JOURNAL OF FUNCTIONAL ANALYSIS 123, 264–307 (1994)

Entropy Minimization, *DAD* Problems, and Doubly Stochastic Kernels

J. M. BORWEIN* AND A. S. LEWIS*

Department of Combinatorics and Optimization, University of Waterloo, Waterloo, Ontario, Canada N2L 3G1

AND

R. D. NUSSBAUM†

Department of Mathematics, Hill Center, Busch Campus, The State University of New Jersey (Rutgers), New Brunswick, New Jersey 08903

Communicated by the Editors

Received February 10, 1993

The classical *DAD* problem asks, for a square matrix A with nonnegative entries, when it is possible to find positive diagonal matrices D_1 and D_2 with D_1AD_2 doubly stochastic. We consider various continuous and measurable generalizations of this problem. Through a fusion of variational and fixed point techniques we obtain strong analogues of the classical results. Our extensions appear inaccessible by either technique separately. © 1994 Academic Press, Inc.

I. VARIATIONAL METHODS FOR INFINITE *DAD* PROBLEMS

1. *Introduction*

Suppose that A is an $n \times n$ matrix with nonnegative entries. The classical *DAD* problem asks when it is possible to find diagonal matrices D_1 and D_2 with positive diagonal entries such that the scaled matrix D_1AD_2 is doubly stochastic (in other words, having row and column sums 1). This problem is an important special case of a large class of matrix scaling problems, with numerous applications in diverse fields. A survey may be found in [24].

* Partially supported by the Natural Sciences and Engineering Research Council of Canada.

† Partially supported by NSFDMS 9105930.

0022-1236/94 $6.00

Necessary and sufficient conditions for the existence and uniqueness (up to a scalar multiple) of the required diagonal matrices have been known since the 1960s. For example, in [4] and independently in [26], it was shown that a necessary and sufficient condition for existence is that A be a direct sum of fully indecomposable matrices. This property is equivalent to the existence of a doubly stochastic matrix with the same pattern of positive entries as A [20].

It is natural to try to extend these elegant results to a continuous setting or a measurable setting. For example, given a nonnegative continuous function k on the unit square and positive continuous functions α and β on the unit interval, when do there exist positive continuous functions f and g on the unit interval such that

$$\begin{aligned} &\int_0^1 f(s)\, k(s, t)\, g(t)\, dt = \alpha(s), \qquad \text{for all} \quad s \in [0, 1], \qquad \text{and} \\ &\int_0^1 f(s)\, k(s, t)\, g(t)\, ds = \beta(t), \qquad \text{for all} \quad t \in [0, 1]? \end{aligned} \tag{1.1}$$

A related problem, in which one assumes $\alpha \equiv \beta \equiv 1$ and k is symmetric and one seeks $f = g$ above, was studied in [15] and later in [12]. See also [10] and Section 4 of [19].

Generally speaking there have been two more-or-less disjoint approaches to these problems: the fixed-point method and the variational method. The fixed point method begins with the observation that finding a solution of, for example, (1.1) is equivalent to finding a fixed point of an associated map Φ which is defined on the interior of a cone in a Banach space. Thus, one tries to prove existence of a fixed point of Φ in the interior of a cone. See, for example, [4, 14, 12]. More recently, this approach has been supplemented by the observation that Φ behaves nicely with respect to Hilbert's projective metric: see [9, 19] and also [17]. This observation yields uniqueness results and geometric convergence of various iterative schemes.

For the variational method, the basic technique has been that of entropy minimization. The idea is to minimize the Boltzmann–Shannon entropy ($x \log x$) of a joint distribution (in the original problem, an $(n \times n)$ matrix, and in (1.1) a density on the unit square), subject to suitable marginal constraints. At least formally, it is not difficult to see that the optimal distribution decomposes into two independent univariate distributions, which provide exactly the desired scaling. In fact these distributions are related directly to the Lagrange multipliers at the optimum, suggesting a practical technique for solving the problem via duality. The desired necessary and sufficient condition is exactly (in the matrix case) the constraint qualification required to apply Lagrangian techniques.

This technique has been quite widely applied in the matrix case (see for example [24]). In the continuous case the idea was used in a special case (with $k>0$) in [13] and in [6], both papers being in a probabilistic framework. However, in Csiszár's paper the underlying optimization is hidden and some of the derivations appear obscure. In particular, the proof of Corollary 3.1 on p. 154 (case (B)) requires a certain subspace to be closed. This point requires careful argument (see [1]) and is even more problematic in some of the generalizations (p. 155). In [19] Csiszár's results were combined with the fixed point approach to show that, if $\alpha \equiv \beta \equiv 1$, a sufficient condition for the existence of a scaling satisfying (1.1) is that k be strictly positive on the main diagonal: $k(s, s)>0$ for all $s \in [0, 1]$. Under these assumptions, results in our Section 4 and [19] also imply that the functions f and g in (1.1) are unique to within positive scalar multiples and can be obtained by an iterative scheme which converges geometrically.

Our principal aims in this paper are to provide, starting from first principles, a rigorous, unified optimization-theoretic framework in which to prove and generalize results on *DAD* problems and to use these theorems to obtain a generalization of the above-mentioned result of Nussbaum. Our initial discussion (Sections 1–3) is restricted to the entropy method, but some of our later theorems (for example Theorem 5.16) involve a combination of the fixed point approach and the entropy method to obtain results which do not follow obviously from either method separately. Our theorems will be phrased in the general framework of regular Borel measures on compact, Hausdorff spaces, so as to include in a unified way the matrix case and (1.1). We shall periodically refer to the matrix case, but we should emphasize that the matrix case is essentially well-understood. Our real interest is in examples like (1.1) under varying assumptions on k, α, and β. If, for example, k is only known to be nonnegative, measurable and essentially bounded and α and β are positive almost everywhere and integrable, (1.1) poses a variety of difficulties, many of which have no analogues in the matrix case.

We begin by studying the general optimization problem of minimizing the Boltzmann–Shannon entropy of a density, subject to abstract linear constraints. The existence of an optimal solution follows from the weak compactness of the level sets of the entropy function, and uniqueness may be deduced by strict convexity.

We next introduce a suitable constraint qualification. The condition we use is a weakened, "quasi-interior" version of the standard Slater condition of convex optimization, of the type studied in [2]. It requires the existence of a strictly positive feasible solution of the optimization problem, which for *DAD* problems becomes exactly the "doubly-stochastic pattern"

property of [20]. Using this constraint qualification we deduce a necessary "optimality" condition which is an asymptotic Lagrange multiplier result.

In order to close the gap between the necessary and the sufficient conditions, and hence to deduce the existence of the required Lagrange multipliers, we need to impose more structure on the problem. Two cases prove tractable. The first is when the abstract linear constraint has finite-dimensional range, giving a "partially finite" convex program. In this case we rederive results in [3], and obtain the known conditions for matrix *DAD* problems. The second case, in which we are particularly interested, is that of marginal constraints. To deduce the existence of multipliers here we apply decomposition results from [1]. In this framework it is not difficult to generalize the result in directions suggested above.

The remaining question concerns conditions ensuring the satisfaction of the "constraint qualification." Basically, the problem is to show that the set of functions $u(s, t)$ satisfying certain constraints is nonempty. If we can prove this, our entropy minimization approach provides a solution to our *DAD* problem. However, when the underlying problem is not finite dimensional, satisfying the constraint qualification may be nontrivial. With the aid of Theorem 5.16 and a fixed point argument, we show how the constraint qualification can be satisfied when $\alpha \equiv \beta$, α is positive and integrable, and k is nonnegative, essentially bounded, and measurable. In a future paper, one of the authors (N.) will show how related ideas can be applied when $\alpha \not\equiv \beta$. Once one has proved existence of a solution of a *DAD* problem, Theorem 5.14 in Section 5 provides a strong result about uniqueness and convergence of an iterative scheme to approximate the solution.

2. Abstract Entropy Minimization

In this section we introduce the Boltzmann–Shannon entropy and consider the optimization problem of minimizing entropy subject to abstract linear constraints. We prove the existence and uniqueness of optimal solutions and give a sufficient condition for optimality. We then present a "quasi-interior" constraint qualification, and under this assumption derive a necessary condition for optimality.

Consider the closed, convex function $\phi: \mathbb{R} \to (-\infty, +\infty]$ defined by

$$\phi(r) := \begin{cases} r \log r - r, & \text{if } r > 0, \\ 0, & \text{if } r = 0, \\ +\infty, & \text{if } r < 0. \end{cases} \tag{2.1}$$

Suppose (P, dp) is a (nonnegative) finite measure space. We define the Boltzmann–Shannon entropy on P as the integral functional $I_\phi: L_1(P, dp) \to (-\infty, +\infty]$ defined by

$$I_\phi(u) := \int_P \phi(u(p))\, dp.$$

THEOREM 2.2. *The function I_ϕ is a well-defined, weakly lower semi-continuous convex function, with weakly compact level sets,*

$$\{u \in L_1(P, dp) \mid I_\phi(u) \leqslant \alpha\}$$

for all $\alpha \in \mathbb{R}$.

Proof. See [21]. The weak compactness property follows either from the fact that the conjugate function $\phi^*(w) = e^w$ is everywhere finite, or may be seen directly from the Dunford–Pettis criterion for compactness in the weak topology on $L^1(P, dp)$. (See, for example, [7] for a discussion of the Dunford–Pettis criterion.) ∎

From the fact that ϕ is strictly convex on $[0, +\infty)$, it follows clearly that I_ϕ is strictly convex on its domain:

$$\operatorname{dom} I_\phi := \{u \in L_1(P) \mid I_\phi(u) < +\infty\}.$$

Note that we may have $0 \leqslant u \in L_1(P)$ and yet $I_\phi(u) = +\infty$. For example, take $P := [0, 1]$ and $dp := e^{-1/p}\, d\lambda$, where λ is Lebesgue measure. Then $u(p) := e^{1/p}$ clearly satisfies $0 \leqslant u \in L_1(P)$ and yet $I_\phi(u) = +\infty$.

Let Z be an arbitrary locally convex (Hausdorff) topological vector space, whose topological dual we denote Z^*. Suppose that $A: L_1(P) \to Z$ is a continuous linear map, with adjoint $A^*: Z^* \to L_\infty(P)$, and suppose that $b \in Z$. The optimization problem that we wish to consider is

$$\text{(EM)} \quad \begin{cases} \inf & I_\phi(u) \\ \text{subject to} & Au = b, \text{ and} \\ & u \in L_1(P). \end{cases}$$

We say $u \in L_1(P)$ is *feasible* if $u \in \operatorname{dom} I_\phi$ and $Au = b$.

COROLLARY 2.3. *Suppose that* (EM) *is* consistent, *meaning there exists a feasible u. Then it has a unique optimal solution.*

Proof. This is a direct application of Theorem 2.2 and the strict convexity of I_ϕ. ∎

LEMMA 2.4. *Suppose* $0 < r_0 \in \mathbb{R}$. *Then for all* $r \in \mathbb{R}$,

$$(r - r_0) \log r_0 \leqslant \phi(r) - \phi(r_0).$$

Proof. This is just the subgradient inequality. ∎

We can now prove sufficient conditions for optimality in (EM).

PROPOSITION 2.5. *Suppose that* u_0 *is feasible for* (EM). *Suppose further that there exists* $\mu \in Z^*$ *with* $A^*\mu = \log u_0$ *a.e. Then* u_0 *is the unique optimal solution of* (EM).

Proof. Since $\log u_0 = A^*\mu \in L_\infty(P)$, $u_0 > 0$ a.e. Suppose $u \in L_1(P)$ and $Au = b$. By Lemma 2.4,

$$\phi(u(p)) - \phi(u_0(p)) \geqslant (u(p) - u_0(p)) \log u_0(p) \qquad \text{a.e.}$$

Integrating over P gives

$$I_\phi(u) - I_\phi(u_0) \geqslant \langle u - u_0, \log u_0 \rangle = \langle u - u_0, A^*\mu \rangle = \langle A(u - u_0), \mu \rangle = 0,$$

so u_0 is optimal. Uniqueness follows from Corollary 2.3. ∎

As usual in optimization, to derive a necessary condition for optimality we require a constraint qualification. The condition we will use is:

(CQ) There exists $\hat{u} > 0$ a.e. which is feasible for (EM).

This type of condition is called in [2] a "quasi-interior" constraint qualification. More specifically, it is shown in [3] that if (CQ) holds then $\hat{u}$ lies in the "quasi relative interior" of dom I_ϕ.

LEMMA 2.6. *For* $r \in (0, +\infty)$, *and* $d \in \mathbb{R}$,

(i) $\lambda^{-1}(\phi(r + \lambda d) - \phi(r)) \downarrow d \log r$ *as* $\lambda \downarrow 0$.

(ii) $\lambda^{-1}\phi(\lambda d) \downarrow -\infty$ *as* $\lambda \downarrow 0$, *for* $d > 0$.

Proof. The lemma follows from the convexity of ϕ and the fact that $\phi'(r; d) = d \log r$. ∎

THEOREM 2.7. *Suppose* (CQ) *holds. Then the unique optimal solution of* (EM), u_0, *satisfies* $u_0 > 0$ *a.e.*

Proof. Suppose the measurable set $P_0 = \{p \in P \mid u_0(p) = 0\}$ has positive measure. By assumption $\phi(\hat{u})$ and $\phi(u_0) \in L_1(P)$, and we have by Lemma 2.6 as $\lambda \downarrow 0$,

$$\phi(\hat{u}) - \phi(u_0) \geq \lambda^{-1}(\phi(u_0 + \lambda(\hat{u} - u_0)) - \phi(u_0))$$
$$\downarrow \begin{cases} -\infty, & \text{a.e. on } P_0, \\ (\hat{u} - u_0)\log u_0, & \text{a.e. on } P_0^c. \end{cases}$$

It follows by the Monotone Convergence Theorem [22] that

$$\lambda^{-1}(I_\phi(u_0 + \lambda(\hat{u} - u_0)) - I_\phi(u_0)) \downarrow -\infty,$$

which is a contradiction, since this quotient is nonnegative by the feasibility of $\hat{u}$ and the optimality of u_0. ∎

We shall denote the weak* closure of a set D by $w^*\text{-cl}(D)$.

LEMMA 2.8. *Let W be a locally convex topological vector space, and suppose $D \subset W$ is convex. Then regarding W as a subspace of W^{**} we have $W \cap w^*\text{-cl}(D) = \text{cl}(D)$, where the closure of D is taken in W.*

Proof. Suppose $w \in W \cap w^*\text{-cl}(D)$, so for some net (w_α) in D, $\langle w_\alpha - w, \mu \rangle \to 0$ for all $\mu \in W^*$. Thus $w_\alpha \to w$ weakly in W, so w lies in the weak closure of D, which equals its closure since D is convex [23]. The opposite conclusion is immediate. ∎

We can now prove necessary conditions for optimality (c.f. [6]).

THEOREM 2.9. *Suppose* (CQ) *holds. Then there exists a unique optimal solution to* (EM), u_0. *Furthermore, $u_0 > 0$ a.e., and there exists a sequence $\mu_0, \mu_2, \ldots \in Z^*$ with $\|u_0(A^*\mu_n - \log u_0)\|_1 \to 0$.*

Proof. The first two statements follow from Theorem 2.7, and, from the proof of this result we deduce that for any feasible u for (EM), if we denote the directional derivative

$$I'_\phi(u_0; u - u_0) = \lim_{\lambda \downarrow 0} \lambda^{-1}(I_\phi(u_0 + \lambda(u - u_0)) - I_\phi(u_0)),$$

then we have

$$0 \leq I'_\phi(u_0; u - u_0) = \int_P [u - u_0] \log u_0. \tag{2.10}$$

Define a linear map $B: L_\infty(P) \to Z$ by $Bh := A(hu_0)$. The continuity of B follows from the continuity of A and the fact that $h \mapsto hu_0$ is continuous

since $\|hu_0\|_1 \leq \|u_0\|_1 \|h\|_\infty$. The adjoint of B, $B^*: Z^* \to L_\infty^*(P)$ can be computed from the fact that for all $h \in L_\infty(P)$ and $\mu \in Z^*$,

$$\langle h, B^*\mu \rangle = \langle Bh, \mu \rangle = \langle A(hu_0), \mu \rangle = \langle hu_0, A^*\mu \rangle = \langle h, u_0 A^*\mu \rangle,$$

so $B^*\mu = u_0 A^*\mu$. Thus the range of B^*, $R(B^*)$, is contained in $L_1(P)$, regarded as a subspace of $L_\infty^*(P)$.

Now suppose h is in the null space of B, $N(B)$. Denote the function with constant value 1 by 1. Then for any ε,

$$A((1+\varepsilon h)u_0) = B(1+\varepsilon h) = B1 = Au_0 = b.$$

If furthermore $|\varepsilon| < \|h\|_\infty^{-1}$ then

$$\begin{aligned} I_\phi((1+\varepsilon h)u_0) &= \int_P \{(1+\varepsilon h)\, u_0 \log((1+\varepsilon h)u_0) - (1+\varepsilon h)u_0\} \\ &= \int_P \{(1+\varepsilon h)(u_0 \log u_0) + ((1+\varepsilon h)\log(1+\varepsilon h) - 1)u_0\} \\ &\leq \|1+\varepsilon h\|_\infty \|u_0 \log u_0\|_1 + \|(1+\varepsilon h)\log(1+\varepsilon h) - 1\|_\infty \|u_0\|_1 \\ &< +\infty. \end{aligned}$$

Thus $(1+\varepsilon h)u_0$ is feasible for (EM), so from (2.10),

$$0 \leq \int_P [(1+\varepsilon h)u_0 - u_0] \log u_0 = \int_P \varepsilon h u_0 \log u_0.$$

Thus for any $h \in N(B)$, $\langle h, u_0 \log u_0 \rangle = 0$, so

$$u_0 \log u_0 \in N(B)^\perp = ({}^\perp R(B^*))^\perp = w^*\text{-cl}(R(B^*))$$

(see, for example, [23]). It follows by Lemma 2.8 that in $L_1(P)$,

$$u_0 \log u_0 \in \text{cl}(R(B^*)),$$

so for some sequence $\mu_1, \mu_2, \ldots \in Z^*$, $\|B^*\mu_n - u_0 \log u_0\|_1 \to 0$, which gives the desired conclusion. ∎

The Lagrangian for the problem (EM) is

$$L(u; \mu) := I_\phi(u) + \langle b - Au, \mu \rangle = I_\phi(u) - \langle u, A^*\mu \rangle + \langle b, \mu \rangle,$$

and the corresponding dual problem is therefore

$$\sup_{\mu \in Z^*} \inf_{u \in L_1} \{\langle b, \mu \rangle + I_\phi(u) - \langle u, A^*\mu \rangle\},$$

which we can write as

$$(\mathrm{EM}^*) \qquad \sup_{\mu \in Z^*} \{\langle b, \mu \rangle - I_\phi^*(A^*\mu)\},$$

where by [21], $I_\phi^*: L_\infty(P) \to (-\infty, +\infty]$ is given by

$$I_\phi^*(w) = I_{\phi^*}(w) = \int_P e^{w(p)}\, dp.$$

The dual problem is thus an unconstrained concave maximization. If u_0 is feasible for (EM), μ is a Lagrange multiplier, or equivalently an optimal dual solution, if

$$0 \in \partial_u L(u_0, \mu) = \partial I_\phi(u_0) - A^*\mu,$$

or $A^*\mu \in \partial\phi(u_0) = \{\log u_0\}$ a.e. (see [21]). Thus Proposition 2.5 may be interpreted as saying that if there exists a multiplier $\mu \in Z$ for u_0, or, in other words,

$$A^*\mu = \log u_0 \qquad \text{a.e.}, \tag{2.11}$$

then u_0 is optimal. On the other hand, Theorem 2.9 says that if (CQ) holds and u_0 is optimal then there exists a sequence of "asymptotic" multipliers $\mu_1, \mu_2, \ldots \in Z^*$ for u_0, meaning

$$A^*\mu_n \to \log u_0 \qquad \text{in } L_1(P, u_0\, dp). \tag{2.12}$$

In order to close the gap between (2.11) and (2.12) we need a suitable closed range assumption.

COROLLARY 2.13. *Suppose* (CQ) *holds,* u_0 *is feasible for* (EM), *and* $R(A^*)$ *is closed as a subspace of* $L_1(P, u_0\, dp)$, (*as holds in particular if* $Z = \mathbb{R}^n$). *Then* u_0 *is optimal for* (EM) *if and only if there exists* $\mu \in Z^*$ *with* $A^*\mu = \log u_0$ *a.e.*

Proof. The range of A^* is a subspace of $L_\infty(P, dp)$ and since $u_0 \in L_1(P, dp)$ we may consider it as a subspace of $L_1(P, u_0\, dp)$. The result now follows from (2.12). If $Z = \mathbb{R}^n$, $R(A^*)$ is finite-dimensional so closed. ∎

The fact that (CQ) implies the existence of a multiplier (or optimal dual solution) when $Z = \mathbb{R}^n$ may be found as part of a more general "partially-finite" theory in [3].

3. Continuous *DAD* Problems

The previous section was concerned with minimizing the entropy of a distribution under abstract linear constraints. Theorem 2.9 gave an asymptotic necessary condition for optimality, under the assumption of a quasi-interior constraint qualification. With a suitable closed range condition (as holds in particular if the constraint map has finite-dimensional range) we obtain a necessary and sufficient condition for optimality (Corollary 2.13). In this section we shall examine the case of marginal constraints, where more care is required. We shall return to the question of satisfying the constraint qualification in what follows.

Suppose (S, ds) and (T, dt) are finite measure spaces with

$$0 \leqslant \alpha \in L_1(S, ds),$$
$$0 \leqslant \beta \in L_1(T, dt), \qquad \text{and}$$
$$0 \leqslant k \in L_1(S \times T, ds\, dt).$$

The marginal problem that we wish to consider is

$$\text{(MOM)} \quad \begin{cases} \inf & \displaystyle\int_{S\times T} \phi(u(s,t))\, k(s,t)\, ds\, dt \\ \text{subject to} & \displaystyle\int_T u(s,t)\, k(s,t)\, dt = \alpha(s), \qquad \text{a.e. on } S, \\ & \displaystyle\int_S u(s,t)\, k(s,t)\, ds = \beta(t), \qquad \text{a.e. on } T, \\ & u \in L_1(S\times T, k\, ds\, dt), \end{cases}$$

where $k\, ds\, dt$ is the measure on $S \times T$ with Radon–Nikodym derivative k, and ϕ is defined as in the previous section. If we write $P := S \times T$ and $dp := k\, ds\, dt$ then (MOM) is exactly of the form (EM), where the constraints are marginal conditions on the density u. We therefore obtain the following result directly from the previous section. The constraint qualification becomes

(CQ1) There exists $\hat{u} > 0$ a.e. $[k\, ds\, dt]$ which is feasible for (MOM).

Theorem 3.1. *If* (MOM) *is consistent then it has a unique optimal solution. If* u_0 *is feasible and there exist* $x \in L_\infty(S, ds)$ *and* $y \in L_\infty(T, dt)$ *with*

$$x(s) + y(t) = \log u_0(s, t) \qquad \text{a.e. } [k\, ds\, dt], \tag{3.2}$$

then u_0 *is the unique optimal solution. Conversely, suppose* (CQ1) *holds.*

Then the unique optimal solution u_0 satisfies $u_0 > 0$ a.e. $[k\, ds\, dt]$ and there exist sequences $x_n \in L_\infty(S, ds)$ and $y_n \in L_\infty(T, dt)$ for $n = 1, 2, \ldots$ with

$$\lim_{n \to \infty} (x_n(s) + y_n(t)) = \log u_0(s, t) \qquad \text{a.e. } [k\, ds\, dt] \tag{3.3}$$

Note. In fact we shall show

$$x_n(s) + y_n(t) \to \log u_0(s, t), \qquad \text{in } L_1(S \times T, u_0 k\, ds\, dt) \tag{3.4}$$

Proof. In the notation of the previous section, we set $Z := L_1(S, ds) \times L_1(T, dt)$, and define $A: L_1(S \times T, k\, ds\, dt) \to L_1(S, ds) \times L_1(T, dt)$ by

$$Au := \begin{pmatrix} \int_T u(s, t)\, k(s, t)\, dt \\ \int_S u(s, t)\, k(s, t)\, ds \end{pmatrix}.$$

It follows that $A^*: L_\infty(S, ds) \times L_\infty(T, dt) \to L_\infty(S \times T, k\, ds\, dt)$ is determined by

$$\begin{aligned} \langle u, A^*(x, y) \rangle &= \langle Au, (x, y) \rangle \\ &= \int_S x(s) \int_T u(s, t)\, k(s, t),\, dt\, ds \\ &\quad + \int_T y(t) \int_S u(s, t)\, k(s, t)\, ds\, dt \\ &= \int_{S \times T} u(s, t)(x(s) + y(t))\, k(s, t)\, ds\, dt \end{aligned}$$

(by Fubini's Theorem), for all u, x, y, so

$$(A^*(x, y))(s, t) = x(s) + y(t) \qquad \text{a.e. } [k\, ds\, dt].$$

Existence and uniqueness follows from Corollary 2.3, and the sufficient condition (3.2) follows from Proposition 2.5. Finally, from Theorem 2.9 it follows that $u_0 > 0$ a.e. $[k\, ds\, dt]$ if (CQ1) holds, and that for some sequences (x_n) and (y_n), (3.4) holds. Taking a subsequence converging pointwise a.e. $[k\, ds\, dt]$ (see [22]) and relabeling gives (3.3). ∎

The *DAD* problem asks when it is possible to find nonnegative functions $f: S \to \mathbb{R}$ and $g: T \to \mathbb{R}$ (generally having some further continuity or measurability properties) satisfying

$$\int_T f(s)\, k(s, t)\, g(t)\, dt = \alpha(s), \qquad \text{a.e. on } S, \qquad \text{and}$$
$$\int_S f(s)\, k(s, t)\, g(t)\, ds = \beta(t), \qquad \text{a.e. on } T. \tag{3.5}$$

When $S = T = \{1, 2, ..., n\}$, with ds and dt counting measure, and α and β are identically 1 then this is the classical *DAD* problem. If we can close the gap between the necessary and sufficient conditions and deduce (3.2) from (3.3) then we obtain the solution of the *DAD* problem: (3.2) implies

$$u_0(s, t) = e^{x(s)} e^{y(t)} \qquad \text{a.e. } [k\, ds\, dt],$$

so if we put $f(s) := e^{x(s)}$ and $g(t) := e^{y(t)}$ then (3.5) follows from the feasibility of u_0.

Whether (3.2) necessarily follows from (3.3) is however not immediately clear. This question is passed over in the analogous result in [6], and is treated in detail in [1].

COROLLARY 3.6. *Suppose* (CQ1) *holds. Then the unique optimal solution* u_0 *of* (MOM) *satisfies* $u_0 > 0$ *a.e.* $[k\, ds\, dt]$ *and there exist functions* $x: S \to \mathbb{R}$ *and* $y: T \to \mathbb{R}$ *satisfying*

$$x(s) + y(t) = \log u_0(s, t) \qquad \text{a.e. } [k\, ds\, dt]. \tag{3.7}$$

Furthermore putting $f(s) := e^{x(s)}$ *and* $g(t) := e^{y(t)}$ *solves the DAD problem* (3.5).

Proof. Equation 3.7 follows from (3.3) and [1]. The remainder of the argument is as above. ∎

Note. The constraint qualification (CQ1) has a very useful "scaling" property. Suppose that on $K = \{(s, t) \mid k(s, t) > 0\}$ there exist constants $m > 0$ and $M > 0$ with $mk \leqslant \tilde{k} \leqslant Mk$. Suppose that (MOM) with k replaced by $\tilde{k}$ has a solution. Then (MOM) itself has a solution. Indeed, if $\tilde{u}$ solves the former problem then

$$u = \begin{cases} \tilde{u}\tilde{k}/k, & \text{if } (s, t) \in K, \\ 0 & \text{otherwise,} \end{cases}$$

has finite entropy and is feasible for the original problem. This simple observation will prove very useful in Part II.

If $\alpha(s) = 0$, a.e. on S_0 then clearly we lose nothing in (3.5) if we replace S with $S \setminus S_0$, and similarly for T. Therefore, we suppose in future:

Assumption 3.8.

$$\alpha(s) > 0, \quad \text{a.e. on } S, \quad \text{and}$$

$$\beta(t) > 0, \quad \text{a.e. on } T.$$

COROLLARY 3.9. *Suppose Assumption* 3.8 *holds. Then* (CQ1) *is both necessary and sufficient for there to exist a feasible solution for* (MOM), *with finite value, of the form* $u(s, t) = f(s)\, g(t)$, *a.e.* $[k\, ds\, dt]$, *with both* f *and* g *strictly positive.*

Proof. If (CQ1) holds, the desired conclusion follows immediately from Corollary 3.6, while the converse is immediate. ∎

Note that at this stage we do not even know whether the functions f, g, x, and y in Corollaries 3.6 and 3.9 are measurable. We return to this point in the next section.

The Finite-Dimensional Case

The case where S and T are finite sets is now particularly straightforward.

COROLLARY 3.10. *Suppose* $0 < \alpha_i$, $\beta_j \in \mathbb{R}$, *and* $0 \leqslant k_{ij} \in \mathbb{R}$, *for all* $i = 1, ..., m$, $j = 1, ..., n$. *Then there exist* $0 < f_i$, $g_j \in \mathbb{R}$ *satisfying*

$$\begin{aligned} \sum_{j=1}^{n} f_i k_{ij} g_j &= \alpha_i, \qquad i = 1, ..., m, \\ \sum_{i=1}^{m} f_i k_{ij} g_j &= \beta_j, \qquad j = 1, ..., n, \end{aligned} \tag{3.11}$$

if and only if there exist v_{ij} *satisfying, for each* i, j,

$$v_{ij} \begin{cases} > 0, & \text{if } k_{ij} > 0, \\ = 0, & \text{if } k_{ij} = 0, \end{cases} \tag{3.12}$$

with

$$\begin{aligned} \sum_{j=1}^{n} v_{ij} &= \alpha_i, \qquad i = 1, ..., m, \\ \sum_{i=1}^{m} v_{ij} &= \beta_j, \qquad j = 1, ..., n. \end{aligned} \tag{3.13}$$

Proof. Suppose (3.11) holds. Then $v_{ij} := f_i k_{ij} g_j$ satisfies (3.12) and (3.13). Conversely, suppose (3.12) and (3.13). Define $S := \{1, ..., m\}$ and $T := \{1, ..., n\}$, both with counting measure, and set

$$\hat{u}_{ij} := \begin{cases} v_{ij}/k_{ij}, & \text{if } k_{ij} > 0, \\ 0, & \text{if } k_{ij} = 0, \end{cases}$$

for each i, j. Then $\hat{u}$ is feasible for (MOM) by (3.13) and satisfies (CQ1) by (3.12). The result now follows by Corollary 3.6. ∎

This result is exactly Corollary 3.3 in [6]: a matrix A with nonnegative entries can be scaled by positive diagonal matrices D_1 and D_2 so $D_1 A D_2$ has prescribed positive row and column sums if and only if there exists a matrix B with nonnegative entries and the same zero pattern as A and with the prescribed row and column sums. This question has been widely studied in the literature (see [24] and also the references preceding), in particular in the case where $m = n$ and the row and column sums are 1. We call a nonnegative square matrix B *fully indecomposable* if there do not exist permutation matrices P and Q such that

$$PBQ = \begin{pmatrix} B_1 & 0 \\ Y & B_2 \end{pmatrix},$$

where B_1 and B_2 are square matrices. By convention a 1×1 matrix is fully indecomposable if and only if it is positive. The results in [20] show that there is a doubly-stochastic matrix with the same pattern of positive entries as A if and only if A is a direct sum of fully indecomposable matrices (after row and column permutations). The results concerning DAD problems in [4, 26] then follow from the above corollary. For another approach to these results, see [8].

The Dual Problem

Following the remarks after Theorem 2.9, the dual problem for (MOM) is

$$\text{(MOM*)} \qquad \begin{cases} \sup \int_S \alpha(s)\, x(s)\, ds + \int_T \beta(t)\, y(t)\, dt \\ \quad - \int_{S \times T} e^{x(s) + y(t)} k(s, t)\, ds\, dt \\ \text{subject to } x \in L_\infty(S, ds), \qquad y \in L_\infty(T, dt), \end{cases}$$

and (x, y) is optimal for (MOM*) if and only if

$$x(s) + y(t) = \log u_0(s, t), \qquad \text{a.e. } [k\, ds\, dt]$$

(where u_0 is optimal for (MOM)), in which case, as before, $f(s) := e^{x(s)}$ and $g(t) := e^{y(t)}$ solves the *DAD* problem (3.5).

Thus one approach to solving (3.5) is to solve the unconstrained concave maximization problem (MOM*). The solution of *DAD* problems is of considerable computational interest. For references, see [24, 25], where this dual approach is described (in the finite-dimensional case), and [5], for example. An early discussion of the continuous case, including algorithmic considerations, may be found in [13].

One simple computational approach is to alternate between maximizing the dual objective function over x and y respectively, keeping the other variable fixed. We could initialize for example by setting $x_0 \equiv 0$ and $y_0 \equiv 0$, and the resulting iteration is

$$x_{2n+1}(s) := \log\left(\left[\int_T e^{y_{2n}(t)} k(s, t)\, dt\right]^{-1} \alpha(s)\right),$$

$$y_{2n+1}(t) := y_{2n}(t),$$

$$x_{2n+2}(s) := x_{2n+1}(s),$$

$$y_{2n+2}(t) := \log\left[\left(\int_S e^{x_{2n+1}(s)} k(s, t)\, ds\right]^{-1} \beta(t)\right).$$

The iterates remain in $L_\infty(S) \times L_\infty(T)$ providing α and β are bounded and bounded uniformly away from 0, and if, for example, (4.3) holds.

Define $u^r(s, t) := e^{x_r(s) + y_r(t)}$. Then it is easy to see that x_r and y_r are chosen so that u^r satisfies the first marginal constraint of (MOM) for odd r and the second for even r. The alternating ascent technique is thus none other than a dual interpretation of the "iterative proportional fitting procedure" described in [13].

The Multivariate Case

Much of the original interest in *DAD* problems arose from the estimation of contingency tables (see, for example, [11]). It was observed in [13] and in [6] that many of the techniques used extend trivially to analogous questions for multivariate distributions. Rather than demonstrate this in generality, we illustrate this by an example from [6].

Suppose (S_i, ds_i), $i = 1, 2, 3, 4$, are finite measure spaces with

$$\alpha_{123} \in L_1(S_1 \times S_2 \times S_3, ds_1\, ds_2\, ds_3),$$

$$\alpha_{124} \in L_1(S_1 \times S_2 \times S_4, ds_1\, ds_2\, ds_4),$$

$$\alpha_{34} \in L_1(S_3 \times S_4, ds_3\, ds_4),$$

all strictly positive a.e., and

$$0 \leqslant k \in L_1(S_1 \times S_2 \times S_3 \times S_4, ds_1\, ds_2\, ds_3\, ds_4).$$

We ask under what conditions it is possible to find strictly positive functions

$$\begin{aligned} f_{123}&: S_1 \times S_2 \times S_3 \to \mathbb{R}, \\ f_{124}&: S_1 \times S_2 \times S_4 \to \mathbb{R}, \\ f_{34}&: S_3 \times S_4 \to \mathbb{R}, \end{aligned} \tag{3.14}$$

so that if we write

$$F(s_1, s_2, s_3, s_4) = f_{123}(s_1, s_2, s_3)\, f_{124}(s_1, s_2, s_4)\, f_{34}(s_3, s_4),$$

then $u := F$ satisfies the conditions

$$\begin{aligned} \int_{S_4} uk\, ds_4 &= \alpha_{123}, \qquad \text{a.e.}, \\ \int_{S_3} uk\, ds_3 &= \alpha_{124}, \qquad \text{a.e.}, \\ \int_{S_1 \times S_2} uk\, ds_1\, ds_2 &= \alpha_{34}, \qquad \text{a.e.} \end{aligned} \tag{3.15}$$

Following the analogous route to Corollary 3.6, we consider the problem

$$\text{(MOM')} \quad \begin{cases} \inf & \displaystyle\int_{\prod_{i=1}^4 S_i} k\phi(u)\, ds_1\, ds_2\, ds_3\, ds_4 \\ \text{subject to} & (3.15), \quad \text{and} \\ & \displaystyle 0 \leqslant u \in L_1\left(\prod_1^4 S_i, k\, ds_1\, ds_2\, ds_3\, ds_4\right). \end{cases}$$

The constraint qualification becomes

$$\text{(CQ1')} \quad \begin{cases} \text{There exists } \hat{u} > 0 \text{ a.e. } [k\, ds_1\, ds_2\, ds_3\, ds_4] \\ \text{which is feasible for (MOM')}, \end{cases}$$

and exactly as in Corollary 3.9 we deduce that we can find functions as in (3.14) solving the multivariate *DAD* problem (3.15) if and only if (CQ1') holds. The only difference in the proof is that the relevant result in [1] is now Theorem 4.4.

4. Measurability, Integrability, and Continuity

Let us recapitulate the main result of the previous section. Under Assumption 3.8, the constraint qualification (CQ1) is both necessary and sufficient for the existence of strictly positive functions $f: S \to \mathbb{R}$ and $g: T \to \mathbb{R}$ such that $u(s, t) := f(s)\, g(t)$ is feasible for (MOM) (implying f and g solve the *DAD* problem (3.5)), with finite value.

It is natural to impose further conditions on f and g, in addition to simply requiring that $f(s)\, g(t) \in L_1(S \times T, k\, ds\, dt)$, which follows from feasibility. We will consider three such conditions: measurability, integrability (see, for example, [6]), and continuity (as in [19]). We could also enquire about the uniqueness of f and g, as in [19], up to multiplication by a constant.

Such questions are considered in [1], so we begin by applying the results there. Throughout this section we suppose that Assumption 3.8 holds, and we write

$$K := \{(s, t) \in S \times T \mid k(s, t) > 0\}, \tag{4.1}$$

defined up to a set of measure zero $[ds\, dt]$.

Theorem 4.2. *Suppose K is a countable union of measurable rectangles (up to a null set $[ds\, dt]$). Then the functions f and g of Corollaries* 3.6 *and* 3.9 *are necessarily measurable. This holds if (S, ds) and (T, dt) are separable metric spaces with associated Borel measures, and K is open, which holds in particular if k is lower semicontinuous.*

Proof. By construction we have $f(s)\, g(t) = u_0(s, t)$ a.e. $[k\, ds\, dt]$, where $u_0 \in L_1(S \times T, k\, ds\, dt)$ is the optimal solution for (MOM). It follows that $f(s)\, g(t) = u_0(s, t)$ a.e. on K $[ds\, dt]$. We now apply the results of [1]. ∎

In order to deduce further properties of f and g we need to impose further conditions on the kernel k. Suppose that S and T are compact Hausdorff spaces with associated regular Borel measures, and consider the following conditions:

For some $\delta > 0$, we have:

(i) For all $\bar{s} \in S$ there exists $\bar{t} \in T$ with $k(s, t) \geqslant \delta$ a.e. $[ds\, dt]$ on a neighbourhood of $(\bar{s}, \bar{t})$,

(ii) For all $\bar{t} \in T$ there exists $\bar{s} \in S$ with $k(s, t) \geqslant \delta$ a.e. $[ds\, dt]$ on a neighbourhood of $(\bar{s}, \bar{t})$. (4.3)

The function k is lower semicontinuous with

(i) For some continuous $\pi: S \to T$,
$k(s, \pi(s)) > 0$, for all $s \in S$,

(ii) For some continuous $\rho: T \to S$,
$k(\rho(t), t) > 0$, for all $t \in T$. (4.4)

$S = T$, k is lower semicontinuous and $k(s, s) > 0$, for all $s \in S$. (4.5)

$S = T$, and for some $\delta > 0$ we have that for all $\bar{s} \in S$,
$k(s, t) \geqslant \delta$ a.e. $[ds\, dt]$ on a neighbourhood of $(\bar{s}, \bar{s})$. (4.6)

LEMMA 4.7. (4.5) $\Rightarrow$ (4.4) $\Rightarrow$ (4.3), *and* (4.6) $\Rightarrow$ (4.3).

Proof. Clearly (4.5) $\Rightarrow$ (4.4), and (4.6) $\Rightarrow$ (4.3). To see that (4.4) $\Rightarrow$ (4.3), note that $k(s, \pi(s))$ is lower-semicontinuous on the compact set S, so for some $\varepsilon_1 > 0$, $k(s, \pi(s)) \geqslant \varepsilon_1$ for all $s \in S$. Similarly, for some $\varepsilon_2 > 0$, $k(\rho(t), t) \geqslant \varepsilon_2$ for all $t \in T$. Now putting $\delta := \frac{1}{2}\min\{\varepsilon_1, \varepsilon_2\}$, $\bar{t} := \pi(s)$ in (i) and $\bar{s} = \rho(\bar{t})$ in (ii) gives (4.3) by the lower semicontinuity of k. ∎

The next result allows us to deduce in addition from Theorem 4.2 that f and g are in fact integrable.

THEOREM 4.8. *Suppose that S and T are compact Hausdorff spaces with associated regular Borel measures of full support (so nonempty open sets have positive measure in S and T respectively). Assume that $f: S \to \mathbb{R}$ and $g: T \to \mathbb{R}$ are measurable and strictly positive a.e. and that $k: S \times T \to \mathbb{R}$ is measurable and non-negative a.e., with $f(s)\, g(t)\, k(s, t) \in L_1(S \times T, ds\, dt)$. If* (4.3) *holds (or* (4.4), (4.5) *or* (4.6)), *then $f \in L_1(S, ds)$ and $g \in L_1(T, dt)$.*

Proof. By Lemma 4.7 we can assume (4.3) holds. By Fubini's theorem we have that $f(s)\, g(t)\, k(s, t) \in L_1(T, dt)$ a.e. $[ds]$, and we know $f(s) > 0$, a.e. $[ds]$, so $g(t)\, k(s, t) \in L_1(T, dt)$ a.e. $[ds]$.

By (4.3) and the assumption of full support we can associate with each $t \in T$, open sets O_t in T (containing t) and U_t in S such that O_t and U_t have positive measure and $k(s, t) \geqslant \delta$, a.e. $[ds\, dt]$ on $U_t \times O_t$. Since $\{O_t : t \in T\}$ is an open cover of T, there is a finite subcover O_{t_i}, $1 \leqslant i \leqslant n$. Since U_{t_i} has positive measure, there exists $s_i \in U_{t_i}$ with $g(t)\, k(s_i, t) \in L_1(T, dt)$. Our construction ensures that

$$\sum_{i=1}^{n} k(s_i, t) \geqslant \delta, \qquad \text{a.e. } [dt] \text{ on } T,$$

so we have

$$0 \leqslant g(t) \leqslant \left(\frac{1}{\delta}\right) \sum_{i=1}^{n} g(t)\, k(s_i, t) \in L_1(T, dt),$$

which implies that $g \in L_1(T, dt)$, as required. The argument for f follows similarly. ∎

Remark. If the measures on S and T are not of full support, there exist compact sets $S_1 \subset S$ and $T_1 \subset T$ such that the measure $[ds]$ restricted to S_1 is of full support and $[dt]$ restricted to T_1 is of full support and $S \backslash S_1$ and $T \backslash T_1$ have measure zero. If k_1 denotes the restriction of k to $S_1 \times T_1$ and if k_1 satisfies (4.3) on $S_1 \times T_1$ then the argument above still proves that $f \in L_1(S, ds)$ and $g \in L_1(T, dt)$.

COROLLARY 4.9. *Suppose (S, ds) and (T, dt) are compact metric spaces with associated regular Borel measures of full support and assume that k is lower semicontinuous and* (4.3) *holds. Then the functions f and g in Corollaries* 3.6 *and* 3.9 *are integrable.*

Proof. We can apply Theorem 4.2 to deduce that f and g are measurable. The result now follows by Theorem 4.8. ∎

In order to move one step further and show f and g are continuous, we need a further lemma.

LEMMA 4.10. *Suppose that (S, ds) and (T, dt) are compact Hausdorff spaces with associated Borel measures, that $k \in C(S \times T)$, and that $f \in L_1(S, ds)$. Define a function $F: T \to \mathbb{R}$ by*

$$F(t) := \int_S f(s)\, k(s, t)\, ds. \tag{4.11}$$

Then F is continuous.

Proof. If S and T are compact metric spaces this is immediate from the uniform continuity of k on the compact metric space $S \times T$. In general we use a standard compactness argument.

Suppose that $t_0 \in T$ and $\varepsilon > 0$. For any $s \in S$ there exist open sets O_s in S (containing s) and U_s in T (containing t_0) such that $|k(r, t) - k(s, t_0)| < \varepsilon/2$, for $r \in O_s$, $t \in U_s$. Since S is compact it has a finite subcover $\bigcup_{i=1}^n O_{s_i}$. Now for any t in $\bigcap_{i=1}^n U_{s_i}$ (an open neighbourhood of t_0) we have, for any $i = 1, \ldots, n$,

$$|k(r, t) - k(r, t_0)| \leqslant |k(r, t) - k(s_i, t_0)| + |k(r, t_0) - k(s_i, t_0)|,$$

for all $r \in S$, so picking i with $r \in O_{s_i}$ gives $|k(r, t) - k(r, t_0)| \leqslant \varepsilon$, for all $r \in S$. Finally we have, for $t \in \bigcap_{i=1}^n U_{s_i}$

$$|F(t)-F(t_0)| = \left|\int_S f(s)(k(s,t)-k(s,t_0))\,ds\right|$$
$$\leqslant \int_S |f(s)|\,|k(s,t)-k(s,t_0)|\,ds$$
$$\leqslant \varepsilon \int_S |f(s)|\,ds,$$

so F is continuous, as required. ∎

COROLLARY 4.12. *Suppose that (S, ds) and (T, dt) are compact metric spaces with associated regular Borel measures of full support, that k is continuous on $S \times T$, that α and β are continuous, and strictly positive a.e. on S and T respectively, and that* (4.3) *holds* (*or* (4.4), (4.5), *or* (4.6)). *Then the functions f and g in Corollaries* 3.6 *and* 3.9 *may be taken to be continuous* (*and strictly positive a.e.*).

Proof. Corollary 4.9 shows that f and g are integrable and we know that if we define F as in (4.11) then F is continuous, and, by feasibility,

$$g(t)\,F(t) = \beta(t), \qquad \text{a.e. } [dt].$$

We claim F is strictly positive. To see this, suppose $\bar{t} \in T$. Since we are assuming (4.3), we have $k(\bar{s}, \bar{t}) > 0$ for some $\bar{s} \in S$, so by continuity, $k(s, \bar{t}) > 0$ for all s in a neighbourhood of $\bar{s}$ (which, since ds has full support, has positive measure). But we know f is strictly positive a.e., so (4.11) shows that $F(\bar{t}) > 0$. Finally, we note $g(t) = \beta(t)/F(t)$, a.e. $[dt]$, and the righthand side is continuous, so we may take g to be continuous. The same argument works for f. ∎

We close this section by discussing some questions of uniqueness of solutions related to the problem (MOM). Assuming (MOM) is consistent, we know by Corollary 2.3 that it has an optimal solution u_0 which is unique among functions $u \in L_1(S \times T, k\,ds\,dt)$ such that $I_\phi(u) < \infty$. Furthermore, by Theorem 3.1, if p and q are measurable with

$$\begin{aligned} &0 < b_1 \leqslant p(s) \leqslant B_1, \qquad \text{a.e. } [ds] \qquad \text{and} \\ &0 < b_2 \leqslant q(t) \leqslant B_2, \qquad \text{a.e. } [dt], \end{aligned} \tag{4.13}$$

and if they solve the *DAD* problem

$$\begin{aligned} &\int_T p(s)\,k(s,t)\,q(t)\,dt = \alpha(s), \qquad \text{a.e. } [ds], \qquad \text{and} \\ &\int_S p(s)\,k(s,t)\,q(t)\,ds = \beta(t), \qquad \text{a.e. } [dt], \end{aligned} \tag{4.14}$$

(where $\alpha \geqslant 0$ a.e., $\beta \geqslant 0$ a.e., $\alpha \in L_1(S, ds)$ and $\beta \in L_1(T, dt)$), then we must have

$$p(s)\, q(t) = u_0(s, t), \qquad \text{a.e. } [k\, ds\, dt]. \tag{4.15}$$

We can now apply uniqueness results in [1, 18, 19] to solutions p and q of the *DAD* problem (4.14). Obviously, if p and q solve (4.14) and λ is a positive constant, then λp and $\lambda^{-1} q$ solve (4.14). Our aim is to prove that this is the only nonuniqueness for p and q, in other words that p and q are unique to within positive scalar multiples. We illustrate with one particular case.

Suppose that S and T are compact Hausdorff spaces with associated regular Borel measures $[ds]$ and $[dt]$, which we assume have full support. Suppose that $\alpha \in L_1(S, ds)$ and $\beta \in L_1(T, dt)$ are positive almost everywhere and that $k \in C(S \times T)$ is nonnegative and satisfies (4.3). Assume that there exist nonnegative functions $f \in L_1(S, ds)$ and $g \in L_1(T, dt)$ satisfying

$$\begin{aligned} &\int_T f(s)\, k(s, t)\, g(t)\, dt = \alpha(s), \qquad \text{a.e. } [ds], \qquad \text{and} \\ &\int_S f(s)\, k(s, t)\, g(t)\, ds = \beta(t), \qquad \text{a.e. } [dt]. \end{aligned} \tag{4.16}$$

If k, α, and β are as above and (CQ1) is satisfied, we know that (MOM) has a unique solution $u_0(s, t) = f_1(s)\, g_1(t)$, a.e. $[k\, ds\, dt]$. If we assume, in addition, that S and T are separable, (which will be true if S and T are metrizable), Theorem 4.2 implies that $f_1(s)$ and $g_1(t)$ are measurable and Theorem 4.8 implies that $f_1 \in L_1(S, ds)$ and $g_1 \in L_1(T, dt)$. Thus, if (CQ1) holds, there exist solutions f and g as in (4.16).

There is a slight subtlety, however. If f and g are as in (4.16), it is immediate from (4.16) and the assumption that α and β are positive a.e. that f and g are positive a.e. It follows from Lemma 4.10, from (4.3) and from the assumption that $[ds]$ and $[dt]$ are of full support that if

$$u(s) = \int_T k(s, t)\, g(t)\, dt \qquad \text{and} \qquad v(t) = \int_S f(s)\, k(s, t)\, ds,$$

then u and v are continuous, strictly positive functions.

It follows that $f(s) = \alpha(s)/u(s)$ and $g(t) = \beta(t)/v(t)$ a.e. In particular, if we define $w(s, t) = f(s)\, g(t)$, we can see that

$$I_\phi(w) < \infty \Leftrightarrow \iint \alpha(s)\, \beta(t) \log(\alpha(s)\, \beta(t))\, k(s, t)\, ds\, dt < \infty.$$

If, for example, $k(s, t) > 0$ for all $(s, t) \in S \times T$, it follows that $I_\phi(w) < \infty$ if and only if

$$\int_S \alpha(s) \log(\alpha(s))\, ds < \infty \qquad \text{and} \qquad \int_T \beta(t) \log(\beta(t))\, dt < \infty. \quad (4.17)$$

If we prove (as we shall) that (4.16) always has unique (to within scalar multiples) positive solutions $f \in L_1(S, ds)$ and $g \in L_1(T, dt)$ when $k(s, t) > 0$ for all $(s, t) \in S \times T$ $(k \in C(S \times T))$, it will follow from our remarks above that (CQ1) is not satisfied in the case when (4.17) fails. (If (CQ1) were satisfied, we could find $w = fg$ with $I_\phi(w) < \infty$.)

On the other hand, if $\alpha \in C(S)$ and $\beta \in C(T)$ and α and β are strictly positive on S and T respectively, we see directly that f and g are continuous and strictly positive and Theorem 3.1 implies that $u_0(s, t) = f(s)\, g(t)$ is the solution of (MOM).

THEOREM 4.18. *Suppose that S and T are compact Hausdorff spaces with regular Borel measures $[ds]$ and $[dt]$ of full support, and either S or T is connected. Suppose $\alpha \in C(S)$ and $\beta \in C(T)$ are both everywhere strictly positive and $k \in C(S \times T)$ is nonnegative and satisfies* (4.3). *If $f \in L_1(S, ds)$ and $g \in L_1(T, dt)$ are nonnegative functions which satisfy* (4.16), *then f and g are unique to within positive scalar multiples.*

Proof. We have already seen that, under the given assumptions, f and g are equal a.e. to continuous and positive functions; Theorem 3.2 implies that $f(s)\, g(t) = u_0(s, t)$ a.e. $[k\, ds\, dt]$ where $u_0(s, t)$ is the unique solution of (MOM). Thus if f_1 and g_1 also solve (4.16) then

$$f_1(s)\, g_1(t) = f(s)\, g(t), \qquad \text{a.e. } [k\, ds\, dt],$$

so, on the open set K defined in (4.1), $f_1(s)\, g_1(t) = f(s)\, g(t)$, a.e. $[ds\, dt]$. Since both sides of the above equation are continuous and ds and dt have full support, we conclude that $f_1(s)\, g_1(t) = f(s)\, g(t)$, for all $(s, t) \in K$. (Note, by the way, that if we assume from the beginning that f and g are continuous and nonnegative functions which satisfy (4.16), then the assumptions that k, α and β are continuous, $k \geqslant 0$ and $\alpha > 0$ and $\beta > 0$, imply that k satisfies (4.3).)

It follows by a simple connectedness argument using (4.3) (see [1]) that there exists a positive constant λ so $f_1(s) = \lambda f(s)$ for all s and $g_1(t) = \lambda^{-1} g(t)$ for all t as required. ∎

It has long been known (see [15], and [12] for the case $S = T = [0, 1]$) that if $(S, ds) = (T, dt)$ and $k(s, t) = k(t, s)$, a.e. $[ds\, dt]$, then there is a

solution of (4.16) with $f = g$. We now show that this is a simple consequence of uniqueness results like Theorem 4.18.

COROLLARY 4.19. *Let assumptions and notation be as in Theorem* 4.18 *and suppose in addition that* $(S, ds) = (T, dt)$, $\alpha(s) = \beta(s)$ *for all* s *and* $k(s, t) = k(t, s)$ *for all* s *and* t *in* S. *Then* (4.16) *has at most one solution* (*up to scalar multiplication*) *and we can assume* $f = g$.

Proof. Uniqueness follows by Theorem 4.18. By symmetry, if $f := f_0$ and $g := g_0$ solves (4.16), then so does $f := g_0$ and $g := f_0$. Uniqueness thus implies that there is a positive scalar λ so $g_0 = \lambda f_0$, and setting $f_1 = g_1 = \sqrt{\lambda} f_0$, we obtain a solution of (4.16) of the desired type. ∎

Uniqueness results like those above depend ultimately on using the form of the optimal solution (4.15) in conjunction with the connectedness of a certain graph associated with the set K defined in (4.1). These techniques can be applied in both the discrete, matrix case, and as above using topological considerations. In Part II we shall adopt a very different approach to the uniqueness question, which also yields further iterative techniques for approximating the solutions f and g of (4.16).

II. A COMBINED FIXED POINT AND VARIATIONAL APPROACH TO INFINITE *DAD* PROBLEMS

5. *Constraint Qualification and Solutions of the DAD Problem*

The key assumption in characterizing solutions to (MOM) was the constraint qualification (CQ1). If (CQ1) is satisfied, Corollary 3.6 implies that there exists a solution to the *DAD* problem in equation (3.5). However, it may not be trivial to verify (CQ1). In this section we want to show how our previous results can be combined with ideas from [18, 15, 12] to prove a generalization of a *DAD* theorem in [19]. The proof we shall give is different from that in [19], and indeed, the argument in [19] does not directly extend to the situation we shall consider. One advantage of the explicit fixed point approach we use in this section is that it produces an iterative technique for solving *DAD* problems which is guaranteed to converge geometrically in great generality.

We begin first with some simpler cases.

THEOREM 5.1. *Suppose that* (S, ds) *and* (T, dt) *are finite measure spaces and that* $\alpha \in L_1(S, ds)$ *is strictly positive a.e.* $[ds]$ *and* $\beta \in L_1(T, dt)$ *is strictly positive a.e.* $[dt]$. *Assume that* $k \in L_1(S \times T, ds\, dt)$ *and that there exists a*

positive scalar δ such that $k(s, t) \geqslant \delta > 0$ a.e. $[ds\, dt]$. If α and β have finite entropy, so

$$\int_S \alpha \log \alpha \, ds < \infty \qquad \text{and} \qquad \int_T \beta \log \beta \, dt < \infty,$$

and if

$$\int_S \alpha(s)\, ds = \int_T \beta(t)\, dt,$$

then the DAD problem (3.5) *has a solution, and the functions f and g in* (3.5) *are measurable and strictly positive a.e.*

Proof. Define $c = \int_S \alpha(s)\, ds > 0$. By the results of Section 3 and Section 4, it suffices to prove that the constraint qualification (CQ1) is satisfied. However, if we define

$$u(s, t) = \frac{\alpha(s)\, \beta(t)}{ck(s, t)},$$

one can easily verify that u is feasible for (MOM). The only point where care is needed is in showing that

$$\int_{S \times T} \phi(u(s, t))\, k(s, t)\, ds\, dt < \infty,$$

where ϕ is given by (2.1), and this follows from the assumption that α and β have finite entropy. ∎

There is a variant of Theorem 5.1 which is actually closer to the questions we shall be considering later in this section.

THEOREM 5.2. *Suppose that (S, ds) and (T, dt) are finite measure spaces and that $\alpha \in L_1(S, ds)$ and $\beta \in L_1(T, dt)$ are positive almost everywhere and $\int \alpha\, ds = \int \beta\, dt$. Assume that $k \in L_\infty(S \times T, ds\, dt)$ and that there exists $\delta > 0$ such that $k(s, t) \geqslant \delta$, a.e. $[ds\, dt]$. Let C_S° (respectively, C_T°) denote the interior of the cone of nonnegative functions in $L_\infty(S)$ (respectively, $L_\infty(T)$). Then there exist $v \in C_S^\circ$ and $w \in C_T^\circ$ such that setting $f := \alpha v$ and $g := \beta w$ satisfies the DAD problem* (3.5). *Furthermore, if $v_1 \in C_S^\circ$, $w_1 \in C_T^\circ$ and $f_1 = \alpha v_1$ and $g_1 = \beta w_1$ also satisfy the DAD problem* (3.5), *then there exists a positive scalar λ such that $v_1 = \lambda v$ and $w_1 = \lambda^{-1} w$.*

Proof. Define $\tilde{k}(s, t) = \alpha(s)\, k(s, t)\, \beta(t)$ and $u(s, t) = (c\tilde{k}(s, t))^{-1}$, where $c = \int \alpha(s)\, ds$. If $\tilde{k}(s, t)$ replaces $k(s, t)$ in problem (MOM), $u(s, t)$ is feasible for (MOM) and $u(s, t) > 0$, a.e. $[ds\, dt]$. It follows from Corollary 3.6 that

there exist functions $v(s)$ and $w(t)$ which are positive almost everywhere such that setting $f := v$ and $g := w$ solves the *DAD* problem (3.5), and Theorem 4.2 implies that v and w are measurable. Equations (3.5) imply that

$$v(s) \int_T k(s, t)\, \beta(t)\, w(t)\, dt = 1, \qquad \text{a.e. } [ds],$$

and since there exist positive constants δ and M such that $\delta \leqslant k(s, t) \leqslant M$, a.e. $[ds\, dt]$, and $v(s) > 0$, a.e. $[ds]$, we conclude that $\beta w \in L_1(T)$. It follows that

$$\left(\frac{1}{M \int \beta(t)\, w(t)\, dt}\right) \leqslant v(s) \leqslant \left(\frac{1}{\delta \int \beta(t)\, w(t)\, dt}\right), \qquad \text{a.e. } [ds],$$

so $v \in L_\infty(S)$. A similar argument shows that $w \in L_\infty(T)$. Theorem 3.1 now implies (since k is positive a.e.) that $v(s)\, w(t) = u_0(s, t)$, a.e. $[ds\, dt]$, where u_0 is the unique solution of (MOM). If v_1 and w_1 are as in the statement of the theorem, the same argument shows that

$$v(s)\, w(t) = u_0(s, t) = v_1(s)\, w_1(t), \qquad \text{a.e. } [ds\, dt].$$

By Fubini's theorem, there exists s_0 such that $v(s_0)\, w(t) = v_1(s_0)\, w_1(t)$, a.e. $[dt]$, which implies that $w(t)/w_1(t) = \lambda = v_1(s_0)/v(s_0)$, a.e. $[dt]$. The same argument shows that $v_1(s)/v(s) = \lambda$, a.e. $[ds]$. ∎

We remind the reader that we have already discussed in Section 3 the classical case of an $n \times n$ nonnegative matrix with $\alpha \equiv \beta \equiv 1$. The more general case, where $S = \{1, 2, ..., m\}$, $T = \{1, 2, ..., n\}$ and α and β are arbitrary nonnegative vectors, is discussed in [14].

We now want to discuss some *DAD* problems which generalize results in Section 4 of [19]. We begin with a lemma related to Theorem 2 in [12].

THEOREM 5.3. *Let (S, σ) be a finite measure space with $\alpha \in L_1(S, \sigma)$, $\alpha(s) > 0$ a.e. $[\sigma]$, $k \in L_\infty(S \times S, \sigma \times \sigma)$ and $k(s, t) \geqslant \delta > 0$ a.e. $[\sigma \times \sigma]$, where δ is a scalar. Let ρ be a scalar with $0 < \rho \leqslant 1$ and let C° denote the interior of the cone of nonnegative functions in $L_\infty(S)$. Define a map $\Phi: C^\circ \to C^\circ$ by*

$$(\Phi(v))(s) = \int k(s, t)\, \alpha(t) \left(\frac{1}{v(t)^\rho}\right) d\sigma(t).$$

Then there is a unique $u \in C^\circ$ such that $\Phi(u) = u$. If $0 < \rho < 1$, and $x \in C^\circ$, then

$$\lim_{j \to \infty} \Phi^j(x) = u.$$

If $\rho = 1$ and $x \in C^\circ$, there is a positive scalar $\lambda(x)$ such that

$$\lim_{j \to \infty} \Phi^{2j}(x) = \lambda(x)u \qquad \text{and} \qquad \lim_{j \to \infty} \Phi^{2j+1}(x) = \lambda(x)^{-1} u.$$

The map $x \to \lambda(x)$ is real analytic on C°.

Proof. We shall use Hilbert's projective metric d and Thompson's metric $\bar{d}$ on C°; see [18, p. 13] for definitions. It is an elementary exercise (see Proposition 1.5, p. 19, in [18]) that if $J(x) = x^{-\rho}$ for $x \in C^\circ$, then

$$d(Jx, Jy) \leqslant \rho d(x, y) \qquad \text{and} \qquad \bar{d}(Jx, Jy) \leqslant \rho \bar{d}(x, y) \qquad \text{for all} \quad x, y \in C^\circ.$$

If we define a linear operator $A: L_\infty(S) \to L_\infty(S)$ by

$$(Ax)(s) = \int k(s, t)\, \alpha(t)\, x(t)\, d\sigma(t),$$

then note that A is a bounded linear operator (with $\|A\| \leqslant \|k\|_\infty \|\alpha\|_1$) and that $A(C^\circ) \subset C^\circ$. It follows (see Proposition 1.5, p. 19, in [18]) that A is nonexpansive with respect to d or $\bar{d}$. If $0 < \rho < 1$, it follows that $\Phi = AJ$ is a contraction mapping of C° into itself with contraction constant $\rho < 1$. A result of [27] implies that $(C^\circ, \bar{d})$ is a complete metric space and that $\bar{d}$ gives the same topology on C° as does the norm on $L_\infty(S)$. Thus the contraction mapping principle implies Theorem 5.1 when $0 < \rho < 1$.

It remains to consider the case $\rho = 1$. If B is a bounded linear operator such that $B(C^\circ) \subset C^\circ$, define $\Delta(B)$ by

$$\Delta(B) = \sup\{d(Bx, By): x, y \in C^\circ\}. \tag{5.4}$$

If $\Delta(B) < \infty$, an old result of Birkhoff implies that $d(Bx, By) \leqslant cd(x, y)$ for all $x, y \in C^\circ$, where $c = \tanh(\Delta(B)/4) < 1$. See [18, p. 43] for references to the literature. Define maps $\Psi: C^\circ \to \mathbb{R}^+$ and $\Phi_1: C^\circ \to C^\circ$ by

$$\Psi(y) = \int_S y(s)\, ds \qquad \text{and} \qquad \Phi_1(x) = \frac{\Phi(x)}{\Psi(\Phi(x))}.$$

If we assume that $\Delta(A) < \infty$, our remarks above show that for all $x, y \in C^\circ$ we have $d(\Phi(x), \Phi(y)) = d(\Phi_1(x), \Phi_1(y)) \leqslant cd(x, y)$, where the constant $c = \tanh(\Delta(A)/4) < 1$. If $\Sigma = \{x \in C^\circ: \Psi(x) = 1\}$, it is well-known that (Σ, d) is a complete metric space: see Theorem 1.2 in [18] and the references to the literature there. Thus (assuming $\Delta(A) < \infty$) Φ_1 is a Lipschitz mapping of (Σ, d) into itself and Φ_1 has Lipschitz constant $c < 1$, so the contraction mapping principle implies that Φ_1 has a unique fixed point ξ in Σ. It follows that $\Phi(\xi) = \lambda(\xi)$, where $\lambda = \Psi(\Phi(\xi))$. If we take $u = \mu\xi$, where $\lambda = \mu^2$, a calculation (using the homogeneity of Φ) gives $\Phi(u) = u$.

Conversely, if $\Phi(u) = u$, a calculation implies that $u/\Psi(u)$ is a fixed point of Φ, so $u/\Psi(u) = \xi$, $(\Psi(u))^2 = \Psi(\Phi(\xi))$, and u is unique.

If we define $f = \Phi^2$, $f: C^\circ \to C^\circ$ is homogeneous of degree one, order-preserving and real analytic. If u is the fixed point of Φ and $L = f'(u)$, the chain rule implies that $L = (-\Phi'(u))^2$, so $\Delta(L) \leqslant \Delta(-\Phi'(u)) = \Delta(A)$. If we prove that $\Delta(A) < \infty$, it follows that $\Delta(L) < \infty$, and Remark 2.4 on p. 44 of [18] implies that the essential spectral radius of L (see [16] for definitions and references to the literature) is strictly less than the spectral radius of L. The final statement of Theorem 5.3 thus follows from Theorem 3.2 on p. 93 in [18].

It remains only to prove that $\Delta(A) < \infty$. This result is known (see [16] for references to the literature), but we sketch a proof for completeness. Let δ and M be positive numbers such that $\delta \leqslant k(s, t) \leqslant M$, a.e. $[\sigma \times \sigma]$. Fubini's theorem implies that there is a set $S_1 \subset S$ such that the measure of the complement of S_1 is zero and if $s \in S_1$, $\delta \leqslant k(s, t) \leqslant M$, a.e. $[d\sigma(t)]$. If $s_1, s_2 \in S_1$ and if we define $\mu = M\delta^{-1}$, it follows that for any $x \in C^\circ$,

$$(Ax)(s_1) = \int k(s_1, t)\,\alpha(t)\,x(t)\,d\sigma(t) \leqslant \mu \int k(s_2, t)\,\alpha(t)\,x(t)\,d\sigma(t) = (\mu Ax)(s_2).$$

If e denotes the function which is identically one, it follows that

$$d(Ax, e) \leqslant \log(\mu),$$

which implies that $\Delta(A) \leqslant 2\log(\mu)$. ∎

Remark 5.5. If $k(s, t) = k(t, s)$ for all s and t and $\rho = 1$, the existence and uniqueness of the fixed point u can be obtained from Theorem 5.2, but the convergence of $\Phi^{2j}(x)$ requires other ideas.

Remark 5.6. In [12] it is assumed that $S = [0, 1]$, $\alpha \equiv 1$ and k is continuous. In that case, the linear map A is compact. Karlin and Nirenberg use the compactness of A and the Schauder fixed point theorem to prove the existence of a fixed point of $\Phi = AJ$ in C°, even for $\rho > 1$. (Uniqueness may fail even in the matrix case for ρ large). In the generality of Theorem 5.3 such a proof is not directly available, because A may not be compact. To see this, take $S = [0, 1]$, $\alpha \equiv 1$ and σ the usual Lebesgue measure. For $j \geqslant 1$, j an integer, define $I_j = [1/2^j, 1/2^{j-1}]$. Define $k(s, t)$ by

$$k(s, t) = 2 + (-1)^m \quad \text{if } s \in I_j \qquad \text{and} \qquad \frac{m}{2^j} \leqslant t < \frac{m+1}{2^j} \quad \text{for } 0 \leqslant m < 2^j.$$

Define $v_j \in L_\infty(S)$ by

$$v_j(t) = (-1)^m \qquad \text{if } \frac{m}{2^j} \leqslant t < \frac{m+1}{2^j}, \qquad 0 \leqslant m < 2^j.$$

A calculation yields

$$(Av_j)(s) = \begin{cases} 1, & \text{if } s \in I_j, \\ 0, & \text{if } s \notin I_j. \end{cases}$$

It follows that

$$\|Av_j - Av_m\|_\infty = 1 \qquad \text{for } j \neq m,$$

so $\{Av_j : j \geq 1\}$ is not precompact and A is not compact.

We want to extend Theorem 5.2 to a case in which k is nonnegative and positive on a neighbourhood of the diagonal of $S \times S$. We shall need a lemma first.

LEMMA 5.7. *Let S and T be compact metric spaces with finite, regular Borel measures ds and dt respectively. Assume that $k_1 \in L_1(S \times T, ds\, dt)$ and define a map $A: L_\infty(T) \to L_1(S)$ by*

$$(Ax)(s) = \int_T k_1(s, t)\, x(t)\, dt.$$

Then A is a compact linear operator.

Proof. It is immediate that A is a bounded linear operator with

$$\|A\| \leq \int_S \int_T |k_1(s, t)|\, ds\, dt,$$

because if $x \in L_\infty(T)$, we have

$$\int_S |(Ax)(s)|\, ds \leq \int_S \int_T |k_1(s, t)|\, |x(t)|\, dt\, ds \leq \|x\|_\infty \int_S \int_T |k_1(s, t)|\, dt\, ds.$$

Recall that if X and Y are any Banach spaces and A_m is a compact linear map from X to Y for $m \geq 1$, and if $\|A_m - A\| \to 0$, where A is a bounded linear map, then A is a compact linear map. Thus in our case it suffices to find a sequence of compact linear maps A_m from $L_\infty(T)$ to $L_1(S)$ which converge to A in norm. Since $ds\, dt$ is a finite regular Borel measure on the compact Hausdorff space $S \times T$, it is known that there exists a sequence $c_m(s, t)$, $m \geq 1$, of continuous, real-valued maps with domain $S \times T$ such that

$$\lim_{m \to \infty} \iint |c_m(s, t) - k_1(s, t)|\, ds\, dt = 0.$$

If we define $A_m: L_\infty(T) \to L_1(S)$ by $(A_m x)(s) = \int c_m(s, t)\, x(t)\, dt$, the first part of the lemma shows that $\|A - A_m\| \leqslant \|c_m - k_1\|_1$, which approaches zero as m approaches ∞.

Thus it suffices to prove that A_m is a compact, linear operator. However, A_m can be considered a bounded, linear map from $L_\infty(T)$ to $C(S)$, and $C(S)$ is continuously imbedded in $L_1(S)$. Thus it suffices to prove that A_m is compact as a map from $L_\infty(T)$ to $C(S)$, and by the Ascoli–Arzela theorem, this will be true if $\{A_m v: v \in L_\infty(T),\ \|v\| \leqslant 1\}$ is equicontinuous. However, equicontinuity follows easily from the fact that c_m is uniformly continuous on $S \times T$. ∎

Our next theorem and remark are generalizations of results of Karlin and Nirenberg [12].

THEOREM 5.8. *Let S be a compact metric space with a finite, regular Borel measure τ of full support. Suppose that $\alpha \in L_1(S, \tau)$ is positive a.e. $[\tau]$ and $k \in L_\infty(S \times S, \tau \times \tau)$ is nonnegative a.e. $[\tau \times \tau]$. Assume that for every $s \in S$ there exists an open set G_s containing s and a positive constant δ_s such that $k(r, t) \geqslant \delta_s$ for almost all $(r, t) \in G_s \times G_s$. Let ρ be a scalar with $0 < \rho \leqslant 1$ and let C° denote the interior of the cone of nonnegative functions in $L_\infty(S)$. Then there exists $u \in C^\circ$ such that*

$$\int_S k(s, t)\, \alpha(t)\, \frac{1}{u(t)^\rho}\, d\tau(t) = u(s), \qquad \text{a.e. } [\tau].$$

Proof. Let Φ, A and J be as defined in the proof of Theorem 5.3. If $0 < \rho < 1$, exactly the proof in Theorem 5.3 shows (assuming only that $A(C^\circ) \subset C^\circ$) that Φ has a unique fixed point u in C° and that $\lim_{j \to \infty} \Phi^j(x) = u$, for all $x \in C^\circ$.

Thus we shall assume $\rho = 1$. As in [12], for each ε with $0 < \varepsilon \leqslant 1$ define $k_\varepsilon(s, t) = k(s, t) + \varepsilon$. Theorem 5.3 implies that there exists a unique $u_\varepsilon \in C^\circ$ such that

$$\int k_\varepsilon(s, t)\, \alpha(t)\, \frac{1}{u_\varepsilon(t)}\, d\tau(t) = u_\varepsilon(s), \qquad \text{a.e. } [\tau].$$

The problem is to take the limit as $\varepsilon \to 0$, and we proceed initially as in [12]. The defining equation for u_ε gives

$$\iint k_\varepsilon(s, t)\, \frac{\alpha(t)}{u_\varepsilon(t)}\, \frac{\alpha(s)}{u_\varepsilon(s)^2}\, d\tau(t)\, d\tau(s) = \int \frac{\alpha(s)}{u_\varepsilon(s)}\, d\tau(s).$$

If G_s is as in the statement of our theorem, we can use the compactness of S to find a finite open covering G_{s_i}, $i = 1, \ldots, m$ and we can then assume that

every G_s equals G_{s_i} for some i, $1 \leqslant i \leqslant m$. We also have that $k(r, t) \geqslant \delta = \min\{\delta_{s_i} : 1 \leqslant i \leqslant m\}$ for all $r, t \in G_{s_i}$, $1 \leqslant i \leqslant m$.

With these conventions we find

$$\delta\left(\int_{G_s} \frac{\alpha(t)}{u_\varepsilon(t)}\, d\tau(t)\right)\left(\int_{G_s} \frac{\alpha(\sigma)}{u_\varepsilon(\sigma)^2}\, d\tau(\sigma)\right) \leqslant \int_S \frac{\alpha(\sigma)}{u_\varepsilon(\sigma)^2}\, d\tau(\sigma).$$

An application of Hölder's inequality gives

$$\int_{G_s} \frac{\alpha}{u_\varepsilon} \leqslant \left(\int_{G_s} \alpha\right)^{1/2} \left(\int_{G_s} \frac{\alpha}{u_\varepsilon^2}\right)^{1/2}.$$

If we define $v(s) = \int_{G_s} \alpha$, we obtain that

$$\int_{G_s} \frac{\alpha}{u_\varepsilon^2} \geqslant \frac{1}{v(s)} \left(\int_{G_s} \frac{\alpha}{u_\varepsilon}\right)^2.$$

It follows that

$$\left(\int_{G_s} \frac{\alpha}{u_\varepsilon}\right)^3 \leqslant \frac{v(s)}{\delta} \int_S \frac{\alpha}{u_\varepsilon} \leqslant c_1 \int_S \frac{\alpha}{u_\varepsilon},$$

where $c_1 = \|\alpha\|_1\, \delta^{-1}$ is independent of ε, for $0 < \varepsilon \leqslant 1$.

For a given ε select i so that

$$\int_{G_{s_i}} \frac{\alpha}{u_\varepsilon} = \max_{1 \leqslant j \leqslant m} \int_{G_{s_j}} \frac{\alpha}{u_\varepsilon}.$$

It follows that

$$\left(\int_{G_{s_i}} \frac{\alpha}{u_\varepsilon}\right)^3 \leqslant m c_1 \int_{G_{s_i}} \frac{\alpha}{u_\varepsilon}.$$

The above inequality implies that there is a constant c_2, independent of ε, such that

$$\int_{G_{s_i}} \frac{\alpha}{u_\varepsilon} \leqslant c_2 \qquad \text{and so} \qquad \int_S \frac{\alpha}{u_\varepsilon} \leqslant m \int_{G_{s_i}} \frac{\alpha}{u_\varepsilon} \leqslant m c_2 = c_3.$$

If we define $M = \|k\|_\infty$, we conclude that

$$u_\varepsilon(s) = \int_S k_\varepsilon(s, t) \frac{\alpha(t)}{u_\varepsilon(t)}\, d\tau(t) \leqslant M c_3 = c_4, \qquad \text{a.e. } [\tau].$$

It follows that, defining $\delta_2 = \min_{1 \leqslant i \leqslant m} \int_{G_{s_i}} \alpha > 0$, we have almost everywhere

$$u_\varepsilon(s) \geqslant \int_{G_s} k_\varepsilon(s, t) \frac{\alpha(t)}{c_4} d\tau(t) \geqslant \delta c_4^{-1} \int_{G_s} \alpha(t)\, d\tau(t) \geqslant \delta \delta_2 c_4^{-1} = c_5 > 0.$$

Notice that c_4 and c_5 are independent of ε, for $0 < \varepsilon \leqslant 1$.

Now let ε_j be a sequence of positive numbers approaching 0. Our previous remarks show that $J(u_{\varepsilon_j})$ is bounded in $L_\infty(S)$ for $j \geqslant 1$, so Lemma 5.7 implies that a subsequence of $AJu_{\varepsilon_j} = u_{\varepsilon_j}$ converges in $L_1(S)$ to some element $u \in L_1(S)$. By relabeling, we can assume that $\lim_{j \to \infty} \|u_{\varepsilon_j} - u\|_1 = 0$, so by taking a further subsequence, we can also assume $\lim_{j \to \infty} u_{\varepsilon_j}(s) = u(s)$, a.e. $[\tau]$.

Since $0 < c_5 \leqslant u_{\varepsilon_j}(s) \leqslant c_4$, a.e. $[\tau]$, we also know that $c_5 \leqslant u(s) \leqslant c_4$, a.e. $[\tau]$. It follows that $\lim_{j \to \infty} (u_{\varepsilon_j}(s))^{-1} = u(s)^{-1}$, a.e. $[\tau]$, and $(u_{\varepsilon_j}(s))^{-1} \leqslant c_5^{-1}$, for all j, a.e. $[\tau]$. Using these facts, it is now easy to see by Lebesgue dominated convergence that, almost everywhere,

$$u(s) = \lim_{j \to \infty} u_{\varepsilon_j}(s) = \lim_{j \to \infty} \int_S k_{\varepsilon_j}(s, t) \frac{\alpha(t)}{u_{\varepsilon_j}(t)} d\tau(t) = \int_S k(s, t) \frac{\alpha(t)}{u(t)} d\tau(t). \quad \blacksquare$$

Remark 5.9. As noted in the proof of Lemma 5.7, there exists a sequence of continuous functions $c_m(s, t)$, which approach $k(s, t)$ in L_1 norm. If $\rho > 1$ and if c_m is defined with some care, the kind of argument in [12] can be applied to prove the existence in C° of u_m such that

$$\int c_m(s, t)\, \alpha(t) \frac{1}{u_m(t)^\rho} d\tau(t) = u_m(s), \qquad \text{a.e. } [\tau].$$

Then, by using the kind of argument as in Theorem 5.8, one can prove that there are positive constants b_1 and b_2 such that $b_1 \leqslant u_m(t) \leqslant b_2$, a.e. $[\tau]$, and some subsequence u_{m_i} converges almost everywhere to $u \in C^\circ$, where

$$\int k(s, t)\, \alpha(t) \frac{1}{u(t)^\rho} d\tau(t) = u(s), \qquad \text{a.e. } [\tau].$$

Thus Theorem 5.8 remains true for $\rho > 1$.

To proceed further we shall need a purely linear result. For any given k in $L_\infty(S \times T)$, α in $L_1(S)$ and β in $L_1(T)$ we can define $A_1 : L_\infty(S) \to L_\infty(T)$ and $A_2 : L_\infty(T) \to L_\infty(S)$ by

$$(A_1 w)(t) = \int_S k(s, t)\, \alpha(s)\, w(s)\, ds \qquad \text{and}$$

$$(A_2 v)(s) = \int_T k(s, t)\, \beta(t)\, v(t)\, dt,$$

and then $B = A_2 A_1$ is given by

$$(Bw)(r) = \int_S c_1(r, s)\, \alpha(s)\, w(s)\, ds, \tag{5.10}$$

where $c_1(r, s) = \int_T k(r, t)\, k(s, t)\, \beta(t)\, dt$. The following condition will be useful.

Assumption 5.11. There exists an integer $m \geqslant 1$ such that for any two points r and s in S there exist points $s_i \in S$ for $0 \leqslant i \leqslant m$, with $s_0 = r$ and $s_m = s$, open neighbourhoods G_i of s_i, for $0 \leqslant i \leqslant m$, and a positive constant $\delta = \delta(r, s)$ with $c_1(s_i, s_{i+1}) \geqslant \delta$ almost everywhere on $G_i \times G_{i+1}$ for $0 \leqslant i \leqslant m$.

LEMMA 5.12. *Let S and T be compact Hausdorff spaces and suppose that ds and dt are regular Borel measures of full support on S and T, respectively. Suppose that $k \in L_\infty(S \times T, ds\, dt)$ and that $k(s, t) \geqslant 0$, a.e. $[ds\, dt]$. Assume that $\alpha \in L_1(S, ds)$ and $\beta \in L_1(T, dt)$ are both positive almost everywhere, and that Assumption* 5.11 *holds. Then if C° denotes the interior of the cone of nonnegative functions in $L_\infty(S)$, d denotes Hilbert's projective metric on C°, and B is given by* (5.10), *we have*

$$\Delta(B^m) \equiv \sup\{d(B^m x, B^m y) : x, y \in C^\circ\} < \infty.$$

Furthermore, we have that $B^m(C \setminus \{0\}) \subset C^\circ$.

If $\bar{r} = \bar{r}(B)$ is the spectral radius of B and $\bar{\rho} = \bar{\rho}(B)$ is the essential spectral radius of B, then $\bar{\rho} < \bar{r}$. Furthermore, there exist $u \in C^\circ$ and $u^* \in (L_\infty(S))^*$ with $Bu = \bar{r}u$, $B^* u^* = \bar{r}u^*$ and $u^*(u) > 0$. The map $\bar{r}I - B$ is Fredholm of index zero and the eigenvalue $\bar{r}$ is isolated in the spectrum of B and has algebraic multiplicity one.

Proof. Lemma 5.12 will follow immediately from Theorem 2.4 and Remark 2.4 on p. 44 in [18] if we prove that $B^m(C \setminus \{0\}) \subset C^\circ$ and $\Delta(B^m) < \infty$. (The assertion that $\bar{r}I - B$ is Fredholm of index zero follows from Remark 2.4 in [18], because Remark 2.4 implies that $\bar{r}I - B$ differs from a linear homeomorphism by a compact linear map.)

It is well-known that for $j \geqslant 1$

$$(B^j w)(r) = \int_S c_j(r, s)\, \alpha(s)\, w(s)\, ds, \qquad \text{where}$$

$$c_j(r, s) = \int_S c_1(r, s_1)\, c_{j-1}(s_1, s)\, ds_1$$

$$= \int_S \int_S \cdots \int_S c_1(r, s_1)\, c_1(s_1, s_2) \cdots c_1(s_{j-1}, s)\, ds_1\, ds_2 \cdots ds_{j-1}.$$

It follows from Assumption 5.11, the above formula for c_m, and the fact that ds has full support that for each r and s in S there exist open neighbourhoods $U = U_{r,s}$ and $V = V_{r,s}$ of r and s, respectively, and $\eta = \eta(r, s) > 0$ such that $c_m(\rho, \sigma) \geqslant \eta > 0$ for almost all $(\rho, \sigma) \in U \times V$.

Now $S \times S$ is a compact Hausdorff space with open cover

$$\{U_{r,s} \times V_{r,s} : (r, s) \in S \times S\},$$

so there exists a finite subcovering $\{U_{r_i,s_i} \times V_{r_i,s_i} : 1 \leqslant i \leqslant n\}$ and

$$c_m(\rho, \sigma) \geqslant \eta_i \equiv \eta(r_i, s_i) > 0,$$

for almost all $(\rho, \sigma) \in U_{r_i,s_i} \times V_{r_i,s_i}$, for $1 \leqslant i \leqslant n$. If $\eta = \min\{\eta_i : 1 \leqslant i \leqslant n\}$, we conclude that $c_m(\rho, \sigma) \geqslant \eta$, a.e. $[ds\, dt]$.

On the other hand, it is clear that there exists a constant M so $c_m(\rho, \sigma) \leqslant M$, a.e. $[ds\, dt]$. If $w \in C \setminus \{0\}$, it follows that, almost everywhere,

$$(B^m w)(r) = \int c_m(r, s)\, \alpha(s)\, w(s)\, ds \geqslant \eta \int \alpha w \qquad \text{and}$$

$$(B^m w)(r) \leqslant M \int \alpha w.$$

This shows that $B^m w \in C^\circ$ and, letting e denote the function which is identically equal to one,

$$d(B^m w_1, B^m w_2) \leqslant d(B^m w_1, e) + d(e, B^m w_2) \leqslant 2 \log(M/\eta),$$

so $\Delta(B^m) \leqslant 2 \log(M/\eta)$. ∎

LEMMA 5.13. *Let notation and assumptions be as in Lemma* 5.12 *except for Assumption* 5.11. *Suppose that S is connected and metrizable. Assume that for every $s \in S$ there exists an open neighbourhood U_s of s, a nonempty*

open set $V_s \subset T$ and a positive constant δ_s such that $k(\sigma, \zeta) \geq \delta_s$ for almost all $(\sigma, \zeta) \in U_s \times V_s$. Then there exists an integer $m \geq 1$ for which Assumption 5.11 *is satisfied, and the operator B satisfies all the conclusions of Lemma* 5.12.

Proof. It suffices to prove that Assumption 5.11 is satisfied. By compactness of S we can cover S by finitely many open sets U_{s_i}, $1 \leq i \leq n$, with corresponding nonempty open sets V_{s_i}, $1 \leq i \leq n$. By assumption we have

$$k(\sigma, \zeta) \geq \delta = \min\{\delta_{s_i} : 1 \leq i \leq n\},$$

for almost all $(\sigma, \zeta) \in U_{s_i} \times V_{s_i}$, $1 \leq i \leq n$. If τ is the measure on T and if we define $\kappa = \min\{\tau(V_{s_i}) : 1 \leq i \leq n\}$, then we find that for almost all $(\sigma_1, \sigma_2) \in U_{s_i} \times U_{s_i}$, $1 \leq i \leq n$, we have

$$c_1(\sigma_1, \sigma_2) = \int_T k(\sigma_1, t)\, k(\sigma_2, t)\, dt \geq \int_{V_{s_i}} k(\sigma_1, t)\, k(\sigma_2, t)\, dt \geq \delta^2 \tau(V_{s_i}) \geq \delta^2 \kappa.$$

Let γ denote the metric on S and for $\rho > 0$ and $s \in S$, let $B_\rho(s)$ denote the open ball of radius ρ and center s. Let ρ_0 denote a "Lebesgue number" of the open cover $\{U_{s_i} : 1 \leq i \leq n\}$, so for any $s \in S$, $B_{\rho_0}(s) \subset U_i$ for some i. Select a fixed number $\rho > 0$ so $2\rho < \rho_0$. Because S is connected and compact, there exists an integer m such that for any r and s in S there exist points $s_i \in S$, for $0 \leq i \leq m$ such that $s_0 = r$, $s_m = s$ and $\gamma(s_i, s_{i+1}) < \rho$ for $0 \leq i \leq m$ (see [28]). For s_i as above, define $G_i = B_\rho(s_i)$. If j is chosen so that $B_{2\rho}(s_i) \subset U_j$, note that

$$G_i \times G_{i+1} \subset B_{2\rho}(s_i) \times B_{2\rho}(s_i) \subset U_j \times U_j,$$

so $c_i(\sigma_1, \sigma_2) \geq \delta^2 \kappa$, almost everywhere on $G_i \times G_{i+1}$.

This proves that Assumption 5.11 holds. ∎

With the aid of Assumption 5.11 and Lemma 5.13 we can generalize earlier results of this section. Under Assumption 5.11 or the conditions of Lemma 5.13 we can prove that *if the DAD problem* (3.5) *has a solution*, it is unique (to within scalar multiples) and the function f in (3.5) can be obtained by an iterative procedure which converges geometrically.

We adopt the following notation. Let A_1, A_2, and c_1 be defined as before (immediately before Assumption 5.11). C_S denotes the cone of nonnegative functions in $L_\infty(S)$ and C_T the cone of nonnegative functions in $L_\infty(T)$, with interiors C_S° and C_T°, respectively. Define $J_S : C_S^\circ \to C_S^\circ$ and $J_T : C_2^\circ \to C_T^\circ$ by

$$(J_S x)(s) = \frac{1}{x(s)} \qquad \text{and} \qquad (J_T y)(t) = \frac{1}{y(t)},$$

and define $\psi(x) = \int_S x(s)\,ds$ for $x \in L_1(S, ds)$. Define $F: C_S^\circ \to C_S^\circ$ and $G: C_S^\circ \to C_S^\circ$ by

$$F = A_2 J_T A_1 J_S \qquad \text{and} \qquad G(x) = \frac{F(x)}{\psi(F(x))}.$$

THEOREM 5.14. *Let S and T be compact Hausdorff spaces and suppose that ds and dt are finite regular Borel measures of full support on S and T respectively. Suppose that $\alpha \in L_1(S, ds)$ is positive, a.e. $[ds]$, $\beta \in L_1(T, dt)$ is positive a.e. $[dt]$ and $\int_S \alpha(s)\,ds = \int_T \beta(t)\,dt$. Assume that $k \in L_\infty(S \times T, ds\,dt)$, $k(s, t) \geqslant 0$, a.e. $[ds\,dt]$ and k satisfies condition (4.3). Then $A_1(C_S^\circ) \subset C_T^\circ$ and $A_2(C_T^\circ) \subset C_S^\circ$. If S is connected and metrizable, S automatically satisfies Assumption 5.11; otherwise, assume it holds.*

With these assumptions the DAD problem (3.5) possesses solutions $f \in L_1(S)$, and $g \in L_1(T)$ with f and g strictly positive almost everywhere if and only if F has an eigenvector $x_0 \in C_S^\circ$. Such an eigenvector x_0 is necessarily a fixed point of F. If $F(x_0) = x_0$ for some $x_0 \in C_S^\circ$ and $y_0 = A_1 J_S x_0$, then $f = \alpha / x_0$ and $g = \beta / y_0$ are L_1 functions which are strictly positive almost everywhere and solve the DAD problem (3.5). Such solutions f and g of (3.5) are (if they exist) unique to within scalar multiples; a fixed point $x_0 \in C_S^\circ$ of F (if it exists) is unique to within scalar multiples. If F has a fixed point $x_0 \in C_S^\circ$, then for any $x \in C_S^\circ$ there exists a scalar $\lambda(x) > 0$ such that

$$\lim_{m \to \infty} F^m(x) = \lambda(x) x_0$$

is a fixed point of F, where convergence is in the L_∞ norm. The map $x \mapsto \lambda(x)$ is real analytic. Furthermore, for any $x \in C_S^\circ$, $G^m(x)$ converges to a fixed point of F in C_S°, and the convergence is geometric.

Proof. The fact that c_1 satisfies Assumption 5.11 if S is connected and metrizable follows from Lemma 5.13 and the assumption that k satisfies (4.3). Because $k \in L_\infty(S \times T)$, k satisfies (4.3) and S and T are compact Hausdorff spaces with corresponding Borel measures ds and dt of full support, it is also not hard to see that $A_1(C_S^\circ) \subset C_T^\circ$ and $A_2(C_T^\circ) \subset C_S^\circ$. We leave the details to the reader.

Suppose that $f \in L_1(S)$ and $g \in L_1(T)$ are strictly positive almost everywhere and solve the *DAD* problem (3.5). By using the facts that $k \in L_\infty(S \times T)$, k satisfies (4.3) and S is compact, it is not hard to see that

$$x_0(s) = \int k(s, t)\, g(t)\, dt$$

defines a function $x_0 \in C_S^\circ$. Similarly, one finds that if

$$y_0(t) = \int k(s, t) f(s)\, ds,$$

then $y_0 \in C_T^\circ$. Equation (3.5) then gives that

$$\int k(s, t) \frac{\alpha(s)}{x_0(s)}\, ds = y_0(t), \qquad \text{a.e. } [dt], \qquad \text{and}$$

$$\int k(s, t) \frac{\beta(t)}{y_0(t)}\, dt = x_0(s), \qquad \text{a.e. } [ds].$$

This shows that $y_0 = A_1 J_S x_0$ and $x_0 = A_2 J_T y_0$, so $x_0 = F(x_0)$.

Conversely, suppose that $F(x_0) = \lambda x_0$ for some $x_0 \in C_S^\circ$. Necessarily, $\lambda > 0$, and writing $y_0 = A_1 J_S x_0$, one obtains

$$y_0(t) = \int k(s, t) \frac{\alpha(s)}{x_0(s)}\, ds, \qquad \text{a.e. } [dt], \qquad \text{and}$$

$$\lambda x_0(s) = \int k(s, t) \frac{\beta(t)}{y_0(t)}\, dt, \qquad \text{a.e. } [ds].$$

These equations yield

$$\beta(t) = \int \frac{\beta(t)}{y_0(t)} k(s, t) \frac{\alpha(s)}{x_0(s)}\, ds, \qquad \text{a.e. } [dt] \qquad \text{and}$$

$$\lambda \alpha(s) = \int \frac{\beta(t)}{y_0(t)} k(s, t) \frac{\alpha(s)}{x_0(s)}\, dt, \qquad \text{a.e. } [ds].$$

Thus, to prove we have a solution of (3.5), it suffices to prove $\lambda = 1$. However, we have

$$\int_T \beta(t)\, dt = \int_T \int_S \frac{\beta(t)}{y_0(t)} k(s, t) \frac{\alpha(s)}{x_0(s)}\, ds\, dt = \lambda \int_S \alpha(s)\, ds,$$

and since we assume $\int_S \alpha(s)\, ds = \int_T \beta(t)\, dt > 0$, $\lambda = 1$.

It follows that to prove uniqueness (within scalar multiples) of L_1 functions f and g which are strictly positive almost everywhere and solve (3.5), it suffices to prove that if F has a fixed point $x_0 \in C_S^\circ$ then all other fixed points of F in C_S° are of the form ρx_0, for $\rho > 0$. Note that F is order-preserving and homogeneous of degree one, so F is nonexpansive with respect to Hilbert's projective metric. Note also that F is C^1 (in fact, real

analytic) on C_S°. Suppose that $F(x_0) = x_0 \in C_S^\circ$, and define $\tilde{B} = F'(x_0)$, the Fréchet derivative of F at x_0. If we set $y_0 = A_1 J_S x_0$, the chain rule implies

$$F'(x_0) = \tilde{B} = \tilde{A}_2 \tilde{A}_1, \qquad \text{where}$$

$$(\tilde{A}_1 u)(t) = \int k(s, t)\, \tilde{\alpha}(s)\, u(s)\, ds, \qquad \text{where} \quad \tilde{\alpha} = \alpha/(x_0)^2, \qquad \text{and}$$

$$(\tilde{A}_2 v)(s) = \int k(s, t)\, \tilde{\beta}(t)\, v(t)\, dt, \qquad \text{where} \quad \tilde{\beta} = \beta/(y_0)^2.$$

Here, $\tilde{A}_1 : L_\infty(S) \to L_\infty(T)$ and $\tilde{A}_2 : L_\infty(T) \to L_\infty(S)$.

It follows as in Lemma 5.12 that $\tilde{B} = \tilde{A}_2 \tilde{A}_1$ is given by

$$(\tilde{B}w)(r) = \int \tilde{c}_1(r, s)\, \tilde{\alpha}(s)\, w(s)\, ds, \qquad \text{where}$$

$$\tilde{c}_1(r, s) = \int k(s, t)\, k(r, t)\, \tilde{\beta}(t)\, dt.$$

It follows that there exist positive constants such that $k_1 \tilde{c}_1(r, s) \leqslant c_1(r, s) \leqslant k_2 \tilde{c}_1(r, s)$, where $c_1(r, s)$ is as in Lemma 5.12. Since c_1 satisfies Assumption 5.11, so does $\tilde{c}_1$. It follows from Lemma 5.12 that the essential spectral radius, $\rho(\tilde{B})$, of $\tilde{B}$ is strictly less than $r(\tilde{B})$, the spectral radius of $\tilde{B}$. Also, there exists an integer m so $\tilde{B}^m(C_S \backslash \{0\}) \subset C_S^\circ$. Theorem 3.2 on p. 93 of [18] now implies that for each $x \in C_S^\circ$, there exists a scalar $\lambda(x) > 0$ such that

$$\lim_{k \to \infty} F^k(x) = \lambda(x) x_0,$$

and the map $x \mapsto \lambda(x)$ is real-analytic. In particular this implies that fixed points of F in C_S° are unique to within scalar multiples. Because F is homogeneous of degree one and order-preserving, F is nonexpansive with respect to Hilbert's projective metric d on C_S° and

$$G^n(x) = \frac{F^n(x)}{\psi(F^n(x))}.$$

The fact that $G^n(x)$ converges geometrically to a fixed point of F follows from Theorem 2.7, p. 78, in [18]. ∎

We observe here that in fact the convergence of F^m demonstrated in the above result is exactly equivalent (after a change of variables) to the convergence of the "iterative proportional fitting procedure" described at the end of Section 3.

Remark 5.15. If $\tilde{B} = F'(x_0)$ is defined as in the proof of Theorem 5.14, Lemma 5.12 actually implies that

1. $\tilde{B}(u) = ru$ for some $u \in C_S^\circ$ and $r = r(\tilde{B})$, the spectral radius of $\tilde{B}$,
2. $(\tilde{B})^*(u^*) = ru^*$ for some $u^* \in (L_\infty(S))^*$ with $u^*(u) > 0$,
3. $rI - \tilde{B}$ is Fredholm of index zero, and
4. the dimension of the null space $N(rI - \tilde{B})$ equals one.

A map $\tilde{B}$ which satisfies these conditions is said to satisfy "condition K–R," see p. 41 in [18]. Let $Y = \{x \in L_\infty(S) \mid \int x(s)\, ds = 0\}$ and let G be as in Theorem 5.14. Because $F'(x_0)$ satisfies condition K–R whenever $F(x_0) = x_0 \in C_S^\circ$, it is a special case of an argument in [18, pp. 50–52] that $(I - G')(x_0) \mid Y$ is one–one and onto Y. This observation will play a crucial role in our further remarks.

We shall now show how the uniqueness ideas in Theorem 5.14 can be combined with the implicit function theorem and our results from Section 3 to prove an existence theorem for solutions of the *DAD* problem (3.5).

THEOREM 5.16. *Let assumptions and notation be as Theorem* 5.14. *Assume that the DAD problem* (3.5) *has a solution* $f \in L_1(S)$ *and* $g \in L_1(T)$ *such that* f *and* g *are positive almost everywhere. Assume that* $\tilde{k} \in L_\infty(S \times T)$, $\tilde{k}(s, t) \geqslant 0$, *a.e.* $[ds\, dt]$, *and* $\tilde{k}$ *satisfies* (4.3). *Define* K *and* $\tilde{K}$ (*up to sets of measure zero*) *by*

$$K = \{(s, t) \mid k(s, t) > 0\} \qquad \text{and} \qquad \tilde{K} = \{(s, t) \mid \tilde{k}(s, t) > 0\}.$$

Assume that there exists a constant M *so* $k(s, t) \leqslant M\tilde{k}(s, t)$ *almost everywhere* (*so* $K \subset \tilde{K}$) *and that* $\tilde{K}$ *is a countable union of measurable rectangles. Then there are functions* $\tilde{f} \in L_1(S)$ *and* $\tilde{g} \in L_1(T)$ *such that* $\tilde{f}$ *and* $\tilde{g}$ *are positive almost everywhere and* $\tilde{f}$ *and* $\tilde{g}$ *give a solution of the DAD problem* (3.5) *for* $\tilde{k}$. *The functions* $\tilde{f}$ *and* $\tilde{g}$ *are unique to within scalar multiples and can be obtained by an iterative procedure as in Theorem* 5.14.

Proof. For $0 \leqslant \lambda \leqslant 1$ define $k_\lambda = (1 - \lambda)k + \lambda\tilde{k}$. Define $A_1^\lambda : L_\infty(S) \to L_\infty(T)$ and $A_2^\lambda : L_\infty(T) \to L_\infty(S)$ by

$$(A_1^\lambda v)(t) = \int_S k_\lambda(s, t)\, \alpha(s)\, v(s)\, ds,$$

and

$$(A_2^\lambda w)(s) = \int_T k_\lambda(s, t)\, \beta(t)\, w(t)\, dt.$$

Define a C^1 map $F: C_S^\circ \times [0,1] \to C_S^\circ$ by $F(x,\lambda) = (A_2^\lambda J_T A_1^\lambda J_S)(x)$. For $x \in L_1(S)$, define $\psi(x) = \int_S x(s)\,ds$ and define Y (a closed linear subspace of $L_\infty(S)$) by $Y = \{x \in L_\infty(S) \mid \psi(x) = 0\}$. By assumption and by Theorem 5.14, there exists $x_0 \in C_S^\circ$ such that $F(x_0, 0) = x_0$ and $\psi(x_0) = 1$. Define

$$G(x,\lambda) = \frac{F(x,\lambda)}{\psi(F(x,\lambda))}, \qquad \text{for} \quad x \in C_S^\circ, \quad 0 \leqslant \lambda \leqslant 1.$$

Define $U = \{y \in Y \mid x_0 + y \in C_S^\circ\}$ and define $H: U \times [0,1] \to Y$ by

$$H(y,\lambda) = x_0 + y - G(x_0 + y, \lambda).$$

By Theorem 5.14, the *DAD* problem (3.5) (with k_λ replacing k in (3.5)) has a solution if there exists $y \in U$ such that $H(y,\lambda) = 0$. However, Remark 5.15 implies that the Fréchet derivative of the map $y \mapsto H(y,0)$ at $y = 0$ is one-one and onto Y. Thus we can apply the implicit function theorem to H: There exist $\lambda_0 > 0$ and a C^1 map $\lambda \mapsto y_\lambda \in Y$ for $0 \leqslant \lambda \leqslant \lambda_0$ such that $x_0 + y_\lambda \in U$ and $H(y_\lambda, \lambda) = 0$. It follows from Theorem 5.14 that if we define

$$f_\lambda = \frac{\alpha}{x_0 + y_\lambda}, \qquad \text{and} \qquad g_\lambda = \frac{\beta}{A_1^\lambda J_S(x_0 + y_\lambda)},$$

then $f_\lambda \in L_1(S)$, $g_\lambda \in L_1(T)$ and f_λ and g_λ are positive almost everywhere and solve (3.5) for k_λ.

For $0 \leqslant \lambda \leqslant 1$ define

$$k_\lambda^*(s,t) = \alpha(s)\, k_\lambda(s,t)\, \beta(t),$$

so $k_1^*(s,t) = \alpha(s)\, \tilde{k}(s,t)\, \beta(t)$. For $0 \leqslant \lambda \leqslant \lambda_0$ define

$$f_\lambda^* = \frac{1}{x_0 + y_\lambda}, \qquad \text{and} \qquad g_\lambda^* = \frac{1}{A_1^\lambda J_S(x_0 + y_\lambda)},$$

and note that $f_\lambda^* \in C_S^\circ$ and $g_\lambda^* \in C_T^\circ$. Take a fixed $\lambda \in (0, \lambda_0)$, and define

$$u^*(s,t) = \begin{cases} f_\lambda^*(s)\, g_\lambda^*(t)\, k_\lambda^*(s,t)/k_1^*(s,t) \\ \quad = f_\lambda^*(s)\, g_\lambda^*(t)\, k_\lambda(s,t)/k_1(s,t), & \text{for} \quad (s,t) \in \tilde{K} \\ 0, & \text{for} \quad (s,t) \notin \tilde{K}. \end{cases}$$

We easily derive from our assumptions on k and $\tilde{k}$ that

$$\lambda f_\lambda^*(s)\, g_\lambda^*(t) \leqslant u^*(s,t) \leqslant [(1-\lambda)M + \lambda]\, f_\lambda^*(s)\, g_\lambda^*(t) \qquad \text{on } \tilde{K}.$$

This implies that $u^*(s, t) > 0$, a.e. $[\tilde{k}\, ds\, dt]$ and that there exist constants μ_1 and μ_2 such that $0 < \mu_1 \leqslant u^*(s, t) \leqslant \mu_2$, a.e. $[ds\, dt]$ on $\tilde{K}$. If we define ϕ as in Section 2, the above inequality shows that

$$\iint \phi(u^*(s, t))\, k_1^*(s, t)\, ds\, dt < \infty.$$

Also u^* has been chosen so that

$$\int u^*(s, t)\, k_1^*(s, t)\, dt = \alpha(s), \qquad \text{a.e. } [ds] \qquad \text{and}$$

$$\int u^*(s, t)\, k_1^*(s, t)\, ds = \beta(t), \qquad \text{a.e. } [dt].$$

It follows that condition (CQ1) of Section 3 is satisfied for k_1^*, so there exists functions f^* and g^* which satisfy

$$\int f^*(s)\, k_1^*(s, t)\, g^*(t)\, dt = \alpha(s), \qquad \text{a.e. } [ds], \qquad \text{and}$$

$$\int f^*(s)\, k_1^*(s, t)\, g^*(t)\, ds = \beta(t), \qquad \text{a.e. } [dt],$$

where $f^*(s)\, g^*(t) \in L_1(S \times T, k_1^*\, ds\, dt)$. The results of Section 4 imply that f^* and g^* are measurable, and the assumption that α and β are positive almost everywhere implies immediately that f^* and g^* are positive almost everywhere. We claim that $f^* \in C_S^\circ$ and $g^* \in C_T^\circ$. If we can prove this, then $\tilde{f} = \alpha f^* \in L_1(S)$ and $\tilde{g} = \beta g^* \in L_1(T)$ will satisfy the *DAD* problem (3.5) for $\tilde{k}$.

To prove that $f^* \in C_S^\circ$, notice that Fubini's theorem gives

$$f^*(s) = \left(\int \tilde{k}(s, t)\, \beta(t)\, g^*(t)\, dt \right)^{-1},$$

where $\int \tilde{k}(s, t)\, \beta(t)\, g^*(t)\, dt < \infty$, a.e. $[ds]$. By (4.3), we can cover S by a finite number of open sets U_i, for $1 \leqslant i \leqslant m$, such that for each i there exists an open set $V_i \subset T$ and a constant $\kappa_i > 0$ such that $k(s, t) \geqslant \kappa_i$ for almost all $(s, t) \in U_i \times V_i$. If $\kappa = \min_i \{\kappa_i\}$, we see that for almost all $s \in U_i$,

$$+\infty > \int_T k(s, t)\, \beta(t)\, g^*(t)\, dt \geqslant \kappa \int_{V_i} \beta(t)\, g^*(t)\, dt.$$

Because β and g^* are positive almost everywhere, we have

$$\delta = \min_{1 \leq i \leq m} \left\{ \int_{V_i} \beta(t)\, g^*(t)\, dt \right\} > 0.$$

We conclude that, for almost all $s \in S$,

$$\int_T k(s, t)\, \beta(t)\, g^*(t)\, dt \geqslant \kappa\delta = \delta_1,$$

and $f^*(s) \leqslant \delta_1^{-1}$. A similar argument shows that $g^*(t)$ is bounded in $L_\infty(T)$, so $g^*(t) \leqslant \delta_2^{-1}$ a.e. $[dt]$. However, this yields

$$f^*(s) = \left(\int_T \tilde{k}(s, t)\, \beta(t)\, g^*(t)\, dt \right)^{-1} \geqslant \|\tilde{k}\|_\infty^{-1}\, \delta_2 \left(\int_T \beta(t)\, dt \right)^{-1}, \qquad \text{a.e. } [ds],$$

so $f^* \in C_S^\circ$. A similar argument shows that $g^* \in C_T^\circ$.

Because we assume that $k \leqslant M\tilde{k}$ almost everywhere and k satisfies the conditions of Theorem 5.14, it is not hard to see that $\tilde{k}$ also satisfies the conditions of Theorem 5.14. This gives the final assertion of the theorem. ∎

By combining Theorems 5.8 and 5.16, we can now give a new existence and uniqueness theorem for the *DAD* problem (3.5) and a geometrically convergent iterative procedure for constructing solutions.

THEOREM 5.17. *Let S be a compact Hausdorff space with a finite, regular Borel measure of full support, τ. Assume that $\tilde{k} \in L_\infty(S \times S, \tau \times \tau)$ is nonnegative almost everywhere and that $\alpha \in L_1(S)$ is positive almost everywhere. For each $s \in S$, assume that there exists an open neighbourhood G_s and a positive number $\delta_s > 0$ such that $\tilde{k}(r, t) \geqslant \delta_s$ for almost all $(r, t) \in G_s \times G_s$. Define*

$$k^*(s, t) = \min\{\tilde{k}(s, t), \tilde{k}(t, s)\} \qquad \text{and} \qquad c_1^*(r, s) = \int k^*(r, t)\, k^*(s, t)\, d\tau(t),$$

and suppose that Assumption 5.11 is satisfied by c_1^*.

Suppose that $\tilde{K} = \{(s, t) \in S \times S \mid \tilde{k}(s, t) > 0\}$ is a countable union of measurable rectangles. Then there exist functions f and $g \in L_1(S)$ which are positive almost everywhere and satisfy

$$\int f(s)\, \tilde{k}(s, t)\, g(t)\, d\tau(t) = \alpha(s) \qquad \text{a.e.} \qquad \text{and}$$

$$\int f(s)\, \tilde{k}(s, t)\, g(t)\, d\tau(s) = \alpha(t) \qquad \text{a.e.}$$

The functions f and g are unique to within scalar multiples and can be approximated by the iterative procedure in Theorem 5.14. *If $\tilde{k}$ and α are continuous, f and g can be chosen continuous.*

Proof. If we apply Theorem 5.8 to k^* and define $f^* = g^* = \alpha/u$, where

$$\int k^*(s, t)\, \alpha(t) \left(\frac{1}{u(t)}\right) d\tau(t) = u(s), \qquad \text{a.e.},$$

we find (using the symmetry of k^*) that

$$\int f^*(s)\, k^*(s, t)\, g^*(t)\, d\tau(t) = \alpha(s) \qquad \text{a.e.} \qquad \text{and}$$

$$\int f^*(s)\, k^*(s, t)\, g^*(t)\, d\tau(s) = \alpha(t) \qquad \text{a.e.}$$

We now apply Theorem 5.16, with k^* replacing k. Our definition of k^* ensures that $k^* \leqslant \tilde{k}$ a.e. The assumptions about G_s and δ_s imply that $\tilde{k}$ and k^* satisfy (4.3). Thus the hypotheses of Theorem 5.16 are satisfied, and we obtain the existence of f and g. The uniqueness of f and g follows from Theorem 5.14, and the continuity of f and g (when $\tilde{k}$ and α are continuous) follows from Corollary 4.12. ∎

If S is connected and metrizable, Theorem 5.17 takes a simpler form.

COROLLARY 5.18. *Let S be a connected, compact, metric space with a finite, regular Borel measure of full support τ. Assume that $\tilde{k} \in L_\infty(S \times S, \tau \times \tau)$ is nonnegative almost everywhere and that $\alpha \in L_1(S)$ is positive everywhere. For each $s \in S$, assume that there exists a positive number δ_s and an open neighbourhood G_s of s such that $\tilde{k}(r, t) \geqslant \delta_s$ for almost all $(r, t) \in G_s \times G_s$. In addition assume that $\tilde{K} = \{(s, t) \mid \tilde{k}(s, t) > 0\}$ is a countable union of measurable rectangles. Then there exist functions f and $g \in L_1(S)$ which satisfy all the conditions of Theorem* 5.17, *so the DAD problem* (3.5) *is uniquely solved for $\tilde{k}$ and $\alpha = \beta$.*

Proof. If k^* is defined as in Theorem 5.17, we have $k^*(r, t) \geqslant \delta_s$ for almost all $(r, t) \in G_s \times G_s$. It follows that k^* satisfies condition (4.3), so Lemma 5.13 implies that c_1^* satisfies Assumption 5.11. Corollary 5.18 now follows immediately from Theorem 5.17. ∎

Conclusion

The techniques leading to the main results in this paper comprise a mixture of variational and fixed point methods. Many of our results appear inaccessible without the combined use of both approaches.

REFERENCES

1. J. M. BORWEIN AND A. S. LEWIS, Decomposition of multivariate functions, *Canad. J. Math.* **44** (1992), 1–20.
2. J. M. BORWEIN AND A. S. LEWIS, Partially finite convex programming, Part I, Duality theory, *Math. Programming B* (1992), 15–48.
3. J. M. BORWEIN AND A. S. LEWIS, Partially finite convex programming, Part II, Explicit lattice models, *Math. Programming B* (1992), 49–84.
4. R. A. BRUALDI, S. V. PARTER, AND H. SCHNEIDER, The diagonal equivalence of a nonnegative matrix to a stochastic matrix, *J. Math. Anal. Appl.* **16** (1966), 31–50.
5. Y. CENSOR AND S. A. ZENIOS, "Interval Constrained Matrix Balancing," Technical Report 89-09-09, University of Pennsylvania, Decision Sciences Department, 1990.
6. I. CSISZÁR, I-divergence geometry of probability distributions and minimization problems, *Annals of Probability* **3** (1975), 146–158.
7. J. DIESTEL, "Sequences and Series in Banach Spaces," Springer-Verlag, New York, 1984.
8. D. Z. DJOKVIĆ, A note on nonnegative matrices, *Proc. Amer. Math. Soc.* **25** (1970), 80–82.
9. F. FRANKLIN AND J. LORENZ, On the scaling of multidimensional matrices, *Linear Algebra Appl.* **114** (1989), 717–735.
10. C. HOBBY AND R. PYKE, Doubly stochastic operators obtained from positive operators, *Pacific J. Math.* **15** (1965), 153–157.
11. C. T. IRELAND AND S. KULLBACK, Contingency tables with given marginals, *Biometrika* **55** (1968), 179–188.
12. S. KARLIN AND L. NIRENBERG, On a theorem of P. Nowosad, *J. Math. Anal. Appl.* **17** (1967), 61–67.
13. S. KULLBACK, Probability densities with given marginals, *Ann. Math. Statist.* **39** (1968), 1236–1243.
14. M. V. MENON AND H. SCHNEIDER, The spectrum of a nonlinear operator associated with a matrix, *Linear Algebra Appl.* **2** (1969), 321–334.
15. P. NOWOSAD, On the integral equation $Kf = 1/f$ arising in a problem in communication, *J. Math. Anal. Appl.* **14** (1966), 484–492.
16. R. D. NUSSBAUM, The radius of the essential spectrum, *Duke J. Math.* **38** (1970), 473–478.
17. R. D. NUSSBAUM, Iterated Nonlinear maps and Hilbert's projective metric: A summary, *in* "Dynamics of Infinite Dimensional Systems" (S.-N. Chow and J. Hale, Eds.), pp. 231–249, Springer-Verlag, New York, 1987.
18. R. D. NUSSBAUM, "Hilbert's Projective Metric and Iterated Nonlinear Maps, I," American Mathematical Society, Providence, RI, 1988.
19. R. D. NUSSBAUM, "Iterated Nonlinear Maps and Hilbert's Projective Metric, II," American Mathematical Society, Providence, RI, 1989.
20. H. PERFECT AND L. MIRSKY, The distribution of positive elements in doubly-stochastic matrices, *J. London Math. Soc.* **40** (1965), 689–698.
21. R. T. ROCKAFELLAR, Integrals which are convex functionals, II, *Pacific J. Math.* **39** (1971), 439–469.
22. W. RUDIN, "Real and Complex Analysis," McGraw–Hill, New York, 1966.
23. W. RUDIN, "Functional Analysis," McGraw–Hill, New York, 1973.
24. M. H. SCHNEIDER, Matrix scaling, entropy minimization and conjugate duality, I, existence conditions, *Linear Algebra Appl.* **114** (1989), 785–813.
25. M. H. SCHNEIDER, Matrix scaling, entropy minimization and conjugate duality, II, the dual problem, *Math. Programming B* **48** (1990), 103–124.

26. R. SINKHORN AND P. KNOPP, Concerning nonnegative matrices and doubly stochastic matrices, *Pacific J. Math.* **21** (1967), 343–348.
27. A. C. THOMPSON, On certain contraction mappings in a partially ordered vector space, *Proc. Amer. Math. Soc.* **14** (1963), 438–443.
28. G. T. WHYBURN, "Topological Analysis," second ed., Princeton Univ. Press, Princeton, NJ, 1964.