ISSN 0032-9460, Problems of Information Transmission, 2019, Vol. 55, No. 2, pp. 101–127. © Pleiades Publishing, Inc., 2019.
Russian Text © A.A. Puhalskii, 2019, published in Problemy Peredachi Informatsii, 2019, Vol. 55, No. 2, pp. 82–111.

The Geometry of Big Queues

A. A. Puhalskii

*Kharkevich Institute for Information Transmission Problems,
Russian Academy of Sciences, Moscow, Russia*
e-mail: puhalski@iitp.ru

Received August 29, 2018; revised January 14, 2019; accepted January 15, 2019

Abstract—In this paper, we use Hamilton equations in order to identify most likely scenarios of long queues being formed in ergodic Jackson networks. The associated Hamiltonians being discontinuous and piecewise Lipschitz, one has to invoke methods of nonsmoooth analysis. Time reversal of the Hamilton equations yields fluid equations for the dual network. Accordingly, the optimal trajectories are time reversals of the fluid trajectories of the dual network. Those trajectories are shown to belong to domains that satisfy a certain condition of being "essential". As an illustration, a two–station Jackson network is looked at. In addition, we prove certain properties of substochastic matrices which may be of interest in their own right.

Key words: queueing theory, Jackson networks, large deviations, large deviation principle, optimal trajectories, Hamilton equations, dual Markov processes, fluid dynamics.

DOI: 10.1134/S0032946019020054

1. OPTIMAL TRAJECTORIES IN JACKSON NETWORKS AND VARIATIONAL PROBLEMS WITH NONFIXED TIME

The Large Deviation Principle (LDP) allows one to determine rough asymptotics of the probabilities of a stochastic process assuming atypical values as well as to identify most likely scenarios of rare events unfolding. To that end, the variational problem of minimising the deviation function (which is also often referred to as "the rate function" or "the action functional") over the trajectories that result in those events needs to be solved. Usually, that is done by analysing Euler–Lagrange equations, cf., e.g., [1,2]. In this paper, with the Jackson network as a case in point, we explore the possibilities offered by solving Hamilton equations.

We concern ourselves with an (exponential) Jackson network of K stations with routing matrix $P=(p_{kl})$ of spectral radius less than unity, with exogenous arrival and service rates at the stations $\lambda_k>0$ and $\mu_k>0$, respectively, see, e.g., [3]. Let $Q_k(t)$ denote the queue length at station k at time t and let $Q(t)=(Q_1(t),\dots,Q_K(t))^T$, with ^{T} standing for the transposition operator. For $J\subset\{1,2,\dots,K\}$, we let F_J denote the set of vectors $x=(x_1,\dots,x_K)^T$ such that $x_i=0$ when $i\in J$ and $x_i>0$ when $i\notin J$. As it is done conventionally, $J^c=\{1,2,\dots,K\}\setminus J$. We also denote $\pi(u)=u\ln u-u+1$ provided $u>0$ and let $\pi(0)=1$. Let $\mathbf{1}_\Gamma$ represent the indicator function of set Γ which equals 1 on Γ and equals 0 outside Γ. We denote by $\mathbb{S}_+^{K\times K}$ the set of (row) substochastic $K\times K$–matrices, by I, the identity $K\times K$–matrix, and by $\mathbb{D}(\mathbb{R}_+,\mathbb{R}_+^K)$, the Skorohod space of right continuous with lefthand limits functions with domain $\mathbb{R}_+$ and range $\mathbb{R}_+^K$, which is equipped with a metric that makes it a complete metric separable space, cf., Jacod and Shiryaev [4], Ethier and Kurtz [5]. The following result is found in [6]. (We assume that the product of the function π and 0 equals zero even when the argument of π equals infinity and that $0/0=0$.)

Theorem 1. *Let $Q(0)=0\in\mathbb{R}_+^K$. Then, as $n\to\infty$, the processes $(Q(nt)/n,\ t\in\mathbb{R}_+)$ satisfy an LDP in $\mathbb{D}(\mathbb{R}_+,\mathbb{R}_+^K)$ with deviation function $\boldsymbol{I}(q)$ that, for absolutely continuous functions $q=$*

$(q(t),\, t \in \mathbb{R}_+)$ *, such that* $q(0) = 0 \in \mathbb{R}_+^K$ *, is of the form*

$$\boldsymbol{I}(q) = \int\limits_0^\infty L(q(t), \dot{q}(t))\, dt,$$

where

$$L(x,y) = \sum_{J \subset \{1,2,\ldots,K\}} \mathbf{1}_{F_J}(x) L_J(y),$$
$$L_J(y) = \inf_{\substack{(a,d,\varrho)\in \mathbb{R}_+^K \times \mathbb{R}_+^K \times \mathbb{S}_+^{K\times K}:\\ y = a + (\varrho^T - I)d}} \psi_J(a,d,\varrho), \tag{1.1}$$

and

$$\psi_J(a,d,\varrho) = \sum_{k=1}^K \pi\Big(\frac{a_k}{\lambda_k}\Big)\lambda_k + \sum_{k\in J^c} \pi\Big(\frac{d_k}{\mu_k}\Big)\mu_k + \sum_{k\in J} \pi\Big(\frac{d_k}{\mu_k}\Big)\mathbf{1}_{(\mu_k,\infty)}(d_k)\mu_k$$
$$+ \sum_{k=1}^K d_k \left[\sum_{\ell=1}^K \pi\Big(\frac{\varrho_{k\ell}}{p_{k\ell}}\Big) p_{k\ell} + \pi \left(\frac{1 - \sum\limits_{\ell=1}^K \varrho_{k\ell}}{1 - \sum\limits_{\ell=1}^K p_{k\ell}} \right) \left(1 - \sum_{\ell=1}^K p_{k\ell} \right) \right]. \tag{1.2}$$

If either the function q *is not absolutely continuous or* $q(0) \neq 0$ *, then* $\boldsymbol{I}(q) = \infty$*.*

In more detail, the theorem states that the sets $\{q \in \mathbb{D}(\mathbb{R}_+, \mathbb{R}^K) : \mathbf{I}(q) \le u\}$ are compact for all $u \ge 0$ and the following inequalities hold:

$$\liminf_{n\to\infty} \frac{1}{n} \ln \mathbf{P}(q \in G) \ge -\inf_{q\in G} \boldsymbol{I}(q)$$

provided G is an open subset of $\mathbb{D}(\mathbb{R}_+, \mathbb{R}_+^K)$ and

$$\limsup_{n\to\infty} \frac{1}{n} \ln \mathbf{P}(q \in F) \le -\inf_{q\in F} \boldsymbol{I}(q)$$

provided F is a closed subset of $\mathbb{D}(\mathbb{R}_+, \mathbb{R}_+^K)$. We note that the LDP of Theorem 1.1 appears also in Atar and Dupuis[7], and in Ignatiouk–Robert [8], with L_J being expressed differently.

The LDP implies that, generally, the logarithm of the probability of $Q(nt)/n$ visiting given set S at large time is close to $-\inf_{q\in S'} \mathbf{I}(q)$, where $S' = \{q : q(0) = 0, q(T) \in S \text{ for some } T > 0\}$. If the latter infimum is attained at a unique function q, then, with a probability close to unity, $Q(nt)/n$ reaches the set S by travelling in a small neighbourhood of that function. In this paper, we are concerned with finding for an ergodic Jackson network function q that starts at the origin and reaches a given point at some time T such that $\int\limits_0^T L(q(t), \dot{q}(t))\, dt$ is minimised. We show that the time reversal of the Hamilton equations for the optimal trajectories yields fluid equations for the dual network. This observation proves, for the Jackson network, "the folklore theorem" that the most probable "large deviation" trajectory of an ergodic Markov process is the time reversal of the fluid limit of the dual process, cf., [9]. (In statistical physics, a similar property is referred to as the extended Onzager – Machlup principle and states that a fluctuation trajectory is the time–reversed relaxation trajectory of the dual dynamic, cf., [10].) We show that the optimal trajectory is a (conditional) law–of–large–numbers limit, provided that the network queues become atypically large. We find necessary and sufficient conditions for a given face F_J to contain a part of an optimal trajectory, hence, a part of a fluid trajectory. The paper concludes with a discussion of geometric aspects of finding optimal trajectories by analysing a two–station ergodic Jackson network. In addition, we establish certain properties of substochastic matrices which may be of interest in their own right.

As functions $L_J(y)$ are convex by Lemma 4.2 in [6], Jensen's inequality implies that in a constant dynamic domain, i.e., in face F_J, it is opimal to move in a straight line. Let us note also that if $J' \subset J''$, then $\psi_{J'}(a,d,\varrho) \geq \psi_{J''}(a,d,\varrho)$, so, $L_{J'}(y) \geq L_{J''}(y)$. Therefore, if the initial and terminal points of a piece of a trajectory belong to F_J, then going in a straight line is "advantageous" to moving in adjacent faces $F_{J'}$, where $J' \subset J$. We will show that, furthermore, the optimal trajectories consist of finitely many straight line segments which belong to faces F_J such that the sets J get smaller, provided one disregards the terminal point, the latter not always conforming to that requirement. As a consequence, each of the faces F_J contains one line segment, at most.

In order to identify optimal trajectories in the faces F_J, we will invoke the following general result. For a locally Lipschitz function $f(y)$, we will denote by $\partial f(y)$ the generalised gradient of f at y, cf., [11]. If the function f is convex, which is the case for the applications in this paper, then the generalised gradient reduces to the subdifferential, cf., [12]. The definition of the normal cone to a nonempty closed set at a point of that set as a dual cone to the tangent cone at that point appears in [11]. For convex sets, it reduces to the definition in [12] that it is a collection of normal vectors to the set at that point, with vector z being called normal to convex set S at point $x \in S$, provided $z \cdot (y-x) \leq 0$ for all $y \in S$. (Here and below, $\cdot$ denotes the inner product. Besides, "a.e." stands for "almost everywhere with respect to the Lebesgue measure".)

Lemma 1. *Let function $L(x,y)$ map $\mathbb{R}^k \times \mathbb{R}^k$ into $\mathbb{R}$, be locally Lipschitz in (x,y) and be convex in y. Let us assume that the function*

$$H(x,p) = \sup_{y \in \mathbb{R}^k} (p \cdot y - L(x,y)) \tag{1.3}$$

is locally Lipschitz (in (x,p)). Let $g(x)$ represent a locally Lipschitz upper semicontinuous function. Let S represent a convex closed subset of $\mathbb{R}^k$ and let $x_0 \in \mathbb{R}^k$. If the minimum of $\int_0^T L(x(t),\dot{x}(t))\,dt$ over $T \geq 0$ and over absolutely continuous functions $x(t)$such that $x(0) = x_0$, $x(T) \in S$ and $g(x(t)) \leq 0$ for all $t \leq T$, is attained by $T = T^*$ and $x(t) = x^*(t)$, then there exist measure μ on $[0,T^*]$, a μ–measurable function $\gamma(t)$ and an absolutely continuous function $p(t)$, which both assume values in $\mathbb{R}^k$, such that the following properties hold:

1 $\gamma(t) \in \partial^{>} g(x^*(t))$ *for μ–almost all $t \in [0,T^*]$ and the support of μ belongs to the set* $\{t \in [0,T^*] : \partial^{>} g(x^*(t)) \neq \varnothing\}$ *, with $\partial^{>} g(x)$ being the convex hull of the limits of sequences γ_i such that $\gamma_i \in \partial g(x_i)$, $x_i \to x$, and $g(x_i) > 0$;*

2 $\begin{bmatrix} -\dot{p}(t) \\ \dot{x}^*(t) \end{bmatrix} \in \partial H\left(x^*(t), p(t) + \int_0^t \gamma(s)\,d\mu(s)\right)$ *a.e. on* $[0,T^*]$;

3 $H\left(x^*(t), p(t) + \int_0^t \gamma(s)\,d\mu(s)\right) = 0$ *on* $[0,T^*]$;

4 $p(T^*) + \int_0^{T^*} \gamma(s)\,d\mu(s) \in -N_S(x^*(T^*))$;

5

$$p(T^*) + \int_0^{T^*} \gamma(s)\,d\mu(s) \in -N_S(x^*(T^*)),$$

where $N_S(x^*(T^*))$ represents the normal cone to S at $x^*(T^*)$.

A proof of the lemma is provided in the Appendix. The lemma is useful because, unlike the Lagrangian L_J, the associated Hamiltonian $H_J(\theta) = \sup_{y\in\mathbb{R}^K}(\theta\cdot y - L_J(y))$ is expressed explicitly, where $\theta\in\mathbb{R}^K$. Let, for $\theta = (\theta_1,\dots,\theta_K)^T$ and $k = 1,2,\dots,K$,

$$h_k(\theta) = e^{-\theta_k}\left(\sum_{\ell=1}^K (e^{\theta_\ell}-1)p_{k\ell}+1\right) - 1. \tag{1.4}$$

The calculations in the Appendix show that

$$L_J(y) = \sup_{\theta\in\mathbb{R}^K}\left(\sum_{k=1}^K \theta_k y_k - \sum_{k=1}^K (e^{\theta_k}-1)\lambda_k - \sum_{k\in J} h_k(\theta)^+\mu_k - \sum_{k\in J^c} h_k(\theta)\mu_k\right), \tag{1.5}$$

where we denote $u^+ = \max(u,0)$. It follows that the function L_J is convex and lower semicontinuous. In addition,

$$H_J(\theta) = \sum_{k=1}^K (e^{\theta_k}-1)\lambda_k + \sum_{k\in J} h_k(\theta)^+\mu_k + \sum_{k\in J^c} h_k(\theta)\mu_k. \tag{1.6}$$

Let us also denote

$$H_0(\theta) = H_\varnothing(\theta) = \sum_{k=1}^K (e^{\theta_k}-1)\lambda_k + \sum_{k=1}^K h_k(\theta)\mu_k. \tag{1.7}$$

The next assertion follows from Lemma 1 with $L(x,y) = L_J(y)$ and $g(x) = \sum_{i\in J} x_i^2$ so that the measure μ equals zero.

Lemma 2. *Optimal motion in the face F_J takes place with constant adjoint variable θ, such that $H_J(\theta) = 0$, whereas an optimal trajectory satisfies the equation*

$$\dot q(t) \in \partial H_J(\theta) \quad a.e. \tag{1.8}$$

This lemma does not address the issue of the existence of optimal trajectories. Unfortunately, by contrast with a fixed time case, in the setup under consideration, the compactness of the set $\{q : \boldsymbol{I}(q) \le u\}$ does not entail the existence of optimal trajectories, where $u \ge 0$. Below, we mostly assume that the optimal trajectories exist and study their properties. The existence and uniqueness will be proved via *ad hoc* arguments.

2. PROPERTIES OF THE HAMILTONIAN

In this section, we identify candidates for optimal adjoint variables θ for motion in face F_J. We need to introduce additional pieces of notation. Given square matrix B, we denote by $B(i\,|\,j)$ the matrix that is obtained by deleting the i-th row and j-th column of B. Similarly, $b_{\cdot\ell}$ denotes the ℓ-th column of B with the ℓ-th entry deleted, and $b_{\ell\cdot}$ denotes the ℓ-th row of B with the ℓ-th entry deleted.

Let vector $\theta^{(m)} = (\theta_1^{(m)},\dots,\theta_K^{(m)})^T \ne 0$ satisfy the equations $h_k(\theta^{(m)}) = 0$, for $k\ne m$, and $H_0(\theta^{(m)}) = 0$, where $m\in\{1,2,\dots,K\}$. By (1.4), the ratios $a_{m\ell} = (e^{\theta_\ell^{(m)}}-1)/(e^{\theta_m^{(m)}}-1)$, for $m\ne\ell$, satisfy the system of equations

$$a_{mk} - \sum_{\ell\ne m} a_{m\ell}p_{k\ell} = p_{km}, \quad k\ne m. \tag{2.1}$$

(We note that if $\theta_m^{(m)} = 0$, then, owing to (1.4), $\theta_k^{(m)} = 0$ for all k .) Equations (2.1) admit the unique solution

$$a_{m\cdot}^T = \big((I-P)(m\,|\,m)\big)^{-1} p_{\cdot m}. \tag{2.2}$$

Also, $a_{mm} = 1$. The $a_{m\ell}$ are specified by these requirements uniquely and are nonnegative. Substitution in the equation $H_0(\theta^{(m)}) = 0$ (cf. (1.7)) implies that

$$e^{\theta_m^{(m)}} = \frac{1 - \sum\limits_{\ell=1}^{K} a_{m\ell} p_{m\ell}}{\sum\limits_{\ell=1}^{K} a_{m\ell}\lambda_\ell}\mu_m. \tag{2.3}$$

Let $\theta^* = (\theta_1^{(1)}, \theta_2^{(2)}, \ldots, \theta_K^{(K)})^T$ and

$$\nu = (I - P^T)^{-1}\lambda, \tag{2.4}$$

where $\lambda = (\lambda_1, \ldots, \lambda_K)^T$. The vector $\nu = (\nu_1, \ldots, \nu_K)^T$ is the vector of mean flow rates through the network's stations in steady state. Let $\rho_m = \nu_m/\mu_m$, where $m \in \{1, 2, \ldots, K\}$. It is a byproduct of the following theorem that the hyperplanes $\{\theta :\ \theta_m = \theta_m^{(m)}\}$ intersect at a point that belongs to the surface $H_0(\theta) = 0$.

Theorem 2. *We have that $H_0(\theta^*) = 0$ and $e^{-\theta_m^{(m)}} = \rho_m$, for $m \in \{1, 2, \ldots, K\}$.*

The proof is preceded by a number of lemmas. Suppose that $B = (b_{ij})$ is a $(K \times K)$-matrix with a nonzero determinant and nonzero principal minors. We use adj to denote an adjoint matrix and use det to denote a determinant. Let us denote by $M_{ij}(\ell\,|\,\ell)$ the (i,j)-minor of $B(\ell\,|\,\ell)$. Given $\ell \neq m$, we introduce

$$f_{m\ell} = \begin{cases} e_m & \text{if } m < \ell, \\ e_{m-1} & \text{if } m > \ell, \end{cases} \tag{2.5}$$

where e_i represents the i-th vector of the standard basis in $\mathbb{R}^{K-1}$.

Lemma 3. *If $\ell \neq m$, then*

$$f_{m\ell}^T \operatorname{adj}(B(\ell\,|\,\ell))b_{\cdot\ell} = (-1)^{m+\ell+1}\det(B(\ell\,|\,m)).$$

Proof. Let $\ell > m$. We have that

$$e_m^T \operatorname{adj}(B(\ell\,|\,\ell))b_{\cdot\ell} = \sum_{j=1}^{\ell-1}(-1)^{m+j}M_{jm}(\ell\,|\,\ell)b_{j\ell} + \sum_{j=\ell+1}^{K}(-1)^{m+j-1}M_{j-1,m}(\ell\,|\,\ell)b_{j\ell}.$$

Since $M_{jm}(\ell\,|\,\ell) = M_{j,\ell-1}(\ell\,|\,m)$ when $j \le \ell - 1$ and $M_{j-1,m}(\ell\,|\,\ell) = M_{j-1,\ell-1}(\ell\,|\,m)$ when $j \ge \ell + 1$, we have that $e_m^T \operatorname{adj}(B(\ell\,|\,\ell))b_{\cdot\ell} = (-1)^{m+\ell+1}\det(B(\ell\,|\,m))$.

Let us assume that $\ell < m$. By an analogous reasoning,

$$\begin{aligned} e_{m-1}^T \operatorname{adj}(B(\ell\,|\,\ell))b_{\cdot\ell} &= \sum_{j=1}^{\ell-1}(-1)^{m+j-1}M_{j,m-1}(\ell\,|\,\ell)b_{j\ell} + \sum_{j=\ell+1}^{K}(-1)^{m+j}M_{j-1,m-1}(\ell\,|\,\ell)b_{j\ell} \\ &= \sum_{j=1}^{\ell-1}(-1)^{m+j-1}M_{j\ell}(\ell\,|\,m)b_{j\ell} + \sum_{j=\ell+1}^{K}(-1)^{m+j}M_{j-1,\ell}(\ell\,|\,m)b_{j\ell} \\ &= (-1)^{m+\ell+1}\det(B(\ell\,|\,m)). \quad \triangle \end{aligned}$$

Lemma 4. *Let $\ell \in \{1, 2, \ldots, K\}$. Then*

$$b_{\ell\ell}\det(B(\ell\,|\,\ell)) - b_{\ell\cdot}\operatorname{adj}(B(\ell\,|\,\ell))b_{\cdot\ell} = \det(B).$$

Proof. Since

$$b_{\ell\cdot}\,\mathrm{adj}(B(\ell\,|\,\ell))b_{\cdot\ell}=\sum_{j\neq\ell}b_{\ell j}f_{j\ell}^T\,\mathrm{adj}(B(\ell\,|\,\ell))b_{\cdot\ell},$$

owing to Lemma 3, we have that

$$b_{\ell\cdot}\,\mathrm{adj}(B(\ell\,|\,\ell))b_{\cdot\ell}=\sum_{j\neq\ell}b_{\ell j}(-1)^{j+\ell+1}\det(B(\ell\,|\,j))=-\det(B)+b_{\ell\ell}\det(B(\ell\,|\,\ell)).\quad\triangle$$

Lemma 5. *The following equations hold:*

$$\frac{b_{m\cdot}\big(B(m\,|\,m)\big)^{-1}b_{\cdot m}}{b_{mm}-b_{m\cdot}\big(B(m\,|\,m)\big)^{-1}b_{\cdot m}}=\sum_{\ell\neq m}\frac{b_{\ell m}f_{m\ell}^T\big(B(\ell\,|\,\ell)\big)^{-1}b_{\cdot\ell}}{b_{\ell\ell}-b_{\ell\cdot}\big(B(\ell\,|\,\ell)\big)^{-1}b_{\cdot\ell}}\tag{2.6}$$

and

$$-\frac{b_{mm}f_{\ell m}^T(B(m\,|\,m))^{-1}b_{\cdot m}}{b_{mm}-b_{m\cdot}(B(m\,|\,m))^{-1}b_{\cdot m}}=-\frac{b_{\ell m}}{b_{\ell\ell}-b_{\ell\cdot}(B(\ell\,|\,\ell))^{-1}b_{\cdot\ell}}+\sum_{\substack{k\neq\ell,\\k\neq m}}\frac{b_{km}f_{lk}^T(B(k\,|\,k))^{-1}b_{\cdot k}}{b_{kk}-b_{k\cdot}(B(k\,|\,k))^{-1}b_{\cdot k}}.\tag{2.7}$$

Proof. On multiplying the numerators and denominators on both sides of (2.6) by the determinants of the matrices being inverted, we have that (2.6) holds if and only if

$$\frac{b_{m\cdot}\,\mathrm{adj}(B(m\,|\,m))b_{\cdot m}}{b_{mm}\det(B(m\,|\,m))-b_{m\cdot}\,\mathrm{adj}(B(m\,|\,m))b_{\cdot m}}=\sum_{\ell\neq m}\frac{b_{\ell m}f_{m\ell}^T\,\mathrm{adj}(B(\ell\,|\,\ell))b_{\cdot\ell}}{b_{\ell\ell}\det(B(\ell\,|\,\ell))-b_{\ell\cdot}\,\mathrm{adj}(B(\ell\,|\,\ell))b_{\cdot\ell}}.\tag{2.8}$$

By Lemma 4, the denominators in (2.8) equal $\det(B)\neq 0$. Hence, one needs to prove that

$$b_{m\cdot}\,\mathrm{adj}(B(m\,|\,m))b_{\cdot m}=\sum_{\ell\neq m}b_{\ell m}f_{m\ell}^T\,\mathrm{adj}(B(\ell\,|\,\ell))b_{\cdot\ell}.$$

An application of Lemma 3 yields

$$\begin{aligned}\sum_{\ell\neq m}b_{\ell m}f_{m\ell}^T\,\mathrm{adj}(B(\ell\,|\,\ell))b_{\cdot\ell}&=\sum_{\ell\neq m}b_{\ell m}(-1)^{m+\ell+1}\det(B(\ell\,|\,m))\\&=-\det(B)+b_{mm}\det(B(m\,|\,m)),\end{aligned}$$

which concludes the proof of (2.6) owing to Lemma 4.

In analogy with the proof of (2.6), (2.7) is equivalent to the equation

$$-b_{mm}f_{\ell m}^T\,\mathrm{adj}(B(m\,|\,m))b_{\cdot m}=-b_{\ell m}\det(B(\ell\,|\,\ell))+\sum_{\substack{k\neq\ell,\\k\neq m}}b_{km}f_{lk}^T\,\mathrm{adj}(B(k\,|\,k))b_{\cdot k}.\tag{2.9}$$

By Lemma 3,

$$\sum_{k\neq\ell}b_{km}f_{lk}^T\,\mathrm{adj}(B(k\,|\,k))b_{\cdot k}=\sum_{k\neq\ell}b_{km}(-1)^{k+\ell+1}\det(B(k\,|\,\ell)).\tag{2.10}$$

Since the sum on the lefthand side of (2.10) is the determinant of the matrix that is obtained from the matrix B by replacing the ℓ-th column with the m-th column, we have that $\sum_{k=1}^{K}b_{km}(-1)^{k+\ell}\det(B(k\,|\,\ell))=0$. Therefore, the righthand side of (2.10) equals $b_{\ell m}\det(B(\ell\,|\,\ell))$. The equation that results is equivalent to (2.9). $\triangle$

Proof of Theorem 2. We start by proving the second part. Let us show that

$$\frac{1}{1-\sum\limits_{\ell=1}^{K} a_{m\ell}p_{m\ell}} = 1 + \sum_{k=1}^{K} \frac{a_{km}p_{km}}{1-\sum\limits_{\ell=1}^{K} a_{k\ell}p_{k\ell}} \tag{2.11}$$

and

$$\frac{a_{m\ell}}{1-\sum\limits_{k=1}^{K} a_{mk}p_{mk}} = \sum_{k=1}^{K} \frac{a_{k\ell}p_{km}}{1-\sum\limits_{\ell'=1}^{K} a_{k\ell'}p_{k\ell'}} \quad \text{provided that } \ell \neq m. \tag{2.12}$$

Since $a_{mm}=1$, equation (2.11) can be written as

$$\frac{\sum\limits_{\ell\neq m} a_{m\ell}p_{m\ell}}{1-p_{mm}-\sum\limits_{\ell\neq m} a_{m\ell}p_{m\ell}} = \sum_{k\neq m} \frac{a_{km}p_{km}}{1-p_{kk}-\sum\limits_{\ell\neq k} a_{k\ell}p_{k\ell}}. \tag{2.13}$$

Equations (2.1), (2.2) and (2.5) imply that $a_{m\ell} = f_{\ell m}^T((I-P)(m\,|\,m))^{-1}p_{\cdot m}$ and $a_{km} = f_{mk}^T((I-P)(k\,|\,k))^{-1}p_{\cdot k}$. Therefore, (2.13) can be written as

$$\frac{p_{m\cdot}\big((I-P)(m\,|\,m)\big)^{-1}p_{\cdot m}}{1-p_{mm}-p_{m\cdot}\big((I-P)(m\,|\,m)\big)^{-1}p_{\cdot m}} = \sum_{k\neq m} \frac{p_{km}f_{mk}^T\big((I-P)(\ell\,|\,\ell)\big)^{-1}p_{\cdot k}}{1-p_{kk}-p_{k\cdot}\big((I-P)(\ell\,|\,\ell)\big)^{-1}p_{\cdot k}},$$

which is a special case of (2.6) in Lemma 5 when $B = I-P$. Equation (2.11) is proved.

Let us prove (2.12). We write this equation as

$$\frac{a_{m\ell}(1-p_{mm})}{1-p_{mm}-\sum\limits_{k\neq m} a_{mk}p_{mk}} = \frac{p_{\ell m}}{1-p_{\ell\ell}-\sum\limits_{\ell'\neq \ell} a_{\ell\ell'}p_{\ell\ell'}} + \sum_{\substack{k\neq \ell,\\ k\neq m}} \frac{a_{k\ell}p_{km}}{1-p_{kk}-\sum\limits_{\ell'\neq k} a_{k\ell'}p_{k\ell'}}, \quad \text{where } \ell\neq m. \tag{2.14}$$

As above, by (2.2) and (2.5), equation (2.14) is equivalent to the equation

$$\frac{(1-p_{mm})f_{\ell m}^T((I-P)(m\,|\,m))^{-1}p_{\cdot m}}{1-p_{mm}-p_{m\cdot}((I-P)(m\,|\,m))^{-1}p_{\cdot m}} = \frac{p_{\ell m}}{1-p_{\ell\ell}-p_{\ell\cdot}((I-P)(\ell\,|\,\ell))^{-1}p_{\cdot \ell}} + \sum_{\substack{k\neq \ell,\\ k\neq m}} \frac{p_{km}f_{lk}^T((I-P)(k\,|\,k))^{-1}p_{\cdot k}}{1-p_{kk}-p_{k\cdot}((I-P)(k\,|\,k))^{-1}p_{\cdot k}}$$

so that is holds owing to (2.7) in Lemma 5. This concludes the proof of (2.12).

Let

$$c_{mk} = \frac{a_{mk}}{1-\sum\limits_{\ell=1}^{K} a_{m\ell}p_{m\ell}} \tag{2.15}$$

and $C = (c_{mk})$. Let us show that

$$C = (I-P^T)^{-1}. \tag{2.16}$$

By (2.15), equation (2.12) assumes the form $c_{m\ell} = \sum\limits_{k=1}^{K} p_{km}c_{k\ell}$ so that the (m,ℓ)-entries of the matrices C and P^TC agree when $m\neq \ell$. Since $a_{mm}=1$, owing to (2.15), (2.11) can be written as $c_{mm} = 1 + \sum_{k=1}^{K} p_{km}c_{km}$. Consequently, the diagonal entries of the matrices C and $I + P^T C$ agree. Thus, $C = I + P^T C$, which proves (2.16). Since $a_{mm} = 1$, we have, by (2.15), that $1 - \sum_{\ell=1}^{K} a_{m\ell}p_{m\ell} = 1/c_{mm}$ and

$$a_{mk} = \frac{c_{mk}}{c_{mm}}. \tag{2.17}$$

Accounting for (2.3), (2.4) and (2.16), we obtain that $e^{\theta_m^{(m)}} = \mu_m \,/\, \sum_{k=1}^{K} c_{mk}\lambda_k = \mu_m/\nu_m = 1/\rho_m$.

In order to conclude the proof of Theorem 2, we note that, by (1.4), (1.7) and the definition of θ^*,

$$H_0(\theta^*) = \sum_{m=1}^{K} (e^{\theta_m^{(m)}} - 1)\Bigg(\lambda_m + \sum_{k=1}^{K} e^{-\theta_k^{(k)}} p_{km}\mu_k - e^{-\theta_m^{(m)}}\mu_m\Bigg).$$

Since $e^{-\theta_k^{(k)}} = \nu_k/\mu_k$ and $\nu = \lambda + P^T\nu$, we have that $\lambda_m + \sum_{k=1}^{K} e^{-\theta_k^{(k)}} p_{km}\mu_k - e^{-\theta_m^{(m)}}\mu_m = 0$. △

Let $\widetilde{\theta}^{(J)} = \big(\widetilde{\theta}_1^{(J)}, \ldots, \widetilde{\theta}_K^{(J)}\big)^T$ be defined by the requirements that $\widetilde{\theta}_k^{(J)} = \theta_k^{(k)}$ when $k \notin J$ and $h_k(\widetilde{\theta}^{(J)}) = 0$ when $k \in J$. It is noteworthy that if $J = \{1, 2, \ldots, K\} \setminus \{m\}$, then $\widetilde{\theta}^{(J)} = \theta^{(m)}$. By (1.6) and (1.7),

$$H_J(\widetilde{\theta}^{(J)}) = H_0(\widetilde{\theta}^{(J)}). \tag{2.18}$$

It will be shown in Lemma 6 that $H_0(\widetilde{\theta}^{(J)}) = 0$. Since $\widetilde{\theta}^{(J)} \cdot x = \theta^* \cdot x$ when $x \in F_J$ and (2.18) holds, we have that, when moving in faces $F_{J_1}, \ldots, F_{J_k}$ with respective adjoint variables $\widetilde{\theta}^{(J)}, \ldots, \widetilde{\theta}^{(J_k)}$ and with displacements $s_1, \ldots, s_k$ during time lengths $t_1, \ldots, t_k$, respectively, the total cost is

$$\begin{aligned}\sum_{\ell=1}^{k} t_l L_{J_\ell}\Big(\frac{s_\ell}{t_\ell}\Big) &= \sum_{\ell=1}^{k} \sup_{\theta\in\mathbb{R}^K} (\theta \cdot s_\ell - t_l H_{J_\ell}(\theta)) \\ &\ge \sum_{\ell=1}^{k} (\widetilde{\theta}^{(J_\ell)} \cdot s_\ell - t_l H_{J_\ell}(\widetilde{\theta}^{(J_\ell)})) \\ &= \sum_{\ell=1}^{k} (\theta^* \cdot s_\ell - t_l H_0(\widetilde{\theta}^{(J_\ell)})) = \theta^* \cdot (s_1 + \ldots + s_k) = \theta^* \cdot r,\end{aligned} \tag{2.19}$$

where r represents the destination. In a general situation, if $q(0) = 0$ and $q(t) = r$, then, owing to the fact that $\dot{q}_k(s) = 0$ for $k \in J$ a.e. on the set $\{s :\ q(s) \in F_J\}$,

$$\begin{aligned}\int_0^t L(q(s), \dot{q}(s))\, ds &= \sum_J \int_0^t \mathbf{1}_{F_J}(q(s)) L_J(\dot{q}(s))\, ds \\ &\ge \sum_J \int_0^t \mathbf{1}_{F_J}(q(s)) (\widetilde{\theta}^{(J)} \cdot \dot{q}(s) - H_0(\widetilde{\theta}^{(J)}))\, ds \\ &= \sum_J \int_0^t \mathbf{1}_{F_J}(q(s)) (\theta^* \cdot \dot{q}(s) - H_0(\widetilde{\theta}^{(J)}))\, ds = \theta^* \cdot r.\end{aligned} \tag{2.20}$$

If a trajectory from O to r is such that (2.19) and (2.20) hold with equalities, i.e., (1.8) hold for $\theta = \widetilde{\theta}^{(J)}$, then that trajectory is optimal. On the other hand, since the functions $H_J(\theta)$ are strictly convex so that the derivatives of those functions are injective, if, for some J such that

$\int_0^t \mathbf{1}_{F_J}(q(s))\,ds > 0$, (1.8) with $\theta = \widetilde{\theta}^{(J)}$ does not hold, then (2.20) holds with strict inequality. Therefore, the lower bound in (2.20) is attained only if the motion in the face F_J a.e. occurs with the adjoint variable $\theta = \widetilde{\theta}^{(J)}$. The next lemma extends the property that $H_0(\theta^*) = 0$.

Lemma 6. *We have that* $H_0(\widetilde{\theta}^{(J)}) = 0$.

Proof. Owing to (1.4), (1.7), the definition of $\widetilde{\theta}^{(J)}$ and the equation $e^{-\theta_m^{(m)}} = \rho_m$,

$$\begin{aligned} H_0(\widetilde{\theta}^{(J)}) &= \sum_{m\notin J} (e^{\theta_m^{(m)}} - 1)\lambda_m + \sum_{m\in J} (e^{\widetilde{\theta}_m^{(J)}} - 1)\lambda_m \\ &\quad + \sum_{k\notin J} \left(e^{-\theta_k^{(k)}} \left(\sum_{m\notin J} (e^{\theta_m^{(m)}} - 1) p_{km} + \sum_{m\in J} (e^{\widetilde{\theta}_m^{(J)}} - 1) p_{km} + 1 \right) - 1 \right) \mu_k \\ &= \sum_{m\notin J} (e^{\theta_m^{(m)}} - 1) \left(\lambda_m + \sum_{k\notin J} e^{-\theta_k^{(k)}} p_{km}\mu_k - e^{-\theta_m^{(m)}} \mu_m \right) \\ &\quad + \sum_{m\in J} (e^{\widetilde{\theta}_m^{(J)}} - 1) \left(\lambda_m + \sum_{k\notin J} e^{-\theta_k^{(k)}} p_{km}\mu_k \right) \\ &= \sum_{m\notin J} (e^{\theta_m^{(m)}} - 1) \left(\lambda_m + \sum_{k\notin J} p_{km}\nu_k - \nu_m \right) + \sum_{m\in J} (e^{\widetilde{\theta}_m^{(J)}} - 1) \left(\lambda_m + \sum_{k\notin J} p_{km}\nu_k \right). \end{aligned} \tag{2.21}$$

Since due to the definition of ν (cf. (2.4)) $\lambda_m + \sum_{k\notin J} p_{km}\nu_k = \nu_m - \sum_{k\in J} p_{km}\nu_k$, we have that

$$\sum_{m\in J} (e^{\widetilde{\theta}_m^{(J)}} - 1) \left(\lambda_m + \sum_{k\notin J} p_{km}\nu_k \right) = \sum_{m\in J} (e^{\widetilde{\theta}_m^{(J)}} - 1)\nu_m - \sum_{k\in J} \nu_k \sum_{m\in J} (e^{\widetilde{\theta}_m^{(J)}} - 1) p_{km}. \tag{2.22}$$

As $h_k(\widetilde{\theta}^{(J)}) = 0$ when $k \in J$, it follows that

$$\sum_{m\in J} (e^{\widetilde{\theta}_m^{(J)}} - 1) p_{km} = e^{\widetilde{\theta}_k^{(J)}} - 1 - \sum_{m\notin J} (e^{\theta_m^{(m)}} - 1) p_{km}. \tag{2.23}$$

Hence, the righthand side of (2.22) equals $\sum_{m\notin J} (e^{\theta_m^{(m)}} - 1) \sum_{k\in J} p_{km}\nu_k$. By (2.21),

$$H_0(\widetilde{\theta}^{(J)}) = \sum_{m\notin J} (e^{\theta_m^{(m)}} - 1) \left(\lambda_m + \sum_{k=1}^{K} p_{km}\nu_k - \nu_m \right).$$

Since, by (2.4), $\lambda_m + \sum_{k=1}^{K} p_{km}\nu_k - \nu_m = 0$, it follows that $H_0(\widetilde{\theta}^{(J)}) = 0$. △

3. ESSENTIAL AND NONESSENTIAL FACES

In this section, we study optimal motion in the faces and state the main result. We will call face F_J essential if there exists a trajectory through F_J that satisfies the assertion of Lemma 2 with $\widetilde{\theta}^{(J)}$. The face is called nonessential otherwise. Since $\dot{q}_k(t) = 0$ a.e. when $k \in J$, if the face F_J is essential, then

$$0 \in \partial_k H_J(\widetilde{\theta}^{(J)}) \quad \text{for} \quad k \in J, \tag{3.1}$$

where $\partial_k H_J(\widetilde{\theta}^{(J)})$ denotes the set that is obtained by projecting the set $\partial H_J(\widetilde{\theta}^{(J)})$ onto the k-th coordinate axis. Let us note that (cf. (1.4))

$$\partial_k h_k(\theta) = -e^{-\theta_k}\left(1 - \sum_{\ell=1}^{K} p_{k\ell} + \sum_{\ell\neq k} e^{\theta_\ell} p_{k\ell}\right) \tag{3.2}$$

and that

$$\partial_\ell h_k(\theta) = e^{-\theta_k} e^{\theta_\ell} p_{k\ell}, \quad \ell \neq k. \tag{3.3}$$

As a consequence,

$$\partial_l h_k(\theta) > 0 \quad \text{if} \quad \ell \neq k \quad \text{and} \quad \partial_k h_k(\theta) < 0. \tag{3.4}$$

We also note that if $h_k(\theta) = 0$, then $\partial h_k(\theta)^+$ is the set of vectors $\alpha \nabla h_k(\theta)$, where $\alpha \in [0,1]$.

Owing to (1.8) with $\theta = \widetilde{\theta}^{(J)}$, (3.2), (3.3) and the equation $h_\ell(\widetilde{\theta}^{(J)}) = 0$ for $\ell \in J$, if the face F_J is essential, then there exist $\alpha_\ell^{(J)} \in [0,1]$, where $\ell \in J$, such that, for $k \in J$ a.e.

$$\begin{aligned}
0 = \dot{q}_k(t) &= e^{\widetilde{\theta}_k^{(J)}}\lambda_k + \sum_{\ell\notin J} \partial_k h_\ell(\widetilde{\theta}^{(J)})\mu_\ell + \sum_{\ell\in J} \alpha_\ell^{(J)} \partial_k h_\ell(\widetilde{\theta}^{(J)})\mu_\ell \\
&= e^{\widetilde{\theta}_k^{(J)}}\lambda_k + \sum_{\ell\notin J} e^{-\theta_\ell^{(\ell)}} e^{\widetilde{\theta}_k^{(J)}} p_{lk}\mu_\ell + \sum_{\substack{\ell\in J,\\ \ell\neq k}} \alpha_\ell^{(J)} e^{-\widetilde{\theta}_\ell^{(J)}} e^{\widetilde{\theta}_k^{(J)}} p_{lk}\mu_\ell \\
&\qquad - \alpha_k^{(J)} e^{-\widetilde{\theta}_k^{(J)}}\left(1 - \sum_{\ell=1}^{K} p_{k\ell} + \sum_{\ell\neq k} e^{\widetilde{\theta}_\ell^{(J)}} p_{k\ell}\right)\mu_k.
\end{aligned}$$

Since $h_k(\widetilde{\theta}^{(J)}) = 0$, we have that

$$1 - \sum_{\ell=1}^{K} p_{k\ell} + \sum_{\ell\neq k} e^{\widetilde{\theta}_\ell^{(J)}} p_{k\ell} = e^{\widetilde{\theta}_k^{(J)}}(1 - p_{kk}). \tag{3.5}$$

Therefore, we need that, for $k \in J$,

$$\lambda_k + \sum_{\ell\notin J} p_{lk}\nu_\ell + \sum_{\ell\in J} \alpha_\ell^{(J)} e^{-\widetilde{\theta}_\ell^{(J)}} p_{lk}\mu_\ell - \alpha_k^{(J)} e^{-\widetilde{\theta}_k^{(J)}}\mu_k = 0.$$

Since $\lambda_k + \sum_{\ell\notin J} p_{lk}\nu_\ell = \nu_k - \sum_{\ell\in J} p_{lk}\nu_\ell$, equivalently,

$$\alpha_k^{(J)} e^{-\widetilde{\theta}_k^{(J)}}\mu_k - \sum_{\ell\in J} \alpha_\ell^{(J)} e^{-\widetilde{\theta}_\ell^{(J)}} p_{lk}\mu_\ell = \nu_k - \sum_{\ell\in J} p_{lk}\nu_\ell.$$

In vector form,

$$(I - P^T)(J^c \,|\, J^c)(\alpha_\ell^{(J)} e^{-\widetilde{\theta}_\ell^{(J)}}\mu_\ell,\ \ell \in J)^T = (I - P^T)(J^c \,|\, J^c)(\nu_\ell,\ \ell \in J)^T.$$

Both here and below the following piece of notation is used: for matrix W, we denote by $W(U \,|\, V)$ the matrix obtained from W by deleting rows and columns numbered with entries of U and V, respectively.

Since the matrix $(I - P^T)(J^c \,|\, J^c)$ is nondegenerate, if, $\ell \in J$, then

$$\alpha_\ell^{(J)} e^{-\widetilde{\theta}_\ell^{(J)}}\mu_\ell = \nu_\ell. \tag{3.6}$$

Hence, in order for the face F_J to be essential, it is necessary that

$$\widetilde{\rho}_\ell^{(J)} \geq \rho_\ell, \quad \ell \in J, \tag{3.7}$$

where

$$\widetilde{\rho}_\ell^{(J)} = e^{-\widetilde{\theta}_\ell^{(J)}}. \tag{3.8}$$

Since by (2.23)

$$(I-P)(J^c\,|\,J^c)((\widetilde{\rho}_\ell^{(J)})^{-1}-1,\ \ell\in J)^T = P(J^c\,|\,J)(\rho_\ell^{-1}-1,\ \ell\notin J)^T,$$

we have that

$$((\widetilde{\rho}_\ell^{(J)})^{-1}-1,\ \ell\in J)^T = \big((I-P)(J^c\,|\,J^c)\big)^{-1}P(J^c\,|\,J)(\rho_\ell^{-1}-1,\ \ell\notin J)^T. \tag{3.9}$$

Since the matrix $\big((I-P)(J^c\,|\,J^c)\big)^{-1}P(J^c\,|\,J)$ has nonnegative entries, $\widetilde{\rho}_\ell^{(J)} \leq 1$, for $\ell \in J$. Besides, the condition (3.7) of being essential equivalently states that, entrywise,

$$(\rho_\ell^{-1}-1,\ \ell\in J)^T \geq \big((I-P)(J^c\,|\,J^c)\big)^{-1}P(J^c\,|\,J)(\rho_\ell^{-1}-1,\ \ell\notin J)^T. \tag{3.10}$$

If $k \notin J$, then by (1.8) with $\theta = \widetilde{\theta}^{(J)}$, (2.4), (3.2), (3.3) and (3.6), for the trajectory in the face F_J with the adjoint variable $\widetilde{\theta}^{(J)}$,

$$\begin{aligned}
\dot{q}_k(t) &= e^{\theta_k^{(k)}}\lambda_k + \sum_{\substack{\ell\notin J,\\ \ell\neq k}} e^{-\theta_\ell^{(\ell)}}e^{\theta_k^{(k)}}p_{lk}\mu_\ell + \sum_{\ell\in J}\alpha_\ell^{(J)}e^{-\widetilde{\theta}_\ell^{(J)}}e^{\theta_k^{(k)}}p_{lk}\mu_\ell \\
&\quad - e^{-\theta_k^{(k)}}\left(1-\sum_{\ell=1}^K p_{k\ell} + \sum_{\ell\neq k} e^{\widetilde{\theta}_\ell^{(J)}}p_{k\ell}\right)\mu_k \\
&= \rho_k^{-1}\left(\lambda_k + \sum_{\ell=1}^K p_{lk}\nu_\ell\right) - \left(1+\sum_{\ell=1}^K (e^{\widetilde{\theta}_\ell^{(J)}}-1)p_{k\ell}\right)\nu_k \\
&= \mu_k - \left(1+\sum_{\ell\notin J}(\rho_\ell^{-1}-1)p_{k\ell} + \sum_{\ell\in J}((\widetilde{\rho}_\ell^{(J)})^{-1}-1)p_{k\ell}\right)\nu_k.
\end{aligned} \tag{3.11}$$

A similar reasoning (or a formal substitution $J = \varnothing$ in the rightmost side of (3.11)) imply that, for all k,

$$\partial_k H_0(\theta^*) = \mu_k - \left(1+\sum_{\ell=1}^K(\rho_\ell^{-1}-1)p_{k\ell}\right)\nu_k, \tag{3.12}$$

so that, for $k \notin J$,

$$\dot{q}_k(t) = \partial_k H_0(\theta^*) + \sum_{\ell\in J}(\rho_\ell^{-1} - (\widetilde{\rho}_\ell^{(J)})^{-1})p_{k\ell}\nu_k. \tag{3.13}$$

Since by (3.5) and (3.8), for $k \in J$, equation (3.12) can be written as

$$\partial_k H_0(\theta^*) = (\rho_k^{-1} - (\widetilde{\rho}_k^{(J)})^{-1})\nu_k - \sum_{\ell=1}^K p_{k\ell}(\rho_\ell^{-1} - (\widetilde{\rho}_\ell^{(J)})^{-1})\nu_k \tag{3.14}$$

and $\dot{q}_k(t) = 0$ a.e., we have that a.e. for $k \in J$

$$\dot{q}_k(t) = \partial_k H_0(\theta^*) - (\rho_k^{-1} - (\widetilde{\rho}_k^{(J)})^{-1})\nu_k + \sum_{\ell\in J}(\rho_\ell^{-1} - (\widetilde{\rho}_\ell^{(J)})^{-1})p_{k\ell}\nu_k. \tag{3.15}$$

Let us recall the definition of the dual Jackson network. It is a Jackson network in which the service rates at the stations are the same as in the original network. It has the same interstation flow rates but they have reverse directions. Hence, appending pieces of notation used for the quantities associated with the original network with "overbar" in order to denote corresponding entities of the dual network, we have that $\overline{\mu}_k = \mu_k$, $\overline{\nu}_k = \nu_k$ and $\overline{\nu}_k\overline{p}_{k\ell} = \nu_\ell p_{lk}$. The network arrival and departure flows are also interchanged: $\overline{\lambda}_k = \nu_k\Big(1 - \sum\limits_{\ell=1}^K p_{k\ell}\Big)$ and $\lambda_k = \nu_k\Big(1 - \sum\limits_{\ell=1}^K \overline{p}_{k\ell}\Big)$. (The following reasoning shows that the matrix $\overline{P} = (\overline{p}_{k\ell})$ is substochastic. Since $\nu = \lambda + P^T\nu$, we have that $\nu \geq P^T\nu$ entrywise. Therefore, $\nu_k \geq \sum\limits_{\ell=1}^K p_{lk}\nu_\ell$, i.e., $\sum\limits_{\ell=1}^K \overline{p}_{k\ell} = \sum\limits_{\ell=1}^K p_{lk}\nu_\ell/\nu_k \leq 1$.) If a Jackson network is in a steady state, then time reversal results in the stationary dual network.

With the introduced notation, by (3.13) and (3.15), if $q(t) \in F_J$, then a.e.

$$\dot q(t) = \nabla H_0(\theta^*) - (I - \overline{P}^T)\varphi^{(J)},$$

where

$$\varphi_\ell^{(J)} = \begin{cases} (\rho_\ell^{-1} - (\widetilde\rho_\ell^{(J)})^{-1})\nu_\ell & \text{if } \ell \in J,\\ 0 & \text{otherwise.}\end{cases} \tag{3.16}$$

In addition, (3.12) assumes the form

$$-\nabla H_0(\theta^*) = \overline{\lambda} - (I - \overline{P}^T)\mu, \tag{3.17}$$

where $\mu = (\mu_1, \ldots, \mu_K)^T$. (We note that the righthand side can be also written as $\nabla\overline{H}_0(0)$.) Let $T > 0$. Introducing $\overline{q}(t) = q(T - t)$, with $0 \leq t \leq T$, we have that, a.e.,

$$\dot{\overline{q}}(t) = \overline{\lambda} - (I - \overline{P}^T)\mu + (I - \overline{P}^T)\dot\varphi(t), \tag{3.18}$$

where

$$\dot\varphi(t) = \sum_J \mathbf{1}_{F_J}(\overline{q}(t))\varphi^{(J)}. \tag{3.19}$$

Since, by (3.16), $\dot\varphi_k(t) = 0$ provided $\overline{q}_k(t) > 0$, (3.18) implies that $\overline{q}(t)$ is an oblique reflection of the function $q(T) + (\overline{\lambda} - (I - \overline{P}^T)\mu)t$, i.e., it is a fluid trajectory that starts at the terminal point r of the queue lengths in a Jackson network with exogenous arrival rates $\overline{\lambda}_k$, service rates μ_k and routing matrix $\overline{P}$. Let us show the converse: if $\overline{q}$ is the fluid trajectory of the dual network that starts at the terminal point r, then its time reversal is an optimal trajectory. Since $\overline{q}$ solves the oblique reflection problem, we have (3.18), where $\varphi(t)$ is an entrywise nondecreasing absolutely continuous function, such that $\overline{q}_k(t)\dot\varphi_k(t) = 0$ s.ђ. By (3.17), if $\overline{q}(t) \in F_J$, then, for restrictions of vectors to J and restrictions of matrices to $J \times J$, a.e. $\dot\varphi_J(t) = (I_{J,J} - P_{J,J}^T)^{-1}(-\nabla H_0(\theta^*)_J)$. By (3.14) and (3.16), $\dot\varphi_J(t) = \varphi_J^{(J)}$, i.e., (3.19) holds. Thus, the time reversal of $\overline{q}$ passes through F_J with adjoint variable $\widetilde\theta^{(J)}$. It follows that the faces F_J in (3.19) are essential. We obtain also that $\varphi_\ell^{(J)} \geq 0$, for $\ell \in J$, and that (3.7) is not only necessary, but also sufficient in order for F_J to be essential because the fluid trajectory that begins in F_J will stay there for some time, provided (3.7) holds. Consequently, the time reversal of an optimal trajectory with adjoint variables $\widetilde\theta^{(J)}$ that goes from some point A to some point B is the fluid trajectory for getting from B to A for the dual network. The converse is true too.

Since the original Jackson netowrk is assumed to be ergodic, the dual network is ergodic too so that the fluid trajectory of the dual network that starts at r reaches eventually the origin. Furthermore, that trajectory consists of finitely many line segments, which belong to faces F_J, with sets J increasing by inclusion, except for, possibly, the "one–point line segment" that is the initial point (cf., e.g., [13, Lemmas 5.3, 5.4]). Therefore, an optimal trajectory from O to r exists, is unique and consists of finitely many line segments which belong to faces F_J, with the sets J decreasing by inclusion, except for, possibly, the "one–point line segment" that is the terminal point (cf. Fig. 9 below). Let T^* represent the length of time that takes the fluid trajectory $\overline{q}(t)$ that starts at the terminal point r to reach the origin. Let $q^*(t) = \overline{q}(T^* - t)$, where $0 \le t \le T^*$. (We recall that q^* attains the lower cost bound, cf., (2.19).) We have proved the following theorem.

Theorem 3. *The trajectory $q^*(t)$ is a unique optimal trajectory up to the time spent at the origin. Optimal motion in face F_J occurs with adjoint variable $\widetilde{\theta}^{(J)}$. The cost of reaching point r when moving along that trajectory is $\int\limits_0^{T^*} L(q^*(t), \dot{q}^*(t))\,dt = \theta^* \cdot r$.*

As a consequence of Theorem 3, we have that the fluid trajectory of the dual network $\overline{q}$ that starts at r reaches the origin by passing through a sequence of essential faces in each of which it spends a positive length of time. If the initial condition belongs to a nonessential face, then the trajectory leaves that face immediately. In an essential face F_J, motion takes place according to the equations $\dot{\overline{q}}_k(t) = -\partial_k H_0(\theta^*) - \sum\limits_{\ell\in J} (\rho_\ell^{-1} - (\widetilde{\rho}_\ell^{(J)})^{-1}) p_{k\ell}\nu_k$ provided $k \notin J$ and $\dot{\overline{q}}_k(t) = 0$ a.e. provided $k \in J$ (cf. (3.11), (3.13)). It is noteworthy that since $\dot{\overline{q}}(t) = \overline{\lambda} + (\overline{P}^T - I)\overline{d}^{(J)}$ when $\overline{q}(t) \in F_J$, we have by (3.18) and (3.19) that $\overline{d}^{(J)} = \mu - \varphi^{(J)}$, where $\overline{d}^{(J)} = (\overline{d}_1^{(J)}, \ldots, \overline{d}_K^{(J)})^T$ denotes the vector of flow rates through the stations. By (3.16), we have that $\overline{d}_\ell^{(J)} = \mu_\ell$ when $\ell \notin J$ and $\overline{d}_\ell^{(J)} = (\widetilde{\rho}_\ell^{(J)})^{-1}\nu_\ell$ when $\ell \in J$. Since $1 \ge \widetilde{\rho}_\ell^{(J)} \ge \rho_\ell$ provided face F_J is essential, we have that $\nu_\ell \le \overline{d}_\ell^{(J)} \le \mu_\ell$. It follows, by (1.1) and (1.2), that for the dual network $\overline{L}_J(\overline{\lambda} + (\overline{P}^T - I)\overline{d}^{(J)}) = 0$. Consequently,

$$\overline{L}(\overline{q}(t), \dot{\overline{q}}(t)) = 0 \quad \text{a.e.} \tag{3.20}$$

The condition for being essential can be made more explicit in the case of a semiaxis. If the face F_J is the semiaxis x_m, then (3.9) takes the form

$$(\widetilde{\rho}_\ell^{(J)})^{-1} - 1 = a_{m\ell}(\rho_m^{-1} - 1), \quad \ell \ne m, \tag{3.21}$$

whereas (3.10) (accounting for (2.2)) takes the form

$$\rho_\ell^{-1} - 1 \ge a_{m\ell}(\rho_m^{-1} - 1), \quad \ell \ne m. \tag{3.22}$$

(Naturally, (3.22) can be obtained directly from (3.6).) Substitution of (3.21) into (3.11) and the fact that $\sum\limits_{\ell=1}^{m} a_{m\ell}p_{m\ell} = 1 - 1/c_{mm}$ (as shown in the proof of Theorem 2) imply that if the semiaxis x_m is essential, then

$$\dot{q}_m^*(t) = \frac{1 - \rho_m}{c_{mm}}\mu_m > 0,$$

as to be expected. The next property of the matrix C is proved in the Appendix.

Lemma 7. *We have that $c_{m\ell} \le c_{mm}$.*

It follows from Lemma 7, (2.17) and (3.22) that if ρ_m is maximal then the semiaxis x_m is essential, i.e., if an optimal trajectory reaches it, it will stay on the semiaxis a positive length of time.

Remark 1. Theorem 2 implies that $\theta^* > 0$ entrywise so that the lower bound in (2.19) and (2.20) is positive for all r if and only if the network in question is ergodic. Therefore, one shouldn't expect the adjoint variables $\widetilde{\theta}^{(J)}$ to be optimal for going from 0 to r in a nonergodic network. On the other hand, the observation that the time reversal of Hamilton equations results in fluid equations is still valid and bounds (2.19) and (2.20) may be useful. For instance, if $\theta^* < 0$ entrywise, then the fluid trajectories of the dual network run off to infinity and the cost for the original network to go from r to the origin is $-\theta^* \cdot r$.

4. CONDITIONAL LAWS OF LARGE NUMBERS. DISCUSSION.

As mentioned earlier, an LDP enables one to derive logarithmic asymptotics of the probabilities of rare events and identify the most likely scenarios for such events to occur. The next two lemmas provide an illustration. We recall that the network is assumed to be ergodic.

Lemma 8. *For arbitrary* $\delta > 0$,

$$\lim_{\varepsilon\to 0}\limsup_{n\to\infty}\mathbf{P}\Big(\sup_{t\le T^*}\Big|\frac{Q(nt)}{n}-q^*(t)\Big|\ge\delta \mid \Big|\frac{Q(nT^*)}{n}-r\Big|\le\varepsilon\Big)=0.$$

In addition,

$$\lim_{\varepsilon\to 0}\limsup_{n\to\infty}\frac{1}{n}\ln\mathbf{P}\Big(\Big|\frac{Q(nT^*)}{n}-r\Big|\le\varepsilon\Big)=\lim_{\varepsilon\to 0}\liminf_{n\to\infty}\frac{1}{n}\ln\mathbf{P}\Big(\Big|\frac{Q(nT^*)}{n}-r\Big|\le\varepsilon\Big)=-\theta^*\cdot r.$$

Proof. In order to prove the first part it is sufficient to prove that

$$\limsup_{\varepsilon\to 0}\limsup_{n\to\infty}\frac{1}{n}\ln\mathbf{P}\Big(\sup_{t\le T^*}\Big|\frac{Q(nt)}{n}-q^*(t)\Big|\ge\delta \mid \Big|\frac{Q(nT^*)}{n}-r\Big|\le\varepsilon\Big)<0. \tag{4.1}$$

By the LDP,

$$\liminf_{n\to\infty}\frac{1}{n}\ln\mathbf{P}\Big(\Big|\frac{Q(nT^*)}{n}-r\Big|\le\varepsilon\Big)\ge-\inf_{q:\,|q(T^*)-r|\le\varepsilon/2}\boldsymbol{I}(q). \tag{4.2}$$

We extend q^* to time epochs in (T^*,∞) according to the law of large numbers. In analogy with (3.20), $\int\limits_{T^*}^{\infty}L(q^*(t),\dot q^*(t))\,dt=0$. Hence $\boldsymbol{I}(q^*)=\int\limits_0^{T^*}L(q^*(t),\dot q^*(t))\,dt$. Since $\boldsymbol{I}(q)$ is lower compact and $q^*(T^*)=r$,

$$\lim_{\varepsilon\to 0}\inf_{q:\,|q(T^*)-r|\le\varepsilon/2}\boldsymbol{I}(q)=\inf_{q:\,q(T^*)=r}\boldsymbol{I}(q)\le\boldsymbol{I}(q^*)=\int\limits_0^{T^*}L(q^*(t),\dot q^*(t))\,dt. \tag{4.3}$$

Similarly,

$$\limsup_{\varepsilon\to 0}\limsup_{n\to\infty}\frac{1}{n}\ln\mathbf{P}\Big(\sup_{t\le T^*}\Big|\frac{Q(nt)}{n}-q^*(t)\Big|\ge\delta,\ \Big|\frac{Q(nT^*)}{n}-r\Big|\le\varepsilon\Big)\le-\inf_{\substack{q:\ \sup_{t\le T^*}|q(t)-q^*(t)|\ge\delta,\\ q(T^*)=r}}\boldsymbol{I}(q).$$

Let the latter infimum be attained at $\widetilde q$. Since $\inf\limits_{q:\,q(T)=r}\int\limits_0^T L(q(t),\dot q(t))\,dt$ is attained at the unique trajectory q^* and since $\widetilde q\ne q^*$, we have that $\boldsymbol{I}(\widetilde q)\ge\int\limits_0^{T^*}L(\widetilde q(t),\dot{\widetilde q}(t))\,dt>\int\limits_0^{T^*}L(q^*(t),\dot q^*(t))\,dt$. Therefore,

$$\limsup_{\varepsilon\to 0}\limsup_{n\to\infty}\frac{1}{n}\ln\mathbf{P}\Big(\sup_{t\le T^*}\Big|\frac{Q(nt)}{n}-q^*(t)\Big|\ge\delta,\ \Big|\frac{Q(nT^*)}{n}-r\Big|\le\varepsilon\Big)<-\int\limits_0^{T^*}L(q^*(t),\dot q^*(t))\,dt.$$

Recalling (4.2) and (4.3) obtains (4.1). The second part of the lemma is essentially implied by (4.2). △

As another application, let us look at the length of time it takes the total queue length to assume a large value: $\tau_n=\inf\Big\{t:\ \sum\limits_{k=1}^K Q_k(nt)\ge nA\Big\}$. We define r as the point at which $\theta^*\cdot x$ attains minimum over all $x=(x_1,\dots,x_K)^T\in\mathbb{R}_+^K$ such that $\sum\limits_{k=1}^K x_k=A$. "In a general position", r is specified uniquely.

Lemma 9. *For all T great enough,*

$$\lim_{n\to\infty}\frac{1}{n}\ln \mathbf{P}(\tau_n \le T) = -\theta^*\cdot r \tag{4.4}$$

and

$$\lim_{n\to\infty}\mathbf{P}\Big(\sup_{(\tau_n-T^*)^+\le t\le\tau_n}\Big|\frac{Q(nt)}{n}-q^*(t-(\tau_n-T^*)^+)\Big|<\delta|\tau_n\le T\Big)=1\,. \tag{4.5}$$

Proof. The argument is similar to the one used in the proof of Lemma 8. Let us show that, for all T great enough,

$$\inf_{q\colon \sup\limits_{t\le T}\sum\limits_{k=1}^K q_k(t)>A}\boldsymbol{I}(q)=\inf_{q\colon \sup\limits_{t\le T}\sum\limits_{k=1}^K q_k(t)\ge A}\boldsymbol{I}(q)=\theta^*\cdot r. \tag{4.6}$$

Given $\varepsilon>0$, let $q^{\varepsilon,*}$ represent the optimal trajectory to get from the origin to $r(1+\varepsilon)$, as defined in the proof of Theorem 3. We assume it has been extended past the time it reaches $r(1+\varepsilon)$ by the law of large numbers. Since before $q^{\varepsilon,*}$ reaches $r(1+\varepsilon)$, that trajectory is a time reversed fluid trajectory of the dual network, there exists $\widehat T>0$ such that all trajectories $q^{\varepsilon,*}$ that correspond to $\varepsilon\in[0,1]$ reach the destination by time $\widehat T$ (cf., e.g., [13, Lemma 5.4, p. 143]). For $T\ge\widehat T$,

$$\begin{aligned}\theta^*\cdot r=\boldsymbol{I}(q^*)&=\inf_{q\colon \sum\limits_{k=1}^K q_k(t)\ge A\text{ for some }t>0}\boldsymbol{I}(q)\le\inf_{q\colon \sup\limits_{t\le T}\sum\limits_{k=1}^K q_k(t)\ge A}\boldsymbol{I}(q)\\&\le\inf_{q\colon \sup\limits_{t\le T}\sum\limits_{k=1}^K q_k(t)>A}\boldsymbol{I}(q)\le\inf_{q\colon \sup\limits_{t\le T}\sum\limits_{k=1}^K q_k(t)\ge A(1+\varepsilon)}\boldsymbol{I}(q)\le\boldsymbol{I}(q^{\varepsilon,*})=\theta^*\cdot r(1+\varepsilon).\end{aligned}$$

Letting ε go to zero obtains (4.6). Since by the LDP,

$$\begin{aligned}\limsup_{n\to\infty}\frac{1}{n}\ln\mathbf{P}(\tau_n\le T)&=\limsup_{n\to\infty}\frac{1}{n}\ln\mathbf{P}\bigg(\sup_{t\le T}\frac{Q(nt)}{n}\ge A\bigg)\le-\inf_{q\colon \sup\limits_{t\le T}\sum\limits_{k=1}^K q_k(t)\ge A}\boldsymbol{I}(q),\\ \text{and }\liminf_{n\to\infty}\frac{1}{n}\ln\mathbf{P}(\tau_n\le T)&=\liminf_{n\to\infty}\frac{1}{n}\ln\mathbf{P}\bigg(\sup_{t\le T}\frac{Q(nt)}{n}\ge A\bigg)\ge-\inf_{q\colon \sup\limits_{t\le T}\sum\limits_{k=1}^K q_k(t)>A}\boldsymbol{I}(q),\end{aligned}$$

we have that (4.4) is implied by (4.6).

In order to prove (4.5), let us introduce $\tau(q)=\inf\{t:\ \sum_{k=1}^K q_k(t)\ge A\}$, where $q=(q(t),\ t\ge 0)\in\mathbb{D}(\mathbb{R}_+,\mathbb{R}_+^K)$. Since the set

$$\bigg\{q\in\mathbb{D}(\mathbb{R}_+,\mathbb{R}_+^K):\sup_{(\tau(q)-T^*)^+\le t\le\tau(q)}\big|q(t)-q^*(t-(\tau(q)-T^*)^+)\big|\ge\delta,\ \sup_{t\le T}\sum_{k=1}^K q_k(t)\ge A\bigg\}$$

contains all its limit points that are continuous functions and since $\boldsymbol{I}(q)=\infty$ provided q is discontinuous,

$$\begin{aligned}&\limsup_{n\to\infty}\frac{1}{n}\ln\mathbf{P}\bigg(\sup_{(\tau_n-T^*)^+\le t\le\tau_n}\Big|\frac{Q(nt)}{n}-q^*(t-(\tau_n-T^*)^+)\Big|\ge\delta,\ \tau_n\le T\bigg)\\&\qquad\le-\inf_{\substack{q\colon \sup\limits_{(\tau(q)-T^*)^+\le t\le\tau(q)}|q(t)-q^*(t-(\tau(q)-T^*)^+)|\ge\delta,\\ \sum\limits_{k=1}^K q_k(t)\ge A\text{ for some }t\in[0,T]}}\boldsymbol{I}(q).\end{aligned}$$

Since the set over which the infimum is taken is closed, the infimum is attained by some function $\breve{q}$. We know that if $\inf\limits_{(q,T'):\,\sum\limits_{k=1}^{K} q_k(T')\ge A} \int\limits_0^{T'} L(q(t),\dot{q}(t))\,dt$ is attained at (q,T') then $T'\ge T^*$ and $q(T'-t)=q^*(T^*-t)$ when $0\le t\le T^*$ (hence, $\tau(q)=T'$). Therefore, if $\tau(\breve{q})<T^*$, then $(\breve{q},\tau(\breve{q}))$ does not attain the latter infimum. Suppose that $\tau(\breve{q})\ge T^*$. Then there exists $t\in[\tau(\breve{q})-T^*,\tau(\breve{q})]$ such that $\breve{q}(t)-q^*(t-\tau(\breve{q})+T^*)\ge\delta$, so, $(\breve{q},\tau(\breve{q}))$ does not attain the infimum either. Thus, $\boldsymbol{I}(\breve{q})\ge\int\limits_0^{\tau(\breve{q})} L(\breve{q}(t),\dot{\breve{q}}(t))\,dt>\int\limits_0^{T^*} L(q^*(t),\dot{q}^*(t))\,dt=\theta^*\cdot r$. Therefore

$$\limsup_{n\to\infty}\frac{1}{n}\ln\mathbf{P}\biggl(\sup_{(\tau_n-T^*)^+\le t\le\tau_n}\biggl|\frac{Q(nt)}{n}-q^*(t-(\tau_n-T^*)^+)\biggr|\ge\delta,\ \tau_n\le T\biggr)<-\theta^*\cdot r.$$

On recalling (4.4), we have that, for all T great enough,

$$\limsup_{n\to\infty}\frac{1}{n}\ln\mathbf{P}\biggl(\sup_{(\tau_n-T^*)^+\le t\le\tau_n}\biggl|\frac{Q(nt)}{n}-q^*(t-(\tau_n-T^*)^+)\biggr|\ge\delta\mid\tau_n\le T\biggr)<0,$$

which implies (4.5). △

There is an insightful explanation as to why q^* should be the time reversal of $\overline{q}$. Following [14] (see also [15]), we note that

$$\begin{aligned}\mathbf{P}\biggl(\sup_{t\le T^*}\biggl|\frac{Q(nt)}{n}-q^*(t)\biggr|\ge\delta\mid\biggl|\frac{Q(nT^*)}{n}-r\biggr|\le\varepsilon\biggr)\\ =\mathbf{P}\biggl(\sup_{t\le T^*}\biggl|\frac{Q(n(T^*-t))}{n}-q^*(T^*-t)\biggr|\ge\delta\mid\biggl|\frac{Q(nT^*)}{n}-r\biggr|\le\varepsilon\biggr)\\ =\mathbf{P}\biggl(\sup_{t\le T^*}\biggl|\frac{\overline{Q}^n(nt)}{n}-\overline{q}(t)\biggr|\ge\delta\mid\biggl|\frac{\overline{Q}^n(0)}{n}-r\biggr|\le\varepsilon\biggr),\end{aligned}$$

where $\overline{Q}^n(t)=Q(nT^*-t)$. If the process $Q(t)$ is stationary, then the process $\overline{Q}^n(t)$ is stationary too and is the process of queue lengths in the time reversed Jackson network. As a consequence, its distribution does not depend on n. Since $\overline{q}(t)$ is the fluid limit of the queue length process in an ergodic Jackson network, it is routine to show that the latter probability tends to zero as $n\to\infty$ and $\varepsilon\to 0$. Similar arguments can be found in [16] and [9], stationarity also playing a key role. (We note that none of those contributions are available in refereed journals.)

Since $\theta^*\cdot r=\inf\limits_{\substack{q:\,q(0)=0,\,q(T)=r\\ \text{for some } T}}\boldsymbol{I}(q)$, the general theory (cf. [1,2]) suggests, if one recalls the second part of Lemma 8, that the stationary distribution of $Q(t)$ has the following asymptotic behaviour:

$$\lim_{\varepsilon\to 0}\limsup_{n\to\infty}\frac{1}{n}\ln\mathbf{P}\biggl(\biggl|\frac{Q(nt)}{n}-r\biggr|<\varepsilon\biggr)=\lim_{\varepsilon\to 0}\liminf_{n\to\infty}\frac{1}{n}\ln\mathbf{P}\biggl(\biggl|\frac{Q(nt)}{n}-r\biggr|<\varepsilon\biggr)=-\theta^*\cdot r.$$

For the setup in question, this property is obvious in that the stationary distribution of $Q(t)$ is known explicitly: $\mathbf{P}(Q(t)=(i_1,\ldots,i_K))=\prod\limits_{k=1}^{K}(1-\rho_k)\rho_k^{i_k}$ (cf., e.g., [3]).

5. AN EXAMPLE: AN ERGODIC TWO-STATION NETWORK

In this section, in order to illustrate the general results, a two–station network is looked at. The analysis is more geometric then. We have that

$$h_1(\theta)=e^{-\theta_1}\bigl((e^{\theta_1}-1)p_{11}+(e^{\theta_2}-1)p_{12}+1\bigr)-1, \tag{5.1}$$

Fig. 1. A case where $\partial_2 H_0(\theta^{(1)}) \le 0$

Fig. 2. A case where $\partial_2 H_0(\theta^{(1)}) > 0$

$$h_2(\theta) = e^{-\theta_2}((e^{\theta_1}-1)p_{21} + (e^{\theta_2}-1)p_{22} + 1) - 1, \tag{5.2}$$

$$\begin{aligned} H_1(\theta) &= (e^{\theta_1}-1)\lambda_1 + (e^{\theta_2}-1)\lambda_2 + h_1(\theta)^+\mu_1 + h_2(\theta)\mu_2,\\ H_2(\theta) &= (e^{\theta_1}-1)\lambda_1 + (e^{\theta_2}-1)\lambda_2 + h_1(\theta)\mu_1 + h_2(\theta)^+\mu_2,\\ H_0(\theta) &= (e^{\theta_1}-1)\lambda_1 + (e^{\theta_2}-1)\lambda_2 + h_1(\theta)\mu_1 + h_2(\theta)\mu_2. \end{aligned} \tag{5.3}$$

We note that $h_1(\theta) = 0$ if and and only if $(e^{\theta_1}-1)/(e^{\theta_2}-1) = p_{12}/(1-p_{11})$, i.e., θ_1 and θ_2 are of the same sign so that the graph of the equation $h_1(\theta) = 0$ is located in quadrants I and III. In addition, $h_1(\theta) > 0$ above and to the left of the graph and $h_1(\theta) < 0$ below and to the right of the graph. Expressing θ_2 as a function of θ_1 in quadrant I via the equation $h_1(\theta) = 0$, we have, on differentiating twice, that $d^2\theta_2/d\theta_1^2 = (1-p_{11})/p_{12}e^{\theta_1-\theta_2} - ((1-p_{11})/p_{12}e^{\theta_1-\theta_2})^2$. Since $(1-p_{11})/p_{12}e^{\theta_1-\theta_2} = 1 + (1-p_{11}-p_{12})/p_{12}e^{-\theta_2} \ge 1$, it follows that θ_2 is a strictly concave function of θ_1 provided $p_{11}+p_{12} < 1$ and is a linear function of θ_1 provided $p_{11}+p_{12} = 1$. Similarly, the graph of $h_2(\theta) = 0$ is located in quadrants I and III, $h_2(\theta) < 0$ above and to the left of the graph, $h_2(\theta) > 0$ below and to the right of the graph, θ_2 is a strictly convex function of θ_1 on the graph $h_2(\theta) = 0$ provided $p_{22}+p_{21} < 1$ and is a linear function of θ_1 provided $p_{22}+p_{21} = 1$. Since $(1-p_{11})/p_{12} \ge p_{21}/(1-p_{22})$, in quadrant I the graph of $h_1(\theta) = 0$ is above the graph of $h_2(\theta) = 0$, being strictly above provided $(1-p_{11})/p_{12} > p_{21}/(1-p_{22})$. In quadrant III the graph of $h_1(\theta) = 0$ is below the graph of $h_2(\theta) = 0$, being strictly below provided $(1-p_{11})/p_{12} > p_{21}/(1-p_{22})$.

Let us determine optimal motion from $(0,0)$ to $(1,0)$ in a straight line. The corresponding Hamiltonian is $H_2(\theta)$ and the state space constraint is $q_2(t) = 0$. By Lemma 1, with $g(q) = q_2^2$ so that $\gamma(t) = 0$, $(\dot q_1(t), 0) = (\dot q_1(t), \dot q_2(t)) \in \partial H_2(\theta)$. Hence, $0 \in \partial_2 H_2(\theta)$ and $\dot q_1(t) \in \partial_1 H_2(\theta)$ (cf. [11, Proposition 2.3.15]). In addition, $H_2(\theta) = 0$. The cost is $1\cdot\theta_1 + 0\cdot\theta_2 = \theta_1$. By (5.1)–(5.3), in order that $\partial_2 H_2(\theta) \ni 0$, it is required that $h_2(\theta) \ge 0$ (because $\partial_2 h_1(\theta) > 0$; cf., e.g., (3.4)).

Let us also note that if $h_2(\theta) = 0$, then

$$\partial H_2(\theta) = \left\{ \begin{pmatrix} e^{\theta_1}\lambda_1 + \partial_1 h_1(\theta)\mu_1 + \alpha\partial_1 h_2(\theta)\mu_2 \\ e^{\theta_2}\lambda_2 + \partial_2 h_1(\theta)\mu_1 + \alpha\partial_2 h_2(\theta)\mu_2 \end{pmatrix},\ \alpha \in [0,1] \right\}.$$

Thus, if $\widehat\theta$ is such that $H_0(\widehat\theta) = 0$, $h_2(\widehat\theta) = 0$, i.e., $\widehat\theta = \theta^{(1)}$, and $\partial_2 H_0(\theta^{(1)}) \le 0$, then $\widehat\theta$ is optimal. If $h_2(\widehat\theta) > 0$, $H_0(\widehat\theta) = 0$ and $\partial_2 H_0(\widehat\theta) = 0$, then $\widehat\theta$ is also optimal. One can see that these two situations are mutually exclusive. They are illustrated in Fig. 1 and Fig. 2, respectively. (We also show vector $\nabla H_0(0)$ because that vector being in quadrant III implies that the network is ergodic.)

Let us analyse optimal trajectories for getting from $(0,0)$ to $r = (r_1, r_2)$. Let us compare the route in a straight line to r and a route that first goes horizontally and then goes straight to r. Let us show that when initially motion occurs horizontally along some vector $s = (s_1, 0)$ with adjoint variable $\widehat\theta = (\widehat\theta_1, \widehat\theta_2)$ such that $\partial_2 H_0(\widehat\theta) = 0$ and $h_2(\widehat\theta) > 0$, as in Fig. 2, then going straight to r does not incur a greater cost. Fig. 3 illustrates our reasoning. The cost of moving along s, at first, and then folowing vector $p = r - s$ equals $\widehat\theta_1 s_1 + \theta(p)\cdot p$, where $\theta(p)$ denotes the point on the curve $H_0(\theta) = 0$ such that the outward normal to the curve at that point is collinear with p. Let us also denote by $\widetilde\theta$ the point on the curve $H_0(\theta) = 0$ whose outward normal is collinear with r. If r does not belong to the horizontal coordinate axis, then $\widetilde\theta \ne \widehat\theta$, because the motion according to $\widehat\theta$ is horizontal. The cost of moving straight along r equals $\widetilde\theta\cdot r = \widetilde\theta\cdot s + \widetilde\theta\cdot p$. As s is orthogonal to

Fig. 3. Going straight to r being advantageous

Fig. 4. Interior motion to the right and upwards

the curve $H_0(\theta)=0$ at $\widehat{\theta}$ and the function $H_0(\theta)$ is strictly convex, so, $(\widetilde{\theta}-\widehat{\theta})\cdot s\le 0$, the inequality being strict provided r does not belong to the horizontal coordinate axis and $s\ne 0$. Analogously, since p is orthogonal to the curve $H_0(\theta)=0$ at $\theta(p)$, we have that $(\widetilde{\theta}-\theta(p))\cdot p\le 0$. Consequently, $\widetilde{\theta}\cdot r\le \theta\cdot s+\theta(p)\cdot p$, the inequality being strict provided r does not belong to the horizontal coordinate axis, i.e., the cost of moving along a nonhorizontal vector r is less than the cost incurred when moving along s initially.

As above, let $\theta^{(1)}$ represent the intersection point of the curves $H_0(\theta)=0$ and $h_2(\theta)=0$, and let $\theta^{(2)}$ represent the intersection point of the curves $H_0(\theta)=0$ and $h_1(\theta)=0$. It has been proved in Theorem 2 that the straight lines $\theta_1=\theta_1^{(1)}$ and $\theta_2=\theta_2^{(2)}$ intersect at point θ^*, which belongs to the curve $H_0(\theta)=0$. The above analysis shows that if $\partial_2H_0(\theta^{(1)})>0$, then moving initiallhy horizontally is not optimal. Furthermore, as we will see, going vertically initially and then crossing the interior of the quadrant in order to land at the destination on the horizontal coordinate axis costs less even if one has to be displaced horizontally from the origin. Similarly, a vertical initial motion is not optimal when $\partial_1H_0(\theta^{(2)})>0$.

Let us suppose that $\partial_2H_0(\theta^{(1)})\le 0$ and $\partial_1H_0(\theta^{(2)})\le 0$ (see Fig. 4). In this case $\partial_1H_0(\theta^{(1)})>0$ and $\partial_2H_0(\theta^{(2)})>0$. Then it is optimal to move horizontally with the adjoint variable $\theta^{(1)}$ and it is optimal to move vertically with $\theta^{(2)}$. The cost of moving horizontally along vector $s=(s_1,0)$ first and moving afterwards along vector p is $\theta_1^{(1)}s_1+\theta(p)\cdot p$. Therefore, as in Fig. 3, if $\widetilde{\theta}_1\le\theta_1^{(1)}$, which is the case if the slope of vector r is greater than the slope of the normal to the curve $H_0(\theta)=0$ at θ^*, accordingly, if $\widetilde{\theta}$ is to the left of θ^*, then motion along r is preferable. Let us suppose that the slope of vector r is less than the slope of the normal to the curve $H_0(\theta)=0$ at θ^*, i.e., $\widetilde{\theta}$ is to the right of θ^*. We look for an optimal horizontal displacement s. We need to minimise $\theta_1^{(1)}s_1+\theta(p)\cdot p$ over $\theta(p)$ assuming that $r=p+s$ and p is proportional to $\nabla H_0(\theta(p))$. Denoting $\theta(p)$ by θ and denoting p by $p(\theta)$, we have that $\theta_1^{(1)}s_1+\theta\cdot p(\theta)=\theta_1^{(1)}(r_1-p_1(\theta))+\theta\cdot p(\theta)=\theta_1^{(1)}r_1-\theta_1^{(1)}p_1(\theta)+\theta\cdot p(\theta)$, where $\varkappa\nabla H_0(\theta)=p(\theta)$, for some $\varkappa>0$. Since $p_2(\theta)=r_2$, $\varkappa\partial_2H_0(\theta)=r_2$. Let us minimise $-\theta_1^{(1)}p_1(\theta)+\theta\cdot p(\theta)$ under the constraints that $H_0(\theta)=0$, $\varkappa\nabla H_0(\theta)=p(\theta)$ and $\varkappa\partial_2H_0(\theta)=r_2$. We have that $-\theta_1^{(1)}p_1(\theta)+\theta\cdot p(\theta)=(\partial_1H_0(\theta)/\partial_2H_0(\theta)(\theta_1-\theta_1^{(1)})+\theta_2)r_2$. Let us show that $\partial_1H_0(\theta)/\partial_2H_0(\theta)(\theta_1-\theta_1^{(1)})+\theta_2$ is decreasing along the arc $[\widetilde{\theta},\theta^*]$. Suppose that θ is located between $\widetilde{\theta}$ and θ^*. Fig. 5 illustrates the argument. Point P corresponds to the adjoint variable θ being minimised over. Let α denote the angle between the normal to $H_0(\theta)=0$ at P and the horizontal through θ. We have that $\partial_1H_0(\theta)/\partial_2H_0(\theta)=\cot\alpha$. By geometry, the angle QCP between the tangent to $H_0(\theta)=0$ at θ and the vertical through $\theta_1^{(1)}$ is α as well. Therefore the line segment CQ that connects $(\theta_1^{(1)},\theta_2)$ and the intersection point of the tangent and the vertical through $\theta_1^{(1)}$ is $(\theta_1-\theta_1^{(1)})\cot\alpha$. Hence, $\partial_1H_0(\theta)/\partial_2H_0(\theta)(\theta_1-\theta_1^{(1)})+\theta_2$ is the length of the vertical line segment CR that goes from the intersection point of the tangent and the vertical through $\theta_1^{(1)}$ to the horizontal coordinate axis. As θ moves counterclockwise along the curve $H_0(\theta)=0$ from $\widetilde{\theta}$, the length of that line segment gets smaller and attains a minimum when θ gets to θ^*. If the motion is continued past θ^*, then the line segment starts to grow longer again. Therefore, if one moves horizontally initially, then it is optimal to move with $\theta^{(1)}$ until the slope of the straight line segment that connects the moving point and point r is the same as the slope of the outward normal at θ^*. Then, one has to move in a straight line with that normal's slope. Since motion along r is a special

Fig. 5. Optimal motion along the normal at θ^*

Fig. 6. Optimal trajectories for the parameters of Fig.4 and various destinations

case of horizontal initial motion for $s_1 = 0$, moving optimally horizontally initially is advantageous to moving along r from the outset. Finally, if one moves vertically at the beginning and moves afterwards toward r, then, since $\widetilde{\theta}_2 < \theta_2^*$, one has a similar situation to the case above where $\widetilde{\theta}_1 < \theta_1^*$, i.e., it is advantageous to move along a straight line from the outset. Thus, it is optimal to move along the horizontal coordinate axis until the straight line that connects the moving point and r is of the same slope as the normal to $H_0(\theta) = 0$ at θ^*. Then, it is optimal to move along that normal. We have that $s_1 = r_1 - p_1 = r_1 - \partial_1 H_0(\theta^*)/\partial_2 H_0(\theta^*) r_2 = r_2(r_1/r_2 - \partial_1 H_0(\theta^*)/\partial_2 H_0(\theta^*))$. The cost is $\theta_1^{(1)} s_1 + \theta^* \cdot (r - s) = \theta^* \cdot r$. Similarly, if the slope of r is greater than the slope of the normal at θ^*, then one has to move vertically initially until the slope to r is the same as the slope of the normal and then move along the normal at θ^*. The cost is $\theta^* \cdot r$. Those scenarios are illustrated in Fig. 6.

If $\partial_2 H_0(\theta^{(1)}) \le 0$ and $\partial_1 H_0(\theta^{(2)}) > 0$, then θ^* is to the left of $\theta^{(2)}$ (see Fig. 7). Since the optimal vertical motion corresponds to the apex of the curve $H_0(\theta) = 0$, it will be inferior to moving straight to r. A similar geometric argument to the one above shows that one has to move horizontally until the slope to r gets equal to the slope of the outward normal at θ^*, after which one has to "return" along that normal. If $\partial_2 H_0(\theta^{(1)}) > 0$ and $\partial_1 H_0(\theta^{(2)}) \le 0$ (see Fig. 8), then one needs to move vertically and then "return". Fig. 9 illustrates optimal trajectories for the parameters of Fig. 8. The horizontal coordinate axis is nonessential then so that the optimal trajectory to a point on that axis passes through the vertical coordinate axis. Anyway, one has to move along one of the coordinate axes in such a way that the vector that connects the moving point to the destination becomes collinear to the normal at θ^*. Then, one has to move straight to the destination. If the network is ergodic then one of those combinations of signs of $\partial_2 H_0(\theta^{(1)})$ and $\partial_1 H_0(\theta^{(2)})$ occurs. The cost is always $\theta^* \cdot r$. As a consequence, if $\partial_2 H_0(\theta^{(1)}) > 0$ (respectively, $\partial_1 H_0(\theta^{(2)}) > 0$), then moving initially horizontally (respectively, vertically) is not a part of the optimal trajectory, i.e., the horizontal coordinate axis (respectively, the vertical coordinate axis) is nonessential. Besides, the optimal route can be found by starting at r and first moving along the antigradient of H_0 at θ^* until one of the coordinate axes is reached after which one moves along that axis to the origin. That axis is necessarily essential. Fig. 10 and 11 illustrate optimal trajectories inside quadrant I parallel to one of the coordinate axes. (Although vector $\nabla H_0(0)$ in Fig. 10 does not lie in quadrant III, one can see that the network is still ergodic.)

APPENDIX

Proof of Lemma 1. If the minimum of $\int_0^T L(x(t), \dot{x}(t))\, dt$ over $T \ge 0$ and over absolutely continuous functions $x(t)$ such that $x(0) = x_0$, $(T, x(T)) \in S'$ and $g(x(t)) \le 0$ for all $t \le T$, where S' is a convex closed subset of $\mathbb{R} \times \mathbb{R}^k$, is attained for time T^* and function $x^*(t)$, then there exist measure μ on $[0, T^*]$, a μ-measurable function $\gamma(t)$ and an absolutely continuous function $p(t)$ such that

1. $\gamma(t) \in \partial^> g(x^*(t))$ for μ-almost all $t \in [0, T^*]$ and the support of the measure μ is contained in the set $\{t \in [0, T^*] : \partial^> g(x^*(t)) \ne \varnothing\}$;
2. $\begin{bmatrix} -\dot{p}(t) \\ \dot{x}^*(t) \end{bmatrix} \in \partial H\left(x^*(t), p(t) + \int_0^t \gamma(s)\, \mu(ds) \right)$ a.e. on $[0, T^*]$;

Fig. 7. Motion inside the quadrant to the left and upwards

Fig. 8. Motion inside the quadrant to the right and downwards

3. There exists constant h such that

$$H\Big(x^*(t), p(t) + \int_0^t \gamma(s)\,\mu(ds)\Big) = h \quad \text{on } [0, T^*];$$

4. $$\begin{bmatrix} -h \\ p(T^*) + \int_0^{T^*} \gamma(t)\,\mu(dt) \end{bmatrix} \in -N_{S'}(T^*, x^*(T^*)).$$

In some more detail, if the set S' is of the form $\{T'\} \times S''$, for some $T' > 0$ and convex closed set $S'' \subset \mathbb{R}^k$, then parts 1 and 2 are implied by [17, Theorem 10.2.1 and the discussion, p. 362–364], as well as by "Hamiltonian dualisation" (cf., [17, Theorem 7.6.5, p. 266]). For general S' one can apply the device of reducing a problem with nonfixed time to a problem with fixed time, as in [11, Proof of Corollary 3.6.1, p. 142] (see also [17, Theorem 8.2.1, p. 290]). It is also possible to apply [11, Corollary 3.6.1, p. 142], with the function f being the additional variable so that ζ is the $(k+2)$-vector, whose last component equals unity and the other components equal zero, implying that λ in [11, Condition 4), p. 142] does not affect the values of the first $k+1$ entries on the lefthand side. One should also use the assertion of [11, Proposition 2.4.4, p. 55], according to which $\partial d_{S'}(T^*, x^*(T^*))$ can be replaced with $N_{S'}(T^*, x^*(T^*))$. In the hypotheses of Lemma 1, $S' = \mathbb{R}_+ \times S$. Therefore the first components of all vectors from $N_{S'}(T^*, x^*(T^*))$ equal zero, e.g., by [11, Theorem 2.5.6, p. 67]. By the transversality condition 4, $h = 0$. The fact that the function H from (1.3) can be used as the function H in [11, Corollary 3.6.1, p. 142], is derived by a similar reasoning as in the proof of Theorem 4.2.2 in [11, б. 156], where the property of the function H being strictly Lipschitz is replaced with the condition of the function L being locally Lipschitz. A similar result to Corollary 3.6.1 in [11, б. 142] is found also in [17, Theoorem 10.5.1, p. 383]. △

Proof of (1.5). Let us note that $\pi(u) = \sup_{\theta\in\mathbb{R}}(u\theta - (e^\theta - 1))$, where $u \ge 0$. Consequently, applying the minimax theorem to (1.1) obtains

$$\begin{aligned}
L_J(y) &= \inf_{\substack{(a,d,\varrho)\in\mathbb{R}_+^K\times\mathbb{R}_+^K\times\mathbb{S}_+^{K\times K}:\\ y=a+(\varrho^T-I)d}} \psi_J(a,d,\varrho)\\
&= \inf_{\substack{(a,d,\varrho)\in\mathbb{R}_+^K\times\mathbb{R}_+^K\times\mathbb{S}_+^{K\times K}:\\ y=a+(\varrho^T-I)d}} \Bigg(\sum_{k=1}^K \sup_\theta(\theta a_k - (e^\theta - 1)\lambda_k)\\
&\quad + \sum_{k\in J^c} \sup_\theta(\theta d_k - (e^\theta-1)\mu_k) + \sum_{k\in J}\sup_\theta(\theta d_k - (e^\theta - 1)\mu_k)\mathbf{1}_{(\mu_k,\infty)}(d_k)\\
&\quad + \sum_{k=1}^K d_k\Bigg[\sum_{\ell=1}^K \sup_\theta(\theta\varrho_{k\ell} - (e^\theta-1)p_{k\ell}) + \sup_\theta\Bigg(\theta\Big(1 - \sum_{\ell=1}^K \varrho_{k\ell}\Big) - (e^\theta - 1)\Big(1 - \sum_{\ell=1}^K p_{k\ell}\Big)\Bigg)\Bigg]\Bigg)\\
&= \inf_{\substack{(a,d,\varrho)\in\mathbb{R}_+^K\times\mathbb{R}_+^K\times\mathbb{S}_+^{K\times K}:\\ y=a+(\varrho^T-I)d}} \sup_{\substack{\theta_k,\vartheta_k,\sigma_{k\ell},\tau_k,\\ k,\ell\in\{1,2,\ldots,K\}}} \Bigg(\sum_{k=1}^K(\theta_k a_k - (e^{\theta_k} - 1)\lambda_k)
\end{aligned}$$

Fig. 9. Optimal trajectories for the parameters of Fig. 8

Fig. 10. Horizontal motion inside the quadrant

$$+\sum_{k\in J^c}(\vartheta_k d_k-(e^{\vartheta_k}-1)\mu_k)+\sum_{k\in J}(\vartheta_k d_k-(e^{\vartheta_k}-1)\mu_k)\mathbf{1}_{(\mu_k,\infty)}(d_k)$$
$$+\sum_{k=1}^K d_k\Bigg[\sum_{\ell=1}^K(\sigma_{k\ell}\varrho_{k\ell}-(e^{\sigma_{k\ell}}-1)p_{k\ell})+\tau_k\Big(1-\sum_{\ell=1}^K\varrho_{k\ell}\Big)-(e^{\tau_k}-1)\Big(1-\sum_{\ell=1}^K p_{k\ell}\Big)\Bigg]\Bigg)$$
$$=\sup_{\substack{\theta_k,\vartheta_k,\sigma_{k\ell},\tau_k,\\ k,\ell\in\{1,2,\ldots,K\}}}\ \inf_{\substack{(a,d,\varrho)\in\mathbb{R}_+^K\times\mathbb{R}_+^K\times\mathbb{S}_+^{K\times K}:\\ y=a+(\varrho^T-I)d}}\Bigg(\sum_{k=1}^K\theta_k a_k+\sum_{k\in J^c}\vartheta_k d_k+\sum_{k\in J}\vartheta_k d_k\mathbf{1}_{(\mu_k,\infty)}(d_k)$$
$$+\sum_{k=1}^K d_k\sum_{\ell=1}^K\sigma_{k\ell}\varrho_{k\ell}+\sum_{k=1}^K d_k\tau_k\Big(1-\sum_{\ell=1}^K\varrho_{k\ell}\Big)-\sum_{k=1}^K(e^{\theta_k}-1)\lambda_k-\sum_{k\in J^c}(e^{\vartheta_k}-1)\mu_k$$
$$-\sum_{k\in J}(e^{\vartheta_k}-1)\mu_k\mathbf{1}(d_k>\mu_k)-\sum_{k=1}^K d_k\sum_{\ell=1}^K(e^{\sigma_{k\ell}}-1)p_{k\ell}-\sum_{k=1}^K d_k(e^{\tau_k}-1)\Big(1-\sum_{\ell=1}^K p_{k\ell}\Big)\Bigg)$$
$$=\sup_{\substack{\theta_k,\vartheta_k,\sigma_{k\ell},\tau_k,\\ k,\ell\in\{1,2,\ldots,K\}}}\ \inf_{\substack{(d,\varrho)\in\mathbb{R}_+^K\times\mathbb{S}_+^{K\times K}:\\ y\ge(\varrho^T-I)d}}\Bigg(\sum_{k=1}^K\theta_k y_k-\sum_{k=1}^K\sum_{\ell=1}^K\theta_\ell\varrho_{k\ell}d_k+\sum_{k=1}^K\theta_k d_k+\sum_{k\in J^c}\vartheta_k d_k$$
$$+\sum_{k\in J}\vartheta_k d_k\mathbf{1}_{(\mu_k,\infty)}(d_k)+\sum_{k=1}^K d_k\sum_{\ell=1}^K\sigma_{k\ell}\varrho_{k\ell}+\sum_{k=1}^K d_k\tau_k\Big(1-\sum_{\ell=1}^K\varrho_{k\ell}\Big)-\sum_{k=1}^K(e^{\theta_k}-1)\lambda_k$$
$$-\sum_{k\in J^c}(e^{\vartheta_k}-1)\mu_k-\sum_{k\in J}(e^{\vartheta_k}-1)\mu_k\mathbf{1}_{(\mu_k,\infty)}(d_k)-\sum_{k=1}^K d_k\sum_{\ell=1}^K(e^{\sigma_{k\ell}}-1)p_{k\ell}$$
$$-\sum_{k=1}^K d_k(e^{\tau_k}-1)\Big(1-\sum_{\ell=1}^K p_{k\ell}\Big)\Bigg).$$

Let us find the infimum over ϱ. One needs to minimise $\sum\limits_{k=1}^K\sum\limits_{\ell=1}^K(-\theta_\ell+\sigma_{k\ell}-\tau_k)d_k\varrho_{k\ell}$ given that $\sum\limits_{k=1}^K\varrho_{k\ell}d_k\le y_\ell+d_\ell$, $\varrho_{k\ell}\ge 0$ and $\sum\limits_{\ell=1}^K\varrho_{k\ell}\le 1$. By the method of Lagrange multipliers (see, e.g., [18, Theorem 6.2.4, p. 196]), there exists a nonnegative vector $\alpha=(\alpha_1,\ldots,\alpha_K)$ such that

$$\sum_{k=1}^K\sum_{\ell=1}^K(-\theta_\ell+\sigma_{k\ell}-\tau_k+\alpha_\ell)d_k\varrho_{k\ell}+\alpha_\ell)d_k\varrho_{k\ell}-\sum_{l=1}^K\alpha_l(y_\ell+d_\ell)$$

is minimised assuming that $\varrho_{k\ell}\ge 0$ and $\sum\limits_{\ell=1}^K\varrho_{k\ell}\le 1$. If $\min\limits_\ell(-\theta_\ell+\sigma_{k\ell}-\tau_k+\alpha_\ell)\le 0$, one takes $\varrho_{km}=1$, where $(-\theta_m+\sigma_{km}-\tau_k+\alpha_m)=\min\limits_\ell(\theta_\ell+\sigma_{k\ell}-\tau_k+\alpha_\ell)$ and $\varrho_{k\ell}=0$ when $\ell\ne m$. If $\min\limits_\ell(-\theta_\ell+\sigma_{k\ell}-\tau_k+\alpha_\ell)>0$, then $\varrho_{k\ell}=0$. It follows that the minimum equals $\sum\limits_k d_k\min\limits_\ell(-\theta_\ell+\sigma_{k\ell}-\tau_k+\alpha_\ell)\wedge 0-\sum\limits_k\alpha_k(y_k+d_k)$, where it is denoted $u\wedge 0=\min(u,0)$. Maximum over α is attained for $\alpha=0$. It is the sought α.

Fig. 11. Vertical motion inside the quadrant

Thus, introducing

$$V_k = \sum_{\ell=1}^{K}(e^{\sigma_{k\ell}}-1)p_{k\ell} + (e^{\tau_k}-1)\Bigl(1-\sum_{\ell=1}^{K}p_{k\ell}\Bigr) + \max_{\ell}(\theta_\ell - \sigma_{k\ell}) \vee (-\tau_k), \tag{A.1}$$

where it is denoted $u \vee v = \max(u,v)$, we have that

$$\begin{aligned}
L_J(y) = &\sup_{\substack{\theta_k,\vartheta_k,\sigma_{k\ell},\tau_k,\\ k,\ell\in\{1,2,\ldots,K\}}} \inf_{d\in\mathbb{R}_+^K}\biggl(\sum_{k=1}^{K}\theta_k y_k - \sum_{k=1}^{K}V_k d_k + \sum_{k\in J}(\vartheta_k \mathbf{1}_{(\mu_k,\infty)}(d_k)+\theta_k)d_k \\
&+ \sum_{k\in J^c}(\theta_k+\vartheta_k)d_k - \sum_{k=1}^{K}(e^{\theta_k}-1)\lambda_k - \sum_{k\in J^c}(e^{\vartheta_k}-1)\mu_k - \sum_{k\in J}(e^{\vartheta_k}-1)\mu_k\mathbf{1}_{(\mu_k,\infty)}(d_k)\biggr) \\
= &\sup_{\substack{\theta_k,\vartheta_k,\sigma_{k\ell},\tau_k:\\ \theta_k+\vartheta_k\ge V_k}} \inf_{d\in\mathbb{R}_+^K}\biggl(\sum_{k=1}^{K}\theta_k y_k - \sum_{k=1}^{K}V_k d_k + \sum_{k\in J}(\vartheta_k \mathbf{1}_{(\mu_k,\infty)}(d_k)+\theta_k)d_k \\
&+ \sum_{k\in J^c}(\theta_k+\vartheta_k)d_k - \sum_{k=1}^{K}(e^{\theta_k}-1)\lambda_k - \sum_{k\in J^c}(e^{\vartheta_k}-1)\mu_k - \sum_{k\in J}(e^{\vartheta_k}-1)\mu_k\mathbf{1}_{(\mu_k,\infty)}(d_k)\biggr) \\
= &\sup_{\substack{\theta_k,\vartheta_k,\sigma_{k\ell},\tau_k:\\ \theta_k+\vartheta_k\ge V_k}} \biggl(\sum_{k=1}^{K}\theta_k y_k + \sum_{k\in J}\mu_k(\theta_k+\vartheta_k-V_k-(e^{\vartheta_k}-1))\wedge 0 - \sum_{k=1}^{K}(e^{\theta_k}-1)\lambda_k \\
&- \sum_{k\in J^c}(e^{\vartheta_k}-1)\mu_k\biggr).
\end{aligned} \tag{A.2}$$

Let us clarify how the latter equation is arrived at. Taking the infimum of the expressions that involve d_k over $d_k > \mu_k$, for $k \in J$, obtains $\mu_k\bigl(-V_k+\theta_k+\vartheta_k-(e^{\vartheta_k}-1)\bigr)$. Taking the infimum over $d_k \le \mu_k$ obtains $\mu_k(\theta_k - V_k)\wedge 0$. Since $\vartheta_k - (e^{\vartheta_k}-1) \le 0$, the minimum of these expressions is $\mu_k\bigl(-V_k+\theta_k+\vartheta_k-(e^{\vartheta_k}-1)\bigr)\wedge 0$.

Let us minimise V_k over $\sigma_{k\ell},\tau_k$. By (A.1), that function is convex in $(\sigma_{k\ell},\tau_k)$. If $p_{k\ell}=0$, for some ℓ, then one can let $\sigma_{k\ell}=\infty$, which enables one to restrict attention to ℓ such that $p_{k\ell}>0$. Similarly, one can exclude τ_k if $\sum_{\ell=1}^{K}p_{k\ell}=1$ by letting $\tau_k=\infty$. Therefore, the minimum on the righthand side of (A.1) is attained. The subdifferential at that point contains the zero vector. Let β_k denote the maximum on the righthand side of (A.1). The subdifferential of the righthand side of (A.1) with respect to $((\sigma_{k\ell},\ \ell\in\{1,2,\ldots,K\}),\tau_k)$ is

$$\biggl((e^{\sigma_{k\ell}}p_{k\ell},\ \ell\in\{1,2,\ldots,K\}),\ e^{\tau_k}\Bigl(1-\sum_{\ell=1}^{K}p_{k\ell}\Bigr)\biggr) - \mathrm{co}\bigl((\mathbf{1}_U(\ell),\ \{\ell\in 1,2,\ldots,K\}),\ \mathbf{1}_{\beta_k}(-\tau_k)\bigr),$$

where co denotes the convex hull and $U$is the set (possibly, empty) of ℓ such that $\theta_\ell - \sigma_{k\ell} = \beta_k$. Since at a point of a minimum the subdifferential contains the zero vector, U contains all ℓ such that $p_{k\ell}>0$ and $\tau_k=-\beta_k$ provided $\sum_{\ell=1}^{K}p_{k\ell}<1$. Hence, there exist nonnegative $\alpha_1,\ldots,\alpha_K,\alpha_{K+1}$, which sum to 1 such that $e^{\sigma_{k\ell}}p_{k\ell}-\alpha_\ell=0$ when $\ell\in U$, $\alpha_\ell=0$ when $\ell\notin U$, and $e^{\tau_k}\Bigl(1-\sum_{\ell=1}^{K}p_{k\ell}\Bigr)-\alpha_{K+1}=0$

provided $\tau_k = -\beta_k$ and $\alpha_{K+1} = 0$, otherwise. Therefore, $\alpha_\ell = e^{\sigma_{k\ell}} p_{k\ell}$ for all $\ell \in \{1, 2, \dots, K\}$ and $\alpha_{K+1} = e^{\tau_k}\Bigl(1 - \sum\limits_{\ell=1}^K p_{k\ell}\Bigr)$. We obtain that $\sum\limits_{\ell=1}^K e^{\sigma_{k\ell}} p_{k\ell} + e^{\tau_k}\Bigl(1 - \sum\limits_{\ell=1}^K p_{k\ell}\Bigr) = 1$. Since $\theta_\ell - \sigma_{k\ell} = \beta_k$ when $p_{k\ell} > 0$ and $\tau_k = -\beta_k$ when $1 - \sum\limits_{\ell=1}^K p_{k\ell} > 0$, we have that $\sum\limits_{\ell=1}^K e^{-\beta_k + \theta_\ell} p_{k\ell} + e^{-\beta_k}\Bigl(1 - \sum\limits_{\ell=1}^K p_{k\ell}\Bigr) = 1$, i.e., $e^{\beta_k} = \sum\limits_{\ell=1}^K e^{\theta_\ell} p_{k\ell} + 1 - \sum\limits_{\ell=1}^K p_{k\ell}$. Finally, $\beta_k = \ln\Bigl(\sum\limits_{\ell=1}^K e^{\theta_\ell} p_{k\ell} + 1 - \sum\limits_{\ell=1}^K p_{k\ell}\Bigr)$ and the minimum of V_k equals $\sum\limits_{\ell=1}^K (e^{-\beta_k + \theta_\ell} - 1) p_{k\ell} + (e^{-\beta_k} - 1)\Bigl(1 - \sum\limits_{\ell=1}^K p_{k\ell}\Bigr) + \beta_k = \beta_k$.

Consequently, the rightmost side of (A.2) equals

$$\sup_{\theta_k, \vartheta_k :\, \theta_k + \vartheta_k \ge \beta_k} \Biggl(\sum_{k=1}^K \theta_k y_k + \sum_{k \in J} \mu_k \bigl(\theta_k + \vartheta_k - \beta_k - (e^{\vartheta_k} - 1)\bigr) \wedge 0 - \sum_{k=1}^K (e^{\theta_k} - 1)\lambda_k - \sum_{k \in J^c} (e^{\vartheta_k} - 1)\mu_k \Biggr). \tag{A.3}$$

If $\theta_k - \beta_k \ge 0$, for $k \in J$, then choosing $\vartheta_k = 0$ shows that the supremum over ϑ_k of the factor at μ_k is zero. Therefore (A.3) can be written as

$$\sup_{\theta_k, \vartheta_k :\, \theta_k + \vartheta_k \ge \beta_k} \Biggl(\sum_{k=1}^K \theta_k y_k + \sum_{k \in J} \mathbf{1}_{(-\infty, \beta_k)}(\theta_k) \mu_k \bigl(\theta_k + \vartheta_k - \beta_k - (e^{\vartheta_k} - 1)\bigr) \wedge 0 - \sum_{k=1}^K (e^{\theta_k} - 1)\lambda_k - \sum_{k \in J^c} (e^{\vartheta_k} - 1)\mu_k \Biggr).$$

Since the function $u - (e^u - 1)$ is decreasing for $u \ge 0$, the supremum over ϑ_k for $k \in J$ is attained at the boundary of the domain defined by the constrains. Equation (1.5) now follows from the definition of $h_k(\theta)$ in (1.4). △

Proof of Lemma 7. Since $C = (I - P^T)^{-1}$, we have that

$$c_{m\ell} = \frac{1}{\det(I - P^T)} (-1)^{m+\ell} M_{\ell m},$$

where $M_{\ell m}$ represents the (ℓ, m)-minor of the matrix $I - P^T$. Let us note that $\det(I - P^T) > 0$. Indeed, $\det I = 1$ and $\det(I - \lambda P^T) \ne 0$ for $\lambda \in [0, 1]$, as the spectral radius of P is less than unity. By continuity, $\det(I - P^T) > 0$. Hence, we need to prove that $(-1)^{m+\ell} M_{\ell m} \le M_{mm}$. Suppose that $\ell = m + 1$. We need to prove that $M_{mm} + M_{m+1,m} \ge 0$. By multilinearity of the determinant, $M_{mm} + M_{m+1,m}$ is the determinant of the matrix $\widetilde I - \widetilde P$, where $\widetilde I$ is the identity $((K-1) \times (K-1))$-matrix and $\widetilde P$ is the $((K-1) \times (K-1))$-matrix that is obtained from P^T by adding rows m and $m+1$, with subsequent deletion of the m-th column. Since the matrix P^T is column substochastic, the matrix $\widetilde P$ is column substochastic too. Therefore, its spectral radius is not greater than unity. It follows that $\det(\widetilde I - \widetilde P) \ge 0$, i.e., $M_{m+1,m} + M_{m,m} \ge 0$. Suppose that $\ell > m + 1$. Sequential transpositions of adjacent rows and columns enables us to move in the matrix $\widetilde I - \widetilde P$ row ℓ and column ℓ in the positions of row $m+1$ and column $m+1$, respectively, keeping the mutual positions of the other rows and columns. This matrix is still of the form $\widetilde I - \widetilde{\widetilde P}$. We have that $\widetilde M_{m+1,m} = (-1)^{\ell - m - 1} M_{\ell m}$, because the sign of a minor changes only when transposing the columns, and $\widetilde M_{mm} = (-1)^{2(\ell - m - 1)} M_{mm}$. Since $\widetilde M_{mm} \ge (-1)\widetilde M_{m+1,m}$, $M_{mm} \ge (-1)^{\ell - m - 2} M_{\ell m} = (-1)^{\ell + m} M_{\ell m}$. The case where $\ell < m$ is treated similarly. △

ACKNOWLEDGEMENT

The author expresses gratitude to S.A. Pirogov and A.N. Rybko for helpful discussions and advice on improving the presentation.

REFERENCES

1. Wentzell, A.D. and Freidlin, M.I., *Fluktuatsii v dinamicheskikh sistemakh pod deistviem malykh sluchainykh vozmushchenii*, Moscow: Nauka, 1979. Tanslated under the title *Random Perturbations of Dynamical Systems*, New York: Springer, 1984.
2. Shwartz, A. and Weiss, A., Large Deviations for Performance Analysis: Queues, Communications, and Computing, London: Chapman & Hall, 1995.
3. Kleinrock, L., *Queueing Systems*, vol. 1: *Theory*, New York: Wiley, 1975. Translated under the title *Teoriya massovogo obsluzhivaniya*, Moscow: Mashinostroenie, 1979.
4. Jacod, J. and Shiryaev, A.N., *Limit Theorems for Stochastic Processes*, Berlin: Springer, 1987. Translated under the title *Predel'nye teoremy dlya sluchainykh protsessov*, Moscow: Fizmatlit, 1994.
5. Ethier, S.N. and Kurtz, T.G., *Markov Processes: Characterization and Convergence*, New York: Wiley, 1986.
6. Puhalskii, A.A., The Action Functional for the Jackson Network, *Markov Process. Related Fields*, 2007, vol. 13, no. 1, pp. 99–136.
7. Atar, R. and Dupuis, P., Large Deviations and Queueing Networks: Methods for Rate Function Identification, *Stochastic Process. Appl.*, 1999, vol. 84, no. 2, pp. 255–296.
8. Ignatiouk-Robert, I., Large Deviations of Jackson Networks, *Ann. Appl. Probab.*, 2000, vol. 10, no. 3, pp. 962–1001.
9. Collingwood, J., Path Properties of Rare Events, *PhD Thesis*, Univ. of Ottawa, Canada, 2015.
10. Bouchet, F., Laurie, J., and Zaboronski, O., Langevin Dynamics, Large Deviations and Instantons for the Quasi-geostrophic Model and Two-Dimensional Euler Equations, *J. Stat. Phys.*, 2014, vol. 156, no. 6, pp. 1066–1092.
11. Clarke, F.H., *Optimization and Nonsmooth Analysis*, New York: Wiley, 1983. Translated under the title *Optimizatsiya i negladkii analiz*, Moscow: Mir, 1988.
12. Rockafellar, R.T., *Convex Analysis*, Princeton Math. Series, vol. 28, Princeton: Princeton Univ. Press, 1970. Translated under the title *Vypuklyi analiz*, Moscow: Mir, 1973.
13. Bramson, M., *Stability of Queueing Networks (Ecole d'Été de Probabilités de Saint-Flour XXXVI-2006)*, Lect. Notes Math., vol. 1950, Berlin: Springer, 2008.
14. Anantharam, V., Heidelberger, R., and Tsoukas, P., *Analysis of Rare Events in Continuous Time Markov Chains via Time Reversal and Fluid Approximation*, IBM Res. Rep. RC 16280, Yorktown Heights, NY, 1990.
15. Shwartz, A. and Weiss, A., Induced Rare Events: Analysis via Large Deviations and Time Reversal, *Adv. in Appl. Probab.*, 1993, vol. 25, no. 3, pp. 667–689.
16. Majewski, K. and Ramanan, K., How Large Queue Lengths Build Up in a Jackson Network, *Preprint*, 2008.
17. Vinter, R.B., *Optimal Control*, Boston: Birkhäuser, 2000.
18. Bazaraa, M.S. and Shetty, C.M., *Nonlinear Programming: Theory and Algorithms*, New York: Wiley, 1979. Translated under the title *Nelineinoe programmirovanie. Teoriya i algoritmy*, Moscow: Mir, 1982.