Transportation approach to some concentration inequalities in product spaces

Amir Dembo[1]

and Ofer Zeitouni[2]

Department of Electrical Engineering, Technion

January 17, 1996. Revised: October 24, 1996

Abstract *Using a transportation approach we prove that for every probability measures* P, Q_1, Q_2 *on* Ω^N *with* P *a product measure there exist r.c.p.d.* ν_j *such that* $\int \nu_j(\cdot|x)dP(x) = Q_j(\cdot)$ *and*

$$\int dP(x) \int \frac{dP}{dQ_1}(y)^\beta \frac{dP}{dQ_2}(z)^\beta (1+\beta(1-2\beta))^{f_N(x,y,z)} d\nu_1(y|x) d\nu_2(z|x) \leq 1 \ ,$$

for every $\beta \in (0, 1/2)$*. Here* f_N *counts the number of coordinates* k *for which* $x_k \neq y_k$ *and* $x_k \neq z_k$*. In case* $Q_1 = Q_2$ *one may take* $\nu_1 = \nu_2$*. In the special case of* $Q_j(\cdot) = P(\cdot|A)$ *we recover some of Talagrand's sharper concentration inequalities in product spaces.*

In [Tal95, Tal96a], Talagrand provides a variety of concentration of measure inequalities which apply in every product space Ω^N (with Ω Polish) equipped with a Borel product (probability) measure P. These inequalities are extremely useful in combinatorial applications such as the longest common/increasing subsequence, in statistical physics applications such as the study of spin glass models, and in areas touching upon functional analysis such as probability in Banach spaces (c.f. [Tal95, Tal96a] and the references therein). The proofs of these inequalities are all based on an induction on N, where in order to prove the concentration of measure result for a generic set $A \subset \Omega^{N+1}$ one applies the induction hypothesis for the N dimensional sets $A(\omega) = \{(y_1, \ldots, y_N) : (y_1, \ldots, y_N, \omega) \in A\}$, $\omega \in \Omega$ fixed, and $B = \cup_\omega A(\omega)$.

Marton, in [Mar96a, Mar96b], building upon [Mar86], extends some of Talagrand's results to the context of contracting Markov chains. In these works concentration inequalities related to the

[1] Current address: Department of Mathematics and Department of Statistics, Stanford University, Stanford, CA 94305. Partially supported by an NSF DMS-9403553 grant and by a US-ISRAEL BSF grant.

[2] Partially supported by a US-ISRAEL BSF grant, and by the Technion E. and J. Bishop research fund.

Key Words: Concentration inequalities, product spaces, transportation.

AMS (1991) subject classifications: 60E15,28A35.

"distance" between a set A and a point x are viewed as consequences of inequalities involving the appropriate "distance" between probability measures, in the special case that some of these measures are supported on A. Since the latter "distance" between measures involves an appropriate coupling, that is, finding the optimal way of "transporting" mass from one measure to another, it has been dubbed the transportation approach. This approach is applied in [De96, Tal96c], where other transportation problems are solved and some new information inequalities are derived in order to recover the inequalities of [Tal95] in their sharpest form and to obtain a few new variants.

A third, different approach to some of the inequalities of [Tal95] via Poincaré and logarithmic Sobolev inequalities is provided in [Led96, BL96].

Common to all inequalities in [Tal95, Tal96a] is their additive nature, allowing for a relatively easy transfer from the transportation problem for $N = 1$ to the general case. In doing so, the relative entropy between probability measures plays a prominent role in [De96, Mar96a, Mar96b, Tal96c].

In contrast, Talagrand in [Tal96b] provides proofs by his induction method of a new family of inequalities, stronger than those of [Tal95, Tal96a], and in which the additive structure of the latter is replaced by a multiplicative one.

In this short note we adapt the transportation approach, avoiding the use of relative entropies, and recover as special cases some of the new inequalities of [Tal96b].

Specifically, for $x = (x_1, x_2, \cdots, x_N) \in \Omega^N$, $y = (y_1, \cdots, y_N) \in \Omega^N$ and $z = (z_1, \cdots, z_N) \in \Omega^N$ let

$$f_N(x, y, z) = \sum_{k=1}^{N} f_1(x_k, y_k, z_k) = \sum_{k=1}^{N} \mathbf{1}_{x_k \neq y_k, x_k \neq z_k} \,. \tag{1}$$

Then, [Tal96b, Theorem 2.1] is the following concentration inequality:

Theorem 1 *For every $N \geq 1$, $\beta \in (0, \infty)$, a product measure P on Ω^N and $A \subset \Omega^N$ there exist*

ν, such that $\nu(A|x) = 1$ for every $x \in \Omega^N$ and for non-negative $a \le \beta/(2\beta+1)$,

$$\int_{\Omega^N} dP(x) \int_{A\times A} (1+a)^{f_N(x,y,z)}\, d\nu(y|x) d\nu(z|x) \ \le P(A)^{-2\beta} \tag{2}$$

This inequality is applied in [Tal96b, Theorem 1.2] to control the tails of the norm of quadratic forms in Rademacher variables.

Let Q_1, Q_2 Borel probability measures on Ω^N, and let ν_1, ν_2 denote any r.c.p.d. such that for $j = 1, 2$

$$\int_{\Omega^N} \nu_j(\cdot|x) dP(x) = Q_j(\cdot), \tag{3}$$

and for $j = 1, 2$, let $\hat{\Omega}_j \subset \Omega^N$ be such that $Q_j(\hat{\Omega}_j) = 1$ and $\frac{dP}{dQ_j}$ exists on $\hat{\Omega}_j$, setting $\nu_j(\hat{\Omega}_j^c|x) = 0$ for every $x \in \Omega^N$.

The next theorem is the main result of this note. As pointed out in Remark 1 below, it implies Talagrand's Theorem 1, at least for a certain range of parameters a, β.

Theorem 2 *For every probability measures P, Q_1, Q_2 on Ω^N, $N \ge 1$ such that P is a product measure, there exist ν_j satisfying (3) such that for every $\beta \in (0, 1/2)$ and non-negative $a \le \beta(1-2\beta)$,*

$$\int_{\Omega^N} dP(x) \int_{\hat{\Omega}_1 \times \hat{\Omega}_2} \frac{dP}{dQ_1}(y)^\beta \frac{dP}{dQ_2}(z)^\beta (1+a)^{f_N(x,y,z)} d\nu_1(y|x) d\nu_2(z|x) \le 1\ . \tag{4}$$

In case $Q_1 = Q_2$ one may take also $\nu_1 = \nu_2$.

Remark 1 Considering (4) with $Q_j(\cdot) = P(\cdot|A)$ for which $\frac{dP}{dQ_j}(y) = P(A)$ for a.e. $y \in A$ and then taking $\nu_1 = \nu_2$ we recover (2) apart from the fact that now only $\beta \in (0, 1/2)$ is allowed with $a \le \beta(1-2\beta)$. In contrast to all previous examples, the constants in the concentration of measure inequality (2) differ from the best constants in the corresponding transportation inequality (4), where the choice $a = \beta(1-2\beta)$ can not be improved upon, even for $N = 1$, $Q_1 = Q_2$. Indeed, for $a > \beta(1-2\beta)$ fix $\delta > 0$ such that $\xi = a(1-\delta) - \beta(1-2\beta) > 0$, denoting $\Delta = (1-\delta)/\delta$. Consider $\Omega = \{0, 1\}$ with $P(\{1\}) = \delta$ and $Q_j(\{1\}) = \delta(1 - \epsilon\Delta)$ for $\epsilon \in (0, \delta)$. Then, for ν_j satisfying (3) we have $w_j = \nu_j(\{0\}|\{1\}) \geq \epsilon\Delta$ with the LHS of (4) being

$$\begin{aligned} G_\epsilon(w_1, w_2) &= (1-\delta)\prod_{j=1}^{2}\left(\frac{1+\epsilon - w_j/\Delta}{(1+\epsilon)^\beta} + \frac{w_j/\Delta - \epsilon}{(1-\epsilon\Delta)^\beta}\right) + \delta\prod_{j=1}^{2}\left(\frac{1-w_j}{(1-\epsilon\Delta)^\beta} + \frac{w_j}{(1+\epsilon)^\beta}\right) \\ &+ a\left((1-\delta)\prod_{j=1}^{2}\frac{(w_j/\Delta - \epsilon)}{(1-\epsilon\Delta)^\beta} + \delta\frac{w_1 w_2}{(1+\epsilon)^{2\beta}}\right) = A(\epsilon)w_1 w_2 - B(\epsilon)(w_1 + w_2) + C(\epsilon) \end{aligned}$$

for smooth functions $A(\epsilon), B(\epsilon) > 0$ and $C(\epsilon)$. Since $B(0) = 0$ and $B'(0) = a\delta < a = A(0)\Delta$, for ϵ sufficiently small the infimum of G_ϵ is obtained at $w_1 = w_2 = \epsilon\Delta$. One checks that $C(0) = 1$, $C'(0) = 0$ and $C''(0)/2 - 2B'(0)\Delta + A(0)\Delta^2 = \Delta\xi$. Thus, $\inf_{w_j} G_\epsilon(w_1, w_2) = 1 + \epsilon^2\Delta\xi + O(\epsilon^3) > 1$ for ϵ sufficiently small.

The same proof we provide allows for extending the inequality (4) to $q > 2$ different measures Q_j and the corresponding ν_j satisfying (3) with $f_1(x_k, y_k^1, \cdots, y_k^q) = \mathbf{1}_{x_k \notin \{y_k^j, j=1,\ldots,q\}}$. The extended inequality then holds for all choices of β, a such that

$$\int_0^1 [(u^{1-\beta} + (1-u)J)^q + a(1-u)^q J^q] d\overline{P}(u) + \int_0^{1^-} v^{q\beta+1} d\overline{Q}(v) \leq 1\,, \tag{5}$$

for every $\overline{P}$ and $\overline{Q}$ satisfying (9) and (10) below and J determined via (11) below. Setting $Q_j = P(\cdot|A)$ one may hope to recover [Tal96b, Theorem 5.1], corresponding to (2) for $q > 2$. However, a necessary condition for (5) to hold is $a \leq \beta q \leq 1$, while the essence of [Tal96b, Theorem 5.1] is that for $q >> 2$ one may take $\beta = 1$ and a/q bounded below away from zero.

Key to the proof of Theorem 2 is the next proposition which is of some independent interest.

Proposition 1 *For every probability measures P and Q on Ω, there exists a r.c.p.d. ν such that*

$$\int_\Omega \nu(\cdot|x)dP(x) = Q(\cdot) \tag{6}$$

and for every $\beta \in (0,1/2)$, $a \leq \beta(1-2\beta)$,

$$I(\nu) \triangleq \int_\Omega dP(x) \left(\left[\int_{\hat{\Omega}} \frac{dP}{dQ}(y)^\beta d\nu(y|x) \right]^2 + a \left[\int_A \frac{dP}{dQ}(y)^\beta d\nu(y|x) \mathbf{1}_{y\neq x} \right]^2 \right) \leq 1\,. \tag{7}$$

Here, $\hat{\Omega}$ is such that $Q(\hat{\Omega}) = 1$ and dP/dQ exists on $\hat{\Omega}$.

Remark 2 Proposition 1 is proved using the r.c.p.d. $\nu^*(\cdot|x)$, independent of β, given in (8) below. This is the same r.c.p.d. used when proving the coupling characterization of the total variation distance (c.f. [BHJ92, page 253]). The condition $a \leq \beta(1-2\beta)$ can not be relaxed by choosing any r.c.p.d. ν that satisfies (6) (see Remark 1 above).

The proof of Proposition 1 relies upon the following technical lemma.

Lemma 1 *Let $h_J(u) = (u^{1-\beta} + (1-u)J)^2 + a(1-u)^2 J^2$. Suppose $\beta < 1/2$ and $0 < a \leq \beta(1-2\beta)$. Then, $h_J(\cdot)$ is concave for every $J \in [1-\beta, 1]$ and $h_J(u) \leq 1$ for every $u \in [0,1]$ and every $J \in [0, 1-\beta]$.*

Proof of Lemma 1: One checks that

$$h_J^{(3)}(x) = 2(1-\beta)\beta x^{-(\beta+2)}[J((1+\beta)+(2-\beta)x) - 2(1-2\beta)x^{1-\beta}]\,.$$

Since $x^{1-\beta} \leq (1-\beta)x + \beta$ for all $x \in [0,1]$ it follows that

$$h_{1-\beta}^{(3)}(x) \geq 2(1-\beta)\beta x^{-(\beta+2)}[(1-\beta)3\beta x + (1+3\beta^2-2\beta)] \geq 0\,,$$

for every $x \in [0,1]$ and $\beta \in [0,0.5]$. Since $h_J^{(3)}(x)$ is monotone increasing in J, it follows that $h_J^{(3)}(x) \geq 0$ for every $J \in [1-\beta,1]$, $x \in [0,1]$. Also,

$$h_J^{(2)}(1) = 2[(1-\beta)(1-2\beta) - 2J(1-\beta) + (a+1)J^2]\,,$$

and for $a \leq \beta(1-2\beta) \leq \beta/(1-\beta)$ both $h_1^{(2)}(1) \leq 0$ and $h_{1-\beta}^{(2)}(1) \leq 0$. Hence, $h_J^{(2)}(1) \leq 0$ for every $J \in [1-\beta,1]$. With $h_J^{(3)}(x) \geq 0$ and $h_J^{(2)}(1) \leq 0$ we deduce that $h_J(\cdot)$ is concave in $[0,1]$ for $J \in [1-\beta,1]$. Finally, $h_J(1) = 1$ while $h'_J(1) = 2(1-\beta-J)$, implying that the concave function $h_{1-\beta}(\cdot)$ is bounded above by $h_{1-\beta}(1) = 1$. Since $h_J(\cdot)$ is monotone increasing in J, the same applies to every $J \in [0,1-\beta]$. $\square$

Proof of Proposition 1: The case of $P = Q$ is trivially settled by taking $\nu(\cdot|x) = \delta_x(\cdot)$. Hence, assume hereafter that $P \neq Q$ and without loss of generality assume also that $a \geq 0$. Let $\tilde{\Omega}$ be such that $P(\tilde{\Omega}) = 1$ and $\frac{dQ}{dP}$ exists on $\tilde{\Omega}$. Set

$$\nu^*(\cdot|x) = \left(1 \wedge \frac{dQ}{dP}(x)\right)\delta_x(\cdot) + \left(1 - \frac{dQ}{dP}(x)\right)_+ \frac{(Q-P)_+(\cdot)}{(Q-P)_+(\Omega)}, \quad x \in \tilde{\Omega}\,, \tag{8}$$

where $(Q-P)_+$ denotes the positive part of the signed measure $Q-P$. Note that $\nu^*(\cdot|x)$ for $x \notin \tilde{\Omega}$ is irrelevant to the proof of Proposition 1. The r.c.p.d. $\nu^*(\cdot|x)$ satisfies (6) since for all $\Gamma \subset \Omega$

$$\begin{aligned}
\int_\Omega \nu^*(\Gamma|x)dP(x) &= \int_{\Gamma\cap\tilde{\Omega}} \left(1 \wedge \frac{dQ}{dP}(x)\right)dP(x) + \frac{(Q-P)_+(\Gamma)}{(Q-P)_+(\Omega)}\int_{\tilde{\Omega}}\left(1 - \frac{dQ}{dP}(x)\right)_+ dP(x) \\
&= P \wedge Q(\Gamma) + \frac{(P-Q)_+(\Omega)}{(Q-P)_+(\Omega)}(Q-P)_+(\Gamma) = Q(\Gamma)\,,
\end{aligned}$$

where $P \wedge Q = P - (P-Q)_+ = Q - (Q-P)_+$. Let $f = \frac{dP}{dQ}$ on $\hat{\Omega}$ and $g = \frac{dQ}{dP}$ on $\tilde{\Omega}$, with $\overline{Q} = Q \circ f^{-1}$ and $\overline{P} = P \circ g^{-1}$ denoting the induced probability measures on $[0,\infty)$. Noting that $fg = 1$ for P

(and Q) a.e. $x \in \hat{\Omega} \cap \tilde{\Omega}$ while $f = 0$ for Q a.e. $x \in \hat{\Omega} \cap \tilde{\Omega}^c$ and $g = 0$ for P a.e. $x \in \hat{\Omega}^c \cap \tilde{\Omega}$, it follows that

$$\int_0^1 u d\overline{P}(u) + \int_0^{1^-} d\overline{Q}(u) = Q(\hat{\Omega}) = 1 \tag{9}$$

and

$$\int_0^1 d\overline{P}(u) + \int_0^{1^-} u d\overline{Q}(u) = P(\tilde{\Omega}) = 1\,. \tag{10}$$

Using the above definitions we see that

$$I(\nu^*) = \int_0^1 h_J(u) d\overline{P}(u) + \int_0^{1^-} v^{2\beta+1} d\overline{Q}(v)$$

with $h_J(\cdot)$ as is in Lemma 1 and

$$J = \frac{\int_0^{1^-} v^\beta (1-v) d\overline{Q}(v)}{\int_0^{1^-} (1-v) d\overline{Q}(v)}\,. \tag{11}$$

In view of Lemma 1, $I(\nu^*) \leq 1$ for every $\overline{Q}$ for which $J \leq 1-\beta$. Fixing $\overline{Q}$ such that $J > (1-\beta)$, the concavity of $h_J(\cdot)$ implies that

$$F(\overline{Q}) \stackrel{\triangle}{=} \sup_{\{\overline{P}:(9) \text{ and } (10) \text{ hold}\}} I(\nu^*) = p h_J(\alpha) + \int_0^{1^-} v^{2\beta+1} d\overline{Q}(v)$$

where

$$\int_0^{1^-} d\overline{Q}(u) = 1 - \alpha p, \quad \int_0^{1^-} u d\overline{Q}(u) = 1 - p\,.$$

Since

$$J = \frac{1}{(1-\alpha)p} \int_0^{1^-} (v^\beta - v^{\beta+1}) d\overline{Q}(v)\,,$$

it follows from Dubbins' theorem [Du62] that suffices in evaluating $\sup_{\overline{Q}} F(\overline{Q})$ to consider atomic $\overline{Q}$ with at most three atoms. Fixing $\alpha, p \in [0,1)$ and the mass $q_i > 0$ of the atoms of $\overline{Q}$ (such that

$\sum_{i=1}^{3} q_i = 1 - \alpha p$), we arrive at

$$F(\overline{Q}) = F(v) = ph_{J(v)}(\alpha) + \sum_{i=1}^{3} v_i^{2\beta+1} q_i$$

where

$$J(v) = \sum_{i=1}^{3} (v_i^{\beta} - v_i^{\beta+1}) q_i / (p - \alpha p)$$

and $v = (v_1, v_2, v_3) \in [0,1)^3$ satisfies the linear constraint $\sum_{i=1}^{3} v_i q_i = 1 - p$.

Note that

$$\frac{1}{q_i} \frac{\partial F}{\partial v_i} = C_{J(v)}(\alpha)(\beta v_i^{\beta-1} - (\beta+1) v_i^{\beta}) + (2\beta+1) v_i^{2\beta}$$

where

$$C_J(\alpha) = 2(\alpha^{1-\beta} + (1+a)(1-\alpha)J) \geq 2\sqrt{h_J(\alpha)}\,.$$

If $C_{J(v)}(\alpha) \leq 2$ then $h_{J(v)}(\alpha) \leq 1$ implying that $F(v) \leq p + \sum_{i=1}^{3} v_i q_i = 1$. Moreover, when $c > 2$ and $0 \leq \beta < 1/2$ the function $v \mapsto c(\beta v^{\beta-1} - (\beta+1) v^{\beta}) + (2\beta+1) v^{2\beta}$ is monotone decreasing on $[0,1]$.

Thus, by Lagrange multipliers, to prove that $\sup_{\substack{\sum_{i=1}^{3} v_i q_i = 1-p \\ 0 \leq v_i < 1}} F(v) \leq 1$ we may assume without loss of generality that $v_1 = 0$ and $v_2 = v_3 = v \in (0,1)$, leading to

$$F(v) = p\, h_J(\alpha) + v^{2\beta+1} q\,,$$

where $J = v^{\beta}(1-v)q/(p - \alpha p)$, $vq = (1-p)$ and $q + \alpha p \leq 1$. Substituting the value of J we see that per given v, q, p, the value of $F(v)$ increases in α, hence the maximum is obtained when $q + \alpha p = 1$. Then, with $p = (1-v)/(1-\alpha v)$ and $q = (1-\alpha)/(1-\alpha v)$ we get $J = v^{\beta}$ and

$$G(v,\alpha) \triangleq \frac{1}{p}(F(v) - 1) = (\alpha^{1-\beta} + (1-\alpha) v^{\beta})^2 + a(1-\alpha)^2 v^{2\beta} - 1 - \frac{1 - v^{2\beta}}{1 - v} v(1-\alpha)$$

Since $\alpha^{1-\beta} + (1-\alpha)(1+a)v^\beta \geq v^\beta$ it follows that

$$\frac{dG}{dv} \geq \frac{(1-\alpha)}{(1-v)^2}[2\beta v^{2\beta-1} + (1-2\beta)v^{2\beta} - 1].$$

By the convexity of $z \mapsto v^z$ it follows that $G(v,\alpha)$ is increasing in v, so that

$$H(\alpha) = \sup_{v\in(0,1)} G(v,\alpha) = G(1,\alpha) = h_1(\alpha) - (1 + 2\beta(1-\alpha)). \quad (12)$$

By Lemma 1, $h_1(\cdot)$ is concave with $h_1(1) = 1, h_1'(1) = -2\beta$. Consequently, (12) implies that $H(\alpha) \leq 0$ for all $\alpha \in [0,1]$. Thus, $F(v) \leq 1$ for every (v,α) and the proof of Proposition 1 is completed. □

Proof of Theorem 2: For ν_j satisfying (3), let

$$G(\nu_1,\nu_2|Q_1,Q_2) = \int_\Omega dP(x)\left[\prod_{j=1}^{2}\int_{\hat{\Omega}_j}\frac{dP}{dQ_j}(y)^\beta d\nu_j(y|x) + a\prod_{j=1}^{2}\int_{\hat{\Omega}_j}\frac{dP}{dQ_j}(y)^\beta \mathbf{1}_{x\neq y}d\nu_j(y|x)\right].$$

By Proposition 1, there exist ν_j satisfying (3) such that $G(\nu_j,\nu_j|Q_j,Q_j) \leq 1$ for every $\beta \in (0,1/2)$ and $a = \beta(1-2\beta)$. Applying the arithmetic-geometric-mean inequality per fixed $x \in \Omega$, we see that $G(\nu_1,\nu_2|Q_1,Q_2) \leq (G(\nu_1,\nu_1|Q_1,Q_1) + G(\nu_2,\nu_2|Q_2,Q_2))/2 \leq 1$. Since $(1+a)^{f_1(x,y,z)} = 1 + a\mathbf{1}_{x\neq y}\mathbf{1}_{x\neq z}$, this proves (4) in the case of $N = 1$.

Suppose now that (4) holds for $N = 1,\ldots,n-1$. Set $N = n$, using the notations $P = P_n = \prod_{k=1}^n P^{(k)}$, $x = (\tilde{x}_n, x_n)$, $y = (\tilde{y}_n, y_n)$, $z = (\tilde{z}_n, z_n)$ and the decomposition $Q_j^{(n)}(\cdot) = Q_j(y_n \in \cdot|\tilde{y}_n)$, $Q_{j,n-1}(\cdot) = Q_j(\tilde{y}_n \in \cdot, \Omega)$ for $j = 1,2$. Since Theorem 2 holds for $N = n-1$ there exist μ_j, $j = 1,2$ such that both

$$\int_{\Omega^{n-1}} dP_{n-1}(\tilde{x}_n)\int\frac{dP_{n-1}}{dQ_{1,n-1}}(\tilde{y}_n)^\beta\frac{dP_{n-1}}{dQ_{2,n-1}}(\tilde{z}_n)^\beta(1+a)^{f_{n-1}(\tilde{x}_n,\tilde{y}_n,\tilde{z}_n)}d\mu_1(\tilde{y}_n|\tilde{x}_n)d\mu_2(\tilde{z}_n|\tilde{x}_n) \leq 1, \quad (13)$$

and

$$\int_{\Omega^{n-1}} \mu_j(\cdot|\tilde{x}_n) dP_{n-1}(\tilde{x}_n) = Q_{j,n-1}(\cdot) \ . \tag{14}$$

Since Theorem 2 also holds for $N = 1$, there exist r.c.p.d. $\nu_j^{(n)}$ which depend upon $\tilde{y}_n, \tilde{z}_n$ and $\tilde{x}_n$ such that almost surely $\mu_j(\cdot|\tilde{x}_n) P_{n-1}(\tilde{x}_n)$ both

$$\int_\Omega dP^{(n)}(x_n) \int \frac{dP^{(n)}}{dQ_1^{(n)}}(y_n)^\beta \frac{dP^{(n)}}{dQ_2^{(n)}}(z_n)^\beta (1+a)^{f_1(x_n,y_n,z_n)} d\nu_1^{(n)}(y_n|x_n) d\nu_2^{(n)}(z_n|x_n) \leq 1 \ , \tag{15}$$

and

$$\int_\Omega \nu_j^{(n)}(\cdot|x_n) dP^{(n)}(x_n) = Q_j^{(n)}(\cdot) \ . \tag{16}$$

Note that (14) and (16) imply that the r.c.p.d. $\nu_j(y|x) = \nu_j^{(n)}(y_n|x_n)\mu_j(\tilde{y}_n|\tilde{x}_n)$ satisfy (3) for $j = 1, 2$. Since $P_n = P^{(n)} \times P_{n-1}$, by (13) and (15) also

$$\int_{\Omega^n} dP_n(x) \int \frac{dP_n}{dQ_1}(y)^\beta \frac{dP_n}{dQ_2}(z)^\beta (1+a)^{f_n(x,y,z)} d\nu_1(y|x) d\nu_2(z|x) =$$

$$\int_{\Omega^{n-1}} dP_{n-1}(\tilde{x}_n) \int \frac{dP_{n-1}}{dQ_{1,n-1}}(\tilde{y}_n)^\beta \frac{dP_{n-1}}{dQ_{2,n-1}}(\tilde{z}_n)^\beta (1+a)^{f_{n-1}(\tilde{x}_n,\tilde{y}_n,\tilde{z}_n)} d\mu_1(\tilde{y}_n|\tilde{x}_n) d\mu_2(\tilde{z}_n|\tilde{x}_n) \cdot$$

$$\int_\Omega dP^{(n)}(x_n) \int \frac{dP^{(n)}}{dQ_1^{(n)}}(y_n)^\beta \frac{dP^{(n)}}{dQ_2^{(n)}}(z_n)^\beta (1+a)^{f_1(x_n,y_n,z_n)} d\nu_1^{(n)}(y_n|x_n) d\nu_2^{(n)}(z_n|x_n) \leq 1 \ .$$

Thus, by induction, (4) holds for all $N \geq 1$. □

References

[BHJ92] Barbour A. D., Holst L. and Janson S. (1992) *Poisson approximation.* Oxford University Press, Oxford.

[BL96] Bobkov S. and Ledoux M. (1996) *Poincaré's inequalities and Talagrand's concentration phenomenon for the exponential distribution.* Preprint.

[De96] Dembo A. (1996) *Information inequalities and concentration of measure.* to appear in Ann. Probab.

[Du62] Dubins L. E. (1962) *On extreme points of convex sets.* J. Math. Anal. Appl. 5, 237–244.

[Led96] Ledoux M. (1996) *On Talagrand's deviation inequalities for product measures.* ESAIM: Probability and Statistics, 1, pp. 63–87.

[Mar86] Marton K. (1986) *A simple proof of the blowing-up Lemma* IEEE Trans. on Information Theory IT-32, 445-446.

[Mar96a] Marton K. (1996) *Bounding $\bar{d}$-distance by information divergence: a method to prove measure concentration.* to appear in the Ann. Probab.

[Mar96b] Marton K. (1996) *A measure concentration inequality for contracting Markov chains.* GAFA, 6, pp. 556–571.

[Tal95] Talagrand, M. (1995), *Concentration of measure and isoperimetric inequalities in product spaces.* Publications Mathématiques de l'I.H.E.S., 81, 73-205.

[Tal96a] Talagrand, M. (1996), *A new look at independence.* Ann. Probab. 24, pp. 1–34.

[Tal96b] Talagrand, M. (1996), *New concentration inequalities in product spaces.* to appear in Invent. Math.

[Tal96c] Talagrand, M. (1996), *Transportation cost for Gaussian and other product measures.* GAFA, 6, pp. 587–600.