

On the Relationship Between Queues and Multipliers

Víctor Valls, Douglas J. Leith
Hamilton Institute, NUI Maynooth

Abstract—We show that the occupancy of appropriate queues can be used as a surrogate for Lagrange multipliers in convex optimisation. Our analysis uses only elementary methods, and is not asymptotic in nature. One immediate consequence is that in network problems the scaled link queue occupancy can be used as multipliers when calculating the dual function. Conversely, the connection with multipliers casts light on the link queue behaviour under optimal decision-making (not just max-weight scheduling). Namely, on links corresponding to active constraints the queue occupancy necessarily grows as step size α is reduced. Importantly, our analysis encompasses nonlinear constraints, and so generalises analysis beyond conventional queueing networks.

Index Terms—convex optimisation, subgradient methods, max-weight scheduling

I. INTRODUCTION

In this paper we consider the use of queue occupancy as a surrogate for Lagrange multipliers. In queueing networks the use of queue occupancy is well established in max-weight scheduling, where scheduling decisions which aim to stabilise queues are explicitly based on queue occupancies. Max-weight methods can be extended to solve a class of convex utility optimisation problems, but as previously noted by *e.g.*, [1] the connection between queues and multipliers in max-weight methods remains largely open. Further, our interest here is in general convex optimisation problems rather than only queueing network problems.

We take as our starting point the observation that the usual subgradient multiplier update $\lambda_{k+1} = [\lambda_k + \alpha(g(\mathbf{z}_k) - b)]^+$ for convex constraint $g(\mathbf{z}_k) \leq b$ has a queue-like form. Namely, rescaling the multiplier by α , the subgradient update can be rewritten equivalently as $Q_{k+1} = [Q_k + g(\mathbf{z}_k) - b]^+$ where $Q_k := \lambda_k/\alpha$, which is identical to a queue update with increment $g(\mathbf{z}_k) - b$. Similarly, we can also consider a sequence of convex constraints $g(\mathbf{z}_k) \leq b_k$ and subgradient multiplier update $\lambda_{k+1} = [\lambda_k + \alpha(g(\mathbf{z}_k) - b_k)]^+$ to accommodate queue-like arrival/departures. This increment is necessarily continuous-valued in convex problems (if \mathbf{z}_k is restricted to take values in a discrete set, the problem is non-convex), and so the queue occupancy Q_k is also continuous-valued. In contrast, in network problems where flows consist of packets, code blocks, vehicles, *etc.*, the queue occupancy is discrete-valued. While an asymptotic fluid analysis might be attempted to establish a relationship between scaled multiplier Q_k and

the occupancy of appropriate network queues, in this paper we show that in fact this is unnecessary. Using elementary methods (no need for sophisticated fluid-limit arguments *etc.*), we present mild conditions under which multiplier λ_k and the occupancy of an appropriate discrete queue are essentially inter-changeable quantities.

One immediate consequence of this result is that in network problems the scaled link queue occupancy can be used as multipliers when calculating the dual function. Conversely, the connection with multipliers casts light on the link queue behaviour under optimal decision-making (not just max-weight scheduling); namely, on links corresponding to active constraints the queue occupancy necessarily grows as step size α is reduced. Importantly, our analysis encompasses nonlinear constraints, and so generalises analysis beyond conventional queueing networks¹.

A. Related Work

Max-weight scheduling was introduced by Tassiulas and Ephremides in their seminal paper [2]. They consider a network of queues with slotted time, an integer number of packet arrivals in each slot and a finite set of admissible scheduling patterns, referred to as *actions*, in each slot. Using a Forster-Lyapunov approach they present a scheduling policy that stabilises the queues provided the external traffic arrivals are strictly feasible. Namely, the scheduling policy consists of selecting the action from a discrete set D at each slot that maximises the queue-length-weighted sum of rates, $\mathbf{x}_k \in \arg \max_{\mathbf{x} \in D} -\mathbf{Q}_k^T \mathbf{A} \mathbf{x}$. Max-weight is extended in a sequence of papers [3], [4], [5], [6] and books [7], [8] to encompass the maximisation of concave utility functions subject to queue stability.

The observation that the multiplier subgradient update has a queue-like form is not new, see for example [9] and later papers. However, we are aware of few rigorous results relating queues and multipliers in convex optimisation. A notable exception is [10], which establishes that under a max-weight schedule a discrete queue update tends on average to drift towards the optimal multiplier value.

¹In a typical queueing network formulation the connectivity between queues is captured via a matrix $\mathbf{A} \in \{-1, 0, 1\}^{n \times n}$, whose i 'th row has a -1 at the i 'th entry, 1 at entries corresponding to queues from which packets are sent to queue i , and 0 entries elsewhere. The queue occupancy then updates according to $Q_{k+1} = [Q_k + \mathbf{A} \mathbf{x}_k + \mathbf{b}_k]^+$, $i = 1, 2, \dots, n$, where the i 'th element of vector $\mathbf{b}_k \in \mathbb{N}^n$ denotes the number of external packet arrivals to queue i at time k .

This material is based upon works supported by the Science Foundation Ireland under Grant No. 11/PI/1177.

The authors are with the Hamilton Institute, NUI Maynooth (e-mail: victor.valls@nuim.ie, doug.leith@nuim.ie).

B. Notation

Vectors and matrices are indicated in bold type. Since we often use subscripts to indicate elements in a sequence, to avoid confusion we usually use a superscript $x^{(i)}$ to denote the i 'th element of a vector \mathbf{x} . The i 'th element of operator $[\mathbf{x}]^{[0, \bar{\lambda}]}$ equals $x^{(i)}$ (the i 'th element of \mathbf{x}) when $x^{(i)} \in [0, \bar{\lambda})$ and otherwise equals 0 when $x^{(i)} < 0$ and $\bar{\lambda}$ when $x^{(i)} \geq \bar{\lambda}$. Note that we allow $\bar{\lambda} = +\infty$, and following standard notation in this case usually write $[x]^+$ instead of $[x]^{[0, \infty)}$. The subgradient of a convex function f at point \mathbf{x} is denoted $\partial f(\mathbf{x})$. For two vectors $\mathbf{x}, \mathbf{y} \in \mathbb{R}^m$ we use element-wise comparisons $\mathbf{x} \succeq \mathbf{y}$ and $\mathbf{y} \succ \mathbf{x}$ to denote when $y^{(i)} \geq x^{(i)}$, $y^{(i)} > x^{(i)}$ respectively for all $i = 1, \dots, m$.

II. DISCRETE AND CONTINUOUS QUEUES

Let sequences \mathbf{z}_k and \mathbf{x}_k , $k = 1, 2, \dots$, consist of points from \mathbb{R}^n . We will often think of \mathbf{z}_k as being continuous-valued and \mathbf{x}_k as being discrete-valued. For example, \mathbf{x}_k might be a suitable projection of \mathbf{z}_k onto the integer lattice. Let matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$ and $\{\mathbf{b}_k\}$ be a sequence of points from \mathbb{R}^m and \mathbf{b} also a point from \mathbb{R}^m . We will often think of \mathbf{b}_k as the packet arrival/departures at time step k and \mathbf{b} as the mean packet arrival/departure rate.

Define the following two queueing iterations:

$$\mathbf{Q}_{k+1} = [\mathbf{Q}_k + \mathbf{A}\mathbf{z}_k - \mathbf{b}]^{[0, \bar{\lambda}]}, \quad (1)$$

$$\tilde{\mathbf{Q}}_{k+1} = [\tilde{\mathbf{Q}}_k + \mathbf{A}\mathbf{x}_k - \mathbf{b}_k]^{[0, \bar{\lambda}]} \quad (2)$$

with $\mathbf{Q}_k, \tilde{\mathbf{Q}}_k \in \mathbb{R}_+^m$, $\mathbf{Q}_k := [Q_k^{(1)}, \dots, Q_k^{(m)}]^T$, $\tilde{\mathbf{Q}}_k := [\tilde{Q}_k^{(1)}, \dots, \tilde{Q}_k^{(m)}]^T$ and initial condition $\mathbf{Q}_1 = \tilde{\mathbf{Q}}_1$. Here, vector \mathbf{Q}_k can be thought of as the occupancy of a set of queues with continuous-valued/fluid-like arrivals/departures $\mathbf{A}\mathbf{z}_{k+1}$ and \mathbf{b}^2 . While vector $\tilde{\mathbf{Q}}_k$ can be thought of as the occupancy of a set of queues with discrete arrivals/departures $\mathbf{A}\mathbf{x}_k$ and \mathbf{b}_k . Note that the maximum queue size $\bar{\lambda}$ may be finite.

When $\mathbf{b}_k = \mathbf{b}$ and $\mathbf{z}_k = \mathbf{x}_k$, $k = 1, 2, \dots$, the two queue updates are of course identical and $\mathbf{Q}_k = \tilde{\mathbf{Q}}_k$. For $\mathbf{z}_k \neq \mathbf{x}_k$ and/or $\mathbf{b}_k \neq \mathbf{b}$, so long as the differences between $\mathbf{z}_k, \mathbf{x}_k$ and between \mathbf{b}_k, \mathbf{b} remain small in an appropriate sense then we might expect that the queue occupancies \mathbf{Q}_k and $\tilde{\mathbf{Q}}_k$ remain close, and indeed this is the case. The required sense in which differences are required to be small is clarified by the following lemma (which corresponds to [11, Proposition 3.1.2]):

Lemma 1. Consider the sequences $x_{k+1} = [x_k + \delta_k]^{[0, \bar{\lambda}]}$ and $y_{k+1} = [y_k + \tilde{\delta}_k]^{[0, \bar{\lambda}]}$, $k = 1, 2, \dots$ where $x_k, y_k \in \mathbb{R}^+$, $\delta_k, \tilde{\delta}_k \in \mathbb{R}$ and $\bar{\lambda} > 0$. Suppose $x_1 = y_1$ and $|\sum_{i=1}^k \delta_i - \tilde{\delta}_i| \leq \epsilon$ for all $k \geq 1$. Then,

$$|x_k - y_k| \leq 2\epsilon, \quad \forall k \geq 1.$$

It can be seen from Lemma 1 that the requirement is that $|\sum_{i=1}^k \delta_i - \tilde{\delta}_i|$ is uniformly bounded for all k . Hence, to bound the difference between queue occupancies \mathbf{Q}_k and $\tilde{\mathbf{Q}}_k$ we can expect to require that $|\sum_{i=1}^k (\mathbf{a}^{(j)})^T (\mathbf{z}_i - \mathbf{x}_i)| \leq \sigma_1$

²Since $\mathbf{A}\mathbf{z}_k$ and \mathbf{b} can both be negative valued we can think of $\mathbf{A}\mathbf{z}_k$ and \mathbf{b} as either arrivals or departures, or any combination of arrivals and departures that is convenient – all that matters is the net queue increment $\mathbf{A}\mathbf{z}_k - \mathbf{b}$.

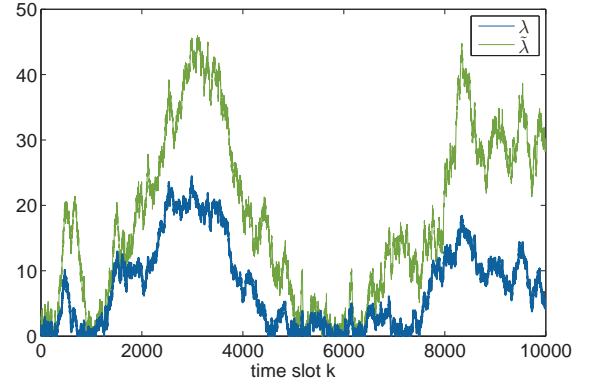


Fig. 1. Example realisations of $\tilde{\lambda}_k$ and λ_k for a queue with service rate $b = 0.51$, i.i.d. equiprobable $\{0,1\}$ arrivals (so mean is 0.5) and parameter $\alpha = 1$.

and $|\sum_{i=1}^k b_i^{(j)} - b^{(j)}| \leq \sigma_2$ for some $\sigma_1, \sigma_2 > 0$ and any $k \geq 1$, where $(\mathbf{a}^{(j)})^T$ denotes the j 'th row of matrix \mathbf{A} . However, rather than stating this result in terms of \mathbf{Q}_k and $\tilde{\mathbf{Q}}_k$ it will prove more convenient later to state it in terms of the corresponding rescaled updates

$$\lambda_{k+1} = [\lambda_k + \alpha(\mathbf{A}\mathbf{z}_k - \mathbf{b})]^{[0, \bar{\lambda}]}, \quad (3)$$

$$\tilde{\lambda}_{k+1} = [\tilde{\lambda}_k + \alpha(\mathbf{A}\mathbf{x}_k - \mathbf{b}_k)]^{[0, \bar{\lambda}]} \quad (4)$$

where $\lambda_k := \alpha\mathbf{Q}_k$ and $\tilde{\lambda}_k := \alpha\tilde{\mathbf{Q}}_k$, $\alpha > 0$. We have the following:

Lemma 2 (Discrete and Continuous Queues). Consider the updates (3) and (4) with $\lambda_1 = \tilde{\lambda}_1$. Suppose that $|\sum_{i=1}^k (\mathbf{a}^{(j)})^T (\mathbf{z}_i - \mathbf{x}_i)| \leq \sigma_1$ and $|\sum_{i=1}^k (b_i^{(j)} - b^{(j)})| \leq \sigma_2$ where $(\mathbf{a}^{(j)})^T$ denotes the j 'th row of matrix \mathbf{A} and $b_i^{(j)}$ denotes the j 'th element of vector \mathbf{b}_i . Then,

$$|\tilde{\lambda}_k^{(j)} - \lambda_k^{(j)}| \leq 2\alpha(\sigma_1 + \sigma_2), \quad k = 1, 2, \dots \quad (5)$$

where $\tilde{\lambda}_k^{(j)}, \lambda_k^{(j)}$ denote, respectively, the j 'th element of vectors $\tilde{\lambda}_k, \lambda_k$.

Note that selecting constant sequence $\mathbf{b}_k = \mathbf{b}$, $k = 1, 2, \dots$ trivially satisfies the conditions of Lemma 2. Also, since $\|\cdot\|_2 \leq \|\cdot\|_1$ it follows immediately from Lemma 2 that

$$\|\tilde{\lambda}_k - \lambda_k\|_2 \leq 2m\alpha(\sigma_1 + \sigma_2), \quad k = 1, 2, \dots \quad (6)$$

This behaviour is illustrated in Figure 1, where it can be seen that the distance between λ_k and $\tilde{\lambda}_k$ remains uniformly bounded over time.

III. APPLICATION TO CONVEX OPTIMISATION

We consider the application of Lemma 2 to convex optimisation. Consider the convex optimisation P :

$$\begin{aligned} & \underset{\mathbf{z} \in C}{\text{minimise}} && f(\mathbf{z}) \\ & \text{subject to} && \mathbf{A}\mathbf{z} \preceq \mathbf{b} \end{aligned}$$

where $f: \mathbb{R}^n \rightarrow \mathbb{R}$ is convex, matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$, vector $\mathbf{b} \in \mathbb{R}^m$ and C a bounded convex subset in \mathbb{R}^n . Let $C_0 :=$

$\{z \in C : \mathbf{A}z \preceq \mathbf{b}\}$ denote the set of feasible points, which we will assume has non-empty relative interior (*i.e.*, a Slater point exists). Let $C^* := \arg \min_{z \in C_0} f(z) \subset C_0$ be the set of optima and $f^* := f(z^*)$, $z^* \in C^*$.

Define Lagrangian $L(z, \lambda) := f(z) + \lambda^T(\mathbf{A}z - \mathbf{b})$ where $\lambda \in \mathbb{R}_+^m$. Since set C has non-empty relative interior, the Slater condition is satisfied and strong duality holds. That is,

$$\min_{z \in C} \max_{\lambda \succeq 0} L(z, \lambda) = \max_{\lambda \succeq 0} \min_{z \in C} L(z, \lambda) = f^*.$$

Further, we have the following boundedness property, which corresponds to [12, Lemma 1]:

Lemma 3. *Let $q(\lambda) := \min_{z \in C} L(z, \lambda)$ and $\lambda^* \in \arg \max_{\lambda \succeq 0} q(\lambda)$. Suppose set C_0 has non-empty relative interior. Then*

$$\|\lambda^*\|_2 \leq \frac{f(\bar{z}) - q(\lambda^*)}{\xi} \quad (7)$$

where $\xi = -\min_{1 \leq j \leq m} g^{(j)}(\bar{z})$, $\bar{z} \in \text{reint}(C_0)$.

From Lemma 3 it follows that there exists a constant $\bar{\lambda} \in [0, \infty)$ such that the optimum $\lambda^* \preceq \bar{\lambda} \mathbf{1} := \bar{\lambda}$. That is, it is sufficient to confine consideration to the Lagrangian on bounded set $0 \preceq \lambda \preceq \bar{\lambda}$ since $\max_{0 \preceq \lambda \preceq \bar{\lambda}} \min_{z \in C} L(z, \lambda) = f^*$.

Before proceeding, we note the following bound that will prove useful:

Lemma 4 (Lower Bound). *Let $f : \mathbb{R}^n \rightarrow \mathbb{R}$, $g^{(j)} : \mathbb{R}^n \rightarrow \mathbb{R}$, $j = 1, \dots, m$ be convex functions and let C be a bounded convex set in \mathbb{R}^n . Further, assume that there exists a constant such that $\max_{z \in C} \|\mathbf{g}(z) - \mathbf{b}\|_2 \leq \sigma_3$ where $\mathbf{g}(z) = [g^{(1)}(z), \dots, g^{(m)}(z)]^T$. Let Lagrangian $L(z, \lambda) := f(z) + \lambda^T(\mathbf{g}(z) - \mathbf{b})$, $\lambda \succeq 0$, and assume that $f^* := L(z^*, \lambda^*)$ is a saddle point, *i.e.*, $L(z^*, \lambda) \leq L(z^*, \lambda^*) \leq L(z, \lambda^*)$, and that $\lambda^* \preceq \bar{\lambda} \mathbf{1}$. Consider update*

$$\lambda_{k+1} = [\lambda_k + \alpha(\mathbf{g}(z_k) - \mathbf{b})]^{[0, \bar{\lambda}]} \quad (8)$$

with constant step size $\alpha > 0$ and $\{z_k\}$ is an arbitrary sequence of points from set C . Then,

$$-\frac{\|\lambda_1 - \lambda^*\|_2^2}{2\alpha k} - \frac{\alpha}{2}\sigma_3^2 \leq \frac{1}{k} \sum_{i=1}^k L(z_i, \lambda_i) - f^*. \quad (9)$$

We are now in a position to consider the following sequence of optimisations $\{\tilde{P}_k^L\}$:

$$z_k \in \arg \min_{z \in C} L(z, \tilde{\lambda}_k) \quad (10)$$

$$= \arg \min_{z \in C} f(z) + \tilde{\lambda}_k^T \mathbf{A}z \quad (11)$$

$$\tilde{\lambda}_{k+1} = [\tilde{\lambda}_k + \alpha(\mathbf{A}x_k - \mathbf{b}_k)]^{[0, \bar{\lambda}]} \quad (12)$$

where $\alpha > 0$ is a step size parameter, $\{\mathbf{b}_k\}$ is a sequence of points from \mathbb{R}^m , $\{x_k\}$ a sequence of points from \mathbb{R}^n . Update (11) is obtained from (10) by retaining only the parts of $L(z, \tilde{\lambda}_k)$ which depend on z *i.e.*, dropping constant terms which do not change the solution to the optimisation. Note that (11) does not involve \mathbf{b} or \mathbf{b}_k .

Observe that when we select $\mathbf{b}_k = \mathbf{b}$, $x_k = z_k$ and $\bar{\lambda} = +\infty$ then sequence $\{\tilde{P}_k^L\}$ corresponds to the standard dual subgradient update:

$$z_k \in \arg \min_{z \in C} L(z, \lambda_k) \quad (13)$$

$$\lambda_{k+1} = [\lambda_k + \alpha(\mathbf{A}z_k - \mathbf{b})]^+ \quad (14)$$

which is known to converge to a ball around the optimum of optimisation P , with the size of the ball depending on the value of step size parameter α . Namely, by Lemma 4 we have the following bound

$$-\frac{\|\lambda_1 - \lambda^*\|_2^2}{2\alpha k} - \frac{\alpha}{2}\sigma_3^2 \leq \frac{1}{k} \sum_{i=1}^k L(z_i, \lambda_i) - f^* \leq 0$$

since $\min_{z \in C} L(z, \lambda_k) = L(z_k, \lambda_k) \leq L(z^*, \lambda_k) \leq L(z^*, \lambda^*) = f^*$ for all k .

Update (10)-(12) generalises this standard subgradient update to allow use of a time-varying, possibly discrete, additive term \mathbf{b}_k and use of an approximate, possibly discrete, action x_k . Provided α is sufficiently small and \mathbf{b}_k , x_k sufficiently well-behaved, by Lemma 2 we know that the multiplier $\tilde{\lambda}_k$ generated by (12) will remain close in value to the continuously valued multiplier λ_k generated by (14). Using this, we can then show that the sequence $\{\tilde{P}_k^L\}$ converges to a ball around the solution of optimisation P :

Theorem 1. *Consider optimisation P . Consider also the associated sequence of optimisations $\{\tilde{P}_k^L\}$. Suppose that sequence $\{\mathbf{b}_k\}$ of points from \mathbb{R}^m satisfies $|(\frac{1}{k} \sum_{i=1}^k b_i^{(j)} - b^{(j)})| \leq \sigma_2/k$, $j = 1, \dots, m$ and sequence x_k satisfies $|\sum_{i=1}^k (\mathbf{a}^{(j)})^T(z_i - x_i)| \leq \sigma_1$, $j = 1, \dots, m$. Then, the sequence of solutions $\{z_k\}$ to the sequence of optimisations $\{\tilde{P}_k^L\}$ satisfies:*

$$-\frac{\|\tilde{\lambda}_1 - \lambda^*\|_2^2}{2\alpha k} - \frac{\alpha}{2}\sigma_3^2 - 2m\alpha(\sigma_1 + \sigma_2)\sigma_3 \leq \frac{1}{k} \sum_{i=1}^k L(z_i, \tilde{\lambda}_i) - f^* \leq 0 \quad (15)$$

where $\sigma_3 := \max_{z \in C} \|\mathbf{A}z - \mathbf{b}\|_2$.

As already noted in Section II, $\tilde{\lambda}_k$ can be thought of as the α scaled occupancy of a set of queues with net increment $\mathbf{A}x_k - \mathbf{b}_k$. That is, use of a discrete time-varying update (12) induces an equivalence between multipliers and the occupancy of an associated set of queues. Note that this observation complements [10, Theorem 1] in the context of max-weight optimisation, which states that a discrete queue update tends on average to drift towards the optimal multiplier value. In network flow problems these queues can be identified with physical link queues. However, since the formulation of network flow constraints is not unique, and changing how these constraints are formulated changes the multiplier update, some care may be necessary to select the constraints in a way that is congruent with the physical queues in a particular network. Further, since the multiplier queue occupancy is given by $\tilde{Q}_k = \tilde{\lambda}_k/\alpha$, for constraints where the multiplier is non-zero at the optimum the associated queue occupancy will necessarily grow as subgradient step size α is decreased. Since a small

step size is generally needed in order to converge to a small ball around the optimum, this indicates that a fundamental trade-off may exist between queue occupancy and optimality.

IV. STOCHASTIC CONSTRAINTS

This optimisation analysis can be readily extended to a useful class of stochastic constraints. Of particular interest, in view of the equivalence which has been established between multiplier updates and queues, is accommodating stochastic queue arrivals/departures. When $\{\mathbf{b}_k\}$ is a sequence such that $\frac{1}{k} \sum_{i=1}^k \mathbf{b}_i$ converges sufficiently quickly to \mathbf{b} , then by Theorem 1 the sequence of non-convex optimisations $\{\tilde{P}_k^L\}$ converges to the solution of optimisation problem P . Since this holds for all admissible sequences $\{\mathbf{b}_k\}$, this of course includes sample paths of a stochastic process.

Let $\{\mathbf{B}_k\}$ be a stochastic process with realisations of \mathbf{B}_k taking values in \mathbb{R}^m and with mean $\mathbf{b} \in \mathbb{R}^m$. Let $p_k := \text{Prob}(\|\frac{1}{k} \sum_{i=1}^k \mathbf{B}_i - \mathbf{b}\|_k \leq m\sigma_2/k)$. Note that, by central limit arguments, for many stochastic processes $\lim_{k \rightarrow \infty} p_k = 0$. Let $\{\mathbf{b}_i\}_{i=1}^k$ denote a realisation of length k and E^k the set of possible realisations of length k . Fraction p_k of these realisations satisfy $\|\frac{1}{k} \sum_{i=1}^k \mathbf{b}_i - \mathbf{b}\|_k \leq m\sigma_2/k$, i.e., fraction p_k of realisations satisfy the conditions of Theorem 1. We therefore have the following corollary to Theorem 1.

Corollary 1. *Consider the setup in Theorem 1. Suppose that sequence $\{\mathbf{b}_k\}$ is a realisation of a stochastic process $\{\mathbf{B}_k\}$ with mean $\mathbf{b} \in \mathbb{R}^m$. Let $p_k := \text{Prob}(\|\frac{1}{k} \sum_{i=1}^k \mathbf{B}_i - \mathbf{b}\|_\infty \leq m\sigma_2/k)$. Then there exists \bar{k} such that with probability p_k for all $k \geq \bar{k}$ the sequence of solutions $\{\mathbf{z}_k\}$ to the sequence of optimisations $\{\tilde{P}_k^L\}$ satisfies (15).*

For example, \mathbf{B}_k might take values in discrete set $E = \{0, 1\}^m$ and we can think of \mathbf{B}_k as random packet arrivals/departures at time step k and \mathbf{b} as the mean packet arrival/departure rate. Note that there is no requirement for stochastic process $\{\mathbf{B}_k\}$ to be i.i.d. or for any of its properties, other than that feasible set $\mathbf{A}\mathbf{z} \preceq \mathbf{b}$ has non-empty relative interior, to be known in advance in order to construct solution sequence $\{\tilde{P}_k^L\}$. Note also that while we require an interior (Slater) point to exist for $\mathbf{A}\mathbf{z} \preceq \mathbf{b}$ we do not require this to be the case for constraint $\mathbf{A}\mathbf{z} \preceq \frac{1}{k} \sum_{i=1}^k \mathbf{B}_i$ for finite k .

V. NONLINEAR CONSTRAINTS

The foregoing is for linear constraints. However, it carries over largely unchanged to nonlinear constraints $g^{(j)}(\mathbf{z}) \leq b^{(j)}$, $j = 1, \dots, m$ (which can be equivalently written in vector notation as $\mathbf{g}(\mathbf{z}) \preceq \mathbf{b}$ where $\mathbf{g}(\mathbf{z}) = [g^{(1)}(\mathbf{z}), \dots, g^{(m)}(\mathbf{z})]^T$) provided sequence $\{\mathbf{x}_k\}$ satisfies $|\sum_{i=1}^k g^{(j)}(\mathbf{z}_i) - g^{(j)}(\mathbf{x}_i)| \leq \sigma_1$. Namely, consider the sequence of optimisations $\{\tilde{P}_k^{NL}\}$:

$$\begin{aligned} \mathbf{z}_k &\in \arg \min_{\mathbf{z} \in C} L(\mathbf{z}, \tilde{\lambda}_k) \\ &= \arg \min_{\mathbf{z} \in C} f(\mathbf{z}) + \tilde{\lambda}_k^T \mathbf{g}(\mathbf{z}) \end{aligned} \quad (16)$$

$$\tilde{\lambda}_{k+1} = [\tilde{\lambda}_k + \alpha(\mathbf{g}(\mathbf{x}_k) - \mathbf{b}_k)]^{[0, \tilde{\lambda}]}. \quad (17)$$

Then we have that:

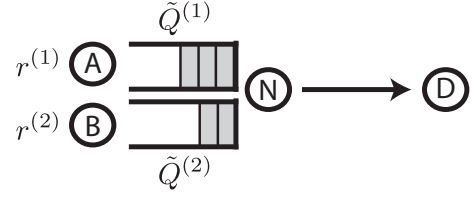


Fig. 2. Illustration of the network used in the example of Section VI.

Theorem 2. *Consider optimisation P with modified constraints $\mathbf{g}(\mathbf{z}) \preceq \mathbf{b}$, and consider associated sequence of non-convex optimisations $\{\tilde{P}_k^{NL}\}$. Suppose that $\sigma_3 := \max_{\mathbf{z} \in C} \|\mathbf{g}(\mathbf{z}) - \mathbf{b}\|_2$ is finite, sequence $\{\mathbf{x}_k\}$ satisfies $|\sum_{i=1}^k g^{(j)}(\mathbf{z}_i) - g^{(j)}(\mathbf{x}_i)| \leq \sigma_1$ and sequence $\{\mathbf{b}_k\}$ of points in \mathbb{R}^m satisfies $|\frac{1}{k} \sum_{i=1}^k b_i^{(j)} - b^{(j)}| \leq \sigma_2/k$, $j = 1, \dots, m$. Then, the sequence of solutions $\{\mathbf{z}_k\}$ to the sequence of optimisations $\{\tilde{P}_k^{NL}\}$ satisfies:*

$$\begin{aligned} & - \frac{\|\tilde{\lambda}_1 - \lambda^*\|_2^2}{2\alpha k} - \frac{\alpha}{2} \sigma_3^2 - 2m\alpha(\sigma_1 + \sigma_2)\sigma_3 \\ & \leq \frac{1}{k} \sum_{i=1}^k L(\mathbf{z}_i, \tilde{\lambda}_i) - f^* \leq 0. \end{aligned} \quad (18)$$

Also, for stochastic nonlinear constraints (where the \mathbf{b}_k are random variables) we have the following:

Corollary 2. *Consider the setup in Theorem 2. Suppose that sequence $\{\mathbf{b}_k\}$ is a realisation of a stochastic process $\{\mathbf{B}_k\}$ with mean $\mathbf{b} \in \mathbb{R}^m$. Let $p_k := \text{Prob}(\|\frac{1}{k} \sum_{i=1}^k \mathbf{B}_i - \mathbf{b}\|_\infty \leq m\sigma_2/k)$. Then with probability p_k for all $k \geq \bar{k}$ the sequence of solutions $\{\mathbf{z}_k\}$ to sequence of optimisations $\{\tilde{P}_k^{NL}\}$ satisfies (18).*

VI. EXAMPLE

Consider the simple network shown in Figure 2. Time is slotted with slots indexed by $k = 1, 2, \dots$ and at each time slot k sources A and B generate packets $s_k^{(j)}$, $j = 1, 2$ with probability $p^{(1)} = 0.66$ and $p^{(2)} = 0.33$ respectively that arrive at queues $\tilde{Q}^{(1)}$ and $\tilde{Q}^{(2)}$ of node N. At each time slot node N selects an action from a discrete set of actions $D := \{(0, 0), (0, 1), (1, 0)\}$, where $(1, 0)$ corresponds to sending a packet from $\tilde{Q}^{(1)}$, $(0, 1)$ to sending a packet from $\tilde{Q}^{(2)}$ and $(0, 0)$ to do nothing. The objective is to minimise the energy utility function $U(\mathbf{z}) := \sum_{j=1}^2 (z^{(j)})^2$ subject to the network rate constraints, where $z^{(1)}$ is the mean transmit rate of packets from source A and $z^{(2)}$ the mean transmit rate of packets from source B.

A. Convex problem formulation

We can formulate the network problem as a convex optimisation problem, i.e.,

$$\begin{aligned} & \underset{\mathbf{z} \in C}{\text{minimise}} && \mathbf{z}^T \mathbf{z} \\ & \text{subject to} && \mathbf{r} \preceq \mathbf{z} \end{aligned}$$

where $\mathbf{r} = [r^{(1)}, r^{(2)}]^T$, $\mathbf{z} = [z^{(1)}, z^{(2)}]^T$ and $C := \text{conv}(D)$. Notice that constraining $\mathbf{z} \in C$ enforces $r^{(1)} + r^{(2)} \leq 1$ in

order for the problem to be feasible. The Lagrangian is given by

$$L(\mathbf{z}, \boldsymbol{\lambda}) = \mathbf{z}^T \mathbf{z} + \boldsymbol{\lambda}^T (\mathbf{r} - \mathbf{z})$$

B. Dual subgradient method with discrete queue updates

We can solve the latter convex optimisation problem using the subgradient method for the dual problem [13]. That is, with updates

$$\begin{aligned} \mathbf{z}_k &\in \arg \min_{\mathbf{z} \in C} \mathbf{z}^T \mathbf{z} + \boldsymbol{\lambda}_k^T (\mathbf{r} - \mathbf{z}) \\ &= \arg \min_{\mathbf{z} \in C} \mathbf{z}^T \mathbf{z} - \boldsymbol{\lambda}_k^T \mathbf{z} \end{aligned} \quad (19)$$

$$\boldsymbol{\lambda}_{k+1} = [\boldsymbol{\lambda}_k + \alpha (\mathbf{r} - \mathbf{z}_k)]^+ \quad (20)$$

However, observe that update (20) requires previous knowledge of the mean rate arrivals. By Theorem 1 we can alternatively choose to solve the network problem using the following updates

$$\begin{aligned} \mathbf{z}_k &\in \arg \min_{\mathbf{z} \in C} \mathbf{z}^T \mathbf{z} + \tilde{\boldsymbol{\lambda}}_k^T (\mathbf{s}_k - \mathbf{z}) \\ &= \arg \min_{\mathbf{z} \in C} \mathbf{z}^T \mathbf{z} - \tilde{\boldsymbol{\lambda}}_k^T \mathbf{z} \end{aligned} \quad (21)$$

$$\tilde{\boldsymbol{\lambda}}_{k+1} = [\tilde{\boldsymbol{\lambda}}_k + \alpha (\mathbf{s}_k - \mathbf{x}_k)]^+ \quad (22)$$

where $\tilde{\boldsymbol{\lambda}}_k = \alpha \tilde{\mathbf{Q}}_k$, $\tilde{\mathbf{Q}}_{k+1} = [\tilde{\mathbf{Q}}_k - \mathbf{x}_k + \mathbf{s}_k]^+$, \mathbf{x}_k is the discrete action from set D selected by node N and \mathbf{s}_k the vector of packet arrivals at the queues at time slot k . Recall that the convergence of the sequence of updates obtained in Theorem 1 relies on the fact that $|\lambda_k^{(j)} - \tilde{\lambda}_k^{(j)}|$ is uniformly upper bounded for all $k \geq 1$, that is, by Lemma 2 we need that the sequences $\{\mathbf{z}_k\}$ and $\{\mathbf{x}_k\}$ remain close in appropriate sense and that $\frac{1}{k} \sum_{i=1}^k \mathbf{s}_i$ converges sufficiently fast to \mathbf{r} . In general, multiple sequences $\{\mathbf{x}_k\}$ might exist, and those can have different properties in terms of average queue backlog, packet delay, etc. Nonetheless, although constructing sequences with different properties is challenging and interesting on its own, this is outside of the scope of the present paper. For this particular example we construct sequence $\{\mathbf{x}_k\}$ as follows

$$\mathbf{x}_k \in \arg \min_{\mathbf{x} \in D} \left\| \sum_{i=1}^k \mathbf{z}_i - \sum_{i=1}^{k-1} \mathbf{x}_i - \mathbf{x} \right\|_{\infty}. \quad (23)$$

Figure 3 and 4 show the values of the multiplier $\boldsymbol{\lambda}_k$ and $\tilde{\boldsymbol{\lambda}}_k$ at each time slot k when we use updates (19) and (20) to solve the optimisation problem with parameter $\alpha = 10^{-2}$ and $\alpha = 10^{-3}$ respectively. It can be observed in both figures that the distance between the continuous and discrete valued multipliers is uniformly bounded, and that we can decrease the distance between both multipliers by choosing α small. Nevertheless, recall that we do not require to know $\boldsymbol{\lambda}$ in the optimisation, and that we can simply obtain $\tilde{\boldsymbol{\lambda}}_k$ by scaling the queue occupancy at time k by α . Figures 5 and 6 show the bound claimed in Theorem 1 for $\alpha = 10^{-2}$ and $\alpha = 10^{-3}$ respectively. The constants used in the lower bound are $\sigma_1 = 1$, $\sigma_2 = 10$, $\sigma_3 = 1$ and $\|\tilde{\boldsymbol{\lambda}}_1 - \boldsymbol{\lambda}^*\|_2^2 = 1$. It can be observed that the bound obtained with $\alpha = 10^{-3}$ is tighter than the one obtained when $\alpha = 10^{-2}$, however, at the cost of slower convergence.

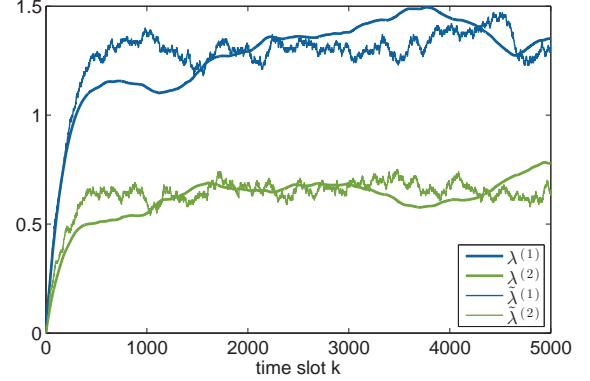


Fig. 3. Illustrating the discrete and continuous valued multipliers when step size $\alpha = 10^{-2}$ is used in the optimisation.

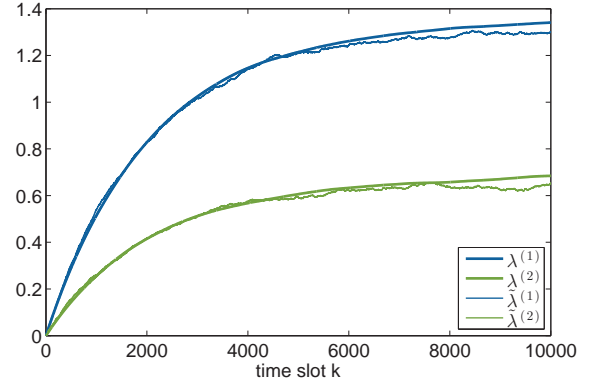


Fig. 4. Illustrating the discrete and continuous valued multipliers when step size $\alpha = 10^{-3}$ is used in the optimisation.

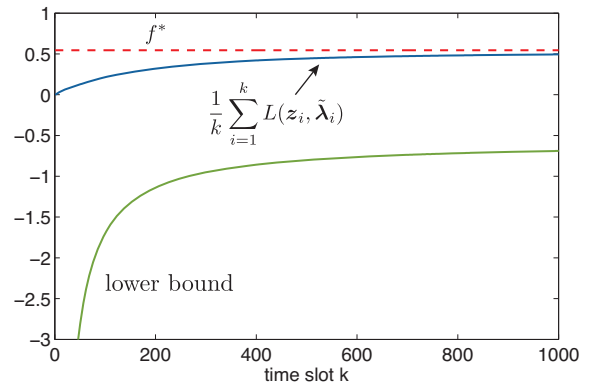


Fig. 5. Illustrating the bound of Theorem 1 when step size $\alpha = 10^{-2}$ is used in the optimisation.

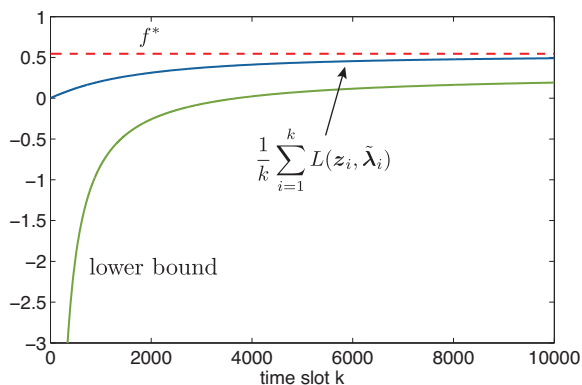


Fig. 6. Illustrating the bound of Theorem 1 when step size $\alpha = 10^{-3}$ is used in the optimisation.

VII. CONCLUSIONS

We show that the occupancy of appropriate queues can be used as a surrogate for Lagrange multipliers in convex optimisation. Our analysis uses only elementary methods, and is not asymptotic in nature. One immediate consequence is that in network problems the scaled link queue occupancy can be used as multipliers when calculating the dual function. Conversely, the connection with multipliers casts light on the link queue behaviour under optimal decision-making (not just max-weight scheduling). Namely, on links corresponding to active constraints the queue occupancy necessarily grows as step size α is reduced. Importantly, our analysis encompasses nonlinear constraints, and so generalises analysis beyond conventional queueing networks.

REFERENCES

- [1] M. Neely, "Distributed and secure computation of convex programs over a network of connected processors," in *Proc DCDIS Conf, Guelph, Ontario*, 2005.
- [2] L. Tassiulas and A. Ephremides, "Stability properties of constrained queueing systems and scheduling policies for maximum throughput in multihop radio networks," *Automatic Control, IEEE Transactions on*, vol. 37, no. 12, pp. 1936–1948, Dec 1992.
- [3] A. Eryilmaz and R. Srikant, "Fair resource allocation in wireless networks using queue-length-based scheduling and congestion control," *Networking, IEEE/ACM Transactions on*, vol. 15, no. 6, pp. 1333–1344, Dec 2007.
- [4] M. Neely, E. Modiano, and C. Rohrs, "Dynamic power allocation and routing for time-varying wireless networks," *Selected Areas in Communications, IEEE Journal on*, vol. 23, no. 1, pp. 89–103, Jan 2005.
- [5] M. Neely, "Energy optimal control for time-varying wireless networks," *Information Theory, IEEE Transactions on*, vol. 52, no. 7, pp. 2915–2934, July 2006.
- [6] M. Neely, E. Modiano, and C. ping Li, "Fairness and optimal stochastic control for heterogeneous networks," *Networking, IEEE/ACM Transactions on*, vol. 16, no. 2, pp. 396–409, April 2008.
- [7] M. Neely, *Stochastic network optimization with application to communication and queueing systems*. Morgan & Claypool Publishers, 2010.
- [8] L. Georgiadis, M. Neely, and L. Tassiulas, *Resource Allocation and Cross-Layer Control in Wireless Networks*. Morgan & Claypool Publishers, 2006.
- [9] X. Lin, N. Shroff, and R. Srikant, "A tutorial on cross-layer optimization in wireless networks," *IEEE J. Selected Areas in Communications*, vol. 24, no. 8, pp. 1452–1463, 2006.
- [10] L. Huang and M. Neely, "Delay reduction via lagrange multipliers in stochastic network optimization," *IEEE Trans Automatic Control*, vol. 56, no. 4, pp. 842–857, 2011.
- [11] S. P. Meyn, *Control techniques for complex networks*. Cambridge University Press, 2008.
- [12] A. Nedić and A. Ozdaglar, "Subgradient methods for saddle-point problems," *Journal of Optimization Theory and Applications*, vol. 142, no. 1, pp. 205–228, 2009.
- [13] D. P. Bertsekas, *Nonlinear programming*. Athena Scientific, 1999.