MIT OpenCourseWare
http://ocw.mit.edu

18.175 Theory of Probability
Fall 2008

For information about citing these materials or our Terms of Use, visit: http://ocw.mit.edu/terms.

Section 19

Strassen's Theorem. Relationships between metrics.

Metric for convergence in probability. Let $(\Omega, \mathcal{B}, \mathbb{P})$ be a probability space, (S, d) - a metric space and $X, Y : \Omega \to S$ - random variables with values in S. The quantity

$$\alpha(X, Y) = \inf\{\varepsilon \geq 0 : \mathbb{P}(d(X, Y) > \varepsilon) \leq \varepsilon\}$$

is called the *Ky Fan metric* on the set $\mathcal{L}^0(\Omega, S)$ of classes of equivalences of such random variables, where two r.v.s are equivalent if they are equal a.s. If we take a sequence

$$\varepsilon_k \downarrow \alpha = \alpha(X, Y)$$

then $\mathbb{P}(d(X, Y) > \varepsilon_k) \leq \varepsilon_k$ and since

$$\mathrm{I}(d(X, Y) > \varepsilon_k) \uparrow \mathrm{I}(d(X, Y) > \alpha),$$

by monotone convergence theorem, $\mathbb{P}(d(X, Y) > \alpha) \leq \alpha$. Thus, the infimum in the definition of $\alpha(X, Y)$ is attained.

Lemma 37 *α is a metric on $\mathcal{L}^0(\Omega, S)$ which metrizes convergence in probability.*

Proof. First of all, clearly, $\alpha(X, Y) = 0$ iff $X = Y$ almost surely. To prove the triangle inequality,

$$\begin{aligned}\mathbb{P}(d(X, Z) > \alpha(X, Y) + \alpha(Y, Z)) &\leq \mathbb{P}(d(X, Y) > \alpha(X, Y)) + \mathbb{P}(d(Y, Z) > \alpha(Y, Z)) \\ &\leq \alpha(Y, Z) + \alpha(Y, Z)\end{aligned}$$

so that $\alpha(X, Z) \leq \alpha(X, Y) + \alpha(Y, Z)$. This proves that α is a metric. Next, if $\alpha_n = \alpha(X_n, X) \to 0$ then for any $\varepsilon > 0$ and large enough n such that $\alpha_n < \varepsilon$,

$$\mathbb{P}(d(X_n, X) > \varepsilon) \leq \mathbb{P}(d(X_n, X) > \alpha_n) \leq \alpha_n \to 0.$$

Conversely, if $X_n \to X$ in probability then for any $m \geq 1$ and large enough $n \geq n(m)$,

$$\mathbb{P}\Big(d(X_n, X) > \frac{1}{m}\Big) \leq \frac{1}{m}$$

which means that $\alpha_n \leq 1/m$ so that $\alpha_n \to 0$.

□

Lemma 38 *For $X, Y \in \mathcal{L}^0(\Omega, S)$, the Levy-Prohorov metric ρ satisfies*

$$\rho(\mathcal{L}(X), \mathcal{L}(Y)) \leq \alpha(X, Y).$$

Proof. Take $\varepsilon > \alpha(X,Y)$ so that $\mathbb{P}(d(X,Y) \geq \varepsilon) \leq \varepsilon$. For any set $A \subseteq S$,

$$\mathbb{P}(X \in A) = \mathbb{P}(X \in A, d(X,Y) < \varepsilon) + \mathbb{P}(X \in A, d(X,Y) \geq \varepsilon) \leq \mathbb{P}(Y \in A^\varepsilon) + \varepsilon$$

which means that $\rho(\mathcal{L}(X), \mathcal{L}(Y)) \leq \varepsilon$. Letting $\varepsilon \downarrow \alpha(X,Y)$ proves the result.

□

We will now prove that, in some sense, the opposite is also true. Let (S,d) be a metric space and $\mathbb{P}, \mathbb{Q}$ be probability laws on S. Suppose that these laws are close in the Levy-Prohorov metric ρ. Can we construct random variables s_1 and s_2, with laws $\mathbb{P}$ and $\mathbb{Q}$, that are define on the same probability space and are close to each other in the Ky Fan metric α? We will construct a distribution on the product space $S \times S$ such that the coordinates s_1 and s_2 have marginal distributions $\mathbb{P}$ and $\mathbb{Q}$ and the distribution is concentrated in the neighborhood of the diagonal $s_1 = s_2$, where s_1 and s_2 are close in metric d, and the size of the neighborhood is controlled by $\rho(\mathbb{P}, \mathbb{Q})$.

Consider two sets X and Y. Given a subset $K \subseteq X \times Y$ and $A \subseteq X$ we define a K-image of A by

$$A^K = \{y \in Y : \exists x \in A, (x,y) \in K\}.$$

A K-*matching* f of X into Y is a one-to-one function $f : X \to Y$ such that $(x, f(x)) \in K$. We will need the following well known matching theorem.

Theorem 45 *If X, Y are finite and for all $A \subseteq X$,*

$$\text{card}(A^K) \geq \text{card}(A) \tag{19.0.1}$$

then there exists a K-matching f of X into Y.

Proof. We will prove the result by induction on $m = \text{card}(X)$. The case of $m = 1$ is obvious. For each $x \in X$ there exists $y \in Y$ such that $(x,y) \in K$. If there is a matching f of $X \setminus \{x\}$ into $Y \setminus \{y\}$ then defining $f(x) = y$ extends f to X. If not, then since $\text{card}(X \setminus \{x\}) < m$, by induction assumption, condition (19.0.1) is violated, i.e. there exists a set $A \subseteq X \setminus \{x\}$ such that $\text{card}(A^K \setminus \{y\}) < \text{card}(A)$. But because we also know that $\text{card}(A^K) \geq \text{card}(A)$ this implies that $\text{card}(A^K) = \text{card}(A)$. Since $\text{card}(A) < m$, by induction there exists a matching of A onto A^K. If there is a matching of $X \setminus A$ into $Y \setminus A^K$ we can combine it with a matching of A and A^K. If not, again by induction assumption, there exists $D \in X \setminus A$ such that $\text{card}(D^K \setminus A^K) < \text{card}(D)$. But then

$$\text{card}\Big((A \cup D)^K\Big) = \text{card}(D^K \setminus A^K) + \text{card}(A^K) < \text{card}(D) + \text{card}(A) = \text{card}(D \cup A),$$

which contradicts the assumption (19.0.1).

□

Theorem 46 *(Strassen) Suppose that (S,d) is a separable metric space and $\alpha, \beta > 0$. Suppose that laws $\mathbb{P}$ and $\mathbb{Q}$ are such that for all measurable sets $F \subseteq S$,*

$$\mathbb{P}(F) \leq \mathbb{Q}(F^\alpha) + \beta \tag{19.0.2}$$

Then for any $\varepsilon > 0$ there exist two non-negative measures η, γ on $S \times S$ such that

1. *$\mu = \eta + \gamma$ is a law on $S \times S$ with marginals $\mathbb{P}$ and $\mathbb{Q}$.*
2. *$\eta(d(x,y) > \alpha + \varepsilon) = 0$.*
3. *$\gamma(S \times S) \leq \beta + \varepsilon$.*
4. *μ is a finite sum of product measures.*

Remark. Condition (19.0.2) is a relaxation of the definition of the Levy-Prohorov metric, one can take any $\alpha, \beta > \rho(\mathbb{P}, \mathbb{Q})$. Conditions 1 - 3 mean that we can construct a measure μ on $S \times S$ such that coordinates x, y have marginal distributions $\mathbb{P}, \mathbb{Q}$, concentrated within distance $\alpha + \varepsilon$ of each other (condition 2) except for the set of measure at most $\beta + \varepsilon$ (condition 3).

Proof. The proof will proceed in several steps.

Case A. We will start with the simplest case which is, however, at the core of everything else. Given small $\varepsilon > 0$, take $n \geq 1$ such that $n\varepsilon > 1$. Suppose that laws $\mathbb{P}, \mathbb{Q}$ are uniform on finite subsets $M, N \subseteq S$ of equal cardinality,

$$\text{card}(M) = \text{card}(N) = n, \ \ \mathbb{P}(x) = \mathbb{Q}(y) = \frac{1}{n} < \varepsilon, \ \ x \in M, y \in N.$$

Using condition (19.0.2), we would like to match as many points from M and N as possible, but only points that are within distance α from each other. To use the matching theorem, we will introduce some auxiliary sets U and V that are not too big, with size controlled by parameter β, and the union of these sets with M and N satisfies a certain matching condition.

Take integer k such that $\beta n \leq k < (\beta + \varepsilon)n$. Let us take sets U and V such that $k = \text{card}(U) = \text{card}(V)$ and U, V are disjoint from M, N. Define

$$X = M \cup U, \ \ Y = N \cup V.$$

Let us define a subset $K \subseteq X \times Y$ such that $(x, y) \in K$ if and only if one of the following holds:

1. $x \in U$,
2. $y \in V$,
3. $d(x, y) \leq \alpha$ if $x \in M, y \in N$.

This means that small auxiliary sets can be matched with any points but only close points, $d(x, y) \leq \alpha$, can be matched in the main sets M and N. Consider a set $A \subseteq X$ with cardinality $\text{card}(A) = r$. If $A \not\subseteq M$ then by 1, $A^K = Y$ and $\text{card}(A^K) \geq r$. Suppose now that $A \subseteq M$ and we would like to show that again $\text{card}(A^K) \geq r$. By (19.0.2),

$$\frac{r}{n} = \mathbb{P}(A) \leq \mathbb{Q}(A^\alpha) + \beta = \frac{1}{n}\text{card}(A^\alpha \cap N) + \beta \leq \frac{1}{n}\text{card}(A^K \cap N) + \beta$$

since by 3, $A^\alpha \subseteq A^K$. Therefore,

$$r = \text{card}(A) \leq n\beta + \text{card}(A^K \cap N) \leq k + \text{card}(A^K \cap N) = \text{card}(A^K),$$

since $k = \text{card}(V)$ and $A^K = V \cup (A^K \cap N)$. By matching theorem, there exists a K-matching f of X and Y. Let

$$T = \{x \in M : f(x) \in N\},$$

i.e. close points, $d(x, y) \leq \alpha$, from M that are matched with points in N. Clearly, $\text{card}(T) \geq n - k$ and for $x \in T$, by 3, $d(x, f(x)) \leq \alpha$. For $x \in M \setminus T$, redefine $f(x)$ to match x with arbitrary points in N that are not matched with points in T. This defines a matching of M onto N. We define measures η and γ by

$$\eta = \frac{1}{n}\sum_{x \in T} \delta(x, f(x)), \ \ \gamma = \frac{1}{n}\sum_{x \in M \setminus T} \delta(x, f(x)),$$

and let $\mu = \eta + \gamma$. First of all, obviously, μ has marginals $\mathbb{P}$ and $\mathbb{Q}$ because each point in M or N appears in the sum $\eta + \gamma$ only once with weight $1/n$. Also,

$$\eta(d(x, f(x)) > \alpha) = 0, \ \ \gamma(S \times S) \leq \frac{\text{card}(M \setminus T)}{n} \leq \frac{k}{n} < \beta + \varepsilon. \tag{19.0.3}$$

Finally, both η and γ are finite sums of point masses which are product measures of point masses.

Case B. Suppose now that $\mathbb{P}$ and $\mathbb{Q}$ are concentrated on finitely many points with rational probabilities. Then we can artificially split all points into "smaller" points of equal probabilities as follows. Let n be such that $n\varepsilon > 1$ and

$$n\mathbb{P}(x), n\mathbb{Q}(x) \in J = \{1, 2, \ldots, n\}.$$

Define a discrete metric on J by $f(i,j) = \varepsilon\,\mathrm{I}(i \neq j)$ and define a metric on $S \times J$ by

$$e\big((x,i),(y,j)\big) = d(x,y) + f(i,j).$$

Define a measure $\mathbb{P}'$ on $S \times J$ as follows. If $\mathbb{P}(x) = \frac{j}{n}$ then

$$\mathbb{P}'\big((x,i)\big) = \frac{1}{n} \quad \text{for} \quad i = 1, \ldots, j.$$

Define $\mathbb{Q}'$ similarly. Let us check that laws $\mathbb{P}', \mathbb{Q}'$ satisfy the assumptions of Case A. Given a set $F \subseteq S \times J$, define

$$F_1 = \{x \in S : (x,j) \in F \text{ for some } j\}.$$

Using (19.0.2),

$$\mathbb{P}'(F) \leq \mathbb{P}(F_1) \leq \mathbb{Q}(F_1^{\alpha}) + \beta \leq \mathbb{Q}'(F^{\alpha+\varepsilon}) + \beta,$$

because $f(i,j) \leq \varepsilon$. By Case A in (19.0.3), we can construct $\mu' = \eta' + \gamma'$ with marginals $\mathbb{P}'$ and $\mathbb{Q}'$ such that

$$\eta'\big(e((x,i),(y,j)) > \alpha + \varepsilon\big) = 0, \;\; \gamma'\big((S \times J) \times (S \times J)\big) < \beta + \varepsilon.$$

Let μ, η, γ be the projections of μ', η', γ' back onto $S \times S$ by the map $((x,i),(y,j)) \to (x,y)$. Then, clearly, $\mu = \eta + \gamma$, μ has marginals $\mathbb{P}$ and $\mathbb{Q}$ and $\gamma(S \times S) < \beta + \varepsilon$. Finally, since

$$e\big((x,i),(y,j)\big) = d(x,y) + f(i,j) \geq d(x,y),$$

we get

$$\eta\big(d(x,y) > \alpha + \varepsilon\big) \leq \eta'\big(e((x,i),(y,j)) > \alpha + \varepsilon\big) = 0.$$

Case C. (General case) Let $\mathbb{P}, \mathbb{Q}$ be the laws on a separable metric space (S,d). Let A be a maximal set such that for all $x, y \in A, d(x,y) \geq \varepsilon$. The set A is countable, $A = \{x_i\}_{i \geq 1}$, because S is separable, and since A is maximal, for all $x \in S$ there exists $y \in A$ such that $d(x,y) < \varepsilon$. Such set A is usually called an ε-packing. Let us create a partition of S using ε-balls around $\{x_i\}$:

$$B_1 = \{x \in S : d(x, x_1) < \varepsilon\}, \;\; B_2 = \{d(x, x_2) < \varepsilon\} \setminus B_1$$

and, iteratively for $k \geq 2$,

$$B_k = \{d(x, x_k) < \varepsilon\} \setminus (B_1 \cup \cdots \cup B_{k-1}).$$

$\{B_k\}_{k \geq 1}$ is a partition of S. Let us discretize measures $\mathbb{P}$ and $\mathbb{Q}$ by projecting them onto $\{x_i\}_{i \geq 1}$:

$$\mathbb{P}'(x_k) = \mathbb{P}(B_k), \;\; \mathbb{Q}'(x_k) = \mathbb{Q}(B_k).$$

Consider any set $F \subseteq S$. For any point $x \in F$, if $x \in B_k$ then $d(x, x_k) < \varepsilon$, i.e. $x_k \in F^{\varepsilon}$ and, therefore,

$$\mathbb{P}(F) \leq \mathbb{P}'(F^{\varepsilon}).$$

Also, if $x_k \in F$ then $B_k \subseteq F^{\varepsilon}$ and, therefore,

$$\mathbb{P}'(F) \leq \mathbb{P}(F^{\varepsilon}).$$

To apply Case B, we need to approximate $\mathbb{P}'$ by a measure on a finite number of points with rationals probabilities. For large enough $n \geq 1$, let

$$\mathbb{P}''(x_k) = \frac{\lfloor n\mathbb{P}'(x_k) \rfloor}{n}.$$

Clearly, as $n \to \infty$, $\mathbb{P}''(x_k) \uparrow \mathbb{P}'(x_k)$. Since only a finite number of points carry non-zero weights $\mathbb{P}''(x_k) > 0$, let x_0 be one of the other points in the sequence $\{x_k\}$. Let us assign to it a probability

$$\mathbb{P}''(x_0) = 1 - \sum_{k \geq 1} \mathbb{P}''(x_k).$$

If we take n large enough so that $\mathbb{P}''(x_0) < \varepsilon/2$ then

$$\sum_{k \geq 0} |\mathbb{P}''(x_k) - \mathbb{P}'(x_k)| \leq \varepsilon.$$

All the relations above also hold true for $\mathbb{Q}, \mathbb{Q}'$ and $\mathbb{Q}''$ that are defined similarly. We can write for $F \subseteq S$

$$\mathbb{P}''(F) \leq \mathbb{P}'(F) + \varepsilon \leq \mathbb{P}(F^{\varepsilon}) + \varepsilon \leq \mathbb{Q}(F^{\varepsilon+\alpha}) + \beta + \varepsilon \leq \mathbb{Q}'(F^{\alpha+2\varepsilon}) + \beta + \varepsilon \leq \mathbb{Q}''(F^{\alpha+2\varepsilon}) + \beta + 2\varepsilon.$$

By Case B, there exists a decomposition $\mu'' = \eta'' + \gamma''$ on $S \times S$ with marginals $\mathbb{P}''$ and $\mathbb{Q}''$ such that

$$\eta''\big(d(x,y) > \alpha + 3\varepsilon\big) = 0, \ \ \gamma''(S \times S) \leq \beta + 3\varepsilon.$$

Let us also assume that the points (x_0, x_i) and (x_i, x_0) for $i \geq 0$ are included in the support of γ''. Since the total weight of these points is at most ε, the total weight of γ'' does no increase much:

$$\gamma''(S \times S) \leq \beta + 5\varepsilon.$$

It remains to redistribute these measures from sequence $\{x_i\}_{i \geq 0}$ to S in a way that recovers marginal distributions $\mathbb{P}$ and $\mathbb{Q}$ and so that not much accuracy is lost. Define a sequence of measures on S by

$$\mathbb{P}_i(C) = \frac{\mathbb{P}(CB_i)}{\mathbb{P}(B_i)} \ \text{ if } \mathbb{P}(B_i) > 0 \text{ and } \mathbb{P}_i(C) = 0 \text{ otherwise}$$

and define $\mathbb{Q}_i$ similarly. The measures $\mathbb{P}_i$ and $\mathbb{Q}_i$ are concentrated on B_i. Define

$$\eta = \sum_{i,j \geq 1} \eta''(x_i, x_j)(\mathbb{P}_i \times \mathbb{Q}_j)$$

The marginals of η satisfy

$$\begin{aligned} u(C) = \eta(C \times S) \quad &\leq \quad \sum_{i,j \geq 1} \eta''(x_i, x_j)\mathbb{P}_i(C) = \sum_{i \geq 1} \eta''(x_i, S)\mathbb{P}_i(C) \\ &\leq \quad \sum_{i \geq 1} \mathbb{P}''(x_i)\mathbb{P}_i(C) \leq \sum_{i \geq 1} \mathbb{P}'(x_i)\mathbb{P}_i(C) = \sum_{i \geq 1} \mathbb{P}(B_i)\mathbb{P}_i(C) = \mathbb{P}(C) \end{aligned}$$

and, similarly,

$$v(C) = \eta(S \times C) \leq \mathbb{Q}(C).$$

Since $\eta''(x_i, x_j) = 0$ unless $d(x_i, x_j) \leq \alpha + 3\varepsilon$, the measure

$$\eta = \sum_{i,j \geq 1} \eta''(x_i, x_j)(\mathbb{P}_i \times \mathbb{Q}_j)$$

is concentrated on the set $\{d(x,y) \leq \alpha + 5\varepsilon\}$ because for $x \in B_i, y \in B_j$,

$$d(x,y) \leq d(x, x_i) + d(x_i, x_j) + d(x_j, y) \leq \varepsilon + \alpha + 3\varepsilon + \varepsilon = \alpha + 5\varepsilon.$$

If $u(S) = v(S) = 1$ then $\eta(S \times S) = 1$ and η has marginals $\mathbb{P}$ and $\mathbb{Q}$ so we can take $\gamma = 0$. Otherwise, take $t = 1 - u(S)$ and define

$$\gamma = \frac{1}{t}(\mathbb{P} - u) \times (\mathbb{Q} - v).$$

It is easy to check that $\mu = \eta + \gamma$ has marginals $\mathbb{P}$ and $\mathbb{Q}$. Also,

$$\gamma(S \times S) = t = 1 - \eta(S \times S) = 1 - \eta''(S \times S) = \gamma''(S \times S) \leq \beta + 5\varepsilon.$$

□

Relationships between metrics. The following relationship between Ky Fan and Levy-Prohorov metrics is an immediate consequence of Strassen's theorem. We already saw that $\rho(\mathcal{L}(X), \mathcal{L}(Y)) \leq \alpha(X, Y)$.

Theorem 47 *If (S, d) is a separable metric space and $\mathbb{P}, \mathbb{Q}$ are laws on S then for any $\varepsilon > 0$ there exist random variables X and Y with distributions $\mathcal{L}(X) = \mathbb{P}$ and $\mathcal{L}(Y) = \mathbb{Q}$ such that*

$$\alpha(X, Y) \leq \rho(\mathbb{P}, \mathbb{Q}) + \varepsilon.$$

If $\mathbb{P}$ and $\mathbb{Q}$ are tight, one can take $\varepsilon = 0$.

Proof. Let us take $\alpha = \beta = \rho(\mathbb{P}, \mathbb{Q})$. Then, by definition of the Levy-Prohorov metric, for any $\varepsilon > 0$ and for any set A,

$$\mathbb{P}(A) \leq \mathbb{Q}(A^{\rho+\varepsilon}) + \rho + \varepsilon.$$

By Strassen's theorem, there exists a measure μ on $S \times S$ with marginals $\mathbb{P}, \mathbb{Q}$ such that

$$\mu\big(d(x, y) > \rho + 2\varepsilon\big) \leq \rho + 2\varepsilon. \tag{19.0.4}$$

Therefore, if X and Y are the coordinates of $S \times S$, i.e.

$$X, Y : S \times S \to S, \;\; X(x, y) = x, \;\; Y(x, y) = y,$$

then by definition of the Ky Fan metric, $\alpha(X, Y) \leq \rho + 2\varepsilon$. If $\mathbb{P}$ and $\mathbb{Q}$ are tight then there exists a compact K such that $\mathbb{P}(K), \mathbb{Q}(K) \geq 1 - \delta$. For $\varepsilon = 1/n$ find μ_n as in (19.0.4). Since μ_n has marginals $\mathbb{P}$ and $\mathbb{Q}$, $\mu_n(K \times K) \geq 1 - 2\delta$, which means that $(\mu_n)_{n \geq 1}$ are uniformly tight. By selection theorem, there exists a convergent subsequence $\mu_{n(k)} \to \mu$. Obviously, μ has marginals $\mathbb{P}$ and $\mathbb{Q}$. Since by construction,

$$\mu_n\Big(d(x, y) > \rho + \frac{2}{n}\Big) \leq \rho + \frac{2}{n}$$

and $\{d(x, y) > \rho + 2/n\}$ is an open set on $S \times S$, by portmanteau theorem,

$$\mu\Big(d(x, y) > \rho + \frac{2}{n}\Big) \leq \liminf_{k \to \infty} \mu_{n(k)}\Big(d(x, y) > \rho + \frac{2}{n(k)}\Big) \leq \rho.$$

Letting $n \to \infty$ we get $\mu(d(x, y) > \rho) \leq \rho$ and, therefore, $\alpha(X, Y) \leq \rho$.

□

This also implies the relationship between the Bounded Lipschitz metric β and Levy-Prohorov metric ρ.

Lemma 39 *If (S, d) is a separable metric space then*

$$\frac{1}{2}\beta(\mathbb{P}, \mathbb{Q}) \leq \rho(\mathbb{P}, \mathbb{Q}) \leq 2\sqrt{\beta(\mathbb{P}, \mathbb{Q})}.$$

Proof. We already proved the second inequality. To prove the first one, given $\varepsilon > 0$ take random variables X and Y such that $\alpha(X, Y) \leq \rho + \varepsilon$. Consider a bounded Lipschitz function f, $||f||_{\mathrm{BL}} < \infty$. Then

$$\begin{aligned} \Big|\int f d\mathbb{P} - \int f d\mathbb{Q}\Big| &= |\mathbb{E}f(X) - \mathbb{E}f(Y)| \leq \mathbb{E}|f(X) - f(Y)| \\ &\leq \|f\|_{\mathrm{L}}(\rho + \varepsilon) + 2\|f\|_\infty \mathbb{P}\Big(d(X, Y) > \rho + \varepsilon\Big) \\ &\leq \|f\|_{\mathrm{L}}(\rho + \varepsilon) + 2\|f\|_\infty(\rho + \varepsilon) \leq 2\|f\|_{\mathrm{BL}}(\rho + \varepsilon). \end{aligned}$$

Thus, $\beta(\mathbb{P}, \mathbb{Q}) \leq 2(\rho(\mathbb{P}, \mathbb{Q}) + \varepsilon)$ and letting $\varepsilon \to 0$ finishes the proof.

□