat.kuleuven.be; paschalis.tsiaflakis@esat.kuleuven.be; and marc.moonen@esat.kuleuven.be). P. Tsiaflakis is also a postdoctoral fellow funded by the Research Foundation—Flanders (FWO). M. Wolkerstorfer is with the FTW Forschungszentrum Telekommunikation Wien GmbH, Donau-City-Straße 1, A-1220 Vienna, Austria (e-mail: wolkerstorfer@ftw.at).

[1] DSM is also classically recognized as a signal level coordination paradigm, see e.g. [1]

$\|\mathbf{d}^k\|_2 = 1$. Here $\|\cdot\|_2$ is the ℓ_2 or Euclidean norm, ρ^k is the radius and $\mathbf{d}^k$ is a direction vector—if $N = 2$, we have $\mathbf{d}^k = \begin{bmatrix}\cos\theta^k & \sin\theta^k\end{bmatrix}^{\mathrm{T}}$, $\theta^k \in [0, \pi/2]$. By observing that the WRS maximization problem is concave in the radius ρ^k, we propose an algorithm that optimizes the power allocation with an exhaustive search on the direction vector while concurrently optimizing for the radius by line search. The benchmark here is the optimal spectrum balancing (OSB) algorithm [11], which is built upon similar concepts but does an exhaustive search on the original, Cartesian coordinates vector. Because we exploit the structure in the radial dimension, we save considerably on computational complexity. We observe that our proposal can be 60 times faster than OSB.

For the second algorithm, we do yet another change of variables. We re-write the decision variables as $\eta^k \mathbf{v}^k$, where $\|\mathbf{v}^k\|_1 = 1$. Here $\|\cdot\|_1$ is the ℓ_1 or taxicab norm, η^k is the radius and $\mathbf{v}^k$ is a direction vector—if $N = 2$, we have $\mathbf{v}^k = \begin{bmatrix}(1-\phi^k) & \phi^k\end{bmatrix}^{\mathrm{T}}$, $\phi^k \in [0, 1]$. This can be interpreted as spherical coordinates in taxicab geometry [20] (as opposed to Euclidean geometry). We show that with this coordinate system it is easier to use a block coordinate descent method, i.e. a sequence of line searches. As in the Euclidean spherical coordinates, concavity on the radial dimension still holds. We show that there is also structure to be exploited in the angle dimensions, i.e. the variables in the direction vector $\mathbf{v}^k$: we identify situations where the line searches for each of the angles are concave or convex. The structure in both the radial and in the angle dimensions are exploited in the second proposed algorithm with good savings in computational complexity. The benchmark here is the iterative spectrum balancing (ISB) algorithm [12], which also uses a block coordinate descent method but does not exploit structure. Through numerous simulations, it is observed that our algorithm performs at least as well as ISB while being 2-15 times faster.

This paper is organized as follows: in Section II we formalize the notation, present the problem in mathematical form and briefly discuss previous work. In Section III we present the first proposed algorithm, followed by analyses of computational complexity and precision. In Section IV-D we present the second algorithm, along with an exposition on how to explore structure on the radial and on the angle dimensions. Section V presents some numerical experiments and finally Section VI presents final remarks.

In this paper, we use lower-case boldface letters to denote vectors, upper-case boldface letters for matrices and calligraphic letters for sets (for example, $\mathbf{a}$, $\mathbf{A}$ and $\mathcal{A}$). We also use $\mathbb{R}_+$ as the set of non-negative real numbers, $\|\cdot\|_2$ as the Euclidean or ℓ_2 norm, $\|\cdot\|_1$ as the taxicab or ℓ_1 norm, $\left|\cdot\right|$ as either absolute value or cardinality of a set, $(\cdot)^{\mathrm{T}}$ as transpose, $(\cdot)^*$ as the conjugate, $\mathrm{E}\left[\cdot\right]$ as expectation, $\log(\cdot)$ as the natural logarithm, $\log_{10}(\cdot)$ as the base 10 logarithm and $\mathrm{unif}(a, b)$ as a uniform random variable in the interval $[a, b]$.

II. System Model and Previous Work

Consider an N user discrete multitone (DMT) system with K Δ_f-spaced tones. We define the set of users as $\mathcal{N} = \{1,\ 2,\ \ldots,\ N\}$ and the set of tones as $\mathcal{K} = \{1,\ \ldots,\ K\}$. Let $\mathbf{P} = \{p_n^k\} \in \mathbb{R}_+^{K\times N}$ be a matrix in which p_n^k is the transmit power of user n on tone k. We also define $\mathbf{p}^k \in \mathbb{R}_+^N$ as the vector containing the powers of all users on tone k, i.e. $\mathbf{p}^k = \begin{bmatrix}p_1^k & \ldots & p_N^k\end{bmatrix}^{\mathrm{T}} \in \mathbb{R}_+^N$ and $\mathbf{p}_n \in \mathbb{R}_+^K$ as the vector containing the powers of user n over all tones, i.e. $\mathbf{p}_n = \begin{bmatrix}p_n^1 & \ldots & p_n^K\end{bmatrix}^{\mathrm{T}} \in \mathbb{R}_+^K$. Let $h_{n,j}^k$ be the channel gain between the transmitter of user j and receiver of user n on tone k. The received signal for user n on tone k is given by

$$y_n^k = h_{n,n}^k x_n^k + \sum_{\substack{j\in\mathcal{N}\\ j\neq n}} h_{n,j}^k x_j^k + z_n^k. \tag{1}$$

Here, we consider the simplifying assumption of perfect DMT block synchronization between users [18], [21], [22]. Also, in our scenario every user operates with single input, single output (SISO) transmission.[2] In (1), x_n^k and y_n^k are respectively the transmitted and received symbols for user n on tone k and z_n^k represents circularly symmetric zero mean complex Gaussian noise. We define $p_n^k = \mathrm{E}\left[x_n^k (x_n^k)^*\right]$ as the transmit power and $\tilde{\sigma}_n^k = \mathrm{E}\left[z_n^k (z_n^k)^*\right]$ as the Gaussian noise power, both relating to user n on tone k.

In this paper, we consider all interference to be Gaussian noise. The data rate for user n on tone k is given by

$$b_n^k = \log\left(1 + \frac{p_n^k}{\sigma_n^k + \sum_{\substack{j\in\mathcal{N}\\ j\neq n}} \alpha_{n,j}^k p_j^k}\right), \tag{2}$$

where $\alpha_{n,j}^k = \Gamma\left|h_{n,j}^k\right|^2 \left|h_{n,n}^k\right|^{-2}$ is the normalized interference channel gain from user j to user n on tone k and $\sigma_n^k = \Gamma\tilde{\sigma}_n^k \left|h_{n,n}^k\right|^{-2}$ is the normalized Gaussian noise power. Also Γ accounts for the SNR gap to capacity, the noise margin and the coding gain [24]. In this paper, we consider continuous data rates. The data rate of user n in bits per second is given by $r_n = {}^{f_s}\!/_{\log(2)} \sum_{k\in\mathcal{K}} b_n^k$, where f_s is the symbol rate.

The problem we focus on is that of maximizing the WRS of the participating users in the network while respecting their per-user PCs. Mathematically, we write

$$\begin{aligned} \max_{\mathbf{P}} \quad & \sum_{n\in\mathcal{N}} \sum_{k\in\mathcal{K}} u_n b_n^k \\ \text{subject to} \quad & \sum_{k\in\mathcal{K}} p_n^k \leq P_n^{\max},\ n \in \mathcal{N} \end{aligned} \tag{3}$$

Here, $P_n^{\max}$ is the PC for user n and the u_n are weights assigned to the users. We call (3) the *DSM problem*. It can be shown that this problem is NP-hard [5].

As mentioned in the introduction, several works have focused on the DSM problem. Refs. [4]–[9] focus primarily on theoretical analyses. These papers are important because they attempt to find structure in the DSM problem, which in turn can be used in the algorithms. In [4], [5] perhaps the most important characteristic of the DSM problem is rigorously formulated: it is established that, although the original problem is non-concave, its duality gap vanishes as the number of tones K increases to infinity. In the same vein, [6] provides an estimate of the duality gap. Ref. [7] provides conditions for the optimal solution of the problem to have a frequency division

[2] Recently there have been efforts to consider the DSM problem in a multiple input, multiple output (MIMO) setting, see e.g. [17], [23]

characteristic. In [8] some situations that allow for concave representation are identified. Some papers use a leakage or spillage criterium [25], [26]. In [9], it is shown that if the interference channel gains are weak enough, the DSM problem can be solved as a geometric program (GP).

Regarding the algorithms, several proposals are available, e.g. [10]–[17]. The algorithms in [11] and [12] are of special interest to this paper. In [11], the optimal spectrum balancing (OSB) algorithm is presented. This algorithm formulates the Lagrange dual of (3) and performs a per-tone exhaustive search for the powers, i.e. it exhaustively finds a $\mathbf{p}^k$ that maximizes the per-tone Lagrangean. This per-tone exhaustive search is done on an N-dimensional grid. If the axes are sampled each with Q points, the total grid has Q^N points. The Lagrangean is then calculated for all grid points and the point that maximizes it is picked. On the minus side the applicability of OSB is hindered for large networks due to the exponentially increasing computational complexity (the grid size increases exponentially with N). On the plus side the two main elements of OSB (the per-tone exhaustive search and the vanishing duality gap in problems with large number of tones) make it approach optimality.

Of the lower computational complexity algorithms proposed so far the iterative spectrum balancing (ISB) algorithm [12] is a well-known example. In this algorithm, the objective function is maximized with a block coordinate descent method, i.e. when optimizing for a given user n, all other users have their powers fixed. The process repeats until convergence. Unlike the algorithms in [13], [14], [18], ISB does not do any approximation of the objective function. ISB has been shown time and again to perform very close to OSB for small and medium scenarios with only a fraction of the computational complexity.

III. DSM with Spherical Coordinates—Exhaustive Search for the Angles

As in [11], we write the Lagrangean of (3) as

$$L(\mathbf{P},\boldsymbol{\lambda}) = \sum_{n\in\mathcal{N}}\sum_{k\in\mathcal{K}} u_n b_n^k - \sum_{n\in\mathcal{N}} \lambda_n \big(\sum_{k\in\mathcal{K}} p_n^k - P_n^{\max}\big). \quad (4)$$

Here $\boldsymbol{\lambda} = \begin{bmatrix}\lambda_1 & \dots & \lambda_N\end{bmatrix}^{\mathrm{T}} \in \mathbb{R}_+^N$ is the vector of Lagrange multipliers associated with the per-user PCs. We formulate the dual problem as

$$\min_{\boldsymbol{\lambda}} \max_{\mathbf{P}} \quad L(\mathbf{P},\boldsymbol{\lambda}),$$

where appropriate values for $\boldsymbol{\lambda}$ should be searched for so that the PCs are respected.

By re-writing (4) as

$$L(\mathbf{P},\boldsymbol{\lambda}) = \sum_{k\in\mathcal{K}} L(\mathbf{p}^k,\boldsymbol{\lambda}) + \sum_{n\in\mathcal{N}} \lambda_n P_n^{\max},$$

where

$$L(\mathbf{p}^k,\boldsymbol{\lambda}) = \sum_{n\in\mathcal{N}} u_n b_n^k - \sum_{n\in\mathcal{N}} \lambda_n p_n^k, \quad (5)$$

we can see that the per-tone maximization of $L(\mathbf{p}^k,\boldsymbol{\lambda})$ also leads to the maximization of (4). We can thus focus on the per-tone maximization of $L(\mathbf{p}^k,\boldsymbol{\lambda})$.

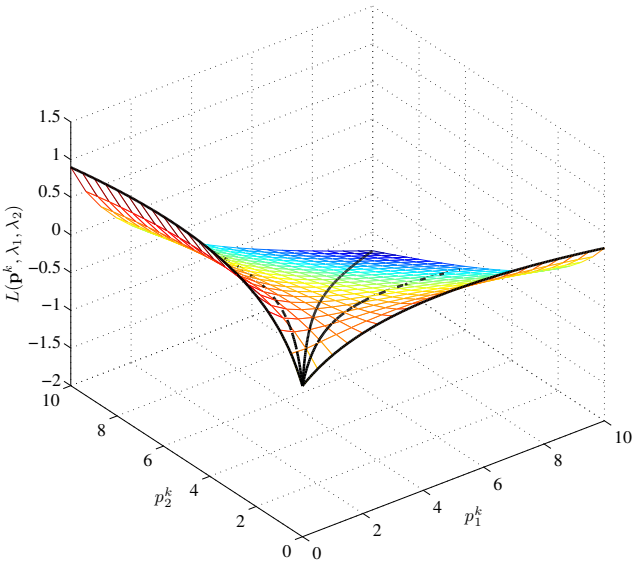

Fig. 1. Illustration of non-concavity of $L(\mathbf{p}^k,\boldsymbol{\lambda})$ for a two user case. We choose $u_1 = u_2 = 0.5$, $\alpha_{1,2}^k = \alpha_{2,1}^k = 1$, $\sigma_1^k = 2$, $\sigma_2^k = 1$, $\lambda_1 = \lambda_2 = 0.15$.

As is well known, the function $L(\mathbf{p}^k,\boldsymbol{\lambda})$ is not concave in $\mathbf{p}^k$ in general. When interference gains $\alpha_{n,j}^k$ are very small, it can be that $L(\mathbf{p}^k,\boldsymbol{\lambda})$ is concave. If interference is sufficiently strong, then $L(\mathbf{p}^k,\boldsymbol{\lambda})$ is maximized with only one active user [7]. We give an example of the latter case in Fig. 1, where $N = 2$. To find the $\mathbf{p}^k$ that maximizes $L(\mathbf{p}^k,\boldsymbol{\lambda})$ for the $N = 2$ case, OSB discretizes the p_1^k-p_2^k plane with Q^2 points, calculates $L(\mathbf{p}^k,\boldsymbol{\lambda})$ on the grid points and picks the best one.

Our approach is different. The first step is to do a change of variables. We write $\mathbf{p}^k$ in spherical coordinates, i.e.

$$\mathbf{p}^k = \rho^k \mathbf{d}^k, \ \rho^k \geq 0, \ \|\mathbf{d}^k\|_2 = 1, \quad (6)$$

where ρ^k is the radius and $\mathbf{d}^k = \begin{bmatrix} d_1^k & \cdots & d_N^k \end{bmatrix} \in \mathbb{R}_+^N$ is a direction vector. Eq. (6) describes the positive quadrant of an N-dimensional sphere. As a more concrete example, consider that $N = 3$. Eq. (6) is written as

$$\mathbf{p}^k = \rho^k \begin{bmatrix} \cos\theta_{[1]}^k \cos\theta_{[2]}^k \\ \cos\theta_{[1]}^k \sin\theta_{[2]}^k \\ \sin\theta_{[1]}^k \end{bmatrix}, \ \rho^k \geq 0, \ \theta_{[1]}^k, \theta_{[2]}^k \in \Big[0, \frac{\pi}{2}\Big]. \quad (7)$$

We write the spherical coordinates representation of a general N-dimensional vector in Appendix A in (28), (29) and (30).

We make two important remarks about (6). First, it is very natural to associate the variables of the DSM problem with users, i.e. p_1^k for user 1 until p_N^k for user N, and literally all previous work has done so.[3] This is not the case with the spherical coordinates. In (7) for example, by changing the radius ρ^k, all user's powers change (in the Cartesian vector). By changing the angle $\theta_{[1]}^k$, the powers of users 1 and 2 change. This motivates our choice not to use subscripts for ρ^k and to use bracketed subscripts for the angles, e.g. $\theta_{[2]}^k$ and $\theta_{[2]}^k$ in (7). For the angles, the bracketed subscripts are best interpreted as directions, not users. Second, notice that

[3] An exception is [23], where spherical coordinates have been used to solve the WRS maximization problem with per-transceiver PCs in a MIMO setting.

although $\mathbf{d}^k$ is N-dimensional, it has only $N-1$ free variables, i.e. $\theta^k_{[1]},\ \theta^k_{[2]},\ \ldots, \theta^k_{[N-1]}$ (see, for example, (7) or, for the general case, (28)-(30)). This is a direct consequence of the ℓ_2 norm constraint in (6).

We continue by redefining some formulas. We re-write (2) and (5) as

$$b^k_n = \log\left(1 + \frac{\rho^k d^k_n}{\sigma^k_n + \sum_{\substack{j\in\mathcal{N}\\ j\neq n}} \alpha^k_{n,j}\rho^k d^k_j}\right) \tag{8}$$

$$L(\rho^k, \boldsymbol{\theta}^k, \boldsymbol{\lambda}) = \sum_{n\in\mathcal{N}} u_n b^k_n - \sum_{n\in\mathcal{N}} \lambda_n \rho^k d^k_n.$$

Here we define $\boldsymbol{\theta}^k = \begin{bmatrix}\theta^k_{[1]} & \ldots & \theta^k_{[N-1]}\end{bmatrix}^{\mathrm{T}} \in \mathbb{R}^{N-1}_+$, $\theta^k_{[i]} \in \left[0, \pi/2\right]$, $i = 1, \ldots, N-1$.

The advantage of using spherical coordinates is that, while keeping $\mathbf{d}^k$ fixed, there is structure to be found on the radial dimension.

Proposition 1. *For a fixed direction $\mathbf{d}^k$, $L(\rho^k, \boldsymbol{\theta}^k, \boldsymbol{\lambda})$ is strictly concave in ρ^k.*

Proof: It suffices to calculate the second derivative in ρ^k and show that

$$\frac{\partial^2 L(\rho^k, \boldsymbol{\theta}^k, \boldsymbol{\lambda})}{\partial(\rho^k)^2} = -\sum_{n\in\mathcal{N}} \frac{u_n d^k_n \sigma^k_n \left(2I^k_n \rho^k (d^k_n + I^k_n) + 2I^k_n \sigma^k_n + d^k_n \sigma^k_n\right)}{\left(I^k_n \rho^k + \sigma^k_n\right)^2 \left((I^k_n + d^k_n)\rho^k + \sigma^k_n\right)^2} < 0. \tag{9}$$

Here

$$I^k_n = \sum_{\substack{j\in\mathcal{N}\\ j\neq n}} \alpha^k_{n,j} d^k_j. \tag{10}$$

Since $d^k_n \geq 0\ \forall n$, all variables in (9) are non-negative. Hence each term of the summation in n is non-negative. Because $\|\mathbf{d}^k\| = 1$, at least one term in the sum is strictly positive. Hence the sum is strictly positive. Because of the minus sign, the second derivative is, for $\rho^k \geq 0$, always strictly negative. ∎

The consequence of Proposition 1 is illustrated in Fig. 1. In the figure, the dark lines show $L(\rho^k, \theta^k, \boldsymbol{\lambda})$ for fixed angles $\theta^k = \{0, \pi/8, \pi/4, \pi/6, \pi/2\}$. One interesting way to interpret this result is the following: for $\theta^k = 0$ and $\theta^k = \pi/2$, it is well known that $L(\rho^k, \theta^k, \boldsymbol{\lambda})$ is concave in ρ^k. After all, these two cases represent situations with a single user. With Proposition 1 we generalize concavity for all other angles $\theta^k \in [0, \pi/2]$.

A. Algorithm

Because of the concavity in the radial dimension, we can save considerably on computational complexity. In this section, we describe a per-tone exhaustive search algorithm that uses the spherical coordinates formulation. Our strategy consists of an exhaustive search only in the variables of the direction vector, i.e. in $\theta^k_{[1]}, \theta^k_{[2]}, \ldots, \theta^k_{[N-1]}$. We construct an $(N-1)$-dimensional grid with each of the continuous axes (in $[0, \pi/2]$) sampled with Q points. A point in this grid corresponds to a vector $\mathbf{d}^k$, hence a direction. For a fixed direction, since optimizing for the radius is a concave line search problem, we write $L(\hat{\rho}^k(\mathbf{d}^k), \mathbf{d}^k, \boldsymbol{\lambda})$, where $\hat{\rho}^k(\mathbf{d}^k)$ is the optimal radius for $\mathbf{d}^k$.

Algorithm 1: OSB-SC

1 Initialize $\boldsymbol{\lambda}$;
2 $\mathcal{D} \leftarrow$ discretized θ-space, with $\theta_{[i]} \in [0, 1] \forall i$;
3 **repeat**
4 Obtain $\hat{\rho}^k(\boldsymbol{\theta}^k)$, $\boldsymbol{\theta}^k \in \mathcal{D},\ k \in \mathcal{K}$;
5 $\hat{\boldsymbol{\theta}}^k = \mathrm{argmax}_{\boldsymbol{\theta}^k\in\mathcal{D}}\ L(\hat{\rho}^k(\boldsymbol{\theta}^k), \boldsymbol{\theta}^k, \boldsymbol{\lambda})\ k \in \mathcal{K}$;
6 $\mathbf{p}^k = \hat{\rho}^k(\boldsymbol{\theta}^k)\mathbf{d}^k, k \in \mathcal{K}$;
7 $\lambda_n = \max\left(\lambda_n + \epsilon(\sum_{k\in\mathcal{K}} p^k_n - P^{\max}_n), 0\right),\ n \in \mathcal{N}$;
8 **until** *until convergence*

A pseudocode is provided in Algorithm 1. We name the algorithm OSB with spherical coordinates (OSB-SC). In line 2, the $(N-1)$-dimensional vector space $\begin{bmatrix}\theta_{[1]} & \cdots & \theta_{[N-1]}\end{bmatrix}^{\mathrm{T}}$, with $\theta_{[i]} \in [0, \pi/2]\ \forall i$, is sampled with Q points on each axis. It is natural to do so uniformly in when all axes are in dB scale [11]. In line 4 we calculate the optimal radius for each fixed direction and fixed $\boldsymbol{\lambda}$. Since this is a one dimensional concave problem, we can solve it with the Newton method. In line 5 the exhaustive search is performed and in line 6 we transform back to Cartesian coordinates. In line 7, the Lagrange multipliers are adjusted (ϵ is a step size). The process repeats until convergence.

B. Complexity

The computational complexity of the OSB-SC algorithm is dominated by the exhaustive search for the angles. Since there is one separate search for each tone, the computational complexity is given by $O(KQ^{N-1})$. As a comparison, OSB has computational complexity given by $O(KQ^N)$.

It should be noted that both OSB and OSB-SC are restricted to cases with small N. For $N > 5$, both algorithms become intractable. This is a direct consequence of the fact that the computational complexity of both algorithms grow exponentially with N. However, with OSB-SC the search is in one less dimension, which means a significant computational complexity reduction.

C. Precision

Empirically we observe that OSB-SC is more precise than OSB. To understand why this is the case, consider Fig. 2. Here, we depict for a two user case the sampling of the $\bar{p}^k_1$-$\bar{p}^k_2$ plane for both algorithms. In the figure, both axes are in dB scale. For this section we use a bar for variables that are represented in dB scale, i.e. $\bar{p}^k_n$. Variables in linear scale are represented without a bar. We use $Q = 5$. OSB samples the space with 25 points in total, shown as the dots in the figure. OSB-SC samples the angle dimension with five angles in total, represented by the vectors $\mathbf{d}_1$ to $\mathbf{d}_5$. While for OSB the search space is restricted to the dots, OSB-SC can use all points on the lines. This is so because the radial dimension is not discretized. The lines cover the continuous plane on average better than the dots. We can quantify this with the following proposition.

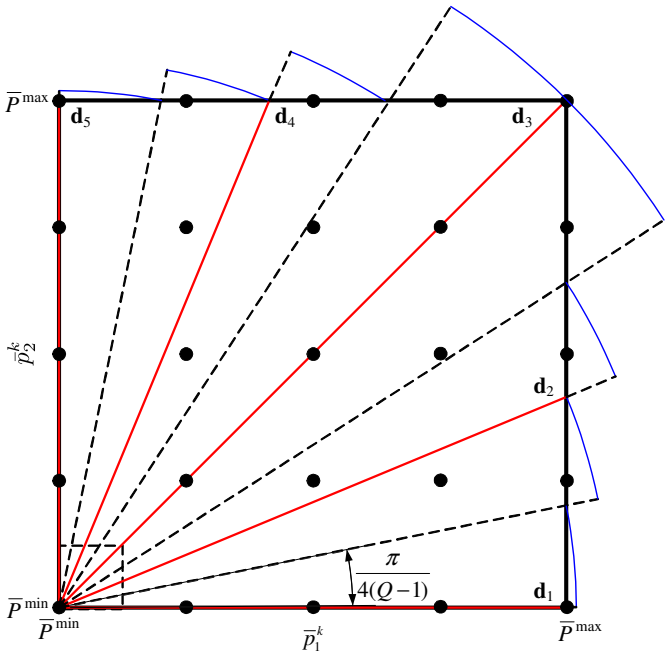

Fig. 2. Per-tone distribution of Q^2 points (for OSB) and Q lines (for OSB-SC) for $Q = 5$. The lines are represented by $\mathbf{d}_1, \ldots, \mathbf{d}_5$. Both axes are in dB scale. The dotted lines demarcate the set of points closer to one line. The square on the lower left corner represent the set of points closer to the point at the origin.

Proposition 2. *Consider $N = 2$. Consider two uniformly distributed, independent random vectors $\bar{\mathbf{p}}_1 \in \mathbb{R}^K$ and $\bar{\mathbf{p}}_2 \in \mathbb{R}^K$ with probability distribution function (pdf) defined by*

$$f(\bar{\mathbf{p}}_n) = \begin{cases} \text{constant}, & \bar{p}_n^k \geq \bar{P}^{\min} \text{ and } \sum\limits_{k\in\mathcal{K}} 10^{\bar{p}_n^k/10} \leq P^{\max} \\ 0, & \text{otherwise} \end{cases} \tag{11}$$

for $n = 1$ and $n = 2$ (for simplicity we consider that both users have the same $P^{\max}$). Define $\bar{P}^{\max} = 10\log_{10} P^{\max}$ with $\bar{P}^{\max} > \bar{P}^{\min}$. For each tone, the $\bar{p}_1^k$-$\bar{p}_2^k$ plane is sampled with Q^2 uniformly distributed grid points and Q uniformly spaced lines. The collection of grid points is represented by the set $\mathcal{Y} = \{\mathbf{y}_1, \ldots, \mathbf{y}_{Q^2}\}$, $\mathbf{y}_i \in \mathbb{R}^2$ $\forall i$ and the lines are represented by $\mathcal{D} = \{\mathbf{d}_1, \ldots, \mathbf{d}_Q\}$, $\mathbf{d}_i \in \mathbb{R}^2$ $\forall i$. Now we define our measure for precision. Consider, respectively, the average squared distance from the random variables to the grid points and from the random variables to the lines.

$$\mathrm{E}\left[d^2_{points}\right] = \mathrm{E}\left[\frac{1}{K}\sum_{k\in\mathcal{K}} \min_{\mathbf{y}\in\mathcal{Y}} \left\|\mathbf{y} - \bar{\mathbf{p}}^k\right\|_2^2\right] \tag{12}$$

$$\mathrm{E}\left[d^2_{lines}\right] = \mathrm{E}\left[\frac{1}{K}\sum_{k\in\mathcal{K}} \min_{\mathbf{d}\in\mathcal{D}} \left\|(\mathbf{d}^{\mathrm{T}}\bar{\mathbf{p}}^k)\mathbf{d} - \bar{\mathbf{p}}^k\right\|_2^2\right]. \tag{13}$$

Here $\bar{\mathbf{p}}^k = \begin{bmatrix}\bar{p}_1^k & \bar{p}_2^k\end{bmatrix}^{\mathrm{T}}$ and, in (13), $(\mathbf{d}^{\mathrm{T}}\bar{\mathbf{p}}^k)$ is the optimal radius for a given $\bar{\mathbf{p}}^k$ (i.e. the projection of $\bar{\mathbf{p}}^k$ in $\mathbf{d}$). Eq. (12) can be upper bounded by

$$\mathrm{E}\left[d^2_{points}\right] \leq \frac{\bar{P}^2}{6(Q-1)^2} \tag{14}$$

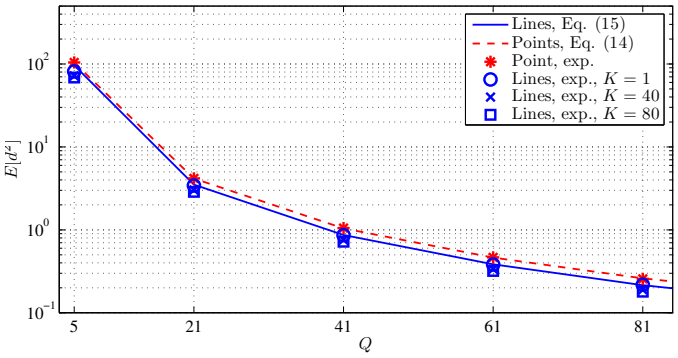

Fig. 3. Average squared distance from random variables to point and lines, with results from both theory and Monte Carlo simulation. Data pertaining to the experiments are identified with 'exp'. For the points, results for different values of K are practically the same, so only one result per Q is shown.

and (13) is upper bounded by

$$\mathrm{E}\left[d^2_{lines}\right] < \bar{P}^2\left(\frac{1}{4} - \frac{Q-1}{2\pi}\sin\left(\frac{\pi}{2(Q-1)}\right)\right) \times \frac{\sum_{i=0}^{Q-2}\cos^{-4}\left(\frac{(i+1)\pi}{4(Q-1)}\right)}{\sum_{i=0}^{Q-2}\cos^{-2}\left(\frac{(i+1)\pi}{4(Q-1)}\right)}, \tag{15}$$

where $\bar{P} = \bar{P}^{\max} - \bar{P}^{\min}$.

The derivations of (14) and (15) are given in Appendix B. The derivation of (14) is straightforward if we decorrelate the per-tone variables. To arrive at (15), we do two relaxations: First, we decorrelate the per-tone variables and, second, we extend the radial dimension to ease the calculation of an integral.

To demonstrate Eqs. (14) and (15), we perform a Monte Carlo simulation. We generate random variables as in (11), a grid with Q^2 equally spaced points and a set of Q equally spaced lines. For the random variables, we consider, without loss of generality, $\bar{P}^{\min} = 0$ and $\bar{P}^{\max} = 100$ (i.e. a per-tone random variable $\bar{p}_n^k$ can have values in the range $[0, 100]$). We calculate the squared distance from the random variables to the points and to the lines and average them over 10^4 realizations. The results of the experiment are given in Fig. (3) (indicated with 'exp'), along with the theoretical results of (14) and (15), for different values of Q and of K. We make two remarks about this experiment. First, we find that the lines (i.e. the OSB-SC grid) are more precise than the points (i.e. the OSB grid). Second, the error for the lines decreases as K increases, while the error for the points remains approximately the same (this is why only two curves are shown for the points in Fig. 3, one curve with the experimental results and one with the upper bound in (14)). That is so because, since the tones are coupled through a total PC, the more tones there are, the less power each individual tone has on average. The error becomes smaller because the lines cover the lower values of power better than larger ones, whereas the points have no spatial preference. This can be seen in Fig. 2, where in the lower left corner of the figure the coverage of the lines is at its best.

We can also look at the complexity/precision tradeoff from a different perspective: we can compare OSB and OSB-SC with

the same computational complexity and measure precision with (14) and (15). Towards this end, we use Q^2 points for OSB and Q^2 lines for OSB-SC. By doing so, the time complexity of the two algorithms is approximately the same, but OSB-SC is significantly more precise. With, say $Q = 40$, we obtain $\mathrm{E}\big[d^2_{\text{points}}\big] \leq 1.0417$ and $\mathrm{E}\big[d^2_{\text{lines}}\big] < 4.8582 \times 10^{-4}$.

Although Proposition 2 is restricted to the $N = 2$ case (the analysis for larger cases is too complicated), we have experimental evidence that point to the fact that the qualitative conclusions of the this section hold for $N > 2$ as well.

IV. DSM with Spherical Coordinates in Taxicab Geometry—Iterative Search for the Angles

Even tough OSB-SC is faster than OSB, the fact remains that it is too complex for large number of users. In view of this, this section presents a lower complexity algorithm that still maintains good performance. Our starting point is to solve the problem with a block coordinate descent method, i.e. instead of jointly optimizing for all variables, we perform a sequence of line searches for each variable at a time. It is an approach that is used, e.g. in the ISB algorithm [12]. To emphasize the differences between our proposal and ISB, we briefly describe the latter in the following paragraph.

Consider the per-tone Lagrangean as a function of, say, p_1^k, and for fixed λ_1

$$L(p_1^k, \lambda_1) = u_1 b_1^k + \sum_{\substack{j \in \mathcal{N} \\ j \neq 1}} u_j b_j^k - \lambda_1 p_1^k. \tag{16}$$

Maximizing (16) in p_1^k corresponds to a so-called *difference of convex* (DC) programming structure: the term b_1^k is concave in p_1^k, while the remaining b_j^k's are convex. The DC structure complicates the problem significantly and, as a consequence, $L(p_1^k, \lambda_1)$ can have multiple local optima. To find the global optimum, the ISB algorithm performs an exhaustive line search in p_1^k and picks the point that maximizes $L(p_1^k, \lambda_1)$. Alternatively, the optimal p_1^k can also be found by writing the stationary condition of (16), i.e. $\partial L(p_1^k, \lambda_1)/\partial p_1^k = 0$. This results in a polynomial of degree $2N-1$. We can find the roots of the polynomial, discard the non-positive ones and calculate the per-tone Lagrangean for the remaining roots and for the border points $p_n^k = 0$ and $p_n^k = P_n^{\max}$. We then pick the point that maximizes the per-tone Lagrangean [27]. Note that ISB does not do approximations to facilitate the line searches. After solving for all variables (i.e. after maximizing $L(p_1^k, \lambda_1)$ in p_1^k until $L(p_N^k, \lambda_N)$ in p_N^k) and all tones we obtain an updated $\mathbf{P}$. An outer loop should then search for appropriate Lagrange multipliers.

For our proposal, we change the decision variables as follows:

$$\mathbf{p}^k = \eta^k \mathbf{v}^k, \ \eta^k \geq 0, \ \|\mathbf{v}^k\|_1 = 1. \tag{17}$$

This describes the positive quadrant of an N-dimensional sphere in taxicab geometry [20]. In (17), η^k is the radius and $\mathbf{v}^k = \begin{bmatrix} v_1^k & \cdots & v_N^k \end{bmatrix} \in \mathbb{R}_+^N$ is a direction vector. As an example, consider $N = 3$. Eq. (17) is written as

$$\mathbf{p}^k = \eta^k \begin{bmatrix} (1-\phi_{[1]}^k)(1-\phi_{[2]}^k) \\ (1-\phi_{[1]}^k)\phi_{[2]}^k \\ \phi_{[1]}^k \end{bmatrix}, \ \eta^k \geq 0, \ \phi_{[1]}^k, \phi_{[2]}^k \in [0,1]. \tag{18}$$

The spherical coordinates in taxicab geometry of a general N-dimensional vector is provided in Appendix A in (31), (32) and (33). We remark that, as in (6), although the vector $\mathbf{v}^k$ is N-dimensional, it has only $N-1$ free variables, i.e. $\phi_{[1]}^k, \ \phi_{[2]}^k, \ \cdots, \ \phi_{[N-1]}^k$.

With (17) in hands, we re-define b_n^k (similarly to (8), just substitute ρ^k with η^k and d_n^k with v_n^k). Next, define $\boldsymbol{\eta} = [\eta^1 \ \cdots \ \eta^K]^{\mathrm{T}} \in \mathbb{R}_+^K$ and $\boldsymbol{\phi} = [\boldsymbol{\phi}_{[1]} \ \cdots \ \boldsymbol{\phi}_{[N]}] \in \mathbb{R}_+^{N-1 \times K}$, where $\boldsymbol{\phi}_{[i]} \in \mathbb{R}_+^K$ is defined similarly to $\boldsymbol{\eta}$, $i = 1, \ldots, N-1$. We re-formulate the DSM problem as

$$\begin{aligned} \max_{\boldsymbol{\eta}, \boldsymbol{\phi}} \ & \sum_{n \in \mathcal{N}} \sum_{k \in \mathcal{K}} u_n b_n^k \\ \text{subject to} \ & \sum_{k \in \mathcal{K}} \eta^k \leq \sum_{n \in \mathcal{N}} P_n^{\max} \\ & \sum_{k \in \mathcal{K}} \eta^k v_n^k \leq P_n^{\max}, \ n \in \mathcal{N} \end{aligned} \tag{19}$$

Here we substitute the N PCs in (3) with their equivalent versions in taxicab geometry. Also, we add a constraint to the sum power (the constraint in $\boldsymbol{\eta}$). In total there are $N+1$ constraints.

The Lagrangean of (19) is given by

$$\begin{aligned} L(\boldsymbol{\eta}, \boldsymbol{\phi}, \bar{\lambda}, \boldsymbol{\lambda}) = & \sum_{n \in \mathcal{N}} \sum_{k \in \mathcal{K}} u_n b_n^k - \bar{\lambda}\Big(\sum_{k \in \mathcal{K}} \eta^k - \sum_{n \in \mathcal{N}} P_n^{\max}\Big) \\ & - \sum_{n \in \mathcal{N}} \lambda_n \Big(\sum_{k \in \mathcal{K}} \eta^k v_n^k - P_n^{\max}\Big). \end{aligned} \tag{20}$$

Here, there are $N+1$ Lagrange multipliers: $\bar{\lambda}$ for the sum power and, as in (4), $\boldsymbol{\lambda} \in \mathbb{R}_+^N$ for the per-user PCs.

Our strategy is to maximize (20) for each of the variables of the problem separately.

The advantage of the taxicab spherical coordinates compared to the Euclidean spherical coordinates is that it is easier to control power for each user with the former than with the latter. Notice that if we change one of the $\phi_{[i]}^k$ in (18), the sum power does not change. This makes it possible to solve first for the radius and then for the angles. In contrast to that, the Euclidean spherical coordinates have some non-linearity to them: by changing one of the angles, the sum power changes.

The advantage of the taxicab spherical coordinates compared to the Cartesian coordinates is that it is easier to find the structure of the problem with them. In the remaining of this section, we show three ways in which we find structure in (19). First, similarly to the results in Section III, we show that the problem is concave in the radial dimension. Second, we show that there is some structure to be found in the angle dimensions too. Given one of the angle variables, we provide conditions for the line search to be concave or convex. Third, we show that, since after the solution for $\boldsymbol{\eta}$ each tone has a sum power constraint, some tones for some angles can be ignored if the sum power constraint is already exhausted.

We detail each of these ways to find structure in the next sections.

A. Solving for the radius

To solve for the radius $\boldsymbol{\eta}$ for fixed $\boldsymbol{\phi}$, $\bar{\lambda}$ and $\boldsymbol{\lambda}$, we write

$$\max_{\boldsymbol{\eta}} L(\boldsymbol{\eta}, \boldsymbol{\phi}, \bar{\lambda}, \boldsymbol{\lambda}),$$

which can be decomposed and solved for each tone separately. The per-tone Lagrangean is given by (here we emphasize the dependency with η^k and $\bar{\lambda}$)

$$L(\eta^k, \bar{\lambda}) = \sum_{n\in\mathcal{N}} u_n b_n^k - \bar{\lambda}\eta^k - \eta^k \sum_{n\in\mathcal{N}} \lambda_n v_n^k.$$

Here, for every tone concavity holds. The demonstration of this fact is almost identical to Proposition 1—in (9) and (10), we only need to replace all d_n^k by v_n^k. Hence, there is no DC structure and we can find the optimal radius with a low complexity line search algorithm.

B. Solving for the angles

To solve for the angle $\boldsymbol{\phi}_{[i]}$, for fixed $\boldsymbol{\eta}$, $\boldsymbol{\phi}_{[j]},\ j\neq i$, $\bar{\lambda}$ and $\boldsymbol{\lambda}$, we write

$$\max_{\boldsymbol{\phi}_{[i]}} L(\boldsymbol{\eta}, \boldsymbol{\phi}, \bar{\lambda}, \boldsymbol{\lambda}),$$

which can be decomposed and solved for each tone separately. The per-tone Lagrangean is given by (here we emphasize the dependency with $\phi_{[i]}^k$ and $\boldsymbol{\lambda}$)

$$L(\phi_{[i]}^k, \boldsymbol{\lambda}) = \sum_{n\in\mathcal{N}} u_n b_n^k - \sum_{n\in\mathcal{N}} \lambda_n \eta^k v_n^k. \tag{21}$$

When we maximize (21) in $\phi_{[i]}^k$, we notice that *given the right conditions, it is possible avoid the DC structure.* To see that, first define

$$b_n^k = \log\left(1 + \frac{a_n^k \phi_{[i]}^k + D_n^k}{c_n^k \phi_{[i]}^k + F_n^k}\right) \tag{22}$$

Here we re-write the formula for the data rate with the emphasis on its dependency on $\phi_{[i]}^k$. In the numerator of the fraction, we write $\eta^k v_n^k = a_n^k \phi_{[i]}^k + D_n^k$ and in the denominator, we write $\sigma_n^k + \sum_{j\neq n} \alpha_{n,j}^k \eta^k v_j^k = c_n^k \phi_{[i]}^k + F_n^k$. If, for example, $N=3$, $n=1$, and $i=1$, we have $a_n^k = -\eta^k(1-\phi_{[2]}^k)$, $D_n^k = \eta^k(1-\phi_{[2]}^k)$, $c_n^k = -\eta^k\alpha_{1,2}^k\phi_{[2]}^k + \eta^k\alpha_{1,3}^k$ and $F_n^k = \sigma_1^k + \eta^k\alpha_{1,2}^k\phi_{[2]}^k$ (see (18)). Notice that all of these are real. It is important to notice that a_n^k and c_n^k can be positive or negative. A simple look-up table is sufficient for calculating these variables.

Now consider the second derivative of (22) in $\phi_{[i]}^k$.

$$\frac{\partial^2 b_n^k}{\partial(\phi_{[i]}^k)^2} = u_n\big(c_n^k D_n^k - a_n^k F_n^k\big) \times \frac{\big(2a_n^k c_n^k + 2(c_n^k)^2\big)\phi_{[i]}^k + a_n^k F_n^k + 2c_n^k F_n^k + c_n^k D_n^k}{(c_n^k\phi_{[i]}^k + F_n^k)^2\big((a_n^k + c_n^k)\phi_{[i]}^k + F_n^k + D_n^k\big)^2} \tag{23}$$

Here we see that the sign of the second derivative is determined by a simple linear function, i.e.

$$g_n^k(\phi_{[i]}^k) = A_n^k \phi_{[i]}^k + B_n^k \tag{24}$$

where

$$A_n^k \triangleq \operatorname{sgn}\big\{(c_n^k D_n^k - a_n^k F_n^k)\big\}\big(2a_n^k c_n^k + 2(c_n^k)^2\big)$$
$$B_n^k \triangleq \operatorname{sgn}\big\{\big((c_n^k D_n^k - a_n^k F_n^k)\big)\big\}\big(a_n^k F_n^k + 2c_n^k F_n^k + c_n^k D_n^k\big)$$

Here $\operatorname{sgn}\{\cdot\}$ is the sign function. The variables A_n^k and B_n^k can be either positive or negative. Hence, given n, k and i, it can be that b_n^k is concave or convex in $\phi_{[i]}^k \in [0,1]$, depending on the sign of (24) in the interval $[0,1]$. This stands in contrast with the case of the Cartesian coordinates and ISB, where, when solving for p_1^k, b_1^k is *always* concave and the b_j^k's, $j\neq 1$, are *always* convex. With the taxicab spherical coordinates, things are more flexible.

To illustrate this flexibility, consider an example with $N=2$. We identify the variables a_n^k, D_n^k, c_n^k and F_n^k.[4] The flexibility lies in the fact that it can be *both* that b_1^k and b_2^k are concave in ϕ^k. A sufficient condition for this is

$$g_n^k(0) \leq 0\ \forall n \rightarrow B_1^k \leq 0 \ \text{ and } \ B_2^k \leq 0$$
$$g_n^k(1) \leq 0\ \forall n \rightarrow A_1^k + B_1^k \leq 0 \ \text{ and } \ A_2^k + B_2^k \leq 0$$

Here, we just check the sign of (24) in the points 0 and 1. After substituting the appropriate values and some manipulations, we find that, if

$$\alpha_{1,2}^k \leq \frac{\sigma_1^k}{2\sigma_1^k + \eta^k} \ \text{ and } \ \alpha_{2,1}^k \leq \frac{\sigma_2^k}{2\sigma_2^k + \eta^k} \tag{25}$$

then b_1^k and b_2^k are concave in ϕ^k. Because the sum of concave function is also concave, (21) is concave. If the conditions in (25) are satisfied, there is no DC structure.

The conditions in (25) add insight and a clear contrast to the case with the Cartesian coordinates. However, they do not exploit all the structure there is. For the N user case, a stronger sufficient condition for (21) to be *concave* in $\phi_{[i]}^k$ is given by the following proposition.

Proposition 3. *If*

$$\sum_{n\in\mathcal{N}} \max\Big(\frac{\partial^2 u_n b_n^k}{\partial(\phi_{[i]}^k)^2}(0),\ \frac{\partial^2 u_n b_n^k}{\partial(\phi_{[i]}^k)^2}(1)\Big) \leq 0 \tag{26}$$

then the per-tone Lagrangean in (21) *is* concave *in* $\phi_{[i]}^k$.

In the same vein, a sufficient condition for (21) to be *convex* in $\phi_{[i]}^k$ is given by the following proposition.

Proposition 4. *If*

$$\sum_{n\in\mathcal{N}} \min\Big(\frac{\partial^2 u_n b_n^k}{\partial(\phi_{[i]}^k)^2}(0),\ \frac{\partial^2 u_n b_n^k}{\partial(\phi_{[i]}^k)^2}(1)\Big) \geq 0 \tag{27}$$

then the per-tone Lagrangean in (21) *is* convex *in* $\phi_{[i]}^k$.

The proofs of Propositions 3 and 4 are given in Appendix C. Its main steps consist of relaxing the second derivative of (21) and simply checking that (23) is either a monotonically increasing or a monotonically decreasing function in $\phi_{[i]}^k$.

[4] For user 1, $a_1^k = -\eta^k$, $D_1^k = \eta^k$, $c_1^k = \eta^k\alpha_{1,2}^k$ and $F_1^k = \sigma_1^k$. For user 2, $a_2^k = \eta^k$, $D_2^k = 0$, $c_2^k = -\eta^k\alpha_{2,1}^k$ and $F_2^k = \eta^k\alpha_{2,1}^k\sigma_2^k$.

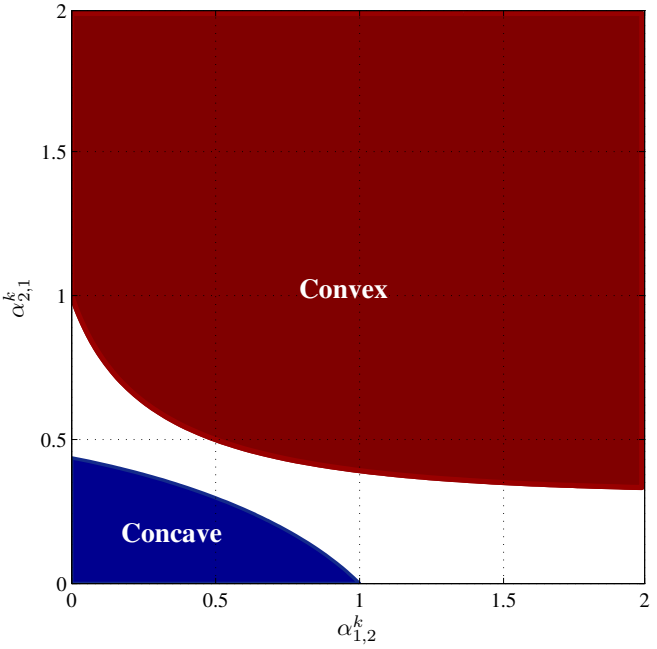

Fig. 4. Depiction of concave and convex regions of the per-tone Lagrangean for a $N = 2$ case as a function of the normalized interference channel coefficients. We choose $u_1 = u_2 = 0.5$, $\eta^k = 1$, $\sigma_1^k = 2$ and $\sigma_2^k = 1$.

Algorithm 2: TaSSO

1 $\boldsymbol{\lambda} = \mathbf{0}_N$;
2 **repeat**
3 **repeat**
4 Solve for $\boldsymbol{\eta}$ with Newton method;
5 Adjust $\bar{\lambda}$;
6 **until** *until convergence*
7 **for** $i = 1,\ \ldots,\ N-1$ **do**
8 Calculate a_n^k, D_n^k, c_n^k and F_n^k $\forall n, k$;
9 $\mathcal{K}_* = \big\{k | k \in \mathcal{K},\ \eta^k > 0 \text{ and } \phi^k_{[j]} \neq 1,\ j < i\big\}$;
10 $\mathcal{K}_{\text{cnc}} = \big\{k \in \mathcal{K}_* | \sum_n \max\big(\frac{\partial^2 u_n b_n^k}{\partial (\phi^k_{[i]})^2}(0), \frac{\partial^2 u_n b_n^k}{\partial (\phi^k_{[i]})^2}(1)\big)\big\} \leq 0$;
11 $\mathcal{K}_{\text{cvx}} = \{k \in \mathcal{K}_* | \sum_n \min\big(\frac{\partial^2 u_n b_n^k}{\partial (\phi^k_{[i]})^2}(0), \frac{\partial^2 u_n b_n^k}{\partial (\phi^k_{[i]})^2}(1)\big)\big\} \geq 0$;
12 $\mathcal{K}_{\text{neither}} = \mathcal{K}_* - \big(\mathcal{K}_{\text{cvx}} + \mathcal{K}_{\text{cnc}}\big)$;
13 **for** $k \in \mathcal{K}_{\text{cnc}}$ **do**
14 Solve with Newton method;
15 **for** $k \in \mathcal{K}_{\text{cvx}}$ **do**
16 Calculate $L(0, \boldsymbol{\lambda})$ and $L(1, \boldsymbol{\lambda})$ in (21);
17 **if** $L(0, \boldsymbol{\lambda}) > L(1, \boldsymbol{\lambda})$ **then** $\phi^k_{[i]} = 0$, **else** $\phi^k_{[i]} = 1$;
18 **for** $k \in \mathcal{K}_{\text{neither}}$ **do**
19 Solve line search;
20 $\mathbf{p}^k = \eta^k \mathbf{v}^k,\ k \in \mathcal{K}$;
21 $\lambda_n = \max\big(\lambda_n + \epsilon(\sum_{k \in \mathcal{K}} p_n^k - P_n^{\max}),\ 0\big),\ n \in \mathcal{N}$;
22 **until** *until convergence*

To make these two propositions more palpable, we return to the example with $N = 2$. By applying (26) and (27), we find regions where the maximization of (21) is concave and convex. These regions are illustrated in Fig. 4 on the $\alpha^k_{1,2}$-$\alpha^k_{2,1}$ plane for fixed η^k, σ_1^k, σ_2^k, u_1 and u_2. For both colored regions, there is no DC structure and the line search can be solved with low complexity. In the unshaded region, neither concavity nor convexity can be established and either a polynomial must be solved or an exhaustive line search has to be performed.

C. Exhausting the sum power

After the solution for the radius, each tone has a sum power constraint to be divided among the users. The share that each user gets is determined by solving for the angles. The third way to exploit structure with the taxicab spherical coordinates relies on this fact. This is best conveyed by two examples.

First, it can be that $\eta^k = 0$. In this case, there is no power to be distributed among the users. Tones where $\eta^k = 0$ can be skipped for all angles.

Second, consider $\eta^k > 0$. For the $N = 3$ case, consider that we have solved for $\boldsymbol{\eta}$ and $\boldsymbol{\phi}_{[1]}$. When solving for $\boldsymbol{\phi}_{[2]}$, tones where $\phi^k_{[1]} = 1$ can be skipped. For these tones, the sum power is already exhausted in the $\phi^k_{[1]}$ direction (see (18)).

Hence, when solving for a given $\boldsymbol{\phi}_{[i]}$ the set of tones that need to be solved for is given by

$$\mathcal{K}_* = \big\{k | k \in \mathcal{K},\ \eta^k > 0 \text{ and } \phi^k_{[j]} \neq 1,\ j < i\big\}.$$

D. Algorithm

We are now ready to summarize our second algorithm. A pseudo-code is shown in Algorithm 2. We call the algorithm **Ta**xicab **S**pherical Coordinates **S**pectrum **O**ptimization (TaSSO). We remark that, as the ISB, TaSSO does not do approximations to facilitate the line searches.

We first solve for the radius, which corresponds to a concave line search problem. The Lagrange multiplier $\bar{\lambda}$ can be found with a simple bisection search. Then we solve for the angles. We begin with $\boldsymbol{\phi}_{[1]}$ and continue until $\boldsymbol{\phi}_{[N-1]}$. For a given i, in lines 8-11 we classify the tones. Here $\mathcal{K}_*$, $\mathcal{K}_{\text{cnc}}$, $\mathcal{K}_{\text{cvx}}$ and $\mathcal{K}_{\text{neither}}$ represent, respectively, the set of tones that need to be solved for, the set of tones where (26) holds, the set of tones where (27) holds, and the set of tones where neither concavity not convexity can be established. Each set has $\phi^k_{[i]}$ optimized in a different way:

If $k \in \mathcal{K}_{\text{cnc}}$, then the problem can be easily solved with e.g. the Newton method.

If $k \in \mathcal{K}_{\text{cvx}}$, the optimal is either $\phi^k_{[i]} = 0$ or $\phi^k_{[i]} = 1$.

If $k \in \mathcal{K}_{\text{neither}}$, then an exhaustive line search is necessary. This can be done with

$$\phi^k_{[i]} = \text{argmax}_{\phi^k_{[i]} \in \mathcal{D}} L(\phi^k_{[i]}, \boldsymbol{\lambda}),$$

where $\mathcal{D}$ contains points in the discretized $[0, 1]$ segment. Alternatively, we can find the stationary condition of (21), i.e. $\partial L(\phi^k_{[i]}, \boldsymbol{\lambda}) / \partial \phi^k_{[i]} = 0$. This results in a polynomial of degree $2N$. We can find the roots of the polynomial, discard the ones that are not in $[0, 1]$, and calculate the per-tone Lagrangean for the remaining roots and for the points $\phi^k_{[i]} = 0$ and $\phi^k_{[i]} = 1$. We then pick the point that maximizes the per-tone Lagrangean. In our implementation, we solve the polynomial.

At the end of the algorithm, we adjust the Lagrange multipliers $\boldsymbol{\lambda}$ (ϵ is a step size). The process repeats until convergence.

E. Computational complexity and convergence

The computational complexity of TaSSO is dominated by the exhaustive line searches. It is difficult to estimate how many times these exhaustive line searches take place because it is not know a priori how many tones in the system fall on the concave or convex categories. Complexity is estimated as $O\big((2N)^3(N-1)\big|\mathcal{K}_{\text{neither}}\big|\big)$, where the term $(2N)^3$ corresponds to the computational complexity of solving a polynomial of degree $2N$. In the worst possible case $\mathcal{K}_{\text{neither}} = \mathcal{K}$, and computational complexity is given by $O\big((2N)^3(N-1)K\big)$. As a comparison, the computational complexity of ISB is given by $O((2N-1)^3NK)$. In the worst case, TaSSO needs to solve $K(N-1)$ polynomials of degree $2N$. ISB always needs to solve KN polynomials of degree $2N-1$.

We see from experiments that even in the worst case TaSSO is faster than ISB. In practice, however, we see that $\mathcal{K}_{\text{neither}} = \mathcal{K}$ almost never happens. In the so-called near-far scenarios (see Section V), sometimes the majority of tones fall in the concave case. We observe from experiments with realistic channels that TaSSO is 2-15 times faster than ISB.

Unfortunately, we do not have a proof of convergence of TaSSO. However, the algorithm has been experimented with extensively and has always been observed to converge.

V. Numerical Simulations

For all simulations in this section, we use $f_s = 4$ kHz and $\Delta_f = 4.3125$ kHz. Also, all algorithms are initialized with $p_n^k = 10^{-6}\ \forall n,k$. The criterion for convergence is $\big|\text{WRS}(i) - \text{WRS}(i-1)\big| \leq N10^{-6}$ for all algorithms, where $\text{WRS}(i)$ is the resulting WRS from iteration i.

A. OSB vs. OSB-SC

We simulate OSB and OSB-SC in a typical near-far, downstream ADSL scenario with 2 users. We illustrate the scenario in Fig. 5. We consider that only the two users on the top of the figure are active. Referring to the figure, we set $l_1 = 5$ km, $l_2 = 3$ km and $d_2 = 4$ km. We use 0.5 mm (24 AWG) cables and noise model ANSI A [28]. We use 20.4 dBm as PC for each user and $\Gamma = 12.8$ dB. We set $u_1 = 0.9$ and $u_2 = 0.1$. We use $Q = 301$ lines for OSB-SC and Q^2 points for OSB.

We see that OSB-SC is emphatically faster than OSB. While the former converges in less than 15 seconds, the latter takes approximately 16 minutes. This represents a saving in computational complexity by a factor of 60. In Fig. 6, we depict the final power allocation of both algorithms. Seen from far they look the same, but, once zoomed in, we can see that OSB-SC is more precise (see Fig. 7 and discussion in Section III-C). The final PSD of OSB is 'blocky', whereas that of OSB-SC is much more smooth.

B. Random Downstream ADSL

In this experiment, we simulate a four user random downstream ADSL scenario. All system parameters are the same as in the previous section. Referring to Fig. 5, we set l_1 to be a random variable with pdf $\text{unif}(4,6)$ km, l_2 to be a random variable with pdf $\text{unif}(3,5)$ km, l_3 to be a random variable with pdf $\text{unif}(2.5,4.5)$ km, and l_4 to be a random variable with pdf $\text{unif}(2,4)$ km. We define $d_1 = 0$ and d_i, $i = 2,3,4$ to be random variables with pdf $\text{unif}(i-1, \min_{j<i} d_j + l_j)$. We define the d_i sequentially, i.e. first d_2 and continue until d_4. We create 100 realizations of this scenario, and for each we calculate the solutions of ISB and TaSSO.

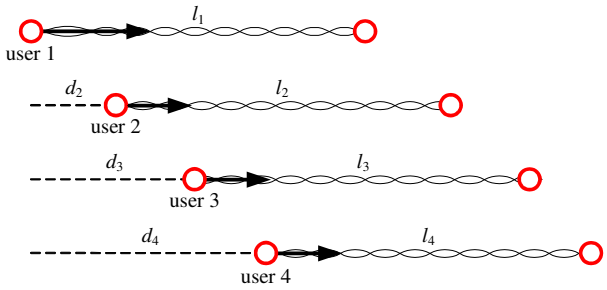

Fig. 5. ADSL downstream scenario.

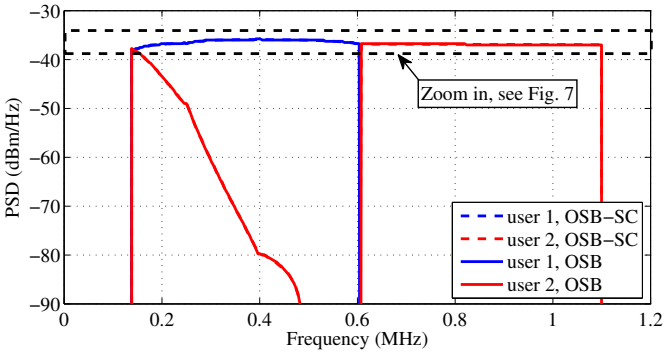

Fig. 6. Power allocation OSB and OSB-SC. Both algorithms reach basically the same result, but OSB-SC is more precise.

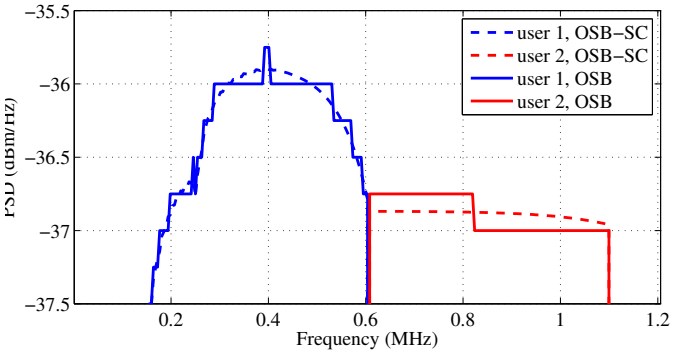

Fig. 7. Zoomed in region indicated in Fig. 6.

The goal of this experiment is to assess the average behavior of ISB and TaSSO in terms of time complexity. The results are shown in Table I.

On average, TaSSO is almost 4 times faster than ISB. The number of line searches each algorithm needs to solve explains why. For TaSSO, on average 257.25 line searches are solved. For ISB, for every realization $KN = 224 \times 4 = 896$ line searches are solved. The two algorithms perform equally well in terms of final WRS.

	ISB	TaSSO
Convergence time, average (sec)	25.04	6.45
Convergence time, st. deviation	5.71	3.14
no. line searches, average	896	271.76
no. line searches, st. deviation	0	70.15

TABLE I
RESULTS FOR THE EXPERIMENT WITH THE RANDOM DOWNSTREAM ADSL SCENARIO.

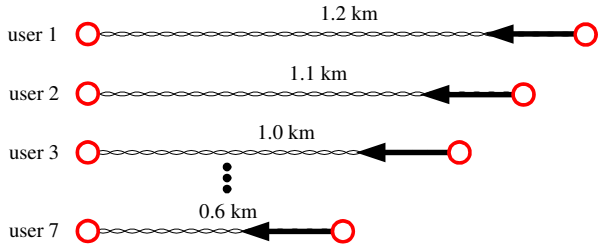

Fig. 8. Upstream VDSL scenario.

C. Upstream VDSL

We also simulate an upstream VDSL scenario. We use 0.4 mm (AWG 26) cables and an $\Gamma = 9.45$ dB. Each user has a total power budget of 11.5 dBm. For each line, noise model ETSI A is adopted with a background noise level of -140 dBm/Hz. We use the FDD 998 frequency bandplan over POTS up to 12 MHz.

The scenario is depicted in Fig. 8. There are 7 users, with line lengths equal to 1.2, 1.1, ..., 0.6 km. We simulate the scenario six times. The first time we consider users 1 and 2 to be active. The second time we consider users 1 to 3 active. We continue until all users are active. For every simulation, we use equal weights for the users. We run ISB and TaSSO and compare their performances.

Both algorithms perform equally well in terms of WRS. The difference is their time complexity, which is depicted in Fig. 9. Here we see that TaSSO is 5 to 15 times faster than ISB. For example, with 7 active users, ISB converges in around 24 minutes, while TaSSO does so in 4 and a half minutes.

TaSSO is faster because it uses structure. This is conveyed by Fig. 10, where we depict the number of line searches each of the algorithms does in its last iteration (we solve polynomials for both algorithms). With TaSSO, these numbers are from 6 to 2.3 times smaller.

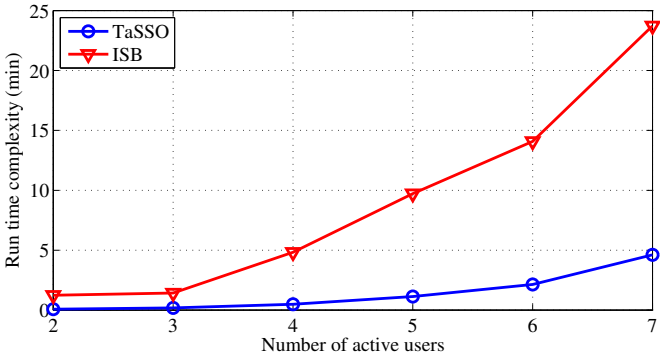

Fig. 9. Time complexity vs. number of active users for the upstream VDSL scenario.

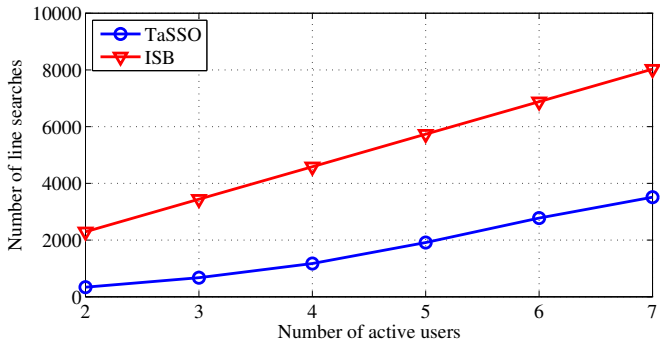

Fig. 10. Number of line searches (in our case, polynomials to be solved) vs. number of active users for the upstream VDSL scenario.

VI. CONCLUSION

In this paper, we have proposed two algorithms for the DSM problem. Both our proposals depart from the re-writing of the decision variables in spherical coordinates, i.e. as a function of a radius and angles. The advantage of this change of variables is that we are able to find structure in the problem.

The first algorithm, called OSB-SC, uses standard (Euclidean) spherical coordinates. We exploit the fact that the problem is concave in the radial dimension. An exhaustive search is done for the remaining variables, i.e. the angles. We see from the experiments that OSB-SC can be up to 60 times faster than the previously proposed OSB.

The second algorithm, called TaSSO, uses spherical coordinates in taxicab geometry. This coordinate system makes it easier to solve the problem in a block coordinate descent method, i.e. a sequence of line searches, where we first solve for the radius and subsequently solve for the angles. This algorithm uses structure in three ways. First, the problem is concave in the radial dimension. Second, there is some limited structure to be found in the angle dimensions. We have established sufficient conditions for the line searches to be convex and concave in each of the angle dimensions. And, third, after the solution for the radius there is a sum power constraint for every tone. We see from the experiments that TaSSO can be 2-15 times faster than the previously proposed ISB.

Of these three ways to exploit structure, we believe that the second (i.e. the structure in the angle dimensions) is both the most important and the most surprising. It is the most important because we see from the experiments that it is the one that accounts for the largest savings in computational complexity. It is the most surprising because it conveys the message that the DSM problem has more structure than previously imagined. We should add that we do not exploit all the structure there is. The sufficient conditions for convexity and concavity we propose have the considerable advantage of being easily verified, but they fall short of mapping the 'real' concave and convex regions, which are much larger. How to better map these 'real' concave and convex regions remains a topic for further research. Perhaps there are also regions of quasi-concavity and quasi-convexity (as defined in [29]).

APPENDIX A
N-DIMENSIONAL SPHERE FORMULAS

To re-write $\mathbf{p}^k \in \mathbb{R}^N$ as (6) for the general case, we write $\mathbf{p}_n^k = \rho_n^k \mathbf{d}_n^k$, where

$$d_1^k = \prod_{j=1}^{N-1} \cos\theta_{[j]}^k \tag{28}$$

$$d_i^k = \prod_{j=1}^{N-i} \cos\theta_{[j]}^k \sin\theta_{[N-i+1]}^k, \quad i = 2, \ldots, N-1 \tag{29}$$

$$d_N^k = \sin\theta_{[1]}^k \tag{30}$$

Here $\theta_{[i]}^k \in [0, \pi/2],\ i = 1, \ldots, N-1$.

To re-write $\mathbf{p}^k$ as (17) for the general case, we have $\mathbf{p}^k = \eta^k \mathbf{v}^k$, where

$$v_1^k = \prod_{j=1}^{N-1} (1 - \phi_{[j]}^k) \tag{31}$$

$$v_i^k = \prod_{j=1}^{N-i} (1 - \phi_{[j]}^k)\phi_{[N-i+1]}^k, \quad i = 2, \ldots, N-1 \tag{32}$$

$$v_N^k = \phi_{[1]}^k \tag{33}$$

Here $\phi_{[i]}^k \in [0, 1],\ i = 1, \ldots, N-1$.

Transforming between Cartesian coordinates on the one hand and N-spherical coordinates in both Euclidean and taxicab geometry on the other hand is straightforward in both directions.

APPENDIX B
PROOF OF PROPOSITION 2

To arrive at (14) and (15), we consider Q even. Without loss of generality, we consider $\bar{P}^{\min} = 0$. For (14), we decorrelate the per-tone variables. We can then focus on the point $\mathbf{y} = [0 \ \ 0]^{\mathrm{T}}$ and the area enclosed by the dotted square close to it. See Fig. 2 for an illustration. Eq. (14) becomes

$$\mathrm{E}\left[d_{\text{points}}^2\right] \leq \mathrm{E}\left[\frac{1}{K}\sum_{k\in\mathcal{K}} (\bar{p}_1^k)^2 + (\bar{p}_2^k)^2\right] = 2\mathrm{E}\left[(\bar{p}_1^1)^2\right] = \frac{\bar{P}}{6(Q-1)^2} \tag{34}$$

Here, we use the fact that the decorrelated per-tone random variables are equally distributed. We can calculate the expected value in $\bar{p}_1^1$ and multiply it by 2. The variable $\bar{p}_1^1$ is uniformly distributed in $(0, \bar{P}/2(Q-1))$. Reaching (34) is then straightforward.

To arrive at (15), we define

$$G(\mathbf{d}, \bar{\mathbf{p}}^k) = \left\|(\mathbf{d}^{\mathrm{T}}\bar{\mathbf{p}}^k)\mathbf{d} - \bar{\mathbf{p}}^k\right\|_2^2 = (\bar{\mathbf{p}}^k)^{\mathrm{T}}\bar{\mathbf{p}}^k - \mathbf{d}^{\mathrm{T}}\bar{\mathbf{p}}^k(\bar{\mathbf{p}}^k)^{\mathrm{T}}\mathbf{d} \tag{35}$$

and

$$\mathrm{E}\left[d_{\text{lines}}^2\right] = \mathrm{E}\left[\frac{1}{K}\sum_{k\in\mathcal{K}} \min_{\mathbf{d}\in\mathcal{D}}\ G(\mathbf{d}, \bar{\mathbf{p}}^k)\right] \tag{36}$$

In (35), we use $\|\mathbf{d}\|_2^2 = \mathbf{d}^{\mathrm{T}}\mathbf{d} = 1$. The most important steps in the derivation are shown in (37)-(42), on the top of the next page. The expectation in (36) is written in integral form as (37). Here the sets $\mathcal{R}_1$ and $\mathcal{R}_2$ represent the hypervolumes where the integration takes place. They are defined as $\mathcal{R}_n = \left\{[\bar{p}_n^1 \ \ \cdots \ \ \bar{p}_n^k] \middle|\ \bar{p}_n^k \geq 0\ \forall k \text{ and } \sum_{k\in\mathcal{K}} 10^{\bar{p}_n^k/10} \leq P^{\max}\right\}$, $n = 1, 2$. The pdf's are represented by $f(\bar{\mathbf{p}}_1)$ and $f(\bar{\mathbf{p}}_2)$. Also $d\bar{\mathbf{p}}_n = d\bar{p}_n^1 \ldots d\bar{p}_n^K$.

In the whole derivation, we do two relaxations. The first is that we decorrelate the per-tone variables, i.e. we define $f(\bar{p}_n^k) = \mathrm{unif}(0, \bar{P})\ \forall n, k$ and $\bar{P} = 10\log_{10} P^{\max}$. The corresponding integral is shown in (38). In (39), we change to spherical coordinates, i.e. $\bar{p}_1^k = r^k\cos\theta^k$ and $\bar{p}_2^k = r^k\sin\theta^k$. We define

$$G(\mathbf{d}, r^k, \theta^k) = (r^k)^2\Big(1 - \mathbf{d}^{\mathrm{T}}\begin{bmatrix}\cos^2\theta^k & \cos\theta^k\sin\theta^k \\ \cos\theta^k\sin\theta^k & \sin^2\theta^k\end{bmatrix}\mathbf{d}\Big) \tag{43}$$

We write the integral in r^k in the area defined by $\mathcal{A}^k = \{r^k | 0 \leq \theta^k \leq \frac{\pi}{2}, r^k = \min(\bar{P}(\cos\theta^k)^{-1},\ \bar{P}(\sin\theta^k)^{-1})\}$ and $d\mathbf{r}$ and $d\boldsymbol{\theta}$ are defined similarly to $d\bar{\mathbf{p}}_n$. In (40), we put the summation in k outside the integrals and notice that, since now the tones are uncorrelated, for a given tone k, we can integrate in $j,\ j \neq k$ easily. After that, we arrive at the left hand side of (41). Here the superscripts are no longer necessary, so we omit them.

From the left hand side of (41) to the right hand side of (41), we do a second relaxation. The integration on the region $\mathcal{A}$ is too difficult, so we extend it to include more points in the radial direction. The region we integrate on is denoted $\mathcal{A}'$ and is given by $\mathcal{A}' = \mathcal{A}_1 \cup \mathcal{A}_2$

$$\mathcal{A}_1 = \bigcup_{i=0}^{Q-2}\Big\{r, \theta \ \Big|\ \frac{i\pi}{4(Q-1)} \leq \theta \leq \frac{(i+1)\pi}{4(Q-1)}\Big], \ r = \bar{P}\cos^{-1}\Big(\frac{(i+1)\pi}{4(Q-1)}\Big)\Big\}$$

$$\mathcal{A}_2 = \bigcup_{i=0}^{Q-2}\Big\{r, \theta \ \Big|\ \frac{\pi}{4} + \frac{i\pi}{4(Q-1)} \leq \theta \leq \frac{\pi}{4} + \frac{(i+1)\pi}{4(Q-1)}, \ r = \bar{P}\sin^{-1}\Big(\frac{\pi}{4} + \frac{i\pi}{4(Q-1)}\Big)\Big\}.$$

We illustrate this region in Fig. 2. It consists of the union of the angular sectors with angular width of $\pi/4(Q-1)$. The radii increase as we approach $\pi/4$. Notice that, as far as the integrals are concerned, the regions $\mathcal{A}_1$ and $\mathcal{A}_2$ are the same. We can thus integrate in $\mathcal{A}_1$ and multiply the result by 2. This is what is shown in the right hand side of (41). The variables r and θ are uniformly distributed in $\mathcal{A}'$, so their joint pdf is given by $f(r, \theta) = 1/A$, where

$$A = \frac{\pi\bar{P}^2}{4(Q-1)}\sum_{i=0}^{Q-2}\cos^{-2}\Big(\frac{(i+1)\pi}{4(Q-1)}\Big). \tag{44}$$

To arrive at (42), we notice that all segments have the same volume if they have the same radius. To make the calculations easier, we consider the segment close to $\mathbf{d} = [1 \ \ 0]^T$ (i.e. $0 \leq \theta \leq \pi/4(Q-1)$) and calculate the integral repeatedly for different values of the radius. We thus re-write (43) as $G(\mathbf{d}, r, \theta) = r^2\sin^2\theta$. In (42), we see that

$$\int_{\theta=0}^{\pi/4(Q-1)} \sin^2\theta d\theta = \frac{\pi}{8(Q-1)} - \frac{1}{4}\sin\Big(\frac{\pi}{2(Q-1)}\Big) \tag{45}$$

$$\mathrm{E}\left[d_{\text{lines}}^2\right] = \underbrace{\int\cdots\int}_{\mathcal{R}_1}\underbrace{\int\cdots\int}_{\mathcal{R}_2}\frac{1}{K}\Big(\sum_{k\in\mathcal{K}}\min_{\mathbf{d}\in\mathcal{D}}\ G(\mathbf{d},\bar{\mathbf{p}}^k)\Big)f(\bar{\mathbf{p}}_1)f(\bar{\mathbf{p}}_2)d\bar{\mathbf{p}}_1 d\bar{\mathbf{p}}_2 \tag{37}$$

$$< \frac{1}{\bar{P}^{2K}}\int_{\bar{p}_1^1=0}^{\bar{P}}\cdots\int_{\bar{p}_1^K=0}^{\bar{P}}\int_{\bar{p}_2^1=0}^{\bar{P}}\cdots\int_{\bar{p}_2^K=0}^{\bar{P}}\frac{1}{K}\Big(\sum_{k\in\mathcal{K}}\min_{\mathbf{d}\in\mathcal{D}}\ G(\mathbf{d},\bar{\mathbf{p}}^k)\Big)d\bar{\mathbf{p}}_1 d\bar{\mathbf{p}}_2 \tag{38}$$

$$= \frac{1}{\bar{P}^{2K}K}\int_{r^1\in\mathcal{A}^1}\int_{\theta^1=0}^{\pi/2}\cdots\int_{r^K\in\mathcal{A}^K}\int_{\theta^K=0}^{\pi/2}\Big(\sum_{k\in\mathcal{K}}\min_{\mathbf{d}\in\mathcal{D}}\ G(\mathbf{d},r^k,\theta^k)\Big)\prod_{k\in\mathcal{K}}r^k d\mathbf{r}d\boldsymbol{\theta} \tag{39}$$

$$= \frac{1}{\bar{P}^{2K}K}\sum_{k\in\mathcal{K}}\int_{r^1\in\mathcal{A}^1}\int_{\theta^1=0}^{\pi/2}\cdots\int_{r^K\in\mathcal{A}^K}\int_{\theta^K=0}^{\pi/2}\min_{\mathbf{d}\in\mathcal{D}}\ G(\mathbf{d},r^k,\theta^k)\prod_{j\in\mathcal{K}}r^j d\mathbf{r}d\boldsymbol{\theta} \tag{40}$$

$$= \frac{1}{\bar{P}^2}\int_{r\in\mathcal{A}}\int_{\theta=0}^{\pi/2}\min_{\mathbf{d}\in\mathcal{D}}\ G(\mathbf{d},r,\theta)rdrd\theta < \frac{2}{A}\sum_{i=0}^{Q-2}\int_{r=0}^{\bar{P}\cos^{-1}\left((i+1)\pi/4(Q-1)\right)}\int_{\theta=i\pi/4(Q-1)}^{(i+1)\pi/4(Q-1)}\min_{\mathbf{d}\in\mathcal{D}}G(\mathbf{d},r,\theta)rdrd\theta \tag{41}$$

$$= \frac{2}{A}\sum_{i=0}^{Q-2}\int_{r=0}^{\bar{P}\cos^{-1}\left((i+1)\pi/4(Q-1)\right)}\int_{\theta=0}^{\pi/4(Q-1)}r^3\sin^2\theta drd\theta \tag{42}$$

and

$$\sum_{i=0}^{Q-2}\int_{r=0}^{\bar{P}\cos^{-1}\left(\frac{(i+1)\pi}{4(Q-1)}\right)}r^3dr = \frac{\bar{P}^4}{4}\sum_{i=0}^{Q-2}\cos^{-4}\Big(\frac{(i+1)\pi}{4(Q-1)}\Big). \tag{46}$$

By substituting (44), (45) and (46) in (42), we complete the proof.

Appendix C
Proof of Propositions 3 and 4

In this appendix, we drop the superscript k for the sake of conciseness. We first establish some properties related to a_n, D_n, c_n and F_n, all of them consequences of the definition of these variables, as follows:

(i) $a_n\phi_{[i]} + D_n \geq 0$ and $c_n\phi_{[i]} + F_n \geq 0$, $\phi_{[i]} \in [0,1]$ $\forall n$;
(ii) $D_n \geq 0$ and $F_n \geq 0$ $\forall n$;
(iii) $a_n + D_n \geq 0$ and $c_n + F_n \geq 0$ $\forall n$
(iv) either $a_n/D_n = -1$ or $a_nD_n = 0$ $\forall n$

Property (i) is due to the fact that by definition $a_n\phi_{[i]} + D_n = \eta v_n^k$, where the right hand side is non-negative. The same is valid for $c_n\phi_{[i]} + F_n$. Properties (ii) and (iii) are direct consequences of (i). Property (iv) is due to the fact that either $a_n\phi_{[i]} + D_n = X(1-\phi_{[i]})$ or $a_n\phi_{[i]} + D_n = X\phi_{[i]}$, X is non-negative constant.

A sufficient condition for concavity of (21) in $\phi_{[i]}$ is given by

$$\max_{\phi_{[i]}\in[0,1]}\Big(\frac{\partial^2 L(\phi_{[i]},\boldsymbol{\lambda})}{\partial(\phi_{[i]})^2}\Big) = \max_{\phi_{[i]}\in[0,1]}\Big(\sum_{n\in\mathcal{N}}\frac{\partial^2 b_n}{\partial(\phi_{[i]})^2}\Big) \leq \sum_{n\in\mathcal{N}}\max_{\phi_{[i]}\in[0,1]}\Big(\frac{\partial^2 u_n b_n}{\partial(\phi_{[i]})^2}\Big) = \sum_{n\in\mathcal{N}}\max\Big(\frac{\partial^2 u_n b_n}{\partial(\phi_{[i]})^2}(0),\frac{\partial^2 u_n b_n}{\partial(\phi_{[i]})^2}(1)\Big) \leq 0 \tag{47}$$

In the same vein, a sufficient condition for convexity of (21) in $\phi_{[i]}$ is

$$\min_{\phi_{[i]}\in[0,1]}\Big(\frac{\partial^2 L(\phi_{[i]},\boldsymbol{\lambda})}{\partial(\phi_{[i]})^2}\Big) = \min_{\phi_{[i]}\in[0,1]}\Big(\sum_{n\in\mathcal{N}}\frac{\partial^2 b_n}{\partial(\phi_{[i]})^2}\Big) \geq \sum_{n\in\mathcal{N}}\min_{\phi_{[i]}\in[0,1]}\Big(\frac{\partial^2 u_n b_n}{\partial(\phi_{[i]})^2}\Big) = \sum_{n\in\mathcal{N}}\min\Big(\frac{\partial^2 u_n b_n}{\partial(\phi_{[i]})^2}(0),\frac{\partial^2 u_n b_n}{\partial(\phi_{[i]})^2}(1)\Big) \geq 0 \tag{48}$$

Define $S_n(\phi_{[i]}) \triangleq {}^{\partial^2 u_n b_n}/_{\partial(\phi_{[i]})^2}(\phi_{[i]})$, and further define

$$S_n(\phi_{[i]}) = \underbrace{\frac{-u_n(a_n+c_n)^2}{\big((a_n+c_n)\phi_{[i]}+D_n+F_n\big)^2}}_{\triangleq S_n^1(\phi_{[i]})} + \underbrace{\frac{u_n(c_n)^2}{\big(c_n\phi_{[i]}+F_n\big)^2}}_{\triangleq S_n^2(\phi_{[i]})} \tag{49}$$

To prove the equalities in (47) and (48), i.e. to prove that $S_n(\phi_{[i]})$ is minimized/maximized either at 0 or 1, we have to show that $S_n(\phi_{[i]})$ is either monotonically increasing or monotonically decreasing for all n.

To show monotonicity of $S_n(\phi_{[i]})$, four cases should be analyzed: (1) $a_n + c_n \geq 0$ and $c_n \leq 0$; (2) $a_n + c_n \leq 0$ and $c_n \geq 0$; (3) $a_n + c_n \leq 0$ and $c_n \leq 0$; and (4) $a_n + c_n \geq 0$ and $c_n \geq 0$. Cases 1 and 2 are easy. For case 1, both $S_n^1(\phi_{[i]})$ and $S_n^2(\phi_{[i]})$ in (49) are monotonically increasing, which means that their sum also is. Similarly, for case 2 both $S_n^1(\phi_{[i]})$ and $S_n^2(\phi_{[i]})$ are monotonically decreasing, which means that their sum also is.

Cases 3 and 4 are more difficult because they are a sum of a monotonically increasing function and a monotonically decreasing function.

Lemma 1. Consider a monotonically increasing function $f_+ : [a,b] \mapsto f_+(x)$ and monotonically decreasing function $f_- : [a,b] \mapsto f_-(x)$. The sum $f_+(x) + f_-(x)$ is monotonically

increasing if

$$\left|\frac{\partial f_+(x)}{\partial x}\right| \geq \left|\frac{\partial f_-(x)}{\partial x}\right|,\ x \in [a,b]$$

Proof. The sum is monotonically increasing if ($\epsilon > 0$ is such that $x + \epsilon \leq b$)

$$\begin{aligned} &f_+(x+\epsilon) + f_-(x+\epsilon) \geq f_+(x) + f_-(x); \\ &\lim_{\epsilon \to 0} \frac{f_+(x+\epsilon) - f_+(x)}{\epsilon} \geq -\lim_{\epsilon \to 0} \frac{f_-(x+\epsilon) - f_-(x)}{\epsilon} \end{aligned} \quad (50)$$

To complete the proof, it suffices to identify the definition of the derivatives and notice that both sides are positive. Hence we can add the absolute value operator without loss of meaning. ■

For the sum $f_+(x) + f_-(x)$ to be monotonically decreasing, we only need to invert the inequality in (50).

In the following, we analyze case 4, i.e. when $S_n^1(\phi_{[i]})$ is monotonically increasing and $S_n^2(\phi_{[i]})$ is monotonically decreasing (the analysis of case 3 is very similar). We identify three subcases to be treated separately, which we call 4.1, 4.2 and 4.3. As a consequence of (ii) and (iv), we have (4.1) $a < 0$, $a = -D_n$; (4.2) $a_n > 0$, $D_n = 0$; and (4.3) $a_n = 0$, $D_n > 0$. For case 4.1, we write

$$\begin{aligned} \left|\frac{\partial S_n^1(\phi_{[i]})}{\partial \phi_{[i]}}\right| &\leq \left|\frac{\partial S_n^2(\phi_{[i]})}{\partial \phi_{[i]}}\right|; \\ \frac{(a_n + c_n)^3}{\left((a_n + c_n)\phi_{[i]} + D_n + F_n\right)^3} &\leq \frac{(c_n)^3}{(c_n\phi_{[i]} + F_n)}, \end{aligned} \quad (51)$$

which means that $S_n(\phi_{[i]})$ is monotonically decreasing. For case 4.2, we write

$$\frac{(a_n + c_n)^3}{\left((a_n + c_n)\phi_{[i]} + F_n\right)^3} \geq \frac{(c_n)^3}{(c_n\phi_{[i]} + F_n)}, \quad (52)$$

which means that $S_n(\phi_{[i]})$ is monotonically increasing. For case 4.3, we write

$$\frac{(c_n)^3}{\left((c_n)\phi_{[i]} + D_n + F_n\right)^3} \leq \frac{(c_n)^3}{(c_n\phi_{[i]} + F_n)}, \quad (53)$$

which means that $S_n(\phi_{[i]})$ is monotonically decreasing. In (51), (52) and (53), we use Properties (ii) and (iii). Cases 3.1, 3.2 and 3.3 are defined similarly and the conclusions are basically the same. This completes the proof.

REFERENCES

[1] G. Ginis and J. M. Cioffi, "Vectored transmission for digital subscriber line systems," *IEEE J. Sel. Areas Commun.*, vol. 20, no. 5, pp. 1085–1104, 2002.

[2] Broadband Forum, "FTTx supercharges broadband deployment," 2013. [Online]. Available: http://www.broadband-forum.org/news/download/pressreleeases/2013/BBF_FTTx13.pdf

[3] P. Ödling, T. Magesacher, S. Höst, P. O. Börjesson, M. Berg, and E. Areizaga, "Fourth generation broadband concept," *IEEE Commun. Mag.*, vol. 47, no. 1, pp. 63–69, 2009.

[4] W. Yu and R. Lui, "Dual methods for nonconvex spectrum optimization of multicarrier systems," *IEEE Trans. Commun.*, vol. 54, no. 7, pp. 1310–1322, 2006.

[5] Z.-Q. Luo and S. Zhang, "Dynamic spectrum management: Complexity and duality," *IEEE Trans. Signal Process.*, vol. 2, no. 1, pp. 57–73, 2009.

[6] ——, "Duality gap estimation and polynomial time approximation for optimal spectrum management," *IEEE Trans. Signal Process.*, vol. 57, no. 7, pp. 2675 – 2689, 2009.

[7] S. Hayashi and Z.-Q. Luo, "Spectrum management for interference-limited multiuser communication systems," *IEEE Trans. Inf. Theory*, vol. 55, no. 3, pp. 1153–1175, 2009.

[8] H. Boche, S. Naik, and T. Alpcan, "Characterization of convex and concave resource allocation problems in interference coupled wireless systems," *IEEE Trans. Sig. Process.*, vol. 59, no. 5, pp. 2382–2394, 2011.

[9] P. Tsiaflakis, C. Tan, Y. Yi, M. Chiang, and M. Moonen, "Optimality certificate of dynamic spectrum management in multi-carrier interference channels," in *IEEE Int. Symp. on Inf. Theory*, Toronto, Canada, 2008.

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

[11] R. Cendrillon, W. Yu, M. Moonen, J. Verlinden, and T. Bostoen, "Optimal multiuser spectrum balancing for digital subscriber lines," *IEEE Trans. Commun.*, vol. 54, no. 5, pp. 922–933, 2006.

[12] R. Cendrillon and M. Moonen, "Iterative spectrum management for digital subscriber lines," in *IEEE Int. Conf. Commun.*, Seoul, Korea, 2005.

[13] J. Papandriopoulos and J. S. Evans, "SCALE: a low-complexity distributed protocol for spectrum balancing in multiuser DSL networks," *IEEE Trans. Inf. Theory*, vol. 55, no. 8, pp. 3711–3724, 2009.

[14] P. Tsiaflakis, M. Diehl, and M. Moonen, "Distributed spectrum management algorithms for multiuser DSL networks," *IEEE Trans. Signal Process.*, vol. 56, no. 10, pp. 4825–4843, 2008.

[15] R. B. Moraes, B. Dortschy, A. Klautau, and J. Rius i Riu, "Semiblind spectrum balancing for DSL," *IEEE Trans. Signal Process.*, vol. 58, no. 7, pp. 3717–3727, 2010.

[16] M. Wolkerstorfer, J. Jaldén, and T. Nordström, "Column generation for discrete-rate multi-user and multi-carrier power control," *IEEE Trans. Commun.*, vol. 60, no. 9, pp. 2712–2722, 2012.

[17] R. B. Moraes, P. Tsiaflakis, J. Maes, and M. Moonen, "DMT MIMO IC rate maximization in DSL with combined signal and spectrum coordination," *IEEE Trans. Signal Process.*, vol. 61, no. 7, pp. 1756–1769, 2013.

[18] W. Yu, "Multiuser water-filling in the presence of crosstalk," in *Inf. Theory and Appl. Workshop*, San Diego, USA, 2007.

[19] Q. Shi, M. Razaviyayn, Z.-Q. Luo, and C. He, "An iteratively weighted MMSE approach to distributed sum-utility maximization for a MIMO interfering broadcast channel," *IEEE Trans. Signal Process.*, vol. 59, no. 9, pp. 4331–4340, 2011.

[20] E. F. Krause, *Taxicab geometry: an adventure in non-Euclidean geometry*. New York (N.Y.): Dover, 1986.

[21] V. M. K. Chan and W. Yu, "Multiuser spectrum optimization for discrete multitone systems with asynchronous crosstalk," *IEEE Trans. Signal Process.*, vol. 55, no. 11, pp. 5425–5435, 2007.

[22] R. B. Moraes, P. Tsiaflakis, and M. Moonen, "Intercarrier interference in DSL networks due to asynchronous DMT transmission," in *IEEE Int. Conf. Acoust., Speech and Signal Process.*, Vancouver, Canada, 2013.

[23] R. B. Moraes, P. Tsiaflakis, J. Maes, and M. Moonen, "DMT MIMO IC rate maximization in DSL with per-transceiver power constraints," *Submitted for publication*, 2013.

[24] T. Starr, J. M. Cioffi, and P. Silverman, *Understanding Digital Subscriber Lines Technology*. Prentice Hall, 1999.

[25] M. Chiang, P. Hande, T. Lan, and C. W. Tan, "Power control in wireless cellular networks," *Found. Trends Netw.*, vol. 2, no. 4, pp. 381–533, Apr. 2008.

[26] M. Sadek, A. Tarighat, and A. H. Sayed, "A leakage-based precoding scheme for downlink multi-user MIMO channels," *IEEE Trans. Wireless Commun.*, vol. 6, no. 5, pp. 1711–1721, 2007.

[27] P. Tsiaflakis and F. Glineur, "A novel class of iterative approximation methods for DSL spectrum optimization," in *IEEE Int. Conf. Commun.*, Ottawa, Canada, 2013.

[28] V. Oksman and J. M. Cioffi, "Noise Models for VDSL Performance Verification," ANSI, ANSI-77E7.4/99.438R2, 1999.

[29] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.