Resource Management in Wireless Networks and the Smart Power Grid

A DISSERTATION
SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL
OF THE UNIVERSITY OF MINNESOTA
BY

Nikolaos Gatsis

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
Doctor of Philosophy

GEORGIOS B. GIANNAKIS, ADVISOR

June 2012

Acknowledgements

Before starting the presentation of the thesis material, I would like to take the opportunity to thank all these people whose invaluable help and support brought this thesis into existence.

My sincere and deepest respect and gratitude goes to my advisor Prof. Georgios B. Giannakis. I would like to thank him for his constant support and encouragement, as well as for offering many hours to help me become a more mature researcher and better understand and appreciate important technical matters in several research areas.

Due thanks go to Professors Mos Kaveh, Tom Luo, Nikos Sidiropoulos, Panos Stinis, and Bruce F. Wollenberg for agreeing to serve on my committee.

The work in this PhD thesis would not have been possible without the help and insightful suggestions of Professors Antonio Marques and Alejandro Ribeiro, whom I collaborated with during the last years. The material of this thesis was also benefited from discussions with current and former members of the SPiNCOM group: Professor Xin Wang, Dr. Seung-Jun Kim, Dr. Alfonso Cano, Dr. Yingqun Yu, Dr. Tairan Wang, Dr. Xiliang Luo, Dr. Vassileios Kekatos, Dr. Daniele Angelosante, Dr. Emiliano Dall'Anese, Dr. Ioannis Schizas, Dr. Eric Msechu, Ketan Rajawat, Gonzalo Mateos, Hao Zhu, Yu Zhang, Juan-Andres Bazerque, Pedro Forrero, Morteza Mardani, and Brian Baingana. I would also like to thank these people for their support.

I would also like to thank my parents for their continued support and encouragement.

Finally, I would like to acknowledge the grants that supported financially our research.

Nikolaos Gatsis Minneapolis, April 25, 2012

Abstract

Optimal resource management is a crucial task in a plethora of scientific fields, including wireless communication and electric power networks, where it ensures efficient operation and user satisfaction. The pressing need to modernize the aging power grid has culminated to a vision encouraging interaction of the end users with the grid through demand response, which amounts to electricity end users adapting their power consumption in response to pricing schemes varying over time (e.g., every hour or day). By the same token, delivering data, voice, and video seamlessly over wireless networks with the quality-of-service demanded by today's multimedia applications requires optimal link-adaptive allocation of the available resources, e.g., power, to the different network nodes and layers. This thesis develops algorithms for (a) scheduling of demand response in the smart power grid, and (b) cross-layer wireless network design.

First, demand response is considered in a multiple-residence setup. The utility company adopts a cost function representing the cost of providing energy to end users. Each residential end user has a base load, two types of adjustable loads, and possibly a storage device. The first load type must consume a specified amount of energy over the scheduling horizon, but the consumption can be adjusted across different slots. Charging a plug-in hybrid electric vehicle is an example. The second type does not entail a total energy requirement, but operation away from a user-specified level results in user dissatisfaction. The research issue amounts to minimizing the electricity provider cost plus the total user dissatisfaction, subject to the individual constraints of the loads. The problem can be solved by a distributed subgradient method. The utility company and the end users exchange information through the Advanced Metering Infrastructure (AMI)—a two-way communication network—in order to converge to the optimal amount of electricity production and the optimal power consumption schedule. The algorithm finds near-optimal schedules even when AMI messages are lost, which can happen in the presence of malfunctions or noise in the communications network. The algorithm amounts to a subgradient iteration with outdated Lagrange multipliers, for which convergence results of wide scope are established.

Next, attention is turned to an energy consumption scheduling problem for a single residential end user, but with an added complexity. Each adjustable load is interruptible in the sense that the load can be either operated (resulting in nonzero power consumption), or not operated (resulting in zero power consumption). The task amounts to minimizing the cost of electricity plus user dissatisfaction, subject to individual load consumption constraints. The resulting problem is nonconvex, but it is shown to have zero duality gap if a continuous-time horizon is considered. This opens up the possibility of using Lagrangian dual algorithms without loss of optimality in order to come up with efficient demand response scheduling schemes.

As regards wireless networking, the challenge is to jointly optimize application-level rates, routes, link capacities, power consumption, and power allocation across frequency tones, neighboring terminals, and fading states. The physical layer is interference-limited, whereby network terminals treat interference as noise. Provably convergent algorithms yield (near-)optimal end-to-end rates, multicommodity flows, link capacities, and average powers. These design variables are obtained offline, and are subsequently used for control during network operation. Moreover, physical layer power allocation algorithms that are seamlessly integrated into layered architectures are developed using successive convex approximations.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

Optimal resource management is a crucial task in a plethora of scientific fields, including the electric power grid and wireless communication networks, where it ensures efficient operation and user satisfaction.

Although the North American power grid has been named the most important engineering achievement of the 20th century by the National Academy of Engineering [1], the power grid today faces major challenges (as well as opportunities) that were not present when it was built. Those pertain to stability, security, consumer choices, and environmental issues [2]. The smart grid *vision* endeavors to capitalize upon state-of-the-art information technologies to address those challenges via three important avenues [3]: (a) intelligent control strategies to efficiently coordinate the power transfer from generation to consumption; (b) advanced sensor and communication infrastructure enabling interaction of end users with the grid; and (c) environment-friendly power technologies, including renewable resources for electricity generation. Realizing this vision and making a significant impact to the smart grid agenda calls for sophisticated resource planning and management of the grid operation.

By the same token, the recent years have seen a rapid growth and demand for fast and error-resilient telecommunication services delivered over wireless networks. To satisfy this demand, quality-of-service provisioning for diverse applications, advanced techniques for coping with the wireless interface, and judicious methods for allocating the available power and bandwidth resources to different network entities, are of paramount importance. To this end, nonlinear optimization has been fruitfully brought into the

design of wireless networks—see e.g., [4–7]—and is instrumental in devising efficient link-adaptive resource allocation algorithms for various wireless networking setups.

This thesis develops algorithms for efficient resource management in both electrical power networks and wireless communication networks. The particular motivation, context, and contributions of this thesis are described in Section 1.1 for power networks and Section 1.2 for wireless networks. The underpinning theme is optimization methodologies, and in particular, Lagrangian duality approaches. The contributions in this respect are also described in the two ensuing sections.

1.1 Demand-Side Management for the Smart Power Grid

The smart grid vision is to modernize the aging power grid infrastructure by capitalizing on the proven sensor, communication, and control technologies of today to address the pressing issues of environment, consumer demand, security and stability of energy generation, transportation, and consumption. One of the directions that are expected to be advanced toward this vision is enabling interaction of end users with the grid [3].

An important resource management task that is instrumental in enabling such interaction is *demand-side management* (DSM). DSM manifests changes in electric usage by the consumers [8], and has a much needed positive impact towards smoothing out the peak demand, increasing the system reliability, reducing generation cost—especially at peak times—and meeting pollution mandates.

DSM can be effected by load control in response to smart time-based, or time-varying, pricing schemes. These schemes are judiciously controlled by the utility companies to elicit desirable energy usage. Load control through pricing has also been termed *demand response* or *load response*, among others; see e.g., [9, 10]. Residential loads have the potential to offer significant benefits to this end, because they consist of loads that can for instance be adjusted—e.g., an air conditioning unit (A/C)—or be deferred for later. The advent of smart grid technologies have also made available energy storage devices (batteries) at the residential level [11]. These can be charged and discharged throughout the day according to residential needs, and constitute an additional device for load control.

DR is facilitated by deployment of *advanced metering infrastructure* (AMI), which

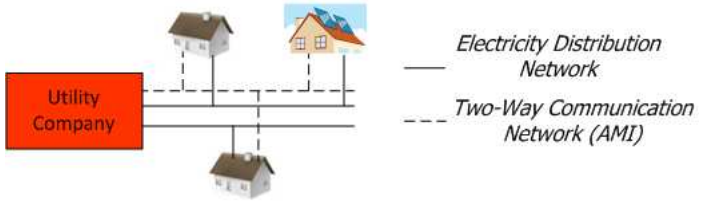


Figure 1.1: Electricity distribution network and AMI.

comprises a two-way communication network between utility companies and the end users (Fig. 1.1) [2, 12]. Smart meters installed at end users' premises are the AMI terminals at the end users' side. These measure not just the total power consumption, but also the power consumption profile throughout the day, and report it to the utility company at regular time intervals. The utility company sends pricing signals to the smart meters through the AMI (real-time pricing), for the smart meters to adjust power consumption profiles of the various residential electric devices, in order to minimize the electricity bill and maximize end user satisfaction.

This thesis focuses on residential DR with emphasis on jointly scheduling of devices with *diverse* power consumption profiles and requirements. Two setups are considered, as explained in the ensuing two subsections. The first is a cooperative DR scheme that involves multiple residential end users (Subsection 1.1.1), while the second pertains to a single residential end user (Subsection 1.1.2).

1.1.1 Cooperative Demand Response Scheduling

Cooperative DR in the thesis deals with optimal energy scheduling for load control of multiple residential loads, which comprise various types of devices, and is the theme of Chapter 2. Economical operation of the utility company is accounted for. Distributed algorithms which leverage the AMI and have guaranteed convergence to optimal schedules even under AMI outages are also developed.

Energy scheduling problems for multiple residences consider end users jointly maximizing the satisfaction from power consumption offset by the total cost of electricity from all residences, which is the *social welfare* [13, 14]. The cost of electricity and the power requests are known ahead of the scheduling horizon. In the aforementioned setup,

the cost of electricity is minimized without accounting for user satisfaction in [15], while an extension where users have limited knowledge about the other users' power requests is studied in [16]. Energy scheduling with distributed storage in a game-theoretic (multi-user) setup is pursued in [17], where each end user maximizes its individual welfare. Scheduling for a single end user with storage is addressed in [18–20].

The present work deals with social welfare maximization for energy scheduling between a utility company and residential end users. It differs from [13–16], which focus on a single type of adjustable load, and typically assume convenient forms of objective functions, such as strictly convex or differentiable. Any convex objective can be accommodated in the present work—not necessarily strictly convex—such as piecewise linear cost of electricity. Each residential end user has two classes of adjustable loads, as well as a storage device. The first class must consume a specified total amount of energy over the scheduling horizon, but the consumption can be adjusted across different slots. An example is plug-in hybrid electric vehicle (PHEV) charging. The second class has adjustable power consumption without a total energy requirement, but operation of the load at reduced power results in dissatisfaction of the end user, with A/C being an example. These two classes reflect the two types of load that can offer residential load response, and it is therefore important to be captured jointly.

The resulting optimization problem is solved through a distributed subgradient algorithm. The algorithm entails exchange of information among the utility company and the end users, and has similar communication requirements as the ones in [13] and [14]. The utility company sends out Lagrange multipliers, which are readily interpreted as pricing signals, and each residence sends back total hourly energy consumption—but not individual appliance consumption. It is established that the distributed algorithm converges to optimal schedules, even if there are AMI lost messages in any of the two directions. In order to establish this result, the overall algorithm is cast in a very general setup as an asynchronous subgradient algorithm with outdated Lagrange multipliers for separable convex programming. General convergence results are established for the Lagrange multipliers and primal optimization variables, by forming running averages (ergodic sequences) of the latter. Related results in the literature deal either with synchronous algorithms [21–23] or asynchronous algorithms but without addressing the issue of primal recovery [24].

1.1.2 Energy Consumption Scheduling

Optimal energy scheduling of interruptible devices in a residential setup is the theme of Chapter 3. The interruptibility alludes to the fact that the device can be turned off and then on depending on the electricity cost, until it provides its service to the end user.

Energy scheduling of noninterruptible devices, which can lead to convex problems has been dealt with in [25–28]. Interruptible devices with discrete power consumption levels—leading to nonconvex problems—have also been considered using different optimization approaches [26, 28–31]. Stochastic counterparts where the task requests or the electricity prices are modeled as random processes have also been considered [28, 32].

As before, two types of adjustable loads are considered. The first must consume a specified total amount of energy over the scheduling horizon, but the consumption can be adjusted across the horizon. The second type of load has adjustable power consumption without a total energy requirement, but operation of the load at reduced power results in dissatisfaction of the end user. Each adjustable load is interruptible in the sense that the load can be either operated (resulting in a nonzero power consumption in a continuous interval), or not operated (resulting in zero power consumption). The resulting formulation is nonconvex, and distinct from the problems with discrete energy levels. Different from other works in the literature, the approach taken here relies on Lagrangian duality. The main findings are that: (a) formulating the problem over a continuous time horizon has zero duality gap; and (b) if the time is discretized, the problem has vanishing duality gap as the discretization becomes finer. The zero duality gap result relies on application of the Lyapunov convexity theorem [33], which has previously found application in digital subscriber line (DSL) networks [34], as well as wireless networks [35, 36].

1.2 Cross-Layer Resource Allocation for Wireless Networks

Cross-layer optimization is a paradigm whereby resources available at different layers of the network—transport, network, link, and physical—are jointly optimized (Fig. 1.2). The relevant optimization variables are rates injected at network nodes, routes, link rates, power, and bandwidth. Relevant approaches must also account for the random nature of the fading channel links. Various criteria are pertinent, such as throughput

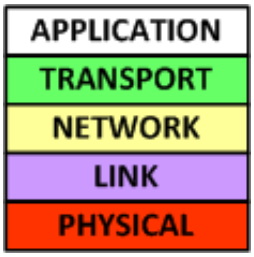


Figure 1.2: Layers in communication networks.

maximization, and delay or power minimization. The "best" operating point is defined as the solution of such an optimization problem, and protocols follow from algorithms used for its solution; see e.g., [4–7].

Chapter 4 of this thesis deals with optimal wireless network design in the presence of fading when the physical layer is interference-limited. All network terminals are allowed to transmit over the available frequency bands and non-intended transmissions are treated as noise. The goal is to find the optimal operating point that maximizes a network-wide performance metric. Finding such an optimal operating point requires determination of (optimal) end-to-end rates, routes, link capacities, average power consumption, as well as power allocation across frequency tones, neighboring terminals, and fading states. The large number of variables involved hints at the difficulty of finding such an operating point. This difficulty is confirmed by the fact that the optimization problem associated with this definition of optimal wireless network is *not* convex.

This complication is typically addressed through the introduction of suitable simplifications. Joint optimization of transport-level rates and physical layer power allocation is considered using a high-signal-to-interference-plus-noise-ratio (SINR) approximation with deterministic channels [37, Sec. 3.4]. A high-SINR assumption is also used to lend tractability to a more general joint congestion control, routing, link rate, and power control problem [38]. The more accurate $\log(1 + \text{SINR})$ objective is also treated in [38] using a trick whereby the total instantaneous transmission power at each node is kept constant, but a way to optimally pick that constant is needed. A heuristic algorithm for the physical layer power allocation, where terminals transmit either at full power or not at all, is pursued in [39]. The strategy of on-off power control—where terminals transmit at their spectral masks or not at all—is shown to be optimal under a low-SINR

assumption in [40], but efficient algorithms to find the optimal power allocations are still lacking.

Alternatively, simplifications and insights can be obtained from the Lagrangian dual problem. Interestingly, it is possible to show that layers emerge naturally from the decomposition of Lagrangians associated with optimal networking formulations [6]. However, these layered architectures might in principle be suboptimal for wireless networks as the lack of convexity entails a positive duality gap. A recent development has shown that, regardless of assumptions on the physical layer, the duality gaps of non-convex wireless network optimization problems are null in the presence of fading, and optimality of separating wireless network design into layers follows easily from the latter result [35]. A consequence of the optimality of layered architectures is that algorithms for optimal resource allocation at the physical layer can be leveraged with relative simplicity to design optimal networks. And in that sense, it upholds the importance of solving physical layer resource allocation problems—a task not addressed in [35].

The present thesis develops a dual subgradient algorithm for cross-layer network optimization. It is shown that despite the lack of convexity of the optimization problem, it is possible to obtain near-optimal primal variables. These near-optimal primal variables are obtained by forming weighted running averages of the primal sequence obtained as a byproduct of the subgradient algorithm. This offers a valuable generalization of ergodic convergence results for convex optimization problems; see e.g., [21, 22] and references therein. The proposed optimization methodology is then used to obtain a simple algorithm for network control. The resultant algorithm comprises two phases: (a) an offline phase, where near-optimal primal variables and Lagrange multipliers are obtained; and (b) an online phase, where power is allocated based on the current channel realization, and routing is performed.

Among other simpler subproblems, each step of the subgradient algorithm necessitates the solution of a sum-rate power allocation problem. This problem has exponential complexity [41] motivating the use of approximate solutions, which constitute another contribution of this thesis. Suitable approximations render the high-SINR scenario tractable, but the low-SINR scenario remains intractable. Successive convex approximations are also developed, using the condensation method, see e.g., [37, Sec. 2.2], and other sophisticated approaches adapted from the DSL literature [42, 43]. Note that

while in e.g., [37,38], approximations are used to modify the primal problem, here they are used to approximately solve subgradient iteration steps.

Chapter 2

Cooperative Demand Response Scheduling

The chapter begins with the development of an asynchronous subgradient algorithm for separable convex programs in Section 2.1. Section 2.2 presents a cooperative demand response (load control) problem for residential end users, and Section 2.3 applies the algorithm of Section 2.1 to the aforementioned problem in a distributed fashion. Numerical tests are presented in Section 2.4, and a brief sensitivity analysis of the load control problem is given in Section 2.5.

The material in this chapter draws from [44, 45].

2.1 Asynchronous subgradient algorithm

Consider functions $f_i : \mathcal{X}_i \to \mathbb{R}$ and $\mathbf{g}_i : \mathcal{X}_i \to \mathbb{R}^K$, $i = 1, \ldots, N$, and the following prototype optimization problem with variables $\mathbf{x} := (\mathbf{x}_1, \ldots, \mathbf{x}_N)$

$$f^* = \min_{\mathbf{x}} \quad \sum_{i=1}^{N} f_i(\mathbf{x}_i) \tag{2.1a}$$

$$\text{subj. to} \quad \sum_{i=1}^{N} \mathbf{g}_i(\mathbf{x}_i) \leq \mathbf{0} \tag{2.1b}$$

$$\mathbf{x}_i \in \mathcal{X}_i,\ i = 1, \ldots, N. \tag{2.1c}$$

Problem (2.1) is separable, that is, the objective and constraints are sums of functions that each depend on different groups of variables; see also [46, Sec. 5.1.6]. Following a standard duality approach, let $\boldsymbol{\lambda}$ denote a K-dimensional Lagrange multiplier vector corresponding to constraint (2.1b). The Lagrangian function, the dual function, and the dual problem corresponding to (2.1) are respectively

$$L(\mathbf{x}, \boldsymbol{\lambda}) = \sum_{i=1}^{N} \left[f_i(\mathbf{x}_i) + \boldsymbol{\lambda}^T \mathbf{g}_i(\mathbf{x}_i) \right] \tag{2.2}$$

$$\rho(\boldsymbol{\lambda}) := \min_{\{\mathbf{x}_i \in \mathcal{X}_i\}_{i=1}^N} L(\mathbf{x}, \boldsymbol{\lambda}) = \sum_{i=1}^{N} \rho_i(\boldsymbol{\lambda}) \tag{2.3}$$

$$\rho^* = \max_{\boldsymbol{\lambda} \geq \mathbf{0}} \rho(\boldsymbol{\lambda}) \tag{2.4}$$

where

$$\rho_i(\boldsymbol{\lambda}) = \min_{\mathbf{x}_i \in \mathcal{X}_i} \{ f_i(\mathbf{x}_i) + \boldsymbol{\lambda}^T \mathbf{g}_i(\mathbf{x}_i) \}, \quad i = 1, \ldots, N. \tag{2.5}$$

The subgradient method aims at solving the dual problem; see e.g. [47, Ch. 8]. It amounts to the iteration

$$\boldsymbol{\lambda}(\ell + 1) = \left[\boldsymbol{\lambda}(\ell) + \beta_\ell \sum_{i=1}^{N} \mathbf{g}_i \left[\mathbf{x}_i(\ell) \right] \right]^+ \tag{2.6}$$

where $\ell = 1, 2, \ldots$ is the iteration index, $[.]^+$ denotes projection to the nonnegative orthant, β_ℓ is a constant stepsize, and

$$\mathbf{x}_i(\ell) \in \arg\min_{\mathbf{x}_i \in \mathcal{X}_i} \{ f_i(\mathbf{x}_i) + \boldsymbol{\lambda}^T(\ell) \mathbf{g}_i(\mathbf{x}_i) \}, \quad i = 1, \ldots, N. \tag{2.7}$$

By definition, the vector $\mathbf{g}_i \left[\mathbf{x}_i(\ell) \right]$ is a subgradient of the function $\rho_i(\boldsymbol{\lambda})$ at $\boldsymbol{\lambda} = \boldsymbol{\lambda}(\ell)$. The iteration is initialized with arbitrary $\boldsymbol{\lambda}(1) \geq \mathbf{0}$.

The following three stepsize rules are considered, each having desirable properties that will be delineated shortly:

(S1) Constant stepsize: $\beta_\ell = \beta > 0$, $\ell = 1, 2, \ldots$

(S2) Nonsummable but square-summable stepsize: There exist sequences $\{\alpha_\ell\}$ and $\{\gamma_\ell\}$ such that (a) $\alpha_\ell \leq \beta_\ell \leq \gamma_\ell$, $\ell = 1, 2, \ldots$; (b) $\alpha_\ell > 0$, $\alpha_\ell \to 0$ as $\ell \to \infty$ and $\sum_{\ell=1}^{\infty} \alpha_\ell = \infty$; and (c) $\gamma_\ell \downarrow 0$ and $\sum_{\ell=1}^{\infty} \gamma_\ell^2 < \infty$.

(S3) Stepsize given by harmonic series: This is a special case of (S2) with $\beta_\ell = a/(\ell + c)$, $\ell = 1, 2, \ldots$, and $a > 0$, $c > 0$ are constants.

The iteration (2.6) is synchronous, because the subgradient components $\mathbf{g}_i\left[\mathbf{x}_i(\ell)\right]$ used to move from $\boldsymbol{\lambda}(\ell)$ to $\boldsymbol{\lambda}(\ell+1)$ depend on the multiplier iterate at ℓ [cf. (2.7)]. The asynchronous subgradient method does not necessarily use the minimizers $\mathbf{x}_i(\ell)$ of the Lagrangian function corresponding to the multiplier at iteration ℓ [cf. (2.7)]. Instead, it is allowed to use subgradient components which correspond to outdated Lagrange multipliers. Specifically, let $1 \leq \tau_i(\ell) \leq \ell$ denote the index of the Lagrange multiplier used in the ith minimization. The asynchronous subgradient method takes the form

$$\boldsymbol{\lambda}(\ell+1) = \left[\boldsymbol{\lambda}(\ell) + \beta \sum_{i=1}^{N} \mathbf{g}_i\left[\mathbf{x}_i\left(\tau_i(\ell)\right)\right]\right]^+ \tag{2.8}$$

where $\mathbf{g}_i\left[\mathbf{x}_i\left(\tau_i(\ell)\right)\right]$ is a subgradient of $\rho_i(\boldsymbol{\lambda})$ at $\boldsymbol{\lambda} = \boldsymbol{\lambda}(\tau_i(\ell))$:

$$\mathbf{x}_i(\tau_i(\ell)) \in \arg\min_{\mathbf{x}_i \in \mathcal{X}_i}\{f_i(\mathbf{x}_i) + \boldsymbol{\lambda}^T(\tau_i(\ell))\mathbf{g}_i(\mathbf{x}_i)\}, \quad i = 1, \ldots, N. \tag{2.9}$$

If $\tau_i(\ell) = \ell$ for all ℓ and i, then the algorithm becomes the standard (synchronous) subgradient method [cf. (2.7)].

The sequence of primal iterates generated from the subgradient method does not necessarily converge. In order to recover (near-)optimal primal variables from the synchronous method, running averages of the primal iterates $\mathbf{x}_i(\ell)$ are formed [21, 22]. For the asynchronous method, the running averages are formed using the primal iterates $\mathbf{x}_i(\tau_i(\ell))$ used for the multiplier updates. These averages have constant weights

$$\bar{\mathbf{x}}_i(\ell) := \frac{1}{\ell} \sum_{l=1}^{\ell} \mathbf{x}_i(\tau_i(l)) \tag{2.10}$$

or weights proportional to the stepsize

$$\hat{\mathbf{x}}_i(\ell) := \frac{1}{\ell} \sum_{l=1}^{\ell} \beta_l \mathbf{x}_i(\tau_i(l)) \tag{2.11}$$

where $B_\ell := \sum_{l=1}^{\ell} \beta_l$. It is also useful to consider averages of the dual iterates in the same fashion as (2.10)

$$\bar{\boldsymbol{\lambda}}(\ell) := \frac{1}{\ell} \sum_{l=1}^{\ell} \boldsymbol{\lambda}(l) \tag{2.12}$$

It should be emphasized that the running averages can be efficiently computed in a recursive fashion, that is,

$$\bar{\mathbf{x}}_i(\ell) = \frac{1}{\ell}\mathbf{x}_i(\tau_i(\ell)) + \frac{\ell-1}{\ell}\bar{\mathbf{x}}(\ell-1) \tag{2.13}$$

or

$$\hat{\mathbf{x}}_i(\ell) = \frac{\beta_\ell}{B_\ell}\mathbf{x}_i(\tau_i(\ell)) + \frac{B_{\ell-1}}{B_\ell}\hat{\mathbf{x}}_i(\ell-1) \tag{2.14}$$

The following two conditions are adopted for the generic problem (2.1) and iterations (2.8).

C2.1. Sets $\mathcal{X}_i$, $i = 1, \ldots, N$ are convex, closed, and bounded. Functions f_i and $\mathbf{g}_i$, $i = 1, \ldots, N$ are convex (componentwise) and continuous. Moreover, there are $\mathbf{x}_i \in \mathcal{X}_i$, $i = 1, \ldots, N$, so that (2.1b) holds as strict inequality (Slater constraint qualification).

C2.2. There exist integers D_i, $i = 1, \ldots, N$, so that $\ell - \tau_i(\ell) \leq D_i$, $\ell = 1, 2, \ldots$

The first condition ensures that problem (2.1) is convex. Moreover, it asserts boundedness of the subgradients. More precisely, with $\|.\|$ denoting the Euclidean norm, there are finite bounds $G, G_1, \ldots, G_N$ so that for all $\mathbf{x}_i \in \mathcal{X}_i$, $i = 1, \ldots, N$,

$$\left\| \sum_{i=1}^{N} \mathbf{g}_i(\mathbf{x}_i) \right\| \leq G; \quad \|\mathbf{g}_i(\mathbf{x}_i)\| \leq G_i. \tag{2.15}$$

The second condition states that the delay $\ell - \tau_i(\ell)$ is upper bounded by a finite number. Note that the condition allows for lost messages to happen infinitely often. In the ensuing propositions, setting $D_i = 0$, $i = 1, \ldots, N$, yields the corresponding results for the synchronous algorithm.

Remark1. The quantities $\tau_i(\ell)$, G, G_i, and D_i are useful in order to analyze the convergence of the algorithm (2.8), as will be seen next. Nevertheless, these quantities are not needed in order to run the iteration (2.8). In particular, the latest primal maximizers available $\mathbf{x}_i(\tau_i(\ell))$ are used in the iteration (2.8) and the averages (2.10), (2.11).

The primal problem (2.1) has an optimal solution, as a consequence of Weierstrass's theorem, because the feasible set is nonempty and compact, and the objective is continuous (cf. C2.1). Moreover, the dual function $\rho(\boldsymbol{\lambda})$ is concave. Condition C1 implies

that it is also finite everywhere; therefore, it is also proper, and continuous everywhere [47, Props. 6.2.1, 1.2.2, 1.4.6]. Moreover, convexity and Slater's constraint qualification in C2.1 imply that there is no duality gap, i.e., $f^* = \rho^*$, and that the set of optimal solutions of the dual problem (2.4) is nonempty and compact [47, Prop. 6.4.3]. The set of optimal dual solutions is exactly what is more commonly known as optimal Lagrange multipliers; see e.g., [47, Props. 6.1.2 and 6.2.3].

The ensuing two subsections provide the convergence of the dual sequences $\{\boldsymbol{\lambda}(\ell)\}$, $\{\bar{\boldsymbol{\lambda}}(\ell)\}$, and the primal sequences $\{\bar{\mathbf{x}}_i(\ell)\}$, $\{\hat{\mathbf{x}}_i(\ell)\}$, respectively, under stepsize rules (S1)–(S3). In the following results, $d(\boldsymbol{x}, \mathcal{X}) := \inf_{\boldsymbol{y}\in\mathcal{X}} \|\boldsymbol{x} - \boldsymbol{y}\|$ denotes the distance of a point $\boldsymbol{x}$ from a closed convex set $\mathcal{X}$.

2.1.1 Dual Convergence

The following lemma asserts that the update direction used in the multiplier updates (2.8) is in fact an ε-subgradient; see e.g., [47, Sec. 4.3] for pertinent definitions.

Lemma 2.1. *Under C2.1 and C2.2, it holds for all $\boldsymbol{\xi} \geq \mathbf{0}$ and $\ell = 1, 2, \ldots$ that*

$$\left[\sum_{i=1}^{N} \mathbf{g}_i[\mathbf{x}_i(\tau_i(\ell))]\right]^T (\boldsymbol{\xi} - \boldsymbol{\lambda}(\ell)) \geq \rho(\boldsymbol{\xi}) - \rho(\boldsymbol{\lambda}(\ell)) - \sum_{i=1}^{N} \varepsilon_i(\ell) \tag{2.16}$$

where in the case of constant stepsize (S1),

$$\varepsilon_i(\ell) = 2\beta G G_i D_i \tag{2.17}$$

or in the case of a diminishing stepsize (S2) or (S3),

$$\varepsilon_i(\ell) = 2G G_i \sum_{j=1}^{D_i} \gamma_{\ell-j} \tag{2.18}$$

and $\varepsilon_i(\ell) \to 0$ as $\ell \to \infty$.

The left-hand side of (2.16) contains the update direction used in (2.8), and is exactly the definition of an ε-subgradient of the dual function at $\boldsymbol{\lambda}(\ell)$. It asserts that ε is constant if a constant stepsize is used, and vanishing if a vanishing stepsize is employed. A result similar to the one in Lemma 2.1 has been developed for an asynchronous subgradient algorithm and only for diminishing stepsizes in [24]. The present

result also covers the case of constant stepsize, and explicitly gives the value of ε as a function of the maximum delays. It is also worth mentioning that a counterpart of the result in Lemma 2.1 has been established for (synchronous) incremental subgradient methods [48]. Nevertheless, these methods are different than the asynchronous method studied here.

Proof of Lemma 2.1. It will be shown that

$$\mathbf{g}_i^T[\mathbf{x}_i(\tau_i(\ell))](\boldsymbol{\xi} - \boldsymbol{\lambda}(\ell)) \geq \rho_i(\boldsymbol{\xi}) - \rho_i(\boldsymbol{\lambda}(\ell)) - \varepsilon_i(\ell) \tag{2.19}$$

from which (2.16) follows readily.

The left-hand side of (2.19) takes the following form, after applying the definition of the subgradient at $\rho_i[\boldsymbol{\lambda}(\tau_i(\ell))]$:

$$\begin{aligned}\mathbf{g}_i^T[\mathbf{x}_i(\tau_i(\ell))](\boldsymbol{\xi} - \boldsymbol{\lambda}(\ell)) &= \mathbf{g}_i^T[\mathbf{x}_i(\tau_i(\ell))](\boldsymbol{\xi} - \boldsymbol{\lambda}(\tau_i(\ell))) + \mathbf{g}_i^T[\mathbf{x}_i(\tau_m(\ell))](\boldsymbol{\lambda}(\tau_i(\ell)) - \boldsymbol{\lambda}(\ell)) \\ &\geq \rho_i(\boldsymbol{\xi}) - \rho_i(\boldsymbol{\lambda}(\tau_i(\ell))) + \mathbf{g}_i^T[\mathbf{x}_m(\tau_i(\ell))](\boldsymbol{\lambda}(\tau_i(\ell)) - \boldsymbol{\lambda}(\ell)) \quad (2.20)\end{aligned}$$

Adding and subtracting the same terms in the right-hand side of (2.20), and applying the definition of the subgradient at $\boldsymbol{\lambda}(\tau_i(\ell) + j)$, it follows that

$$\begin{aligned}\mathbf{g}_i^T[\mathbf{x}_i(\tau_i(\ell))](\boldsymbol{\xi} - \boldsymbol{\lambda}(\ell)) &\geq \rho_i(\boldsymbol{\xi}) - \rho_i(\boldsymbol{\lambda}(\ell)) \\ &\quad + \sum_{j=1}^{\ell-\tau_i(\ell)} (\rho_i[\boldsymbol{\lambda}(\tau_i(\ell) + j)] - \rho_i[\boldsymbol{\lambda}(\tau_i(\ell) + j - 1)]) \\ &\quad + \sum_{j=1}^{\ell-\tau_i(\ell)} \mathbf{g}_i^T[\mathbf{x}_i(\tau_i(\ell))](\boldsymbol{\lambda}(\tau_i(\ell) + j - 1) - \boldsymbol{\lambda}(\tau_i(\ell) + j)) \quad (2.21) \\ &\geq \rho_i(\boldsymbol{\xi}) - \rho_i(\boldsymbol{\lambda}(\ell)) \\ &\quad - \sum_{j=1}^{\ell-\tau_i(\ell)} \mathbf{g}_i^T[\mathbf{x}(\tau_i(\ell) + j)](\boldsymbol{\lambda}(\tau_i(\ell) + j - 1) - \boldsymbol{\lambda}(\tau_i(\ell) + j)) \\ &\quad + \sum_{j=1}^{\ell-\tau_m(\ell)} \mathbf{g}_i^T[\mathbf{x}_i(\tau_i(\ell))](\boldsymbol{\lambda}(\tau_i(\ell) + j - 1) - \boldsymbol{\lambda}(\tau_i(\ell) + j)). \quad (2.22)\end{aligned}$$

Applying the Cauchy-Schwartz inequality to the right-hand side of (2.22) leads to

$$\mathbf{g}_i^T[\mathbf{q}_i(\tau_i(\ell))](\boldsymbol{\xi} - \boldsymbol{\lambda}(\ell)) \geq \rho_i(\boldsymbol{\xi}) - \rho_i(\boldsymbol{\lambda}(\ell))$$

$$- \sum_{j=1}^{\ell-\tau_i(\ell)} \left(\|\mathbf{g}_i[\mathbf{x}(\tau_i(\ell)+j)]\| + \|\mathbf{g}_i[\mathbf{x}_i(\tau_i(\ell))]\|\right) \cdot \|\boldsymbol{\lambda}(\tau_i(\ell)+j-1) - \boldsymbol{\lambda}(\tau_i(\ell)+j)\|. \quad (2.23)$$

From the subgradient iteration (2.8) at $\boldsymbol{\lambda}(\tau_m(\ell)+j-1)$ and the nonexpansive property of the projection [47, Prop. 2.2.1], it follows easily that

$$\|\boldsymbol{\lambda}(\tau_i(\ell)+j-1) - \boldsymbol{\lambda}(\tau_i(\ell)+j)\| \leq \beta_{\tau_i(\ell)+j-1} G. \quad (2.24)$$

Combining (2.23) with (2.24) and the bounds in (2.15), it is deduced that

$$\mathbf{g}_i^T[\mathbf{x}_i(\tau_i(\ell))](\boldsymbol{\xi} - \boldsymbol{\lambda}(\ell)) \geq \rho_i(\boldsymbol{\xi}) - \rho_i(\boldsymbol{\lambda}(\ell)) - 2GG_i \sum_{j=1}^{\ell-\tau_i(\ell)} \beta_{\tau_i(\ell)+j-1}. \quad (2.25)$$

Recalling also that $\ell - \tau_i(\ell) \leq D_i$ (cf. C2.2), it holds for a constant stepsize that

$$2GG_i \sum_{j=1}^{\ell-\tau_i(\ell)} \beta_{\tau_i(\ell)+j-1} \leq 2GG_i\beta D_i =: \varepsilon_i(\ell). \quad (2.26)$$

For diminishing stepsizes, the fact that the stepsize sequence is majorized by the sequence $\{\gamma_\ell\}$, and also the condition $\ell - \tau_i(\ell) \leq D_i$ imply that

$$\begin{aligned} 2GG_m \sum_{j=1}^{\ell-\tau_i(\ell)} \beta_{\tau_i(\ell)+j-1} &\leq 2GG_i \sum_{j=1}^{\ell-\tau_m(\ell)} \gamma_{\tau_i(\ell)+j-1} \\ &\leq 2GG_i \sum_{j=1}^{D_i} \gamma_{\ell-j} =: \varepsilon_i(\ell) \end{aligned}$$

which vanishes because $\gamma_\ell \to 0$ as $\ell \to \infty$.

The next propositions deal with convergence of Lagrange multiplier iterates with constant and diminishing stepsizes, respectively.

Proposition 2.1. *Under C2.1, C2.2, and with stepsize (S1), the following hold:*

(a) *The sequence* $\{\bar{\boldsymbol{\lambda}}(\ell)\}$ *satisfies*

$$\liminf_{\ell\to\infty} \rho(\bar{\boldsymbol{\lambda}}(\ell)) \geq \rho^* - \frac{1}{2}\beta G^2 - 2\beta G \sum_{i=1}^{N} G_i D_i. \quad (2.27)$$

(b) *Define the set* $\tilde{\Lambda}$ *of near-optimal dual solutions as*

$$\tilde{\Lambda} := \left\{ \boldsymbol{\lambda} \geq \mathbf{0} \,\middle|\, \rho(\boldsymbol{\lambda}) \geq \rho^* - \frac{1}{2}\beta G^2 - 2\beta G \sum_{i=1}^{N} G_i D_i \right\}. \tag{2.28}$$

The sequence $\{\boldsymbol{\lambda}(\ell)\}$ *is bounded, and the sequence* $\{\bar{\boldsymbol{\lambda}}(\ell)\}$ *has limit points, all of which are in the set* $\tilde{\Lambda}$*. Moreover, it holds that*

$$\lim_{\ell \to \infty} d(\bar{\boldsymbol{\lambda}}(\ell), \tilde{\Lambda}) = 0. \tag{2.29}$$

The previous proposition states essentially that the dual averages $\{\bar{\boldsymbol{\lambda}}(\ell)\}$ converge to a point which corresponds to a near-optimal dual value. Proximity to the optimal value is quantified by $\frac{1}{2}\beta G^2 + 2\beta G \sum_{i=1}^{N} G_i D_i$, which decreases linearly with the stepsize, and also decreases as the delays D_i decrease. This "discrepancy" can be made as small as desired by choosing a sufficiently small stepsize β.

Proof of Proposition 2.1. The proof relies on application of [49, Th. 4.1]. The function f of [49] corresponds to $\rho(\boldsymbol{\lambda})$ in the present setup, and the set S to $\{\boldsymbol{\lambda} \geq \mathbf{0}\}$. The theorem can be applied because the dual function is a proper concave function, continuous everywhere, and moreover, the set of optimal Lagrange multipliers is nonempty and compact, as explained earlier. This falls into the coercive case.

The claims of Proposition 2.1 are deduced from [49, Th. 4.1] as follows:

(a) Apply parts (i) and (iv). The quantity δ corresponds to $\frac{1}{2}\beta G^2 + 2\beta G \sum_{i=1}^{N} D_i G_i$ in the present case, by application of Lemma 2.1.

(b) Part (ii) implies that the sequence $\{\boldsymbol{\lambda}(\ell)\}$ is bounded. The remaining claims follow readily from part (iv).

The next propositions deal with the convergence of the Lagrange multiplier iterates with diminishing stepsizes.

Proposition 2.2. *Under C2.1, C2.2, and with stepsize (S2) or (S3), the sequence* $\{\boldsymbol{\lambda}(\ell)\}$ *converges to an optimal dual solution.*

Proposition 2.2 asserts that the iterates $\{\boldsymbol{\lambda}(\ell)\}$ converge to an optimal Lagrange multiplier vector $\boldsymbol{\lambda}^*$. This result extends [24, Prop. 6.1], which only deals with convergence of $\rho(\boldsymbol{\lambda}(\ell))$ but not of the iterates $\boldsymbol{\lambda}(\ell)$ themselves.

Proof of Proposition 2.2. Convergence can be shown using [49, Th. 3.3] or [23, Th. 8]. The essential requirement in order to apply those results is to show that

$$\sum_{\ell=1}^{\infty} \beta_\ell \sum_{i=1}^{N} \varepsilon_i(\ell) < \infty. \tag{2.30}$$

The latter can be shown using (2.18) and the properties of the sequence $\{\gamma_\ell\}$ majorizing $\{\beta_\ell\}$. Specifically, it holds for $i = 1, \ldots, N$ (and an integer $M \geq D_i$ if needed) that

$$\sum_{\ell=M}^{\infty} \beta_\ell \varepsilon_i(\ell) \leq 2GG_i \sum_{\ell=M}^{\infty} \gamma_\ell \sum_{j=1}^{D_i} \gamma_{\ell-j} \tag{2.31}$$

$$\leq 2GG_i \sum_{\ell=M}^{\infty} \sum_{j=1}^{D_i} \gamma_{\ell-j}^2 \tag{2.32}$$

where (2.31) holds because $\beta_\ell \leq \gamma_\ell$, while (2.32) follows from the fact that $\{\gamma_\ell\}$ is monotonically decreasing. The series in (2.32) is finite, because $\{\gamma_\ell\}$ is square-summable. Therefore, (2.30) holds too.

2.1.2 Primal Convergence

The following lemma provides an upper bound on the objective function evaluated at the primal averages.

Lemma 2.2. *The following holds for the running averages* (2.10) *under C2.1 and C2.2:*

$$\sum_{i=1}^{N} f_i(\bar{\mathbf{x}}_i(\ell)) \leq \rho^* + \sum_{i=1}^{N} \frac{1}{\ell} \sum_{l=1}^{\ell} \varepsilon_i(l) - \frac{1}{\ell} \sum_{l=1}^{\ell} \boldsymbol{\lambda}^T(l) \left[\sum_{i=1}^{N} \mathbf{g}_i[\mathbf{x}_i(\tau_i(l))] \right] \tag{2.33}$$

where $\varepsilon_m(l)$ is given by (2.17) *if the constant stepsize (S1) is used, or by* (2.18) *if the diminishing stepsize (S3) is used.*

Similarly, the following holds for the running averages (2.11) *under C2.1 and C2.2 and for the stepsize (S2):*

$$\sum_{i=1}^{N} f_i(\hat{\mathbf{x}}_i(\ell)) \leq \rho^* + \sum_{i=1}^{N} \frac{1}{B_\ell} \sum_{l=1}^{\ell} \varepsilon_i(l) - \frac{1}{B_\ell} \sum_{l=1}^{\ell} \beta_l \boldsymbol{\lambda}^T(l) \left[\sum_{i=1}^{N} \mathbf{g}_i[\mathbf{x}_m(\tau_i(l))] \right] \tag{2.34}$$

where $\varepsilon_i(l)$ is given by (2.18).

Proof of Lemma 2.2. It holds due to the convexity of $\sum_{i=1}^{N} f_i$ that

$$\sum_{i=1}^{M} f_i(\bar{\mathbf{x}}_i(\ell)) \leq \frac{1}{\ell} \sum_{l=1}^{\ell} \sum_{i=1}^{N} f_i[\mathbf{x}_i(\tau_i(l))]. \tag{2.35}$$

Adding and subtracting identical terms to the right-hand side of (2.35) leads to

$$\begin{aligned} \sum_{i=1}^{N} f_i(\bar{\mathbf{x}}_i(l)) \leq \frac{1}{\ell} \sum_{l=1}^{\ell} \sum_{i=1}^{N} \left[f_i[\mathbf{x}_i(\tau_i(\ell)) + \boldsymbol{\lambda}^T(l) \mathbf{g}_i[\mathbf{x}_i(\tau_i(l))] \right] \\ - \frac{1}{\ell} \sum_{l=1}^{\ell} \boldsymbol{\lambda}^T(l) \sum_{i=1}^{N} \mathbf{g}_i[\mathbf{x}_i(\tau_i(l))]. \end{aligned} \tag{2.36}$$

Consider the m-th summand in the second term on the right-hand side of (2.36). Using the definition of the function $\rho_m(\boldsymbol{\lambda})$ and (2.19), it holds that

$$\begin{aligned} f_i[\mathbf{x}_i(\tau_i(\ell))] + \boldsymbol{\lambda}^T(l) \mathbf{g}_i[\mathbf{x}_i(\tau_i(l))] &= f_i[\mathbf{x}_i(\tau_i(\ell))] + \boldsymbol{\lambda}^T(\tau_i(l)) \mathbf{g}_i[\mathbf{x}_i(\tau_i(l))] \\ &\quad - \mathbf{g}_i^T[\mathbf{x}_i(\tau_i(l))][\boldsymbol{\lambda}(\tau_i(l)) - \boldsymbol{\lambda}(l)] \\ &\leq \rho_i[\boldsymbol{\lambda}(\tau_i(l))] + \rho_i(\boldsymbol{\lambda}(l)) - \rho_i[\boldsymbol{\lambda}(\tau_i(l))] + \varepsilon_i(l) \end{aligned} \tag{2.37}$$

where $\varepsilon_m(l)$ is given by (2.17) or (2.18).

Introducing (2.37) into (2.36), and using the fact that $\rho(\boldsymbol{\lambda}(l)) \leq \rho^*$, establishes (2.33). The proof of (2.34) is analogous. □

The preceding proof adapts methods used in the proof of [22, Prop. 1(b)]. The additional difficulty here lies on the fact that the subgradients used in the updates correspond to outdated Lagrange multipliers. To overcome this issue, the preceding proof is constructed in order to leverage Lemma 2.1.

The following lemma provides a lower bound on the objective function evaluated at the primal averages.

Lemma 2.3. *The following holds under C2.1 and C2.2, and any stepsize rule:*

$$\sum_{i=1}^{N} f_m(\bar{\mathbf{x}}_i(\ell)) \geq \rho^* - \|\boldsymbol{\lambda}^*\| \left\| \left[\sum_{i=1}^{N} \mathbf{g}_i(\bar{\mathbf{x}}_i(\ell)) \right]^+ \right\| \tag{2.38}$$

where $\boldsymbol{\lambda}^*$ is any dual optimal solution.

Inequality (2.38) *holds also if $\bar{\mathbf{x}}(\ell)$ is replaced by $\hat{\mathbf{x}}(\ell)$.*

This result follows readily from [22, Prop. 1(c)].

The following proposition characterizes the convergence properties of the primal averages under stepsize rule (S1).

Proposition 2.3. *Under C2.1, C2.2, and with stepsize (S1), the following hold:*

(a) *The primal averages* $\{\bar{\mathbf{x}}(\ell)\}$ *in* (2.10) *satisfy*

$$\text{(i)} \quad \lim_{\ell\to\infty} \left\| \left[\sum_{m=1}^{M} \mathbf{g}_i(\bar{\mathbf{x}}_i(\ell)) \right]^+ \right\| = 0 \tag{2.39}$$

$$\text{(ii)} \quad \limsup_{\ell\to\infty} \left(\sum_{i=1}^{N} f_i(\bar{\mathbf{x}}_i(\ell)) \right) \leq f^* + \frac{1}{2}\beta G^2 + 2\beta G \sum_{i=1}^{N} G_i D_i \tag{2.40}$$

$$\text{(iii)} \quad \liminf_{\ell\to\infty} \left(\sum_{i=1}^{N} f_i(\bar{\mathbf{x}}_i(\ell)) \right) \geq f^*. \tag{2.41}$$

(b) *Define the set* $\tilde{\mathcal{R}}$ *of near-optimal solutions of* (2.1) *as*

$$\tilde{\mathcal{R}} = \left\{ \{\mathbf{x}_i \in \mathcal{X}_i\}_{i=1}^N \,\middle|\, \sum_{i=1}^{N} \mathbf{g}_i(\mathbf{x}_i) \leq \mathbf{0} \ \sum_{i=1}^{N} f_i(\mathbf{x}_i) \leq f^* + \frac{1}{2}\beta G^2 + 2\beta G \sum_{i=1}^{N} G_i D_i \right\}. \tag{2.42}$$

The sequence $\{\bar{\mathbf{x}}_1(\ell), \ldots, \bar{\mathbf{x}}_N(\ell)\}$ *has limit points, all of which are in the set* $\tilde{\mathcal{R}}$. *Moreover, it holds that*

$$\lim_{\ell\to\infty} d((\bar{\mathbf{x}}(\ell), \bar{\mathbf{x}}_1(\ell), \ldots, \bar{\mathbf{x}}_N(\ell)), \tilde{\mathcal{R}}) = 0. \tag{2.43}$$

The last proposition establishes that the primal averages become asymptotically feasible and asymptotically near-optimal. Specifically, the term inside the norm in part (a)-(i) is the violation of the constraint $\sum_{i=1}^{N} \mathbf{x}_i(\bar{\mathbf{x}}_i(\ell)) \leq \mathbf{0}$, and this converges to zero. Moreover, near-optimality is measured by the quantity $\frac{1}{2}\beta G^2 + 2\beta G \sum_{i=1}^{N} G_i D_i$, which decreases with the stepsize β and the delays D_m. Note that the same quantity appeared in Proposition 2.1, and hence, there is a symmetry between the primal and dual convergence results with constant stepsize. Proposition 2.3 generalizes results of [22, Sec. 4.2], which deals with a synchronous subgradient algorithm.

Proof of Proposition 2.3. (a)-(i) Convexity of the constraint function implies that

$$\sum_{i=1}^{N} \mathbf{g}_i(\bar{\mathbf{x}}_i(\ell)) \leq \frac{1}{\ell} \sum_{l=1}^{\ell} \left[\mathbf{g}_0(\mathbf{s}(l)) + \sum_{i=1}^{N} \mathbf{g}_i[\mathbf{x}_i(\tau_i(l))] \right]. \tag{2.44}$$

It is deduced from the Lagrange multiplier updates (2.8) that for all $l = 1, 2, \ldots$,

$$\mathbf{g}_0(\mathbf{s}(l)) + \sum_{m=1}^{M} \mathbf{g}_m[\mathbf{q}_m(\tau_m(l))] \leq \frac{\boldsymbol{\lambda}(l+1) - \boldsymbol{\lambda}(l)}{\beta}. \tag{2.45}$$

It follows from (2.44) after substituting (2.45) that

$$\sum_{i=1}^{M} \mathbf{g}_i(\bar{\mathbf{x}}_i(\ell)) \leq \frac{\boldsymbol{\lambda}(\ell+1)}{\beta\ell} \tag{2.46}$$

from which (2.39) follows because the sequence $\{\boldsymbol{\lambda}(\ell)\}$ is bounded.

(a)-(ii) The Lagrange multiplier updates (2.8) using the nonexpansive property of the projection imply that

$$\begin{aligned} \|\boldsymbol{\lambda}(l+1)\|^2 &\leq \left\| \boldsymbol{\lambda}(l) + \beta \left[\sum_{i=1}^{N} \mathbf{g}_i[\mathbf{x}_i(\tau_i(l))] \right] \right\|^2 \\ &\leq \|\boldsymbol{\lambda}(l)\|^2 + \beta^2 G^2 + 2\beta \boldsymbol{\lambda}^T(l) \left[\sum_{i=1}^{N} \mathbf{g}_i[\mathbf{x}_i(\tau_i(l))] \right] \end{aligned} \tag{2.47}$$

and therefore,

$$-\boldsymbol{\lambda}^T(l) \left[\sum_{i=1}^{N} \mathbf{g}_i[\mathbf{x}_i(\tau_i(l))] \right] \leq \frac{\|\boldsymbol{\lambda}(l)\|^2 - \|\boldsymbol{\lambda}(l+1)\|^2}{2\beta} + \frac{\beta G^2}{2}. \tag{2.48}$$

Upon substituting $\rho^* = f^*$ (zero duality gap), (2.17), and (2.48) into (2.33), and taking lim sup on both sides of the resulting inequality, (2.40) follows.

(a)-(iii) The result follows by taking lim inf in (2.38), using part (a)-(i), and the fact that the dual optimal set is bounded.

(b) Since the sets $\mathcal{X}_i$, $i = 1, \ldots, N$, are compact, the sequence $\{\bar{\mathbf{x}}_1(\ell), \ldots, \bar{\mathbf{x}}_N(\ell)\}$ has limit points in $\mathcal{X}_1 \times \ldots \times \mathcal{X}_N$. Let $\{\tilde{\mathbf{x}}_1, \ldots, \tilde{\mathbf{x}}_N\}$ be one such limit point. Due to the continuity of the projection $[.]^+$ and the norm $\|.\|$, part (a)-(i) implies that

$$\left\| \left[\sum_{i=1}^{N} \mathbf{g}_i(\tilde{\mathbf{x}}_i) \right]^+ \right\| \leq 0 \tag{2.49}$$

or equivalently, $\{\tilde{\mathbf{x}}_1, \ldots, \tilde{\mathbf{x}}_N\}$ satisfies (2.1b).

Moreover, $\{\tilde{\mathbf{s}}, \tilde{\mathbf{q}}_1, \ldots, \tilde{\mathbf{q}}_M\}$ satisfies

$$\sum_{i=1}^{N} f_i(\tilde{\mathbf{x}}_i) \leq f^* + \frac{\beta G^2}{2} + 2\beta G \sum_{i=1}^{M} G_i D_i. \tag{2.50}$$

The latter holds due to part (a)-(ii) and the fact that lim sup is the supremum among all subsequential limits of a sequence; see e.g., [50, Def. 3.16]. It also follows in a similar way from part (a)-(iii) that the left-hand side of (2.50) is lower-bounded by f^*. It is therefore concluded that $\{\tilde{\mathbf{x}}_1, \ldots, \tilde{\mathbf{x}}_N\} \in \tilde{\mathcal{R}}$.

In order to show (2.43), note that there is a subsequence of $\{\bar{\mathbf{x}}_1(\ell), \ldots, \bar{\mathbf{x}}_N(\ell)\}$ indexed by $\mathcal{L} \subset \mathbb{N}$ so that (see e.g., [50, Th. 3.17(a)])

$$\limsup_{\ell\to\infty} d_{\tilde{\mathcal{R}}}(\bar{\mathbf{x}}_1(\ell), \ldots, \bar{\mathbf{x}}_N(\ell)) = \lim_{\ell\to\infty, \ell\in\mathcal{L}} d_{\tilde{\mathcal{R}}}(\bar{\mathbf{x}}_1(\ell), \ldots, \bar{\mathbf{x}}_N(\ell)). \tag{2.51}$$

The right-hand side of (2.51) is zero (restricting to a further subsequence if necessary) due to the continuity of the distance function and the fact that the limit of every convergent subsequence is in $\tilde{\mathcal{R}}$. Eq. (2.43) follows readily. □

The following proposition characterizes the convergence of the primal averages with weights proportional to the stepsize under the diminishing stepsize rule.

Proposition 2.4. *Under C2.1, C2.2, and with stepsize (S2), the following hold:*

(a) *The primal averages* $\{\hat{\mathbf{x}}_i(\ell)\}$ *in* (2.11) *satisfy*

$$\text{(i)} \quad \lim_{\ell\to\infty} \left\| \left[\sum_{i=1}^{N} \mathbf{g}_i(\hat{\mathbf{x}}_i(\ell)) \right]^+ \right\| = 0 \tag{2.52}$$

$$\text{(ii)} \quad \lim_{\ell\to\infty} \left(\sum_{i=1}^{N} f_m(\hat{\mathbf{x}}_i(\ell)) \right) = f^*. \tag{2.53}$$

(b) *The sequence* $\{\hat{\mathbf{x}}_1(\ell), \ldots, \hat{\mathbf{x}}_N(\ell)\}$ *has limit points. All such limit points are in the set of optimal solutions of* (2.1)*, which is denoted by* $\mathcal{R}^*$*. Moreover, it holds that*

$$\lim_{\ell\to\infty} d(\hat{\mathbf{x}}_1(\ell), \ldots, \hat{\mathbf{x}}_N(\ell), \mathcal{R}^*) = 0. \tag{2.54}$$

The previous proposition reveals that the primal averages with weights proportional to the stepsize will be asymptotically feasible and asymptotically optimal. It extends [21, Th. 1], which deals with a synchronous case. The proof of Proposition 2.4 is completely analogous to the proof of Proposition 2.3, and it is presented next noting only the points of difference.

Proof of Proposition 2.4. (a)-(i) Similar to (2.46), it can be shown that

$$\sum_{i=1}^{N} \mathbf{g}_i(\hat{\mathbf{x}}_i(\ell)) \leq \frac{\boldsymbol{\lambda}(\ell+1)}{B_\ell}. \tag{2.55}$$

(a)-(ii) Following the steps in the proof of Proposition 2.3(a)-(ii), it can be shown that

$$\sum_{i=1}^{N} f_m(\hat{\mathbf{x}}_i(\ell)) \leq f^* + \frac{G^2}{2B_\ell} \sum_{l=1}^{\ell} \beta_l^2 + \sum_{i=1}^{N} \frac{1}{B_\ell} \sum_{l=1}^{\ell} \beta_l \epsilon_i(l) - \frac{\|\boldsymbol{\lambda}(1)\|^2}{B_\ell}. \tag{2.56}$$

It holds using Toeplitz's lemma—see e.g., [49, Lemma 2.2]—that

$$\lim_{\ell\to\infty} \frac{1}{B_\ell} \sum_{l=1}^{\ell} \beta_l^2 = 0; \quad \lim_{\ell\to\infty} \frac{1}{B_\ell} \sum_{l=1}^{\ell} \beta_l \epsilon_i(l) = 0. \tag{2.57}$$

where it was used that $\beta_l \leq \gamma_l \to 0$ and $\epsilon_i(l) \to 0$ (cf. Lemma 2.1), as $l \to \infty$. Taking lim sup on both sides of (2.56) and using (2.57), it is deduced that

$$\limsup_{\ell\to\infty} \left(\sum_{i=1}^{N} f_i(\hat{\mathbf{x}}_i(\ell)) \right) \leq f^*. \tag{2.58}$$

A bound as in part (a)-(iii) of Proposition 2.3 holds also for $\hat{\mathbf{x}}_i(\ell)$ by taking the lim inf in (2.38). Upon combining the latter two results, (2.53) follows readily.

(b) This part is analogous to the one in Proposition 2.3. □

It should be noted that Proposition 2.4 could also follow directly from [23, Th. 20]. The proof presented here is an alternative approach, based on the methods of Proposition 2.3.

The following proposition characterizes the convergence of the primal averages with constant weights under the stepsize rule (S3).

Proposition 2.5. *Under C2.1, C2.2, and with stepsize (S3), the following hold.*

(a) *The primal averages* $\{\bar{\mathbf{x}}_i(\ell)\}$ *satisfy*

$$\text{(i)} \ \lim_{\ell\to\infty} \left\| \left[\sum_{i=1}^{N} \mathbf{g}_i(\bar{\mathbf{g}}_i(\ell)) \right]^+ \right\| = 0 \tag{2.59}$$

$$\text{(ii)} \ \lim_{\ell\to\infty} \left(\sum_{i=1}^{N} f_i(\bar{\mathbf{x}}_i(\ell)) \right) = f^*. \tag{2.60}$$

(b) *The sequence* $\{\bar{\mathbf{x}}_1(\ell),\ldots,\bar{\mathbf{x}}_N(\ell)\}$ *has limit points, all of which are in the set* $\mathcal{R}^*$ *(defined in Proposition 2.4). Moreover, it holds that*

$$\lim_{\ell\to\infty} d((\bar{\mathbf{x}}_1(\ell),\ldots,\bar{\mathbf{x}}_N(\ell)),\mathcal{R}^*)=0. \tag{2.61}$$

This proposition asserts that the primal averages with constant weights are asymptotically feasible and asymptotically optimal. A corresponding result for the synchronous case is given by [21, Th. 2], which actually adopts a more general stepsize rule than (S3).

Proof of Proposition 2.5. (a)-(i) As with (2.45), it is deduced from the Lagrange multiplier updates that for all $l=1,2,\ldots,$

$$\sum_{i=1}^{N}\mathbf{g}_i[\mathbf{x}_i\left(\tau_i(l)\right)]\leq\frac{\boldsymbol{\lambda}(l+1)-\boldsymbol{\lambda}(l)}{\beta_l}. \tag{2.62}$$

Using (2.62) into (2.44), it follows that

$$\sum_{i=1}^{N}\mathbf{g}_i(\bar{\mathbf{x}}_i\left(\ell\right))\leq\frac{1}{\ell}\sum_{j=1}^{\ell}\frac{\boldsymbol{\lambda}(j+1)-\boldsymbol{\lambda}(j)}{\beta_j}. \tag{2.63}$$

The ensuing claim will be proved next, from which (2.59) follows due to the continuity of the projection and norm functions.

Claim: It holds that

$$\lim_{\ell\to\infty}\frac{1}{\ell}\sum_{j=1}^{\ell}\frac{\boldsymbol{\lambda}(j+1)-\boldsymbol{\lambda}(j)}{\beta_j}=\mathbf{0}. \tag{2.64}$$

To prove the claim, recall that the sequence $\{\boldsymbol{\lambda}(\ell)\}$ converges to an optimal dual solution $\boldsymbol{\lambda}^*$ (cf. Proposition 2.2). The quantity of which the limit is taken in (2.64) is equivalently written in the following form involving the particular $\boldsymbol{\lambda}^*$:

$$\begin{aligned}\frac{1}{\ell}\sum_{j=1}^{\ell}\frac{\boldsymbol{\lambda}(j+1)-\boldsymbol{\lambda}(j)}{\beta_j}=&\frac{1}{\ell}\sum_{j=2}^{\ell}\left(\frac{1}{\beta_j}-\frac{1}{\beta_{j-1}}\right)(\boldsymbol{\lambda}(j)-\boldsymbol{\lambda}^*)\\&+\frac{1}{\ell}\frac{1}{\beta_\ell}(\boldsymbol{\lambda}^*-\boldsymbol{\lambda}(\ell+1))-\frac{1}{\ell}\frac{1}{\beta_1}(\boldsymbol{\lambda}^*-\boldsymbol{\lambda}(1)).\end{aligned} \tag{2.65}$$

Substituting $\beta_j=a/(c+j)$ and introducing the triangle inequality imply that

$$\left\| \frac{1}{\ell} \sum_{j=1}^{\ell} \frac{\boldsymbol{\lambda}(j+1) - \boldsymbol{\lambda}(j)}{\beta_j} \right\| \leq \frac{1}{\ell} \sum_{j=2}^{\ell} \frac{1}{a} \|\boldsymbol{\lambda}(j) - \boldsymbol{\lambda}^*\| + \frac{c+\ell}{a\ell} \|\boldsymbol{\lambda}^* - \boldsymbol{\lambda}(\ell+1)\| + \frac{c+1}{a\ell} \|\boldsymbol{\lambda}^* - \boldsymbol{\lambda}(1)\|. \quad (2.66)$$

Upon rearranging the right-hand side of (2.66), it is deduced that

$$\left\| \frac{1}{\ell} \sum_{j=1}^{\ell} \frac{\boldsymbol{\lambda}(j+1) - \boldsymbol{\lambda}(j)}{\beta_j} \right\| \leq \frac{1}{\ell} \sum_{j=1}^{\ell} \frac{1}{a} \|\boldsymbol{\lambda}(j) - \boldsymbol{\lambda}^*\| + \frac{c+\ell}{a\ell} \|\boldsymbol{\lambda}^* - \boldsymbol{\lambda}(\ell+1)\| + \frac{c}{a\ell} \|\boldsymbol{\lambda}^* - \boldsymbol{\lambda}(1)\|. \quad (2.67)$$

Each term in the right-hand side of (2.67) has limit zero as $\ell \to \infty$ (the first one due to Toeplitz's lemma). Therefore, the claim holds.

(a)-(ii) Eq. (2.33) will be employed to show that a bound as in (2.58) holds for $\bar{\mathbf{s}}(\ell)$ and $\bar{\mathbf{q}}_m(\ell)$. Combining the latter with the lim inf on (2.38) will lead to (2.60).

It holds from the Lagrange multiplier updates in a fashion similar to (2.47) and (2.48) that

$$-\frac{1}{\ell} \sum_{l=1}^{\ell} \boldsymbol{\lambda}^T(l) \left[\sum_{i=1}^{N} \mathbf{g}_i[\mathbf{x}_i(\tau_i(l))] \right] \leq \frac{1}{2} \frac{1}{\ell} \sum_{l=1}^{\ell} \frac{\|\boldsymbol{\lambda}(l)\|^2 - \|\boldsymbol{\lambda}(l+1)\|^2}{\beta_l} + \frac{G^2}{2} \frac{1}{\ell} \sum_{l=1}^{\ell} \beta_l. \quad (2.68)$$

The second term has limit zero as $\ell \to \infty$, while in a fashion similar to part (i), the following is true for the first term.

Claim: It holds that

$$\lim_{\ell \to \infty} \frac{1}{\ell} \sum_{l=1}^{\ell} \frac{\|\boldsymbol{\lambda}(l)\|^2 - \|\boldsymbol{\lambda}(l+1)\|^2}{\beta_l} = 0. \quad (2.69)$$

The desired relationship follows by taking the lim sup in (2.33) and using the previous results.

(b) The proof is identical to the one of Proposition 2.4. □

The ensuing section presents a cooperative demand response problem, to which the algorithm of the present section can be applied.

2.2 Cooperative Load Control Formulation

Consider M residences provided with electricity from the same utility company. Each residence has a smart meter that communicates with the various devices per residence, and also with the utility company through the AMI. It is supposed that the cost structure of energy provided by the utility company is determined in advance for a given time period $\{1, \ldots, T\}$. Each time slot of the scheduling horizon can represent e.g., one hour, with $T = 24$ corresponding to one day.

2.2.1 Residential Appliances

A base residential non-deferrable load is considered along with two classes of devices with adjustable power, denoted respectively as $\mathcal{K}_m^{(1)}$ and $\mathcal{K}_m^{(2)}$, where $m = 1, \ldots, M$ indexes the residences. The base load at slot t is denoted by p_{m0}^t, $t = 1, \ldots, T$, and can be e.g., lights or computers.

The devices of classes $\mathcal{K}_m^{(1)}$ and $\mathcal{K}_m^{(2)}$ are generically indexed by k, and p_{mk}^t denotes the power consumption of device k over slot t. Note that the term power here represents essentially consumption over the fixed duration of the slot, and therefore has units of kilowatt hours. The particular characteristics of those classes are as follows:

- Class $\mathcal{K}_m^{(1)}$ contains devices with a prescribed energy requirement E_{mk} that has to be completed between slots S_{mk} (start time) and T_{mk} (termination time). An example is PHEV charging, where the user may specify the charging to start e.g., at midnight, and finish by a morning hour. Power consumption vectors $\mathbf{p}_{mk} := (p_{mk}^1, \ldots, p_{mk}^T)$ across slots are constrained to be in the set

$$\mathcal{P}_{mk} := \Bigg\{ \mathbf{p}_{mk} \in \mathbb{R}^T \Big| \sum_{t=S_{mk}}^{T_{mk}} p_{mk}^t = E_{mk}; \\ p_{mk}^t \in [p_{mk}^{\min}, p_{mk}^{\max}],\ t = S_{mk}, \ldots, T_{mk}; p_{mk}^t = 0,\ \text{o/w} \Bigg\}. \quad (2.70)$$

  The sought power consumption variables belonging to the interval $[p_{mk}^{\min}, p_{mk}^{\max}]$ (with $p_{mk}^{\min} \geq 0$) for each slot t in order to meet E_{mk} over the horizon, will be the result of the optimization formulation.

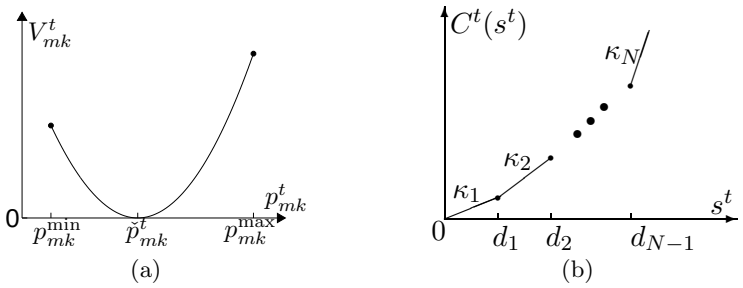


Figure 2.1: (a) Example of disutility function $V^t_{mk}(p^t_{mk})$. The feasible range is $[p^{\min}_{mk}, p^{\max}_{mk}]$, while the value $\check{p}^t_{mk}$ represents a desirable set point with minimal dissatisfaction. (b) Example of cost function $C^t(s^t)$. The cost is piecewise linear, consisting of N line segments, with slopes $\kappa_1, \ldots, \kappa_N$, and breakpoints $d_0 = 0, d_1, \ldots, d_{N-1}$.

- Class $\mathcal{K}^{(2)}_m$ includes devices operating with power in $[p^{\min}_{mk}, p^{\max}_{mk}]$, but without a total energy requirement. Instead, a *disutility* function $V^t_{mk}(p^t_{mk})$ is introduced to capture dissatisfaction of the end-user for operating away from a nominal point. The premise is that the end-user may choose to operate away from a nominal point, if this can reduce the electricity bill, as determined by the optimization formulation. An example from this class is an A/C unit. The disutility function is selected to be convex, and may vary with time to reflect the variable importance of operating the device across time. A disutility function example is illustrated in Fig. 2.1a. The power consumption vectors $\mathbf{p}_{mk} := (p^1_{mk}, \ldots, p^T_{mk})$ for class 2 devices are constrained to be in the set

$$\mathcal{P}_{mk} := \{\mathbf{p}_{mk} \in \mathbb{R}^T | p^t_{mk} \in [p^{\min}_{mk}, p^{\max}_{mk}], \\ t = S_{mk}, \ldots, T_{mk};\ p^t_{mk} = 0,\ \text{o/w}\}. \quad (2.71)$$

  When a time range $\{S_{mk}, \ldots, T_{mk}\}$ over which the device will be operated is given, it holds that $V^t_{mk}(p^t_{mk}) = 0$ for $t < S_{mk}$ or $t > T_{mk}$.

The notation $\mathbf{p}_m := \{\mathbf{p}_{mk}\}_{k \in \mathcal{K}^{(1)}_m \cup \mathcal{K}^{(2)}_m}$ collects the power consumptions of all devices of residence m. Note that the feasible set where $\mathbf{p}_m$ belongs to is convex.

2.2.2 Residential Storage Device Model

Residence m is also allowed to have a battery. Let $x_m^t \geq 0$, $t = 1, ..., T$, be the state of charge of the battery (i.e., energy available) at the end of slot t; and $x_m^{\max}$ the capacity of the battery, so that $x_m^t \leq x_m^{\max}$. The available energy at the beginning of the horizon is denoted by x_m^0.

The battery can be either charged or discharged during slot t. In the latter case, the energy stored in the battery is supplied to the residential appliances. The optimal decision will be the result of the optimization formulation. Let b_m^t, $t = 1, \ldots, T$, denote the energy drawn from or provided to the battery at slot t, where $b_m^t < 0$ in the former case, and $b_m^t > 0$ in the latter case. The charge/discharge variables as well as the energy stored in the battery are related by the dynamical equation

$$x_m^t = x_m^{t-1} + b_m^t, \quad t = 1, \ldots, T. \tag{2.72}$$

Variables b_m^t are constrained in three different ways:

i) Variables b_m^t are limited by maximum charge $b_m^{\text{ch}} > 0$ and discharge $b_m^{\text{dis}} < 0$ rates, i.e., $b_m^{\text{dis}} \leq b_m^t \leq b_m^{\text{ch}}$.

ii) The battery-supplied energy is no more than the current energy consumption, i.e., $b_m^t + \sum_k p_{mk}^t \geq 0$ for all t and m, where the summation over k includes the base load and all devices of classes 1 and 2 of end-user m.

iii) Each battery has efficiency $\eta_m \in (0, 1)$, meaning that if x_m^{t-1} is stored at the end of slot $t-1$, the discharge at slot t is limited by $b_m^t \geq -\eta_m x_m^{t-1}$, $t = 1, \ldots, T$.

The set of feasible $\mathbf{b}_m := (b_m^1, \ldots, b_m^T)$ per residence m is given by the stated constraints collected in the following set:

$$\begin{aligned}\mathcal{B}_m(\mathbf{p}_m) = \{\mathbf{b}_m \in \mathbb{R}^T | b_m^{\text{dis}} \leq b_m^t \leq b_m^{\text{ch}},\, b_m^t + \sum\nolimits_k p_{mk}^t \geq 0,\\ \text{and } b_m^t \geq -\eta_m x_m^{t-1},\ \text{for some } x_m^t \in [0, x_m^{\max}] \text{ such that}\\ x_m^t = x_m^{t-1} + b_m^t,\, x_m^T \geq x_m^{\min};\, t = 1, \ldots, T\}.\end{aligned} \tag{2.73}$$

In (2.73), a constraint for the final state of charge x_m^T, which is available at the beginning of the next horizon, has been included; while the initial state x_m^0 is known. The b_m^t's and p_{mk}^t's are coupled for end-user m through the constraint $b_m^t + \sum_k p_{mk}^t \geq 0$, and this is indicated by the dependence of $\mathcal{B}_m$ on $\mathbf{p}_m$. Note also that $\mathcal{B}_m(\mathbf{p}_m)$ is a convex set.

2.2.3 Social Welfare Maximization

Let $C^t(\cdot)$ denote the pricing function representing the cost of electricity over slot t. This is the cost that the utility incurs in order to provide electricity to the end-user. For instance, this cost can be determined by the company's bidding to the wholesale market, or it can represent distribution network operating costs; see e.g., [12–16, 51, 52].

Consider also a non-adjustable base load p_0^t from other end-users in the system—due to e.g., end-users not participating in the DSM program, or a commercial load. Then, the cost of electricity over slot t is given by $C^t(\sum_m(\sum_k p_{mk}^t + b_m^t) + p_0^t)$, where the summation over m is over all residences. Let s^t denote a variable upper-bounding $\sum_m(\sum_k p_{mk}^t + b_m^t) + p_0^t$, which is interpreted as the power provided by the utility company, and define $\mathbf{s} := (s^1, \ldots, s^T)$. Let also $\mathbf{p}$ and $\mathbf{b}$ collect all $\mathbf{p}_m$ and $\mathbf{b}_m$, respectively, for all m. Recall that p_{m0}^t and p_0^t for all m and t are not optimization variables, and hence they are not included in $\mathbf{p}_m$ or $\mathbf{p}$. The multi-residential load control task amounts to minimizing the total cost of electricity as well as the total dissatisfaction of the end-users, that is

$$\min_{\mathbf{p},\mathbf{b},\mathbf{s}} \quad \sum_{t=1}^{T} C^t(s^t) + \sum_{t=1}^{T}\sum_{m=1}^{M}\sum_{k\in\mathcal{K}_m^{(2)}} V_{mk}^t(p_{mk}^t) \tag{2.74a}$$

$$\text{subj. to} \quad \sum_{m=1}^{M}\left(\sum_k p_{mk}^t + b_m^t\right) + p_0^t \le s^t,\ t = 1, \ldots, T \tag{2.74b}$$

$$0 \le s^t \le s^{\max}, \quad t = 1, \ldots, T \tag{2.74c}$$

$$\mathbf{p}_{mk} \in \mathcal{P}_{mk},\ \mathbf{b}_m \in \mathcal{B}_m(\mathbf{p}_m),$$
$$k \in \mathcal{K}_m^{(1)} \cup \mathcal{K}_m^{(2)},\ m = 1, \ldots, M. \tag{2.74d}$$

A constraint ensuring a safety cap $s^{\max}$ upon the total power consumption has been introduced in (2.74c), taking into account security and reliability considerations for the distribution network from the utility company to the residences. Note that the model includes as special case the situation where there are no batteries at all; in this case, variables b_m^t and x_m^t as well as the constraint $\mathbf{b}_m \in \mathcal{B}_m(\mathbf{p}_m)$ will simply not be present.

The function $C^t(s^t)$ is chosen to be convex, continuous, and strictly increasing (see Fig. 2.1b for an example). Because of the latter, (2.74b) holds with equality at the

optimal point, matching the power supply with the demand. Moreover, (2.74) is a convex optimization problem. It is supposed that there are feasible p_{mk}^t, b_m^t, and s^t, so that (2.74b) holds as strict inequality. This is the standard Slater's constraint qualification, ensuring zero duality gap and existence of optimal Lagrange multipliers.

It is clear that problem (2.74) is of the generic form (2.1), and therefore the results of Section 2.1 can be applied. The exact algorithm and its distributed implementation are detailed in Section 2.3.

The objective in (2.74a) is the opposite of social welfare. Therefore, problem (2.74) amounts to maximizing social welfare of end-users and the utility company. Pricing interpretations related to (2.74) are described in the ensuing subsection, while a sensitivity analysis of the social welfare with respect to the parameters of the cost function is provided in Section 2.5.

2.2.4 Lagrangian Duality and Economic Interpretation

Let λ^t denote the Lagrange multiplier corresponding to (2.74b), and $\boldsymbol{\lambda} := (\lambda^1, \ldots, \lambda^T)$. Then, keeping the rest of the constraints implicit, the Lagrangian function for (2.74) is

$$L(\boldsymbol{\lambda}, \mathbf{p}, \mathbf{b}, \mathbf{s}) = \sum_{t=1}^{T} C^t(s^t) + \sum_{t=1}^{T} \sum_{m=1}^{M} \sum_{k \in \mathcal{K}_m^{(2)}} V_{mk}^t(p_{mk}^t) + \sum_{t=1}^{T} \lambda^t \left[\sum_{m=1}^{M} \left(\sum_{k} p_{mk}^t + b_m^t \right) + p_0^t - s^t \right]. \quad (2.75)$$

The dual function and the dual problem take the form

$$\rho(\boldsymbol{\lambda}) = \min_{\substack{0 \le s^t \le s^{\max},\, t=1,\ldots,T \\ \mathbf{p}_{mk} \in \mathcal{P}_{mk},\, \mathbf{b}_m \in \mathcal{B}_m(\mathbf{p}_m) \\ k \in \mathcal{K}_m^{(1)} \cup \mathcal{K}_m^{(2)},\, m=1,\ldots,M}} L(\boldsymbol{\lambda}, \mathbf{p}, \mathbf{b}, \mathbf{s}) \quad (2.76)$$

$$\rho^* = \max_{\boldsymbol{\lambda} \ge \mathbf{0}} \rho(\boldsymbol{\lambda}). \quad (2.77)$$

Because problem (2.74) is convex and has zero duality gap, standard duality theory can be used to interpret how the Lagrange multipliers act as pricing signals coordinating the utility company with the end-users; see also [46, Sec. 5.1.6] describing Lagrange multipliers as a coordination mechanism and [13] for a related DSM application.

Specifically, let $\mathbf{p}^*$, $\mathbf{b}^*$, and $\mathbf{s}^*$ denote the optimal solution of (2.74), and $\boldsymbol{\lambda}^*$ the optimal Lagrange multipliers corresponding to the dual solution [cf. (2.77)]. Variables $\mathbf{p}^*$, $\mathbf{b}^*$, and $\mathbf{s}^*$ are also the minimizers in (2.76) for $\boldsymbol{\lambda} = \boldsymbol{\lambda}^*$, i.e., $\rho(\boldsymbol{\lambda}^*) = L(\boldsymbol{\lambda}^*, \mathbf{p}^*, \mathbf{b}^*, \mathbf{s}^*)$; see e.g., [47, Prop. 6.2.5]. Using the fact that the primal and dual optimal values are the same, the following holds:

$$\begin{aligned}\sum_{t=1}^{T} C^t(s^{t*}) + \sum_{t=1}^{T}\sum_{m=1}^{M}\sum_{k\in\mathcal{K}_m^{(2)}} V_{mk}^t(p_{mk}^{t*}) &= \rho(\boldsymbol{\lambda}^*) \\ &= \underbrace{\sum_{t=1}^{T}[C^t(s^{t*}) - \lambda^{t*}s^{t*}]}_{\text{net cost for utility}} + \underbrace{\sum_{t=1}^{T}\lambda^{t*}p_0^{t*}}_{\text{cost for base load}} \\ &\quad + \sum_{m=1}^{M}\underbrace{\sum_{t=1}^{T}\left[\lambda^{t*}\left(\sum_k p_{mk}^{t*} + b_m^{t*}\right) + \sum_{k\in\mathcal{K}_m^{(2)}} V_{mk}^t(p_{mk}^{t*})\right]}_{\text{aggregate cost for residence } m}\end{aligned} \tag{2.78}$$

where the last equality follows upon straightforward rearrangements of the terms in the Lagrangian function.

The Lagrange multiplier λ^{t*} can be interpreted as the price charged from the utility company at slot t. The term "net cost for utility" in (2.78) represents the cost the utility incurs to provide electricity, minus the revenue from selling this electricity to end-users. Therefore, the units of Lagrange multipliers can be interpreted to be monetary units/kWh. The term "aggregate cost for residence m" in (2.78) represents the payment to the utility company $\sum_t \lambda^{t*}(\sum_k p_{mk}^{t*} + b_m^{t*})$ plus the disutility experienced by the end-user. It is instrumental in this interpretation to recall that the total power consumed at slot t, namely, $\sum_m(\sum_k p_{mk}^{t*} + b_m^{t*}) + p_0^{t*}$, is exactly the electricity provided by the utility, s^{t*}; i.e., (2.74b) holds as equality.

2.3 Distributed Scheme

The multi-residence load control problem (2.74) is separable, meaning that the objective (2.74a) and the constraint (2.74b) comprise sums of functions that depend only on some (but not all) optimization variables. In particular, there are $1+M$ groups of variables, namely, $\mathbf{s}$ and the power consumption variables $\mathbf{p}_m$ and $\mathbf{b}_m$ for each residence m.

The methods and results of Section 2.1 are applicable, and the synchronous algorithm is given next.

The Lagrangian minimization (2.76)—which defines the dual function and is also part of the subgradient iterations—is easily seen to be decomposable to individual minimizations with respect to $\mathbf{s}$, and also with respect to the $\mathbf{p}_m$ and $\mathbf{b}_m$ for each residence m. Specifically, the subgradient method consists of the following iterations, indexed by $\ell = 1, 2, \ldots$ and initialized with arbitrary $\lambda^t(1) \geq 0$:

$$s^t(\ell) \in \underset{0 \leq s^t \leq s^{\max}}{\arg\min} \{C^t(s^t) - \lambda^t(\ell) s^t\}, \quad t = 1, \ldots, T \tag{2.79a}$$

$$\{\mathbf{p}_m(\ell), \mathbf{b}_m(\ell)\} \in \underset{\substack{\mathbf{p}_{mk} \in \mathcal{P}_{mk},\, \mathbf{b}_m \in \mathcal{B}_m(\mathbf{p}_m) \\ k \in \mathcal{K}_m^{(1)} \cup \mathcal{K}_m^{(2)}}}{\arg\min} \sum_{t=1}^{T} \left[\lambda^t(\ell) \left(\sum_k p_{mk}^t + b_m^t \right) + V_{mk}^t(p_{mk}^t) \right], m = 1, \ldots, M \tag{2.79b}$$

$$\lambda^t(\ell+1) = \left[\lambda^t(\ell) + \beta_\ell \left[\sum_{m=1}^{M} \left(\sum_k p_{mk}^t(\ell) + b_m^t(\ell) \right) + p_0^t - s^t(\ell) \right] \right]^+ \quad t = 1, \ldots, T \tag{2.79c}$$

where β_ℓ is the stepsize. Recall also that the summation over k involves the base load (for which $p_{m0}^t(\ell) \equiv p_{m0}^t$), and all appliances at residence m.

The form of the updates (2.79) readily suggests a distributed implementation. In particular, the utility broadcasts at every iteration ℓ the value of the Lagrange multipliers $\{\lambda^t(\ell)\}_{t=1}^T$ to all residential smart meters through the AMI. These Lagrange multipliers are needed to solve (2.79b) with respect to $\mathbf{p}_m$ and $\mathbf{b}_m$ at the smart meter of residence m. Each residence sends back to the utility company the values of $\{\sum_k p_{mk}^t(\ell) + b_m^t(\ell)\}_{t=1}^T$, which correspond to the total power consumption per hour. In this way, residences do not reveal the individual appliance consumption or battery storage profiles. The exchange of information is illustrated in Fig. 2.2.

The minimization (2.79a) with respect to s^t takes place at the utility company at every iteration ℓ. The Lagrange multiplier updates (2.79c) also take place at the utility, which has knowledge of the non-adjustable load p_0^t, and then the process is repeated. The distributed load control algorithm runs ahead of the scheduling horizon. If the parameters change at a slot during the horizon, then problem (2.74) with the new parameters can be solved for the remaining slots.

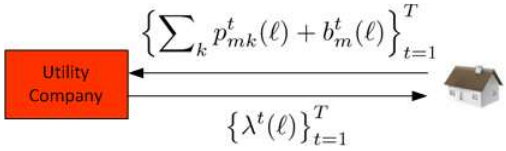


Figure 2.2: Exchange of information between utility company and residence m.

The problems in (2.79a) and (2.79b) are convex. Therefore, they can be solved efficiently and locally at the utility company or at the smart meter of residence m. More details on the particular algorithms that can be used are provided next.

The minimization in (2.79a) involves a single variable and a box constraint; therefore, it is easy to solve in closed form. Specifically, if the cost function has a derivative with inverse $(C^{t\prime})^{-1}$, then the solution is

$$s^t(\ell) = \min\{\max\{(C^{t\prime})^{-1}(\lambda^t(\ell)), 0\}, s^{\max}\}. \tag{2.80}$$

If the cost is piecewise linear as in Fig. 2.1b, and without loss of generality $d_{N-1} < s^{\max} = d_N$, then the solution is

$$s^t(\ell) = \begin{cases} 0 & \text{if } \lambda^t(\ell) < \kappa_1 \\ d_i & \text{if } \kappa_i \le \lambda^t(\ell) < \kappa_{i+1},\ i = 1, \dots, N-1 \\ s^{\max} & \text{if } \kappa_N \le \lambda^t(\ell). \end{cases} \tag{2.81}$$

In the absence of a storage element at residence m, the minimization (2.79b) decouples into *per device* minimizations as follows:

$$\mathbf{p}_{mk}^t(\ell) \in \underset{\mathbf{p}_{mk} \in \mathcal{P}_{mk}}{\arg\min} \sum_{t=1}^T \lambda^t(\ell) p_{mk}^t, \quad k \in \mathcal{K}_m^{(1)} \tag{2.82a}$$

$$p_{mk}^t(\ell) \in \underset{p_{mk}^t \in [p_{mk}^{\min}, p_{mk}^{\max}]}{\arg\min} \{V_{mk}^t(p_{mk}^t) + \lambda^t(\ell) p_{mk}^t\},$$
$$t = S_{mk}, \dots, T_{mk},\ k \in \mathcal{K}_m^{(2)}. \tag{2.82b}$$

The minimization in (2.82b) is analogous to the one in (2.79a). The minimization in (2.82a) can be solved easily using the algorithm listed in Table 2.1.

In the presence of a storage element, the minimization (2.79b) couples all the devices and the storage element, due to the constraint $b_m^t + \sum_k p_{mk}^t \ge 0$ [cf. (2.73)]. Note that all constraints are linear. The problem is convex, and can be solved by an interior point

Table 2.1: Power consumption scheduling for class 1 devices.

Arrange $\lambda^{S_{mk}}, \ldots, \lambda^{T_{mk}}$ in increasing order

Define the function $\tilde{t} : \{1, \ldots, T_{mk} - S_{mk} + 1\} \to \{S_{mk}, \ldots, T_{mk}\}$ so that $\tilde{t}(n)$ returns the time index of the nth smallest Lagrange multplier among $\lambda^{S_{mk}}, \ldots, \lambda^{T_{mk}}$

$n \leftarrow 1$

$\texttt{accumEnergy} \leftarrow 0$

while $\texttt{accumEnergy} < E_{mk}$ **do**

  if $E_{mk} - (\texttt{accumEnergy} + p_{mk}^{\max}) > p_{mk}^{\min}$ **then**

    $p_{mk}^{\tilde{t}(n)}(\ell) \leftarrow p_{mk}^{\max}$

    $n \leftarrow n + 1$

    $\texttt{accumEnergy} \leftarrow \texttt{accumEnergy} + p_{mk}^{\tilde{t}(n)}(\ell)$

  else

    $p_{mk}^{\tilde{t}(n)}(\ell) \leftarrow E_{mk} - (\texttt{accumEnergy} + p_{mk}^{\min})$

    $p_{mk}^{\tilde{t}(n+1)}(\ell) \leftarrow p_{mk}^{\min}$

  end if

end while

algorithm. In particular, it is a linear or a quadratic program, if the disutility function V_{mk}^t is respectively linear or quadratic.

Note that in practice, the following can happen to an AMI message, $\lambda^t(\ell)$ or $\sum_k p_{mk}^t(\ell) + b_m^t(\ell)$ [cf. 2.2], at slot ℓ:

- It might not be transmitted, e.g., if there is a malfunction.
- It might not be received due to e.g., noise in the communications network, especially if there is a wireless network connecting the residential smart meters with the utility company.

Both situations are referred to as *lost AMI messages*, and are modeled next with outdated Lagrange multipliers, in the setup of Section 2.1.

Lost AMI messages cause power consumption data $\sum_k p_{mk}^t + b_m^t$ used for the Lagrange multiplier updates in (2.79c) to correspond to outdated Lagrange multipliers.

Specifically, iterates $\sum_k p_{mk}^t + b_m^t$ corresponding to the most recent Lagrange multipliers available are used in (2.79c), as described next.

- Suppose that the message $\{\lambda^t(\ell)\}_{t=1}^T$ for residence m at slot ℓ is lost. Let $\tau_m(\ell) \leq \ell$ be the index of the most recent Lagrange multiplier available. Then, the message transmitted to the utility company is $\{\sum_k p_{mk}^t(\tau_m(\ell)) + b_m^t(\tau_m(\ell))\}_{t=1}^T$, which corresponds to the minimization of the Lagrangian in (2.79b) with $\boldsymbol{\lambda}^t(\tau_m(\ell))$ instead of $\boldsymbol{\lambda}^t(\ell)$. See also Table 2.2 for an example.
- Suppose on the other hand that the message $\{\lambda^t(\ell)\}_{t=1}^T$ for residence m at slot ℓ is received, but the transmitted message $\{\sum_k p_{mk}^t(\ell) + b_m^t(\ell)\}_{t=1}^T$ is lost. Then, the most recent received message will be used instead, which effectively corresponds to the Lagrangian minimizer at an outdated Lagrange multiplier $\boldsymbol{\lambda}(\tau_m(\ell))$.

In a nutshell, if messages in either communication direction are lost, the algorithm is effectively using $\sum_{m=1}^M \left(\sum_k p_{mk}^t(\tau_m(\ell)) + b_m^t(\tau_m(\ell))\right) + p_0^t - s^t(\ell)$ as the update direction in (2.79c).

The primal averages are formed as follows for the case of constant weights

$$\bar{\mathbf{s}}(\ell) = \frac{1}{\ell}\sum_{l=1}^{\ell} \mathbf{s}(l) \tag{2.83}$$

$$\bar{\mathbf{p}}_m(\ell) = \frac{1}{\ell}\sum_{l=1}^{\ell} \mathbf{p}_m(\tau_m(l)), \qquad \bar{\mathbf{b}}_m(\ell) = \frac{1}{\ell}\sum_{l=1}^{\ell} \mathbf{b}_m(\tau_m(l)) \tag{2.84}$$

and similarly for the weights proportional to the stepsize.

The punchline is that the results of Section 2.1 cover the convergence of the distributed algorithm, even under lost AMI messages. The subgradient norm bounds in those results of take the form

$$G_m = \sqrt{T}\left(\sum_k p_{mk}^{\max} + \max\{b_m^{\text{ch}}, |b_m^{\text{dis}}|\}\right) \tag{2.85}$$

$$G = \sqrt{T}\left(s^{\max} + \sum_k p_{mk}^{\max} + \max\{b_m^{\text{ch}}, |b_m^{\text{dis}}|\}\right) \tag{2.86}$$

where $p_{m0}^{\max} = \max_{t\in\{1,\ldots,T\}} p_{m0}^t$.

Note that in general, primal averaging is necessary if the primal objective is *not strictly convex*. This can be the case here for two practically relevant reasons: (a) the

Table 2.2: Example of AMI message progression, with x denoting a lost message. The Lagrange multipliers received at and the power consumption data transmitted from residence m are shown, using the shorthand notation $P_m^t(\ell) := \sum_k p_{mk}^t(\ell) + b_m^t(\ell)$.

Iter. ℓ	1	2	3	4	5	...
Lagr. mult. rx	$\lambda^t(1)$	x	x	$\lambda^t(4)$	x	...
Power cons. tx	$P_m^t(1)$	$P_m^t(1)$	$P_m^t(1)$	$P_m^t(4)$	$P_m^t(4)$	...
$\tau_m(\ell)$	1	1	1	4	4	...

objective can include functions which are not strictly convex in their argument, as with e.g., a piecewise linear $C^t(s^t)$ (cf. Fig. 2.1b); and (b) the objective is not a function of all optimization variables, namely, it does not involve the p_{mk}^t's for $k \in \mathcal{K}_m^{(1)}$ or the b_m^t's.

The constant (S1) as well as the diminishing stepsize rules (S2) and (S3) have their own merits. Specifically, although convergence with constant stepsize is within a ball (near-optimal), it is typically faster than convergence with a diminishing stepsize. Note further that the diminishing stepsize (S3) allows recovery of the primal variables with the averaging scheme (2.10) and (2.84), which can be preferable for two reasons: (a) it is the simplest averaging scheme; and (b) the scheme in (2.10) and (2.84) uses constant weights, and therefore has the potential to converge faster than the one using vanishing weights.

2.4 Numerical Tests

Two sets of numerical results are presented; one showing properties of the optimal load control and scheduling in a 6-user scenario (Subsection 2.4.1), and the other highlighting economic interpretations in a considerably larger scale scenario (Subsection 2.4.2).

The time horizon in both tests is 24 hours, corresponding to 8am, 9am, 10am, etc., until 7am of the next day. Each end-user has a base load p_{m0}^t, which is drawn randomly from a uniform distribution with limits shown in Fig. 2.3 as fraction of 3 kWh. The latter value corresponds to 50% of a typical peak residential load of 6 kW [52, Figs. 2.5–2.7]. On top of the base load there is adjustable residential load. The detailed setup and results for each test case are described next.

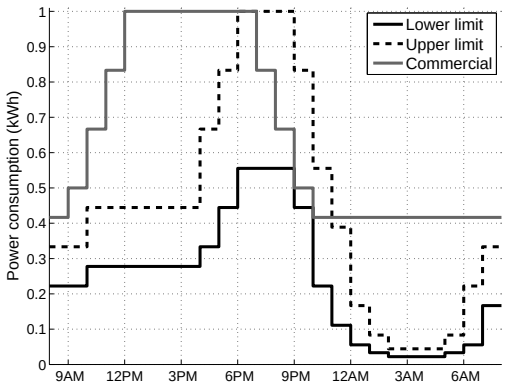


Figure 2.3: Upper and lower limits for must-run residential load as a fraction of 3 kWh; and commercial load as a fraction of 980 kWh.

2.4.1 Test Case 1: Load Control and Scheduling

For the scenario with 6 end-users, the characteristics of the residential devices are listed in Table 2.3. The class 1 device can be for instance a PHEV, which has battery of 10 kWh for a 5-mile drive [53], and has to be charged during the night. A typical value for maximum charging rate is 1.2 kW [54], giving a maximum hourly consumption of 1.2 kWh. The disutility function selected for the class 2 device is $V_{m2}^t = w_m^t(p_{m2}^{\max} - p_{m2}^t)^2$. This can e.g., be a room A/C unit that is to be operated during the day and in the evening, and has cooling capacity 10 000 BTU and energy efficiency ratio (EER) 10 [55], entailing a maximum hourly power consumption in the range of 1 kWh. The selected consumption parameters yield a daily residential power consumption in the order of 10 kWh to 30 kWh, which is representative of a household in the U.S. [56]. The cost function is $C^t(s^t) = 0.2(s^t)^2$ for all t (similar to [15]), and $s^{\max} = 40$ kWh. There is no additional base load p_0^t. The stepsize is $\beta_\ell = 1/(10 + \ell)$, which satisfies (S3), while averages (2.10) and (2.84) are used.

Load scheduling

The optimal schedules for the 6 residences are depicted in Figs. 2.4a–2.4f, and they verify the intuition. Specifically, the power consumption of the class 1 device takes its largest value between hours 1am and 5am for all end-users, in a fashion complementary

Table 2.3: Parameters of residential devices for Test Case 1. End-users 1–5 have a class 1 device, and end-users 1–3, 5, and 6 have a class 2 device. The units of $p_{mk}^{\min}$, $p_{mk}^{\max}$, and E_{mk} are kilowatt hours.

	User 1	User 2	User 3	User 4	User 5	User 6
E_{m1}	10	12	14	10	11	
$p_{m1}^{\max}$	1.4	1.4	1.5	1.4	1.4	
$p_{m1}^{\min}$	0	0	0	0	0	
S_{m1}	9pm	9pm	7pm	8pm	7pm	
T_{m1}	7am	6am	6am	6am	6am	
w_m^t	14	16	18		22	24
$p_{m2}^{\max}$	1.2	1.2	1.4		1.2	1
$p_{m2}^{\min}$	0	0	0		0	0
S_{m2}	11am	1pm	4pm		2pm	12pm
T_{m2}	11pm	9pm	10pm		8pm	8pm

to the base load. The power consumption for class 1 device of end-user 3 increases sharply after 10pm, as opposed to the more gradual increase for the other end-users. The reason is that E_{31} is the highest among the E_{m1}'s. It is also interesting to note that the power consumption of class 1 device for end-user 1 increases from 6am to 7am, while it is constrained to be zero at 7am for the remaining end-users (cf. Table 2.3). It can be deduced that the drop at hour 6am and then the rise at hour 7am for end-user 1 helps to smooth out the total power consumption across all users.

Adding a battery

Now suppose that end-user 1 also has a storage device with capacity $x_1^{\max} = 3.2$ kWh as in [17], and remaining parameters $b_1^{\text{ch}} = 0.5$ kWh, $b_1^{\text{dis}} = -0.5$ kWh, $\eta_1 = 0.95$, $x_1^0 = 1$ kWh, and $x_1^{\min} = 1$ kWh.

The total hourly power consumption for end-user 1 is shown in Fig. 2.5. The battery is charged at the hours where the consumption is seen to have increased over the one without battery. A clear effect of adding the storage device is that the peak power consumption of end-user 1 is reduced. Specifically, the battery is discharged during the hours 6pm–9pm, when the peak demand occurs. Moreover, Table 2.4 lists the optimal objective value and its constituent costs. As the battery increases the scheduling degrees of freedom of the system, the resulting costs are smaller.

Table 2.4: Results for Test Case 1

	Obj. value	$\sum_{t=1}^{T} C^t(s^{t*})$	$\sum_{\substack{t,m \\ k\in\mathcal{K}_m^{(2)}}} V_{mk}^t(p_{mk}^{t*})$
Without battery	614.09	591.7	22.387
With battery	601.37	579.96	21.413

Asynchronicity effects

Asynchronicity is now introduced (in the setup without battery) by setting $\tau_m(\ell) = \ell - \text{mod}\ (\ell - 1, D_m + 1)$, $\ell = 1, 2, \ldots$, for an integer $D_m \geq 0$. This has essentially the effect that only the minimizers $\mathbf{p}_m$ at slots $1, 2(D_m+1), 3(D_m+1), \ldots$, are used in the updates for all users (the units of D_m are slots). Therefore, condition C2 is satisfied. Setting $D_m = 0$ recovers the synchronous case.

The evolution of the objective value evaluated at the running averages $\bar{\mathbf{p}}_m(\ell)$ and $\bar{\mathbf{s}}(\ell)$ is depicted in Fig. 2.6 for different values of D_m. According to Proposition 2.5, convergence to the optimal value occurs under all delays, which is verified in the figure. Note though that convergence under larger delays in the subgradient updates is characterized by a higher overshoot and a larger lag.

2.4.2 Test Case 2: Scenario With Distribution System Data

A system with 420 end-users and an additional commercial base load is tested. The hourly commercial load p_0^t is depicted in Fig. 2.3, where the maximum corresponds to 980 kWh. The parameters of the residential devices are listed in Table 2.5, similar to the Test Case 1. Every end-user has a class 2 device, while 80% of the end-users (randomly selected) have a class 1 device. Finally, the peak base residential load and the commercial load coincide at 8 pm (cf. [52, Fig. 2.7]).

The parameters are selected as follows. The nominal total active power load across the three phases in the IEEE 123-node test feeder is approximately 3500 kW [57]. Supposing a peak residential load of 6 kW, 420 residential end-users correspond to 72% of the peak load, while the remaining 28% is the peak commercial load of 980 kWh. Moreover, the class 2 device can be thought to represent an A/C unit. In this case, $p_{m2}^{\max}$ corresponds to roughly 40% of the nominal peak residential load of 6 kW. The

Table 2.5: Parameters of residential devices for Test Case 2.

E_{m1} (kWh)	Uniform on {10, 11, 12}
$p_{m1}^{\max}$ (kWh)	Uniform on {1.4, 1.5 }
$p_{m1}^{\min}$ (kWh)	0
S_{m1}	Uniform on {5pm, 6pm, 7pm, 8pm, 9pm}
T_{m1}	5am w.p. 1/4, 6am w.p. 1/2, 7am w.p. 1/4
w_m^t	Uniform on {8, 9, 10, 11, 12}, same for all t
$\check{p}_{m2}^t/p_{m2}^{\max}$	Equally spaced from 0.1 to 0.8 for $\{S_{m2}, \ldots, T_{m2}-6\}$; 1 for $\{T_{m2}-5, \ldots, T_{m2}-2\}$; 0.8 for $t = T_{m2}-1$; 0.4 for $t = T_{m2}$
$p_{m2}^{\max}$ (kWh)	Uniform on {2.2, 2.3, 2.4, 2.5, 2.6, 2.7}
$p_{m2}^{\min}$ (kWh)	0
S_{m2}	Uniform on {10am, 11am, 12pm}
T_{m2}	Uniform on {10pm, 11pm, 12am, 1am }

disuitility function has the form $V_{m2}^t = w_m^t(\check{p}_{m2}^t - p_{m2}^t)^2$. Moreover, the parameters S_{m2}, T_{m2}, and $\check{p}_{m2}^t$, follow a typical daily profile [52, Fig. 2.4].

The cost function is $C^t(s^t) = 7 \cdot 10^{-4}(s^t)^2$ for all t—if the coefficient $7 \cdot 10^{-4}$ has units cents/(kWh)^2, then this cost function gives an incremental cost of approximately 49\$/MWh at 3500 kWh, which falls in the range of values used in [12]. The remaining parameters are $s^{\max} = 5000$ kWh and $\beta_\ell = 0.1/(5+\ell)$, while running averages (2.10) and (2.84) are used.

The resulting optimal Lagrange multipliers $\{\lambda^{t*}\}_{t=1}^{24}$ are shown in Fig. 2.7. It is interesting to observe that the magnitude of the Lagrange multipliers follows the variation of the total load across the 24 hours, which is depicted in Fig. 2.8 (thick grey curve). For example, they peak at 6pm, which is also when the highest power consumption occurs.

The results of the proposed formulation are also compared with a scheme where the residential loads are not controlled, and the residential power consumption is elicited by a flat pricing scheme. Specifically, with reference to (2.78), suppose that $\{\lambda^{t*}\}$ are substituted by other quantities, π, the same for all t, playing the role of prices. In this case, the power consumed by the end-users is computed as follows. For class 1 devices, the power consumption at slots $S_{mk}, \ldots, T_{mk}$ is $p_{mk}^t = E_{mk}/(T_{mk} - S_{mk} + 1)$, and 0 otherwise. For class 2 devices, the power is obtained from the minimization in (2.82b) with π instead of $\lambda^t(\ell)$. The power provided matches the total power consumption; that

Table 2.6: Results for Test Case 2

	Obj. value	C-part	V-part	$\sum_{t=1}^{T} s^t$	Load Factor
Optimal	49264	48007	1257.5	37295	0.56
Fixed pricing	51341	50038	1302.8	37255	0.53

is, $s^t = \sum_{m,k} p_{mk}^t + p_0^t$.

Fig. 2.9 shows the objective value—as defined in the left-hand side of (2.78)—for the two schemes. Recall that the objective value is to be minimized, as it represents the opposite of the social welfare. A range of prices covering the magnitudes of the optimal Lagrange multipliers is used. It can be seen that the optimal load control scheme is the one with the largest social welfare for the whole range of fixed prices.

It is noted in Fig. 2.9 that the objective value is minimized approximately for $\pi = 3$. The value $\pi = 3$ is also used to compute the total power consumption depicted with the thin black line in Fig. 2.8. The power consumption resulting from the optimal load control scheme is much smoother than the one from the fixed pricing scheme; compare for instance the values of the peak power consumptions. Smoothing out the aggregate load is a major attractive feature of (2.74).

Table 2.6 lists the objective value and its constituent costs, as well as the total power consumption and the load factor, obtained from the two schemes by averaging 10 Monte Carlo realizations. For the fixed pricing scheme, the value of π minimizing the objective value (in a fashion similar to Fig. 2.9) was chosen. The load factor is defined as the ratio $\sum_t s^t/(\max_t s^t \cdot T)$, and the closer it is to 1, the smoother the power consumption is [52, p. 58]. Observe that all costs are smaller under the optimal scheme, while the load factor is improved, although the total power consumption has increased.

Finally, the effect of the disutility functions on the total power consumption is investigated. Specifically, the parameters of Table 2.5 are used, while 100% of the end-users have class 1 loads. Moreover, the disutility function V_k^t of randomly selected sets of end-users is set to zero. This implies that certain end-users would be willing to completely curtail power consumption from class 2 devices. Five Monte Carlo utilizations are used to obtain the results of Fig. 2.10 and Table 2.7. Specifically, fig. 2.10 depicts the total

Table 2.7: Load factor with different disutility functions

Percentage of end-users with $V_k^t \equiv 0$	0%	33%	66%	100 %
Load Factor	0.57	0.59	0.61	0.64

power consumption elicited by the optimal load scheduling algorithm for different percentages of the end-users adopting zero disutility function. The associated load factor is given in Table 2.7. Clearly, the total power consumption is flatter (and the load factor is larger) when a larger percentage of end-users has zero disutility.

2.5 Sensitivity Analysis

This section studies how the objective value in (2.74a) will change, if the parameters of the optimization problem change. For instance, if the cost function is $C^t(s^t) = a(s^t)^2 + bs^t$ and the parameter a increases to $a + \Delta a$, then how much does the optimal social welfare change?

Explicit formulas for the directional derivatives of the optimal objective value in (2.74) with respect to changes in certain parameters of the cost functions $C^t(s^t)$ and $V_{mk}^t(p_{mk}^t)$ are provided. These are useful in three respects. First, the sign of the derivative is enough to assess if the objective value is increasing or decreasing in the parameter. Second, by comparing the derivatives with respect to different parameters, it is possible to identify the most influential parameters. And third, the value of the derivative multiplied by the parameter change approximately gives the change in the objective value. The sensitivity analysis presented here relies on the formulas of [58, Coroll. 2.3.8] and [59].

For simplicity in exposition, no storage devices are considered. Since the objective is the negative of the social welfare, the notation NSW is used for the optimal objective value of (2.74). Now, suppose that the cost $C^t(s^t)$ depends on a scalar parameter r; this is denoted by $C^t(s^t; r)$. For example, if $C^t(s^t)$ is quadratic, r could be the coefficient of $(s^t)^2$. As the optimal value of (2.74) depends on r, this is denoted as $\mathsf{NSW}(r)$.

All assumptions on problem (2.74) stated so far (convexity, Slater constraint qualification, etc.) hold throughout. Suppose that s^{t*} and p_{mk}^{t*} are the solutions to (2.74) for a given nominal value r_0 of the parameter. Two sensitivity results will be given next,

depending on whether $C^t(s^t; r)$ is differentiable (e.g., linear or quadratic) or not (e.g., piecewise linear).

Suppose that $C^t(s^t; r)$ is continuously differentiable in both s^t and r. Under a certain regularity condition, which will be explained shortly, the derivative of $\mathsf{NSW}(r)$ at r_0 is given by

$$\frac{\partial \mathsf{NSW}(r_0)}{\partial r} = \frac{\partial C^t(s^{t*}; r_0)}{\partial r}. \tag{2.87}$$

Similarly, if instead of $C^t(s^t)$, one is interested in changes in a parameter of the disutility function $V^t_{mk}(p^t_{mk}; r)$, then the following holds under differentiability of V^t_{mk}:

$$\frac{\partial \mathsf{NSW}(r_0)}{\partial r} = \frac{\partial V^t_{mk}(p^{t*}_{mk}; r_0)}{\partial r}. \tag{2.88}$$

In order to state the regularity condition, view every constraint of problem (2.74) (i.e., (2.74b) as well as $\sum_t p^t_{mk} = E_{mk}$, $p^t_{mk} \geq p^{\min}_{mk}$, $p^t_{mk} \leq p^{\max}_{mk}$, $s^t \geq 0$, $s^t \leq s^{\max}$) as function of $\mathbf{s}$ and $\mathbf{p}$, and consider the gradient vector of each of those. The regularity condition requires that the gradient vectors corresponding to active constraints (i.e., constraints holding as equality) are linearly independent [58, CQ3, p. 24]. It can be shown to hold in the present setup, if $0 < s^{t*} < s^{\max}$ for all t, and if for every class 1 device there is a t such that $p^{\min}_{mk} < p^{t*}_{mk} < p^{\max}_{mk}$, which are mild conditions.

Consider now a piecewise linear $C^t(s^t)$. Specifically, suppose that $C^t(s^t) = \max_{i=1\ldots,N}\{\kappa_i s^t + b_i\}$ (cf. Fig.2.1b). To analyze the sensitivities with respect to κ_i and b_i, a standard linear programming trick is applied. Auxiliary variables u^t, $t = 1, \ldots, T$ are introduced, so that the cost $\sum_t C^t(s^t)$ is replaced by $\sum_t u^t$, while the following constraints are added to problem (2.74):

$$\kappa_i s^t + b_i \leq u^t, \quad i = 1, \ldots, N,\ t = 1, \ldots, T. \tag{2.89}$$

Let ζ^t_i be Lagrange multipliers corresponding to (2.89). Under a regularity condition, the derivatives of the objective value at nominal values κ_{i0} and b_{i0} are

$$\frac{\partial \mathsf{NSW}(\kappa_{i0})}{\partial \kappa_i} = \sum_{t=1}^{T} \zeta^{t*}_i s^{t*}; \quad \frac{\partial \mathsf{NSW}(b_{i0})}{\partial b_i} = \sum_{t=1}^{T} \zeta^{t*}_i. \tag{2.90}$$

The regularity condition for this case is the Mangasarian-Fromovitz constraint qualification [58, CQ1, p. 23], which holds under the same conditions as in the differentiable case.

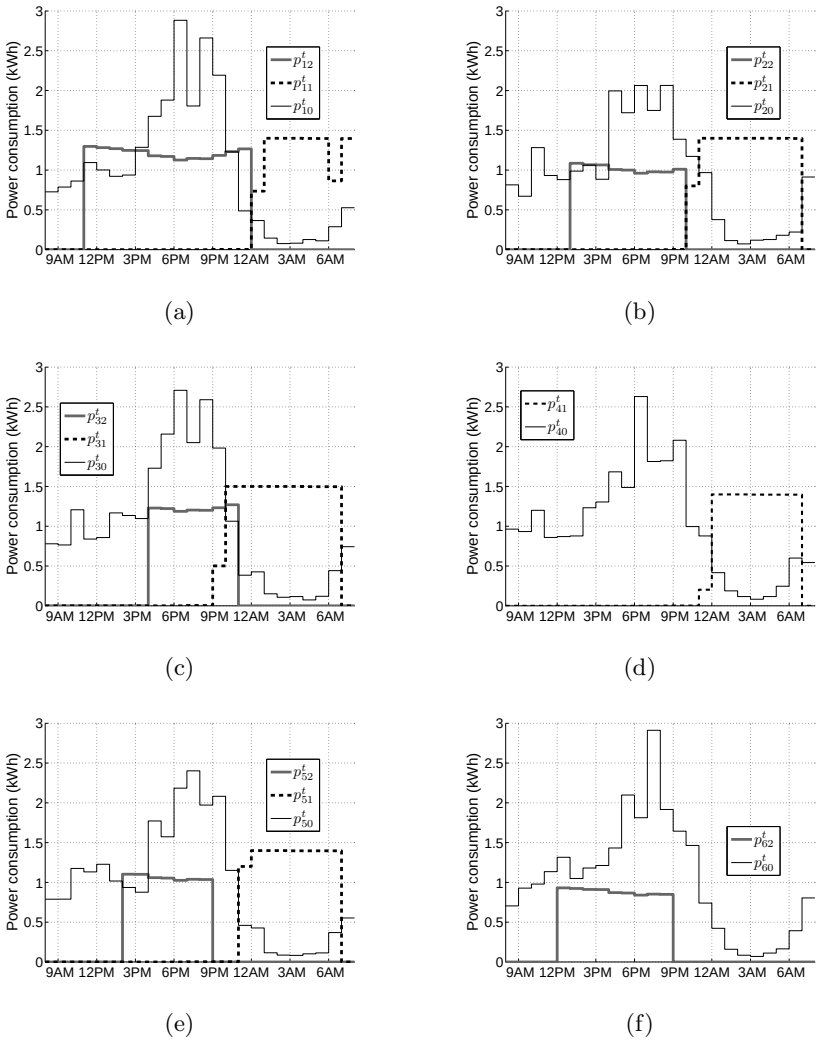


Figure 2.4: Schedule for the 6 end-users.

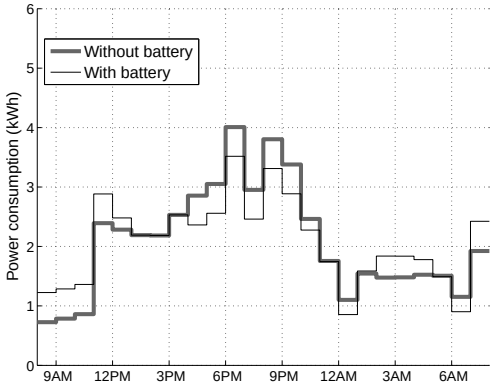


Figure 2.5: Power consumption of end-user 1 with and without battery.

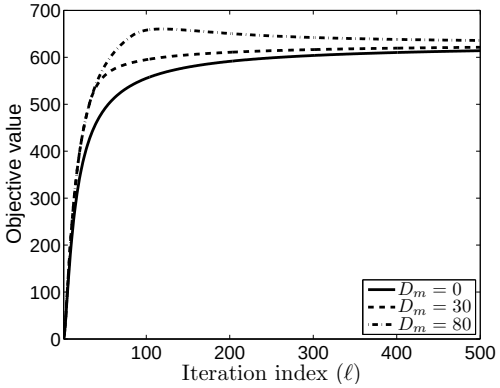


Figure 2.6: Objective value evaluated at running averages, $\sum_{t=1}^{T} C^t(\bar{s}^t(\ell)) + \sum_{t=1}^{T}\sum_{m=1}^{M}\sum_{k\in\mathcal{K}_m^{(2)}} V_{mk}^t(\bar{p}_{mk}^t(\ell))$, for different delays.

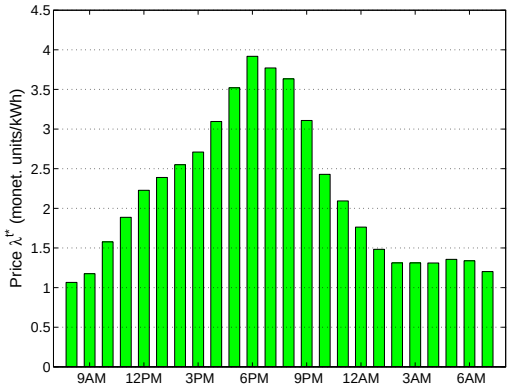


Figure 2.7: Optimal Lagrange multipliers.

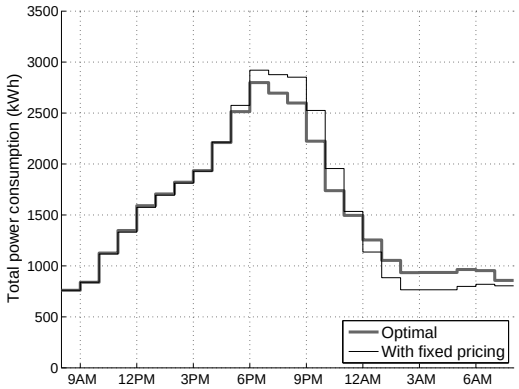


Figure 2.8: Total power consumption elicited from the optimal load control scheme and a fixed pricing scheme.

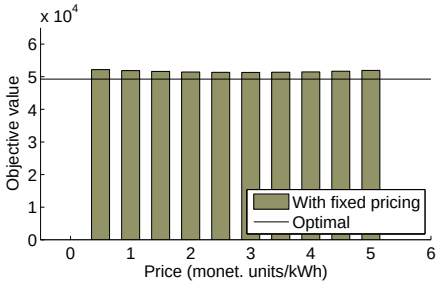


Figure 2.9: Objective value (opposite of social welfare) for the optimal load control scheme and a fixed pricing scheme.

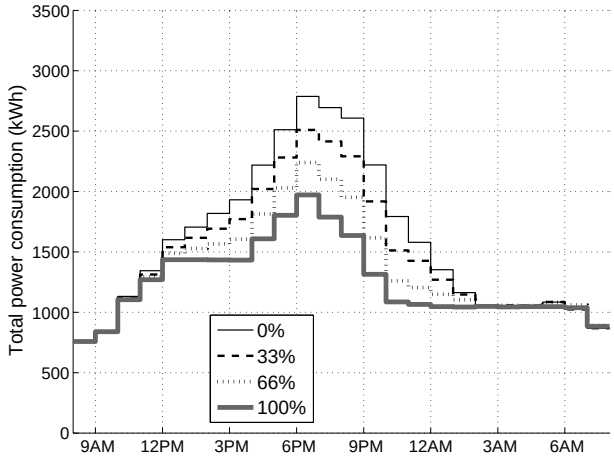


Figure 2.10: Total power consumption when different percentages of the end-users adopt zero disutility ($V_k^t \equiv 0$).

Chapter 3

Energy Consumption Scheduling

Section 3.1 establishes that certain nonconvex programs with special structure have zero duality gap, and connections to related results in the literature are made. The results of Section 3.1 are applied to a nonconvex energy consumption scheduling problem over a continuous time interval in Section 3.2. Additional duality results for a discretized version of the problem are established in Section 3.3. A solution algorithm alongside numerical results are presented in Section 3.4.

The material in this chapter draws from [36, 60].

3.1 Strong Duality in Nonconvex Programs with Special Structure

Consider a probability space $(\Omega, \mathcal{D}, P_{\boldsymbol{\omega}})$, where Ω is a subset of a Euclidean space and $\boldsymbol{\omega}$ is a random vector. The set $\mathcal{D}$ is the σ-field of Borel sets in Ω and $P_{\boldsymbol{\omega}}$ the distribution of $\boldsymbol{\omega}$, which is a probability measure on $\mathcal{D}$. The choice of the Ω will depend on the application, as will be seen in the present chapter and the next chapter.

The optimization problem of interest entails two types of variables: the vector $\mathbf{x}$ constrained to a subset $\mathcal{X}$ of the Euclidean space $\mathbb{R}^{d_{\mathbf{x}}}$; and the function $\mathbf{p} : \Omega \to \mathbb{R}^{d_{\mathbf{p}}}$ belonging to an appropriate set of functions $\mathcal{P}$. Henceforth, the function variable will be denoted by $\mathbf{p}$ instead of $\mathbf{p}(\boldsymbol{\omega})$, for brevity. For any $\boldsymbol{\omega} \in \Omega$, let $\Pi(\boldsymbol{\omega})$ be a subset of

$\mathbb{R}^{d_{\mathbf{p}}}$. Then $\mathcal{P}$ is defined as the set of functions taking values in $\Pi(\boldsymbol{\omega})$.

$$\mathcal{P} := \{\mathbf{p} : \Omega \to \mathbb{R}^{d_{\mathbf{p}}} \text{ measurable} \,|\, \mathbf{p}(\boldsymbol{\omega}) \in \Pi(\boldsymbol{\omega}) \text{ for almost all } \boldsymbol{\omega} \in \Omega\}. \tag{3.1}$$

Consider the following optimization problem:

$$\begin{aligned} f^* = \quad & \min_{\mathbf{x},\mathbf{p}} \quad f_1(\mathbf{x}) + \mathbb{E}[f_2(\mathbf{p}(\boldsymbol{\omega}), \boldsymbol{\omega})] && \text{(3.2a)} \\ & \text{subj. to} \quad \mathbf{g}_1(\mathbf{x}) + \mathbb{E}[\mathbf{g}_2(\mathbf{p}(\boldsymbol{\omega}), \boldsymbol{\omega})] \leq \mathbf{0} && \text{(3.2b)} \\ & \qquad\quad \mathbf{x} \in \mathcal{X},\ \mathbf{p} \in \mathcal{P}. && \text{(3.2c)} \end{aligned}$$

where $\mathbf{g}_1$ and $\mathbf{g}_2$ are $\mathbb{R}^K$-valued functions describing K constraints.

The following conditions regarding (3.2) are supposed to hold:

C3.1. Constraint set $\mathcal{X}$ is convex, closed, bounded, and in the interior of the domains of functions $f_1(\mathbf{x})$ and $\mathbf{g}_1(\mathbf{x})$. Set $\Pi(\boldsymbol{\omega})$ is closed, bounded, and in the interior of the domains of functions $f_2(., \boldsymbol{\omega})$ and $\mathbf{g}_2(., \boldsymbol{\omega})$ for all $\boldsymbol{\omega}$.

C3.2. Functions f_1 and $\mathbf{g}_1$ are convex (entrywise). Functions $f_2(\mathbf{p}(\boldsymbol{\omega}), \boldsymbol{\omega})$ and $\mathbf{g}_2(\mathbf{p}(\boldsymbol{\omega}), \boldsymbol{\omega})$ are integrable whenever $\mathbf{p}$ is measurable. Furthermore, there is a $G > 0$ such that for all $\mathbf{x} \in \mathcal{X}$ and $\mathbf{p} \in \mathcal{P}$,

$$\|\mathbf{g}_1(\mathbf{x}) + \mathbb{E}[\mathbf{g}_2(\mathbf{p}(\boldsymbol{\omega}), \boldsymbol{\omega})]\| \leq G. \tag{3.3}$$

C3.3. Random vector $\boldsymbol{\omega}$ is continuous;[1] and

C3.4. There exist $\mathbf{x}' \in \mathcal{X}$ and $\mathbf{p}' \in \mathcal{P}$ such that (3.2b) holds as strict inequality (Slater constraint qualification).

While (3.2) is not convex in general, it is separable. Let $\boldsymbol{\lambda}$ denote a Lagrange multiplier vector corresponding to constraints (3.2b). The Lagrangian (keeping constraints (3.2c) implicit) and the dual function are, respectively

$$L(\mathbf{x}, \mathbf{p}, \boldsymbol{\lambda}) = f_1(\mathbf{x}) + \mathbb{E}[f_2(\mathbf{p}(\boldsymbol{\omega}), \boldsymbol{\omega})] + \boldsymbol{\lambda}^T \Big(\mathbf{g}_1(\mathbf{x}) + \mathbb{E}[\mathbf{g}_2(\mathbf{p}(\boldsymbol{\omega}), \boldsymbol{\omega})] \Big) \tag{3.4}$$

$$\rho(\boldsymbol{\lambda}) := \min_{\mathbf{x} \in \mathcal{X},\, \mathbf{p} \in \mathcal{P}} L(\mathbf{x}, \mathbf{p}, \boldsymbol{\lambda}). \tag{3.5}$$

[1] Formally, this is equivalent to saying that $P_{\boldsymbol{\omega}}$ is absolutely continuous with respect to the Lebesgue measure on $\mathbb{R}^{d_{\mathbf{p}}}$. In more practical terms, it means that $\boldsymbol{\omega}$ has a probability density function without deltas.

Conditions C3.1 and C3.2 ensure that the domain of $\rho(\boldsymbol{\lambda})$ is $\mathbb{R}^K$. The dual problem becomes

$$\rho^* := \min_{\boldsymbol{\lambda} \geq \mathbf{0}} \rho(\boldsymbol{\lambda}). \tag{3.6}$$

Define next the vector function

$$\mathbf{v}(\mathbf{p}(\boldsymbol{\omega}), \boldsymbol{\omega}) = \begin{bmatrix} f_2(\mathbf{p}(\boldsymbol{\omega}), \boldsymbol{\omega}) \\ \mathbf{g}_2(\mathbf{p}(\boldsymbol{\omega}), \boldsymbol{\omega}) \end{bmatrix}. \tag{3.7}$$

As $\mathbf{p}$ varies in $\mathcal{P}$, define the range of $\mathbb{E}[\mathbf{v}(\mathbf{p}(\boldsymbol{\omega}), \boldsymbol{\omega})]$ as

$$\mathcal{R} := \left\{ \mathbf{r} \in \mathbb{R}^K \,|\mathbf{r} = \mathbb{E}[\mathbf{v}(\mathbf{p}(\boldsymbol{\omega}), \boldsymbol{\omega})] \text{ for some } \mathbf{p} \in \mathcal{P} \right\}. \tag{3.8}$$

The following lemma demonstrating the convexity of $\mathcal{R}$ plays a central role in establishing the zero duality gap of (3.2).

Lemma 3.1. *If C3.1-C3.3 hold, then the set $\mathcal{R}$ is convex.*

The proof relies on Lyapunov's convexity theorem [33]. An extension of Lyapunov's theorem [33, Extension 1] has been applied to show zero duality gap of power control problems in DSL [34]. This extension however does not apply here, as indicated in the ensuing proof. In a related contribution [35], it is shown that the perturbation function of a wireless networking problem that is a special case of (3.2) is convex; the claim of Lemma 3.1 though is different. A result corresponding to Lemma 3.1 has been established in [36] for a problem of the form (3.2) with $f_2 \equiv 0$ and $\Pi(\boldsymbol{\omega})$ independent of ω, that is, $\Pi(\boldsymbol{\omega}) = \Pi$ for all $\boldsymbol{\omega} \in \Omega$.

Proof of Lemma 3.1. Let $\mathbf{r}_1$ and $\mathbf{r}_2$ denote arbitrary points in $\mathcal{R}$, and let $\alpha \in (0, 1)$ be arbitrary. By the definition of $\mathcal{R}$, there are functions $\mathbf{p}_1$ and $\mathbf{p}_2$ in $\mathcal{P}$ such that

$$\mathbf{r}_1 = \int \mathbf{v}(\mathbf{p}_1(\boldsymbol{\omega}), \boldsymbol{\omega}) dP_{\boldsymbol{\omega}} \text{ and } \mathbf{r}_2 = \int \mathbf{v}(\mathbf{p}_2(\boldsymbol{\omega}), \boldsymbol{\omega}) dP_{\boldsymbol{\omega}}. \tag{3.9}$$

Now define

$$\mathbf{u}(E) := \begin{bmatrix} \int_E \mathbf{v}(\mathbf{p}_1(\boldsymbol{\omega}), \boldsymbol{\omega}) dP_{\boldsymbol{\omega}} \\ \int_E \mathbf{v}(\mathbf{p}_2(\boldsymbol{\omega}), \boldsymbol{\omega}) dP_{\boldsymbol{\omega}} \end{bmatrix}, \quad E \in \mathcal{D}. \tag{3.10}$$

The set function $\mathbf{u}(E)$ is a nonatomic vector measure on $\mathcal{D}$, because $P_{\mathbf{h}}$ is nonatomic (cf. C3.3) and the functions $\mathbf{v}(\mathbf{p}_1(\boldsymbol{\omega}), \boldsymbol{\omega})$ and $\mathbf{v}(\mathbf{p}_2(\boldsymbol{\omega}), \boldsymbol{\omega})$ are integrable (cf. C3.2); see [61]

for definitions. Hence, Lyapunov's theorem applies to $\mathbf{u}(E)$; see also [33, Extension 1] and [35, Lemma 1].

Specifically, consider a null set N in $\mathcal{D}$, i.e., a set with $P_{\boldsymbol{\omega}}(N) = 0$, and the whole space $\Omega \in \mathcal{D}$. It holds that $\mathbf{u}(N) = \mathbf{0}$ and $\mathbf{u}(\Omega) = [\mathbf{r}_1^T, \mathbf{r}_2^T]^T$. For the chosen α, Lyapunov's theorem asserts that there exists a set $E_\alpha \in \mathcal{D}$ such that (E_α^c denotes the complement of E_α)

$$\mathbf{u}(E_\alpha) = \alpha \mathbf{u}(\Omega) + (1-\alpha)\mathbf{u}(N) = \alpha \begin{bmatrix} \mathbf{r}_1 \\ \\ \mathbf{r}_2 \end{bmatrix} \tag{3.11a}$$

$$\mathbf{u}(E_\alpha^c) = \mathbf{u}(\Omega) - \mathbf{u}(E_\alpha) = (1-\alpha) \begin{bmatrix} \mathbf{r}_1 \\ \\ \mathbf{r}_2 \end{bmatrix}. \tag{3.11b}$$

Now using these E_α and E_α^c, define

$$\mathbf{p}_\alpha(\boldsymbol{\omega}) = \begin{cases} \mathbf{p}_1(\boldsymbol{\omega}), & \boldsymbol{\omega} \in E_\alpha \\ \mathbf{p}_2(\boldsymbol{\omega}), & \boldsymbol{\omega} \in E_\alpha^c. \end{cases} \tag{3.12}$$

It is easy to show that $\mathbf{p}_\alpha \in \mathcal{P}$. In particular, the function $\mathbf{p}_\alpha(\boldsymbol{\omega})$ can written as $\mathbf{p}_\alpha(\boldsymbol{\omega}) = \mathbf{p}_1(\boldsymbol{\omega})\mathbf{1}_{E_\alpha} + \mathbf{p}_2(\boldsymbol{\omega})\mathbf{1}_{E_\alpha^c}$, where $\mathbf{1}_E$ is the indicator function of a set $E \in \mathcal{D}$. Hence it is measurable, as sum of measurable functions. Moreover, we have that $\mathbf{p}_\alpha(\boldsymbol{\omega}) \in \Pi(\boldsymbol{\omega})$ for almost all $\boldsymbol{\omega}$, because $\mathbf{p}_1(\boldsymbol{\omega})$ and $\mathbf{p}_2(\boldsymbol{\omega})$ satisfy this property. The need to show $\mathbf{p}_\alpha \in \mathcal{P}$ makes [33, Extension 1] not directly applicable here.

Thus, $\mathbf{p}_\alpha(\boldsymbol{\omega}) \in \mathcal{P}$ and satisfies [cf. (3.11)]

$$\begin{aligned} \int \mathbf{v}(\mathbf{p}_\alpha(\boldsymbol{\omega}), \boldsymbol{\omega}) dP_{\boldsymbol{\omega}} &= \int_{E_\alpha} \mathbf{v}(\mathbf{p}_1(\boldsymbol{\omega}), \boldsymbol{\omega}) dP_{\boldsymbol{\omega}} + \int_{E_\alpha^c} \mathbf{v}(\mathbf{p}_2(\boldsymbol{\omega}), \boldsymbol{\omega}) dP_{\boldsymbol{\omega}} \\ &= \alpha \mathbf{r}_1 + (1-\alpha)\mathbf{r}_2. \end{aligned} \tag{3.13}$$

Therefore, $\alpha \mathbf{r}_1 + (1-\alpha)\mathbf{r}_2 \in \mathcal{R}$.

Finally, the zero duality gap result follows from Lemma 3.1 and is stated in the following proposition.

Proposition 3.1. *If C3.1–C3.4 hold, then problem* (3.2) *has zero duality gap, i.e., $f^* = \rho^*$. Furthermore, the values f^* and ρ^* are finite, the dual problem* (3.6) *has an optimal solution, and the set of optimal solutions of* (3.6) *is bounded.*

Proof. Conditions C3.1 and C3.2 imply that the optimal primal value f^* is finite. Consider the set

$$\mathcal{W} := \big\{(w_1, \ldots, w_K, z) \in \mathbb{R}^{K+1} \;\mid\; f_1(\mathbf{x}) + \mathbb{E}[f_2(\mathbf{p}(\boldsymbol{\omega}), \boldsymbol{\omega})] \leq z,$$

$$\mathbf{g}_1(\mathbf{x}) + \mathbb{E}[\mathbf{g}_2(\mathbf{p}(\boldsymbol{\omega}), \boldsymbol{\omega})] \leq \mathbf{w} \text{ for some } \mathbf{x} \in \mathcal{B}_{\mathbf{y}},\ \mathbf{p} \in \mathcal{P}\}\,. \tag{3.14}$$

Using Lemma 3.1, it is easy to verify that set $\mathcal{W}$ is convex. The rest of the proof follows that of [46, Prop. 5.3.1 and 5.1.4], using the finiteness of P and Slater constraint qualification (cf. C3.4).

The boundedness of the optimal dual set is a standard result for convex problems under Slater constraint qualification and finiteness of optimal primal value; see e.g., [47, Prop. 6.4.3] or [22, p. 1762]. The proof holds also in the present setup since f^* is finite, $f^* = \rho^*$, and C3.4 holds.

The results of the present section are applied to a nonconvex energy consumption scheduling problem for a single residential end user in the ensuing section.

3.2 Continuous-Time Demand Response Formulation

Consider a residential end user with a smart meter communicating with various devices at the residence. The smart meter also communicates with the utility company through the AMI. The utility company has announced the prices for the scheduling horizon (e.g., next day) ahead of time. Let $[0, T]$ denote the scheduling horizon, and $C(s(t), t)$ the cost charged to the end user at time t, where $s(t)$ is the total residential power consumption. Note that the cost is time-varying, and hence explicitly incorporates the real-time pricing paradigm of the future grid.

In the DR formulation, it is imperative to capture diverse end user requirements, as well as different classes of devices and tasks involved in the scheduling. Here, we consider a base residential load that is not to be shifted, and two classes of interruptible devices, denoted respectively as $\mathcal{K}_1$ and $\mathcal{K}_2$. The base load is denoted by $p_0(t)$, and can be lights or computers. Moreover, let $p_k(t)$ be the power consumption of device k, where $k \in \mathcal{K}_1$ or $k \in \mathcal{K}_2$ (by convention, $0 \notin \mathcal{K}_1$ and $0 \notin \mathcal{K}_2$). The particular characteristics of those classes are as follows:

1. Class $\mathcal{K}_1$ contains devices with a prescribed energy requirement $\check{E}_k$ that has to be completed over a duration $\check{T}_k$, starting from time $\check{\alpha}_k$. The power consumed is constrained to $\{0\} \cup [\check{p}_{k\,\min}, \check{p}_{k\,\max}]$ with $\check{p}_{k\,\min} > 0$ capturing the situation that a device cannot operate at arbitrarily low power. Examples are operating a pool

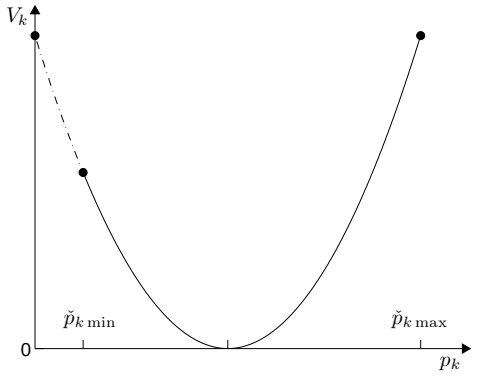


Figure 3.1: Example of a convex disutility function. The point where the function achieves is minimum can be a desirable set-point by the end user.

pump or charging a PHEV. The energy constraint can be written as

$$\frac{1}{T}\int_{\check{\alpha}_k}^{\check{\alpha}_k+\check{T}_k} p_k(t)dt = \check{E}_k,\ k \in \mathcal{K}_1. \tag{3.15}$$

where due to the normalization factor $1/T$, the quantities $p_k(t)$ and $\check{E}_k$ have units of power.

2. Class $\mathcal{K}_2$ includes devices operating with power in $\{0\} \cup [\check{p}_{k\,\min}, \check{p}_{k\,\max}]$. There is not total energy requirement, but instead, a disutility function $V_k(p_k(t), t)$ that is introduced to capture dissatisfaction of the end user for operating away from a nominal point, as in Fig. 3.1 for example. The premise is that the end user may choose to operate away from a nominal point, if this can reduce the electricity bill, as will be seen shortly in the optimization formulation. An example in this class is an A/C unit. As with class $\mathcal{K}_1$, it is possible to define an interval $[\check{\alpha}_k, \check{\alpha}_k + \check{T}_k]$ over which the device is to be operated. In this case, we have that $V_k(p_k(t), t) = 0$ for $t < \check{\alpha}_k$ or $t \geq \check{\alpha}_k + \check{T}_k$.

The interruptible nature is apparent due to the constraint set $\{0\} \cup [\check{p}_{k\,\min}, \check{p}_{k\,\max}]$ for the instantaneous power consumption. The case of uninterruptible tasks can be accommodated by considering a constraint set $[\check{p}_{k\,\min}, \check{p}_{k\,\max}]$ with $\check{p}_{k\,\min} > 0$ over the desired hours of operation.

Let $\mathcal{K}$ denote the set of all adjustable devices ($\mathcal{K}_1 \cup \mathcal{K}_2$) and the base load. The *continuous-time* residential DR scheduling amounts to solving the following optimization

problem:

$$\mathsf{P} = \min \quad \frac{1}{T}\int_0^T C\left(\sum_{k\in\mathcal{K}} p_k(t), t\right) dt + \frac{1}{T}\int_0^T \sum_{k\in\mathcal{K}_2} V_k(p_k(t), t)dt \tag{3.16a}$$

$$\text{subj. to} \quad \frac{1}{T}\int_0^T p_k(t)dt \geq \check{E}_k,\ k \in \mathcal{K}_1 \tag{3.16b}$$

$$p_k(t) \in \{0\} \cup [\check{p}_{k\,\min}, \check{p}_{k\,\max}],\ t \in [\check{\alpha}_k, \check{\alpha}_k + \check{T}_k],\ k \in \mathcal{K}_1 \cup \mathcal{K}_2 \tag{3.16c}$$

$$p_k(t) = 0,\ t \notin [\check{\alpha}_k, \check{\alpha}_k + \check{T}_k],\ k \in \mathcal{K}_1 \cup \mathcal{K}_2 \tag{3.16d}$$

$$\text{variables} \quad p_k(t)\ (0 \leq t \leq T,\ k \in \mathcal{K}_1 \cup \mathcal{K}_2).$$

The optimization variables in (3.16) are the power consumptions of the adjustable devices, and are functions of time over the interval $[0, T]$, as indicated in the last line of formulation (3.16). As such, they are infinite-dimensional. Note that (3.15) has been relaxed to inequality in (3.16b)—this is valid in practice under Condition C3.5, which will be given shortly. The factor $(1/T)$ multiplying all integrals in (3.16) is only for symmetry with the discrete-time case, as will be seen later. Regarding units, $p_k(t)$ and $\check{E}_k$ are in Watts.

Problem (3.16) is nonconvex due to constraint (3.16c). The following assumptions are made:

C3.5. Functions $C\left(\sum_k p_k, t\right)$ and $V_k(p_k, t)$ with respect to their first argument are defined, continuous, and convex, over $\times_{k\in\mathcal{K}_1\cup\mathcal{K}_2}[0, p_{k\,\max}]$ for all $t \in [0, T]$. The former is also strictly increasing in its first argument. Moreover, functions $C\left(\sum_k p_k(t), t\right)$ and $V_k(p_k(t), t)$ are integrable whenever function $p_k(t)$ is (Borel) measurable. Finally, for constraint (3.16b) to be practically meaningful, the condition $\check{p}_{k\,\min}\check{T}_k < \check{E}_k T$ is supposed to hold.

C3.6. There exist $\{p_k(t)\}$, $k \in \mathcal{K}_1$ satisfying (3.16c) so that (3.16b) holds as strict inequality.

Both conditions are mild and are expected to hold in DR setups. The fact that $C(s, t)$ is strictly increasing in s together with the condition $\check{p}_{k\,\min}\check{T}_k < \check{E}_k T$ imply that (3.16b) holds as *equality* at the optimal point. The inequality $\check{p}_{k\,\min}\check{T}_k < \check{E}_k T$ essentially means that operating the device $k \in \mathcal{K}_1$ at its lowest power level over the specified interval is not enough in order to provide the required energy. Condition C3.6 corresponds to the standard Slater constraint qualification.

Taking a Lagrangian dual approach, let λ_k denote the Lagrange multiplier corresponding to (3.16b); $\boldsymbol{\lambda}$ be a vector collecting all Lagrange multipliers; and $\boldsymbol{p}(t)$ be a vector collecting all power consumptions $p_k(t)$, $k \in \mathcal{K}_1 \cup \mathcal{K}_2$. Let $\Pi(t)$ denote the $(|\mathcal{K}_1| + |\mathcal{K}_2|)$-dimensional region representing the instantaneous constraints (3.16c) and (3.16d). This region is a function of t because for each k, the constraint set may change depending on whether $t \in [\check{\alpha}_k, \check{\alpha}_k + \check{T}_k]$. Keeping those constraints implicit, the Lagrangian function corresponding to (3.16) takes the following form, after straightforward arrangements:

$$L(\boldsymbol{\lambda}, \boldsymbol{p}(t)) = \frac{1}{T} \int_0^T \left(C\left(\sum_{k \in \mathcal{K}} p_k(t), t \right) + \sum_{k \in \mathcal{K}_2} V_k(p_k(t), t) - \sum_{k \in \mathcal{K}_1} \lambda_k p_k(t) \right) dt + \sum_{k \in \mathcal{K}_1} \lambda_k \check{E}_k. \quad (3.17)$$

The dual function and the dual problem are, respectively,

$$q(\boldsymbol{\lambda}) = \min_{\boldsymbol{p}(t) \in \Pi(t)} L(\boldsymbol{\lambda}, \boldsymbol{p}(t)) \quad (3.18)$$

$$\mathsf{D} = \max_{\boldsymbol{\lambda} \geq \mathbf{0}} q(\boldsymbol{\lambda}). \quad (3.19)$$

Weak duality implies that $\mathsf{D} \leq \mathsf{P}$. Despite nonconvexity, the following result asserts that there is no duality gap.

Proposition 3.2. *Strong duality holds under conditions C3.5 and C3.6, i.e.,*

$$\mathsf{D} = \mathsf{P}. \quad (3.20)$$

Intuitively, continuous-time averaging smooths out the non-convexity, and thus eliminates the duality gap induced by (3.16c). The proof relies on the results of Section 3.1.

Proof of Proposition 3.2. Problem (3.16) has the generic form of (3.2) where $\boldsymbol{\omega}$ is t and $\Omega = [0, T]$ here, and the distribution $P_{\boldsymbol{\omega}}$ is the uniform distribution over the interval $[0, T]$. Conditions C3.5 and C3.6 imply conditions C3.1–C3.4. □

Proposition 3.2 implies that solving the dual problem is optimal. Two questions arise naturally: (a) how the dual problem can be solved efficiently, and (b) how the solution of the dual problem can be leveraged in order to give a solution of the primal

problem. These questions are addressed in Section 3.4. To this end, it will be helpful to examine the duality properties of a discretized version of (3.16), the subject of the ensuing section.

3.3 Discrete-Time Demand Response Formulation

Consider a partition of $[0,T]$ into N intervals, $\{0, \frac{T}{N}, \frac{2T}{N}, \ldots, T\}$. Let $p_k^n := p_k[(n-1)T/N]$, $C^n(.) := C[.,(n-1)T/N]$, and $V_k^n(.) := V_k[.,(n-1)T/N]$, for $n = 1, \ldots, N$. It is supposed that the partition is fine enough so that the time instants $\check{\alpha}_k$ and $\check{\alpha}_k + \check{T}_k$ are members of the partition. The *discrete-time* version of (3.16) is formulated by replacing the integrals with left Riemann sums, and takes the following form

$$\mathsf{P}_N = \min \quad \frac{1}{N}\sum_{n=1}^{N} C^n\left(\sum_k p_k^n\right) + \frac{1}{N}\sum_{n=1}^{N}\sum_{k\in\mathcal{K}_2} V_k^n(p_k^n) \tag{3.21a}$$

$$\text{subj. to} \quad \frac{1}{N}\sum_{n=1}^{N} p_k^n \geq \check{E}_k,\ k \in \mathcal{K}_1 \tag{3.21b}$$

$$p_k^n \in \{0\} \cup [\check{p}_{k\,\min}, \check{p}_{k\,\max}],\ k \in \mathcal{K}_1 \cup \mathcal{K}_2$$

$$n \in \frac{N}{T}\cdot\{\check{\alpha}_k, \ldots, \check{\alpha}_k + \check{T}_k - 1\} \tag{3.21c}$$

$$p_k^n = 0,\ k \in \mathcal{K}_1 \cup \mathcal{K}_2, n \notin \frac{N}{T}\cdot\{\check{\alpha}_k, \ldots, \check{\alpha}_k + \check{T}_k - 1\} \tag{3.21d}$$

$$\text{variables} \quad p_k^n\ (n \in \{1,\ldots,N\},\ k \in \mathcal{K}_1 \cup \mathcal{K}_2).$$

Problem (3.21) is nonconvex in general, due to constraint (3.21c). Nevertheless, it is a mixed integer program with special separable structure [62, Sec. 5.6.1]. The dual problem can be formulated analogously to the continuous-time case by introducing a Lagrange multiplier for constraint (3.21b), and keeping the rest of the constraints implicit. Let D_N denote the optimal value of this dual problem. The following duality gap estimate asserts that the duality gap vanishes as the partition size increases.

Proposition 3.3. *For the duality gap of* (3.21)*, under conditions C3.5 and C3.6, it holds that*

$$0 \leq \mathsf{P}_N - \mathsf{D}_N = O\left(\frac{1}{N}\right). \tag{3.22}$$

Proof of Proposition 3.3. Problem (3.21) can be written in the form of the separable program of [62, p. 371] with the correspondences listed in Table 3.1, where for $k \in \mathcal{K}_1 \cup \mathcal{K}_2$

Table 3.1: Correspondences between a prototype separable program and the discrete-time DR problem (3.21)

Separable program	Discrete-time DR problem
I	N
i	n
p_i	$\lvert\mathcal{K}_1 \cup \mathcal{K}_2\rvert$
x_i	$\{p_k^n\}_{k\in\mathcal{K}_1\cup\mathcal{K}_2}$
X_i	$\times_{k\in\mathcal{K}_1\cup\mathcal{K}_2}\mathcal{P}_k^n$
$f_i(x_i)$	$C^n(\sum_k p_k^n) + \sum_{k\in\mathcal{K}_2} V_k^n(p_k^n)$
m	$\lvert\mathcal{K}_1\rvert$
$h_i(x_i)$	$\{h_k^n(p_k^n)\}_{k\in\mathcal{K}_1}$, where $h_k^n(p_k^n) = \begin{cases} -\frac{p_k^n}{N}, & n \in \frac{N}{T}\cdot\{\check{\alpha}_k,\ldots,\check{\alpha}_k+\check{T}_k-1\} \\ 0, & \text{otherwise} \end{cases}$
b	$\{-\check{E}_k\}_{k\in\mathcal{K}_1}$

[cf. the instantaneous power constraints (3.21c) and (3.21d)]

$$\mathcal{P}_k^n = \begin{cases} \{0\} \cup [\check{p}_{k\min}, \check{p}_{k\max}], \\ \qquad n \in \frac{N}{T}\cdot\{\check{\alpha}_k,\ldots,\check{\alpha}_k+\check{T}_k-1\} \\ \{0\}, \quad n \notin \frac{N}{T}\cdot\{\check{\alpha}_k,\ldots,\check{\alpha}_k+\check{T}_k-1\}. \end{cases} \tag{3.23}$$

The main idea is that the desired duality gap estimate (3.22) follows from the related estimate of [62, p. 376]. In order to use the aforementioned result, the following conditions need to be verified for the problem at hand:

1. The functions $f_i(x_i)$ and $h_i(x_i)$ need to be defined over the convex hull of X_i. In the present case, this condition holds for the objective function $C^n(\sum_k p_k^n) + \sum_{k\in\mathcal{K}_2} V_k^n(p_k^n)$ due to C3.5, while it is obviously true for the constraint function $-\frac{p_k^n}{N}$.
2. Assumptions (A1)–(A3) in [62, Sec. 5.6.1] should hold. In particular, problem (3.21) has a solution and thus (A1) holds, because the objective is continuous and the feasible set is nonemtpy and compact. It is also clear that (A2) holds due to the continuity in condition C3.5 and the fact that the instantaneous power consumption constraints are a compact set. Finally, it is not difficult to verify

$$\begin{array}{ccccc} \mathsf{P} & = & \mathsf{D} & & \\ {\scriptstyle ?}\uparrow & & \uparrow{\scriptstyle ?} & & \\ 0 \leq \mathsf{P}_N & - & \mathsf{D}_N & \to 0 & \end{array}$$

Figure 3.2: Primal and dual values of continuous- and discrete-time DR problems.

that for each n, the vector that satisfies (A3) has entries $k \in \mathcal{K}_1 \cup \mathcal{K}_2$ given by

$$\begin{cases} \max\limits_{k \in \mathcal{K}_1 \cup \mathcal{K}_2} \check{p}_{k\max}, & \text{if } n \in \frac{N}{T} \cdot \{\check{\alpha}_k, \ldots, \check{\alpha}_k + \check{T}_k - 1\} \\ 0, & \text{otherwise.} \end{cases}$$

3. Finally, it has to be verified that the quantities $\sup\{f_i(\mathbf{x}_i)|\mathbf{x}_i \in X_i\}$ and $\inf\{f_i(\mathbf{x}_i)|\mathbf{x}_i \in X_i\}$ are respectively upper and lower bounded uniformly as I varies. In other words, the aforementioned quantities need to have bounds that do not depend on I (the length of discretization). In the present case, $\max_{i=1,\ldots,I} \sup\{f_i(\mathbf{x}_i)|\mathbf{x}_i \in X_i\}$ is upper bounded by

$$\sup_{t \in [0,T]} \left\{ C\left(\check{p}_{k\max}, t\right) + \sum_{k \in \mathcal{K}_2} V_k(\check{p}_{k\max}, t) \right\}$$

which does not depend on N. Replacing sup by inf in the previous expression gives a lower bound for $\min_{i=1,\ldots,I} \inf\{f_i(\mathbf{x}_i)|\mathbf{x}_i \in X_i\}$.

□

Proposition 3.3 asserts that the duality gap of (3.21) vanishes as $N \to \infty$. It is worth noting that Proposition 2 does not imply that P_N converges to the optimal value P of (3.16) as $N \to \infty$; see also Fig. 3.2. This result would be useful in approximating the continuous-time problem with the discrete-time one. Nevertheless, the results of Proposition 3.2 and 3.3 illustrated in Fig. 3.2 point to this direction. Related results developed for hard nonconvex programs or other continuous-time formulations may be useful to this end; see e.g., [63, 64] and references therein.

Motivated by the duality gap results of Propositions 3.2 and 3.3, the next section develops an algorithm to solve (3.19), which relies on the discretized problem (3.21).

3.4 Solution Technique

The subgradient method will be employed to solve the dual problem (3.19) in Subsection 3.4.1. The motivation is that solving the dual problem is optimal due to Proposition 3.2. Numerical tests are presented in Subsection 3.4.2.

3.4.1 Subgradient Algorithm

The method iterates between two steps, namely (a) Lagrangian minimization, in the same fashion as (3.18); and (ii) Lagrange multiplier update.

Let ℓ be the iteration index, and $\boldsymbol{\lambda}(0) \geq \mathbf{0}$ be the initial Lagrange multiplier vector. The Lagrangian minimization step takes the form

$$\boldsymbol{p}^{\dagger}(t;\ell) \in \arg \min_{\boldsymbol{p}(t)\in\Pi(t)} L(\boldsymbol{\lambda}(\ell), \boldsymbol{p}(t)) \tag{3.24}$$

The Lagrange multiplier $\boldsymbol{\lambda}(\ell)$ is constant for the minimization in (3.24). Note also that a function of time is sought at each iteration.

The following lemma asserts that the minimization decomposes into minimizations per time instant t.

Lemma 3.2. *The value $\boldsymbol{p}^{\dagger}(t;\ell)$ can be obtained as the solution of the following optimization problem at $t \in [0,T]$*

$$\begin{aligned}
\min \quad & C\left(\sum_{k\in\mathcal{K}} p_k, t\right) + \sum_{k\in\mathcal{K}_2} V_k(p_k,t) - \sum_{k\in\mathcal{K}_1} \lambda_k(\ell) p_k && \text{(3.25a)}\\
\text{subj. to} \quad & p_k \in \{0\} \cup [\check{p}_{k\,\min}, \check{p}_{k\,\max}],\ t \in [\check{\alpha}_k, \check{\alpha}_k + \check{T}_k],\ k \in \mathcal{K}_1 \cup \mathcal{K}_2 && \text{(3.25b)}\\
& p_k = 0,\ t \notin [\check{\alpha}_k, \check{\alpha}_k + \check{T}_k],\ k \in \mathcal{K}_1 \cup \mathcal{K}_2 && \text{(3.25c)}\\
\text{variables} \quad & p_k \quad (k \in \mathcal{K}_1 \cup \mathcal{K}_2).
\end{aligned}$$

The key idea to note is that the optimization variable in (3.25) is a *vector*, as opposed to a function in (3.24). Lemma 3.2 relies on minimizing the integrand in the Lagrangian function per t, which is reminiscent of the procedure to obtain the well-known waterfilling in wireless communications [65, Ch. 4]. The proof of Lemma 1 follows next.

Proof of Lemma 3.2. Let $t \in [0,T]$ be arbitrary, and let $\boldsymbol{p}^\dagger(t;\ell)$ be the solution of (3.25) given $\boldsymbol{\lambda}(\ell)$ for the particular t. Moreover, let $\boldsymbol{p}(t) \in \Pi(t)$ be arbitrary. It holds by definition that

$$C\left(\sum_{k\in\mathcal{K}} p_k^\dagger(t;\ell),t\right) + \sum_{k\in\mathcal{K}_2} V_k(p_k^\dagger(t;\ell),t) - \sum_{k\in\mathcal{K}_1} \lambda_k(\ell) p_k^\dagger(t;\ell)$$
$$\leq C\left(\sum_{k\in\mathcal{K}} p_k(t),t\right) + \sum_{k\in\mathcal{K}_2} V_k(p_k(t),t) - \sum_{k\in\mathcal{K}_1} \lambda_k(\ell) p_k(t) \quad (3.26)$$

Integrating the latter over $[0,T]$ yields

$$L(\boldsymbol{\lambda}(\ell),\boldsymbol{p}^\dagger(t;\ell)) \leq L(\boldsymbol{\lambda}(\ell),\boldsymbol{p}(t;\ell)) \quad (3.27)$$

Therefore, $\boldsymbol{p}^\dagger(t;\ell)$ is the function minimizing the Lagrangian in (3.24). □

It is difficult to perform the minimization for all $t \in [0,T]$ in practice, and we will shortly attempt to bypass this issue by performing the minimization over a partition of $[0,T]$.

Regarding solvers of (3.25), note that if the function $C(.,t)$ is linear, then the minimization also decomposes into *per device* problems, which are straightforward to solve in closed form.

For more general costs, problem (3.25) becomes a convex optimization problem if the on/off status of devices ($p_k = 0$ or $p_k \in [\check{p}_{k\,\min}, \check{p}_{k\,\max}]$) per device is fixed. The number of on/off combinations is in general $2^{|\mathcal{K}_1|+|\mathcal{K}_2|}$, which may not be prohibitively large. The convex problem that has to be solved given the on/off status has special structure which can further be leveraged to devise an efficient solution. Specifically, it involves box constraints and typically a combination of linear, piecewise linear, or quadratic functions as objective.

It is also possible to obtain a (nearly) closed-form solution if the cost is quadratic, $C(s,t) = as^2 + bs$ for all t, the disutility function is quadratic, $V_k(p_k,t) = w(\check{p}_{k\,\max} - p_k)^2$, and there are two interruptible devices, one per class with consumptions $p_1(t)$ and $p_2(t)$, respectively. Table 3.2 lists the related expressions, which are not difficult to derive from the optimality conditions of convex problems with box constraints [46, Example 2.1].

Having determined the primal minimizers $\boldsymbol{p}^\dagger(t;\ell)$, the Lagrange multipliers are updated as follows (β_ℓ is the iteration-dependent stepsize):

$$\lambda_k(\ell+1) = \max\left\{\lambda_k(\ell) - \beta_\ell\left(\frac{1}{T}\int_0^T p_k^\dagger(t;\ell)dt - \check{E}_k\right)\right\}. \tag{3.28}$$

In order to evaluate the integral in (3.28), the functions $p_k^\dagger(t;\ell)$ are needed. The integral can be approximated from values of $p_k^\dagger(t;\ell)$ at the partition $\{0, \frac{T}{N}, \frac{2T}{N}, \ldots, T\}$ as follows:

$$\int_0^T p_k^\dagger(t;\ell)dt \approx \frac{1}{N}\sum_{n=1}^{N} p_k^\dagger[(n-1)T/N;\ell]. \tag{3.29}$$

The approach of partitioning $[0,T]$ in order to facilitate (3.24) and (3.28) can be interpreted as application of the subgradient method to the dual of the discrete-time problem. If a nonsummable but square-summable stepsize is used, the method will converge to the optimal Lagrange multipliers for (3.21) [47, Sec. 8.2]. The final step is to recover a primal solution from the dual solution.

This is an issue that requires further research. A possible approach is to substitute the optimal Lagrange multipliers into (3.25) and minimize the Lagrangian function at a chosen partition. The power consumption schedule obtained this way will be optimal if it satisfies any of the following two conditions: (a) it is primal feasible, and satisfies the complementary slackness condition [46, Prop. 5.1.4]; or (b) it is the unique minimizer of the Lagrangian [66, p. 248]. This schedule can be considered to be piecewise constant over the continuous time-interval $[0,T]$. The challenge facing this method is that it is not straightforward to show satisfaction of either condition in the general case. Numerical tests illustrating the merits of this method are presented in the ensuing subsection.

3.4.2 Numerical Test

Scheduling of 2 devices (one from each class) over $T = 5$ hours is considered here. The power consumptions are denoted as $p_1(t)$ and $p_2(t)$. There is also a base load, $p_0(t) = 0.2\,\text{kW}$ for $0 \le t < 2$, and $p_0(t) = 0$ otherwise. The device parameters are $\check{p}_{1\,\min} = \check{p}_{2\,\min} = 0.1\,\text{kW}$, $\check{p}_{1\,\max} = \check{p}_{2\,\max} = 1\,\text{kW}$, and $\check{E}_1 = 0.3\,\text{kW}$. The cost is $C(s,t) = 0.01s^2 + 0.8s$ for all t (where s is in kW). Device 2 is constrained to be off during the first and the last hour, i.e., $\check{\alpha}_2 = 1$ and $\check{\alpha}_2 + \check{T}_2 = 4$. The disutility function is $V_2(p_2,t) = (\check{p}_{2\,\max} - p_2)^2$ for $1 \le t < 4$, and 0 otherwise. The stepsize is selected to

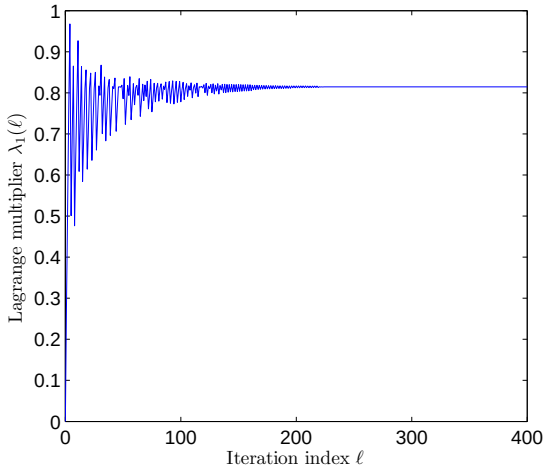


Figure 3.3: Convergence of Lagrange multiplier

be $\beta_\ell = 10/(\ell + 10)$, $\ell \geq 1$, and the partition size $N = 10$. The algorithm of Table I was used for minimizing the Lagrangian. There is a single Lagrange multiplier in this problem. Fig. 3.3 depicts the convergence of the subgradient method.

The final schedule is obtained by minimizing the Lagrangian function given the optimal Lagrange multipliers, and is depicted in Fig. 3.4. It can be observed that for both devices the instantaneous power consumption constraints are satisfied. Device 1 is off between hours 1 and 2, because the total power consumption from the base load and device 2 is already high. It is interesting to note that the area under the grey curve normalized by T has the value of $\check{E}_1 = 0.3$. Therefore, the solution is feasible, and also satisfies (3.21b) and (3.16b). The power consumption schedule in this case is optimal for (3.21), since it satisfies the necessary and sufficient optimality conditions [46, Prop. 5.1.4].

Increasing the partition size is also investigated. The results are shown in Fig. 3.5. For all partition sizes, the resulting power consumption schedule is feasible, and in fact optimal (as in the previous discussion related to Fig. 3.4). The dual value is obtained in a straightforward manner from the subgradient method. It is observed that the two values are identical even for finite N. If these are the same as P, then the behavior depicted in Fig. 3.5 indicates that there is a piecewise constant solution to (3.16) in the present case. The behavior in Fig. 3.5 is consistent with Proposition 3.3.

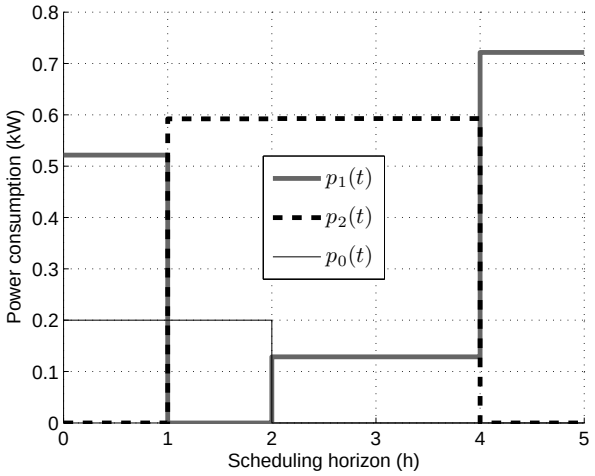


Figure 3.4: Schedule of 3 devices.

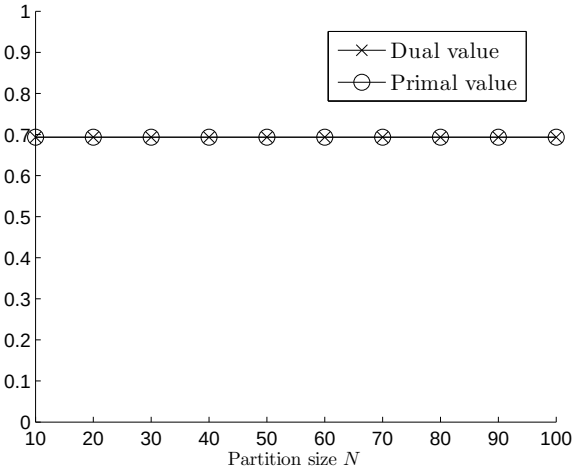


Figure 3.5: Duality gap between primal value P_N and dual value D_N.

Table 3.2: Algorithm for Lagrangian minimization with 2 devices given Lagrange multipliers.

if 1 is on and 2 is off **then**
  $p_1^\dagger = \left[\frac{\lambda - b}{2a} - p_0(t)\right]_{\check{p}_{1\min}}^{\check{p}_{1\max}}$
end if
if 2 is on and 1 is off **then**
  $p_2^\dagger = \left[\frac{2w\check{p}_{2\max} - (b + 2ap_0(t))}{2a + 2w}\right]_{\check{p}_{2\min}}^{\check{p}_{2\max}}$
end if
if 1 and 2 are on **then**

`//Case 1: Check if` $\check{p}_{1\min} < p_1^\dagger < \check{p}_{1\max}$
  $p_2 = \left[\check{p}_{2\max} - \frac{\lambda}{2w}\right]_{\check{p}_{2\min}}^{\check{p}_{2\max}}$
  $p_1 = \frac{\lambda - b}{2a} - (p_0(t) + p_2)$
  if $\check{p}_{1\min} < p_1 < \check{p}_{1\max}$ **then**
    $p_1^\dagger = p_1$; $p_2^\dagger = p_2$
  end if

`//Case 2: Check if` $p_1^\dagger = \check{p}_{1\max}$
  $p_2 = \left[\frac{2w\check{p}_{2\max} - [b + 2a(p_0(t) + \check{p}_{1\max})]}{2a + 2w}\right]_{\check{p}_{2\min}}^{\check{p}_{2\max}}$
  if $2a(p_0(t) + \check{p}_{1\max} + p_2) + b - \lambda \leq 0$ **then**
    $p_1^\dagger = \check{p}_{1\max}$; $p_2^\dagger = p_2$
  end if

`//Case 3: Check if` $p_1^\dagger = \check{p}_{1\min}$
  $p_2 = \left[\frac{2w\check{p}_{2\max} - [b + 2a(p_0(t) + \check{p}_{1\min})]}{2a + 2w}\right]_{\check{p}_{2\min}}^{\check{p}_{2\max}}$
  if $2a(p_0(t) + \check{p}_{1\min} + p_2) + b - \lambda \leq 0$ **then**
    $p_1^\dagger = \check{p}_{1\min}$; $p_2^\dagger = p_2$
  end if
end if

Chapter 4

Cross-Layer Optimization in Fading Networks

Section 4.1 develops a subgradient algorithm for the solution of problem (3.2). The challenge is to recover optimal primal variables, despite the nonconvexity. Section 4.2 presents a cross-layer optimization problem for wireless networks, and Section 4.3 applies the subgradient method of Section 4.1 to the solution of the aforementioned cross-layer optimization problem. A strategy for online network control is also presented in Section 4.3. One of the tasks required by the subgradient method is optimal power allocation at the physical layer, and Section 4.4 develops algorithms for this purpose based on successive convex approximations. Section 4.5 presents numerical tests.

The material in this chapter draws from [67, 68].

4.1 Subgradient algorithm

The present section presents a subgradient algorithm for the solution of problem (3.2), which is repeated here for convenience.

$$f^* = \min_{\mathbf{x},\mathbf{p}} \quad f_1(\mathbf{x}) + \mathbb{E}[f_2(\mathbf{p}(\boldsymbol{\omega}), \boldsymbol{\omega})] \tag{3.2a}$$

$$\text{subj. to} \quad \mathbf{g}_1(\mathbf{x}) + \mathbb{E}[\mathbf{g}_2(\mathbf{p}(\boldsymbol{\omega}), \boldsymbol{\omega})] \leq \mathbf{0} \tag{3.2b}$$

$$\mathbf{x} \in \mathcal{X},\ \mathbf{p} \in \mathcal{P}. \tag{3.2c}$$

Recall that the Lagrangian and the dual function are, respectively

$$L(\mathbf{x},\mathbf{p},\boldsymbol{\lambda}) = f_1(\mathbf{x}) + \mathbb{E}[f_2(\mathbf{p}(\boldsymbol{\omega}),\boldsymbol{\omega})] + \boldsymbol{\lambda}^T\Big(\mathbf{g}_1(\mathbf{x}) + \mathbb{E}[\mathbf{g}_2(\mathbf{p}(\boldsymbol{\omega}),\boldsymbol{\omega})]\Big) \tag{3.4}$$

$$\rho(\boldsymbol{\lambda}) = \min_{\mathbf{x}\in\mathcal{X},\,\mathbf{p}\in\mathcal{P}} L(\mathbf{x},\mathbf{p},\boldsymbol{\lambda}). \tag{3.5}$$

The subgradient iteration proceeds as follows:

$$\boldsymbol{\lambda}(\ell+1) = [\boldsymbol{\lambda}(\ell) + \beta_\ell(\mathbf{g}_1(\mathbf{x}(\ell)) + \mathbb{E}[\mathbf{g}_2(\mathbf{p}(\boldsymbol{\omega};\ell),\ell)])]^+ \tag{4.1}$$

where

$$\mathbf{x}(\ell) \in \arg\min_{\mathbf{x}\in\mathcal{X}} \big\{f_1(\mathbf{x}) + \boldsymbol{\lambda}^T(\ell)\mathbf{g}_1(\mathbf{x})\big\} \tag{4.2a}$$

$$\mathbf{p}(\boldsymbol{\omega};\ell) \in \arg\min_{\mathbf{p}\in\Pi(\boldsymbol{\omega})} \{f_2(\mathbf{p},\boldsymbol{\omega}) + \boldsymbol{\lambda}^T(\ell)\mathbf{g}_2(\mathbf{p},\boldsymbol{\omega})\},\ \boldsymbol{\omega}\in\Omega. \tag{4.2b}$$

Note that the minimization in (4.2b) returns a function of ω which is in the set $\mathcal{P}$ [cf. (3.1)]. The iteration is initialized with arbitrary $\boldsymbol{\lambda}(1) \geq \mathbf{0}$.

There are two stepsize rules of interest here:

(S1) Constant stepsize: $\beta_\ell = \beta > 0$, $\ell = 1, 2, \ldots$

(S2) Nonsummable but square-summable stepsize: $\beta_\ell > 0$, $\ell = 1, 2, \ldots$; $\sum_\ell^\infty \beta_\ell = \infty$ and $\sum_{\ell=1}^\infty \beta_\ell^2 < \infty$.

Dual convergence follows for both stepsize rules from classical results in the literature; see e.g., [47, Ch. 8], [49] (or Propositions 2.1 and 2.1 specialized to the synchronous case). The challenge is to obtain near-optimal primal variables, considering that problem (3.2) is nonconvex in general.

To this end, running averages of the iterates $\mathbf{x}(\ell)$ are formed as follows:

$$\bar{\mathbf{x}}(\ell) := \frac{1}{B_\ell}\sum_{l=1}^{\ell} \beta_l \mathbf{x}(l). \tag{4.3}$$

It is clear that (4.3) becomes $\bar{\mathbf{x}}(\ell) = \frac{1}{\ell}\sum_{l=1}^{\ell}\mathbf{x}(l)$ when the stepsize is constant. Certain results and proofs will be given for the general $\bar{\mathbf{x}}(\ell)$ defined in (4.3), and will be specialized to the two cases of stepsizes (S1) and (S2) when needed.

The following lemma provides bounds on the constraint violation and the objective evaluated at the running averages (4.3).

Lemma 4.1. *Let C3.1–C3.4 hold and $\boldsymbol{\lambda}^*$ denote any optimal Lagrange multiplier vector. Then, there exists a sequence $\mathring{\mathbf{p}}(\boldsymbol{\omega};\ell)$ in $\mathcal{P}$ such that for $\ell = 1, 2, \ldots$ it holds that*

$$\text{(i)} \ \left\| \left[\mathbf{g}_1(\bar{\mathbf{x}}(\ell)) + \mathbb{E}[\mathbf{g}_2(\mathring{\mathbf{p}}(\boldsymbol{\omega};\ell),\boldsymbol{\omega})]\right]^+ \right\| \leq \frac{||\boldsymbol{\lambda}(\ell+1)||}{B_\ell} \tag{4.4}$$

$$\text{(ii)} \ f_1(\bar{\mathbf{x}}(\ell)) + \mathbb{E}[f_2(\mathring{\mathbf{p}}(\boldsymbol{\omega};\ell),\boldsymbol{\omega})] \geq \rho^* - \frac{||\boldsymbol{\lambda}(1)||^2}{2B_\ell} - \frac{G^2}{2B_\ell}\sum_{l=1}^{\ell}\beta_l^2 \tag{4.5}$$

$$\text{(iii)} \ f_1(\bar{\mathbf{x}}(\ell)) + \mathbb{E}[f_2(\mathring{\mathbf{p}}(\boldsymbol{\omega};\ell),\boldsymbol{\omega})] \leq \rho^* + ||\boldsymbol{\lambda}^*|| \left\| \left[\mathbf{g}_1(\bar{\mathbf{x}}(\ell)) + \mathbb{E}[\mathbf{g}_2(\mathring{\mathbf{p}}(\boldsymbol{\omega};\ell),\boldsymbol{\omega})]\right]^+ \right\|. \tag{4.6}$$

Part (i) of Lemma 4.1 provides an upper bound on the constraint violation caused by the sequence $\{\bar{\mathbf{x}}(\ell), \mathring{\mathbf{p}}(\boldsymbol{\omega};\ell)\}$ as a function of the Langrange multiplier iterates $\boldsymbol{\Lambda}(\ell)$. Parts (ii) and (iii) give lower and upper bounds for the objective function value at $\bar{\mathbf{x}}(\ell)$. The proof of Lemma 4.1 follows. It adapts methods from [22] (which deals with convex programming), and combines them Lemma 3.1.

Proof of Lemma 4.1. Consider the iterates $\mathbf{x}(l)$ and $\mathbf{p}(\boldsymbol{\omega};l)$, $l = 1, 2, \ldots, \ell$ generated by the subgradient method. Due to Lemma 3.1, there exists $\mathring{\mathbf{p}}(\boldsymbol{\omega};\ell) \in \Pi(\boldsymbol{\omega})$ for all $\boldsymbol{\omega} \in \Omega$ such that

$$\frac{1}{B_\ell}\sum_{l=1}^{\ell}\beta_l \begin{bmatrix} f_2(\mathbf{p}(\boldsymbol{\omega};l),\boldsymbol{\omega}) \\ \mathbf{g}_2(\mathbf{p}(\boldsymbol{\omega};l),\boldsymbol{\omega}) \end{bmatrix} = \begin{bmatrix} f_2(\mathring{\mathbf{p}}(\boldsymbol{\omega};\ell),\boldsymbol{\omega}) \\ \mathbf{g}_2(\mathring{\mathbf{p}}(\boldsymbol{\omega};\ell),\boldsymbol{\omega}) \end{bmatrix} \tag{4.7}$$

The proof of the parts (i)–(iii) follow next.

(i) It holds for $l = 1, 2, \ldots$ from the subgradient iteration [cf. (4.1)] that

$$\boldsymbol{\lambda}(l+1) = \boldsymbol{\lambda}(l) + \beta_l\big(\mathbf{g}_2(\mathbf{x}(l)) + \mathbb{E}[\mathbf{g}_2(\mathbf{p}(\boldsymbol{\omega};l),\boldsymbol{\omega})]\big) \tag{4.8}$$

Summing up (4.8) for $l = 1, \ldots, \ell$, dividing by B_ℓ, and using the fact that $\boldsymbol{\lambda}(1) \geq \mathbf{0}$ yields

$$\frac{1}{B_\ell}\sum_{l=1}^{\ell}\beta_l\mathbf{g}(\mathbf{x}(l)) + \frac{1}{B_\ell}\sum_{l=1}^{\ell}\beta_l\mathbb{E}[\mathbf{g}_2(\mathbf{p}(\boldsymbol{\omega};l),\boldsymbol{\omega})] \leq \frac{\boldsymbol{\lambda}(\ell+1)}{B_l}. \tag{4.9}$$

Due to the convexity of function $\mathbf{g}(.)$, it holds that $\mathbf{g}(\bar{\mathbf{x}}(l)) \leq \frac{1}{B_l}\sum_{l=1}^{\ell}\beta_l\mathbf{g}(\mathbf{x}(l))$. Using the latter and (4.7) into (4.9) it follows that

$$\mathbf{g}_1(\bar{\mathbf{x}}(\ell)) + \mathbb{E}[\mathbf{g}_2(\mathring{\mathbf{p}}(\boldsymbol{\omega};\ell),\boldsymbol{\omega})] \leq \frac{\boldsymbol{\lambda}(\ell+1)}{B_\ell} \tag{4.10}$$

Because $\boldsymbol{\lambda}(\ell+1) \geq \mathbf{0}$ and $[.]^+ \geq \mathbf{0}$, it is easy to deduce that

$$\left[\mathbf{g}_1(\bar{\mathbf{x}}(\ell)) + \mathbb{E}[\mathbf{g}_2(\mathring{\mathbf{p}}(\boldsymbol{\omega};\ell),\boldsymbol{\omega})]\right]^+ \leq \frac{\boldsymbol{\lambda}(\ell+1)}{B_\ell} \tag{4.11}$$

and then, (4.4) follows by taking the norm of both sides of (4.11).

(ii) Due to the convexity of $f(.)$, it holds that $f(\bar{\mathbf{x}}(\ell)) \leq \frac{1}{B_\ell}\sum_{l=1}^{\ell}\beta_l f(\mathbf{x}(l))$. Adding and subtracting the same terms to the right-hand side of the latter and also invoking (4.7), one arrives at

$$\begin{aligned} f(\bar{\mathbf{x}}(\ell)) + \mathbb{E}[f_2(\mathring{\mathbf{p}}(\boldsymbol{\omega};\ell),\boldsymbol{\omega})] \leq & \frac{1}{B_\ell}\sum_{l=1}^{\ell}\beta_l\left[f(\mathbf{x}(l)) + \mathbb{E}[f_2(\mathbf{p}(\boldsymbol{\omega};l),\boldsymbol{\omega})]\right] \\ & - \frac{1}{B_\ell}\sum_{l=1}^{\ell}\beta_l\boldsymbol{\lambda}(l)^T\big(\mathbf{g}_1(\mathbf{x}(l)) + \mathbb{E}[\mathbf{g}_2(\mathbf{p}(\boldsymbol{\omega};l),\boldsymbol{\omega})]\big) \\ & + \frac{1}{B_\ell}\sum_{l=1}^{\ell}\beta_l\boldsymbol{\lambda}(l)^T\big(\mathbf{g}_1(\mathbf{x}(l)) + \mathbb{E}[\mathbf{g}_2(\mathbf{p}(\boldsymbol{\omega};l),\boldsymbol{\omega})]\big). \end{aligned} \tag{4.12}$$

It follows from the definition of the dual function [cf. (3.5) and (3.6)] that

$$f_1(\mathbf{x}(l)) + \mathbb{E}[f_2(\mathbf{p}(\boldsymbol{\omega};l),\boldsymbol{\omega})] + \boldsymbol{\Lambda}(l)^T\big(\mathbf{g}_1(\mathbf{x}(l)) + \mathbb{E}[\mathbf{g}_2(\mathbf{p}(\boldsymbol{\omega};l),\boldsymbol{\omega})]\big) \leq \rho^* \tag{4.13}$$

Using the latter into (4.12) it follows that

$$f(\bar{\mathbf{x}}(l)) + \mathbb{E}[f_2(\mathring{\mathbf{p}}(\boldsymbol{\omega};\ell),\boldsymbol{\omega})] \leq \rho^* - \frac{1}{B_\ell}\sum_{l=1}^{\ell}\beta_l\boldsymbol{\lambda}(l)^T\big(\mathbf{g}_1(\mathbf{x}(l)) + \mathbb{E}[\mathbf{g}_2(\mathbf{p}(\boldsymbol{\omega};l),\boldsymbol{\omega})]\big). \tag{4.14}$$

Moreover, it follows from (4.1) and the nonexpansive property of the projection that

$$\|\boldsymbol{\lambda}(l+1)\|^2 \leq \big\|\boldsymbol{\lambda}(l) + \beta_l\big(\mathbf{g}_1(\mathbf{x}(l)) + \mathbb{E}[\mathbf{g}_2(\mathbf{p}(\boldsymbol{\omega};l),\boldsymbol{\omega})]\big)\big\|^2 \tag{4.15}$$

and thus,

$$\begin{aligned} \|\boldsymbol{\Lambda}(l+1)\|^2 \leq & \|\boldsymbol{\lambda}(l)\|^2 + \beta_l^2\left\|\mathbf{g}_1(\mathbf{x}(l)) + \mathbb{E}[\mathbf{g}_2(\mathbf{p}(\boldsymbol{\omega};l),\boldsymbol{\omega})]\right\|^2 \\ & + 2\beta_l\boldsymbol{\lambda}(l)^T\left(\mathbf{g}_1(\mathbf{x}(l)) + \mathbb{E}[\mathbf{g}_2(\mathbf{p}(\boldsymbol{\omega};l),\boldsymbol{\omega})]\right). \end{aligned} \tag{4.16}$$

Summing (4.16) for $l = 1,\ldots,\ell$, and dividing by $2B_\ell$ yields

$$\begin{aligned} & -\frac{1}{B_\ell}\sum_{l=1}^{\ell}\beta_l\boldsymbol{\Lambda}(l)^T\big(\mathbf{g}(\mathbf{x}(l)) + \mathbb{E}[\mathbf{g}(\mathbf{p}(\boldsymbol{\omega};l),\boldsymbol{\omega})]\big) \\ & \quad \leq -\frac{1}{2B_\ell}\sum_{l=1}^{\ell}\beta_l^2\left\|\mathbf{g}_1(\mathbf{x}(l)) + \mathbb{E}[\mathbf{g}(\mathbf{p}_2(\boldsymbol{\omega};l),\boldsymbol{\omega})]\right\|^2 + \frac{||\boldsymbol{\lambda}(\ell+1)||^2 - ||\boldsymbol{\lambda}(1)||^2}{2B_\ell}. \end{aligned} \tag{4.17}$$

Combining (4.14) with (4.17) and introducing the bound on the subgradient norm, (4.5) follows.

(iii) This result follows readily from [22, Prop. 1(c)]. □

It should be emphasized that the previous lemma asserts the existence of $\mathring{\mathbf{p}}(\boldsymbol{\omega};\ell)$, but does not provide a way to obtain it.

The following proposition characterizes the convergence properties of sequence $\{\bar{\mathbf{x}}(\ell), \mathring{\mathbf{p}}(\boldsymbol{\omega};\ell)\}$ for the the constant stepsize rule (S1). It is the counterpart of Proposition 2.3 to the present setup.

Proposition 4.1. *Under C3.1–C3.4 and stepsize rule (S1), sequence* $\{\bar{\mathbf{x}}(\ell), \mathring{\mathbf{p}}(\boldsymbol{\omega};\ell)\}$ *satisfies*

$$\text{(i)} \lim_{\ell\to\infty} \Big\| \big[\mathbf{g}_1(\bar{\mathbf{x}}(\ell)) + \mathbb{E}[\mathbf{g}_2(\mathring{\mathbf{p}}(\boldsymbol{\omega};\ell),\boldsymbol{\omega})]\big]^+ \Big\| = 0 \tag{4.18}$$

$$\text{(ii)} \limsup_{\ell\to\infty} \left(f_1(\bar{\mathbf{x}}(\ell)) + \mathbb{E}[f_2(\mathring{\mathbf{p}}(\boldsymbol{\omega};\ell),\boldsymbol{\omega})]\right) \le f^* + \frac{\beta G^2}{2} \tag{4.19}$$

$$\text{(iii)} \liminf_{\ell\to\infty} \left(f_1(\bar{\mathbf{x}}(\ell)) + \mathbb{E}[f_2(\mathring{\mathbf{p}}(\boldsymbol{\omega};\ell),\boldsymbol{\omega})]\right) \ge f^* \tag{4.20}$$

Proof of Proposition 4.1. (i) The sequence $||\boldsymbol{\lambda}(\ell)||$ generated by the subgradient method with constant stepsize is bounded when the dual optimal set is compact; see e.g., [49]. But the optimal dual set is indeed compact in the present setup [cf. Proposition 3.1]. Hence, (4.18) follows from (4.4) by taking the lim sup as $\ell \to \infty$.

(ii) Using $\beta_\ell = \beta$ and $\rho^* = f^*$ in (4.5) and taking lim sup, (4.19) follows.

(iii) The set of optimal dual solutions is bounded. Eq. (4.18) then follows from (4.6) after invoking $\rho^* = f^*$ and (4.18). □

The following proposition characterizes the convergence properties of sequence $\{\bar{\mathbf{x}}(\ell), \mathring{\mathbf{p}}(\boldsymbol{\omega};\ell)\}$ for the the nonsummable but square-summable stepsize rule (S2). It is the counterpart of Proposition 2.4 to the present setup.

Proposition 4.2. *Under C3.1–C3.4 and stepsize rule (S2), sequence* $\{\bar{\mathbf{x}}(\ell), \mathring{\mathbf{p}}(\boldsymbol{\omega};\ell)\}$ *satisfies*

$$\text{(i)} \lim_{\ell\to\infty} \Big\| \big[\mathbf{g}_1(\bar{\mathbf{x}}(\ell)) + \mathbb{E}[\mathbf{g}_2(\mathring{\mathbf{p}}(\boldsymbol{\omega};\ell),\boldsymbol{\omega})]\big]^+ \Big\| = 0 \tag{4.21}$$

$$\text{(ii)} \lim_{\ell\to\infty} \left(f_1(\bar{\mathbf{x}}(\ell)) + \mathbb{E}[f_2(\mathring{\mathbf{p}}(\boldsymbol{\omega};\ell),\boldsymbol{\omega})]\right) = f^*. \tag{4.22}$$

Proof of Proposition 4.2. (i) Under the considered stepsize rule, sequence $\{\boldsymbol{\lambda}(\ell)\}$ converges to an optimal dual solution; see e.g., [47, Prop. 8.2.6]. Hence, it is bounded, and the result follows by taking limits in (4.4).

(ii) Using $\rho^* = f^*$ and $\lim_{\ell\to\infty}(1/B_\ell)\sum_{l=1}^{\ell}\beta_l^2 = 0$ in (4.5) and taking *limsup*, it follows that

$$\limsup_{\ell\to\infty}\left(f_1(\bar{\mathbf{x}}(\ell)) + \mathbb{E}[f_2(\mathring{\mathbf{p}}(\boldsymbol{\omega};\ell),\boldsymbol{\omega})]\right) \leq f^*. \tag{4.23}$$

Moreover, in a fashion similar to (4.20), it holds that

$$\liminf_{\ell\to\infty}\left(f_1(\bar{\mathbf{x}}(\ell)) + \mathbb{E}[f_2(\mathring{\mathbf{p}}(\boldsymbol{\omega};\ell),\boldsymbol{\omega})]\right) \geq f^*. \tag{4.24}$$

Upon combining the latter two inequalities, (4.22) follows. □

It should be emphasized that although $\mathring{\mathbf{p}}(\boldsymbol{\omega};\ell)$ is not known, the running averages $\bar{\mathbf{x}}(\ell)$ are straightforward to obtain, and are guaranteed to verify (near-)optimality together with $\mathring{\mathbf{p}}(\boldsymbol{\omega};\ell)$.

The subgradient method of the present section can be applied to optimal wireless network design. The optimal wireless networking problem is presented in the ensuing section, and the particular form of the algorithm is detailed in Section 4.3.

4.2 Cross-Layer Optimization Problem

This section formulates the cross-layer resource allocation problem for wireless fading networks. Its solution and the resulting network control strategy will be the themes of Sections 4.3 and 4.4.

Consider a wireless network as in [5] and [35], comprising a set of terminals (nodes) denoted by $\mathcal{V}$. Two nodes i, j form a link when they can communicate with each other; here, the communication between nodes is bidirectional. The set of nodes that a node $i \in \mathcal{V}$ can communicate with, forms the neighborhood of i that is denoted by $\mathcal{N}(i)$. The network operates in a time-slotted fashion.

Packets generated exogenously at each node correspond to possibly different applications (such as video, voice, file transfer), and are destined for different nodes. Such packet streams are called commodities, and are generically indexed by k. The destination node associated with each commodity k, is denoted by $\text{dest}(k) \in \mathcal{V}$. Each node serves commodities that have other nodes as destination. Clearly, commodities that have as destination node i are never actually generated at node i. Hence, the set of commodities that can be generated exogenously at node i is $\mathcal{K}_i := \{k | i \neq \text{dest}(k)\}$.

The transport layer protocol of the network is responsible for maintaining the end-to-end flow of packets from each node i to the destination of every commodity k that is generated exogenously at node i. The average rate of such a flow is denoted by a_i^k.

Moreover, there are also flows endogenous to the network layer. Specifically, each node i will transmit to its neighbors $j \in \mathcal{N}(i)$ packets of commodity $k \in \mathcal{K}_i$ at an average rate denoted by r_{ij}^k, called network-layer rate or multicommodity flow. Rates r_{ij}^k are essentially routing variables, because they dictate how packets of various commodities are forwarded to the outgoing links of node i. As such, they pertain to the network layer. Hence, at node i there are exogenous packet arrivals with rate a_i^k from the transport layer, and endogenous packet arrivals with rates r_{ji}^k from neighbors $j \in \mathcal{N}(i)$. Packets also leave node i with rates r_{ij}^k towards neighbors $j \in \mathcal{N}(i)$. Queue stability considerations dictate these rates to satisfy a *flow conservation* constraint

$$a_i^k \leq \sum_{j \in \mathcal{N}(i)} r_{ij}^k - \sum_{j \in \mathcal{N}(i), j \neq \text{dest}(k)} r_{ji}^k \qquad \forall\, i,\, \forall\, k \in \mathcal{K}_i. \tag{4.25}$$

In addition, link (i, j) "carries" total average rate $\sum_{k \in \mathcal{K}_i} r_{ij}^k$ that cannot exceed the average (i.e., ergodic) capacity of the link, denoted by c_{ij}. This gives rise to the *link capacity* constraint

$$\sum_{k \in \mathcal{K}_i} r_{ij}^k \leq c_{ij} \qquad \forall\, i,\, \forall\, j \in \mathcal{N}(i). \tag{4.26}$$

Link capacities c_{ij} in wireless networks are not fixed, but vary with time and depend on the specific physical and medium access control (MAC) layers, as well as on the allocation of network resources. In the considered network model, terminals have a set of tones $\mathcal{F} := \{1, \ldots, F\}$ available for transmission. It is further assumed that link (i, j) over tone f has power gain coefficient h_{ij}^f, which captures fading effects. The focus of the present work is on *non-orthogonal* medium access, where different terminals are allowed to use the same frequency to transmit, treating other terminals' transmissions as noise. Network resources here are power allocations $p_{ij}^f(\mathbf{h})$ for link (i, j) over tone f as a function of the power gains h_{ij}^f collected in the vector $\mathbf{h}$. The channel vector $\mathbf{h}$ may change from slot to slot, or, even at a faster scale, depending on the coherence time. The fading process is stationary and ergodic, but otherwise has arbitrary correlation, as is typically the case in practice. Let $\mathbb{E}$ denote expectation over the stationary distribution of $\mathbf{h}$.

Link capacities are dictated by the SINR; the instantaneous SINR of link (i,j) over tone f is

$$\gamma_{ij}^f := \frac{h_{ij}^f p_{ij}^f(\mathbf{h})}{\sigma_j^f + \sum_{(k,l)\in\mathcal{I}_{ij}} h_{kj}^f p_{kl}^f(\mathbf{h})} \tag{4.27}$$

where σ_j^f is the noise variance at node j over tone f, and $\mathcal{I}_{ij}$ denotes the set of links causing interference to (i,j). This set consists of the links carrying: (i) incoming transmissions for j over f from nodes other than i; (ii) outgoing transmissions from j; and (iii) transmissions originating from nodes in $\mathcal{N}(j)$ intended for nodes other than j. Specifically, this set takes the form

$$\mathcal{I}_{ij} := \big\{(k,l)\colon k\in\mathcal{N}(j)\backslash\{i\}, l\in\mathcal{N}(k);\ (i,l)\colon l\in\mathcal{N}(i)\backslash\{j\};\ (j,l)\colon l\in\mathcal{N}(j)\big\}. \tag{4.28}$$

Based on (4.28), the term $h_{jj}^f \sum_{l\in\mathcal{N}(j)} p_{jl}^f$ in the denominator in (4.27) represents self-interference to receiving node j from transmissions originating from j. In order to discourage this self-interference, and thus ensure half-duplex operation, h_{jj}^f is set to a high value. Furthermore, interference from "far-away" links, corresponding to (k,l) with $k\in\mathcal{N}(j)$, $k\neq j$ and $l\in\mathcal{N}(k)$, $l\neq j$, is neglected. Note that γ_{ij}^f depends on power allocations $p_{ij}^f(\mathbf{h})$ as well as on $\mathbf{h}$.

Supposing capacity-achieving codebooks, the ergodic capacity c_{ij} of link (i,j) is

$$c_{ij} = \mathbb{E}\left[\sum_{f\in\mathcal{F}} \ln\big(1+\gamma_{ij}^f\big)\right]. \tag{4.29}$$

Two kinds of power constraints are considered. The first ones are instantaneous spectral mask constraints, expressed as $0\le p_{ij}^f(\mathbf{h}) \le p_{ij\,\max}^f$. The second kind pertains to the average power consumed by a node in the network. Specifically, the average power per node i is defined as

$$p_i := \mathbb{E}\left[\sum_{j\in\mathcal{N}(i)}\sum_{f\in\mathcal{F}} p_{ij}^f(\mathbf{h})\right] \tag{4.30}$$

and is assumed constrained by a power budget $p_{i\,\max}$, i.e., $0\le p_i\le p_{i\,\max}$ for all i.

Average exogenous rates a_i^k in practical networks lie within application-specific bounds expressed as $a_{i\,\min}^k \le a_i^k \le a_{i\,\max}^k$. Furthermore, the network designer may wish to impose upper bounds on link capacities c_{ij} and multicommodity flows r_{ij}^k, that

is, $0 \leq c_{ij} \leq c_{ij\,\max}$ and $0 \leq r_{ij}^k \leq r_{i\,\max}$. The notation $\mathbf{x}$ will be used to denote the collection of all average variables, i.e., a_i^k, r_{ij}^k, c_{ij}, p_i for all $i \in \mathcal{V}$, $j \in \mathcal{N}(i)$, $k \in \mathcal{K}_i$. Then, all previously mentioned average and instantaneous "box" constraints can be conveniently written using the set notation

$$\begin{aligned}\mathcal{B}:=\{\mathbf{x}, \mathbf{p}(\mathbf{h}) : 0 \leq p_{ij}^f(\mathbf{h}) \leq p_{ij\,\max}^f, 0 \leq p_i \leq p_{i\,\max},\\ a_{i\,\min}^k \leq a_i^k \leq a_{i\,\max}^k, 0 \leq c_{ij} \leq c_{ij\,\max}, 0 \leq r_{ij}^k \leq r_{i\,\max}\}.\end{aligned} \tag{4.31}$$

Higher exogenous arrival rates a_i^k are in general more desirable, a fact captured by utility functions $U_i^k(a_i^k)$ that are selected strictly increasing and concave. Moreover, average power consumption will be penalized by cost functions $V_i(p_i)$, selected strictly increasing and convex. The optimal operating point of the wireless network refers to average variables a_i^k, r_{ij}^k, c_{ij}, p_i and instantaneous power allocations $p_{ij}^f(\mathbf{h})$, such that constraints (4.25), (4.26), (4.29), (4.30) and (4.31) are respected, while the total network utility is maximized and the total cost minimized; that is,

$$\mathsf{P} = \max_{\{\mathbf{x}, \mathbf{p}(\mathbf{h})\} \in \mathcal{B}} \sum_{i, k \in \mathcal{K}_i} U_i^k(a_i^k) - \sum_i V_i(p_i) \tag{4.32a}$$

$$\text{subj. to } a_i^k \leq \sum_{j \in \mathcal{N}(i)} r_{ij}^k - \sum_{\substack{j \in \mathcal{N}(i) \\ j \neq \text{dest}(k)}} r_{ji}^k \; \forall k \in \mathcal{K}_i, \; \forall i \tag{4.32b}$$

$$\sum_{k \in \mathcal{K}_i} r_{ij}^k \leq c_{ij} \qquad \forall j \in \mathcal{N}(i), \; \forall i \tag{4.32c}$$

$$c_{ij} \leq \mathbb{E}\left[\sum_f \ln\left(1 + \gamma_{ij}^f\right)\right] \quad \forall j \in \mathcal{N}(i), \; \forall i \tag{4.32d}$$

$$\mathbb{E}\left[\sum_{j \in \mathcal{N}(i)} \sum_{f \in \mathcal{F}} p_{ij}^f(\mathbf{h})\right] \leq p_i \; \forall i. \tag{4.32e}$$

where inequalities (4.32d) and (4.32e) replaced the equalities (4.29) and (4.30). Note that variables a_i^k, r_{ij}^k, and c_{ij} have units nats/(sec $\cdot$ Hz) [cf. (4.29)], while variables p_i and $p_{ij}^f(\mathbf{h})$ have units W/Hz.

It is worth stressing at this point that in practice, the generated traffic traverses the network in bits or packets, which are stored at every node in queues. In particular, each node keeps a queue at the network layer for every commodity that can enter the node, and a queue at the physical layer for every outgoing link. For instance, the average

arrival rate to the queue of node i for commodity $k \in \mathcal{K}_i$ is $a_i^k + \sum_{j\in\mathcal{N}(i), j\neq \mathrm{dest}(k)} r_{ji}^k$, while the average departure rate is $\sum_{j\in\mathcal{N}(i)} r_{ij}^k$. The constraints in (4.32) guarantee that all the queues in the network are stable [5, Ch. 3].

It is clear that (4.32) has the generic form of (3.2), where the set Ω is here the set of all possible channel values $\mathbf{h}$, and $f_2 \equiv 0$. Therefore, all results of Sections 3.1 and 4.1 are applicable, when the fading distribution is continuous, as is the case with Rayleigh or Rice fading. It should be mentioned that zero duality gap for a networking problem with generic capacity functions has been established in [35]. Algorithmic solutions to (4.32) based on the subgradient method of Section 4.1 are developed in the ensuing section.

4.3 Subgradient Method and Network Control

The subgradient method for optimal wireless network design is detailed in Subsection 4.3.1. A more detailed description of the queueing operations and the network functionality are given in Subsection 4.3.2, where the solution of (4.32) is utilized in order to obtain a simple strategy for network control.

4.3.1 Subgradient Method

Let ν_i^k, ξ_{ij}, λ_{ij}, μ_i be Lagrange multipliers for constraints (4.32b), (4.32c), (4.32d) and (4.32e), respectively. The box constraints (4.31) are kept implicit. Also let $\mathbf{\Lambda}$ collectively denote all Lagrange multipliers. The Lagrangian function of (4.32) reduces after straightforward re-arrangements to

$$\begin{aligned} L(\mathbf{\Lambda}, \mathbf{x}, \mathbf{p}(\mathbf{h})) = & \sum_{i,k\in\mathcal{K}_i} \Big(U_i^k(a_i^k) - \nu_i^k a_i^k \Big) + \sum_i \left(\mu_i p_i - V_i(p_i)\right) \\ & + \sum_{\substack{i,j\in\mathcal{N}(i) \\ f\in\mathcal{F}}} \mathbb{E}\big[\lambda_{ij} \ln\big(1+\gamma_{ij}^f\big) - \mu_i p_{ij}^f(\mathbf{h})\big] + \sum_{i,j\in\mathcal{N}(i)} \left(\xi_{ij} - \lambda_{ij}\right) c_{ij} \\ & + \sum_{\substack{i,k\in\mathcal{K}_i, j\in\mathcal{N}(i) \\ j\neq \mathrm{dest}(k)}} \big(\nu_i^k - \nu_j^k - \xi_{ij}\big) r_{ij}^k + \sum_{\substack{i,k\in\mathcal{K}_i, j\in\mathcal{N}(i) \\ j=\mathrm{dest}(k)}} \big(\nu_i^k - \xi_{ij}\big) r_{ij}^k. \end{aligned} \tag{4.33}$$

It should be noted that the last sum in (4.33) might not be present. The dual function and the dual problem are, respectively,

$$\rho(\mathbf{\Lambda}) := \max_{(\mathbf{x},\mathbf{p}(\mathbf{h}))\in\mathcal{B}} L(\mathbf{\Lambda}, \mathbf{x}, \mathbf{p}(\mathbf{h})) \tag{4.34}$$

$$\rho^* = \min_{\mathbf{\Lambda} \geq \mathbf{0}} \rho(\mathbf{\Lambda}). \tag{4.35}$$

The dual problem (4.35) is solved via subgradient iterations. Let t be the iteration index. The sequence $\mathbf{\Lambda}(t)$ obtained from the subgradient method, with initial $\mathbf{\Lambda}(0) \geq \mathbf{0}$, is

$$(\mathbf{x}(t), \mathbf{p}(\mathbf{h}; t)) \in \underset{(\mathbf{x}, \mathbf{p}(\mathbf{h})) \in \mathcal{B}}{\arg\max}\ L(\mathbf{\Lambda}(t), \mathbf{x}, \mathbf{p}(\mathbf{h})) \tag{4.36a}$$

$$\mathbf{\Lambda}(t+1) = \Big[\mathbf{\Lambda}(t) + \beta_t\big(\mathbf{g}(\mathbf{x}(t)) + \mathbb{E}[\mathbf{g}(\mathbf{p}(\mathbf{h}; t), \mathbf{h})]\big)\Big]^+ \tag{4.36b}$$

where the inclusion symbol ($\in$) in (4.36a) covers the possibility of multiple maximizers and β_t is the stepsize (varying with t in general).

Now apply the $\max_{(\mathbf{x}, \mathbf{p}(\mathbf{h})) \in \mathcal{B}}$ operator to (4.33), and note that (4.33) is a sum of terms where each term depends on different variables, and (4.31) are box constraints. The Lagrangian maximization becomes

$$a_i^k(t) \in \underset{a_{i\min}^k \leq a_i^k \leq a_{i\max}^k}{\arg\max} \big[U_i^k(a_i^k) - \nu_i^k(t) a_i^k\big] \tag{4.37a}$$

$$r_{ij}^k(t) \in \underset{0 \leq r_{ij}^k \leq r_{i\max}}{\arg\max} \begin{cases} \big(\nu_i^k(t) - \nu_j^k(t) - \xi_{ij}(t)\big) r_{ij}^k & \text{if } j \neq \text{dest}(k) \\ \big(\nu_i^k(t) - \xi_{ij}(t)\big) r_{ij}^k & \text{if } j = \text{dest}(k) \end{cases} \tag{4.37b}$$

$$c_{ij}(t) \in \underset{0 \leq c_{ij} \leq c_{ij\max}}{\arg\max} \big[(\xi_{ij}(t) - \lambda_{ij}(t)) c_{ij}\big] \tag{4.37c}$$

$$p_i(t) \in \underset{0 \leq p_i \leq p_{i\max}}{\arg\max} \big[\mu_i(t) p_i - V_i(p_i)\big] \tag{4.37d}$$

$$\mathbf{p}(\mathbf{h}; t) \in \underset{\substack{0 \leq p_{ij}^f \leq p_{ij\max}^f \\ \forall\, i, \forall\, j \in \mathcal{N}(i), \forall\, f}}{\arg\max} \sum_{i,f,j \in \mathcal{N}(i)} \big[\lambda_{ij}(t) \ln\big(1 + \gamma_{ij}^f\big) - \mu_i(t) p_{ij}^f\big]. \tag{4.37e}$$

Eq. (4.37e) is obtained by noting that the part of (4.33) which involves the $\mathbb{E}[.]$ operator can be maximized if the term inside the expectation is maximized for each fading state $\mathbf{h}$. The optimization variables p_{ij}^f in the right hand side of (4.37e) are not functions of $\mathbf{h}$; but the maximizer in the left hand side depends on $\mathbf{h}$, because the SINR depends on $\mathbf{h}$.

The decomposition of the Lagrangian maximization [cf. (4.36a)] into (4.37a)–(4.37e) can be understood as separation of the solution of the wireless networking problem into conventional layers [35]. In particular, (4.37a) solves the flow control problem at the

transport layer; (4.37b) performs routing at the network layer; (4.37c) and (4.37d) address the link rate control and (average) power control at the data link layer; and (4.37e) solves the power allocation at the physical layer.

The multiplier updates take the form

$$\nu_i^k(t+1) = \left[\nu_i^k(t) + \beta_t \left(a_i^k(t) - \sum_{j\in\mathcal{N}(i)} r_{ij}^k(t) + \sum_{\substack{j\in\mathcal{N}(i)\\ j\neq \mathrm{dest}(k)}} r_{ji}^k(t)\right)\right]^+ \tag{4.38a}$$

$$\xi_{ij}(t+1) = \left[\xi_{ij}(t) + \beta_t \left(\sum_{k\in\mathcal{K}_i} r_{ij}^k(t) - c_{ij}(t)\right)\right]^+ \tag{4.38b}$$

$$\lambda_{ij}(t+1) = \left[\lambda_{ij}(t) + \beta_t \left(c_{ij}(t) - \mathbb{E}\left[\sum_f \ln\bigl(1+\gamma_{ij}^f(t)\bigr)\right]\right)\right]^+ \tag{4.38c}$$

$$\mu_i(t+1) = \left[\mu_i(t) + \beta_t \left(\mathbb{E}\left[\sum_{j\in\mathcal{N}(i),f} p_{ij}^f(\mathbf{h};t)\right] - p_i(t)\right)\right]^+ . \tag{4.38d}$$

In order to perform iterations (4.38), the solution of (4.37) is required. Each of the problems (4.37a)–(4.37d) involves a single variable, concave objective, and box constraints; thus, their solution as a function of the Lagrange multipliers is straightforward. But the solution of (4.37e) poses significant challenges. In particular, (4.37e) carries similarities to the problem of power control in deterministic DSL channels, see e.g., [42, 43] and references therein. This problem may have exponential complexity in the number of links [41]. Section 4.4 develops an approximate solution of (4.37e). One remark is now due on the dual updates (4.38).

Remark2. The subgradients in (4.38a) and (4.38b) are easily determined, once the solution of (4.37) is found. On the other hand, (4.38c) and (4.38d) involve the expectation $\mathbb{E}[.]$. This can be evaluated efficiently through Monte Carlo methods using two possible approaches. The first method assumes that the distribution of $\mathbf{h}$ is known (e.g., Rayleigh or Rician), and independent realizations of $\mathbf{h}$ are drawn from this distribution offline. For each realization, (4.37e) is solved and the quantities inside the expectations in (4.38c)–(4.38d) are formed. Then the expectations are estimated through the sample averages of the quantities inside the expectations. The second method assumes that measurements of the channel vector $\mathbf{h}$ are available, and uses them in exactly the same way as the first method. The second method works even if the measurements

are correlated—which may happen in the present context when the fading process is correlated across time.

In order to complete the design, it is also important to obtain the optimal (primal) solution of (4.32), apart from the optimal dual variables. These are obtained by forming the running averages $\bar{a}_i^k$, $\bar{r}_{ij}^k$, $\bar{c}_{ij}$, and $\bar{p}_i$.

The flow of the subgradient algorithm is summarized next, where the solution of (4.37a)–(4.37d) is shown explicitly. The utilities $U_i^k(a_i^k)$ and $V_i(p_i)$ are assumed non-linear, e.g., $U_i^k(a_i^k) = \ln a_i^k$, for which the inverses $\left(U_i^{k\prime}\right)^{-1}(.)$ and $(V_i')^{-1}(.)$ exist. Algorithmic solutions to the power allocation at PHY [cf. (4.37e) or line 12 next] are given in Section 4.4. In Section 4.3.2, a simple strategy for network control based on the optimal solution of (4.32) is described.

1: Initialize Lagrange multipliers $\nu_i^k(0) = \xi_{ij}(0) = \lambda_{ij}(0) = 0$, $\mu_i(0) = 0$

2: Initialize running averages $\bar{a}_i^k(0) = \bar{r}_{ij}^k(0) = \bar{c}_{ij}(0) = 0$, $\bar{p}_i(0) = 0$

3: **for** $t = 0, 1, 2, \ldots, N$ **do**

4: **for all** nodes $i \in \mathcal{V}$, neighbors $j \in \mathcal{N}(i)$, and flows $k \in \mathcal{K}_i$ **do**

5: Update arrival rates $a_i^k(t) \leftarrow \left[\left(U_i^{k\prime}\right)^{-1}(\nu_i^k(t))\right]_{a_{i\min}^k}^{a_{i\max}^k}$

6: Update routing variables $r_{ij}^k(t) \leftarrow \begin{cases} 0 & \text{if } j \neq \text{dest}(k) \text{ and } \nu_i^k(t) - \nu_j^k(t) - \xi_{ij}(t) \leq 0 \\ & \text{or if } j = \text{dest}(k) \text{ and } \nu_i^k(t) - \xi_{ij}(t) \leq 0 \\ r_{i\max} & \text{if } j \neq \text{dest}(k) \text{ and } \nu_i^k(t) - \nu_j^k(t) - \xi_{ij}(t) \geq 0 \\ & \text{or if } j = \text{dest}(k) \text{ and } \nu_i^k(t) - \xi_{ij}(t) \geq 0 \end{cases}$

7: Update link capacities $c_{ij}(t) \leftarrow \begin{cases} 0 & \text{if } \xi_{ij}(t) - \lambda_{ij}(t) \leq 0 \\ c_{ij\max} & \text{if } \xi_{ij}(t) - \lambda_{ij}(t) \geq 0 \end{cases}$

8: Update average power $p_i(t) \leftarrow \left[(V_i')^{-1}(\mu_i(t))\right]_0^{p_{i\max}}$

9: **end for**

10: Obtain R samples of $\mathbf{h}$, $\{\mathbf{h}^{(r)}\}_{r=1}^R$:

Draw R independent realizations of $\mathbf{h}$ from its distribution, or use R available measurements

11: **for** channel realizations $r = 1, \ldots, R$ **do**

12: Power allocation at PHY

$$\mathbf{p}\left(\mathbf{h}^{(r)};t\right) \in \underset{\substack{0\le p_{ij}^f \le p_{ij\,\max}^f \\ \forall\, i, \forall\, j\in\mathcal{N}(i), \forall\, f}}{\arg\max} \sum_{i,f,j\in\mathcal{N}(i)} \left[\lambda_{ij}(t)\ln\left(1+\gamma_{ij}^{f\,(r)}\right) - \mu_i(t)p_{ij}^f(\mathbf{h}^{(r)})\right]$$

13: **end for**

14: **for all** nodes $i \in \mathcal{V}$, neighbors $j \in \mathcal{N}(i)$, and flows $k \in \mathcal{K}_i$ **do**

15: Approximate ergodic capacity $\hat{C}_{ij} \leftarrow \frac{1}{R}\sum_{r=1}^{R}\sum_f \ln\left(1+\gamma_{ij}^{f\,(r)}(t)\right)$

16: Approximate average power $\hat{P}_i \leftarrow \frac{1}{R}\sum_{r=1}^{R}\sum_{j\in\mathcal{N}(i),f} p_{ij}^f\left(\mathbf{h}^{(r)};(t)\right)$

17: Update dual variables

$$\nu_i^k(t+1) \leftarrow \left[\nu_i^k(t) + \beta_t\left(a_i^k(t) - \textstyle\sum_{j\in\mathcal{N}(i)} r_{ij}^k(t) + \sum_{\substack{j\in\mathcal{N}(i)\\ j\neq \mathrm{dest}(k)}} r_{ji}^k(t)\right)\right]^+$$

$$\xi_{ij}(t+1) \leftarrow \left[\xi_{ij}(t) + \beta_t\left(\textstyle\sum_{k\in\mathcal{K}_i} r_{ij}^k(t) - c_{ij}(t)\right)\right]^+$$

$$\lambda_{ij}(t+1) \leftarrow \left[\lambda_{ij}(t) + \beta_t\left(c_{ij}(t) - \hat{C}_{ij}\right)\right]^+$$

$$\mu_i(t+1) \leftarrow \left[\mu_i(t) + \beta_t\left(\hat{P}_i - p_i(t)\right)\right]^+$$

18: **end for**

19: **for all** nodes $i \in \mathcal{V}$, neighbors $j \in \mathcal{N}(i)$, and flows $k \in \mathcal{K}_i$ **do**

20: Update running averages

$$\bar{a}_i^k(t+1) \leftarrow \tfrac{1}{K_{t+1}}\left[K_t\bar{a}_i^k(t) + \beta_t a_i^k(t)\right]$$

$$\bar{r}_{ij}^k(t+1) \leftarrow \tfrac{1}{K_{t+1}}\left[K_t\bar{r}_{ij}^k(t) + \beta_t r_{ij}^k(t)\right]$$

$$\bar{c}_{ij}(t+1) \leftarrow \tfrac{1}{K_{t+1}}\left[K_t\bar{c}_{ij}(t) + \beta_t c_{ij}(t)\right]$$

$$\bar{p}_i(t+1) \leftarrow \tfrac{1}{K_{t+1}}\left[K_t\bar{p}_i(t) + \beta_t p_i(t)\right]$$

21: **end for**

22: **end for**

4.3.2 A strategy for network control

This section describes a network control strategy that utilizes the solution of problem (4.32)—that is, the optimal $a_i^{k\,*}$, $r_{ij}^{k\,*}$, c_{ij}^*, p_i^*, $\mathbf{p}^*(\mathbf{h})$. The network operates in time slots, which are indexed by $\ell = 1, 2, 3, \ldots$, and network control amounts to determining how the various flows and powers are allocated per time slot. The overall algorithm proceeds in two phases: (a) an offline phase, where near-optimal primal variables and Lagrange multipliers are obtained; and (b) an online phase, where power is allocated

based on the current channel realization, and routing is performed. Both phases are run at a central network controller. The two phases are detailed next.

Offline phase

During this phase, the subgradient algorithm is implemented along with the running averages, as explained in Sections 4.3.1. The purpose is to obtain near-optimal primal variables and Lagrange multipliers. In practice, a specific number (N) of subgradient iterations will be run. Hence, what is referred to as (near-)optimal a_i^{k*}, $r_{ij}^{k\,*}$, c_{ij}^*, and p_i^* are in fact the values of the running averages $\bar{a}_i^k(N)$, $\bar{r}_{ij}^k(N)$, $\bar{c}_{ij}(N)$, and $\bar{p}_i(N)$ at the last iteration. Similarly, the values of $\mathbf{\Lambda}(N)$ (for diminishing stepsize) or $\bar{\mathbf{\Lambda}}(N)$ (for constant stepsize) at the last iteration will yield the (near-)optimal Lagrange multipliers to be used during the online phase. Note also that Propositions 4.1 and 4.2 do not provide a method to obtain power allocations $\mathbf{p}^*(\mathbf{h})$. These will be obtained during the online phase.

The offline phase is run before communication takes place. The central controller must either know the channel distribution or have channel measurements, in order to estimate the expectations in (4.38c) and (4.38d), as explained in Remark 2. Note that even when the offline phase requires considerable computation, this does not imply overhead during the network operation. Moreover, the offline phase needs to be re-run only when the network setup or the channel statistics change.

Online phase

This phase pertains to the network operation.

Each node i keeps a queue for each commodity $k \in \mathcal{K}_i$ at the network layer, and a queue for each neighbor $j \in \mathcal{N}(i)$ at the physical layer; see Fig. 4.1. Every network layer queue accepts exogenous traffic—from the transport layer—with instantaneous rate $\check{a}_i^k(\ell)$. Every physical layer queue sends to the corresponding neighbor bits with instantaneous rate $\check{c}_{ij}(\ell)$, which depends on the instantaneous power allocation and the fading; this effect will be described in detail later. Each queue operates in a first-in-first-out (FIFO) fashion, and has unlimited storage space.

There is an interface connecting the network layer and physical layer queues at node i. This interface is responsible for routing, because it takes bits from each network layer

queue, and places them into the physical layer queues. The individual rate from the kth network layer queue to the jth physical layer queue is denoted by $\check{r}_{ij}^k(\ell)$. Then, bits leave the kth network layer queue with instantaneous rate $\sum_{j\in\mathcal{N}(i)} \check{r}_{ij}^k(\ell)$, and arrive at the jth physical layer queue with rate $\sum_{k\in\mathcal{K}_i} \check{r}_{ij}^k(\ell)$. Note that the data in the physical layer queues are labeled with the commodity index k, which determines their destination. The bits $\check{c}_{ij}(\ell)$ from node i to node j are placed into the corresponding network layer queues of node j according to their label, except those with $\text{dest}(k) = j$, which have arrived at their destination and are not placed in any queue. Now, consider this operation for the bits $\check{c}_{mi}(\ell)$ arriving to i from its neighbors $m \in \mathcal{N}(i)$; the endogenous arrivals at the kth network layer queue will be $\sum_{m\in\mathcal{N}(i),m\neq\text{dest}(k)} \check{c}_{mi}^k(\ell)$, where the $\check{c}_{mi}^k(\ell)$ are determined by splitting $\check{c}_{mi}(\ell)$. Note that if the time slot duration is normalized to one, all the aforementioned instantaneous rates also play the role of arrival or service processes for their respective queues.

A network control algorithm must determine $\check{a}_i^k(\ell)$, $\check{r}_{ij}^k(\ell)$, and $\check{c}_{ij}(\ell)$ per time slot ℓ. Variables $\check{c}_{ij}(\ell)$ also depend on the fading, which of course cannot be controlled. Here, $\check{a}_i^k(\ell)$, $\check{r}_{ij}^k(\ell)$, and $\check{c}_{ij}(\ell)$ will be determined by the optimal solution of (4.32), that is, a_i^{k*}, ${r_{ij}^k}^*$, and $\mathbf{p}^*(\mathbf{h})$. Specifically, since an optimal $\mathbf{p}^*(\mathbf{h})$ is not readily available, the solution of (4.37e) with the optimal Lagrange multipliers ${\lambda_{ij}}^*$ and μ_i^* will be used whenever $\mathbf{p}^*(\mathbf{h})$ is mentioned.

In the online operation, the central controller knows the current realization of $\mathbf{h}$, and has the a_i^{k*}, ${r_{ij}^k}^*$, ${\lambda_{ij}}^*$, and μ_i^* available from the offline phase. It can then obtain $\mathbf{p}^*(\mathbf{h})$ for the current channels—using the solvers of (4.37e) detailed in Section 4.4. These quantities are used per time slot as described next.

The network operates under the condition that the random traffic arrival process has the optimal long-term average, i.e., $\lim_{M\to\infty} \frac{1}{M} \sum_{\ell=1}^{M} \check{a}_i^k(\ell) = a_i^{k*}$. This operating condition for $\check{a}_i^k(\ell)$ is adopted because a_i^{k*} is the optimal operating point of the network, determined by problem (4.32). Supposing as usual that there are always enough packets available at the transport layer of every node (infinitely backlogged transport layer [5]), this condition is ensured if the controller admits packets with rate $\check{a}_i^k(\ell)$ having the correct long-term average a_i^{k*}. As a special case, one can set $\check{a}_i^k(\ell) = a_i^{k*}$ for all ℓ.

Moreover, the routing variables are set to the optimal multicommodity flows ${r_{ij}^k}^*$,

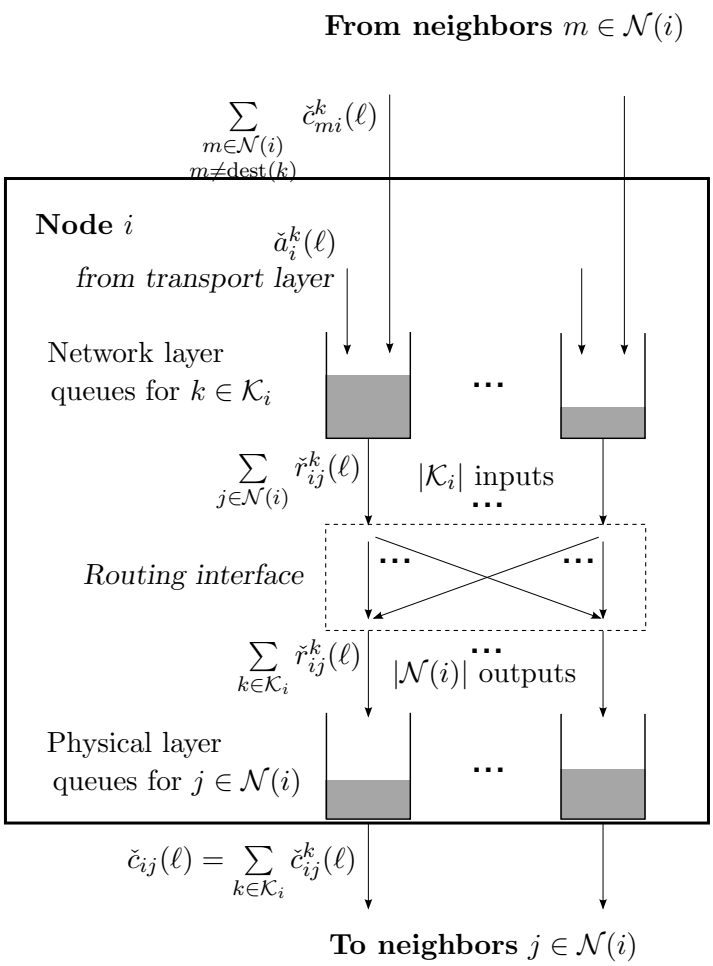


Figure 4.1: Queues at node i and connections to neighbors.

i.e.,

$$\check{r}^k_{ij}(\ell) = r^{k\,*}_{ij}, \quad \ell = 1, 2, \ldots. \tag{4.39}$$

The power allocation $\mathbf{p}^*(\mathbf{h})$ will be used whenever the fading state is $\mathbf{h}$. The physical layer rate $\check{c}_{ij}(\ell)$ at each time slot depends on whether the fading is slow or fast. In the case of slow fading, the channels $\mathbf{h}$ will be (approximately) constant over the time slot, and will have value denoted by $\mathbf{h}(\ell)$. The instantaneous physical layer rate $\check{c}_{ij}(\ell)$ is

$$\check{c}_{ij}(\ell) = \sum_{f\in\mathcal{F}} \ln\bigl[1 + \gamma^f_{ij}(\mathbf{h}(\ell), \mathbf{p}^*(\mathbf{h}(\ell)))\bigr], \quad \ell = 1, 2, \ldots \tag{4.40}$$

where the notation $\gamma^f_{ij}(\mathbf{h}, \mathbf{p}(\mathbf{h}))$ means that the channels have value $\mathbf{h}$ in (4.27), and

the power allocation is $\mathbf{p}(\mathbf{h})$. In the case of fast fading, the channels $\mathbf{h}$ change over the time slot. If the time slot duration is much larger than the coherence time of $\mathbf{h}$, then the instantaneous physical layer rate $\check{c}_{ij}(\ell)$ can be approximated by

$$\check{c}_{ij}(\ell) = \mathbb{E}\left[\sum_{f\in\mathcal{F}} \ln\bigl(1+\gamma_{ij}^{f}(\mathbf{h},\mathbf{p}^{*}(\mathbf{h}))\bigr)\right], \quad \ell = 1,2,\ldots. \tag{4.41}$$

It is possible that the queues do not have as many bits as required by the service processes $\sum_{j\in\mathcal{N}(i)} \check{r}_{ij}^{k}(\ell)$ and $\check{c}_{ij}(\ell)$; then, they simply transmit as many as available. Furthermore, it is important to stress the role of the physical layer queues in the interaction between the routing decisions $\check{r}_{ij}^{k}(\ell)$ and the instantaneous capacity supported by the physical layer $\check{c}_{ij}(\ell)$. These buffers effectively store the bits that cannot be transmitted if the instantaneous capacity $\check{c}_{ij}(\ell)$ drops below the rate that the layer above wants to "pump out", that is, $\sum_{k\in\mathcal{K}_i} \check{r}_{ij}^{k}(\ell)$.

To build intuition about the algorithm, note that the long-term average rates $\frac{1}{M}\sum_{\ell=1}^{M} \check{a}_{i}^{k}(\ell)$ and $\frac{1}{M}\sum_{\ell=1}^{M} \check{r}_{ij}^{k}(\ell)$ converge to the optimal $a_i^{k\,*}$ and $r_{ij}^{k\,*}$. Moreover, the endogenous rates $\check{c}_{ji}^{k}(\ell)$ will satisfy $\sum_{\ell=1}^{M} \check{c}_{ji}^{k}(\ell) \leq \sum_{\ell=1}^{M} \check{r}_{ji}^{k}(\ell)$, because all packets placed endogenously into the network layer queues of node i must have been routed to i from its neighbors. Hence, (4.32b) will be satisfied by the long-term averages of the respective processes; that is, $\lim_{M\to\infty} \frac{1}{M}\sum_{\ell=1}^{M} \check{a}_{i}^{k}(\ell) + \lim_{M\to\infty} \frac{1}{M}\sum_{\ell=1}^{M} \check{c}_{ji}^{k}(\ell) \leq \lim_{M\to\infty} \frac{1}{M}\sum_{\ell=1}^{M} \check{r}_{ij}^{k}(\ell)$. A similar conclusion is true for (4.32c) with $c_{ij} = c_{ij}^{*}$. Recall that $a_i^{k\,*}$, $r_{ij}^{k\,*}$, c_{ij}^{*}, p_i^{*}, $\mathbf{p}^{*}(\mathbf{h})$ are optimal solutions of (4.32). Also, as long as the solution of (4.37e) with the optimal Lagrange multipliers is close to the optimal power allocation, the long-term average link capacities $\frac{1}{M}\sum_{\ell=1}^{M} \check{c}_{ij}(\ell)$ and long-term average power consumptions converge, respectively, to their expected values $\mathbb{E}\bigl[\sum_{f\in\mathcal{F}} \ln\bigl(1+\gamma_{ij}^{f}(\mathbf{h},\mathbf{p}^{*}(\mathbf{h}))\bigr)\bigr]$ and $\mathbb{E}\bigl[\sum_{f\in\mathcal{F}}\sum_{j\in\mathcal{N}(i)} p_{ij}^{f\,*}(\mathbf{h})\bigr]$ under both fading scenarios, since the fading process is stationary and ergodic. Then, (4.32d) and (4.32e) are satisfied with $c_{ij} = c_{ij}^{*}$ and $p_i = p_i^{*}$.

Remark3. The present framework has certain practical limitations, which are to some degree common to several works developing online algorithms for network optimization and control. Specifically, a central controller is not desirable in large or ad hoc networks, yet it is commonly utilized; see e.g., [5,6,39], and references therein. Another limitation is that the network controller needs access to the current fading state without delay, and is again typical; see for instance, the framework of [5] and [39]. Developing efficient

methods for acquiring channel estimates accurately and fast poses a challenge. Furthermore, solving for the power allocation per fading state using (4.37e) may incur high computational burden, even when the machinery of convex optimization is employed, as discussed in Section 4.4. Most online cross-layer resource allocation algorithms deal with a similar task at the physical and link layers; see e.g., [4, 5, 39]. This task is usually referred to as scheduling and is well known to be difficult for several physical and link layer models, such as the one here; approximate solutions that are suboptimal but considerably less computationally demanding are welcome.

4.4 Power allocation at the PHY

This section deals with the solution of (4.37e). Let $\mathbf{p}^f$ denote the power allocations p_{ij}^f of all links over tone f; also define $\mathcal{P}_{\text{mask}}^f := \{\mathbf{p}^f | 0 \leq p_{ij}^f \leq p_{ij\,\max}^f \,\forall\, j \in \mathcal{N}(i),\, \forall\, i\}$. To start, note that (4.37e) can be decomposed into separate problems, one per tone f, i.e.,

$$\max_{\mathbf{p}^f \in \mathcal{P}_{\text{mask}}^f} \sum_{i,j \in \mathcal{N}(i)} \left[\lambda_{ij}(t) \ln\left(1+\gamma_{ij}^f\right) - \mu_i(t) p_{ij}^f\right]. \tag{4.42}$$

Note that t and the Lagrange multipliers are fixed now. Next, certain methods for the solution of (4.42) are described.

Exact methods

Exhaustive search in the space of $\mathbf{p}^f$ can be used in principle to find the global maximizer of (4.42). The major limitation however, is that exhaustive search incurs exponential complexity in the number of variables in $\mathbf{p}^f$. Another approach is to use a branch-and-bound method. Such a method is pursued in [69, Sec. 5], and its adaptation for the solution of (4.42) is straightforward. The worst-case complexity of this method is also exponential, but in practice it may work well, in the sense that it can find the global maximizers with much lower computation requirements than the exhaustive search.

Low-SINR and high-SINR approximations

The low-SINR approximation uses $\ln\left(1+\gamma_{ij}^f\right) \approx \gamma_{ij}^f$ in (4.42). The resulting optimization problem can be written after straightforward manipulations as a signomial program,

which is an intractable non-convex problem (see [37, Sec. 2.2.5] for definitions). Its solution will not be further pursued here, because it requires extensive computation, while it will only be valid for a limited range of SINRs. Moreover, results on the optimality of on-off power allocation in the low-SINR region (see e.g., [40] and references in [4, Sec. III-E]) may not be directly applicable here, because (4.42) is formally different than the problems in those works.

On the other hand, the high-SINR approximation uses $\ln\left(1+\gamma_{ij}^{f}\right) \approx \ln \gamma_{ij}^{f}$ in (4.42). Using the transformation $p_{ij}^{f} = e^{y_{ij}^{f}}$, the problem reduces to a convex one in the variables y_{ij}^{f}. This transformation is standard in geometric programming, as well as in power control problems [37]. In the case of (4.42) however, the resulting problem is not a geometric program. Nonetheless, it can be solved efficiently by any algorithm used for convex problems, e.g., gradient descent or interior-point methods [66]. Moreover, the solution of the high-SINR approximation may be a step within successive approximation methods, which are described next.

Successive convex approximations

The premise here is that a sequence of power allocations $\mathbf{p}^{f}(\tau)$, $\tau \geq 1$, is obtained, where each $\mathbf{p}^{f}(\tau)$ is the solution of an optimization problem that approximates (4.42). Since this problem must be easy to solve, it is constructed to ensure convexity. This is done essentially by substituting the objective function in (4.42) with a convex function. The resulting problem can be solved efficiently by any algorithm used for convex problems, e.g., gradient descent or interior-point methods; or even by an algorithm developed specifically for that problem. Three methods to obtain problems approximating (4.42) are described here, inspired from the DSL literature [42], [43], and the condensation method [37, Sec. 2.2]. In all these cases the function replacing the objective in (4.42) is a lower bound of the objective for all values of p_{ij}^{f}. Moreover, the convex function selected at step τ is constructed using the previous solution $\mathbf{p}^{f}(\tau-1)$. The idea is that the final power allocation obtained from this procedure offers an approximate solution to (4.42). The details of each method are described next.

First, a solution to (4.42) is pursued based on the SCALE algorithm [42]. The lower bound for the objective function in (4.42) is based on the inequality $\alpha_{ij}^{f} \ln \gamma_{ij}^{f} + \beta_{ij}^{f} \leq \ln(1+\gamma_{ij}^{f})$ for all γ_{ij}^{f}, where $\alpha_{ij}^{f} = \check{\gamma}/(1+\check{\gamma})$ for some arbitrary $\check{\gamma} > 0$, and β_{ij}^{f} is properly

selected [42, eq. (2)]. The inequality becomes equality for $\gamma_{ij}^f = \check{\gamma}$. The approximation to (4.42) at step τ amounts to

$$\max_{\mathbf{p}^f \in \mathcal{P}_{\text{mask}}^f} \sum_{i,j \in \mathcal{N}(i)} \left[\lambda_{ij}(t)\big(\alpha_{ij}^f(\tau) \ln \gamma_{ij}^f + \beta_{ij}^f(\tau)\big) - \mu_i(t) p_{ij}^f\right]. \tag{4.43}$$

Because the β_{ij}^f are constants, hereafter they will not be taken into account in problem (4.43). Using the transformation $p_{ij}^f = e^{y_{ij}^f}$, (4.43) can be recast as a convex optimization problem.

The solution of (4.42) based on (4.43) proceeds as follows. In each step $\tau \geq 2$, the weights are computed as $\alpha_{ij}^f(\tau) = \gamma_{ij}^f(\tau-1)/(1+\gamma_{ij}^f(\tau-1))$, where the previous solution $\mathbf{p}^f(\tau-1)$ was used in $\gamma_{ij}^f(\tau-1)$. Then (4.43) is solved exactly with the fixed $\alpha_{ij}^f(\tau)$ to obtain $\mathbf{p}^f(\tau)$; and this sequence of weight updating and solving (4.43) is repeated. This procedure ends when the sequence $\{\mathbf{p}^f(\tau)\}$ obtained as solutions of (4.43) converges, or practically, after a prescribed number of approximations. At $\tau = 1$, $\{\mathbf{p}^f(1)\}$ is obtained from (4.43) with $\alpha_{ij}^f(1) = 1$, which coincides with the high-SINR approximation. Any algorithm can be used for solving (4.43); recall that any locally optimal solution of (4.43) will be globally optimal, due to the equivalence of (4.43) to a convex problem. It is also possible to derive an iterative algorithm tailored specifically for (4.43); see [42], [67] for details.

The successive approximation procedure based on SCALE is summarized in Table 4.1, where Steps 2–7 correspond to the solution of the high-SINR approximation.

The second method to obtain a convex approximation of (4.42) is based on rewriting $\ln\big(1+\gamma_{ij}^f\big)$ as $\ln\big(h_{ij}^f p_{ij}^f + \sigma_j^f + \sum_{(k,l)\in\mathcal{I}_{ij}} h_{kj}^f p_{kl}^f\big) - \ln\left(\sigma_j^f + \sum_{(k,l)\in\mathcal{I}_{ij}} h_{kj}^f p_{kl}^f\right)$, and noting that the second summand with its sign is convex in $\mathbf{p}^f$. But any convex function can be lower bounded by an affine function [66, Sec. 3.1.3]. Specifically, given a fixed $\check{\mathbf{p}}^f$, the following inequality holds for all $\check{\mathbf{p}}^f \geq \mathbf{0}$ for properly selected constants b_{ij}^f [66, eq. (3.2)]:

$$-\ln\left(\sigma_j^f + \sum_{(k,l)\in\mathcal{I}_{ij}} h_{kj}^f p_{kl}^f\right) \geq -\sum_{(k,l)\in\mathcal{I}_{ij}} a_{ij;kl}^f p_{kl}^f + b_{ij}^f, \quad a_{ij;kl}^f = \frac{h_{kj}^f}{\sigma_j^f + \sum_{(m,n)\in\mathcal{I}_{ij}} h_{mj}^f \check{p}_{mn}^f}. \tag{4.44}$$

Using (4.44) and omitting the b_{ij}^f, (4.42) becomes at step τ the following convex problem

Table 4.1: Power allocation algorithm based on SCALE. T is the maximum number of successive approximations after the 1st (high-SINR) approximation ($\mathrm{T} \geq 1$); and in Steps 7 and 13 any algorithm can be used.

1: **for all** tones $f \in \mathcal{F}$ **do**

High-SINR approximation

2: Set $\tau = 1$

3: **for all** nodes $i \in \mathcal{V}$ and neighbors $j \in \mathcal{N}(i)$ **do**

4: Initialize weights $\alpha_{ij}^f(1) = 1$

5: Initialize power allocations $0 < p_{ij}^f(0) \leq p_{ij\,\max}^f$

6: **end for**

7: Solve high-SINR approximation, using $\mathbf{p}^f(0)$ as initialization

$$\mathbf{p}^f(1) \in \arg\max_{\mathbf{p}^f \in \mathcal{P}_{\text{mask}}^f} \sum_{i,j \in \mathcal{N}(i)} \left[\lambda_{ij}(t) \ln \gamma_{ij}^f - \mu_i(t) p_{ij}^f\right]$$

Successive approximations

8: **repeat**

9: $\tau \leftarrow \tau + 1$

10: **for all** nodes $i \in \mathcal{V}$ and neighbors $j \in \mathcal{N}(i)$ **do**

11: Update weights using $\mathbf{p}^f(\tau - 1)$, $\alpha_{ij}^f(\tau) \leftarrow \frac{\gamma_{ij}^f(\tau-1)}{1+\gamma_{ij}^f(\tau-1)}$

12: **end for**

13: Solve the approximation at step τ, using $\mathbf{p}^f(\tau - 1)$ as initialization

$$\mathbf{p}^f(\tau) \in \arg\max_{\mathbf{p}^f \in \mathcal{P}_{\text{mask}}^f} \sum_{i,j \in \mathcal{N}(i)} \left[\lambda_{ij}(t) \alpha_{ij}^f(\tau) \ln \gamma_{ij}^f - \mu_i(t) p_{ij}^f\right]$$

14: **until** $\mathbf{p}^f(\tau)$ converges **or** $\tau > \mathrm{T}$

15: **end for**

in $\mathbf{p}^f$:

$$\max_{\mathbf{p}^f \in \mathcal{P}_{\text{mask}}^f} \sum_{i,j \in \mathcal{N}(i)} \left[\lambda_{ij}(t) \ln\big(h_{ij}^f p_{ij}^f + \sigma_j^f + \sum\nolimits_{(k,l) \in \mathcal{I}_{ij}} h_{kj}^f p_{kl}^f\big) - \sum_{(k,l) \in \mathcal{I}_{ij}} a_{ij;kl}^f(\tau) p_{kl}^f - \mu_i(t) p_{ij}^f\right]. \tag{4.45}$$

Similar to the method based on SCALE, the solution of (4.42) based on (4.44)–(4.45) proceeds as follows. In each step $\tau \geq 1$, the coefficients $a_{ij;kl}^f(\tau)$ are computed based on (4.44) using the previous solution $\mathbf{p}^f(\tau - 1)$. Then problem (4.45) is solved exactly with the fixed $a_{ij;kl}^f(\tau)$ to obtain $\mathbf{p}^f(\tau)$, using any algorithm for convex problems. The procedure is initialized with power allocation $\mathbf{p}^f(0)$ within the spectral mask. The algorithm is summarized in Table 4.2.

Table 4.2: Power allocation algorithm based on affine approximation to the logarithm of noise-plus-interference. T is the maximum number of successive approximations ($\mathrm{T} \geq 1$); and in Step 13 any algorithm for convex programs can be used.

1: **for all** tones $f \in \mathcal{F}$ **do**
2: Set $\tau = 0$
3: **for all** nodes $i \in \mathcal{V}$ and neighbors $j \in \mathcal{N}(i)$ **do**
4: Initialize power allocations $0 \leq p_{ij}^f(0) \leq p_{ij\,\max}^f$
5: **end for**
6: **repeat**
7: $\tau \leftarrow \tau + 1$
8: **for all** nodes $i \in \mathcal{V}$ and neighbors $j \in \mathcal{N}(i)$ **do**
9: **for all** links $(k,l) \in \mathcal{I}_{ij}$ [causing interference to (i,j)] **do**
10: Update coefficients using $\mathbf{p}^f(\tau - 1)$, $a_{ij;kl}^f(\tau) \leftarrow \frac{h_{kj}^f}{\sigma_j^f + \sum_{(m,n)\in\mathcal{I}_{ij}} h_{mj}^f p_{mn}^f(\tau-1)}$
11: **end for**
12: **end for**
13: Solve convex approximation at step τ
$\mathbf{p}^f(\tau) \in \arg\max_{\mathbf{p}^f \in \mathcal{P}_{\mathrm{mask}}^f}$
$$\sum_{i,j\in\mathcal{N}(i)} \left[\lambda_{ij}(t) \ln\left(h_{ij}^f p_{ij}^f + \sigma_j^f + \sum_{(k,l)\in\mathcal{I}_{ij}} h_{kj}^f p_{kl}^f \right) - \sum_{(k,l)\in\mathcal{I}_{ij}} a_{ij;kl}^f(\tau) p_{kl}^f - \mu_i(t) p_{ij}^f \right]$$
14: **until** $\mathbf{p}^f(\tau)$ converges **or** $\tau \geq \mathrm{T}$
15: **end for**

The third method is based on rewriting (4.42) in an equivalent form such that an upper bound of a ratio of posynomials appears as constraint (see [66, p. 160] for definitions). This form is

$$\min_{\mathbf{p}^f, \mathbf{u}^f, s} \quad \prod_{i,j\in\mathcal{N}(i)} (u_{ij}^f)^{\lambda_{ij}(t)} \cdot s \tag{4.46a}$$

$$\text{subj. to} \quad \frac{\sigma_j^f + \sum_{(k,l)\in\mathcal{I}_{ij}} h_{kj}^f p_{kl}^f}{u_{ij}^f \left(h_{ij}^f p_{ij}^f + \sigma_j^f + \sum_{(k,l)\in\mathcal{I}_{ij}} h_{kj}^f p_{kl}^f \right)} \leq 1 \quad \forall j \in \mathcal{N}(i),\ \forall i \tag{4.46b}$$

$$\exp\left(\sum\nolimits_{i,j\in\mathcal{N}(i)} \mu_i(t) p_{ij}^f\right) \leq s \quad \forall j \in \mathcal{N}(i),\ \forall i \tag{4.46c}$$

$$0 \leq p_{ij}^f \leq p_{ij\,\max}^f \quad \forall j \in \mathcal{N}(i),\ \forall i \tag{4.46d}$$

where u_{ij}^f, s are auxiliary variables, and $\mathbf{u}^f$ collects u_{ij}^f for all links. Constraint (4.46b)

is a ratio of posynomials and is non-convex. The approach is to approximate the posynomial in the denominator with a monomial. This is called condensation of the posynomial and constitutes a standard method to convexify constraints which involve an upper bound on the ratio of posynomials [37, Sec. 2.2.5]. Specifically, given fixed $\check{u}_{ij}^f$, $\check{p}_{ij}^f$ and $\check{p}_{kl}^f$ for $(k,l) \in \mathcal{I}_{ij}$, the monomial approximating (in fact lower-bounding for all $p_{ij}^f > 0$, $u_{ij}^f > 0$) the denominator in (4.46b) for link (i,j) is

$$\left(\frac{h_{ij}^f p_{ij}^f u_{ij}^f}{\delta_{ij}^f}\right)^{\delta_{ij}^f} \cdot \left(\frac{\sigma_j^f u_{ij}^f}{\eta_{ij}^f}\right)^{\eta_{ij}^f} \cdot \prod_{(k,l)\in\mathcal{I}_{ij}} \left(\frac{h_{kj}^f p_{kl}^f u_{ij}^f}{\theta_{ij;kl}^f}\right)^{\theta_{ij;kl}^f}, \tag{4.47}$$

$$\delta_{ij}^f = \frac{h_{ij}^f \check{p}_{ij}^f \check{u}_{ij}^f}{\check{\omega}}, \eta_{ij}^f = \frac{\sigma_j^f \check{u}_{ij}^f}{\check{\omega}}, \theta_{ij;kl}^f = \frac{h_{kj}^f \check{p}_{kl}^f \check{u}_{ij}^f}{\check{\omega}}, \check{\omega} = \check{u}_{ij}^f \left(h_{ij}^f \check{p}_{ij}^f + \sigma_j^f + \sum_{(k,l)\in\mathcal{I}_{ij}} h_{kj}^f \check{p}_{kl}^f\right). \tag{4.48}$$

Using the aforementioned approximation, and applying the transformation $p_{ij}^f = e^{y_{ij}^f}$, $u_{ij}^f = e^{v_{ij}^f}$, $s = e^z$, the resulting problem is convex in the variables y_{ij}^f, v_{ij}^f, z (but not a geometric program).

The algorithm to solve (4.42) based on (4.46)–(4.47) proceeds as follows. In each step $\tau \geq 1$, the parameters $\delta_{ij}^f(\tau)$, $\eta_{ij}^f(\tau)$, and $\theta_{ij;kl}^f(\tau)$ are obtained for all links based on (4.48) using the previous solution $\mathbf{p}^f(\tau-1)$, $\mathbf{u}^f(\tau-1)$. Next, problem (4.46) is solved, where the denominator in (4.46b) is first substituted by the monomial in (4.47) with parameters $\delta_{ij}^f(\tau)$, $\eta_{ij}^f(\tau)$, and $\theta_{ij;kl}^f(\tau)$. Note that this problem can be solved by any algorithm for programs which are convex in the variables y_{ij}^f, v_{ij}^f, z. The procedure is initialized with power allocation $\mathbf{p}^f(0)$ within the spectral mask. The algorithm is summarized in Table 4.3.

Remark4. The algorithms presented here involve centralized computation. This is expected, because the objective function and the interference couple the power allocation across terminals [cf. (4.37e) or (4.42)], and a (near-)optimal power allocation is sought. Moreover, while problem (4.42) may in general incur exponential complexity in the number of links, the methods based on successive convex approximations have polynomial complexity.

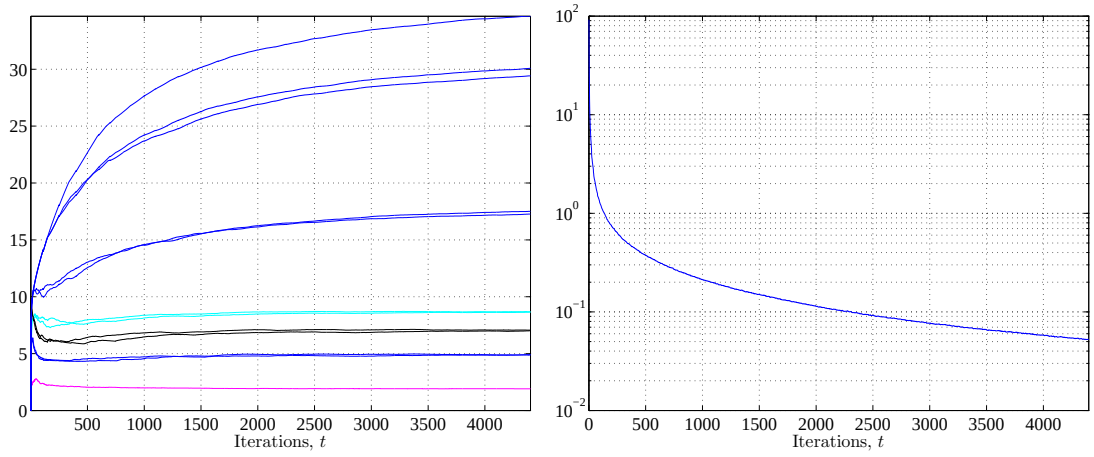
4.5 Numerical tests

The algorithms developed in the previous sections will be tested on a wireless network consisting of 8 terminals placed on a 300m $\times$ 100m area, as shown in Fig. 4.3. In Section 4.5.1, the subgradient algorithm of Section 4.3 is used to obtain near-optimal Lagrange multipliers, and near-optimal primal variables a_i^k, r_{ij}^k, c_{ij}, p_i (offline phase). Then, simulation results from running the network with the strategy of Section 4.3.2 are presented in Section 4.5.2 (online phase).

4.5.1 Near-optimal design

Two different designs are pursued. One is obtained from the subgradient method together with the algorithm based on SCALE (cf. Table 4.1), while the other uses the high-SINR approximation instead. Each terminal in the network is destination of a single commodity ($\text{dest}(k) = k$ for $k = 1, \ldots, 8$); hence, the set of commodities that enter each terminal is $\mathcal{K}_i = \mathcal{V} \backslash \{i\}$. The rest of the parameters are listed in Table 4.4. The values of channel gains and noise have the implication that, even with no interference and transmission power 1 W/Hz, the terminals operate at low to moderate SINR. First, the results obtained from the subgradient method combined with the algorithm based on SCALE are presented; and then, they are compared with those from the high-SINR approximation. In both cases, MATLAB's implementation of sequential quadratic programming [70] was used to solve the approximating problem (4.43). The algorithm based on SCALE (Table 4.1) was run with multiple initializations in Step 5. Specifically, each initialization $\mathbf{p}^f(0)$ had the following form: one entry took the value of the spectral mask, and all other entries took the value -100 dB; this was repeated for every entry. After the algorithm of Table 4.1 had run with those initializations, the solution giving the best objective value in Step 13 was used for the next subgradient step. On the other hand, random initialization was used in Step 5 for the case where just the high-SINR approximation was used for the design; in particular, uniform over the interval $[0, p_{ij\,\max}^f/5]$, independently for each link and tone.

The subgradient method with constant stepsize gives near-optimal Lagrange multipliers, by forming running averages of the dual iterates, $\bar{\mathbf{\Lambda}}(s) := \frac{1}{s}\sum_{t=0}^{s-1} \mathbf{\Lambda}(t)$; and also



(a) (b)

Figure 4.2: (a) Convergence of running averages of Lagrange multipliers corresponding to node 1. Averages $\bar{\nu}_1^k(t)$, $k = 2, \ldots, 8$ are shown in blue; $\bar{\xi}_{1j}(t)$, $j = 2, 3$ in black; $\bar{\lambda}_{1j}(t)$, $j = 2, 3$ in cyan; and $\bar{\mu}_1(t)$ in magenta. (b) Quantity

$$\left[\sum_{i,k \in \mathcal{K}_i} \left(\left[\bar{a}_i^k(t) - \sum_{j \in \mathcal{N}(i)} \bar{r}_{ij}^k(t) + \sum_{j \in \mathcal{N}(i)}^{j \neq \text{dest}(k)} \bar{r}_{ji}^k(t) \right]^+ \right)^2 + \sum_{i,j \in \mathcal{N}(i)} \left(\left[\sum_{k \in \mathcal{K}_i} \bar{r}_{ij}^k(t) - \bar{c}_{ij}(t) \right]^+ \right)^2 \right]^{1/2}$$

capturing the violation of constraints (4.32b) and (4.32c).

near-optimal arrival rates, multicommodity flows, link capacities, and average powers, by forming the corresponding running averages, $\bar{a}_{ij}^k(t)$, $\bar{r}_{ij}^k(t)$, $\bar{c}_{ij}(t)$, and $\bar{p}_i(t)$ (cf. Proposition 4.1). Those are depicted in Figures 4.2, 4.3, 4.4, as explained next.

Fig. 4.2a shows an example of the convergence of the averaged dual iterates; not all Lagrange multipliers can be shown here due to space limitations. It is interesting to note that convergence occurs even with a few Monte Carlo realizations. Moreover, Fig. 4.2b depicts the norm of the violation of the constraints (4.32b) and (4.32c) evaluated at the running averages of the respective variables, which is seen to be decreasing [cf. Proposition 4.1(i)].

Multicommodity flows r_{ij}^k define routes for each commodity. Consider commodity $k = 8$ consisting of traffic transmitted to terminal 8. Terminal 8 is depicted in Fig. 4.3 as a square. The remaining terminals are shown as circles with area proportional to the total (endogenous plus exogenous) incoming rate to that terminal having terminal 8 as destination, i.e., $a_i^8 + \sum_{j \in \mathcal{N}(i), j \neq 8} r_{ji}^8$. Those incoming rates are also given in Fig. 4.3,

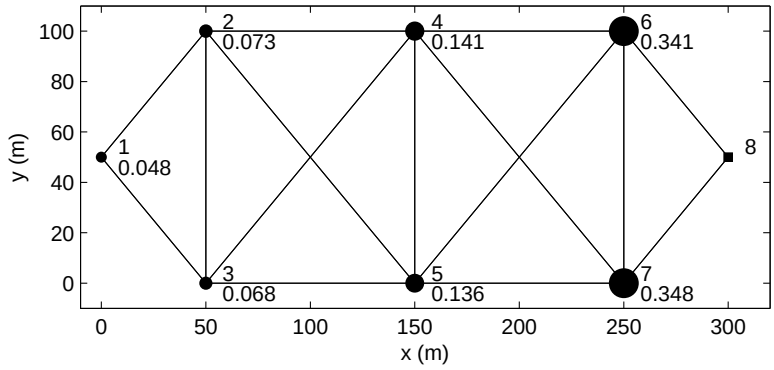


Figure 4.3: A wireless network with its physical dimensions indicated; and incoming rates with node 8 as destination ($a_i^8 + \sum_{j\in\mathcal{N}(i), j\neq 8} r_{ji}^8$), shown in bps/Hz below the node numbers. The area of each circle is proportional to these incoming rates.

and represent the amount of traffic for terminal 8 that each node handles. Observe that packets accumulate as we move from terminal 1 (the furthest from 8) to the closest terminals 6 and 7.

Fig. 4.4 shows link capacities. The width of each arrow is proportional to the average capacity c_{ij} of the corresponding link. The color of each arrow is obtained by linearly mapping the capacities on a scale of gray (see bar in Fig. 4.4), where black corresponds to the largest capacity in the network. Notice that certain links have larger capacity than others. For instance, the links corresponding to the edges of the "squares", e.g., (2,4) and (5,4), have larger capacities than the "diagonal" links such as (3,4). This is expected, because the distance between terminals 3 and 4 is larger than the distance between 2 and 4, or, between 5 and 4.

Consider now comparing the previous results with those obtained when the high-SINR approximation is used for power allocation at the physical layer. Note that (4.43) with $\alpha_{ij}^f = 1$ is solved at each step, but the true rate, $\ln\big(1 + \gamma_{ij}^f\big)$, is used for the subgradient [cf. (4.38c)]. The capacities c_{ij} of all links obtained with the algorithm based on SCALE, given in Fig. 4.4, are larger than the ones obtained using the high-SINR approximation after 5000 subgradient iterations. Specifically, Table 4.5 shows the latter as percentage of the former; this percentage can be as low as 5%. This shows the suboptimality of the high-SINR approximation for the particular design.

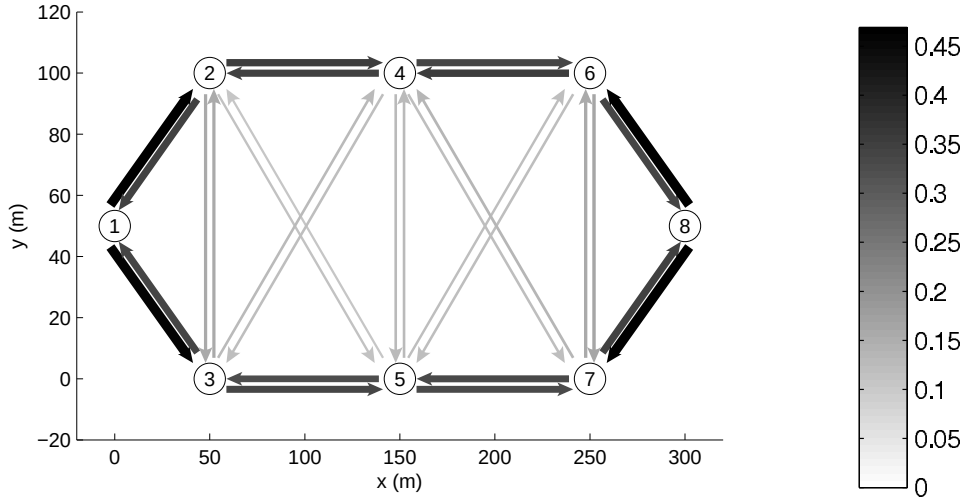


Figure 4.4: Link capacities. The width of each arrow is proportional to the average capacity of the corresponding link. The color of each arrow is obtained by linearly mapping the capacities on a scale of gray (see bar on the right). Black corresponds to the largest capacity in the network. The numbers on the colorbar are in units of bps/Hz.

4.5.2 Network simulation

Simulation results from running the network with the strategy of Section 4.3.2 (online phase) are presented here. The network and its parameters are the same as those in Section 4.5.1, where also quantities needed for the online phase—primal variables and Lagrange multipliers—have been obtained. In what follows, some conventions about the network operation used here are briefly described, and then, the simulation results are presented.

Each node maintains the queues indicated in Fig. 4.1, as described in Section 4.3.2. The numbers of bits entering or leaving the queues are conventionally assumed to take continuous values. This means equivalently that the packet sizes are very small as compared to the number of bits that the network control algorithm specifies to be moved at each slot.

The bits entering the physical layer queues in some slot, can be served in the same slot, if the queue becomes empty within that slot. This happens because these bits

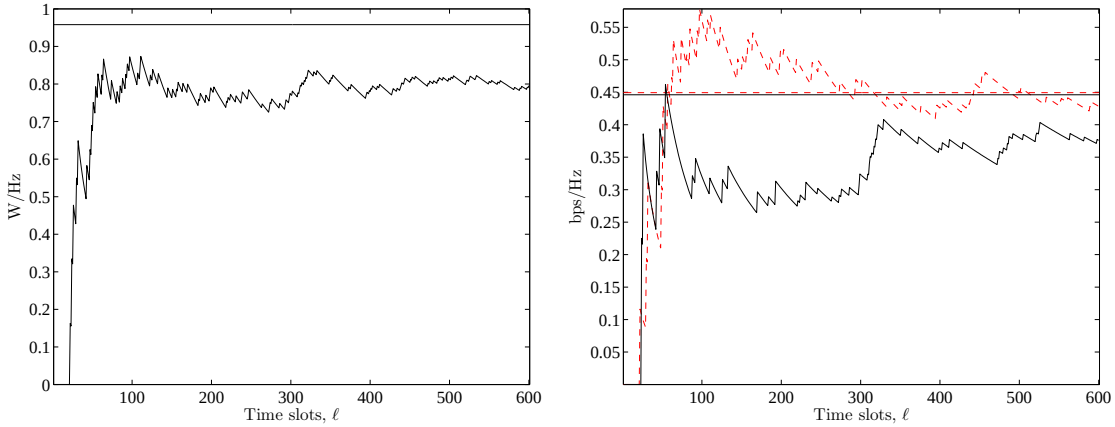
arrive from the network layer queues of the same node. The same convention is adopted for the bits exogenously generated at node i and entering the network layer queues. Now regarding the endogenous arrivals, the bits transmitted during the *previous* slot from the neighbors, can be served in the current time slot.

Hence, at every slot in each node, the exogenous arrivals of that slot and the endogenous arrivals of the previous slot (with destinations other than the present node) enter the network layer queues. Then, the bits leaving the queues are determined, according to $\sum_{j\in\mathcal{N}(i)} \check{r}_{ij}^k(\ell)$, where $\check{r}_{ij}^k(\ell)$ are given by (4.39), and the final values of $\bar{r}_{ij}^k(t)$ (cf. Section 4.5.1) are used in place of ${r_{ij}^k}^*$. If a network layer queue does not have so many bits, a random permutation of the neighbor indices is taken at the routing interface, in order to determine which physical layer queues will receive bits. Then, the number of bits leaving the physical layer queues and transmitted to the neighbors is determined by $\check{c}_{ij}(\ell)$ (how $\check{c}_{ij}(\ell)$ is computed will be explained shortly).

The exogenous arrivals are determined based on ${a_i^k}^*$, obtained as final values of $\bar{a}_i^k(t)$. In particular, the exogenous arrivals $\check{a}_i^k(\ell)$ are i.i.d. across slots, nodes, and commodities, drawn from a uniform distribution. The mean of the distribution is a large fraction of the ${a_i^k}^*$, as given in Table 4.4, and the support is 20% of ${a_i^k}^*$. The reason why the average arrival rates here are not exactly ${a_i^k}^*$, is the following: the values of a_i^k and r_{ij}^k obtained in the previous subsection are not such that (4.32b) is exactly feasible; see Fig. 4.2b. More iterations of the subgradient method would suffice so that even higher arrival rates than those of Table 4.4 can be used.

The fading is slow, i.e., the channel vector keeps the same value, $\mathbf{h}(\ell)$, for the duration of the slot. The fading process is Rayleigh, i.i.d. across slots and independent across nodes, following the distribution given in Table 4.4. The instantaneous power allocation for all links is determined based on $\mathbf{h}(\ell)$. The solution of (4.42) is used here, denoted as $\mathbf{p}(\mathbf{h}(\ell);\ell)$, where the final values of $\bar{\mathbf{\Lambda}}(s)$ are the Lagrange multipliers in (4.42), the same for all ℓ. For the solution of (4.42), the algorithm based on SCALE is used, with parameters given in the last rows of Table 4.4. Once $\mathbf{p}(\mathbf{h}(\ell);\ell)$ is determined, $\check{c}_{ij}(\ell)$ is determined by (4.40) using $\mathbf{p}(\mathbf{h}(\ell);\ell)$ instead of $\mathbf{p}^*(\mathbf{h}(\ell))$.

Fig. 4.5a shows the long-term average power consumed at node 1, computed as follows. Let $\check{p}_i(\ell) := \sum_{j,f} p_{ij}^f(\mathbf{h}(\ell);\ell)$ be the power consumed at node i at slot ℓ. The long-term average power consumed at node i is then $\frac{1}{\ell}\sum_{l=1}^{\ell} \check{p}_i(l)$. Fig. 4.5a shows



(a) (b)

Figure 4.5: (a) Long-term average power consumed at node 1. The horizontal line shows the optimal average power. The other trajectory corresponds to $\frac{1}{\ell}\sum_{l=1}^{\ell}\check{p}_i(l)$. (b) Long-term average capacity of links between node 1 and its neighbors. The horizontal lines show respectively the optimal link capacities of link (1,2) (dotted line) and link (1,3) (solid line); those are also depicted in Fig. 4.4. The other trajectories correspond to $\frac{1}{\ell}\sum_{l=1}^{\ell}\check{c}_{12}(l)$ and $\frac{1}{\ell}\sum_{l=1}^{\ell}\check{c}_{13}(l)$.

that this is smaller than the average power computed by the subgradient method (final value of $\bar{p}_i(t)$). In other words, the network control strategy keeps the average power consumption approximately at its designed level. The discrepancy shown in Fig. 4.5a can be explained, if we recall that the Lagrange multipliers used for the computation of $\mathbf{p}(\mathbf{h}(\ell);\ell)$ are not exactly optimal, but rather obtained from the subgradient method with *constant* stepsize, which enures convergence within a ball of the optimal ones. Finally, similar comments hold for the long-term link capacities, shown in Fig. 4.5b.

Table 4.3: Power allocation algorithm based on condensation. T is the maximum number of successive approximations (T $\geq$ 1); and in Step 16 any algorithm for convex programs can be used.

1: **for all** tones $f \in \mathcal{F}$ **do**
2: Set $\tau = 0$
3: **for all** nodes $i \in \mathcal{V}$ and neighbors $j \in \mathcal{N}(i)$ **do**
4: Initialize power allocations $0 \leq p_{ij}^f(0) \leq p_{ij\,\max}^f$
5: Initialize auxiliary variable $u_{ij}^f(0)$ such that $\frac{\sigma_j^f + \sum_{(k,l)\in\mathcal{I}_{ij}} h_{kj}^f p_{kl}^f(0)}{u_{ij}^f(0)\left(h_{ij}^f p_{ij}^f(0) + \sigma_j^f + \sum_{(k,l)\in\mathcal{I}_{ij}} h_{kj}^f p_{kl}^f(0)\right)} \leq 1$
6: **end for**
7: Initialize auxiliary variable $s(0)$ such that $\exp\left(\sum_{i,j\in\mathcal{N}(i)} \mu_i(t) p_{ij}^f(0)\right) \leq s(0)$
8: **repeat**
9: $\tau \leftarrow \tau + 1$
10: **for all** nodes $i \in \mathcal{V}$ and neighbors $j \in \mathcal{N}(i)$ **do**
11: Update parameters δ_{ij}^f, η_{ij}^f of monomial approximation using $\mathbf{p}^f(\tau-1)$ and $\mathbf{u}^f(\tau-1)$

$$\delta_{ij}^f(\tau) \leftarrow \frac{h_{ij}^f p_{ij}^f(\tau-1) u_{ij}^f(\tau-1)}{u_{ij}^f(\tau-1)\left(h_{ij}^f p_{ij}^f(\tau-1) + \sigma_j^f + \sum_{(k,l)\in\mathcal{I}_{ij}} h_{kj}^f p_{kl}^f(\tau-1)\right)}$$

$$\eta_{ij}^f(\tau) \leftarrow \frac{\sigma_j^f u_{ij}^f(\tau-1)}{u_{ij}^f(\tau-1)\left(h_{ij}^f p_{ij}^f(\tau-1) + \sigma_j^f + \sum_{(k,l)\in\mathcal{I}_{ij}} h_{kj}^f p_{kl}^f(\tau-1)\right)}$$

12: **for all** links $(k,l) \in \mathcal{I}_{ij}$ [causing interference to (i,j)] **do**
13: Update parameters $\theta_{ij;kl}^f$ of monomial approximation using $\mathbf{p}^f(\tau-1)$, $\mathbf{u}^f(\tau-1)$

$$\theta_{ij;kl}^f(\tau) \leftarrow \frac{h_{kj}^f p_{kl}^f(\tau-1) u_{ij}^f(\tau-1)}{u_{ij}^f(\tau-1)\left(h_{ij}^f p_{ij}^f(\tau-1) + \sigma_j^f + \sum_{(k,l)\in\mathcal{I}_{ij}} h_{kj}^f p_{kl}^f(\tau-1)\right)}$$

14: **end for**
15: **end for**
16: Solve convex approximation in the variables $y_{ij}^f = \ln p_{ij}^f$, $v_{ij}^f = \ln u_{ij}^f$, $z = \ln s$ at step τ to obtain $\mathbf{p}^f(\tau)$

$$\min_{\mathbf{p}^f,\mathbf{u}^f,s} \prod_{i,j\in\mathcal{N}(i)} (u_{ij}^f)^{\lambda_{ij}(t)} + s$$

$$\text{subj. to } \frac{\left(\sigma_j^f + \sum_{(k,l)\in\mathcal{I}_{ij}} h_{kj}^f p_{kl}^f\right)\left(\delta_{ij}^f(\tau)\right)^{\delta_{ij}^f(\tau)}\left(\eta_{ij}^f(\tau)\right)^{\eta_{ij}^f(\tau)} \prod_{(k,l)\in\mathcal{I}_{ij}} \left(\theta_{ij;kl}^f(\tau)\right)^{\theta_{ij;kl}^f(\tau)}}{\left(h_{ij}^f p_{ij}^f u_{ij}^f\right)^{\delta_{ij}^f(\tau)}\left(\sigma_j^f u_{ij}^f\right)^{\eta_{ij}^f(\tau)} \prod_{(k,l)\in\mathcal{I}_{ij}} \left(h_{kj}^f p_{kl}^f u_{ij}^f\right)^{\theta_{ij;kl}^f(\tau)}} \leq 1$$

$$\exp\left(\sum\nolimits_{i,j\in\mathcal{N}(i)} \mu_i(t) p_{ij}^f\right) \leq s$$

$$0 \leq p_{ij}^f \leq p_{ij\,\max}^f \quad \forall j \in \mathcal{N}(i),\ \forall i$$

17: **until** $\mathbf{p}^f(\tau)$ converges **or** $\tau \geq$ T
18: **end for**

Table 4.4: Parameters used in the numerical tests

Network parameters and utility functions	
F	1
h_{ij}^f	Exponential with mean $\bar{h}_{ij}^f = 0.1(d_{ij}/d_0)^{-2}$ for all (i,j) and f; $d_0 = 20\,\text{m}$, d_{ij} = distance between nodes i and j in meters; links are reciprocal, i.e., $h_{ij}^f = h_{ji}^f$ for all (i,j) and f
h_{jj}^f	10 dB, deterministic for all j and f
σ_j^f	$\min_{i\in\mathcal{N}(j)} \bar{h}_{ij}^f$ (W/Hz) for all j and f
$p_{i\,\max}$	5 W/Hz for all i
$p_{ij\,\max}^f$	5 W/Hz for all (i,j) and f
$a_{i\,\max}^k$	14 bps/Hz for all i and $k \in \mathcal{K}_i$
$a_{i\,\min}^k$	$1.96 \cdot 10^{-4}$ bps/Hz for all i and $k \in \mathcal{K}_i$
$c_{ij\,\max}$	$\max_{i,j\in\mathcal{N}(i)} \breve{C}_{ij}$ for all (i,j) where $\breve{C}_{ij}$ is the max. capacity of link (i,j) without interference; obtained via waterfilling under $\mathbb{E}\big[\sum_f p_{ij}^f(\mathbf{h})\big] \le p_{i\,\max}$
$r_{i\,\max}$	$c_{ij\,\max}$ for all i
$U_i^k(a_i^k)$	$\ln a_i^k$ for all i and $k \in \mathcal{K}_i$
$V_i(p_i)$	p_i^2 for all i
Parameters for design based on SCALE	
β_t	1
R	3 independent realizations
T	5, up to iteration 2000; 6, thereafter
Algorithm for (4.43)	MATLAB's Sequential Quadratic Programming default parameters, except `TolFun` $= 10^{-2}$, `TolCon` $= 10^{-5}$ lower bound 10^{-10} was used for all variables
Parameters for design based on high-SINR approximation	
β_t	1
R	3 independent realizations
Algorithm for (4.43)	same as above
Simulation parameters	
Average arrival rates (% of a_i^{k*} for (i,k))	96% for (1,2), (1,3), (2,1), (2,3), (3,1), (3,2); 73% for (1,8) 90% for (1,4), (1,5), (4,1), (5,1); 77% for (1,6), (1,7); 93% for (2,4), (4,2), (3,5), (5,3); 92% for (2,5), (3,4), (4,3), (5,2); 80% for (2,6), (2,7), (3,6), (6,2), (6,3), (7,2); 95% for (4,5), (5,4); 75% for (6,1), (7,1); similar for pairs not listed, by the symmetry of the network
Algorithm for (4.43)	same as above
T	6

Table 4.5: Certain link capacities achieved with high-SINR approximation as percentage of those achieved with the algorithm based on SCALE. The final values of $\bar{c}_{ij}(t)$ (weighted running averages) are used for the computation. Values are similar for links symmetric to those shown.

(1,2)	(4,2)	(4,3)	(4,5)
3.9	5.2	14.7	14.4
(2,1)	(2,3)	(2,4)	(2,5)
5.2	11.7	5.2	15.9

Chapter 5

Summary and Future Directions

This thesis presented algorithms for optimal demand response in electrical power networks and cross-layer resource allocation in wireless networks. The common theme was the use of Lagrangian duality in order to come up with efficient algorithms.

Demand response amounts to adapting the end users' power consumption to time-based pricing signals. The focus was on residential end users. The first DR model was load control among multiple residences and the electricity provider. Two different types of residential devices offering load response are considered. The first type must consume a specified amount of energy over a prescribed horizon, but the consumption can be adjusted across the horizon. The second type does not have a total energy requirement, but operation of the device provides satisfaction to the end user, resulting in elastic power consumption. Dual decomposition yielded a distributed algorithm. The utility company and the end users exchange through the AMI Lagrange multipliers and hourly consumption data in order to converge to the optimal schedule. The algorithm was shown to find near-optimal schedules even when AMI messages are lost.

The latter result followed by modeling the algorithm as a subgradient method for separable convex programming that utilizes outdated multipliers. Convergence of the dual and the primal iterates was established for constant and diminishing stepsize rules. Strict convexity is not needed for convergence, but running averages of the primal iterates are formed. The main condition enabling convergence was that the delay between the current multiplier iterate and the multipliers used to obtain the subgradients for the updates is bounded.

The second DR model pertained to energy consumption scheduling for a single residential end user, when the electricity cost structure is known ahead of the horizon. Again, the end user has a "must-run" load, and two types of adjustable loads. The first type must consume a specified total amount of energy over the scheduling horizon, but the consumption can be adjusted across the horizon. The second type of load has adjustable power consumption without a total energy requirement, but operation of the load at reduced power results in dissatisfaction of the end user. The main feature of the second formulation is that each adjustable load is interruptible in the sense that the load can be either operated (resulting in nonzero power consumption), or not (resulting in zero power consumption). The on/off feature of the problem results in nonconvexity. Two issues regarding duality of this problem are positively resolved. First, the problem is shown to have zero duality gap, if it is formulated over a continuous time interval. Second, a regular discretized version of the problem is shown to have vanishing duality gap as the partition size of the continuous-time interval grows. Numerical experiments illustrate the usefulness of the subgradient method as a solver leveraging these results.

The zero duality gap was in fact proved for a generic optimization problem which involves two types of variables: a vector and a function. Convexity with respect to the vector variable is assumed. As long as the function variable appears in a function under an integral and weighted by a continuous distribution, the possible nonconvexity is "smoothed-out," and the duality gap is zero. The subgradient method can be used to solve the dual problem. It was shown that forming running averages of the vector variable iterates produced by the subgradient method can yield a near-optimal solution to the problem.

As regards wireless networking, this thesis dealt with algorithm development for cross-layer design of wireless networks in the presence of fading. The network terminals at the physical layer treat interference as noise, and therefore, the instantaneous link capacities are given by the classical $\log(1 + \text{SINR})$ model. The problem variables are long-term end-to-end rates, multicommodity flows (routing variables), link capacities, average power consumption, and power allocation across links and tones that is a function of the fading state. The problem has zero duality gap, and a subgradient algorithm was developed for the solution of the dual problem. The subgradient method involves power allocation at the physical layer, which is a nonconvex problem. Different

power allocation options were developed relying on successive convex approximations, and seamlessly integrated into the subgradient method. Furthermore, optimal design variables of the network, that is end-to-end rates, multicommodity flows, link capacities and average powers, were obtained. A network control strategy utilizing the optimal solution of the cross-layer design problem was also outlined, along with simulation results.

The work presented in this thesis only addresses certain aspects of resource management in power and communication networks. It can be extended in several directions, listed below.

- **Cooperative demand response for end users with distributed generation.** Part of the smart grid vision is to bring generation closer to the end user. It is advocated that end users will have small-scale generation, which can come from conventional sources (e.g., diesel generators, fuel cells) or from renewable sources (e.g., solar, wind). This idea introduces additional scheduling variables to the system. It is also possible that distributed generation is not only used to provide (active) power to the end users, but also to provide services to the distribution systems such as voltage regulation. Moreover, the random and unpredictable nature of the renewable resources calls for robust optimization formulations.
- **Energy consumption scheduling combining interruptible as well as deferrable tasks.** It has already been demonstrated in Chapter 3 that scheduling with interruptible tasks poses significant challenges, and calls for algorithms that leverage the "hidden convexity" of the problem. An additional pertinent feature is to account for tasks that are deferrable, where the starting time is a variable, but once started, these tasks cannot be interrupted and must be completed. Such device profiles give rise to integer programming constraints as well, and therefore efficient algorithms are needed.
- **Distributed power allocation algorithms at the physical layer in wireless ad hoc networks.** The successive convex approximation algorithms in Chapter 4 require centralized computation. But ad hoc networks call for distributed algorithms which do not rely on a central controller in order to compute efficient power allocations. Obtaining near-optimal solutions to the physical layer subproblem in a distributed fashion while accounting for the coupling induced by

interference, is a challenging task.

- **Distributed resource management algorithms relying on incremental subgradient methods or multiple subgradients (e.g., heavy-ball or cutting-plane methods).** Such algorithms have the potential to offer faster convergence than the subgradient methods of Chapters 2 and 4. The issues are distributed implementation, performance under asynchronous operation, and recovery of the primal variables, in particular for nonconvex problems as those studied in Chapter 4.

References

[1] W. A. Wulf. Great achievements and grand challenges. *The Bridge*, 30(3/4):5–10, Fall 2010.

[2] The smart grid: An introduction. U.S. Department of Energy, 2008. `http://www.oe.energy.gov/SmartGridIntroduction.htm`.

[3] Smart grid basics. Federal Smart Grid Task Force, July 14, 2010. `http://www.smartgrid.gov/basics`.

[4] X. Lin, N. B. Shroff, and R. Srikant. A tutorial on cross-layer optimization in wireless networks. *IEEE J. Sel. Areas Commun.*, 24(8):1452–1463, Aug. 2006.

[5] L. Georgiadis, M. J. Neely, and L. Tassiulas. Resource allocation and cross-layer control in wireless networks. *Foundations and Trends in Networking*, 1(1):1–144, 2006.

[6] M. Chiang, S. H. Low, A. R. Calderbank, and J. C. Doyle. Layering as optimization decomposition: A mathematical theory of network architectures. *Proc. of the IEEE*, 95(1):255–312, Jan. 2007.

[7] S. Shakkottai and R. Srikant. Network optimization and control. *Foundations and Trends in Networking*, 2(3):279–379, 2007.

[8] C. W. Gellings and J. H. Chamberlin. *Demand-Side Management: Concepts and Methods*. The Fairmont Press, Lidburn, GA, 1988.

[9] Demand response. U.S. Department of Energy, Office of Electricity Delivery and Energy Reliability. `http://www.oe.energy.gov/demand.htm`.

[10] International Energy Agency. *The Power to Choose: Demand Response in Liberalised Electricity Markets*. OECD Publishing, Paris, France, 2003.

[11] Grid 2030: A national vision for electricity's second 100 years. U.S. Department of Energy, Office of Electric Transmission and Distribution, July 2003. `http://www.oe.energy.gov/DocumentsandMedia/Electric_Vision_Document.pdf`.

[12] A. M. Giacomoni, S. M. Amin, and B. F. Wollenberg. Reconfigurable interdependent infrastructure systems: Advances in distributed sensing, modeling, and control. In *Proc. American Control Conf.*, San Francisco, CA, June–July 2011.

[13] P. Samadi, A.-H. Mohsenian-Rad, R. Schober, V. W. S. Wong, and J. Jatskevich. Optimal real-time pricing algorithm based on utility maximization for smart grid. In *Proc. 1st IEEE Intl. Conf. Smart Grid Communications*, pages 415–420, Gaithersburg, MD, Oct. 2010.

[14] L. Chen, N. Li, S. H. Low, and J. C. Doyle. Two market models for demand response in power networks. In *Proc. 1st IEEE Intl. Conf. Smart Grid Communications*, pages 397–402, Gaithersburg, MD, Oct. 2010.

[15] A.-H. Mohsenian-Rad, V. S. W. Wong, J. Jatskevich, R. Schober, and A. Leon-Garcia. Autonomous demand side management based on game-theoretic energy consumption scheduling for the future smart grid. *IEEE Trans. Smart Grid*, 1(3):320–331, Dec. 2010.

[16] S. Caron and G. Kesidis. Incentive-based energy consumption scheduling algorithms for the smart grid. In *Proc. 1st IEEE Intl. Conf. Smart Grid Communications*, pages 391–396, Gaithersburg, MD, Oct. 2010.

[17] P. Vytelingum, T. D. Voice, S. D. Ramchurn, A. Rogers, and N. R. Jennings. Agent-based micro-storage management for the smart grid. In *Proc. 9th Int. Conf. Autonomous and Multiagent Systems*, pages 39–46, Toronto, Canada, May 2010.

[18] B. Daryanian, R. E. Bohn, and R. D. Tabors. Optimal demand-side response to electricity spot prices for storage-type customers. *IEEE Trans. Power Syst.*, 4(3):897–903, Aug. 1989.

[19] L. Exarchakos, M. Leach, and G. Exarchakos. Modelling electricity storage systems management under the influence of demand-side management programmes. *Int. J. Energy Res.*, 33(1):62–76, Jan. 2009.

[20] X. Guan, X. Zu, and Q.-S. Jia. Energy-efficient buildings facilitated by microgrid. *IEEE Trans. Smart Grid*, 1(3):243–252, Dec. 2010.

[21] T. Larsson, M. Patriksson, and A.-B. Strömberg. Ergodic, primal convergence in dual subgradient schemes for convex programming. *Math. Programming*, 86(2):283–312, 1999.

[22] A. Nedić and A. Ozdaglar. Approximate primal solutions and rate analysis for dual subgradient methods. *SIAM J. Optim.*, 19(4):1757–1780, 2009.

[23] T. Larsson, M. Patriksson, and A.-B. Strömberg. On the convergence of conditional ε-subgradient methods for convex programs and convex-concave saddle-point problems. *Eur. J. Oper. Res.*, 151(3):461–473, 2003.

[24] K. C. Kiwiel and P. O. Lindberg. Parallel subgradient methods for convex optimization. In D. Butnariu, Y. Censor, and S. Reich, editors, *Inherently Parallel Algorithms in Feasibility and Optimization*, pages 335–344. Elsevier Science B.V., Amsterdam, Netherlands, 2001.

[25] F. C. Schweppe, B. Daryanian, and R. D. Tabors. Algorithms for a spot price responding algorithm. *IEEE Trans. Power Systems*, 4(2):507–516, May 1989.

[26] A.-H. Mohsenian-Rad and A. Leon-Garcia. Optimal residential load control with price prediction in real-time electricity pricing environments. *IEEE Trans. Smart Grid*, 1(2):120–133, Sept. 2010.

[27] A. J. Conejo, J. M. Morales, and L. Baringo. Real-time demand response model. *IEEE Trans. Smart Grid*, 1(3):236–242, Dec. 2010.

[28] I. Koutsopoulos and L. Tassiulas. Control and optimization meet the smart power grid: Scheduling of power demands for optimal energy management. 2010. `http://arxiv.org/abs/1008.3614`.

[29] S. Hatami and M. Pedram. Minimizing the electricity bill of cooperative users under a quasi-dynamic pricing model. In *Proc. 1st IEEE Int. Conf. Smart Grid Communications*, pages 421–426, Gaithersburg, MD, Oct. 2010.

[30] L. D. Ha, S. Ploix, E. Zamai, and M. Jacomino. Tabu search for the optimization of household energy consumption. In *Proc. IEEE Int. Conf. Information Reuse and Integration*, pages 86–92, Waikoloa Village, HI, Sept. 2006.

[31] M. A. A. Pedrasa, T. D. Spooner, and I. F. MacGill. Coordinated scheduling of residential distributed energy resources to optimize smart home energy services. *IEEE Trans. Smart Grid*, 1(2):134–143, Sept. 2010.

[32] D. O'Neill, M. Levorato, A. Goldsmith, and U. Mitra. Residential demand response using reinforcement learning. In *Proc. 1st IEEE Int. Conf. Smart Grid Communications*, pages 409–414, Gaithersburg, MD, Oct. 2010.

[33] D. Blackwell. On a theorem of Lyapunov. *Ann. Math. Statist.*, 22(1):112–114, Mar. 1951.

[34] Z.-Q. Luo and S. Zhang. Dynamic spectrum management: Complexity and duality. *IEEE J. Sel. Topics Signal Process.*, 2(1):57–73, Feb. 2008.

[35] A. Ribeiro and G. B. Giannakis. Separation theorems of wireless networking. *IEEE Trans. Inf. Theory*, 56(9):4488–4505, Sep. 2010.

[36] K. Rajawat, N. Gatsis, and G. B. Giannakis. Cross-layer designs in coded wireless fading networks with multicast. *IEEE/ACM Trans. Netw.*, 19(5):1276–1289, Oct. 2011.

[37] M. Chiang. Geometric programming for communication systems. *Foundations and Trends in Communications and Information Theory*, 2(1/2):1–156, 2005.

[38] Y. Xi and E. M. Yeh. Node-based optimal power control, routing, and congestion control in wireless networks. *IEEE Trans. Inf. Theory*, 54(9):4081–4106, Sep. 2008.

[39] M. J. Neely, E. Modiano, and C. E. Rohrs. Dynamic power allocation and routing for time-varying wireless networks. *IEEE J. Select. Areas Commun.*, 23(1):89–103, Jan. 2005.

[40] B. Radunovic and J.-Y. Le Boudec. Power control is not required for wireless networks in the linear regime. In *Proc. 6th IEEE Int. Symp. World of Wireless Mobile and Multimedia Networks (WoWMoM)*, pages 417–427, Taormina, Italy, Jun. 2005.

[41] S. Hayashi and Z.-Q. Luo. Spectrum management for interference-limited multiuser communication systems. *IEEE Trans. Inf. Theory*, 55(3):1153–1175, Mar. 2009.

[42] J. Papandriopoulos and J. S. Evans. SCALE: A low complexity distributed protocol for spectrum balancing in multiuser DSL networks. *IEEE Trans. Inf. Theory*, 55(8):3711–3724, Aug. 2009.

[43] P. Tsiaflakis, M. Diehl, and M. Moonen. Distributed spectrum management algorithms for multiuser DSL networks. *IEEE Trans. Signal Process.*, 56(10):4825–4843, Oct. 2008.

[44] N. Gatsis and G. B. Giannakis. Cooperative multi-residence demand response scheduling. In *Proc. 45th Conf. Information Sciences and Systems*, Baltimore, MD, Mar. 2011.

[45] N. Gatsis and G. B. Giannakis. Residential load control: Distributed scheduling and convergence with lost AMI messages. *IEEE Trans. Smart Grid*, 3(2):770–786, June 2012.

[46] D. P. Bertsekas. *Nonlinear Programming.* Athena Scientific, Belmont, MA, 1999.

[47] D. P. Bertsekas, A. Nedić, and A. Ozdaglar. *Convex Analysis and Optimization.* Athena Scientific, Belmont, MA, 2003.

[48] A. Nedić and D. P. Bertsekas. Incremental subgradient methods for nondifferentiable optimization. *SIAM J. Optim.*, 12(1):109–138, 2001.

[49] K. C. Kiwiel. Convergence of approximate and incremental subgradient methods for convex optimization. *SIAM J. Optim.*, 14(3):807–840, 2004.

[50] W. Rudin. *Principles of Mathematical Analysis.* McGraw-Hill, New York, NY, 3rd edition, 1976.

[51] A. J. Wood and B. F. Wollenberg. *Power Generation, Operation, and Control*. John Wiley & Sons, New York, NY, 2nd edition, 1996.

[52] H. Lee Willis. *Power Distribution Planning Reference Book*. CRC Press, New York, NY, 2nd edition, 2004.

[53] Charging plug-in hybrid and all-electric vehicles at home. U.S. Department of Energy, Office of Energy Efficiency and Renewable Energy, Alternative Fuels and Advanced Vehicles Data Center, Jan. 28, 2011. `http://avt.inel.gov/pdf/phev/phevInfrastructureReport08.pdf`.

[54] K. Morrow, D. Karner, and J. Francfort. Plug-in hybrid electric vehicle charging infrastructure review. Final Report, Battelle Energy Alliance, Contract No. 58517, Nov. 2008. `http://avt.inel.gov/pdf/phev/phevInfrastructureReport08.pdf`.

[55] Appliance Energy Data. Federal Trade Commission. `http://www.ftc.gov/appliancedata`.

[56] End-use consumption of electricity 2001. U.S. Department of Energy, U.S. Energy Information Administration, Apr. 20, 2009. `http://www.eia.gov/emeu/recs/recs2001/enduse2001/enduse2001.html`.

[57] IEEE Distribution Planning Working Group Report. Radial distribution test feeders. *IEEE Trans. Power Syst.*, 6(3):975–985, Aug. 1991.

[58] A. V. Fiacco. *Introduction to Sensitivity and Stability Analysis in Nonlinear Programming*. Academic Press, New York, NY, 1983.

[59] E. Castillo, A. Conejo, R. Minguez, and C. Castillo. A closed formula for local sensitivity analysis in mathematical programming. *Engineering Optimization*, 38(1):93–112, Jan. 2006.

[60] N. Gatsis and G. B. Giannakis. Residential demand response with interruptible tasks: Duality and algorithms. In *Proc. 50th IEEE Conf. Decision and Control and European Control Conf.*, pages 1–6, Orlando, FL, Dec. 2011, invited.

[61] N. Dinculeanu. *Vector Measures*. Pergamon Press, Oxford, U.K., 1967.

[62] D. P. Bertsekas. *Constrained Optimization and Lagrange Multiplier Methods*. Athena Scientific, Belmont, MA, 1996.

[63] Z.-Q. Luo and S. Zhang. Duality gap estimation and polynomial time approximation for optimal spectrum management. *IEEE Trans. Signal Process.*, 57(7):2675–2689, July 2009.

[64] X. Wang, S. Zhang, and D. Y. Yao. Separated continuous conic programming: Strong duality and an approximation algorithm. *SIAM J. Control Optim.*, 48(4):2118–2138, 2009.

[65] A. Goldsmith. *Wireless Communications*. Cambridge University Press, New York, NY, 2005.

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

[67] N. Gatsis, A. Ribeiro, and G. B. Giannakis. Cross-layer optimization of wireless fading ad-hoc networks. In *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing*, pages 2353–2356, Taipei, Taiwan, Apr. 2009.

[68] N. Gatsis, A. Ribeiro, and G.B. Giannakis. A class of convergent algorithms for resource allocation in wireless fading networks. *IEEE Trans. Wireless Commun.*, 9(5):1808–1823, May 2010.

[69] P. Tsiaflakis, J. Vangorp, M. Moonen, and J. Verlinden. A low complexity optimal spectrum balancing algorithm digital subscriber lines. *Signal Processing*, 87(7):1735–1753, Jul. 2007.

[70] Constrained nonlinear optimization. MATLAB Documentation, The MathWorks. `www.mathworks.com/access/helpdesk/help/toolbox/optim/ug/brnoxzl.html`.