Optimal layered architectures of wireless networks

Alejandro Ribeiro and Georgios B. Giannakis†

***Abstract*—A general wireless networking problem is formulated to jointly optimize end-to-end user rates, routes, link capacities, transmitted power, and power allocation across subcarriers and fading states. It has been recently shown that such general wireless networking problems in the presence of fading, while non-convex, have zero Lagrangian duality gap. In this paper the optimality of two different layered architectures is derived as a consequence of this result.**

I. Introduction

Communication networks are deployed to transport information from generating sources to intended destinations. There are different ways in which a given network can accomplish this task, but among them there are some that optimize properly chosen criteria. Optimal design is one of the most promising alternatives for future wireless networks [1], [2]. This paper is concerned with fundamental properties of such optimal wireless networks. In particular the optimality of certain layered architectures is proved.

The first optimal layered architecture considered coincides with the conventional layers used in wired networks. The second optimal architecture is based on layers and layer interfaces. Layers maintain variables of interest to the network, while interfaces maintain auxiliary variables. Layers exchange variables only with adjacent interfaces and interfaces only with adjacent layers. Over time the network finds an optimal operating point that maximizes a given utility. Although the architectures presented here are novel, similar architectures have been reported elsewhere, see e.g., [3]. The main contribution of this paper is to show their optimality in the presence of fading.

II. Optimal wireless network

Consider an ad-hoc wireless network composed of J user terminals $\{T_i\}_{i=1}^J$. Terminal T_i wants to deliver packets for different application level flows generically denoted by k, with the flow k intended for destination T^k. Network connectivity is modeled with a graph $\mathcal{G}(v,e)$ with vertices $v := [1, J]$ and edges $e \in \mathcal{E}$ connecting pairs of vertices (i,j) when and only when T_i and T_j can communicate with each other. Adjacency of i is denoted as $n(i) := \{j : (i,j) \in \mathcal{E}\}$. Each terminal $\{T_j\}_{j\in n(i)}$ that can communicate with T_i is referred to as a neighbor and the set of all neighbors as T_i's neighborhood. Network nodes communicate using a set of frequency tones $f \in \mathcal{F}$. The channel from T_i to T_j is denoted as h_{ij}^f and modeled as a random variable. Channel gains of all network links are collected in the vector $\mathbf{h}$.

Terminals T_i select various variables that determine the flow of information through the network. For given channel realizations $\mathbf{h}$, terminal T_i determines a power profile $p_{ij}^f(\mathbf{h})$ used for sending packets to T_j on the tone f when the channel vector realization is $\mathbf{h}$. Power profiles determine T_i's power consumption p_i and the capacity c_{ij} of the $T_i \to T_j$ link. For every flow k, T_i sends packets to neighboring terminals $\{T_j\}_{j\in n(i)}$ at an average rate r_{ij}^k. Likewise it receives packets from neighbors at a rate r_{ji}^k. Finally, variables a_i^k determine the rate at which T_i accepts packets of the flow k from applications. These variables are not independent of each other. They must satisfy constraints that will be explained shortly [cf. (2)-(4)].

†A. Ribeiro is with the Dept. of Electrical and Systems Engineering, University of Pennsylvania. G. B. Giannakis is with the Dept. of Electrical and Computer Engineering, University of Minnesota. Email: aribeiro@ese.upenn.edu, georgios@ece.umn.edu.

Network design calls for selecting powers p_i, link capacities c_{ij}, rates r_{ij}^k, arrivals a_i^k and power profiles $p_{ij}^f(\mathbf{h})$ that are optimal in some sense. Thus, concave $U_i^k(a_i^k)$ and convex $V_i(p_i)$ functions, respectively, representing the value of rate a_i^k and the cost of power p_i are introduced. The optimal operating point is then defined as the set of variables solving the optimization problem

$$P = \max \sum_{i,k} U_i^k(a_i^k) - \sum_i V_i(p_i) \tag{1}$$

$$c_{ij} \leq \mathrm{E}_{\mathbf{h}}\left[\sum_{f\in\mathcal{F}} C_{ij}\left(\mathbf{h}^f, \mathbf{p}^f(\mathbf{h})\right)\right] \tag{2}$$

$$p_i \geq \mathrm{E}_{\mathbf{h}}\left[\sum_{j\in\mathcal{N}(i)}\sum_{f\in\mathcal{F}} p_{ij}^f(\mathbf{h})\right] \tag{3}$$

$$a_i^k \leq \sum_{j\in n(i)}\left(r_{ij}^k - r_{ji}^k\right), \qquad \sum_k r_{ij}^k \leq c_{ij}. \tag{4}$$

Of the two constraints in (4) the first one requires the rate a_i^k at which packets are accepted from applications to be smaller than the difference between the aggregate departure rates (to neighbors) $\sum_{j\in n(i)} r_{ij}^k$ and arrival rates (from neighbors) $\sum_{j\in n(i)} r_{ji}^k$. The second constraint requires the total rate $\sum_k r_{ij}^k$ sent from T_i to T_j for all flows to be smaller than the link's capacity c_{ij}. The constraint in (3) states that the average power consumption p_i is obtained by summing over all links $j \in n(i)$ and tones $f \in \mathcal{F}$ and taking expected value over channel realizations $\mathbf{h}$. The capacity constraint in (2) is a similar average over fading states and tones. The function $C_{ij}(\mathbf{h}^f, \mathbf{p}^f(\mathbf{h}))$ maps channels and powers into link capacities so that the capacity $c_{ij}^f(\mathbf{h})$ of the link $T_i \to T_j$ on the tone f is $C_{ij}(\mathbf{h}^f, \mathbf{p}^f(\mathbf{h}))$. The function $C(\cdot)$ is determined by terminal's capabilities and operating conditions. If, e.g., terminals perform single user detection, link capacity is determined by the signal to noise plus interference ratio (SINR). Refer to [4] for a more detailed account of the model in (1)-(4).

All problem variables have to be non-negative, but this is left implicit in (1)-(4). Also implicit in (1)-(4) are power constraints $p_i \leq p_{\max i}$ and $p_{ij}^f(\mathbf{h}) \leq p_{\max}$, arrival rate requirements $a_{\min i}^k \leq a_i^k \leq a_{\max i}^k$ and upper bound constraints $c_{ij} \leq c_{\max}$ and $r_{ij}^k \leq r_{\max}$ on link capacities and link flow rates. These constraint define a box B of feasible variables. They will be implicit in general and make explicit when demanded by clarity. For future reference define the vector valued power distribution $\mathbf{p}(\mathbf{h})$ with components $p_{ij}^f(\mathbf{h})$ and $\mathbf{X}$ the set of primal variables c_{ij}, p_i, r_{ij}^k and a_i^k for all possible subindexes – i.e., all i and all $j \in n(i)$ for c_{ij}, all i for p_i and so on. Further define $f(\mathbf{X})$ as the utility function in (1) and $\mathbf{h}\left[\mathbf{X}, \mathbf{p}(\mathbf{h})\right] \geq \mathbf{0}$ the constraints (2)-(4) so that (1)-(4) can be written in generic form as

$$P = \max_{(\mathbf{X},\mathbf{p}(\mathbf{h}))\in B} f(\mathbf{X}); \qquad \text{st } \mathbf{h}\left[\mathbf{X}, \mathbf{p}(\mathbf{h})\right] \geq \mathbf{0} \tag{5}$$

where $B := \left\{(\mathbf{X}, \mathbf{p}(h)) : 0 \leq p_{ij}^f(\mathbf{h}) \leq p_{\max}, 0 \leq p_i \leq p_{\max i}, a_{\min i}^k \leq a_i^k \leq a_{\max i}^k, 0 \leq c_{ij} \leq c_{\max},\ 0 \leq r_{ij}^k \leq r_{\max}\right\}$ is the box outlined above.

The function $C(\cdot)$ in (2), and as a consequence $\mathbf{h}\left[\mathbf{X}, \mathbf{p}(\mathbf{h})\right]$ in (5), is not concave in general. Therefore, (1) is a difficult optimization problem. This difficulty notwithstanding, properties of wireless networks can be derived from properties of (1). For this purpose introduce multipliers

$\mathbf{\Lambda}$ and the Lagrangian

$$\mathcal{L}[\mathbf{X}, \mathbf{p}(\mathbf{h}), \mathbf{\Lambda}] = f(\mathbf{X}) + \mathbf{\Lambda}^T \mathbf{h}\left[\mathbf{X}, \mathbf{p}(\mathbf{h})\right]. \tag{6}$$

The dual function is obtained by maximizing the Lagrangian over the primal variables

$$g[\mathbf{\Lambda}] = \max_{(\mathbf{X}, \mathbf{p}(\mathbf{h})) \in B} \mathcal{L}[\mathbf{X}, \mathbf{p}(\mathbf{h}), \mathbf{\Lambda}] \tag{7}$$

where the maximization is performed over the set of primal variables satisfying the box constraints. The dual problem is finally defined as

$$D = \min_{\mathbf{\Lambda} \geq 0} g[\mathbf{\Lambda}]. \tag{8}$$

Because (1)-(4) is non-convex it may be the case that $D > P$, implying that working with the dual problem entails loss of optimality. However, it has been proved that as long as the cumulative distribution function of the fading channels is continuous the duality gap is null [4]:

Theorem 1 (Ribeiro-Giannakis 2008) *Let P denote the solution of the primal problem* (1) *and D the solution of its dual in* (8)*. If the channel cumulative distribution function (cdf) is continuous, then*

$$P = D. \tag{9}$$

Because of Theorem 1 the dual problem in (8) can be solved in lieu of the primal problem (5). This paper explains how this result can be exploited to show the optimality of separating the optimization of wireless networks into layers. Subsequently, the paper studies subgradient descent algorithms and shows that it induces an optimal separation in layers and layer interfaces.

III. Optimality of layering

An implication of Theorem 1 is the optimality of conventional layering in wireless networking problems. As is usually the case, the Lagrangian exhibits a separable structure in the sense that it can be written as a sum of terms that depend on a few primal variables. Define components of the Lagrange multipliers $\mathbf{\Lambda}$ so that λ_{ij} is associated with the capacity constraints in (2), μ_i with the power constraint in (3), and ν_i^k and ξ_{ij} with the flow and rate constraints in (4). The Lagrangian can then be written as

$$\begin{aligned}\mathcal{L}\Big[\mathbf{X}, \mathbf{p}(\mathbf{h}), \mathbf{\Lambda}\Big] &= \sum_{i,k} U_i^k(a_i^k) - \sum_i V_i(p_i) \\ &+ \sum_{i,j} \lambda_{ij} \left[\mathrm{E}_{\mathbf{h}} \left[\sum_{f \in \mathcal{F}} C_{ij}\Big(\mathbf{h}^f, \mathbf{p}^f(\mathbf{h})\Big) \right] - c_{ij} \right] \\ &+ \sum_i \mu_i \left[p_i - \mathrm{E}_{\mathbf{h}} \left[\sum_{j \in \mathcal{N}(i)} \sum_{f \in \mathcal{F}} p_{ij}^f(\mathbf{h}) \right] \right] \\ &+ \sum_{i,k} \nu_i^k \left[\sum_{j \in \mathcal{N}(i)} \Big(r_{ij}^k - r_{ji}^k\Big) - a_i^k \right] + \sum_{ij} \xi_{ij} \left[c_{ij} - \sum_k r_{ij}^k \right].\end{aligned} \tag{10}$$

Rearranging terms in (10) and assuming that the optimal dual argument $\mathbf{\Lambda}^*$ is available, we can write

$$\begin{aligned}\mathcal{L}[\mathbf{X}, \mathbf{p}(\mathbf{h}), \mathbf{\Lambda}^*] &= \sum_{i,k} \Big(U_i^k(a_i^k) - \nu_j^{k\,*} a_j^k\Big) + \sum_i \Big(\mu_i^* p_i - V_i(p_i)\Big) \\ &+ \sum_{i,j} \Big(\xi_{ij}^* - \lambda_{ij}^*\Big) c_{ij} + \sum_{i,j,k} \Big(\nu_i^{k\,*} - \nu_j^{k\,*} - \xi_{ij}^*\Big) r_{ij}^k \\ &+ \mathrm{E}_{\mathbf{h}} \left[\sum_{i,j,f} \lambda_{ij}^* C_{ij}\Big(\mathbf{h}^f, \mathbf{p}^f(\mathbf{h})\Big) - \mu_i^* p_{ij}^f(\mathbf{h}) \right].\end{aligned} \tag{11}$$

Layers at terminal T_i

$a_i^{k\,*} \leftarrow \max_{a_{\min\, i}^k \leq a_i^k \leq a_{\max\, i}^k} \left[U_i^k(a_i^k) - \nu_i^{k\,*} a_i^k\right]$	$a_i^{k\,*}$ for all k
$r_{ij}^{k\,*} \leftarrow \max_{0 \leq r_{ij}^k \leq r_{\max}} \left[\left(\nu_i^{k\,*} - \nu_j^{k\,*} - \xi_{ij}^*\right) r_{ij}^k\right]$	$r_{ij}^{k\,*}$ for $j \in n(i)$ and all k
$c_{ij}^* \leftarrow \max_{0 \leq c_{ij} \leq c_{\max}} \left[\left(\xi_{ij}^* - \lambda_{ij}^*\right) c_{ij}\right]$	c_{ij}^* for $j \in n(i)$
$p_i^* \leftarrow \max_{0 \leq p_i \leq p_{\max\, i}} \left[\mu_i^* p_i - V_i(p_i)\right]$	p_i^*
$\mathbf{p}^*(\mathbf{h}) \leftarrow \max_{0 \leq p_{ij}^f(\mathbf{h}) \leq p_{\max}} \left[\sum_{i,j,f} \lambda_{ij}^* C_{ij}\Big(\mathbf{h}^f, \mathbf{p}^f(\mathbf{h})\Big) - \mu_i^* p_{ij}^f(\mathbf{h})\right]$	

Fig. 1. Having zero duality gap the wireless networking problem can be separated in layers without loss of optimality. Therefore, we can consider separate optimization problems to determine arrival rates $a_i^{k\,*}$, link rates $r_{ij}^{k\,*}$, link capacities c_{ij}^*, and average transmitted power p_i^*. The physical layer problem can be further separated in per-fading-state subproblems. It cannot, alas, be separated in per-terminal problems for general link capacity functions $C_{ij}\Big(\mathbf{h}^f, \mathbf{p}^f(\mathbf{h})\Big)$. Thus, the challenge in wireless networking is not as much in cross-layer optimization as in cross-terminal optimization of the physical layer.

The zero duality gap implies that if $\mathbf{\Lambda}^*$ is known we can, instead of solving (1), solve the (separable) problem

$$P = D = d[\mathbf{\Lambda}^*] = \max_{\mathbf{X}, \mathbf{p}(\mathbf{h})} \mathcal{L}[\mathbf{X}, \mathbf{p}(\mathbf{h}), \mathbf{\Lambda}^*] \tag{12}$$

where the maximization is constrained to the $\mathbf{X}$ and $\mathbf{p}(\mathbf{h})$ that satisfy the box constraints.

Because primal variables are decoupled in the Lagrangian $\mathcal{L}[\mathbf{X}, \mathbf{p}(\mathbf{h}), \mathbf{\Lambda}^*]$ [cf. (11)], the maximization required in (12) can be split into smaller maximization problems involving less variables. This separability can be used to prove the next theorem; see [4] for the proof.

Theorem 2 Layer separability. *Let λ_{ij}^*, μ_i^*, $\nu_i^{k\,*}$, and ξ_{ij}^* denote the optimal dual variables that solve* (8)*. Consider the sub-problems*

$$P(a_i^k) = \max_{a_{\min\, i}^k \leq a_i^k \leq a_{\max\, i}^k} \left[U_i^k(a_i^k) - \nu_i^{k\,*} a_i^k\right] \tag{13}$$

$$P(r_{ij}^k) = \max_{0 \leq r_{ij}^k \leq r_{\max}} \left[\left(\nu_i^{k\,*} - \nu_j^{k\,*} - \xi_{ij}^*\right) r_{ij}^k\right] \tag{14}$$

$$P(c_{ij}) = \max_{0 \leq c_{ij} \leq c_{\max}} \left[\left(\xi_{ij}^* - \lambda_{ij}^*\right) c_{ij}\right] \tag{15}$$

$$P(p_i) = \max_{0 \leq p_i \leq p_{\max\, i}} \left[\mu_i^* p_i - V_i(p_i)\right]. \tag{16}$$

Define further the per-fading state optimal power allocation problem

$$\mathbf{p}(\mathbf{h}) = \max_{0 \leq p_{ij}^f(\mathbf{h}) \leq p_{\max}} \sum_{i,j,f} \lambda_{ij}^* C_{ij}\Big(\mathbf{h}^f, \mathbf{p}^f(\mathbf{h})\Big) - \mu_i^* p_{ij}^f(\mathbf{h}) \tag{17}$$

Then, the optimal utility yield P in (1) *is given by*

$$P = \sum_{i,k} P(a_i^k) + \sum_{i,j,k} P(r_{ij}^k) + \sum_{i,j} P(c_{ij}) + \sum_i P(p_i) + \mathrm{E}_{\mathbf{h}}\Big[P[\mathbf{p}(\mathbf{h})]\Big] \tag{18}$$

i.e., the primal problem (1) *can be separated into the (sub-) problems* (13)-(17) *without loss of optimality.*

The rate problem in (13) dictates the amount of traffic allowed into the network. It therefore solves the flow control problem at the transport

layer; see Fig. 1. Likewise, (14) represents the network layer routing problem, (15) determines link-level capacities at the data link layer and (16) is the (link layer) average power control problem. Eq. (17) represents resource allocation at the physical layer. Therefore, it is a consequence of Theorem 2 that layering, in the sense of problem separability as per (18) is optimal in faded wireless networks.

IV. Computation of optimal Lagrange multipliers

Solving the optimal wireless networking problem in (1) can be reduced to finding the optimal dual variables $\mathbf{\Lambda}^*$ of (8). Because the dual function $g(\mathbf{\Lambda})$ is convex, descent algorithms can be used to find $\mathbf{\Lambda}^*$. However, $g(\mathbf{\Lambda})$ need not be differentiable, and it certainly will not be in some cases. The challenge is therefore to find such descent direction. This prompts the definition of subgradient that we introduce next.

Definition 1 Subgradient *We say that $\check{\boldsymbol{g}}(\mathbf{\Lambda}_0)$ is a subgradient of the convex dual function $g(\mathbf{\Lambda})$ at $\mathbf{\Lambda}=\mathbf{\Lambda}_0$ if for every $\mathbf{\Lambda}\geq 0$ we have*

$$g(\mathbf{\Lambda}) \geq g(\mathbf{\Lambda}_0) + \check{\boldsymbol{g}}^T(\mathbf{\Lambda}_0)(\mathbf{\Lambda}-\mathbf{\Lambda}_0). \tag{19}$$

The hyperplane $g(\mathbf{\Lambda}_0) + \check{\boldsymbol{g}}^T(\mathbf{\Lambda}_0)(\mathbf{\Lambda}-\mathbf{\Lambda}_0) = 0$ defined by the subgradient direction $\check{\boldsymbol{g}}(\mathbf{\Lambda}_0)$ and the point $(\mathbf{\Lambda}_0, g(\mathbf{\Lambda}_0))$ supports $g(\mathbf{\Lambda})$ in the sense that it touches $g(\mathbf{\Lambda})$ at $\mathbf{\Lambda}=\mathbf{\Lambda}_0$ and is below $g(\mathbf{\Lambda})$ at any other point. The fundamental property of a subgradient is that it always points toward the optimal argument. Formally, let $\mathbf{\Lambda}=\mathbf{\Lambda}^*$ in (19), and reorder terms to obtain

$$\check{\boldsymbol{g}}^T(\mathbf{\Lambda}_0)(\mathbf{\Lambda}_0-\mathbf{\Lambda}^*) \geq g[\mathbf{\Lambda}_0] - g(\mathbf{\Lambda}^*) = g[\mathbf{\Lambda}_0] - D \geq 0 \tag{20}$$

where we replaced $g(\mathbf{\Lambda}^*)=D$ and use the fact that D is the minimum value of $g(\mathbf{\Lambda})$. Given that the inner product of $\check{\boldsymbol{g}}^T(\mathbf{\Lambda}_0)(\mathbf{\Lambda}_0-\mathbf{\Lambda}^*)$ is positive, (20) proves that the angle between $\check{\boldsymbol{g}}^T(\mathbf{\Lambda}_0)$ and $\mathbf{\Lambda}_0-\mathbf{\Lambda}^*$ is less than $\pi/2$. Therefore, the negative of the subgradient points "towards", i.e., with an angle of less than $\pi/2$ radians, the optimal argument.

A subgradient of the dual function can be obtained from the arguments that maximize the Lagrangian for given $\mathbf{\Lambda}$ multipliers as detailed by the following theorem. This as well as subsequent results in Theorems 4 - 6 are known for finite-dimensional optimization problems, [5]. We present them here for the (infinite-dimensional) variational problem (5). The proofs here are patterned after those in [5]; see [4].

Theorem 3 *With $\mathbf{\Lambda}_0 \geq 0$ an arbitrary dual variable and $\mathbf{X}^\dagger(\mathbf{\Lambda}_0)$ primal variables that maximize the Lagrangian function in* (6) *for $\mathbf{\Lambda}=\mathbf{\Lambda}_0$*

$$\Big(\mathbf{X}^\dagger(\mathbf{\Lambda}_0), \mathbf{p}^\dagger(\mathbf{h},\mathbf{\Lambda}_0)\Big) \in \arg \max_{(\mathbf{X},\mathbf{p}(\mathbf{h}))\in B} \mathcal{L}[\mathbf{X},\mathbf{p}(\mathbf{h}),\mathbf{\Lambda}_0]. \tag{21}$$

Then a subgradient of the dual function at $\mathbf{\Lambda}=\mathbf{\Lambda}_0$ is given by

$$\check{\boldsymbol{g}}(\mathbf{\Lambda}_0) = \mathbf{h}\left[\mathbf{X}^\dagger(\mathbf{\Lambda}_0), \mathbf{p}^\dagger(\mathbf{h},\mathbf{\Lambda}_0)\right]. \tag{22}$$

In general, there is more than one argument maximizing (21). Therefore the $\arg\max$ operator does not specify a value but a set, as signified by the $\in$ symbol in (21). We interpret $\mathbf{X}^\dagger(\mathbf{\Lambda})$ as any element of this set.

A. Subgradient descent algorithm

A descent algorithm to compute optimal multipliers $\mathbf{\Lambda}^*$ and minimum dual value $D=P$ is obtained using the subgradient of the dual function described in Theorem 3. With iterations indexed on t, start with given dual variables $\mathbf{\Lambda}(t)$ and compute arguments $[\mathbf{X}(t),\mathbf{p}(\mathbf{h},t)]$ that maximize the Lagrangian in (6),

$$\begin{aligned}[\mathbf{X}(t),\mathbf{p}(\mathbf{h},t)] &\in \arg \max_{(\mathbf{X},\mathbf{p}(\mathbf{h}))\in B} \mathcal{L}[\mathbf{X},\mathbf{p}(\mathbf{h}),\mathbf{\Lambda}(t)] \\ &= \arg \max_{(\mathbf{X},\mathbf{p}(\mathbf{h}))\in B} f(\mathbf{X}) + \mathbf{\Lambda}^T(t)\mathbf{h}[\mathbf{X},\mathbf{p}(\mathbf{h})] \end{aligned} \tag{23}$$

Using (22) we have that a subgradient of the dual function at $\mathbf{\Lambda}=\mathbf{\Lambda}(t)$ is given by $\check{\boldsymbol{g}}(t) := \check{\boldsymbol{g}}[\mathbf{\Lambda}(t)] = \mathbf{h}[\mathbf{X}(t),\mathbf{p}(\mathbf{h},t)]$. Therefore, the dual variable is updated as

$$\begin{aligned}\mathbf{\Lambda}(t+1) &= \Big[\mathbf{\Lambda}(t) - \epsilon_t \check{\boldsymbol{g}}(t)\Big]^+ \\ &= \Big[\mathbf{\Lambda}(t) - \epsilon_t \mathbf{h}[\mathbf{X}(t),\mathbf{p}(\mathbf{h},t)]\Big]^+. \end{aligned} \tag{24}$$

where $[\cdot]^+$ denotes the componentwise maximum of 0 and the value between parenthesis and ϵ_t is a properly selected step-size; see Theorems 5 and 6. Because the negative of the subgradient $-\check{\boldsymbol{g}}(t)$ points towards $\mathbf{\Lambda}^*$ it is expected that iterates of (24) are progressively closer to $\mathbf{\Lambda}^*$. As the following standard result shows, this is indeed true in some sense.

Theorem 4 *Consider the subgradient descent iteration in* (24) *and define the dual value at iteration t as $g(t) := g[\mathbf{\Lambda}(t)]$. Let $G := \max_{(\mathbf{X},\mathbf{p}(\mathbf{h}))\in B} \big\|\mathbf{h}[\mathbf{X},\mathbf{p}(\mathbf{h})]\big\|$ be a bound on the norm of the subgradient of the dual function. The 2-norm distances $\big\|\mathbf{\Lambda}(t)-\mathbf{\Lambda}^*\big\|$ of iterates $\mathbf{\Lambda}(t)$ to the optimal argument $\mathbf{\Lambda}^*$ at times t and $t+1$ satisfy the relation*

$$\big\|\mathbf{\Lambda}(t+1)-\mathbf{\Lambda}^*\big\|^2 \leq \big\|\mathbf{\Lambda}(t)-\mathbf{\Lambda}^*\big\|^2 + \epsilon_t^2 G^2 - 2\epsilon_t\big[g(t)-D\big]. \tag{25}$$

Because all primal variables are constrained to the bounded region B, the bound G on the subgradient norm is finite. Given that D denotes the minimum of $g(t)$ it is clearly true that $g(t)-D\geq 0$. Thus, at each iteration the distance between the current dual iterate $\mathbf{\Lambda}(t)$ and the optimal dual variable $\mathbf{\Lambda}^*$ is reduced by (at least) $2\epsilon_t\big[g(t)-D\big]$ and increased by (at most) $\epsilon_t^2 G^2$. For small ϵ_t we expect the reduction $2\epsilon_t\big[g(t)-D\big]$ to dominate the increase $\epsilon_t^2 G^2$ and consequently for $\mathbf{\Lambda}(t)$ to approach $\mathbf{\Lambda}^*$.

For fixed step size $\epsilon_t=\epsilon$ for all t, however, there is a limit on how close $\mathbf{\Lambda}(t)$ can come to $\mathbf{\Lambda}^*$. For any given ϵ, $\epsilon^2 G^2$ will eventually become larger than $2\epsilon\big[g(t)-D\big]$ preventing the optimality gap $\big[g(t)-D\big]$ to go to zero. This is not a limitation of the analysis but a consequence of the fact that for non-differentiable functions the norm of the subgradient $\|\check{\boldsymbol{g}}(\mathbf{\Lambda})\|$ does not necessarily vanish as $\mathbf{\Lambda}$ approaches $\mathbf{\Lambda}^*$. Therefore, the iteration in (IV-A) is not convergent. Rather, the iterates $\mathbf{\Lambda}(t)$ approach $\mathbf{\Lambda}^*$ until $\epsilon^2 G^2$ starts dominating $2\epsilon\big[g(t)-D\big]$.

This motivates the use of vanishing step-size sequences, i.e., $\lim_{t\to\infty}\epsilon_t = 0$, so that as the duality gap $\big[g(t)-D\big]$ approaches zero, so does ϵ_t. This allows for $2\epsilon_t\big[g(t)-D\big]$ to always dominate $\epsilon_t^2 G^2$ leading to the following classical convergence result.

Theorem 5 *Consider the subgradient descent iteration in* (24) *with vanishing step sizes ϵ_t. Require the sum of step sizes to be divergent, i.e.,*

$$\sum_{t=1}^{\infty}\epsilon_t = \infty, \quad \lim_{t\to\infty}\epsilon_t = 0. \tag{26}$$

Then, the limit of the sequence of iterates $\mathbf{\Lambda}(t)$ exists and

$$\lim_{t\to\infty}\mathbf{\Lambda}(t) = \mathbf{\Lambda}^*. \tag{27}$$

The conditions (26) on the step-size sequence are certainly minimal. E.g., sequences of the form $\epsilon_t = \epsilon_1/(t+\epsilon_2)^\alpha$ with $\alpha>0$ for arbitrary positive constants ϵ_1 and ϵ_2 satisfy (26). Nonetheless, constant step sizes $\epsilon_t=\epsilon$ for all t, are still desirable in some cases. In this case it can be proven that as $t\to\infty$, $\mathbf{\Lambda}(t)$ "stays close" to $\mathbf{\Lambda}^*$.

Theorem 6 *Consider the subgradient descent iteration in* (24) *with constant step sizes $\epsilon_t=\epsilon$ for all t. With $G := \max_{(\mathbf{X},\mathbf{p}(\mathbf{h}))\in B} \big\|\mathbf{h}[\mathbf{X},\mathbf{p}(\mathbf{h})]\big\|$ the subgradient norm bound of Theorem 4, it holds:*

(i) *The best dual value at time t,* $g_{\text{best}}(t) := \min_{s\in[0,t]} g(t)$*, converges to a value within* $\epsilon G^2/2$ *of the optimum* D*, i.e.,*

$$\lim_{t\to\infty} g_{\text{best}}(t) - D \leq \epsilon G^2/2. \tag{28}$$

(ii) *The average of the dual iterates* $\bar{\mathbf{\Lambda}}(t) := (1/t)\sum_{s=1}^{t}\mathbf{\Lambda}(s)$*, converges to a point whose optimality gap is less than* $\epsilon G^2/2$*, i.e.,*

$$g\left[\lim_{t\to\infty}\bar{\mathbf{\Lambda}}(t)\right] - D \leq \epsilon G^2/2. \tag{29}$$

As commented after Theorem 4, the subgradient descent algorithm (23) - (24) does not necessarily converge for fixed step sizes. Nonetheless, a reasonable approximation to $\mathbf{\Lambda}^*$ is achieved by $\mathbf{\Lambda}_{\text{best}}(t)$ defined as the argument for which $g[\mathbf{\Lambda}_{\text{best}}(t)] = g_{\text{best}}(t)$. The quality of this approximation is measured in the optimality gap $g[\mathbf{\Lambda}_{\text{best}}(t)] - D$ that can be made arbitrarily small with adequately selected step size ϵ.

By its own definition $\mathbf{\Lambda}_{\text{best}}(t)$ is the best approximation to $\mathbf{\Lambda}^*$ that can be obtained by (23) - (24) with fixed step sizes. Finding $\mathbf{\Lambda}_{\text{best}}(t)$, though, requires access to the dual values $g(t)$, which might not be available; see e.g., Section IV-B. In such circumstances a similarly good approximation to $\mathbf{\Lambda}^*$ is the average $\bar{\mathbf{\Lambda}}(t)$ of iterates $\mathbf{\Lambda}(t)$.

B. Layers and layer interfaces

Implementing the subgradient descent iteration (23)-(24) uncovers details in the interaction between layers. Returning to the explicit Lagrangian $\mathcal{L}[\mathbf{X}, \mathbf{p}(\mathbf{h}), \mathbf{\Lambda}(t)]$ notation used in (10) and reordering terms in the same manner as done for $\mathbf{\Lambda} = \mathbf{\Lambda}^*$ in (11) the primal iteration in (23) becomes

$$\begin{aligned}\mathcal{L}[\mathbf{X}, \mathbf{p}(\mathbf{h}), \mathbf{\Lambda}(t)] =& \sum_{i,k}\Big(U_i^k(a_i^k) - \nu_j^k(t)a_j^k\Big) + \sum_i\Big(\mu_i(t)p_i - V_i(p_i)\Big)\\ &+ \sum_{i,j}\Big(\xi_{ij}(t) - \lambda_{ij}(t)\Big)c_{ij} + \sum_{i,j,k}\Big(\nu_i^k(t) - \nu_j^k(t) - \xi_{ij}(t)\Big)r_{ij}^k\\ &+ \mathrm{E}_{\mathbf{h}}\left[\sum_{i,j,f}\lambda_{ij}(t)C_{ij}\Big(\mathbf{h}^f, \mathbf{p}^f(\mathbf{h})\Big) - \mu_i(t)p_{ij}^f(\mathbf{h})\right].\end{aligned} \tag{30}$$

Proceeding with the same line of argument used in Theorem 2, the maximization required for the primal iteration (23) can be separated in specific subproblems associated with each of these variables. The elements of $\mathbf{X}(t)$ in (23) are thus

$$a_i^k(t) = \max_{a_{\min\, i}^k \leq a_i^k \leq a_{\max\, i}^k}\left[U_i^k(a_i^k) - \nu_i^k(t)a_i^k\right] \tag{31}$$

$$r_{ij}^k(t) = \max_{0\leq r_{ij}^k \leq r_{\max}}\left[\Big(\nu_i^k(t) - \nu_j^k(t) - \xi_{ij}(t)\Big)r_{ij}^k\right] \tag{32}$$

$$c_{ij}(t) = \max_{0\leq c_{ij}\leq c_{\max}}\left[\left(\xi_{ij}(t) - \lambda_{ij}(t)\right)c_{ij}\right] \tag{33}$$

$$p_i(t) = \max_{0\leq p_i \leq p_{\max\, i}}\left[\mu_i(t)p_i - V_i(p_i)\right] \tag{34}$$

Also, in the last term of (30), the maximization can be brought into the expected value operator. The elements of the power distribution $\mathbf{p}(\mathbf{h}; t)$ in (23) can thus be computed separately for each fading state $\mathbf{h}$, i.e.,

$$\mathbf{p}(\mathbf{h}; t) = \max_{0\leq p_{ij}^f(\mathbf{h}) \leq p_{\max}}\left[\sum_{i,j,f}\lambda_{ij}(t)C_{ij}\Big(\mathbf{h}^f, \mathbf{p}^f(\mathbf{h})\Big) - \mu_i(t)p_{ij}^f(\mathbf{h})\right]. \tag{35}$$

The subgradient $\breve{\mathbf{g}}(t) = \mathbf{h}[\mathbf{X}(t), \mathbf{p}(\mathbf{h}; t)]$ can be likewise separated. The components of the vector function $\mathbf{h}[\mathbf{X}(t), \mathbf{p}(\mathbf{h}; t)]$ are as specified in (2)-(4). Therefore, the dual iteration (24) can be written explicitly as

$$\nu_i^k(t+1) = \nu_i^k(t) + \epsilon_t\left[\sum_{j\in n(i)}\Big(r_{ij}^k(t) - r_{ji}^k(t)\Big) - a_i^k(t)\right] \tag{36}$$

$$\xi_{ij}(t+1) = \xi_{ij}(t) + \epsilon_t\left[c_{ij}(t) - \sum_k r_{ij}^k(t)\right] \tag{37}$$

$$\lambda_{ij}(t+1) = \lambda_{ij}(t) + \epsilon_t\left[\mathrm{E}_{\mathbf{h}}\left[\sum_{f\in\mathcal{F}}C_{ij}\Big(\mathbf{h}^f, \mathbf{p}^f(\mathbf{h}; t)\Big)\right] - c_{ij}(t)\right] \tag{38}$$

$$\mu_i(t+1) = \mu_i(t) + \epsilon_t\left[p_i - \mathrm{E}_{\mathbf{h}}\left[\sum_{j\in\mathcal{N}(i)}\sum_{f\in\mathcal{F}}p_{ij}^f(\mathbf{h}; t)\right]\right] \tag{39}$$

The argument to be optimized in (31) is solely parameterized by $\nu_i^k(t)$. Thus, given the multiplier $\nu_i^k(t)$ associated with the flow conservation constraint, $a_i^k(t)$ is determined. Likewise, $r_{ij}^k(t)$ is determined by flow conservation multipliers $\nu_i^k(t)$ and $\nu_j^k(t)$ and link capacity constraints multipliers $\xi_{ij}(t)$. In general, all the primal iterations (31)-(35) depend on multipliers associated with no more than two types of constraints. The dual iterations (38)-(39) have a similar property. The update of $\mu_i(t)$ in (39) for instance, depends on the total power $p_i(t)$ and the power distribution $\mathbf{p}(\mathbf{h}; t)$. In general, the multipliers' updates depend on no more than two different types of primal variables.

The fact that primal and dual variable updates depend on only two types of variables prompts a interpretation of (31)-(39) in terms of layers and layer interfaces. The flow control problem (31) is associated with the transport layer, the link rate problem (32) with the routing layer, link capacity (33) and power control (34) problems are solved at the link layer and power distribution (35) pertains to the physical layer. Because the dual variables in (31)-(35) are not optimal, it becomes necessary to communicate variables across layer interfaces. These interfaces are defined by the dual variable updates (36)-(39). Thus, the update of multipliers $\nu_i^k(t)$ in (36) defines the interface between the network and transport layer and (37) the link to network layer interface. Because there are two problems being solved at the link layer, (38) defines the interface between the physical layer and the link capacity subproblem and (39) between physical layer and power control subproblem.

Fig. 2 shows a schematic representation of the layers and their interfaces. At the bottom of the stack the physical layer solves (35) to find the power distribution $\mathbf{p}(\mathbf{h}; t)$. Due to coupling that in general is introduced by the function $C_{ij}\left(\mathbf{h}^f, \mathbf{p}^f(\mathbf{h})\right)$ the physical layer optimization cannot be separated in per-terminal optimization problems and is therefore represented as a common substrate supporting per-terminal stacks. To compute $\mathbf{p}(\mathbf{h}; t)$ the physical layer receives multipliers $\lambda_{ij}(t)$ and $\mu_i(t)$ from the physical-link interface.

At the link layer each terminal maintains variables representing the average link capacities $c_{ij}(t)$ to neighbors T_j, $j \in n(i)$ and the average transmitted power $p_i(t)$. These are computed by solving (33) and (34). In turn, this requires dual variables $\lambda_{ij}(t)$ and $\mu_i(t)$ communicated from the physical-link interface and $\xi_{ij}(t)$ communicated from the link-network interface.

As is true for physical and link, all layers compute network variables of interest based on dual variables received from adjacent interfaces. That way, the network layer maintains variables r_{ij}^k for neighbors $j \in n(i)$ and flows k that determine local routing decisions. These are updated as per (32) using multipliers $\xi_{ij}(t)$ received from the link-network interface and ν_i^k and ν_j^k, $j \in n(i)$ from the network-transport interface. The transport layer, finally, keeps variables a_i^k determining the rate at which packets pertaining to the k-th flow are accepted into the network by terminal T_i. These are updated as per (31) using multipliers ν_i^k received from the network-transport interface.

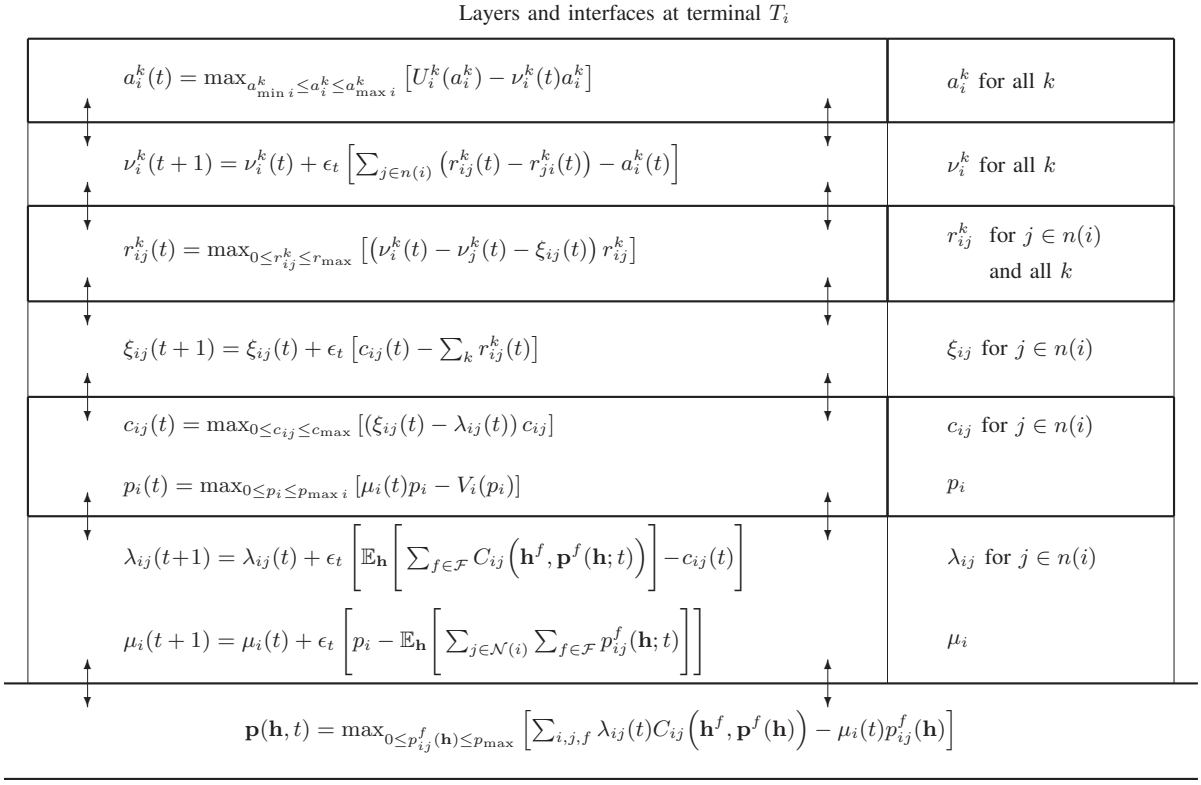

Fig. 2. The subgradient descent iteration (31)-(39) can be interpreted in terms of layers and layer interfaces. Layers keep variables of interest to the network, e.g., link transmission rates r_{ij}^k at the network layer, that they update according to primal iterations (31)-(35). Layer interfaces maintain (auxiliary) dual variables updated as per the dual iterations (36)-(39). Communication of variables across layers and interfaces is restricted to adjacent entities; i.e., layers receive variables from, and transmit to, adjacent interfaces. Interfaces exchange variables with adjacent layers. Note that in general the physical layer optimization problem cannot be separated in per-terminal problems.

Interfaces in turn, update dual variables using information received from adjacent layers. The physical-link interface computes dual variables $\lambda_{ij}(t)$ for $j \in n(i)$ and $\mu_i(t)$. This is fitting because the multipliers $\lambda_{ij}(t)$ and $\mu_i(t)$ are respectively associated with the link capacity (2) and power (3) constraints that relate physical-level variables $\mathbf{p}(\mathbf{h})$ and link-level quantities c_{ij} and p_i. The updates (38) and (39) carried at the physical-link interface require variables $\mathbf{p}(\mathbf{h};t)$ communicated from the physical layer and variables $c_{ij}(t)$ and $p_i(t)$ from the link layer.

Likewise, the link-network interface keeps one multiplier $\xi_{ij}(t)$ per neighbor T_j, $j \in n(i)$. These are associated with the rate constraints in (4) that couple link variables c_{ij} and network variables r_{ij}^k. Updates of $\xi_{ij}(t)$ are specified in (37), being determined by variables $c_{ij}(t)$ and $r_{ij}^k(t)$ respectively communicated from the link and network layers. The network-transport interface, finally, maintains dual variables $\nu_i^k(t)$ associated with the flow conservation constraints in (4) that couple network r_{ij}^k and transport a_i^k variables. These $\nu_i^k(t)$ variables are updated as per (36) using $r_{ij}^k(t)$ and $a_i^k(t)$ received from the network and transport layer respectively.

As time progresses, interfaces' variables $\lambda_{ij}(t)$, $\mu_i(t)$, $\xi_{ij}(t)$ and $\nu_i^k(t)$ converge to optimal multipliers λ_{ij}^*, μ_i^*, ξ_{ij}^* and $\nu_i^{k\,*}$ [cf. Theorem 5] – or a point close to them if the step size ϵ_t is fixed [cf. Theorem (6)] – enabling computation of optimal network variables $\mathbf{p}^*(\mathbf{h})$, p_i^*, c_{ij}^* $r_{ij}^{k\,*}$ and $a_i^{k\,*}$.

V. Concluding remarks

General wireless networking optimization problems in the presence of fading are non-convex but have zero duality gap [4]. This paper has described how the separability of wireless networking problems into layers follows from this result. An architecture based in layers and layer interfaces was shown to be optimal as a consequence of the implementation of a subgradient descent algorithm for the dual function. Similar architectures have been reported elsewhere. The main contribution of this paper is to show that this architecture is optimal[1].

References

[1] L. Georgiadis, M. J. Neely, and L. Tassiulas, "Resource allocation and cross-layer control in wireless networks," *Foundations and Trends in Networking*, vol. 1, no. 1, pp. 1–144, 2006.

[2] X. Lin, N. B. Shroff, and R. Srikant, "A tutorial on cross-layer optimization in wireless networks," *IEEE J. Sel. Areas Commun.*, vol. 24, no. 8, pp. 1452–1463, August 2006.

[3] M. Chiang, S. H. Low, R. A. Calderbank, and J. C. Doyle, "Layering as optimization decomposition," *Proc IEEE*, vol. 95, no. 1, pp. 255–312, January 2007.

[4] Alejandro Ribeiro and Georgios B. Giannakis, "Separation theorems of wireless networking," *IEEE Trans. Inf. Theory*, December 2007 (submitted), Available at http:// www.seas.upenn.edu/˜aribeiro/.

[5] Naum Zuselevich Shor, *Minimization Methods for Non-Differentiable Functions*, Springer-Verlag, Berlin, Heilderberg, Germany, 1985.

[1]Work in this paper was prepared through collaborative participation in the Communications and Networks Consortium sponsored by the U. S. Army Research Laboratory under the Collaborative Technology Alliance Program, Cooperative Agreement DAAD19-01-2-0011. The U. S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation thereon.