# DISTRIBUTED OPTIMIZATION OF COUPLED SYSTEMS WITH APPLICATIONS TO NETWORK UTILITY MAXIMIZATION

*Chee Wei Tan, Daniel P. Palomar, and Mung Chiang*

Electrical Engineering Department, Princeton University, NJ 08544

**ABSTRACT**

In Network Utility Maximization (NUM) problems, it is generally assumed that user utilities are uncoupled, i.e., each utility depends only on local variables. Then the coupling in constraint functions among users sharing common resources can be decoupled by standard methods such as dual decomposition. However, in problems where cooperation or competition is modeled through the objective function, such as rate allocation in clustered system and power control in interference limited system, each utility may depend not only on its local variables but also on the local variables of other utilities. Applications of this coupled utility model include wireless power control and DSL spectrum management, where the utilities are functions of the Signal-to-Interference Ratios (SIR) that depend on the transmit powers of other users. We present a systematic approach of consistency pricing to decouple NUM problems with coupled utilities, obtaining distributed algorithms that efficiently handle couplings in utilities with two alternative timescales, as well as a method to reduce message passing overhead in the case of interference-based coupling.

## 1. INTRODUCTION

An important approach to design a network system is by formulating the design as the aggregate maximization of the utilities of all the nodes subject to physical and economic constraints in the network. This is referred to as Network Utility Maximization (NUM). For example, consider a communication network with $L$ links, each with a fixed capacity of $c_l$ bits per transmission, and $S$ sources or nodes, each transmitting at a source rate of $x_s$ bits per transmission. Each source $s$ emits one flow, using a fixed set of links $L(s)$ in its path, and has a utility function $U_s(x_s)$. The basic version of NUM is the problem of maximizing the total utility $\sum_s U_s(x_s)$, over the source rates $\mathbf{x}$, subject to linear flow constraints $\sum_{s:l\in L(s)} x_s \leq c_l$ for all links $l$ [1]:

$$\begin{array}{ll} \underset{\mathbf{x}\geq 0}{\text{maximize}} & \sum_s U_s(x_s) \\ \text{subject to} & \sum_{s:l\in L(s)} x_s \leq c_l, \quad \forall l, \end{array} \tag{1}$$

This work was supported by the United States NSF Grants CCF-0448012 and CNS-0427677; and by the Fulbright Program and the Ministry of Education and Science of Spain.

where the variables are $\mathbf{x}$. Note that the variables are coupled through the linear flow constraints, but each utility is a function only of local variables (uncoupled utilities). It is assumed that the utilities $U_s$ are strictly concave functions. Much research effort has been put in the design of distributed algorithms for NUM [2]. The main ingredient to obtain distributed algorithms is the *decomposition techniques*, widely used in optimization theory [2, 3] (see [3] for an overview of distributed algorithms based on primal and dual decomposition approaches).

The majority of the utility problem formulations considered in the literature concern uncoupled utilities where the local variables corresponding to one node do not directly disturb the utilities of the other nodes. Systems with competition or cooperation, however, do not satisfy this assumption and the utilities are indeed coupled. An example of cooperation model can be found in networks where nodes form clusters and the utility obtained by each node depends on the rate allocated to others within the same cluster. An example of competition model is wireless power control and Digital Subscriber Line (DSL) spectrum management of copper wires in a cable binder, where the utilities are functions of the Signal-to-Interference Ratios (SIR)s that are dependent on the transmit powers of other users.

This paper's focus on *coupled utility* is new even though *coupled constraints* are standard in NUM. We present a systematic approach to deal with coupled utilities in a distributed and efficient way by using proper combination of existing distributed algorithms for uncoupled NUM problems and decomposition techniques from optimization theory. The key idea to deal with coupling in the objective function is to introduce auxiliary variables and equality constraints, thus transferring the coupling in the objective function to coupling in the constraints, which can be decoupled by dual decomposition and solved by introducing *consistency pricing* in addition to the *link congestion pricing* inherent in the dual decomposition approach of (1). Moreover, our *cluster based coupled NUM* approach has implication on the timescale of network protocol operations that use both consistency and link pricing. Table 1 summarizes the key ideas behind coupled NUM. Our main results are contained in the first column of Table 1 whereas the ideas in the second column are already well known in basic NUM.

**Table 1**. A Summary of coupled NUM with Implications on Network Design.

| | *Competition or Cooperation* | *Flow Constraints or Resource Sharing* |
|---|---|---|
| (A) Problem | Coupled objective | Coupled constraint |
| (B) Approach | Local copies of variables | Dual decomposition |
| (C) Overhead | Local consistency pricing | Global congestion pricing |

The paper is organized as follows. Section 2 contains the main contribution of the paper: an efficient distributed algorithm to solve general coupled NUM problems. Particularly, two different decomposition techniques lead to different interpretations in the design of network protocols. Section 3 presents an application of the proposed method: an optimal power control scheme to minimize queueing delay in wireless networks using consistency pricing.

## 2. DECOMPOSITION APPROACH AND DISTRIBUTED ALGORITHMS

### 2.1. Problem Formulation

Consider the basic NUM problem utilities which not only depend on local variables but also on variables of other utilities. The problem formulation is

$$\begin{array}{ll} \underset{\{\mathbf{x}_k\}}{\text{maximize}} & \sum_{k=1}^{K} U_k\left(\mathbf{x}_k, \{\mathbf{x}_l\}_{l\in\mathcal{L}(k)}\right) \\ \text{subject to} & \mathbf{x}_k \in \mathcal{X}_k \qquad \forall k, \\ & \sum_{k=1}^{K} \mathbf{g}_k\left(\mathbf{x}_k\right) \leq \mathbf{c} \end{array} \tag{2}$$

where the (strictly concave) utilities $U_k$ depend on a local vector variable $\mathbf{x}_k$ and on variables of other utilities $\mathbf{x}_l$ for $l \in \mathcal{L}(k)$ (i.e., coupled utilities), $\mathcal{L}(k)$ is the set of nodes coupled with the $k$th utility, the sets $\mathcal{X}_k$ are arbitrary convex sets representing local constraints, and the coupling constraining function $\sum_k \mathbf{g}_k(\mathbf{x}_k)$ is not necessarily linear, but still convex. Note that this model has two types of coupling: coupling constraints (slightly more general than the linear constraints in the basic NUM model in (1)) and coupled utilities (absent in the basic NUM model).

Of particular interest is the case where the coupling between utilities is through an interference term that contains an additive convex combination of the coupling variables:

$$U_k\left(\mathbf{x}_k, \{\mathbf{x}_l\}_{l\in\mathcal{L}(k)}\right) = U_k\left(\mathbf{x}_k, \mathbf{i}_k\right) \tag{3}$$

where $\mathbf{i}_k = \sum_{l\in\mathcal{L}(k)} \mathbf{h}_{kl}(\mathbf{x}_l)$ and the $\mathbf{h}_{kl}$'s are convex functions. Note that, by definition of *interference*, each utility $U_k(\mathbf{x}_k, \mathbf{i}_k)$ is decreasing in $\mathbf{i}_k$. This case allows a simpler implementation as shown in the next subsection. The interference term has a physical implication in practice where network nodes such as DSL modems already have the capability to measure locally the total interference from other competing network nodes.

### 2.2. Distributed Algorithm Based on Consistency Pricing

The key idea to tackle coupled utilities is to introduce auxiliary variables and additional equality constraints, thus transferring the coupling in the objective function to coupling in the constraints, which can be decoupled by dual decomposition and solved by introducing additional *consistency pricing*. It is reasonable to assume that if two nodes have their individual utilities dependent on each other's local variables, then there must be some communication channels in which they can locally exchange pricing messages. It turns out that the global link congestion price update of the canonical distributed algorithm is not affected by the local *consistency price* updates, which can be conducted via these local communication channels among the nodes.

The key step is to introduce in problem (2) auxiliary variables $\mathbf{x}_{kl}$ for the coupled arguments in the utility functions and additional equality constraints to enforce consistency:

$$\begin{array}{ll} \underset{\{\mathbf{x}_k\},\{\mathbf{x}_{kl}\}}{\text{maximize}} & \sum_k U_k\left(\mathbf{x}_k, \{\mathbf{x}_{kl}\}_{l\in\mathcal{L}(k)}\right) \\ \text{subject to} & \mathbf{x}_k \in \mathcal{X}_k \qquad \forall k, \\ & \sum_k \mathbf{g}_k\left(\mathbf{x}_k\right) \leq \mathbf{c}, \\ & \mathbf{x}_{kl} = \mathbf{x}_l, \quad \forall k, l \in \mathcal{L}(k), \end{array} \tag{4}$$

where $\mathbf{x}_k$ are local variables at the $k$th node. Next, to obtain a distributed algorithm, we take a dual decomposition approach by relaxing all the coupling constraints in problem (4):

$$\begin{array}{ll} \underset{\{\mathbf{x}_k\},\{\mathbf{x}_{kl}\}}{\text{maximize}} & \sum_k U_k\left(\mathbf{x}_k, \{\mathbf{x}_{kl}\}_{l\in\mathcal{L}(k)}\right) + \boldsymbol{\lambda}^T\left(\mathbf{c} - \sum_k \mathbf{g}_k\left(\mathbf{x}_k\right)\right) \\ & + \sum_{k,l\in\mathcal{L}(k)} \boldsymbol{\gamma}_{kl}^T\left(\mathbf{x}_l - \mathbf{x}_{kl}\right) \\ \text{subject to} & \mathbf{x}_k \in \mathcal{X}_k \qquad \forall k, \\ & \mathbf{x}_{kl} \in \mathcal{X}_l \qquad \forall k, l \in \mathcal{L}(k) \end{array} \tag{5}$$

where $\boldsymbol{\lambda}$ are the *link prices* and the $\boldsymbol{\gamma}_{kl}$'s are the *consistency prices*. By exploiting the *separability* structure of the Lagrangian, the Lagrangian is separated into many subproblems where maximization is done using local variables (the $k$th subproblem uses *only* variables with the first subscript index $k$). The optimal value of (5) for a given set of $\boldsymbol{\gamma}_{kl}$'s and $\boldsymbol{\lambda}$ defines the dual function $g(\{\boldsymbol{\gamma}_{kl}\}, \boldsymbol{\lambda})$. The dual problem is thus given by

$$\underset{\{\boldsymbol{\gamma}_{kl}\},\boldsymbol{\lambda}}{\text{minimize}} \quad g(\{\boldsymbol{\gamma}_{kl}\}, \boldsymbol{\lambda}) \quad \text{subject to} \quad \boldsymbol{\lambda} \geq \mathbf{0}. \tag{6}$$

It is worthwhile noting that (6) is equivalent to

$$\underset{\boldsymbol{\lambda}}{\text{minimize}} \left( \underset{\{\boldsymbol{\gamma}_{kl}\}}{\text{minimize}} \quad g(\{\boldsymbol{\gamma}_{kl}\}, \boldsymbol{\lambda}) \right) \quad \text{subject to} \quad \boldsymbol{\lambda} \geq \mathbf{0}. \tag{7}$$

Solving the dual function (either (6) or (7)) is equivalent to solving the original problem.

In one approach, (6) is solved by simultaneously updating the prices (both the link prices and the consistency prices) using a subgradient algorithm. In another approach through formulation (7), however, the inner minimization is fully performed (by repeatedly updating the set of $\boldsymbol{\gamma}_{kl}$'s) for each update of $\boldsymbol{\lambda}$. This latter approach implies two timescales: a fast timescale in which each cluster updates the corresponding consistency prices and a slow timescale in which the network updates the link prices; whereas the former approach has just one timescale. The alternative of two timescales has an interest from a practical perspective since consistency prices can be exchanged very quickly over local communication channels only by nodes that are coupled together – a result attributed to close physical proximity in practice. The dual decomposition using (7) is termed *cluster based NUM*. Our proposed solution is summarized next.

Problem (2), where the utilities $U_k$ are strictly concave, the sets $\mathcal{X}_k$ are arbitrary convex sets, and the constraining functions $\mathbf{g}_k$ are convex, can be optimally solved (for sufficiently small stepsizes by the following distributed algorithm:

**Algorithm:** At each iteration $t$,

- Step 1: The link prices are updated as

$$\boldsymbol{\lambda}(t+1) = \left[ \boldsymbol{\lambda}(t) - \alpha \left( \mathbf{c} - \sum_k \mathbf{g}_k(\mathbf{x}_k) \right) \right]^+ \tag{8}$$

  and are then broadcasted to the nodes. Note each component of $\boldsymbol{\lambda}$ can be updated by each link distributively.

- Step 2: each $k$th node updates the consistency prices (at a faster or same timescale as the update of $\boldsymbol{\lambda}(t)$) as

$$\boldsymbol{\gamma}_{kl}(t+1) = \boldsymbol{\gamma}_{kl}(t) - \alpha \left( \mathbf{x}_l(t) - \mathbf{x}_{kl}(t) \right), l \in L(k) \tag{9}$$

  and then broadcast them to the coupled nodes within the cluster, and

- Step 3: the $k$th node, for all $k$, locally solves the problem

$$\begin{array}{ll} \underset{\mathbf{x}_k, \{\mathbf{x}_{kl}\}r}{\text{maximize}} & U_k\left(\mathbf{x}_k, \{\mathbf{x}_{kl}\}_{l \in \mathcal{L}(k)}\right) - \boldsymbol{\lambda}^T \sum_k \mathbf{g}_k(\mathbf{x}_k) + \\ & \left(\sum_{l:k \in \mathcal{L}(l)} \boldsymbol{\gamma}_{lk}\right)^T \mathbf{x}_k - \sum_{l \in \mathcal{L}(k)} \boldsymbol{\gamma}_{kl}^T \mathbf{x}_{kl} \\ \text{subject to} & \mathbf{x}_k \in \mathcal{X}_k \\ & \mathbf{x}_{kl} \in \mathcal{X}_l \qquad \forall l \in \mathcal{L}(k) \end{array} \tag{10}$$

  where $\{\mathbf{x}_{kl}\}_{l \in \mathcal{L}(k)}$ are auxiliary local variables for the $k$th node.

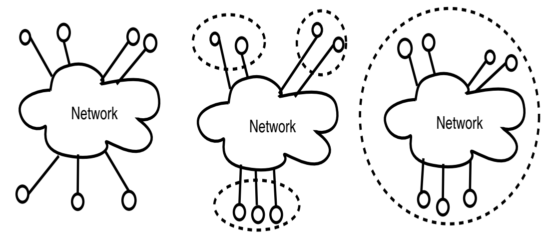


**Fig. 1**. Three network illustrations in terms of coupling: (a) all uncoupled utilities; (b) partially coupled utilities within clusters; (c) fully coupled utilities. (The dotted lines indicate coupling and where the consistency prices are exchanged.)

It is possible to consider an asyncronous update in Step 2 that still converges to the optimal solution (c.f. [2]). It is important to note that the update of each $\boldsymbol{\gamma}_{kl}$ can be locally done at the $k$th node with knowledge of the coupling variable $\mathbf{x}_l$ (through measurement or estimation) and its own local auxiliary variable $\mathbf{x}_{kl}$.

Summarizing, all the nodes advertise their local variables $\mathbf{x}_k$ (not the auxiliary ones $\mathbf{x}_{kl}$); each link $j$, for all $j$, updates and signals the $j$th component of $\boldsymbol{\lambda}$ to all the links; each node updates the corresponding $\boldsymbol{\gamma}_{kl}$'s (with knowledge of the variables $\mathbf{x}_k$ of the coupled nodes) and signals it to the coupled nodes (such a message passing overhead within each cluster is the price to pay to decouple the coupled utilities).

See Figure 1 for an illustration of three scenarios depending on the coupling: (a) uncoupled utilities (solved with a standard dual-based algorithm); (b) partially coupled utilities on a cluster basis (solved with the standard dual-based algorithm plus an additional message passing within each cluster); and (c) fully coupled utilities (solved with the standard dual-based algorithm with additional message passing in the entire network).

### 2.3. Interference-Dependent Utilities

Consider now an interesting special case in which the coupling is through an interference term lumping together the coupling variables as in (3). The (convex) problem with auxiliary variables $\mathbf{i}_k$ for the coupled variables is

$$\begin{array}{ll} \underset{\{\mathbf{x}_k\},\{\mathbf{i}_k\}}{\text{maximize}} & \sum_k U_k(\mathbf{x}_k, \mathbf{i}_k) \\ \text{subject to} & \mathbf{x}_k \in \mathcal{X}_k \qquad \forall k, \\ & \sum_k \mathbf{g}_k(\mathbf{x}_k) \leq \mathbf{c}, \\ & \mathbf{i}_k \geq \sum_{l \in \mathcal{L}(k)} \mathbf{h}_{kl}(\mathbf{x}_l) \quad \forall k \end{array} \tag{11}$$

where the interference inequality constraint is satisfied with equality at an optimal point since utilities are decreasing in the interference term. The only modification to our earlier algorithm is that the update of the consistency prices in (9) is replaced by:

$$\gamma_k (t+1) = \left[ \gamma_k (t) - \alpha \left( \mathbf{i}_k - \sum_{l \in \mathcal{L}(k)} \mathbf{h}_{kl} (\mathbf{x}_l) \right) \right]^+ \tag{12}$$

which can be done at the $k$th node with knowledge of the local variables and of the linear combination of the coupling variables from other nodes.

By leveraging the structure of the interference term, only one consistency price is needed for each interference term (which may contain many coupled variables), substantially reducing the amount of signaling. Indeed, message passing overhead, measured by the number of consistency prices to update in (9) and (12), is of the order $O(N^2)$ and $O(N)$, respectively, where $N$ is the number of nodes in a cluster.

## 3. OPTIMAL POWER CONTROL IN BROADCAST CHANNELS USING CONSISTENCY PRICING

In this section, we illustrate an application of consistency pricing to power control scheme in a multiuser wireless network with Gaussian broadcast channel which corresponds to the case of fully coupled utilities in Figure 1(c). The network has $L$ logical links (equivalently, transceiver pairs). Transmit powers for each user are denoted by $p_1, \ldots, p_L$. The Signal-to-Interference Ratio (SIR) for the receiver on logical link $l$ is

$$\mathsf{SIR}_l(\mathbf{p}) = \frac{p_l G_{ll}}{\sum_{j \neq l}^{L} p_j G_{lj} + n_l}, \tag{13}$$

where $G_{lj}$ are the channel gains from transmitter $j$ to receiver $l$, and $n_l$ is the additive white Gaussian noise for receiver $l$. Attainable data rates at each logical link is given (ignoring constants) by $c_l(\mathbf{p}) = \log(1+\mathsf{SIR}_l(\mathbf{p}))$ bits per transmission. It is desired to minimize the total queueing delay in the wireless network which assumes that flows arrive according to some statistical assumption and are served in a single buffer at each logical link in the wireless network with a total power constraint $P_T$. Formally, we have

$$\begin{array}{ll} \underset{\mathbf{t},\mathbf{p} \geq \mathbf{0}}{\text{minimize}} & \sum_l \frac{t_l}{c_l(\mathbf{p}) - t_l} \\ \text{subject to} & t_l \leq c_l(\mathbf{p}), \quad \forall l, \\ & p_l \leq p_{\max}, \quad \forall l, \\ & \sum_l p_l \leq P_T. \end{array} \tag{14}$$

where the objective function is a sum of queueing delays at each link. It is well known that the set of achievable rates $t_l, \forall l$, in the Gaussian broadcast channel and the power vector $\mathbf{p}$ are not jointly convex in the capacity region, i.e., the set of constraints in (14) is not convex [4]. Furthermore, the objective function is a nonconvex function of $\mathbf{t}$ and $\mathbf{p}$, and is tightly coupled in $\mathbf{p}$.

However, with a suitable change of variables, the above problem can be transformed into a convex optimization problem. Specifically, by a log transformation of variables $\tilde{t}_l = \log t_l$ and $\tilde{p}_l = \log p_l, \forall l$, the capacity region of the Gaussian broadcast channel is jointly convex in the transformed variables $\tilde{\mathbf{t}}$ and $\tilde{\mathbf{p}}$[1]. In our transformation, we need to ensure both convexity (for global optimality) and decomposition (for distributed solution). Introducing an additional variable $\tilde{\nu}$ and transferring the objective function to the constraints, (14) is transformed into the following convex optimization.

$$\begin{array}{ll} \underset{\tilde{\nu},\tilde{\mathbf{t}},\tilde{\mathbf{p}} \geq \mathbf{0}}{\text{minimize}} & \sum_l e^{\tilde{\nu}_l} \\ \text{subject to} & (e^{\tilde{t}_l - \tilde{\nu}_l} + e^{\tilde{t}_l})/c_l(\tilde{\mathbf{p}}) \leq 1, \quad \forall l, \\ & e^{\tilde{t}_l}/c_l(\tilde{\mathbf{p}}) \leq 1, \quad \forall l, \\ & e^{\tilde{p}_l} \leq p_{\max}, \quad \forall l, \quad \sum_l e^{\tilde{p}_l} \leq P_T. \end{array} \tag{15}$$

We can resolve the coupling in the total power constraint in (15) using the primal or dual decomposition approaches in [3]. If we use dual decomposition, at Step 1 in our general algorithm, we have

$$\lambda(t+1) = \left[\lambda(t) - \alpha\left(P_T - \textstyle\sum_l e^{\tilde{p}_l}\right)\right]^+.$$

The consistency pricing algorithm is used to resolve the coupling in $\tilde{\mathbf{p}}$ in the first two constraints of (15)[2]. Specifically, auxiliary variables $p_{lj}^R$ and additional consistency constraints $p_{lj}^R = G_{lj} p_j, \forall l, j$ are introduced in (15) which, after a log transformation, gives $\tilde{p}_{lj}^R = \tilde{G}_{lj} + \tilde{p}_j, \forall l, j$. At Step 2 in our general algorithm, the $l$th logical link now estimates the effective received power from the $j$th interfering link, $p_j^{est}(t)$ (Ideal estimation yields $p_j^{est}(t) = G_{lj} p_j, \forall j$), and updates the dual variable of the $j$th consistency constraint by

$$\gamma_{lj}(t+1) = \gamma_{lj}(t) + \alpha\left(\tilde{p}_{lj}^R(t) - \log p_j^{est}(t)\right).$$

## 4. REFERENCES


[1] F. P. Kelly, A. Maulloo, and D. Tan. Rate control for communication networks: Shadow prices, proportional fairness, and stability. *Journal of Oper. Res. Soc.*, 49(3):237–252, Mar. 1998.

[2] Dimitri P. Bertsekas and John N. Tsitsiklis. *Parallel and Distributed Computation: Numerical Methods*. Prentice Hall, Englewood Cliffs, NJ, 1989.

[3] D. P. Palomar and M. Chiang. Alternative decompositions for distributed maximization of network utility: Framework and applications. *Proc. of IEEE Infocom 2006*, Apr. 2006.

[4] D. N. Tse. Optimal power allocation over parallel gaussian broadcast channels. *Proc. Int. Symp. Information Theory*, Jun. 1997.


[1]Beside the work in [4], the above log transformation approach provides an alternate *convex* characterization of the rate region of the multiuser Gaussian broadcast channels.

[2]To obtain a convex transformation, the interference-dependent term in Section 2.3 cannot be used in (14).