SECTION 5

Stability

Oftentimes an interesting process can be put together from simpler processes, to which the combinatorial methods of Section 4 apply directly. The question then becomes one of stability: Does the process inherit the nice properties from its component pieces? This section provides some answers for the case of processes $\boldsymbol{\sigma} \cdot \mathbf{f}$ indexed by subsets of Euclidean space.

Throughout the section $\mathcal{F}$ and $\mathcal{G}$ will be fixed subsets of $\mathbb{R}^n$, with envelopes $\mathbf{F}$ and $\mathbf{G}$ and $\boldsymbol{\sigma} = (\sigma_1, \ldots, \sigma_n)$ will be a vector of independent random variables, each taking the values ± 1 with probability $1/2$. In particular, $\boldsymbol{\sigma}$ will be regarded as the generic point in the set $\mathcal{S}$ of all n-tuples of ± 1's, under its uniform distribution $\mathbb{P}_\sigma$. The problem is to determine which properties of $\mathcal{F}$ and $\mathcal{G}$ are inherited by classes such as

$$\begin{aligned}
\mathcal{F} \oplus \mathcal{G} &= \{\mathbf{f} + \mathbf{g} : \mathbf{f} \in \mathcal{F}, \mathbf{g} \in \mathcal{G}\}, \\
\mathcal{F} \vee \mathcal{G} &= \{\mathbf{f} \vee \mathbf{g} : \mathbf{f} \in \mathcal{F}, \mathbf{g} \in \mathcal{G}\}, \\
\mathcal{F} \wedge \mathcal{G} &= \{\mathbf{f} \wedge \mathbf{g} : \mathbf{f} \in \mathcal{F}, \mathbf{g} \in \mathcal{G}\}, \\
\mathcal{F} \odot \mathcal{G} &= \{\mathbf{f} \odot \mathbf{g} : \mathbf{f} \in \mathcal{F}, \mathbf{g} \in \mathcal{G}\}.
\end{aligned}$$

The reader might want to skip the material in the subsection headed "General Maximal Inequalities". It is included in this section merely to illustrate one of the more recent developments in the subject; it is based on the paper by Ledoux and Talagrand (1989). For most applications to asymptotic problems, the simpler results contained in the first two subsections seem to suffice.

Pseudodimension. This property is stable only for the formation of unions, pointwise maxima, and pointwise minima.

Suppose that both $\mathcal{F}$ and $\mathcal{G}$ have pseudodimension at most V. Then, for every $\mathbf{t}$ in $\mathbb{R}^k$ and every k less than n, Lemma 4.6 asserts that the projection of $\mathcal{F}$ can occupy at most

$$m = \binom{k}{0} + \cdots + \binom{k}{V}$$

of the orthants around $\mathbf{t}$, and similarly for $\mathcal{G}$. For any two vectors $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ in $\mathbb{R}^k$, the orthants of $\mathbf{t}$ occupied by $\boldsymbol{\alpha} \vee \boldsymbol{\beta}$ and $\boldsymbol{\alpha} \wedge \boldsymbol{\beta}$ are uniquely determined by the orthants occupied by $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$. (The same cannot be said for $\boldsymbol{\alpha} + \boldsymbol{\beta}$ or $\boldsymbol{\alpha} \odot \boldsymbol{\beta}$.) Thus the projections of $\mathcal{F} \vee \mathcal{G}$ and $\mathcal{F} \wedge \mathcal{G}$ each occupy at most m^2 different orthants. It is even easier to show that the union $\mathcal{F} \cup \mathcal{G}$ occupies at most $2m$ orthants. If k could be chosen so that $m^2 < 2^k$, this would imply that none of the projections surrounds $\mathbf{t}$. So, we need to find a k such that

$$\left[\binom{k}{0} + \cdots + \binom{k}{V}\right]^2 < 2^k.$$

On the left-hand side we have a polynomial of degree $2V$, which increases much more slowly with k than the 2^k on the right-hand side. For k large enough the inequality will be satisfied. Just knowing that such a k exists is good enough for most applications, but, for the sake of having an explicit bound, let us determine how k depends on V.

Introduce a random variable X with a $\text{Bin}(k, 1/2)$ distribution. The desired inequality is equivalent to

$$\Big[\mathbb{P}\{X \geq k - V\}\Big]^2 < 2^{-k}.$$

Bound the left-hand side by

$$\Big[9^{-(k-V)}\mathbb{P}9^X\Big]^2 = 81^{-(k-V)}25^k,$$

then choose $k = 10V$ to make the bound less than 2^{-k} for every V. [It is possible to replace 10 by a smaller constant, but this has no advantage for our purposes.]

(5.1) Lemma. *If both $\mathcal{F}$ and $\mathcal{G}$ have pseudodimension at most V, then all of $\mathcal{F} \cup \mathcal{G}$ and $\mathcal{F} \vee \mathcal{G}$ and $\mathcal{F} \wedge \mathcal{G}$ have pseudodimension less than* $10V$. □

Unfortunately neither sums nor products share this form of stability.

Packing Numbers. Stability properties for packing or covering numbers follow easily from the triangle inequality: we construct approximating subclasses $\{\mathbf{f}_i\}$ for $\mathcal{F}$ and $\{\mathbf{g}_j\}$ for $\mathcal{G}$, and then argue from inequalities such as

$$|\mathbf{f} \vee \mathbf{g} - \mathbf{f}_i \vee \mathbf{g}_j|_2 \leq |\mathbf{f} - \mathbf{f}_i|_2 + |\mathbf{g} - \mathbf{g}_j|_2.$$

In this way we get covering number bounds

$$N_2(\epsilon + \delta, \mathcal{F} \square \mathcal{G}) \leq N_2(\epsilon, \mathcal{F}) N_2(\delta, \mathcal{G}),$$

where $\square$ stands for either $+$ or $\vee$ or $\wedge$. The corresponding bounds for packing numbers,

$$D_2(2\epsilon + 2\delta, \mathcal{F} \square \mathcal{G}) \leq D_2(\epsilon, \mathcal{F}) D_2(\delta, \mathcal{G}),$$

follow from the inequalities that relate packing to covering. An even easier argument would establish a stability property for the packing numbers for the union $\mathcal{F} \cup \mathcal{G}$.

Pointwise products are more interesting, for here we need the flexibility of bounds valid for arbitrary rescaling vectors. Let us show that the covering numbers for the set $\mathcal{F} \odot \mathcal{G}$ of all pairwise products $\mathbf{f} \odot \mathbf{g}$ satisfy the inequality

$$N_2(\epsilon + \delta, \boldsymbol{\alpha} \odot \mathcal{F} \odot \mathcal{G}) \leq N_2(\epsilon, \boldsymbol{\alpha} \odot \mathbf{G} \odot \mathcal{F})\, N_2(\delta, \boldsymbol{\alpha} \odot \mathbf{F} \odot \mathcal{G}), \tag{5.2}$$

which implies the corresponding inequality for packing numbers

$$D_2(2\epsilon + 2\delta, \boldsymbol{\alpha} \odot \mathcal{F} \odot \mathcal{G}) \leq D_2(\epsilon, \boldsymbol{\alpha} \odot \mathbf{G} \odot \mathcal{F})\, D_2(\delta, \boldsymbol{\alpha} \odot \mathbf{F} \odot \mathcal{G}).$$

Choose approximating points $\boldsymbol{\alpha} \odot \mathbf{G} \odot \mathbf{f}_1, \ldots, \boldsymbol{\alpha} \odot \mathbf{G} \odot \mathbf{f}_r$ for $\boldsymbol{\alpha} \odot \mathbf{G} \odot \mathcal{F}$, and points $\boldsymbol{\alpha} \odot \mathbf{F} \odot \mathbf{g}_1, \ldots, \boldsymbol{\alpha} \odot \mathbf{F} \odot \mathbf{g}_s$ for $\mathbf{F} \odot \mathcal{G}$. We may assume each $\mathbf{f}_i$ lies within the box defined by the envelope $\mathbf{F}$, and each $\mathbf{g}_j$ lies within the box defined by $\mathbf{G}$. For an $\boldsymbol{\alpha} \odot \mathbf{f} \odot \mathbf{g}$ in the set $\boldsymbol{\alpha} \odot \mathcal{F} \odot \mathcal{G}$, and appropriate $\mathbf{f}_i$ and $\mathbf{g}_j$,

$$\begin{aligned}
&|\boldsymbol{\alpha} \odot \mathbf{f} \odot \mathbf{g} - \boldsymbol{\alpha} \odot \mathbf{f}_i \odot \mathbf{g}_j|_2 \\
&\qquad \leq |\boldsymbol{\alpha} \odot \mathbf{f} \odot \mathbf{g} - \boldsymbol{\alpha} \odot \mathbf{f}_i \odot \mathbf{g}|_2 + |\boldsymbol{\alpha} \odot \mathbf{f}_i \odot \mathbf{g} - \boldsymbol{\alpha} \odot \mathbf{f}_i \odot \mathbf{g}_j|_2 \\
&\qquad \leq |\boldsymbol{\alpha} \odot \mathbf{f} \odot \mathbf{G} - \boldsymbol{\alpha} \odot \mathbf{f}_i \odot \mathbf{G}|_2 + |\boldsymbol{\alpha} \odot \mathbf{F} \odot \mathbf{g} - \boldsymbol{\alpha} \odot \mathbf{F} \odot \mathbf{g}_j|_2 \\
&\qquad \leq \epsilon + \delta.
\end{aligned}$$

Inequality (5.2) fits well with the bounds from Section 4.

(5.3) LEMMA. *Suppose $\mathcal{F}$ and $\mathcal{G}$ are subsets of $\mathbb{R}^n$ for which*

$$D_1(\epsilon|\boldsymbol{\alpha} \odot \mathbf{F}|_1, \boldsymbol{\alpha} \odot \mathcal{F}) \leq A(1/\epsilon)^W,$$
$$D_1(\epsilon|\boldsymbol{\alpha} \odot \mathbf{G}|_1, \boldsymbol{\alpha} \odot \mathcal{G}) \leq A(1/\epsilon)^W,$$

for $0 < \epsilon \leq 1$ and every rescaling vector $\boldsymbol{\alpha}$ of nonnegative weights. Then, for every such $\boldsymbol{\alpha}$,

$$N_2(\epsilon|\boldsymbol{\alpha} \odot \mathbf{F} \odot \mathbf{G}|_2, \boldsymbol{\alpha} \odot \mathcal{F} \odot \mathcal{G}) \leq A^2(8/\epsilon^2)^{2W} \qquad \textit{for } 0 < \epsilon \leq 1. \tag{5.4}$$

A similar inequality holds for the packing numbers.

PROOF. The set $\mathcal{H} = \boldsymbol{\alpha} \odot \mathcal{F} \odot \mathcal{G}$ has envelope $\mathbf{H} = \boldsymbol{\alpha} \odot \mathbf{F} \odot \mathbf{G}$, whose ℓ_2 norm,

$$|\mathbf{H}|_2 = \left(\sum_i \alpha_i^2 F_i^2 G_i^2\right)^{1/2} = \left(|\mathbf{H} \odot \mathbf{H}|_1\right)^{1/2},$$

provides the natural scaling factor. From inequality (5.2) and Lemma 4.9, which relates ℓ_1 and ℓ_2 packing numbers, we get

$$\begin{aligned}
N_2(\epsilon|\mathbf{H}|_2, \mathcal{H}) &\leq N_2(\tfrac{1}{2}\epsilon|\mathbf{H}|_2, \boldsymbol{\alpha} \odot \mathbf{G} \odot \mathcal{F})\, N_2(\tfrac{1}{2}\epsilon|\mathbf{H}|_2, \boldsymbol{\alpha} \odot \mathbf{F} \odot \mathcal{G}) \\
&\leq D_1(\tfrac{1}{8}\epsilon^2|\mathbf{H}|_2^2, \mathbf{H} \odot \boldsymbol{\alpha} \odot \mathbf{G} \odot \mathcal{F})\, D_1(\tfrac{1}{8}\epsilon^2|\mathbf{H}|_2^2, \mathbf{H} \odot \boldsymbol{\alpha} \odot \mathbf{F} \odot \mathcal{G}).
\end{aligned}$$

The set $\mathbf{H} \odot \boldsymbol{\alpha} \odot \mathbf{G} \odot \mathcal{F}$ has envelope $\mathbf{H} \odot \mathbf{H}$, which has ℓ_1 norm $|\mathbf{H}|_2^2$, and likewise for the set $\mathbf{H} \odot \boldsymbol{\alpha} \odot \mathbf{F} \odot \mathcal{G}$. With the uniform bounds on D_1 packing numbers applied to the last two factors we end up with the asserted inequality. □

The results in this subsection are actually examples of a more general stability property involving *contraction maps.* A function $\boldsymbol{\lambda}$ from $\mathbb{R}^n$ into another Euclidean space is called an ℓ_2–contraction if it satisfies the inequality

$$|\boldsymbol{\lambda}(\mathbf{f}) - \boldsymbol{\lambda}(\mathbf{g})|_2 \leq |\mathbf{f} - \mathbf{g}|_2 \qquad \text{for all } \mathbf{f},\, \mathbf{g} \text{ in } \mathbb{R}^n.$$

For such a map $\boldsymbol{\lambda}$, it is easy to show that

$$D_2(\epsilon, \boldsymbol{\lambda}(\mathcal{F})) \le D_2(\epsilon, \mathcal{F}).$$

When applied to various cartesian products, for various maps $\boldsymbol{\lambda}$ from $\mathbb{R}^{2n}$ into $\mathbb{R}^n$, this would reproduce the bounds stated above.

General maximal inequalities. It is perhaps most natural—or at least most elegant—to start from the assumption that we are given bounds on quantities such as $\mathbb{P}_\sigma \Phi(\sup_{\mathcal{F}} |\boldsymbol{\sigma} \cdot \mathbf{f}|)$, for a convex, increasing nonnegative function Φ on $\mathbb{R}^+$. The bounds might have been derived by a chaining argument, based on inequalities for packing numbers, but we need not assume as much.

Without loss of generality we may assume sets such as $\mathcal{F}$ to be compact: by continuity, the supremum over $\mathcal{F}$ in each of the asserted inequalities will be equal to the supremum over the closure of $\mathcal{F}$; and the inequalities for unbounded $\mathcal{F}$ may be obtained as limiting cases of the inequalities for a sequence of bounded subsets of $\mathcal{F}$. Also we may assume that the zero vector belongs to $\mathcal{F}$.

The stability property for sums follows directly from the convexity of Φ:

$$(5.5)\quad \mathbb{P}_\sigma \Phi\left(\sup_{\mathcal{F},\mathcal{G}} |\boldsymbol{\sigma} \cdot (\mathbf{f} + \mathbf{g})| \right) \le \mathbb{P}_\sigma \Phi\left(\sup_{\mathcal{F}} |\boldsymbol{\sigma} \cdot \mathbf{f}| + \sup_{\mathcal{G}} |\boldsymbol{\sigma} \cdot \mathbf{g}| \right)$$
$$\le \frac{1}{2}\mathbb{P}_\sigma \Phi\left(2\sup_{\mathcal{F}} |\boldsymbol{\sigma} \cdot \mathbf{f}| \right) + \frac{1}{2}\mathbb{P}_\sigma \Phi\left(2\sup_{\mathcal{F}} |\boldsymbol{\sigma} \cdot \mathbf{g}| \right).$$

To eliminate the extra factors of 2 from the last two terms (or from similar terms later in this section) we could apply the same argument to the rescaled function $\Phi_0(x) = \Phi(x/2)$.

More subtle is the effect of applying a contraction operation to each coordinate of the vectors in $\mathcal{F}$. Suppose we have maps $\lambda_i : \mathbb{R} \to \mathbb{R}$ such that

$$(5.6)\qquad \lambda_i(0) = 0 \quad \text{and} \quad |\lambda_i(s) - \lambda_i(t)| \le |s - t| \qquad \text{for all real } s, t.$$

They define a contraction map on $\mathbb{R}^n$ pointwise, $\boldsymbol{\lambda}(\mathbf{f}) = (\lambda_1(f_1), \ldots, \lambda_n(f_n))$.

(5.7) THEOREM. *For every subset $\mathcal{F}$ of $\mathbb{R}^n$, and contraction maps λ_i,*

$$\mathbb{P}_\sigma \Phi\left(\sup_{\mathcal{F}} |\boldsymbol{\sigma} \cdot \boldsymbol{\lambda}(\mathbf{f})| \right) \le \frac{3}{2}\mathbb{P}_\sigma \Phi\left(2\sup_{\mathcal{F}} |\boldsymbol{\sigma} \cdot \mathbf{f}| \right),$$

where $\boldsymbol{\lambda}(\mathbf{f}) = (\lambda_1(f_1), \ldots, \lambda_n(f_n))$. □

Before proceeding to the proof, let us see how the theorem can be applied.

(5.8) EXAMPLE. We can build the class $\mathcal{F} \vee \mathcal{G}$ (or $\mathcal{F} \wedge \mathcal{G}$) using sums and contractions, based on the representation

$$f_i \vee g_i = (f_i - g_i)^+ + g_i.$$

Arguing as for (5.5) we get a bound for the set of all differences $\mathbf{f} - \mathbf{g}$. With the contraction maps $\lambda_i(s) = s^+$ we get a bound for the set of vectors with components $(f_i - g_i)^+$, which we combine with the bound for $\mathcal{G}$ using (5.5). □

(5.9) EXAMPLE. If we impose the condition that $|f_i| \leq 1$ and $|g_i| \leq 1$ for all components of all vectors in $\mathcal{F}$ and $\mathcal{G}$, then we can build $\mathcal{F} \odot \mathcal{G}$ using sums and contractions, based on the representation

$$f_i g_i = \tfrac{1}{4}(f_i + g_i)^2 - \tfrac{1}{4}(f_i - g_i)^2.$$

Stability for sums (and differences) gives bounds for the sets of vectors with components $1/2(f_i \pm g_i)$. With the contraction map $\lambda_i(s) = 1/2 \min(1, s^2)$ we get a suitable bound for both the squared terms, which we again combine by means of inequality (5.5). □

As the first step towards the proof of Theorem 5.7 we must establish a stronger result for a special case, using only elementary properties of Φ.

(5.10) LEMMA. *If $\mathcal{F}$ lies within the positive orthant of $\mathbb{R}^n$,*

$$\mathbb{P}_\sigma \Phi\left(\sup_{\mathcal{F}} |\boldsymbol{\sigma} \cdot \boldsymbol{\lambda}(\mathbf{f})|\right) \leq \mathbb{P}_\sigma \Phi\left(\sup_{\mathcal{F}} |\boldsymbol{\sigma} \cdot \mathbf{f}|\right)$$

for contraction maps λ_i, as in (5.6).

PROOF. It would suffice to consider the effect of the contractions one coordinate at a time. We would first show that

$$\mathbb{P}_\sigma \Phi\left(\sup_{\mathcal{F}} \Big|\sum_{i<n} \sigma_i f_i + \sigma_n \lambda_n(f_n)\Big|\right) \leq \mathbb{P}_\sigma \Phi\left(\sup_{\mathcal{F}} |\boldsymbol{\sigma} \cdot \mathbf{f}|\right).$$

Then we could argue similarly for the $(n-1)^{st}$ coordinate—replacing the right-hand side by the quantity now on the left-hand side, and replacing f_{n-1} on the left-hand side by $\lambda_{n-1}(f_{n-1})$—and so on.

Let us establish only the inequality for the n^{th} coordinate. Argue conditionally on $\sigma_1, \ldots, \sigma_{n-1}$. To simplify the notation, write λ instead of λ_n, write $x(\mathbf{f})$ for the contribution from the first $n-1$ coordinates, and write $y(\mathbf{f})$ for f_n. Then we need to show that

$$\begin{aligned}(5.11)\qquad \Phi\left(\sup_{\mathcal{F}} |x(\mathbf{f}) + \lambda(y(\mathbf{f}))|\right) &+ \Phi\left(\sup_{\mathcal{F}} |x(\mathbf{f}) - \lambda(y(\mathbf{f}))|\right) \\ &\leq \Phi\left(\sup_{\mathcal{F}} |x(\mathbf{f}) + y(\mathbf{f})|\right) + \Phi\left(\sup_{\mathcal{F}} |x(\mathbf{f}) - y(\mathbf{f})|\right).\end{aligned}$$

The argument will be broken into four cases. Suppose the supremum in the first term on the left-hand side is achieved at $\mathbf{f}_0$ and for the second term at $\mathbf{f}_1$. That is, if $x_0 = x(\mathbf{f}_0)$ and so on,

$$\begin{aligned}(5.12)\qquad |x_0 + \lambda(y_0)| &\geq |x(\mathbf{f}) + \lambda(y(\mathbf{f}))| \\ |x_1 - \lambda(y_1)| &\geq |x(\mathbf{f}) - \lambda(y(\mathbf{f}))|\end{aligned}$$

for all $\mathbf{f}$ in $\mathcal{F}$. For the first two cases we will need only to appeal to the facts: $\Phi(t)$ is an increasing function of t on $\mathbb{R}^+$; both y_0 and y_1 are nonnegative; and

$$(5.13)\qquad |\lambda(y_i)| = |\lambda(y_i) - \lambda(0)| \leq |y_i| = y_i \qquad \text{for } i = 0, 1,$$

as a consequence of the contraction property for λ.

For notational convenience, extend the function Φ by symmetry to the whole real line: $\Phi(-t) = \Phi(t)$. Then it will be enough to show that in each case at least one of the following inequalities holds:

$$\Phi(x_0 + \lambda(y_0)) + \Phi(x_1 - \lambda(y_1)) \le \begin{cases} \Phi(x_0 + y_0) + \Phi(x_1 - y_1) \\ \Phi(x_1 + y_1) + \Phi(x_0 - y_0) \end{cases} \tag{5.14}$$

First case: if $x_0 + \lambda(y_0) \ge 0 \ge x_1 - \lambda(y_1)$, then

$$\Phi(x_0 + \lambda(y_0)) \le \Phi(x_0 + y_0),$$
$$\Phi(x_1 + \lambda(y_1)) \le \Phi(x_1 - y_1).$$

Second case: if $x_0 + \lambda(y_0) \le 0 \le x_1 - \lambda(y_1)$, then

$$\Phi(x_0 + \lambda(y_0)) \le \Phi(x_0 - y_0),$$
$$\Phi(x_1 - \lambda(y_1)) \le \Phi(x_1 + y_1).$$

At least one of the inequalities in (5.14) is clearly satisfied in both these cases.

For the other two cases, where $x_0 + \lambda(y_0)$ and $x_1 - \lambda(y_1)$ have the same sign, we need the following consequence of the convexity of Φ: *if $\alpha \le \beta$ and $\beta \ge 0$ and $0 \le s \le t$, then*

$$\Phi(\beta + t) - \Phi(\beta) - \Phi(\alpha + s) + \Phi(\alpha) \ge 0. \tag{5.15}$$

If $s = 0$ this inequality reasserts that Φ is an increasing function on $\mathbb{R}^+$. If $s > 0$ it follows from the convexity inequality

$$\frac{\Phi(\alpha + s) - \Phi(\alpha)}{s} \le \frac{\Phi(\beta + t) - \Phi(\beta)}{t}$$

and the nonnegativity of the ratio on the right-hand side.

Third case: if $x_0 + \lambda(y_0) \ge 0$ and $x_1 - \lambda(y_1) \ge 0$, then invoke inequality (5.15) with

$$\alpha = x_1 - y_1, \quad \beta = x_0 + \lambda(y_0), \quad s = y_1 - \lambda(y_1), \quad t = y_0 - \lambda(y_0) \quad \text{if } y_0 \ge y_1,$$
$$\alpha = x_0 - y_0, \quad \beta = x_1 - \lambda(y_1), \quad s = y_0 + \lambda(y_0), \quad t = y_1 + \lambda(y_1) \quad \text{if } y_0 < y_1.$$

The inequalities (5.12) and (5.13) give $\alpha \le \beta$ in each case, and the inequality $s \le t$ follows from the contraction property

$$|\lambda(y_1) - \lambda(y_0)| \le \begin{cases} y_0 - y_1 & \text{if } y_0 \ge y_1, \\ y_1 - y_0 & \text{if } y_0 < y_1. \end{cases}$$

Fourth case: if $x_0 + \lambda(y_0) \le 0$ and $x_1 - \lambda(y_1) \le 0$, then invoke (5.15) with

$$\alpha = -x_1 - y_1, \quad \beta = -x_0 - \lambda(y_0), \quad s = y_1 + \lambda(y_1), \quad t = y_0 + \lambda(y_0) \quad \text{if } y_0 \ge y_1,$$
$$\alpha = -x_0 - y_0, \quad \beta = -x_1 + \lambda(y_1), \quad s = y_0 - \lambda(y_0), \quad t = y_1 - \lambda(y_1) \quad \text{if } y_0 < y_1.$$

The required inequalities $\alpha \le \beta$ and $s \le t$ are established as in the third case. $\square$

PROOF OF THEOREM 5.7. Notice that $\lambda_i(f_i) = \lambda_i(f_i^+) + \lambda_i(-f_i^-)$, because either

$$f_i \geq 0 \quad \text{and} \quad \lambda_i(f_i) = \lambda_i(f_i^+) \quad \text{and} \quad \lambda_i(-f_i^-) = \lambda_i(0) = 0,$$

or

$$f_i \leq 0 \quad \text{and} \quad \lambda_i(f_i) = \lambda_i(-f_i^-) \quad \text{and} \quad \lambda_i(f_i^+) = \lambda_i(0) = 0.$$

Convexity of Φ gives the inequality

$$\begin{aligned}\mathbb{P}_\sigma\Phi\left(\sup_{\mathcal{F}}\left|\sum_{i\leq n}\sigma_i[\lambda_i(f_i^+) + \lambda_i(-f_i^-)]\right|\right)\\ \leq \frac{1}{2}\mathbb{P}_\sigma\Phi\left(2\sup_{\mathcal{F}}\left|\sum_{i\leq n}\sigma_i\lambda_i(f_i^+)\right|\right) + \frac{1}{2}\mathbb{P}_\sigma\Phi\left(2\sup_{\mathcal{F}}\left|\sum_{i\leq n}\sigma_i\lambda_i(-f_i^-)\right|\right).\end{aligned}$$

Lemma 5.10 shows that the right-hand side increases if $\lambda_i(f_i^+)$ is replaced by f_i^+ and $\lambda_i(-f_i^-)$ is replaced by $-f_i^-$. (For $-f_i^-$, note that $\lambda(-t)$ is also a contraction mapping.) Argue from convexity of Φ and the inequality $f_i^+ = 1/2(f_i + |f_i|)$ that

$$\mathbb{P}_\sigma\Phi\left(2\sup_{\mathcal{F}}\left|\sum_{i\leq n}\sigma_i f_i^+\right|\right) \leq \frac{1}{2}\mathbb{P}_\sigma\Phi\left(2\sup_{\mathcal{F}}\left|\sum_{i\leq n}\sigma_i f_i\right|\right) + \frac{1}{2}\mathbb{P}_\sigma\Phi\left(2\sup_{\mathcal{F}}\left|\sum_{i\leq n}\sigma_i |f_i|\right|\right),$$

with a similar inequality for the contribution from the $-f_i^-$ term. The proof will be completed by an application of the Basic Combinatorial Lemma from Section 1 to show that

$$\mathbb{P}_\sigma\Phi\left(2\sup_{\mathcal{F}}\left|\sum_{i\leq n}\sigma_i |f_i|\right|\right) \leq 2\,\mathbb{P}_\sigma\Phi\left(2\sup_{\mathcal{F}}\left|\sum_{i\leq n}\sigma_i f_i\right|\right).$$

Because Φ is increasing and nonnegative, and $\mathcal{F}$ contains the zero vector,

$$\mathbb{P}_\sigma\Phi\left(2\sup_{\mathcal{F}}\left|\sum_{i\leq n}\sigma_i |f_i|\right|\right) \leq \mathbb{P}_\sigma\Phi\left(2\sup_{\mathcal{F}}\sum_{i\leq n}\sigma_i |f_i|\right) + \mathbb{P}_\sigma\Phi\left(2\sup_{\mathcal{F}}\sum_{i\leq n}(-\sigma_i)|f_i|\right).$$

The two expectations on the right-hand side are equal; it will suffice if we bound the first of them by the corresponding quantity with $|f_i|$ replaced by f_i.

To do this, let us construct, by means of the Basic Combinatorial Lemma, a one-to-one map θ from $\mathcal{S}$ onto itself such that

$$\sup_{\mathcal{F}}\sum_{i\leq n}\sigma_i|f_i| \leq \sup_{\mathcal{F}}\sum_{i\leq n}\theta(\boldsymbol{\sigma})_i f_i. \tag{5.16}$$

For each $\boldsymbol{\sigma}$ in $\mathcal{S}$, the compactness of $\mathcal{F}$ ensures existence of a vector $\mathbf{f}^\sigma$ for which the left-hand side of (5.16) equals

$$\sum_{i\leq n}\sigma_i|f_i^\sigma|.$$

Define the map η from $\mathcal{S}$ into itself by

$$\eta(\boldsymbol{\sigma})_i = \begin{cases} +1 & \text{if } \sigma_i = +1 \text{ and } f_i^\sigma \geq 0, \\ -1 & \text{otherwise.}\end{cases}$$

For every $\boldsymbol{\sigma}$ we have $\eta(\boldsymbol{\sigma}) \leq \boldsymbol{\sigma}$. The Basic Combinatorial Lemma gives a one-to-one map θ that has $\theta(\boldsymbol{\sigma}) \wedge \boldsymbol{\sigma} = \eta(\boldsymbol{\sigma})$. In particular, $\theta(\boldsymbol{\sigma})_i$ is equal to $+1$ if both $\sigma_i = +1$ and $f_i^\sigma \geq 0$, and equal to -1 if $\sigma_i = +1$ and $f_i^\sigma < 0$. Thus

$$\begin{aligned}
\sum_{i\leq n} \sigma_i |f_i^\sigma| &= \sum_{\sigma_i=+1} \theta(\boldsymbol{\sigma})_i f_i^\sigma - \sum_{\sigma_i=-1} |f_i^\sigma| \\
&\leq \sum_{i\leq n} \theta(\boldsymbol{\sigma})_i f_i^\sigma \\
&\leq \sup_{\mathcal{F}} \sum_{i\leq n} \theta(\boldsymbol{\sigma})_i f_i,
\end{aligned}$$

as asserted by (5.16). Because θ is one-to-one, the random vector $\theta(\boldsymbol{\sigma})$ has a uniform distribution under $\mathbb{P}_\sigma$, and

$$\mathbb{P}_\sigma \Phi\left(2 \sup_{\mathcal{F}} \sum_{i\leq n} \sigma_i |f_i|\right) \leq \mathbb{P}_\sigma \Phi\left(2 \sup_{\mathcal{F}} \theta(\boldsymbol{\sigma}) \cdot \mathbf{f}\right) = \mathbb{P}_\sigma \Phi\left(2 \sup_{\mathcal{F}} \boldsymbol{\sigma} \cdot \mathbf{f}\right),$$

as required. □

REMARKS. The last subsection corresponds to a small fraction of the Ledoux and Talagrand (1989) paper. Ledoux and Talagrand (1990, Chapter 4) have further refined the method of proof. Except perhaps for the stability result for covering numbers of products, the rest of the section merely collects together small results that have been derived many times in the literature.