Invent. math. 102, 179–208 (1990)

Inventiones mathematicae

Gaussian kernels have only Gaussian maximizers*

Elliott H. Lieb
Department of Mathematics, Princeton University, Princeton, NJ 08544, USA

Oblatum 19-XII-1989

Abstract. A Gaussian integral kernel $G(x, y)$ on $\mathbf{R}^n \times \mathbf{R}^n$ is the exponential of a quadratic form in x and y; the Fourier transform kernel is an example. The problem addressed here is to find the sharp bound of G as an operator from $L^p(\mathbf{R}^n)$ to $L^q(\mathbf{R}^n)$ and to prove that the $L^p(\mathbf{R}^n)$ functions that saturate the bound are necessarily Gaussians. This is accomplished generally for $1 < p \leqq q < \infty$ and also for $p > q$ in some special cases. Besides greatly extending previous results in this area, the proof technique is also essentially different from earlier ones. A corollary of these results is a fully multidimensional, multilinear generalization of Young's inequality.

I. Introduction

The classic Hausdorff-Young-Titchmarsh [T] inequality for Fourier integrals states that for $1 \leqq p \leqq 2$ the Fourier transform on $L^p(\mathbf{R}^n)$ is a bounded map into $L^{p'}(\mathbf{R}^n)$ with a bound that is at most 1; here $1/p' + 1/p = 1$. In 1961 Babenko [BA] showed that when p' is an even integer greater than 2 and $n = 1$ the bound is in fact less than 1, and he determined its value. This bound is achieved for Gaussian functions and Babenko states, but does not demonstrate explicitly, that Gaussians are the only functions with this property. Babenko's method was to apply analytic function theory to the Euler-Lagrange equation associated with the maximization problem.

The Fourier integral is but one example of a transform given by a Gaussian integral kernel $G(x, y)$, i.e., the exponential of a quadratic plus linear form in x and y. In the Fourier transform case in $\mathbf{R}^n$ the kernel is $G(x, y) = \exp\{-2i(x, y)\}$. Another well known example in $\mathbf{R}^n$ is the purely real operator $\tilde{\mathscr{G}} = \exp\{t\Delta + 2tx \cdot \nabla\}$ on Gauss space (with measure $d\mu = \exp\{-|x|^2\}\,dx$) investigated by Nelson [N1; N2] as an operator from $L^p(\mathbf{R}^n, d\mu)$ to $L^q(\mathbf{R}^n, d\mu)$. In

* Work partially supported by U.S. National Science Foundation grant PHY-85-15288-A03

terms of Lebesgue measure, this amounts to considering the kernel

$$G(x, y) = \exp\left\{ -\frac{1}{q}|x|^2 + \frac{1}{p}|y|^2 - \frac{|y - cx|^2}{(1 - c^2)} \right\}$$

from $L^p(\mathbf{R}^n)$ to $L^q(\mathbf{R}^n)$ for $0 \leqq c = e^{-t} < 1$. Nelson defined the operator $\mathscr{G}$ by $(\mathscr{G}f)(x) = \int G(x, y) f(y) dy$ and showed that $\mathscr{G}$ is bounded from $L^p(\mathbf{R}^n)$ to $L^q(\mathbf{R}^n)$ when $p \leqq q$ if and only if $(q - 1)c^2 \leqq p - 1$; he also derived the explicit value of the bound—which again is achieved when f is a Gaussian. This is the famous hypercontractivity theorem. [In [N1] Nelson showed that $\mathscr{G}$ is bounded if c is small enough; Glimm [GL] used this fact plus the spectral gap in the generator to show that $\tilde{\mathscr{G}}$ is a contraction on Gauss space for some still smaller c. Finally Nelson [N2] proved the sharp bound as stated above. In 1976 Neveu [NE] and Brascamp and Lieb [BL] found other proofs, and Simon [SI] found a proof for $p = 2$ and $q = 2, 4, 6, 8 \ldots$. Recently, Carlen and Loss [CL] have used their method of competing symmetries to construct another proof of the hypercontractivity theorem.] However, Nelson's method seems incapable of showing that Gaussians are the *only* maximizers; the proof of this fact, as well as a completely different proof (using rearrangement inequalities) of the hypercontractivity theorem was given by Brascamp and Lieb [BL]. The method in [CL] also yields uniqueness. Nelson's original proof used stochastic integrals and Gaussian processes in $\mathbf{R}^n$ (in fact it even extends to infinite dimensions). Segal [S] showed how to use Minkowski's inequality [HLP] to reduce the $\mathbf{R}^n$ case of Nelson's kernel to the $\mathbf{R}^1$ case; he also showed that $\tilde{\mathscr{G}}$ is a contraction on Gauss space for small c. The $\mathbf{R}^1$ case was simplified by Gross [G] who showed the equivalence of hypercontractivity with logarithmic Sobolev inequalities and built up one-dimensional Gauss measure from two-point measures via the central limit theorem. See the survey by Davies et al. [DGS].

In his important 1975 paper, Beckner [B1; B2] used the Nelson-Gross machinery and the Hermite semigroup to settle the question raised by Babenko. By using the tensor product structure of Fourier transforms and an application of Minkowski's inequality related to, but distinct from, Segal's [S], he reduced the $\mathbf{R}^n$ case to the $\mathbf{R}^1$ case. He also showed that for all $1 \leqq p \leqq 2$ the sharp constant in the Hausdorff-Young-Titchmarsh inequality is given by Gaussian functions—as found by Babenko. However, this method also leaves open the question of whether Gaussian functions are the *only* maximizers.

Since then the Nelson-Gross-Beckner method has been extended to other complex (as distinct from purely real or purely imaginary) Gaussian kernels in $\mathbf{R}^n$ (i.e., the complex Mehler kernel) [C; E; J; W]. In this paper the *general* problem in $\mathbf{R}^n$ in the $p \leqq q$ case will be settled by a *completely different method* and, moreover, *the maximizers will be shown to be Gaussian functions*. Some of the $p > q$ cases will be settled as well. Before discussing the earlier results in detail it is necessary to define the problem more completely.

The most general **Gaussian kernel** on $\mathbf{R}^n \times \mathbf{R}^n$ is

$$G(x, y) = \exp\left\{ -(x, Ax) - (y, By) - 2(x, Dy) + 2\left(L, \begin{pmatrix} x \\ y \end{pmatrix}\right)\right\} \tag{1.1}$$

and its action on complex valued, measurable functions $f: \mathbf{R}^n \to \mathbf{C}$, is formally given by

$$(\mathscr{G}f)(x) = \int G(x, y) f(y) \mathrm{d}y \,. \tag{1.2}$$

In (1.1) A, B and D are (complex) $n \times n$ matrices with A and B being symmetric while L is a vector in $\mathbf{C}^{2n}$. The Fourier transform corresponds to $A = B = 0$, $L = 0$ and $D = iI$, with I denoting the identity.

Notation. If α and β are vectors in $\mathbf{C}^n$ then $(\alpha, \beta) \equiv \sum_{i=1}^n \alpha_i\beta_i$ and *not* $\sum_{i=1}^n \bar{\alpha}_i\beta_i$. Lebesgue integration over $\mathbf{R}^n$ is denoted simply by $\int dx$ whenever the n in question is clear from the context. The $L^p(\mathbf{R}^n)$ norm of a measurable function f will be denoted by $\|f\|_p$, i.e., $\{\int |f(x)|^p dx\}^{1/p}$. The notation

$$\begin{pmatrix} A & D \\ D^T & B \end{pmatrix} = M + iN \tag{1.3}$$

will also be used, where M and N are real, symmetric $2n \times 2n$ matrices. The sole condition imposed on G is that M is positive semidefinite. G is said to be **nondegenerate** if M is positive definite, while G is said to be **degenerate** if M has a zero eigenvalue. The Fourier transform kernel and Nelson's kernel with $(q-1)c^2 = (p-1)$ are examples of degenerate kernels. The operator $\mathcal{G}$ should perhaps be written $\mathcal{G}_G$, but this will not be done since the pairing of $\mathcal{G}$ and G will always be clear from the context.

The linear operator $\mathcal{G}$ associated to G will be studied as an operator from $L^p(\mathbf{R}^n)$ to $L^q(\mathbf{R}^n)$ for $1 < p < \infty$ and $1 < q < \infty$. (The cases p or $q = 1$ or ∞ can also be analyzed by the methods of this paper but they will be omitted since these cases involve extra technical considerations.) When G is nondegenerate the definition of $\mathcal{G}$ in (1.2) makes sense (by Hölder's inequality) but if G is degenerate then (1.2) is meaningless unless f is also in $L^p(\mathbf{R}^n) \cap L^1(R^n)$. Assuming that $\mathcal{G}$, when restricted to $L^p(\mathbf{R}^n) \cap L^1(\mathbf{R}^n)$, is bounded from $L^p(\mathbf{R}^n)$ to $L^q(\mathbf{R}^n)$ then, for any $f \in L^p(\mathbf{R}^n)$, $\mathcal{G}f \in L^q(\mathbf{R}^n)$ is uniquely defined by taking *any* sequence $f_j \in L^p(\mathbf{R}^n) \cap L^1(\mathbf{R}^n)$ that converges to f in $L^p(\mathbf{R}^n)$ and then noting that $\mathcal{G}f = \lim_{j\to\infty} \mathcal{G}f_j$ is well defined since $\mathcal{G}f_j$ is a Cauchy sequence in $L^q(\mathbf{R}^n)$. This definition is well known and is, in fact, the way that the Fourier transform is defined when $1 < p \leqq 2$.

Associated to G and the numbers p and q with $1 \leqq p \leqq \infty$ and $1 \leqq q \leqq \infty$ is the ratio

$$\mathcal{R}_{p\to q}(f) = \frac{\|\mathcal{G}f\|_q}{\|f\|_p} \tag{1.4}$$

for $f \in L^p(\mathbf{R}^n)$, $f \neq 0$ and, in case G is degenerate, $f \in L^1(\mathbf{R}^n)$ as well. The norm of $\mathcal{G}$ from $L^p(\mathbf{R}^n)$ to $L^q(\mathbf{R}^n)$ is defined to be

$$C_{p\to q} = \sup_f \mathcal{R}_{p\to q}(f) \tag{1.5}$$

in which the supremum is over the class of f's just stated. In case $0 \neq f \in L^p(\mathbf{R}^n)$ and $C_{p\to q} < \infty$ and

$$\|\mathcal{G}f\|_q = C_{p\to q}\|f\|_p$$

(using the above definition of $\mathcal{G}f$ as a limit when G is degenerate) then f is said to be a **maximizer** for $\mathcal{G}$ (or for G). If there is any ambiguity about the G under discussion (e.g., in Theorem 3.3) the notation $\mathcal{R}_{p\to q}(G, f)$ and $C_{p\to q}(G)$ will be used.

Functions from $\mathbf{R}^n$ to $\mathbf{C}$ of the form

$$g(x) = \mu \exp\{-(x, Jx) + (l, x)\} \tag{1.6}$$

with $0 \neq \mu \in \mathbf{C}$, $l \in \mathbf{C}^n$ and J a symmetric $n \times n$ matrix with $\mathrm{Re}(J)$ positive *definite* will be called **Gaussian functions.** In case $L = 0$ in (1.1) or $l = 0$ in (1.6) then G (resp. g) will be called a **centered Gaussian kernel** (resp. **function**). If A, B, D and L in (1.1) are real then G is said to be a **real Gaussian kernel.** Likewise, if J and l (but not necessarily μ) in (1.6) are real then g is said to be a **real Gaussian function.**

A preliminary simplification of G can be made. Without loss of generality it can be assumed that A and B are real matrices because the imaginary part of B can be absorbed into f in (1.4) without changing $\|f\|_p$. The imaginary part of A can be omitted without changing $\|\mathscr{G}f\|_q$. For the same reason the vector L can be assumed to be real. Furthermore, when G is nondegenerate then we can also set L (which is now real) equal to zero. The reason is simply that the affine change of variables $\binom{x}{y} \mapsto \binom{x}{y} - V$, with V being the unique solution of the equation $MV = L$ in $\mathbf{R}^{2n}$, eliminates the real linear term from (1.1) and merely changes $C_{p\to q}$ into $C_{p\to q}\exp\{(L, V)\}$. When G is degenerate, L can also be eliminated in the same way provided $MV = L$ has a solution. Because $\mathrm{Rank}(M) < 2n$ in the degenerate case, such a solution conceivably might not exist, but it turns out that a solution does indeed exist whenever $\mathscr{G}$ is bounded. This is the content of Lemma 2.2 below. Therefore, *without loss of generality, the only* G's *that need to be studied are those for which*

(*i*) A and B are real, symmetric $n \times n$ matrices,
(*ii*) $L = 0$, i.e., G is centered.

These assumptions will be made in the theorems in this paper.

On the other hand, suppose that the supremum of $\mathscr{R}_{p\to q}(f)$ in (1.5) is taken over Gaussian functions *only* (which are automatically in $L^p(\mathbf{R}^n)$ for every p). Then, according to Lemma 2.3 below, only centered Gaussian functions need be considered in (1.5). This is a considerable simplification that is not altogether obvious and it is important in the application of Theorem 4.1 which states that this restricted supremum is all that need be considered.

The results of this paper can be summarized as follows. Three cases are treated. With the assumptions (i) and (ii) above,

(*A*) D is real and $1 < p < \infty$ and $1 < q < \infty$
(*B*) D is imaginary and either $1 < p \leqq 2$ and $1 < q < \infty$ or else $1 < p < \infty$, and $2 \leqq q < \infty$.
(*C*) D is complex and $1 < p \leqq q < \infty$.

If G is *nondegenerate* then $\mathscr{R}_{p\to q}$ has exactly one maximizer and it is a centered Gaussian function. These are Theorems 3.2, 3.3 and 3.4.

If G is *degenerate* then in all cases

$$C_{p\to q} = \sup_g \mathscr{R}_{p\to q}(g)\,, \tag{1.7}$$

where the supremum is over centered Gaussian functions. This is Theorem 4.1. Furthermore, if the supremum in (1.7) is achieved for some Gaussian function then, when $p < q$, *every maximizer is a Gaussian function*—as Theorems 4.5 shows. Theorem 4.3 gives a sufficient condition for the achievement of the maximum in (1.7) in the degenerate case; in Case (A) it is necessary as well. Thus, Case (A) is settled completely: in the degenerate case with $p \geqq q$ there is no maximizer of any kind, while if $p < q$ all maximizers are Gaussian functions.

In general, the question of the existence of a maximizer in the degenerate case is a subtle one. For the Fourier transform (which is both Case (B) and (C)), *every* function in $L^2(\mathbf{R}^n)$ is a maximizer when $p = q = 2$; on the other hand the seemingly harmless modification of the Fourier transform in 4.2(5) below is bounded but has *no maximizer of any kind* when $q = p' \geqq 2$. When $q = p' > 2$ the Fourier transform on $\mathbf{R}^1$ has a three real parameter family of maximizers, $f(y) = \exp\{-Jy^2 + ly\}$ with $J > 0$ and $l \in \mathbf{C}$. When $p < q$ the convolution kernel $G(x, y) = \exp\{-(x-y)^2\}$ on $\mathbf{R}^1$ has a one real parameter family of maximizers, $f(y) = \exp\{-Jy^2 + ly\}$ with $l \in \mathbf{R}$ and $J = \frac{p'}{q'} - 1$; when $p = q$, G is bounded but there is no maximizer (see 4.2 below). There does not seem to be any simple rule. In simple cases (which include all the standard ones in $\mathbf{R}^n$ and all the cases in $\mathbf{R}^1$) the existence of a Gaussian maximizer in (1.7) can be decided by computation. Otherwise, (1.7) reduces to a complicated algebraic problem and precise conditions are not given here. Moreover it is not even proved that the absence of a Gaussian maximizer in (1.7) precludes the existence of a non-Gaussian maximizer—although a conjecture to this effect is made in 4.4.

All these results extend to Gaussian kernels on $\mathbf{R}^m \times \mathbf{R}^n$, in which A is $m \times m$, B is $n \times n$, D is $m \times n$ and $L \in \mathbf{C}^{m+n}$. The proof is given in Sect. V. This generalization, while it is an easy one, does occur in applications, e.g., the entropy bound for coherent states in [L1].

Multilinear Gaussian forms are discussed in Sect. VI and it is proved there that the methods and results of Sects. II–V carry through for real forms. As an application of the real multilinear result in Sect. 6.1, the *fully multidimensional Young inequality* for K functions (which was left unresolved in [BL], p. 162) is proved in 6.2. The method of proof is, of course, quite different from that in [BL]; there, rearrangement inequalities were used and they were not flexible enough to encompass the fully multidimensional case.

The relationship of the results of this paper to earlier results on Gaussian kernels (beyond [BA; N1; N2; B1; B2]) can be summarized as follows. In 1976 Brascamp and Lieb [BL] found the norm for Case (A) in $\mathbf{R}^n$ (Theorem 7) and proved that Gaussian functions are the unique maximizer in $\mathbf{R}^1$ in the degenerate case (Theorem 13); this latter proof easily extends to $\mathbf{R}^n$ and to the nondegenerate case. In fact, by a simple change of variables (see the proof of Theorem 4.3 below) the $\mathbf{R}^n$ Case (A) reduces to a simple tensor product of $\mathbf{R}^1$ kernels. In 1979 Coifman et al. [C] used Beckner's result and an interpolation technique to deduce the norm for the complex Mehler kernel in $\mathbf{R}^1$ for $q = p' \geqq 2$ (which is in Case (C)). In the same year Weissler [W] extended Nelson's and Beckner's results to the complex Mehler kernel in $\mathbf{R}^1$ with the exception of $2 < p \leqq q < 3$ and $\frac{3}{2} < p \leqq q < 2$. In 1988, Epperson [E] found the norm for the following nondegenerate cases in $\mathbf{R}^1$: Case (C), Case (B), the case $p \geqq 2 \geqq q$. He also found the norm for certain $\mathbf{R}^1$ cases $q < p < 2$ and $2 < q < p$ with sufficiently nondegenerate kernels (Theorem 2.10), and for the $\mathbf{R}^1$ degenerate Case (C) if $A > 0$ and $B > 0$ (corresponding to Theorem 4.3 here).

The only complex cases in $\mathbf{R}^n$ that were known prior to Epperson's work were the simple tensor products of $\mathbf{R}^1$ kernels; these could be analyzed for $p \leqq q$ via Minkowski's inequality, as shown by Beckner [B1; B2]. Epperson was able to handle the nondegenerate Case (C) for which there is an $n \times n$ complex symmetric matrix W with $\|W\| \leqq 1$ such that $A = W(I - W^2)^{-1}W - \frac{1}{q}I$, $B = (I - W^2)^{-1} - \frac{1}{p'}I$ and $D = W(I - W^2)^{-1}$. Here, I is the identity matrix.

It will be seen from the above summary that all the previous cases, except for Epperson's $\mathbf{R}^1$ cases of $p \geqq 2 \geqq q$ and the special $q < p < 2$ and $2 < q < p$ cases, are covered in the cases (A), (B) and (C) treated in this paper. Moreover cases (A), (B) and (C) are resolved here in full $\mathbf{R}^n$ generality (i.e., not only for simple n-fold tensor products of $\mathbf{R}^1$ kernels). The main methodological point of this paper, however, is that all the previous results, except for [BL] and [BA], ultimately rely on the Nelson-Gross machinery which, while it is natural in its original context of quantum field theory and Gauss measures, is conceptually complicated in the context of general Gaussian kernels with Lebesgue measure. The two settings (Gauss measure and Lebesgue measure) for Gaussian kernels are mathematically equivalent, however, and the choice is a matter of taste. Lebesgue measure is used in this paper because it is felt that it is more natural to retain translation invariance (e.g., in the Fourier transform). Prior to Epperson's work all results in the field, except for [BL] and [BA] came from translating Gauss measure bounds for products of complex $\mathbf{R}^1$ Mehler kernels into $\mathbf{R}^n$ results via Beckner's Minkowski lemma. The proofs here use only Minkowski's inequality and simple facts about analytic functions (which appear to be unrelated to Babenko's use of analyticity—the Euler-Lagrange equation is not used).

Basically there is one idea that runs through Theorems 3.1, 3.3 and 4.5, although the technicalities are different in each. The main idea is to study $\mathscr{G} \otimes \mathscr{G}$ from $L^p(\mathbf{R}^{2n})$ to $L^q(\mathbf{R}^{2n})$ and use Minkowski's inequality. By considering the $\mathscr{G} \otimes \mathscr{G}$ maximizer $F(y_1, y_2) = f\left(\frac{y_1 + y_2}{\sqrt{2}}\right) f\left(\frac{y_1 - y_2}{\sqrt{2}}\right)$, where f is a maximizer for $\mathscr{G}$, it is possible to conclude that f must be a Gaussian. It will be noted that some of the proofs are long, and so it may appear at first that their structure is not really very simple. To a large extent the length is due to the fact that proving uniqueness raises technical considerations that would be absent if only inequalities are proved, e.g., it is not sufficient here to prove the inequalities for a dense set of smooth functions.

Apart from the extension to $\mathbf{R}^n$ (which is handled here in a natural way) the main new theorem in this paper is that a maximizer *must be* a Gaussian, and it is unique in the nondegenerate case. In the degenerate case $C_{p \to q}(G)$ is determined by examining only Gaussian functions and, if a Gaussian maximizer exists, every maximizer is a Gaussian. This is Theorem 4.5 and it can be useful as in [L1] and [L2]. Except for the real case [BL], it was previously known that Gaussian functions were *among* the maximizers. The one exception to this rule was pointed out by Beckner (private communication) for the Fourier transform from $L^p(\mathbf{R}^n)$ to $L^{p'}(\mathbf{R}^n)$ with the restriction $p' \geqq 4$. His proof that a maximizer must be a Gaussian function in this case uses a result in [BL]; the proof is

$$\|f\|_p^2 \geqq \mu_r \|f * f\|_r \geqq \mu_r (C_r^B)^n \|\widehat{f * f}\|_{r'} = \mu_r (C_r^B)^n \|(\hat{f})^2\|_{r'} = \mu (C_r^B)^n \|\hat{f}\|_{p'}^2$$

with $r' = p'/2 \geqq 2$, with $(C_r^B)^n$ being the sharp Beckner (or Babenko) constant for the Fourier transform (denoted by $\widehat{\ }$), and with μ_r being the sharp constant in Young's convolution inequality which was derived simultaneously in [B1, B2] and in [BL]. A Gaussian function $f(y) = \exp\{-Jy^2 + ly\}$ with $J > 0$ and $l \in \mathbf{C}$ gives equality above. However, [BL] (Theorem 13) proved that such functions are the *only* ones that give equality in Young's inequality.

It is a pleasure to acknowledge my debt to Eric Carlen. He helped to stimulate my interest in this problem and to understand the literature in the field. He also critically examined the work as it took shape. Thanks are also due to the Institute for Advanced Study for its hospitality during part of this work, and to Michael Loss for valuable discussions.

II. Some basic properties of Gaussians

2.1. Lemma (nondegenerate Gaussian kernels are compact and have maximizers). *Let G be a centered, nondegenerate Gaussian kernel in $\mathbf{R}^n \times \mathbf{R}^n$ as in* (1.1) *with M in* (1.3) *positive definite and $L = 0$. Let $1 < p < \infty$ and $1 < q < \infty$. Then $\mathscr{G}$ in* (1.2) *is a compact operator from $L^p(\mathbf{R}^n)$ to $L^q(\mathbf{R}^n)$ and there is at least one maximizer $f \in L^p(\mathbf{R}^n)$(i.e., $\mathscr{R}_{p\to q}(f) = C_{p\to q}$).*

Every such maximizer $f\colon \mathbf{R}^n \to \mathbf{C}$, has the following three properties, in which α and β are positive constants that depend on G, p and q but not on f.

(*a*) *There is an entire analytic function of order at most 2, $m\colon \mathbf{C}^n \to \mathbf{C}$, such that $f(x) = |m(x)|^{p'} m(x)^{-1}$ for $x \in \mathbf{R}^n$. Here $1/p + 1/p' = 1$. Moreover, for $z \in \mathbf{C}^n$,*

$$|m(z)| \leqq \alpha \| f \|_p^{p-1} \exp\{\beta |z|^2\} .$$

(*b*) *The function $|f|^{2(p-1)}$ from $\mathbf{R}^n$ to $\mathbf{R}$ has an extension to an entire analytic function from $\mathbf{C}^n$ to $\mathbf{C}$ whose order is at most 2. If $g\colon \mathbf{C}^n \to \mathbf{C}$ is this extension then for $z \in \mathbf{C}^n$*

$$|g(z)| \leqq \alpha \| f \|_p^{2(p-1)} \exp\{\beta |z|^2\} .$$

(*c*) *For $x \in \mathbf{R}^n$*

$$|f(x)| \leqq \alpha \| f \|_p \exp\{ -\beta(x, x)\} .$$

Finally, if $f_j \in L^p(\mathbf{R}^n)$ for $j = 1, 2, 3, \ldots$ is an L^p bounded maximizing sequence for G(i.e., $\mathscr{R}_{p\to q}(f_j) \to C_{p\to q}$) then there is a function $f \in L^p(\mathbf{R}^n)$ and a subsequence $j(1), j(2), \ldots$ such that $f_{j(k)} \to f$ strongly in $L^p(\mathbf{R}^n)$ as $k \to \infty$. If $f \not\equiv 0$ (i.e., if $\| f_j \|_p \to 0$ as $j \to \infty$) then f is a maximizer.

Proof. For any $f \in L^p(\mathbf{R}^n)$, Hölders inequality can be used to deduce

$$|(\mathscr{G} f)(x)| \leqq T(x) \| f \|_p \tag{1}$$

with $T(x) = \| G(x, \cdot) \|_{p'}$. Simple computation shows that there are positive numbers γ and δ depending only on G and p such that $|T(x)| \leqq \gamma \exp\{ -\delta(x, x)\}$. The fact that G is nondegenerate is crucial for this result. The fact that $T \in L^1(\mathbf{R}^n) \cap L^\infty(\mathbf{R}^n)$ shows that $\mathscr{G}$ is bounded from $L^p(\mathbf{R}^n)$ to $L^q(\mathbf{R}^n)$. Now suppose that $f_j \in L^p(\mathbf{R}^n)$ is a sequence that converges weakly in $L^p(\mathbf{R}^n)$ to some $f \in L^p(\mathbf{R}^n)$ as $j \to \infty$. Since, for each $x \in \mathbf{R}^n$, $G(x, \cdot)$ is in $L^{p'}(\mathbf{R}^n)$, it follows that $(\mathscr{G} f_j)(x) \to (\mathscr{G} f)(x)$ as $j \to \infty$ for each $x \in \mathbf{R}^n$. It can be assumed that the f_j and f satisfy $\| f_j \|_p$ and $\| f \|_p \leqq C$ for some $C > 0$ and hence, from (1), the functions $\mathscr{G} f_j$ and $\mathscr{G} f$ are bounded pointwise by the function CT. Since $T \in L^q(\mathbf{R}^n)$, $\| \mathscr{G} f_j - \mathscr{G} f \|_q \to 0$ by dominated convergence. Thus $\mathscr{G}$ takes weakly convergent sequences in $L^p(\mathbf{R}^n)$ into strongly convergent sequences in $L^q(\mathbf{R}^n)$, and so $\mathscr{G}$ is compact.

Now let f_j be a bounded maximizing sequence, i.e., $\mathscr{R}_{p\to q}(f_j) \to C_{p\to q}$ as $j \to \infty$. We can assume $\|f_j\|_p = 1$ for each j. By the Banach-Alaoglu theorem, there is an $f \in L^p(\mathbf{R}^n)$ and a subsequence $j(1), j(2), \ldots$ such that $f_j \to f$ weakly in $L^p(\mathbf{R}^n)$. As is well known, $\|f\|_p \leqq 1$. Then, by the strong convergence proved above

$$C_{p\to q} = \lim_{k\to\infty} \|\mathscr{G} f_{j(k)}\|_q = \|\mathscr{G} f\|_q \leqq C_{p\to q}\|f\|_p \leqq C_{p\to q}\,.$$

This implies that $\|f\|_p = 1$ and that f is a maximizer. Moreover, the fact that $\|f\|_p = 1$ implies (by the uniform convexity of the L^p norm) that $f_{j(k)}$ converges to f strongly in $L^p(\mathbf{R}^n)$. Thus, the first and last assertions of the lemma have been proved.

It remains to prove that a maximizer f satisfies conditions (a), (b) and (c) and it suffices to assume that $\|f\|_p = 1$. There is a function $h \in L^{q'}(\mathbf{R}^n)$ such that $\|h\|_{q'} = 1$ and $C_{p\to q} = \|\mathscr{G} f\|_q = \int h(x)(\mathscr{G} f)(x)\mathrm{d}x$. Let

$$m(y) = \int G(x, y)h(x)\mathrm{d}x = \mathrm{e}^{-(y,By)} \int \mathrm{e}^{-(x,Ax)-2(x,Dy)} h(x)\mathrm{d}x \tag{2}$$

so that, as in the proof of (1) above, $|m(y)| \leqq W(y) \equiv \mu \exp\{-\nu(y, y)\}$ for suitable positive numbers μ and ν which depend only G and q. Hölder's inequality implies that the function $(x, y) \mapsto h(x)G(x, y)f(y)$ is in $L^1(\mathbf{R}^n \times \mathbf{R}^n)$, and Fubini's theorem then implies that $\|\mathscr{G} f\|_q = \int m(y)f(y)\mathrm{d}y$. If $m(y) \equiv |m(y)|\exp\{\mathrm{i}\theta(y)\}$, the optimum choice for f is $f(y) = [|m(y)|/\|m\|_{p'}]^{p'-1} \exp\{-\mathrm{i}\theta(y)\}$, for otherwise $\mathscr{R}_{p\to q}(f)$ can be increased.

The function $m: \mathbf{R}^n \to \mathbf{C}$ has an extension to an entire analytic function on $\mathbf{C}^n$ of order at most 2. This can be seen easily from the representation (2) above and Hölder's inequality; if $y_j = u_j + \mathrm{i}v_j$ for $j = 1, \ldots, n$ and $D = E + \mathrm{i}H$ with u_j, v_j, E and H real then

$$\begin{aligned}|m(y)| &\leqq \exp\{(v, Bv) - (u, Bu)\}\left[\int \exp\{-q'(x, Ax) - 2q'(x, Eu) + 2q'(x, Hv)\}\mathrm{d}x\right]^{1/q'}\\ &= (\text{const.})\exp\{(v, Bv) - (u, Bu) + (Eu - Hv, A^{-1}(Eu - Hv))\}\,.\end{aligned}$$

Thus $|m(y)| \leqq (\text{const.})\exp\{(\text{const.})[(u, u) + (v, v)]\}$ which implies that the order of m is at most 2. This establishes conclusion (a). Since m is entire, the function $y \mapsto m^*(y) \equiv \overline{m(\bar{y})}$ (with the bar denoting complex conjugate) is also entire, and hence $N(y) \equiv m(y)m^*(y)$ is also entire with order at most 2 and with a pointwise bound that is independent of f. However, when $y \in \mathbf{R}^n$(i.e., $v_j = 0$ for all j) then $N(y) = |m(y)|^2$. Conclusion (b) is then an immediate consequence of the relation between f and m which implies that for $y \in \mathbf{R}^n$, $|f(y)|^{2(p-1)} = \|m\|_{p'}^{-2}|m(y)|^2 = \|m\|_{p'}^{-2}N(y)$; thus $|f|^{2(p-1)}$ has an analytic extension of order at most 2, namely $\|m\|_{p'}^{-2}N$. It only has to be shown that $\|m\|_{p'}^{-2}$ is universally bounded, but this follows from the relation $C_{p\to q} = \|\mathscr{G} f\|_q = \int mf = \|m\|_{p'}$.

Conclusion (c) follows from the fact that when $y \in \mathbf{R}^n$ then $|f(y)| = [|m(y)|/\|m\|_{p'}]^{p'-1} \leqq W(y)^{p'-1}\|m\|_{p'}^{1-p'}$. □

The next two lemmas validate the assertion in Sect. I that linear terms can be eliminated from Gaussians.

2.2. Lemma (elimination of linear terms from Gaussian kernels). *Let G be the (degenerate or nondegenerate) Gaussian kernel given in* (1.1) *with positive definite or semidefinite real quadratic form M in* (1.3) *and with real linear term $L \in \mathbf{R}^{2n}$. Let G_0 denote the Gaussian kernel with no linear term, which is obtained from G by setting*

$L=0$, *i.e.*, $G_0(x, y) = G(x, y)\exp\left\{-2\left(L, \begin{pmatrix} x \\ y \end{pmatrix}\right)\right\}$. *Let* $1 \leqq p \leqq \infty$ *and* $1 \leqq q \leqq \infty$. *Then the following conditions are equivalent.*

(*i*) $\mathscr{G}_0$ *is bounded from* $L^p(\mathbf{R}^n)$ *to* $L^q(\mathbf{R}^n)$ *and the equation* $MV = L$ *has a solution* $V \in \mathbf{R}^{2n}$.

(*ii*) $\mathscr{G}$ *is bounded from* $L^p(\mathbf{R}^n)$ *to* $L^q(\mathbf{R}^n)$.

In case these conditions are both satisfied the relation between the norms is

$$C_{p\to q}(G) = C_{p\to q}(G_0)\exp\{(L, V)\}\,.$$

The number (L, V) *is uniquely defined even if the vector* V *is not unique.* $\mathscr{G}$ *has a unique maximizer if and only if* $\mathscr{G}_0$ *has one.*

Proof. (i) $\Rightarrow$ (ii). This was explained in Sect. I. Simply change variables; writing $V = \begin{pmatrix} a \\ b \end{pmatrix}$, let $x \to x + a$ and $y \to y + b$. Then

$$G \to G_0\exp\left\{(L, V) - 2\mathrm{i}\left(\begin{pmatrix} x \\ y \end{pmatrix}, NV\right) - \mathrm{i}(V, NV)\right\}.$$

The imaginary terms above do not affect the norm. Since M is Hermitian L must be orthogonal to $\mathscr{K} \equiv$ kernel of $M \subset \mathbf{R}^{2n}$, while any two solutions V_1 and V_2 differ by an element of $\mathscr{K}$. Thus, (L, V) is unique. This change of variables also shows that $\mathscr{G}$ has a unique maximizer if and only if $\mathscr{G}_0$ has one.

(ii) $\Rightarrow$ (i). Suppose that $MV = L$ has no solution. Then, since M is Hermitean, L is not orthogonal to $\mathscr{K}$ and thus there is a vector $W = \begin{pmatrix} s \\ t \end{pmatrix} \in \mathscr{K}$ such that the number $P = (W, L)$ is positive. Make the change of variables $x \to x + s$ and $y \to y + t$. Then, since $MW = 0$, G becomes

$$\tilde{G}(x, y) = G(x, y)\exp\left\{-\mathrm{i}(W, NW) - 2\mathrm{i}\left(\begin{pmatrix} x \\ y \end{pmatrix}, NW\right) + 2P\right\}.$$

The change of variables is an isometry so the norm of $\tilde{\mathscr{G}}$ is the same as the norm of $\mathscr{G}$ and, since the imaginary terms are irrelevant, we have $C_{p\to q}(G) = C_{p\to q}(\tilde{G}) = e^{2P}C_{p\to q}(G)$. This is a contradiction since $C_{p\to q}(G) \neq 0$. Thus $MV = L$ has a solution and the same change of variables can be made as before to derive the relation between the norms of $\mathscr{G}$ and $\mathscr{G}_0$. $\square$

2.3. Lemma (elimination of linear terms from maximizers). *Let* G *be a centered Gaussian kernel* (*degenerate or nondegenerate*) *and let* $1 < p < \infty$ *and* $1 < q < \infty$. *Assume* $\mathscr{G}: L^p(\mathbf{R}^n) \to L^q(\mathbf{R}^n)$ *is bounded* (*which is automatically true in the nondegenerate case*). *If* $g(x) = \exp\{-(x, Jx) + (l, x)\}$ *is a Gaussian function that maximizes* $\mathscr{R}_{p\to q}(g)$ *among all Gaussian functions then* $g_0(x) = \exp\{-(x, Jx)\}$ *is also a maximizer. Moreover, if* $\mathscr{R}_{p\to q}(g)$ *does not have a maximizer among Gaussian functions* (*which can happen only if* G *is degenerate*) *then the supremum of* $\mathscr{R}_{p\to q}(g)$ *over Gaussian functions equals the supremum over centered Gaussian functions. Finally, if* G *is nondegenerate then* $g = g_0$, *i.e.*, $l = 0$, *and therefore* g *is centered.*

Proof. Consider the functions $g_\lambda(x) = \exp\{-(x, Jx) + \lambda(l, x)\}$ with λ a real parameter. Clearly $g_\lambda \in L^p(\mathbf{R}^n)$ for all λ and, by a well known property of Gaussian integrals,

$$\|g_\lambda\|_p = \|g_0\|_p e^{\alpha\lambda^2} \quad \text{and} \quad \|\mathcal{G}g_\lambda\|_q = \|\mathcal{G}g_0\|_q e^{\beta\lambda^2}$$

for some real constants α and β. There are three cases to be considered:

(*i*) $\alpha > \beta$. By setting $\lambda = 0$, $\mathcal{R}_{p\to q}$ is increased, i.e., $\mathcal{R}_{p\to q}(g_0) > \mathcal{R}_{p\to q}(g)$. This means that g is not a maximizer—which is a contradiction.
(*ii*) $\alpha < \beta$. By letting λ tend to infinity we conclude that $\mathcal{R}_{p\to q}$ (and hence also $\mathcal{G}$) is unbounded—which is a contradiction.
(*ii*) $\alpha = \beta$. In this case g_λ is a maximizer for every λ and hence g_0 is a maximizer, as claimed.

These considerations prove all but the last sentence of the lemma.

If G is nondegenerate it is possible to go further. Consider the following sequence of functions with $\lambda = j$, namely $h_j = Z_j g_j$ for $j = 1, 2, 3, \ldots$, where the numbers Z_j are chosen so that $\|h_j\|_p = 1$ for each j. This is a bounded maximizing sequence and, by a trivial modification of the last part of Lemma 2.1 (using the fact that a nonzero $L^p(\mathbf{R}^n)$ weak limit of Gaussian functions is a Gaussian function), there is a nonzero Gaussian function $h \in L^p(\mathbf{R}^n)$ and a subsequence $j(1), j(2), \ldots$ such that $h_{j(k)} \to h$ strongly in $L^p(\mathbf{R}^n)$ as $k \to \infty$. If $l \neq 0$, however, it is easy to check that $h_j \to 0$ weakly in $L^p(\mathbf{R}^n)$ as $j \to \infty$. This contradicts the supposed strong convergence to a nonzero function. □

III. Nondegenerate gaussian kernels

A main ingredient in the following theorems is Minkowski's inequality for integrals. It was exploited by Beckner [B1; B2] to prove that the sharp bound for the tensor product of two operators (e.g., Fourier transforms) is often the product of the individual bounds. In particular, the bound for the Fourier transform from $L^p(\mathbf{R}^n)$ to $L^{p'}(\mathbf{R}^n)$ is $(C_p^B)^n$, where C_p^B is the sharp constant for $\mathbf{R}^1$. A proof of Minkowski's inequality can be found in [HLP]. Of crucial importance here is the sharp form in which the necessary and sufficient condition for equality is specified; this condition was not used before to analyze Gaussian kernels.

3.1. Lemma (Minkowski's inequality). *Let* $f\colon \mathbf{R}^n \times \mathbf{R}^m \to [0, \infty]$ *be Lebesgue measurable and let* $1 \leqq r < \infty$. *Suppose that the measurable function* M, *defined for almost every* $x \in \mathbf{R}^n$ *by*

$$M(x) \equiv \int_{\mathbf{R}^m} f(x, y)^r \,\mathrm{d}y\,,$$

is finite for almost every x *and that* $M^{1/r} \in L^1(\mathbf{R}^n)$. *Then the measurable function*

$$N(y) \equiv \int_{\mathbf{R}^n} f(x, y)\,\mathrm{d}x$$

is finite for almost every $y \in \mathbf{R}^m$ *and*

$$\left\{\int_{\mathbf{R}^m} N^r\right\}^{1/r} \leqq \int_{\mathbf{R}^n} M^{1/r}\,. \qquad (*)$$

Furthermore, if $r > 1$ and if there is equality in $(*)$ then there are nonnegative, measurable functions $A \in L^1(\mathbf{R}^n)$ and $B \in L^r(\mathbf{R}^m)$ such that

$$f(x, y) = A(x)B(y)$$

for almost every $(x, y) \in \mathbf{R}^n \times \mathbf{R}^m$.

Remark. This lemma extends to an arbitrary pair of measure spaces (X, μ) and (Y, ν) in place of $(\mathbf{R}^n, \mathrm{d}x)$ and $(\mathbf{R}^m, \mathrm{d}y)$ when μ and ν are sigma finite.

As a first application of Minkowski's inequality the uniqueness of maximizers for real, nondegenerate Gaussian kernels for *all p* and *q* will be proved. This is Case (A) of Section I. It is to be noted that the order of integration in Theorem 3.2 is as in [S] and is opposite to that of Theorem 3.4 and opposite to the order in Beckner's lemma. Analyticity considerations play only a subsidiary role in Theorem 3.2 and can be bypassed if desired, but they are important later. Theorem 3.2 was already essentially contained in [BL] Theorems 7 and 13. The following proof is offered because (*i*) it is different from the [BL] approach and (*ii*) it illustrates the techniques of the present paper.

3.2. Theorem (unique Gaussian maximizer for all p and q in the real nondegenerate case). *Let G be a real, nondegenerate, centered Gaussian kernel, i.e., the matrix N in (1.3) is zero. Let $1 < p < \infty$ and $1 < q < \infty$. Then $\mathscr{G}$ has exactly one maximizer, f, (up to a multiplicative constant) from $L^p(\mathbf{R}^n)$ to $L^q(\mathbf{R}^n)$ and f is a real, centered Gaussian, i.e., $f(x) = \exp\{-(x, Jx)\}$ with J being a real, positive definite matrix.*

Proof. Consider the linear operator $\mathscr{G}^{(2)} = \mathscr{G} \otimes \mathscr{G} : L^p(\mathbf{R}^{2n}) \to L^q(\mathbf{R}^{2n})$ given by the Gaussian kernel $G^{(2)}((x_1, x_2), (y_1, y_2)) = G(x_1, y_1)G(x_2, y_2)$ with x_1, x_2, y_1 and y_2 in $\mathbf{R}^n$. The first goal is to prove that $C_{p\to q}(G^{(2)}) = C_{p\to q}(G)^2$. If $F \in L^p(\mathbf{R}^{2n})$ then $(y_1, y_2) \mapsto G^{(2)}((x_1, x_2), (y_1, y_2))\, F(y_1, y_2)$ is in $L^1(\mathbf{R}^{2n})$ for every (x_1, x_2) because $G^{(2)}$ is nondegenerate. Fubini's theorem and Minkowski's inequality yield

$$\|\mathscr{G}^{(2)}F\|_q^q = \int \Big\{ \int \Big| \int \Big(\int G(x_1, y_1)G(x_2, y_2)F(y_1, y_2)\mathrm{d}y_1 \Big)\mathrm{d}y_2 \Big|^q \mathrm{d}x_1 \Big\}\, \mathrm{d}x_2 \tag{1}$$

$$\leqq \int \Big\{ \int \Big[\int G(x_2, y_2)|K(x_1, y_2)|\mathrm{d}y_2 \Big]^q \mathrm{d}x_1 \Big\} \mathrm{d}x_2$$

$$\text{(with } K(x_1, y_2) = \int G(x_1, y_1)F(y_1, y_2)\mathrm{d}y_1) \tag{2}$$

$$\leqq \int \Big\{ \int \Big[\int G(x_2, y_2)^q |K(x_1, y_2)|^q \mathrm{d}x_1 \Big]^{1/q} \mathrm{d}y_2 \Big\}^q \mathrm{d}x_2 \tag{3}$$

$$\leqq (C_{p\to q}(G))^q \int \Big\{ \int G(x_2, y_2)\Big[\int |F(y_1, y_2)|^p \mathrm{d}y_1 \Big]^{1/p} \mathrm{d}y_2 \Big\}^q \mathrm{d}x_2 \tag{4}$$

$$\leqq (C_{p\to q}(G))^{2q} \Big\{ \int\int |F(y_1, y_2)|^p \mathrm{d}y_1\, \mathrm{d}y_2 \Big\}^{q/p}. \tag{5}$$

(Notes: (2) $\to$ (3) is Minkowski's inequality. (3) $\to$ (4) uses $C_{p\to q}(G) \geqq \mathscr{R}_{p\to q}(F(\cdot, y_2))$ for each y_2. (4) $\to$ (5) uses $C_{p\to q}(G) \geqq \mathscr{R}_{p\to q}((\int |F(y_1, \cdot)|^p \mathrm{d}y_1)^{1/p})$. The fact that $G(x, y) \geqq 0$ is crucial. Here the x_1 integration was done before the x_2 integration; in Theorem 3.4 the x_2 integration will be done first.) Inequalities (1)–(5) establish that $C_{p\to q}(G^{(2)}) \leqq C_{p\to q}(G)^2$. Clearly, by considering F's of the product form $F(y_1, y_2) = h(y_1)h(y_2)$, the reverse inequality is obtained, and so the goal is reached.

Suppose now that $F : \mathbf{R}^{2n} \to \mathbf{C}$ is a maximizer for $G^{(2)}$. Since $G^{(2)}$ is nondegenerate, it has a maximizer by Lemma 2.1. Since $G(x, y) > 0$ for all x and y, it is clear that $F = \lambda|F|$ and $|\lambda| = 1$, for otherwise replacing F by $|F|$ will increase the quotient $\mathscr{R}_{p\to q}$ for $G^{(2)}$. It can be assumed henceforth that $F \geqq 0$. Since F is a maximizer all the inequalities in (1)–(5) must be equalities. Equality of (2) and (3) implies, by Lemma 3.1, that for almost every x_2 there are measurable functions A_{x_2} and $B_{x_2}: \mathbf{R}^n \to [0, \infty)$ such that

$$G(x_2, y_2)K(x_1, y_2) = A_{x_2}(x_1)B_{x_2}(y_2) \tag{6}$$

for almost every x_1 and y_2. Since $G > 0$, this equation can be divided by $G(x_2, y_2)$ to obtain $K(x_1, y_2) = A_{x_2}(x_1)E_{x_2}(y_2)$ with $E_{x_2}(y) \equiv B_{x_2}(y)/G(x_2, y)$. However, $K(x_1, y_2)$ is independent of x_2 and therefore if any particular value of x_2 is chosen for which (6) holds for almost every x_1 and y_2, and if the functions A and $E: \mathbf{R}^n \to [0, \infty)$ are defined by $A \equiv A_{x_2}$ and $E \equiv E_{x_2}$ for this value of x_2, then

$$K(x_1, y_2) = A(x_1)E(y_2)$$

for almost every x_1 and y_2. If this equation is multiplied by $G(x_2, y_2)$ and integrated over y_2 the result is

$$(\mathscr{G}^{(2)}F)(x_1, x_2) = A(x_1)Z(x_2)$$

for almost every x_1 and x_2 with $Z = \mathscr{G}E$. Since $G^{(2)} > 0$, both A and Z are strictly positive functions.

There is a function $H \in L^{q'}(\mathbf{R}^{2n})$ with $\|H\|_{q'} = 1$, such that $\|\mathscr{G}^{(2)}F\|_q = \int H \cdot \mathscr{G}^{(2)}F$. In fact

$$H(x_1, x_2) = (\text{const.})[(\mathscr{G}^{(2)}F)(x_1, x_2)]^{q-1} = (\text{const.})A(x_1)^{q-1}Z(x_2)^{q-1} .$$

The point here is that H is a product function. Then, as in the proof of Lemma 2.1, F satisfies

$$F(y_1, y_2) = (\text{const.})\left\{\iint G(x_1, y_1)G(x_2, y_2)H(x_1, x_2)\mathrm{d}x_1\,\mathrm{d}x_2\right\}^{p'-1} = \alpha(y_1)\beta(y_2) \tag{7}$$

for some positive function α and $\beta: \mathbf{R}^n \to [0, \infty)$. In brief, F *must be* a product function, and this fact is crucial for the next step.

One example of a maximizer is $F(y_1, y_2) = f(y_1)f(y_2)$, where f is an $L^p(\mathbf{R}^n) \to L^q(\mathbf{R}^n)$ maximizer for G (whose existence is guaranteed by Lemma 2.1). For the reason given before about F, we can and do assume that $f(x) \geqq 0$ for all $x \in \mathbf{R}^n$.

A more interesting maximizer is

$$F(y_1, y_2) = f\left(\frac{y_1 - y_2}{\sqrt{2}}\right)f\left(\frac{y_1 + y_2}{\sqrt{2}}\right). \tag{8}$$

Here, the essential property of $O(2)$ rotation invariance of products of centered Gaussians and of Lebesgue measure is being exploited. If θ is any fixed angle and if x'_1, x'_2, y'_1, y'_2 in $\mathbf{R}^n$ are defined by $x'_1 = x_1 \cos\theta - x_1 \sin\theta$, $x'_2 = x_1 \sin\theta + x_2 \cos\theta$, $y'_1 = y_1 \cos\theta - y_2 \sin\theta$, $y'_2 = y_1 \sin\theta + y_2 \cos\theta$, the $O(2)$ invariance of Lebesgue measure is that $\mathrm{d}x_1\mathrm{d}x_2 = \mathrm{d}x'_1\mathrm{d}x'_2$ and $\mathrm{d}y_1\mathrm{d}y_2 = \mathrm{d}y'_1\mathrm{d}y'_2$. The $O(2)$ invariance of centered Gaussian functions is that $g(x_1)g(x_2) = g(x'_1)g(x'_2)$, while for centered Gaussian kernels $G(x_1, y_1)G(x_2, y_2) = G(x'_1, y'_1)G(x'_2, y'_2)$. With the choice $\theta = \pi/4$, these observations lead to (8). Combining (7) and (8),

$$f\left(\frac{y_1 - y_2}{\sqrt{2}}\right)f\left(\frac{y_1 + y_2}{\sqrt{2}}\right) = \alpha(y_1)\beta(y_2) . \tag{9}$$

for almost every y_1 and y_2.

Equation (9) implies that f is a Gaussian. Instead of proving this in full generality for $L^p(\mathbf{R}^n)$ functions, as is done by Carlen [CA], it is easier to simplify the proof here by taking the $2(p-1)^{th}$ power of (9) and by taking advantage of the analyticity result Lemma 2.1(b). Introducing $h = f^{2(p-1)}$, $\gamma = \alpha^{2(p-1)}$ and $\delta = \beta^{2(p-1)}$, it is seen from (9) (by fixing y_2) that γ is analytic; likewise δ is analytic. Thus, (9) holds for *all* y_1 and y_2 because when two analytic functions on $\mathbf{C}^n \times \mathbf{C}^n$ agree almost everywhere on $\mathbf{R}^n \times \mathbf{R}^n$ then they agree everywhere. Furthermore f never vanishes for real y because if $f(Y) = 0$ then, setting $y_1 = y_2 + \sqrt{2}\,Y$, we would have that $0 = \gamma(y_2 + \sqrt{2}\,Y)\delta(y_2)$ for all y_2; this is impossible, given that γ and δ are analytic, unless $\gamma \equiv 0$ or $\delta \equiv 0$, which contradicts the assumption that $f \not\equiv 0$. Thus, the logarithms of h, γ and δ are real analytic and

$$\ln\left[h\left(\frac{y_1 - y_2}{\sqrt{2}}\right)\right] + \ln\left[h\left(\frac{y_1 + y_2}{\sqrt{2}}\right)\right] = \ln[\gamma(y_1)] + \ln[\delta(y_2)]\,. \tag{10}$$

If ∂_i denotes the derivative with respect to the i^{th} coordinate, and ∂_i with respect to y_1 and ∂_j with respect to y_2 is taken in (10), then

$$(\partial_i\partial_j \ln h)\left(\frac{y_1 - y_2}{\sqrt{2}}\right) = (\partial_i\partial_j \ln h)\left(\frac{y_1 + y_2}{\sqrt{2}}\right),$$

which implies that the function $\partial_i\partial_j \ln h$ is a constant (call it $4(1-p)J_{ij}$) and therefore $\ln[f(y)] = \frac{1}{2(p-1)}\ln[h(y)] = -(y, Jy) + (l, y)$ for some vector l. According to Lemma 2.3, $l = 0$ since G is centered and nondegenerate. This completes the proof that f must be a centered Gaussian.

It remains to prove that f is unique (i.e., the matrix J above is unique). One way would be to compute $\mathscr{R}_{p\to q}(\exp\{-(x, Jx)\})$ for G and then deduce that there is only one optimum J. A very much easier route is to suppose that there are two maximizers f^1 and f^2 with $f^i(y) = \exp\{-(y, J^i y)\}$. Then, for the same reason as before ($O(2)$ symmetry) the function

$$F(y_1, y_2) = f^1\left(\frac{y_1 - y_2}{\sqrt{2}}\right) f^2\left(\frac{y_1 + y_2}{\sqrt{2}}\right) \tag{11}$$

is a maximizer for $\mathscr{G}^{(2)}$. There are two ways in which this implies that $f^1 = f^2$. The first is to use (7), namely F must be a product function, and to note that this product structure is true if and only if $J^1 = J^2$. The second way is to note that since the F in (11) is never zero and, since (3) $\to$ (4) must be an equality, we have that the function $y_1 \mapsto h_{y_2}(y_1) \equiv F(y_1, y_2)$ must be a maximizer for $\mathscr{G}$ for almost every y_2. Although the function h_{y_2} is a Gaussian for each y_2, the Gaussian will have a linear term for each $y_2 \neq 0$ unless $J^1 = J^2$. However, Lemma 2.3 precludes the existence of such a linear term, so $J^1 = J^2$. $\square$

The next theorem concerns Case (B) of Sect. I.

3.3. Theorem (unique Gaussian maximizers in the imaginary, nondegenerate case). *Let G be a centered, nondegenerate Gaussian kernel with a real diagonal part and a purely imaginary off-diagonal part, i.e.,*

$$G(x, y) = \exp\{-(x, Ax) - (y, By) - 2i(x, Dy)\}$$

where A, B and D are real $n \times n$ matrices and A and B are positive definite. Let $1 < p \leqq 2$ and $1 < q < \infty$ or else $1 < p < \infty$ and $2 \leqq q \leqq \infty$. Then, in either case, $\mathscr{G}$ has exactly one maximizer, f, (up to a multiplicative constant) from $L^p(\mathbf{R}^n)$ to $L^q(\mathbf{R}^n)$ and this f is a real, centered Gaussian, i.e., $f(x) = \exp\{-(x, Jx)\}$ with J being a real, positive definite matrix.

Proof. Assume at first that D is nonsingular. Since A and B are positive definite there are nonsingular real matrices U and V so that the change of variables $x \to Ux$ and $y \to Vy$ changes A and B to the identity matrix, I, that is $I = U^T A U = V^T B V$, where T denotes transpose. Then $(x, Dy) \to (x, \tilde{D}y)$ with $\tilde{D} = U^T D V$. The polar decomposition of $\tilde{D}$ is $\tilde{D} = W|\tilde{D}|$, where W is orthogonal and $|\tilde{D}|$ is positive definite (the assumption that D is nonsingular is used here). Then there is an orthogonal matrix Y such that $Y^T|\tilde{D}|Y$ is diagonal and there is a real diagonal matrix Z such that $ZY^T|\tilde{D}|YZ = I$. Now make one more change of variables: $x \to WYZx$ and $y \to YZy$ so that $(x, \tilde{D}y) \to (WYZx,\ W|\tilde{D}|YZy) = (x, y)$ and $(x, x) = (x, Ix) \to (WYZx,\ WYZx) = (x, Z^2x)$ and $(y, y) \to (YZy, YZy) = (y, Z^2y)$. These two changes of variables affect $\mathscr{R}_{p\to q}$ in a trivial way (involving only p and q and the determinants of U, V and Z) and, most importantly, take Gaussian functions into Gaussian functions. In short, it can be assumed without loss of generality that G has the canonical form

$$G(x, y) = \exp\{-(x, Ax) - (y, Ay) - 2i(x, y)\}\ , \tag{1}$$

where A is positive definite and diagonal.

By duality $C_{p\to q}(G) = C_{q'\to p'}(G^T)$ with $G^T(x,\ y) \equiv G(y, x) = G(x, y)$, so it suffices to consider only the case $1 < p \leqq 2$ and $1 < q < \infty$. It is easily seen that $(\mathscr{G}f)(x) = \exp\{-(x, Ax)\}\hat{h}(x)$ where $\hat{h}$ is the Fourier transform of the function $h(y) = \exp\{-(y, Ay)\}f(y)$. Since $f \in L^p(\mathbf{R}^n)$ it has a Fourier transform $\hat{f}$, and Beckner's theorem (which will also be proved here in Theorem 4.1 and 4.2(1)) states that $\|\hat{f}\|_{p'} \leqq (C_p^B)^n \|f\|_p$, where Beckner's constant C_p^B is the sharp constant for the $p \to p'$ norm of the Fourier transform in $\mathbf{R}^1$. By the convolution formula, $\hat{h}$ satisfies

$$\hat{h}(x) = \mu \int \exp\{-(x - y, A^{-1}(x - y))\}\hat{f}(y)dy\ ,$$

where $\mu > 0$ is a constant which depends only on A. Therefore $(\mathscr{G}f)(x) = \mu(\tilde{\mathscr{G}}\hat{f})(x)$ where $\tilde{G}$ is the real, centered, nondegenerate Gaussian

$$\tilde{G}(x, y) = \exp\{-(x, Ax) - (x - y, A^{-1}(x - y))\}\ . \tag{2}$$

Thus

$$\mathscr{R}_{p\to q}(G, f)\|f\|_p = \mu\mathscr{R}_{p'\to q}(\tilde{G}, \hat{f})\|\hat{f}\|_{p'} \leqq \mu C_{p'\to q}(\tilde{G})\|\hat{f}\|_{p'} \leqq \mu C_{p'\to q}(\tilde{G})(C_p^B)^n\|f\|_p\ , \tag{3}$$

from which it follows that $C_{p\to q}(G) \leqq \mu(C_p^B)^n C_{p'\to q}(\tilde{G})$. However, equality can be achieved in (3) in exactly one way (up to a multiplicative constant). By Theorem 3.2 there is exactly one choice for $\hat{f}$ that will make the first inequality in (3) into an equality. This $\hat{f}$ is a real, centered Gaussian, $\hat{f}(x) = \exp\{-(x, Jx)\}$. Its inverse transform f is also a real, centered Gaussian, i.e., $f(x) = (\text{const.})\exp\{-(x, J^{-1}x)\}$. The second inequality in (3) (Beckner's) is an equality for any *real* Gaussian (in particular, our f), and therefore f is the unique maximizer as asserted in the theorem.

In case D is singular, a change of variables similar to the above replaces the canonical form (1) by

$$G(x, y) = \exp\{-(x, Ax) - (y, Ay) - 2i(x, Py)\}$$

where $P = \begin{pmatrix} I & 0 \\ 0 & 0 \end{pmatrix}$ is a diagonal projection onto $\mathbf{R}^m$ (with $m < n$ being the rank of D) and $A = \begin{pmatrix} \alpha & 0 \\ 0 & I \end{pmatrix}$ with α positive definite, $m \times m$ and diagonal. Writing $x \in \mathbf{R}^n$ as (x_1, x_2) with $x_1 \in \mathbf{R}^m$ and $x_2 \in \mathbf{R}^{n-m}$, define $g: \mathbf{R}^m \to \mathbf{C}$ by

$$g(y_1) = \int_{\mathbf{R}^{n-m}} \exp\{-(y_2, y_2)\} f(y_1, y_2) \mathrm{d}y_2 \,, \tag{4}$$

and $G^*: \mathbf{R}^m \times \mathbf{R}^m \to (0, \infty)$ by $G^*(x_1, y_1) = \exp\{-(x_1, \alpha x_1) - (y_1, \alpha y_1) - 2i(x_1, y_1)\}$. Then, using Fubini's theorem and the same analysis as before with α and m in place of A and n, and with $\tilde{G}^*: \mathbf{R}^m \times \mathbf{R}^m \to (0, \infty)$ as given in (2),

$$\mathscr{R}_{p\to q}(G, f) \| f \|_{L^p(\mathbf{R}^n)} = v \mathscr{R}_{p\to q}(G^*, g) \| g \|_{L^p(\mathbf{R}^m)} \leqq v\mu C_{p'\to q}(\tilde{G}^*)(C_p^B)^m \| g \|_{L^p(\mathbf{R}^m)} \tag{5}$$

where v is the $L^q(\mathbf{R}^{n-m})$ norm of the Gaussian function $\exp\{-(x_2, x_2)\}$. As was the case for (3) and the subsequent argument, equality in (5) is uniquely achieved by a real, centered Gaussian

$$g(y_1) = \exp\{-(y_1, Ey_1)\} \,, \tag{6}$$

where E is a real, positive definite $m \times m$ matrix. By Hölder's or Minkowski's inequality, it follows from (4) that

$$\| g \|_{L^p(\mathbf{R}^m)} \leqq N \| f \|_{L^p(\mathbf{R}^n)} \tag{7}$$

where N is the $L^{p'}(\mathbf{R}^{n-m})$ norm of the Gaussian function $\exp\{-(y_2, y_2)\}$. Equality in (7) is compatible with (6) if and only if $f(y) = f(y_1, y_2) = (\text{const.})\exp\{-y_1, Ey_1) - (p'-1)(y_2, y_2)\}$. The reason is that equality in (7) requires that $f(y_1, y_2) = \lambda(y_1)\exp\{-(p'-1)(y_2, y_2)\}$ for almost every $y_1 \in \mathbf{R}^m$. Then, computing the integral in (4), one finds that $\exp\{-(y_1, Ey_1)\} = g(y_1) = (\text{const.})\, \lambda(y_1)$. □

Finally, Case (C) of Sect. I will be considered.

3.4. Theorem (unique Gaussian maximizer when $p \leqq q$ in the general nondegenerate case). *Let G be a centered, nondegenerate Gaussian kernel and let $1 < p \leqq q < \infty$. Then $\mathscr{G}$ has exactly one maximizer, f, (up to a multiplicative constant) from $L^p(\mathbf{R}^n)$ to $L^q(\mathbf{R}^n)$ and f is a centered Gaussian function.*

Proof. As in the proof of Theorem 3.2, the key is to study the kernel $G^{(2)} = G \otimes G$ by means of Minkowski's inequality. Now, however, the x_2 integration is done first. Thus, for $F \in L^p(\mathbf{R}^{2n})$,

$$\| \mathscr{G}^{(2)} F \|_q^q = \int \{ \int | \int G(x_2, y_2)(\int G(x_1, y_1) F(y_1, y_2) \mathrm{d}y_1) \mathrm{d}y_2 |^q \mathrm{d}x_2 \} \mathrm{d}x_1 \tag{1}$$

$$= \int \{ \int | \int G(x_2, y_2) K(x_1, y_2) \mathrm{d}y_2 |^q \mathrm{d}x_2 \} \mathrm{d}x_1$$

$$\text{(with } K(x_1, y_2) = \int G(x_1, y_1)\, F(y_1, y_2) \mathrm{d}y_1) \tag{2}$$

$$\leqq (C_{p\to q}(G))^q \int \{ \int | K(x_1, y_2) |^p \mathrm{d}y_2 \}^{q/p} \mathrm{d}x_1 \tag{3}$$

$$\leqq (C_{p\to q}(G))^q \{ \int [\int | K(x_1, y_2) |^q \mathrm{d}x_1]^{p/q} \mathrm{d}y_2 \}^{q/p} \tag{4}$$

$$\leqq (C_{p\to q}(G))^{2q} \{ \iint | F(y_1, y_2) |^p \mathrm{d}y_1 \mathrm{d}y_2 \}^{q/p} \,. \tag{5}$$

(Notes: (2) → (3) uses $C_{p\to q}(G) \geqq \mathscr{R}_{p\to q}(K(x_1,\cdot))$ for each x_1. (3) → (4) is Minkowski's inequality for the exponent $r = q/p \geqq 1$ and the function $|K(x_1, y_2)|^p$. (4) → (5) uses $C_{p\to q}(G) \geqq \mathscr{R}_{p\to q}(F(\cdot, y_2))$ for each y_2.) This inequality (along with consideration of F's of the form $F(y_1, y_2) = h(y_1)h(y_2)$) shows that

$$C_{p\to q}(G^{(2)}) = C_{p\to q}(G)^2 \ .$$

Suppose now that $F: \mathbf{R}^{2n} \to \mathbf{C}$ is a maximizer for $\mathscr{G}^{(2)}$. Since $G^{(2)}$ and G are nondegenerate, maximizers exist for each of them by Lemma 2.1. Then all the inequalities in (1)–(5) must be equalities. In particular, inequality (4) → (5) implies that the function $y_1 \mapsto F(y_1, y_2)$ must either be the zero function or it must be a maximizer for $\mathscr{G}$ for almost every $y_2 \in \mathbf{R}^n$. (It is well known that this function is in $L^p(\mathbf{R}^n)$ for almost every y_2.) As in the proof of Theorem 3.2, the $O(2)$ invariance of $G^{(2)}$ implies that the function given by

$$F(y_1, y_2) = f\left(\frac{y_1 - y_2}{\sqrt{2}}\right) f\left(\frac{y_1 + y_2}{\sqrt{2}}\right) \tag{6}$$

is a maximizer for $\mathscr{G}^{(2)}$ when f is a maximizer for $\mathscr{G}$, as will henceforth be assumed. Thus, for almost every z in $\mathbf{R}^n$, the function $g_z(y) = F(y, z)$ is in $L^p(\mathbf{R}^n)$ and either (a) it is a maximizer for $\mathscr{G}$ or (b) g_z is the zero function. The second possibility (b) can be excluded by Lemma 2.1 (b). If $g_z \equiv 0$ then, from (6), $f(w) = 0$ for all w in some set $A \subset \mathbf{R}^n$ of positive Lebesgue measure. But $|f|^{2(p-1)}$ is analytic and this is impossible unless $f \equiv 0$. Thus it can be assumed that g_z is indeed a maximizer for almost every z, i.e., $g_z \not\equiv 0$.

In fact g_z is an $L^p(\mathbf{R}^n)$ maximizer for *every* $z \in \mathbf{R}^n$. To prove this assertion, fix z and let $z_1, z_2, \ldots$ be any sequence in $\mathbf{R}^n$ such that $z_j \to z$ as $j \to \infty$ and such that g_{z_j} is an $L^p(\mathbf{R}^n)$ maximizer for each j. Such a sequence exists because g_z is a maximizer for z's in a dense set. Define $h_j(y) = Z_j g_{z_j}(y)$ where Z_j is chosen so that $\|h_j\|_p = 1$ for each j. By Lemma 2.1, there is a subsequence (still denoted by h_j) and a maximizing function $h \in L^p(\mathbf{R}^n)$ such that $h_j \to h$ strongly as $j \to \infty$. By passing to a further subsequence this convergence can also be assumed to be pointwise almost everywhere. However, translation is a continuous operation in $L^p(\mathbf{R}^n)$ and thus, by passing to a further subsequence, $f((y + z_j)/\sqrt{2})$ converges pointwise to $f((y + z)/\sqrt{2})$ for almost every y. Likewise, by passing to a further subsequence, $f((y - z_j)/\sqrt{2})$ converges pointwise to $f((y - z)/\sqrt{2})$ for almost every y. It follows then that the maximizer h satisfies

$$h(y) = f\left(\frac{y - z}{\sqrt{2}}\right) f\left(\frac{y + z}{\sqrt{2}}\right) \lim_{j\to\infty} Z_j$$

for almost every y. Therefore $\lim_{j\to\infty} Z_j$ exists and g_z is a maximizer for every $z \in \mathbf{R}^n$.

Our first application of this result will be the proof that there is a Gaussian maximizer. Take $z = 0$ so that $f^{(2)}(y) \equiv f(y/\sqrt{2})^2$ is a maximizer. Then apply the same conclusion to $f^{(2)}$ so that $f^{(4)}(y) \equiv f(y/2)^4$ is also a maximizer. Repeating this indefinitely, the sequence of $L^p(\mathbf{R}^n)$ functions given by

$$g_j(y) = N_j f\left(\frac{y}{\sqrt{j}}\right)^j \tag{7}$$

is a sequence of maximizers for $j = 2, 4, 8, 16, \ldots$. The number N_j is chosen in each case so that $\|g_j\|_p = 1$. Using Lemma 2.1 again we infer the existence of a subsequence (still denoted by j) and a maximizer g such that $g_j \to g$ strongly in $L^p(\mathbf{R}^n)$ and pointwise almost everywhere. Our goal will be to prove that g is a Gaussian. This can be inferred from the central limit theorem, but the following argument is more direct and will be needed later for the proof that *every* maximizer is a Gaussian.

The first step is to prove that $f(0) \neq 0$. Recall from Lemma 2.1(b) that $R \equiv |f|^{2(p-1)}$ is analytic. Likewise $S \equiv |g|^{2(p-1)}$ is also analytic and

$$S(y) = \lim_{j\to\infty} N_j^{2(p-1)} R\left(\frac{y}{\sqrt{j}}\right)^j \tag{8}$$

for almost all $y \in \mathbf{R}^n$. Since $S_j(y) \equiv N_j^{2(p-1)} R(y/\sqrt{j})^j$ is the $2(p-1)^{th}$ power of the modulus of a maximizer with unit $L^p(\mathbf{R}^n)$ norm (namely g_j), Lemma 2.1(b) states that the analytic extension of S_j is uniformly bounded on compact subsets of $\mathbf{C}^n$. The almost everywhere convergence in (8) then implies (by Vitali's theorem) that (8) holds for all $y \in \mathbf{C}^n$ and that all partial derivatives with respect to y of the sequence of functions S_j also converge as $j \to \infty$ to the corresponding derivatives of S. However, it is easily seen by Leibniz's rule that if $R(0) = 0$ then *every* derivative of S_j at $y = 0$ converges to zero as $j \to \infty$. This is impossible unless $S(y)$ vanishes identically, which contradicts the fact that $\|g\|_p = 1$.

The second step is to prove that g is a Gaussian. By Lemma 2.1(a), for $y \in \mathbf{R}^n$, $f(y) = |m(y)|^{p'}/m(y)$, where $m: \mathbf{C}^n \to \mathbf{C}$ is entire analytic. Since $f(0) \neq 0$, also $m(0) \neq 0$ and hence there is a neighborhood U of $0 \in \mathbf{C}^n$ on which f has an analytic extension and on which f is never zero. [Reason: $m_1(y) \equiv \mathrm{Re}(m(y))$ can be written as a Taylor series for $y \in \mathbf{R}^n$, and so can $m_2(y) \equiv \mathrm{Im}(m(y))$. Consequently m_1 and m_2 extend to entire functions. Then $(m_1^2 + m_2^2)^{p'/2}$ is analytic on U since $m_1(0)^2 + m_2(0)^2 = |m(0)|^2 \neq 0$.] Therefore f has a logarithm, H, which is analytic on U, i.e., $f(y) = f(0)\exp\{H(y)\}$. The function H can be written as $H(y) = (V, y) - (y, Jy) + O(y^3)$ for some $V \in \mathbf{C}^n$ and J a symmetric matrix. For each $y \in \mathbf{R}^n$, the point $y/\sqrt{j}$ lies in U for all sufficiently large j and therefore, by (7),

$$g(y) = \lim_{j\to\infty} N_j f(0)^j \exp\{\sqrt{j}(V, y) - (y, Jy) + O(y^3 j^{-1/2})\}$$

for almost every $y \in \mathbf{R}^n$. The factor $\exp\{O(y^3 j^{-1/2})\}$ converges to 1 as $j \to \infty$ and thus

$$g(y) = \exp\{-(y, Jy)\} \lim_{j\to\infty} N_j f(0)^j \exp\{\sqrt{j}(V, y)\}\ .$$

Clearly this last limit can exist for almost every y if and only if $V = 0$ and $N_j f(0)^j$ has a finite limit (which cannot be zero since $\|g\|_p = 1$). This proves that g must be a Gaussian as claimed (and hence $\mathrm{Re}(J)$ is positive definite) but we also note that the argument also proves the following three statements: Whenever f is a maximizer then (*i*) f is analytic in some complex neighborhood of 0; (*ii*) $f(0) \neq 0$; (*iii*) $(\partial f/\partial y^i)(0) = 0$, for $i = 1, \ldots, n$.

The second assertion of the theorem is that every other maximizer, f, is proportional to the one just found, namely $g(y) = \exp\{-(y, Jy))\}$. Instead of (6) take

$$F(y_1, y_2) = g\left(\frac{y_1 - y_2}{\sqrt{2}}\right) f\left(\frac{y_1 + y_2}{\sqrt{2}}\right)$$

which is obviously also a maximizer for $\mathscr{G}^{(2)}$. By the same reasoning as before, F has the property that $y \mapsto k_z(y) \equiv F(y, z)$ is a maximizer for each fixed $z \in \mathbf{R}^n$. By the three statements just made above, we conclude that k_z is analytic near 0,

$$k_z(0) \neq 0 \quad \text{and} \quad (\partial k_z/\partial y^i)(0) = 0 .$$

This is equivalent to the statement that for every $z \in \mathbf{R}^n$, f is analytic near $z/\sqrt{2}$, $f(z/\sqrt{2}) \neq 0$ and

$$(\partial f/\partial y^i)\left(\frac{z}{\sqrt{2}}\right) = [-Jz]_i f\left(\frac{z}{\sqrt{2}}\right),$$

which shows that $f = g$. □

IV. Degenerate Gaussian kernels

In the three cases (A), (B) and (C) of Sect. I, which correspond to Theorems 3.2, 3.3 and 3.4, every nondegenerate Gaussian kernel has a unique maximizer which is a Gaussian function. By taking suitable limits the following formula 4.1(∗), which is one of the main results of this paper, can be deduced for the $L^p(\mathbf{R}^n)$ to $L^q(\mathbf{R}^n)$ norm of degenerate kernels. This formula is, of course, trivially true in the nondegenerate case.

4.1. Theorem (the sharp bound for degenerate kernels). *Let G be a centered Gaussian kernel as in* (1.1) *with $L = 0$ and let p and q satisfy the appropriate conditions given in* (*A*), (*B*) *or* (*C*) *of Sect. I, according to the properties of G. Then $\mathscr{G}$ is bounded from $L^p(\mathbf{R}^n)$ to $L^q(\mathbf{R}^n)$ if and only if the following supremum is finite, in which case the supremum is equal to $C_{p\to q}$.*

$$\sup_g \mathscr{R}_{p\to q}(g) = C_{p\to q} , \tag{$*$}$$

where the supremum is taken over all centered Gaussian functions, and in Cases (*A*) *and* (*B*) *they can be restricted to be real.*

Proof. For each $\varepsilon > 0$ let $h_\varepsilon(x) \equiv \exp\{-\varepsilon(x, x)\}$ and define $G_\varepsilon(x, y) \equiv G(x, y)h_\varepsilon(x)h_\varepsilon(y)$, which is nondegenerate. Correspondingly, there is the linear operator $\mathscr{G}_\varepsilon$. For each $f \in L^p(\mathbf{R}^n)$

$$\begin{aligned}\mathscr{R}_{p\to q}(G_\varepsilon, f)\|f\|_p = \|h_\varepsilon \mathscr{G}(h_\varepsilon f_\varepsilon)\|_q < \|\mathscr{G}(h_\varepsilon f_\varepsilon)\|_q \leqq C_{p\to q}(G)\|h_\varepsilon f_\varepsilon\|_p \\ < C_{p\to q}(G)\|f\|_p .\end{aligned} \tag{1}$$

This proves that $C_{p\to q}(G_\varepsilon) < C_{p\to q}(G)$.

On the other hand, assuming that $C_{p\to q}(G) < \infty$, for each $\delta > 0$ there is an $f^\delta \in L^p(\mathbf{R}^n)$ with $\|f^\delta\|_p = 1$ such that $\|\mathscr{G}f^\delta\|_q > C_{p\to q}(G) - \delta$. Then

$$C_{p\to q}(G_\varepsilon) \geqq \mathscr{R}_{p\to q}(G_\varepsilon, f^\delta) = \|\mathscr{G}_\varepsilon(f^\delta)\|_q = \|h_\varepsilon \mathscr{G}(h_\varepsilon f^\delta)\|_q . \tag{2}$$

As $\varepsilon \to 0$, $h_\varepsilon f^\delta \to f^\delta$ strongly in $L^p(\mathbf{R}^n)$, so $\mathscr{G}(h_\varepsilon f^\delta) \to \mathscr{G}(f^\delta)$ strongly in $L^q(\mathbf{R}^n)$. This implies that $h_\varepsilon \mathscr{G}(h_\varepsilon f^\delta) \to \mathscr{G}(f^\delta)$ strongly in $L^q(\mathbf{R}^n)$ as well, and thus, from (2), $\liminf_{\varepsilon\to 0} C_{p\to q}(G_\varepsilon) \geqq C_{p\to q}(G) - \delta$. Since δ is arbitrary, and in view of (1),

$$\lim_{\varepsilon\to 0} C_{p\to q}(G_\varepsilon) = C_{p\to q}(G) . \tag{3}$$

A similar argument shows that (3) holds even if $C_{p\to q}(G) = \infty$.

Now let g_ε denote the maximizer for G_ε, which is a centered Gaussian function. Assume $\|g_\varepsilon\|_p = 1$. Then

$$\|h_\varepsilon g_\varepsilon\|_p C_{p\to q}(G) \geqq \|h_\varepsilon g_\varepsilon\|_p \mathscr{R}_{p\to q}(G, h_\varepsilon g_\varepsilon) = \|\mathscr{G}(h_\varepsilon g_\varepsilon)\|_q$$
$$\geqq \|h_\varepsilon \mathscr{G}(h_\varepsilon g_\varepsilon)\|_q = \|\mathscr{G}_\varepsilon(g_\varepsilon)\|_q = C_{p\to q}(G_\varepsilon)\,. \tag{4}$$

Assuming $\mathscr{G}$ to be bounded, (4) together with (3) and the fact that $\|h_\varepsilon g_\varepsilon\|_p \leqq \|g_\varepsilon\|_p = 1$ implies that $\|h_\varepsilon g_\varepsilon\|_p \to 1$ as $\varepsilon \to 0$. Then

$$C_{p\to q}(G) = \lim_{\varepsilon\to 0} C_{p\to q}(G_\varepsilon) \leqq \lim_{\varepsilon\to 0} \mathscr{R}_{p\to q}(G, h_\varepsilon g_\varepsilon) \leqq C_{p\to q}(G)\,. \tag{5}$$

This proves the theorem in the bounded case since $h_\varepsilon g_\varepsilon$ is a Gaussian function (which is real in Cases (A) and (B)).

In case $\mathscr{G}$ is unbounded, (4) and (5) imply that

$$\infty = \lim_{\varepsilon\to 0} C_{p\to q}(G_\varepsilon) \leqq \lim_{\varepsilon\to 0} \|h_\varepsilon g_\varepsilon\|_p \mathscr{R}_{p\to q}(G, h_\varepsilon g_\varepsilon) \leqq \lim_{\varepsilon\to 0} \mathscr{R}_{p\to q}(G, h_\varepsilon g_\varepsilon)$$

which proves the theorem since $h_\varepsilon g_\varepsilon$ is a Gaussian function. □

4.2. Remarks and examples. Theorem 4.1(∗) is a formula for the $L^p(\mathbf{R}^n) \to L^q(\mathbf{R}^n)$ norm of $\mathscr{G}$. The same formula is, of course, also valid for nondegenerate kernels, but in that case we are assured that there is precisely one g that achieves the supremum. In the degenerate case a maximizer may not exist—*even if $\mathscr{G}$ is bounded*—as the examples below show. In any event, the evaluation of this formula is, in general, a difficult nonlinear algebraic exercise, although it is simple in many applications.

For example, when $G(x, y) = \exp\{2i(x, y)\}$ (the Fourier transform kernel), it is easy to deduce from 4.1(∗) that G is bounded if and only if $q = p' \geqq 2$, in which case a Gaussian function is a maximizer if and only if it has the form

$$g(x) = \mu \exp\{-(x, Jx) + (l, x)\}$$

with J positive definite, real and symmetric and $l \in \mathbf{C}$. Both J and l are arbitrary. This g is not necessarily centered even though G is. In the degenerate case it is not asserted that every maximizer must be centered when G is centered. The sharp constant is then $C_{p\to p'} = (C_p^B)^n$ with

$$C_p^B = \pi^{1/p'} p^{1/2p} (p')^{-1/2p'}\,. \tag{1}$$

[Note: The Fourier transform is an example of *both* Cases (B) and (C). While the proof of Theorem 3.3 (Case (B)) required 4.1(∗) and 4.2(1), the proof of Theorem 3.4 (Case (C)) did not. Therefore no circular reasoning is involved because 3.4 ⇒ 4.1(∗) for Case (C) ⇒ 4.2(1) ⇒ 3.3 ⇒ 4.1(∗) for Case (B).]

Another example is the (real convolution operator $G(x, y) = \exp\{-\lambda(x-y, x-y)\}$ which, using Theorem 4.1, turns out to be bounded if and only if $p \leqq q$ (see [BL] Section 4 for more details). There is a maximizing Gaussian function if and only if $p < q$ and it must have the form

$$g(x) = \exp\{-J(x, x) + (l, x)\} \tag{2}$$

with $J = \lambda\left(\frac{p'}{q'} - 1\right)$ and with $l \in \mathbf{R}^n$ arbitrary. Also

$$(C_{p\to q})^{2/n} = \left(\frac{\pi}{\lambda}\right)^{1/p' + 1/q} (p' - q')^{1/p - 1/q} p^{1/p} (q')^{1/p'} q^{-1/q} (p')^{-1/q'} . \quad (3)$$

When $p = q$ the limiting value $C_{p\to q} = \pi/\lambda$ is correct but, since $J = 0$ in this case, there is no Gaussian maximizer. Indeed, there is *no maximizer of any kind* in this case. To prove this, note that $G(x, y) = H(x - y)$ with $H(x) = \exp\{-\lambda(x, x)\}$ and $\int H(x - y) f(y) \mathrm{d}y = \int f(x - y) H(y) \mathrm{d}y$. Then, by Minkowski's inequality,

$$\left\{\int \left|\int f(x - y) H(y) \mathrm{d}y\right|^p \mathrm{d}x\right\}^{1/p} \leqq \int \left\{\int |f(x - y)|^p H(y)^p \mathrm{d}x\right\}^{1/p} \mathrm{d}y$$
$$= \|f\|_p \int H(y) \mathrm{d}y = \left(\frac{\pi}{\lambda}\right)^{n/2} \|f\|_p . \quad (4)$$

Since the condition in Lemma 3.1 for equality is clearly not satisfied, and since $(\pi/\lambda)^{n/2}$ has already been shown to be the sharp bound, a maximizer cannot exist.

A second example of a degenerate G that is bounded but does not have a maximizer is the following modification of the Fourier transform in $\mathbf{R}^1$ with $\lambda > 0$.

$$G_\lambda(x, y) = \exp\{-\lambda y^2 - 2\mathrm{i}xy\} . \quad (5)$$

It is easily verified for all p that $\mathcal{R}_{p\to q}(g)$ is unbounded on *complex* Gaussian functions when $q < 2$. Thus, it can be assumed that $q \geqq 2$, which places us in Case (B) of Sect. I. If $f_J(x) = \exp\{-Jx^2\}$ is an arbitrary Gaussian function, one finds that when $q \geqq 2$ the optimum choice is J real and

$$[\mathcal{R}_{p\to q}(f_J)]^2 = \pi^{1/q + 1/p'} p^{1/p} q^{-1/q} J^{1/p} (\lambda + J)^{-1/q'} . \quad (6)$$

By maximizing this with respect to J one finds that $C_{p\to q}$ is finite whenever $p \geqq q'$ and $C_{p\to q} = \infty$ when $p < q'$. If $p = q'$ there is no J that maximizes the right side of (6) (i.e., $J \to \infty$), although the right side is bounded. Indeed, *there is no maximizer* of any kind when $p = q'$. If there were a maximizer $f \in L^p(\mathbf{R}^1)$ then, by imitating the proof of Theorem 4.1, it is easily seen that $C_{p\to p'}(G_{\mu\lambda}) > C_{p\to p'}(G_\lambda)$ when $0 < \mu < 1$. This contradicts the conclusion of Theorem 4.1 which states that the supremum over J of the right side of (6) correctly gives $C_{p\to p'}(G_\lambda)$ for every λ, but this supremum is obviously independent of λ.

These examples motivate the following theorem.

4.3. Theorem (a condition for Gaussian maximizers). *Let G be a degenerate Gaussian kernel with the property that the $n \times n$ real, symmetric matrices A and B in* (1.1) *are both positive definite. If* $1 < p \leqq q < \infty$ *then $\mathcal{G}$ is bounded from $L^p(\mathbf{R}^n)$ to $L^q(\mathbf{R}^n)$. If, additionally, $p < q$ then $\mathcal{G}$ has a maximizer which is a Gaussian function.*

If G is also real then obviously A and B must be positive definite if $\mathcal{G}$ is bounded at all. In this real, degenerate case $\mathcal{G}$ is unbounded when $1 < q < p < \infty$ and $\mathcal{G}$ has no maximizer of any kind when $1 < p = q < \infty$.

Proof. It can be assumed that G is centered and, as in the proof of Theorem 3.3, we can use the fact that A and B are positive definite to change variables so that $G(x, y)$ is brought into the canonical form

$$G(x, y) = \exp\{-(x, x) - (y, y) - 2(x, Ey) - 2\mathrm{i}(x, Hy)\} , \quad (1)$$

where E and H are real matrices and E is also diagonal. In the real case $H = 0$. Since $M = \begin{pmatrix} I & E \\ E & I \end{pmatrix}$ must be positive semidefinite, the eigenvalues $e_1, \ldots, e_n$ of E must be in the interval $[-1, 1]$. Since G is degenerate at least one of the e_i's (say e_1) is $+1$ or -1 and, by changing y to $-y$ if necessary, we can assume that $e_1 = -1$. Thus, $G(x, y)$ contains the factor $\exp\{-(x_1 - y_1)^2\}$.

In the real case, $H = 0$, G in (1) is seen to be a tensor product of operators on $\mathbf{R}^1$, i.e., $G(x, y) = G_1(x_1, y_1) \ldots G_n(x_n, y_n)$. If $p > q$ the operator $\mathscr{G}_1$ corresponding to e_1 is unbounded, as shown in 4.2, so $\mathscr{G}$ is unbounded as well. In case $p \leqq q$ the Minkowski inequality argument in the first part of the proof of Theorem 3.2 (applied sequentially to $\mathscr{G}_1, \mathscr{G}_2, \ldots, \mathscr{G}_n$) shows that any maximizer, F, for $\mathscr{G}$ must be of the product form, i.e., $F(y_1, \ldots, y_n) = f_1(y_1) \ldots f_n(y_n)$ and each f_i must be a maximizer for the corresponding $\mathscr{G}_i$. When $p = q$, however, $\mathscr{G}_1$ does not have a maximizer as stated in 4.2 and therefore $\mathscr{G}$ has no maximizer. When $p < q$ we know from 4.2 that each $\mathscr{G}_i$ has a Gaussian maximizer g_i. Since $\prod_i g_i(x_i)$ is a Gaussian function on $\mathbf{R}^n$, the proof for the real case is complete.

For the general case with $p \leqq q$, let $G^0(x, y)$ be the real kernel given by (1) but with H set equal to zero and let $\mathscr{G}^0$ denote the corresponding operator. If $f \in L^p(\mathbf{R}^n) \cap \mathrm{L}^1(\mathbf{R}^n)$ then clearly $\mathscr{R}_{p\to q}(G, f) \leqq \mathscr{R}_{p\to q}(G^0, F)$, where $F \equiv |f|$. Since $\mathscr{G}^0$ is bounded when $p \leqq q$, then $\mathscr{G}$ is also bounded. Referring now to Theorem 4.1 let G_ε be the kernel defined in that proof, i.e., $G_\varepsilon(x, y) = G(x, y)h_\varepsilon(x)h_\varepsilon(y)$ with $h_\varepsilon(x) = \exp\{-\varepsilon(x, x)\}$, and let g_ε denote its unique Gaussian maximizer with $\|g_\varepsilon\|_p = 1$. Let $g_\varepsilon(x) = \mu_\varepsilon \exp\{-(x, J_\varepsilon x) - i(x, K_\varepsilon x)\}$ with J and K real, symmetric and with J_ε positive definite. Define $g_\varepsilon^0(x) = \mu_\varepsilon \exp\{-x, J_\varepsilon x)\}$. Let $\varepsilon \to 0$ through the sequence $\varepsilon = 1/j$ with $j = 1, 2, 3, \ldots$ There is a subsequence of the j's (which we continue to denote by j) such that the eigenvectors of J_ε and K_ε have limits as $j \to \infty$ (because the manifold $O(n)$ is compact). The corresponding eigenvalues of J_ε must be uniformly bounded away from 0 and ∞ since otherwise $\mathscr{R}_{p\to q}(G^0, g_\varepsilon^0)$ will converge to zero, as the following computation shows.

Apart from irrelevant constants, $\|g_\varepsilon^0\|_p = |J_\varepsilon|^{-1/p}$, where $|\cdot|$ denotes the determinant. Also, $\|\mathscr{G}^0(g_\varepsilon^0)\|_q = |J_\varepsilon + I|^{-1}|I - E(J_\varepsilon + I)^{-1}E|^{-1/q}$. Using the fact that $|I - M^T M| = |I - MM^T|$ for any real matrix, M, we have $|I - E(J_\varepsilon + I)^{-1}E| = |I - (J_\varepsilon + I)^{-1/2}E^2(J_\varepsilon + I)^{-1/2}| = |J_\varepsilon + I|^{-1}|J_\varepsilon + I - E^2| \geqq |J_\varepsilon + I|^{-1}|J_\varepsilon|$. Therefore

$$\mathscr{R}_{p\to q}(G^0, g_\varepsilon^0) \leqq |J_\varepsilon|^{1/p - 1/q}|J_\varepsilon + I|^{-1/q'} = \prod_{i=1}^{n} (J_\varepsilon^i)^{1/p - 1/q}(1 + J_\varepsilon^i)^{-1/q'},$$

where the J_ε^i's are the eigenvalues of J_ε. Since $p < q$ the function $t \mapsto t^{1/p - 1/q}(1 + t)^{-1/q'}$ is bounded and goes to zero as $t \to 0$ or $t \to \infty$.

The possibility that $\mathscr{R}_{p\to q}(G^0, g_\varepsilon^0) \to 0$ is not allowed by 4.1(*) and the fact that $\mathscr{R}_{p\to q}(G^0, g_\varepsilon^0) \geqq \mathscr{R}_{p\to q}(G, g_\varepsilon)$. Thus we can pass to a further subsequence such that J_ε has a positive definite limit J as $\varepsilon \to 0$. This implies that μ_ε also has a finite, nonzero limit μ.

The eigenvalues of K_ε must also stay bounded away from infinity for otherwise g_ε would tend weakly to zero in $L^p(\mathbf{R}^n)$ and then the function $\mathscr{G}(g_\varepsilon)$ would tend to zero pointwise. (This is so because the function $y \mapsto G(x, y)\exp\{-\frac{1}{2}(y, Jy)\}$ is in $L^{p'}(\mathbf{R}^n)$ for each x.) But $\mathscr{G}(g_\varepsilon)$ is bounded above pointwise by $\mathscr{G}^0(g_\varepsilon^0)$, and the pointwise convergence to zero would imply by dominated convergence that $\mathscr{G}(g_\varepsilon)$ converges to zero in $L^q(\mathbf{R}^n)$ norm. Thus, by passing to a further subsequence, J_ε and

K_ε have limits J and K. From this it follows that g_ε converges strongly in $L^p(\mathbf{R}^n)$ norm to $g(x) = \mu \exp\{-(x, Jx) - i(x, Kx)\}$.

The Gaussian function g is the desired maximizer for $\mathscr{G}$. First note that $h_\varepsilon g \to g$ in $L^p(\mathbf{R}^n)$ norm as $\varepsilon \to 0$. Also $g_\varepsilon \to g$, and thus we can write $h_\varepsilon g_\varepsilon = g + \Delta_\varepsilon$ with $\delta_\varepsilon \equiv \|\Delta_\varepsilon\|_p \to 0$ as $\varepsilon \to 0$. Then, since $\mathscr{G}$ is bounded,

$$C_{p\to q}(G) \geqq \mathscr{R}_{p\to q}(G, g) \geqq \mathscr{R}_{p\to q}(G, h_\varepsilon g_\varepsilon) - C_{p\to q}\delta_\varepsilon .$$

Taking the limit $\varepsilon \to 0$,

$$C_{p\to q}(G) \geqq \mathscr{R}_{p\to q}(G, g) \geqq \limsup_{\varepsilon\to 0} \mathscr{R}_{p\to q}(G, h_\varepsilon g_\varepsilon) .$$

But by Eq. (5) of the proof of Theorem 4.1, this latter limit equals $C_{p\to q}(G)$. □

4.4. Remarks and conjectures. Formula 4.1(∗) gives the sharp bound. The question that is incompletely resolved here is whether there is a Gaussian maximizer in the degenerate case or, indeed, any maximizer at all. In the cases of most interest (e.g., Nelson's kernel of Sect. I and the Fourier transform) the existence of a Gaussian maximizer can easily be verified by simple computation. The general case is algebraically complex, although Theorem 4.3 does give a criterion for a Gaussian maximizer and it completely settles the case of real Gaussian kernels. Indeed, as shown in 4.2, a maximizer need not exist even if $\mathscr{G}$ is bounded.

The examples given here lead to the following *conjectures.*

(1) If there is a maximizer for cases (A), (B) or (C) of Sect. I then there is a Gaussian maximizer.
(2) There is a maximizer in these cases if and only if the unique Gaussian maximizer g_ε for the mollified kernel $G_\varepsilon(x, y) = G(x, y)h_\varepsilon(x)h_\varepsilon(y)$ defined in the proof of Theorem 4.1 has a strong limit g in $L^p(\mathbf{R}^n)$ as $\varepsilon \to 0$.

Maximizers need not be unique, as shown in 4.2, but *if there is any Gaussian maximizer for $p < q$ then every maximizer is a Gaussian.* This is Theorem 4.5, and it completely settles the Fourier transform case, for example. (Note that when $p = q = 2$, every function in $L^2(\mathbf{R}^n)$ is a maximizer for the Fourier transform and thus there is at least one case in which there are maximizers that are not Gaussians.)

Theorem 4.5 also completely settles the real Case (A) because, by Theorem 4.3, no maximizer exists in this case when $p \geqq q$ and a Gaussian maximizer does exist when $p < q$.

4.5. Theorem (when $p < q$, a Gaussian maximizer implies all maximizers are Gaussians). *Let $1 < p < q < \infty$ and let G be a degenerate Gaussian kernel. Assume that $\mathscr{G}$ is a bounded operator from $L^p(\mathbf{R}^n)$ to $L^q(\mathbf{R}^n)$ and that g is a Gaussian function that is a maximizer for $\mathscr{G}$. If $f \in L^p(\mathbf{R}^n)$ is another maximizer for $\mathscr{G}$ then f is also a Gaussian (but f is not necessarily proportional to g and f is not necessarily centered even if G is).*

Proof. Step 1. According to Lemmas 2.2 and 2.3 it can be assumed without loss of generality that both G and g are centered. As in the proof of Theorem 3.4, we study the kernel $G^{(2)} = G \otimes G$. For $F \in L^p(\mathbf{R}^{2n}) \cap L^1(\mathbf{R}^{2n})$ the inequalities (1)–(5) there are valid and we conclude that $C_{p\to q}(G^{(2)}) = (C_{p\to q})^2$, where $C_{p\to q} \equiv C_{p\to q}(G)$.

Step 2. If $f \in L^p(\mathbf{R}^n)$ is a maximizer for $\mathscr{G}$ then, using $O(2)$ invariance again,

$$F(y_1, y_2) \equiv f\left(\frac{y_1 + y_2}{\sqrt{2}}\right) g\left(\frac{y_1 - y_2}{\sqrt{2}}\right) \tag{1}$$

is obviously a maximizer for $\mathscr{G}^{(2)}$ if f is also in $L^1(\mathbf{R}^n)$, in which case $F \in L^1(\mathbf{R}^{2n})$. If $f \notin L^1(\mathbf{R}^n)$ consider the mollified function $F_j(y_1, y_2) = F(y_1, y_2) \exp\{-(y_1 + y_2, y_1 + y_2)/j\}$ for $j = 1, 2, \ldots$, which is in $L^1(\mathbf{R}^{2n})$. Clearly $F_j \to F$ strongly in $L^p(\mathbf{R}^{2n})$ as $j \to \infty$. The function $\mathscr{G}^{(2)} F_j$ can be computed as a $\mathrm{d}y_1 \mathrm{d}y_2$ integral of $G^{(2)} F_j$ and the result (using the $O(2)$ invariance of $G^{(2)}$ and a change of variables) is

$$(\mathscr{G}^{(2)} F_j)(x_1, x_2) = (\mathscr{G} f_j)\left(\frac{x_1 + x_2}{\sqrt{2}}\right) \cdot (\mathscr{G} g)\left(\frac{x_1 - x_2}{\sqrt{2}}\right)$$

with $f_j(y) = f(y) \exp\{-2(y, y)/j\}$. Now the q^{th} power integral of $\mathscr{G}^{(2)} F_j$ can be computed by changing variables again and the result is $\| \mathscr{G}^{(2)} F_j \|_q = \| \mathscr{G} f_j \|_q \| \mathscr{G} g \|_q$. However $\| \mathscr{G} f_j \|_q \to \| \mathscr{G} f \|_q = C_{p \to q} \| f \|_p$ as $j \to \infty$ since $f_j \to f$ in $L^p(\mathbf{R}^n)$ norm, and we conclude that $\| \mathscr{G}^{(2)} F \|_q = \lim_{j \to \infty} \| \mathscr{G}^{(2)} F_j \|_q$ (by definition) $= (C_{p \to q})^2 \| F \|_p$, so that F is indeed a maximizer for $\mathscr{G}^{(2)}$.

Step 3. Since g is a Gaussian, it is obvious that the function $z \mapsto F(z, y)$ is in $L^p(\mathbf{R}^n) \cap L^1(\mathbf{R}^n)$ for each y and therefore that

$$K(x, y) \equiv \int G(x, z) F(z, y) \mathrm{d}z \tag{2}$$

is well defined for each x and y in $\mathbf{R}^n$. Since $\mathscr{G}$ is a bounded operator, the function $x \mapsto K(x, y)$ is in $L^q(\mathbf{R}^n)$ for each y. We now assert that the function $y \mapsto K(x, y)$ is in $L^p(\mathbf{R}^n)$ for almost every $x \in \mathbf{R}^n$ and that this function satisfies

$$\{ \int [\int |K(x, y)|^q \mathrm{d}x]^{p/q} \mathrm{d}y \}^{q/p} = \int \{ \int |K(x, y)|^p \mathrm{d}y \}^{q/p} \mathrm{d}x \,, \tag{3}$$

with the understanding that both sides of (3) are finite. Formally, this assertion is a consequence of inequality (3) → (4) in the proof of Theorem 3.4 and the fact that all the inequalities (1)–(5) must be equalities since $F(y_1, y_2)$ is a maximizer for $\mathscr{G}^{(2)}$. If $F \in L^1(\mathbf{R}^{2n})$ this would be correct, but if $F \notin L^1(\mathbf{R}^{2n})$ a proof is needed.

Set $F_j(y_1, y_2) = F(y_1, y_2) \exp\{-(y_2, y_2)/j\}$ for $j = 1, 2, \ldots$. Clearly $F_j \in L^1(\mathbf{R}^{2n})$ and $F_j \to F$ strongly in $L^p(\mathbf{R}^{2n})$ as $j \to \infty$. (Note that this F_j is not the same one as in step 2.) Let $K_j(x, y)$ be as in (2) with F replaced by F_j, so that $K_j(x, y) = K(x, y) \exp\{-(y, y)/j\}$. The inequalities (1)–(5) in the proof of Theorem 3.4 are then valid with F replaced by F_j. As $j \to \infty$ the left side of these inequalities, namely $\| \mathscr{G}^{(2)} F_j \|_q^q$, converges to $\| \mathscr{G}^{(2)} F \|_q^q = (C_{p \to q})^{2q} \| F \|_p^q \equiv (C_{p \to q})^q Z$ since F is a maximizer. Likewise, the right side, namely $(C_{p \to q})^{2q} \| F_j \|_p^q$ also converges to $(C_{p \to q})^q Z$ since $F_j \to F$. Therefore the numbers

$$B_j \equiv \{ \int [\int |K_j(x, y)|^q \mathrm{d}x]^{p/q} \mathrm{d}y \}^{q/p} - \int \{ \int |K_j(x, y)|^p \mathrm{d}y \}^{q/p} \mathrm{d}x \tag{4}$$

(which are nonnegative by Minkowski's inequality) must converge to zero as $j \to \infty$. Moreover, each term in B_j is bounded by $(C_{p \to q})^q \| F_j \|_p^q < Z$, and each term converges to Z as $j \to \infty$ (because of inequalities (1)–(5)). The first term in B_j is

$$A_j \equiv \left\{ \int [\int |K(x, y)|^q \mathrm{d}x]^{p/q} \exp\left\{ -\frac{p}{j}(y, y) \right\} \mathrm{d}y \right\}^{q/p}$$

and, by the monotone convergence theorem, A_j converges to $A \equiv$ (the left side of (3)). Therefore $A = Z$. The second term in B_j is

$$D_j \equiv \int \left\{ \int |K(x, y)|^p \exp\left\{ -\frac{p}{j}(y, y) \right\} \mathrm{d}y \right\}^{q/p} \mathrm{d}x \,.$$

The inner integral (call it $E_j(x)$) converges (by monotone convergence) to $E(x) \equiv \int |K(x, y)|^p \mathrm{d}y$. The function E is measurable since it is the monotone limit of measurable functions E_j. Then $\int \{E_j\}^{q/p}$ converges to $\int \{E\}^{q/p}$ by monotone convergence, so D_j converges to the right side of (3). But, as stated above, D_j also converges to Z, so the two sides of (3) are equal and $E(x)$ is finite for almost every x, as asserted.

Step 4. Since $q > p$, the strong form of Minkowski's inequality and the equality in (3) implies the existence of measurable functions α and $\beta: \mathbf{R}^n \to [0, \infty)$ such that

$$|K(x, y)| = \alpha(x)\beta(y) \tag{5}$$

for almost every x and y in $\mathbf{R}^n$. Writing $G(x, y) = \exp\{-(x, Ax) - (y, By) - 2(x, Dy)\}$ as usual (with A and B real, symmetric, positive definite), and writing $g(y) = \exp\{-(y, Jy)\}$ (with J symmetric and $\mathrm{Re}(J)$ positive definite) a simple computation gives

$$\begin{aligned} K(x, y) = \exp\{ &-(x, Ax) + (D^T x, (B + J)^{-1} D^T x) \\ &- (y, Jy)\} Q((B + J)y - D^T x) \end{aligned} \tag{6}$$

with $Q: \mathbf{C}^n \to \mathbf{C}$ given by

$$Q(w) = \exp\{-(w, (B + J)^{-1} w)\} \int f\left(\frac{z}{\sqrt{2}}\right) \exp\{-(z, (B + \tfrac{1}{2} J)z) + 2(z, w)\} \, \mathrm{d}z \,. \tag{7}$$

Evidently Q is an entire analytic function of order at most 2. Define the function $M: \mathbf{R}^{2n} \to \mathbf{C}$ by $M(x, y) = Q((B + J)y - D^T x)$. Plainly, since Q is entire M has an extension to an entire analytic function from $\mathbf{C}^{2n}$ to $\mathbf{C}$; call this extension N. The function $N^*: \mathbf{C}^{2n} \to \mathbf{C}$ defined by $N^*(x, y) = \overline{N(\bar{x}, \bar{y})}$ for x and $y \in \mathbf{C}^n$ is also entire analytic, and thus $P \equiv NN^*$ is entire analytic as well. It is also true that $P(x, y) = |M(x, y)|^2$ when x and y are in $\mathbf{R}^n$. From (5) and (6),

$$P(x, y) = \gamma(x)\delta(y) \tag{8}$$

for almost every x and y in $\mathbf{R}^n$, and where γ and $\delta: \mathbf{R}^n \to [0, \infty)$ are the measurable functions given by $\gamma(x) = \alpha(x)^2 \exp\{2(x, Ax) - 2\mathrm{Re}((D^T x, (B + J)^{-1} D^T x))\}$ and $\delta(y) = \beta(y)^2 \exp\{2\mathrm{Re}((y, Jy))\}$. If y_0 is a value of y such that $\delta(y_0) \neq 0$ and such that (8) holds for almost every x, we see by substituting this y_0 in (8) that γ has an extension to an entire analytic function. Likewise, δ has an extension. Thus (8) holds for *every* x and y in $\mathbf{C}^n$ (because if two entire functions agree almost everywhere on $\mathbf{R}^n \times \mathbf{R}^n$ then they agree on all of $\mathbf{C}^n \times \mathbf{C}^n$).

Now suppose that $\gamma(x_0) = 0$ for some x_0 in $\mathbf{C}^n$. Then, by (8), $P(x_0, y) = 0$ for every $y \in \mathbf{C}^n$, which implies that for each y either (*i*) $N(x_0, y) = 0$ or (*ii*) $N^*(x_0, y) = 0$. This, in turn, means that for each $y \in \mathbf{C}^n$ either (i) $N(x_0, y) \equiv Q((B + J)y - D^T x_0) = 0$ or (ii) $N(\bar{x}_0, \bar{y}) \equiv Q((B + J)\bar{y} - D^T \bar{x}_0) = 0$. Necessarily, either case (i) holds for all y in some set $S \subset \mathbf{C}^n$ of positive $2n$-dimensional Lebesgue measure $\mathscr{L}^{2n}$ or case (ii) holds in some set S of positive $\mathscr{L}^{2n}$ measure. As y ranges over S both $(B + J)y$ and $(B + J)\bar{y}$ range over sets of positive $\mathscr{L}^{2n}$ measure (because $\mathrm{Re}(B + J)$ is positive definite and therefore $Rank(B + J) = n$). An analytic function that vanishes on a set of positive $\mathscr{L}^{2n}$ measure vanishes identically, and thus Q would vanish identically if $\gamma(x_0) = 0$. This contradicts the fact that $K(x, y)$ is not identically zero. Thus, the assumption that $\gamma(x_0) = 0$ is not possible, and it will be assumed henceforth that $\gamma(x) \neq 0$ for all $x \in \mathbf{C}^n$.

Define the set $A = \{y \in \mathbf{R}^n : \delta(y) \neq 0\} \subset \mathbf{R}^n$. This set A has positive n-dimensional Lebesgue measure $\mathscr{L}^n$, for otherwise $K(x, y) = 0$, $\mathscr{L}^{2n}$ almost everywhere. (In fact $\mathscr{L}^n(\mathbf{R}^n \sim A) = 0$ because δ is analytic and δ does not vanish identically, but this fact is not needed.) For $y \in A$, the function $Z_y : \mathbf{C}^n \to \mathbf{C}$ defined by $Z_y(x) = K(x, y)$ is entire analytic of order at most 2 and never zero (because $\gamma(x)$ is never zero). Then Z_y has the form

$$Z_y(x) = K(x, y) = \exp\{-(x, T_y x) - (R_y, x) + \mu_y\}, \tag{9}$$

where T_y is a complex, symmetric matrix, $R_y \in \mathbf{C}^n$ and $\mu_y \in \mathbf{C}$ (all of which depend on y). I thank Eric Carlen for the simple proof of this fact, which is that Z_y, being zero free, has an entire analytic logarithm, i.e., $Z_y = \exp\{H_y\}$. Then, since Z_y has order at most 2, $|H_y(x)|$ is bounded above by (const.) $|x|^2$. By a well known argument using Cauchy's integral formula, H_y must be a polynomial whose order is at most 2, i.e., Z_y has the form stated in (9).

Step 5. As noted in step 2, the function $x \mapsto K(x, y)$ is in $L^q(\mathbf{R}^n)$ for almost every $y \in \mathbf{R}^n$. By (4) $\to$ (5) of Theorem 3.4, the function $z \mapsto F(z, y)$ (which is in $L^p(\mathbf{R}^n)$ for almost every y) must be a maximizer of $\mathscr{R}_{p\to q}$ for almost every y. (Note that $z \mapsto F(z, y)$ cannot be the zero function for any y since g never vanishes.) Thus there is at least one point $y_0 \in \mathbf{R}^n$ such that $\delta(y_0) \neq 0$ and (9) holds and such that $z \mapsto F(z, y_0)$ is a maximizer in $L^p(\mathbf{R}^n)$. Fix this y_0 henceforth and denote the matrix in (9) simply by T. There is then a function $h \in L^{q'}(\mathbf{R}^n)$ with $\|h\|_{q'} = 1$ such that

$$\int h(x)K(x, y_0)dx = \|K(\cdot, y_0)\|_q = C_{p\to q}\|F(\cdot, y_0)\|_p. \tag{10}$$

Since $K(\cdot, y_0) \in L^q(\mathbf{R}^n)$, the matrix T must satisfy $\mathrm{Re}(T)$ is positive definite and therefore $K(\cdot, y_0)$ is a Gaussian. The optimum h satisfies $h(x) = (\text{const.})\,|K(x, y_0)|^q/K(x, y_0)$ for $x \in \mathbf{R}^n$ and therefore h is also a Gaussian (and hence $h \in L^1(\mathbf{R}^n)$). As remarked in step 3, $F(\cdot, y_0)$ is in $L^1(\mathbf{R}^n)$. Therefore the function $(x, y) \mapsto h(x)G(x, y)F(y, y_0)$ is in $L^1(\mathbf{R}^{2n})$ and Fubini's theorem can be applied to (10). Thus,

$$\int h(x)K(x, y_0)dx = \int \left\{\int h(x)G(x, z)dx\right\} F(z, y_0)dz. \tag{11}$$

Since h is a Gaussian, the inner integral in (11) (call it $k(z)$) is also a Gaussian. Since $F(\cdot, y_0)$ is a maximizer, $F(z, y_0) = (\text{const.})|k(z)|^{p'}/k(z) \equiv r(z)$ for almost every $z \in \mathbf{R}^n$. Clearly r is a Gaussian and, by (1)

$$f\left(\frac{z + y_0}{\sqrt{2}}\right) g\left(\frac{z - y_0}{\sqrt{2}}\right) = r(z) \tag{12}$$

for almost every $z \in \mathbf{R}^n$. Setting $z = w - y_0$, (12) yields $f(w/\sqrt{2}) = r(w - y_0)/g((w - 2y_0)/\sqrt{2})$, which is a Gaussian (in w) as asserted in the theorem. □

V. Gaussian kernels from $L^p(\mathbf{R}^n)$ to $L^q(\mathbf{R}^m)$

This section consists essentially of a simple remark, but it can be a useful one in applications, e.g., in [L1]. Let G be a Gaussian kernel on $\mathbf{R}^m \times \mathbf{R}^n$ with $m \neq n$, i.e., $G(x, y)$ is given by (1.1) with A $m \times m$ symmetric, B $n \times n$ symmetric, D $m \times n$ and

$L \in C^{m+n}$, and with M in (1.3) a positive semidefinite $(m+n) \times (m+n)$ matrix. Evidently Lemmas 2.1, 2.2 and 2.3 continue to hold in this case, and it can be assumed without loss of generality that A and B are real and $L = 0$. The linear operator $\mathscr{G}$ from $L^p(\mathbf{R}^n)$ to $L^q(\mathbf{R}^m)$ and the norm $C_{p\to q}(G)$ are defined, *mutatis mutandis*, as in Sect. I. The remark is the following.

5.1. Theorem (extension to $m \neq n$). *Let G be a Gaussian kernel on $\mathbf{R}^m \times \mathbf{R}^n$ as defined above. Then all the preceding theorems and lemmas in this paper holds, mutatis mutandis, in this more general case.*

Proof. Suppose $m < n$ and extend G to a Gaussian kernel, $\tilde{G}$, on $\mathbf{R}^n \times \mathbf{R}^n$ by

$$\tilde{G}(x, y) = h(x_1)G(x_2, y)$$

where $x \in \mathbf{R}^n$ is written as (x_1, x_2) with $x_1 \in \mathbf{R}^{n-m}$ and $x_2 \in \mathbf{R}^m$, and where $h(x_1) \equiv \exp\{-(x_1, x_1)\}$. Let $\tilde{\mathscr{G}}$ be the corresponding operator from $L^p(\mathbf{R}^n)$ to $L^q(\mathbf{R}^n)$. Note that $\tilde{G}$ has the same properties as G, i.e., the degeneracy or nondegeneracy of $\tilde{G}$ is the same as that of G; $\tilde{G}$ is in Case (A), (B) or (C) if G is; the $n \times n$ matrix $\begin{pmatrix} I & 0 \\ 0 & A \end{pmatrix}$ is positive definite if and only if A is. Also, $\tilde{\mathscr{G}}$ is unbounded if $\mathscr{G}$ is, and it will be assumed henceforth that $\mathscr{G}$ is bounded.

If $f \in L^p(\mathbf{R}^n)$ then evidently, as functions in $L^q(\mathbf{R}^n)$, $(\tilde{\mathscr{G}}f)(x) = h(x_1)(\mathscr{G}f)(x_2)$ and thus $\|\tilde{\mathscr{G}}f\|_{L^p(\mathbf{R}^n)} = \|h\|_{L^q(\mathbf{R}^{n-m})}\|\mathscr{G}f\|_{L^q(\mathbf{R}^m)}$. This proves that $C_{p\to q}(\tilde{G}) = C_{p\to q}(G)\|h\|_{L^q(\mathbf{R}^{n-m})}$ and that f is a maximizer for $\tilde{\mathscr{G}}$ if and only if f is a maximizer for $\mathscr{G}$. This concludes the $m < n$ case.

If $m > n$ duality can be used: $C_{p\to q}(G) = C_{q'\to p'}(G^T)$ where $G^T(x, y) = G(y, x)$. This changes the $m > n$ case into the $m < n$ case and, since all the theorems in this paper are "duality invariant", the $m > n$ case is proved. □

Remark. Clearly the proof of Theorem 5.1 is such that if other cases with $m = n$ are settled in the future then Theorem 5.1 for $m \neq n$ holds for those cases as well.

VI. Multilinear forms in the real case and Young's inequality

After Sects. I to V were completed, Eric Carlen suggested that the same methods should yield similar results for real multilinear forms. Indeed this is so and the proof is outlined here (the omitted details are merely a repetition of those given before). Some remarks about the complex case will also be made here. Finally, Theorem 6.2 contains an application of the result in Sect. 6.1 for real multilinear forms: The truly multidimensional generalization of Young's inequality, which was surmised in [BL, p. 162], will be proved.

6.1. **Multilinear forms.** For $i = 1, 2, \ldots, K$ let n_i be a positive integer and let x_i denote a point in $\mathbf{R}^{n_i}$. The point $X = (x_1, \ldots, x_K)$ denotes a point in $\mathbf{R}^N$ with $N = \sum_{i=1}^K n_i$. Let $G(X)$ be a "Gaussian kernel", i.e.,

$$G(X) = \exp\left\{-\sum_{i=1}^{K}\sum_{j=1}^{K}(x_i, A_{ij}x_j) + 2(L, X)\right\},$$

where A_{ij} is a $n_i \times n_j$ matrix with $A_{ij} = A_{ji}^T$, and where $L \in \mathbf{C}^N$. The $N \times N$ symmetric matrix A is the matrix whose blocks are the A_{ij}'s and G is said to be **nondegenerate** if $M \equiv \mathrm{Re}(A)$ is positive definite. Otherwise $M \geqq 0$ and G is **degenerate**.

Let $P = (p_1, \ldots, p_K)$ satisfy $1 < p_i < \infty$ for each i. The **multilinear form** is

$$\mathscr{G}(f_1, \ldots, f_K) = \int G(x_1, \ldots, x_K) \prod_{i=1}^{K} f_i(x_i) \mathrm{d}x_1 \ldots \mathrm{d}x_K , \tag{1}$$

where the integration is over $\mathbf{R}^{n_1} \times \mathbf{R}^{n_2} \times \ldots \times \mathbf{R}^{n_K}$ and each $f_i \in L^{p_i}(\mathbf{R}^{n_i})$. The problem is to evaluate

$$C_P = \sup \mathscr{G}(f_1, \ldots, f_K) \tag{2}$$

where the supremum is over f_i's with $\| f_i \|_{p_i} = 1$. As before, if G is degenerate we have to take $f_i \in L^{p_i}(\mathbf{R}^{n_i}) \cap L^1(\mathbf{R}^{n_i})$ and then take limits.

The cases treated in Sects. I to V correspond to $K = 2$ with $p_1 = p$ and $p_2 = q'$. The case $K = 1$ is trivial—by Hölder's inequality.

Lemma 2.1 is easily generalized to the complex, nondegenerate multilinear case; the details are left to the reader. The conclusion of Lemma 2.1 holds for each f_i in a maximizing set $(f_1, \ldots, f_K)$. The conclusions (a), (b) and (c) follow by fixing all the f_j's with $j \neq i$ and then investigating the dependence of $\mathscr{G}(f_1, \ldots, f_K)$ on f_i.

Lemma 2.2 obviously carries through as well; that is A_{ii} can be assumed to be real and G can be assumed to be centered, i.e., $L = 0$. Likewise Lemma 2.3 carries through: When G is centered (i.e., $L = 0$) and when the supremum in (2) is restricted to Gaussian functions f_i, then each f_i can be taken to be centered and, in the nondegenerate case, each f_i *must be* centered.

Let us now turn to the real case, i.e., each A_{ij} is real and $L = 0$. *Theorem* 3.2 *for the nondegenerate case carries through for every choice* of P. The maximizing K-tuple $(f_1, \ldots, f_K)$ is unique (up to multiplicative constants) and each $f_i(x) = \exp\{-(x, J_i x)\}$ with J_i being real and positive definite. To prove that f_1, say, has this property we write (with $q = p_1'$)

$$C_P = \sup \| \tilde{\mathscr{G}}(f_2, \ldots, f_K) \|_q$$

where the supremum is on $f_2, \ldots, f_K$ with $\| f_j \|_{p_j} = 1$ and where

$$\tilde{\mathscr{G}}(f_2, \ldots, f_K)(x_1) = \int G(x_1, \ldots, x_K) \prod_{j=2}^{K} f_j(x_j) \mathrm{d}x_2 \ldots \mathrm{d}x_K .$$

As before, we replace $\tilde{\mathscr{G}}$ by $\tilde{\mathscr{G}}^{(2)} = \tilde{\mathscr{G}} \otimes \tilde{\mathscr{G}}$ and $f_2(x_2), \ldots, f_K(x_K)$ by $F_2(x_2, y_2), \ldots, F_K(x_K, y_K)$ with $F_j \in L^{p_j}(\mathbf{R}^{2n_j})$. To imitate the inequalities (1)–(5) in Theorem 3.2, define

$$K(x_1, y_2, \ldots, y_K) = \int G(x_1, x_2, \ldots, x_K) \prod_{j=2}^{K} F_j(x_j, y_j) \mathrm{d}x_2 \ldots \mathrm{d}x_K .$$

Then, proceeding as in (1)–(5) (and with the F_j nonnegative for the same reason as before)

$$\begin{aligned}
\| \tilde{\mathscr{G}}^{(2)}(F_2, \ldots, F_K) \|_q^q &= \int [\int G(y_1, \ldots, y_K) K(x_1, y_2, \ldots, y_K) \mathrm{d}y_2 \ldots \mathrm{d}y_K]^q \mathrm{d}x_1 \mathrm{d}y_1 \\
&\leqq \int \{ \int G(y_1, \ldots, y_K) [\int K(x_1, y_2, \ldots, y_K)^q \mathrm{d}x_1]^{1/q} \\
&\quad \times \mathrm{d}y_2 \ldots \mathrm{d}y_K \}^q \mathrm{d}y_1 \\
&\leqq (C_P)^q \int \{ \int G(y_1, \ldots, y_K) \prod_{j=2}^{K} h_j(y_j) \mathrm{d}y_2 \ldots \mathrm{d}y_K \}^q \mathrm{d}y_1 \\
&\quad (\text{with } h_j(y) = [\int F_j(x, y)^{p_j} \mathrm{d}x]^{1/p_j}) \\
&\leqq (C_P)^{2q} \prod_{j=2}^{K} \| F_j \|_{p_j}^q .
\end{aligned}$$

As before, Minkowski's inequality implies that

$$K(x_1, y_2, \ldots, y_K) = A(x_1)E(y_2, \ldots, y_K),$$

which then implies that

$$\tilde{\mathscr{G}}^{(2)}(F_2, \ldots, F_K)(x_1, y_1) = A(x_1)Z(y_1).$$

However, $\tilde{\mathscr{G}}^{(2)}(F_2, \ldots, F_K) = F_1^{p_1-1}$, and hence F_1 is a product function. The rest of the proof is identical to the proof of Theorem 3.2.

By taking limits, the analogue of Theorem (4.1) hold in the degenerate case whenever it is known that the nondegenerate case has a Gaussian maximizer K-tuple. *In particular, Theorem* 4.1 *holds in the real case.* The analogue of 4.1($*$) is that C_p is given by (2) with the supremum restricted to centered Gaussian functions. Likewise, Theorem 4.3 extends to the multilinear case under the same assumption about the nondegenerate case; the analogous hypothesis is that each A_{ii} is positive definite. *In particular Theorem* 4.3 *holds in the real case.*

These results can be used to derive the sharp constants in the fully multidimensional generalized Young's inequality. Recall that Young's original inequality states that if $f \in L^p(\mathbf{R}^n)$ and $g \in L^q(\mathbf{R}^n)$ then $f * g \in L^r(\mathbf{R}^n)$ with $1/p + 1/q = 1 + 1/r$; here $*$ denotes convolution. The sharp constant in this inequality was derived simultaneously by Beckner [B1, B2] and by Brascamp and Lieb [BL]. Another way to state the inequality is that

$$\int_{\mathbf{R}^n}\int_{\mathbf{R}^n} h(x)f(x-y)g(y)\,\mathrm{d}x\,\mathrm{d}y \leqq C\|h\|_{r'}\|f\|_p\|g\|_q \tag{3}$$

with $1/p + 1/q + 1/r' = 2$. The Beckner, Brascamp-Lieb result is that C can be determined by restricting f, g and h to be Gaussian functions. (These, in fact, are the only maximizers, as shown in [BL].)

Young's inequality (3) was generalized in several ways in [BL]. The first way is to allow an arbitrary number of functions $f_1, \ldots, f_K$ instead of merely three as in (3). These are functions from $\mathbf{R}^n$ to $\mathbf{C}$ and $f_j \in L^{p_j}(\mathbf{R}^n)$. This is Theorem 7 of [BL]. The integration is then over $(\mathbf{R}^m)^n$ and the arguments of the f_j's are taken to be $((a_1^j, x_1), \ldots, (a_n^j, x_n)) \in \mathbf{R}^n$, where $a_l^j \in \mathbf{R}^m$ are specified vectors and $x_i \in \mathbf{R}^m$. Unfortunately, this is not a fully mn-dimensional generalization of the $n = 1$ result because $\mathbf{R}^{mn}$ is split unnaturally into $(\mathbf{R}^m)^n$. Following Theorem 7 in [BL] we asked whether the full generalization is possible and Theorem 6.2 below gives it.

A second generalization was the incorporation of a fixed Gaussian function in the integral, as in Theorem 6 of [BL]. Again, the Gaussian in [BL] was completely general when $n = 1$, but not otherwise. In Theorem 6.2 it is completely general.

6.2. Theorem (fully generalized Young's inequality). *Fix $K > 1$, $n_1, \ldots, n_K$ and $p_1, \ldots, p_K > 1$ as before. Let $M \geqq 1$ be an integer and let B_i (for $i = 1, \ldots, K$) be a linear mapping from $\mathbf{R}^M$ to $\mathbf{R}^{n_i}$. For nonnegative functions $f_1, \ldots, f_K$, with $f_i \in L^{p_i}(\mathbf{R}^{n_i})$ consider*

$$I(f_1, \ldots, f_K) = \int_{\mathbf{R}^M} \prod_{i=1}^{K} f_i(B_i x)\,\mathrm{d}x. \tag{1}$$

More generally, let $g: \mathbf{R}^M \to \mathbf{R}^+$, $g(x) = \exp\{-(x, Jx)\}$, be a fixed, centered, real Gaussian function and consider

$$I_g(f_1, \ldots, f_K) = \int_{\mathbf{R}^M} \prod_{i=1}^{K} f_i(B_i x)g(x)\,\mathrm{d}x. \tag{2}$$

Let

$$C_g \equiv \sup_{f_1, \ldots, f_K} \{I_g(f_1, \ldots, f_K) : \|f_i\|_{p_i} = 1\} \tag{3}$$

and similarly for C *(with* I_g *replaced by* I*). Then*

$$C_g = \sup\{I_g(f_1, \ldots, f_K) : f_1, \ldots, f_K \text{ are real, centered Gaussian functions with } \|f_i\|_{p_i} = 1\}\,, \tag{4}$$

and similarly for C.

Proof. Suppose the theorem is false and that the right side of (4) (call it D_g) is strictly smaller than C_g. (Alternatively, $D < C$.) Then there are nonnegative summable functions that are *not all* Gaussians, $f_1, \ldots, f_K$ of unit L^{p_i} norm, such that $I_g(f_1, \ldots, f_K) > D_g$ (or $I(f_1, \ldots, f_K) > D$).

Consider the functions $f_i^{(l)} : \mathbf{R}^{n_i} \to \mathbf{R}^+$ given by $f_i^{(l)} \equiv f_i * g_i^{(l)}$ for l a positive integer, where $g_i^{(l)}(x) \equiv (l/\pi)^{n_i/2} \exp\{-l(x, x)\}$ is an $L^1(\mathbf{R}^{n_i})$ normalized Gaussian function. We note that $\|f_i^{(l)}\|_{p_i} \leqq 1$ and that $f_i^{(l)} \to f_i$ in $L^{p_j}(\mathbf{R}^{n_j})$ as $l \to \infty$. By passing to a subsequence (henceforth still denoted by l) we can assume that $f_i^{(l)}(x) \to f_i(x)$ for almost every x in $\mathbf{R}^{n_i}$.

Evidently we can assume that $M \geqq \max\{n_1, \ldots, n_K\}$ and that the rank of B_i is n_i for all i. Otherwise, I or I_g involves knowledge of some f_i only on a hyperplane in $\mathbf{R}^{n_i}$ and this means that I or I_g can be made arbitrarily large (with all f_i's being Gaussian functions) while preserving $\|f_i\|_{p_i} = 1$; the theorem would then be true in this case because both sides of (4) would be infinite. Similarly, the mapping $W \equiv J + \sum_{i=1}^K B_j^* B_j$ (with $*$ denoting adjoint) from $\mathbf{R}^M$ to $\mathbf{R}^M$ is positive definite; otherwise I_g can again be made arbitrarily large with Gaussian f_i's. A similar condition holds for I with $J = 0$. Since B_i is linear and has full rank n_i, the almost everywhere pointwise convergence of $f_i^{(l)}$ to f_i in $\mathbf{R}^{n_i}$ implies that $f_i^{(l)}(B_i x) \to f_i(B_i x)$ for almost every x in $\mathbf{R}^M$.

By Fatou's lemma

$$C_g' \equiv \liminf_{l \to \infty} I_g(f_i^{(l)}, \ldots, f_K^{(l)}) \geqq I_g(f_1, \ldots, f_K) > D_g \tag{5}$$

and similarly for C' (with I in place of I_g). By Fubini's theorem, however,

$$I_g(f_1^{(l)}, \ldots, f_K^{(l)}) = \int_{\mathbf{R}^N} G_g^{(l)}(y_1, \ldots, y_K) \prod_{i=1}^K f_i(y_i) \mathrm{d}y_1 \ldots \mathrm{d}y_K\,. \tag{6}$$

Here $N = \sum_{i=1}^K n_i$ as in Sect. 6.1, $y_i \in \mathbf{R}^{n_i}$, and $G_g^{(l)}$ is the centered Gaussian kernel

$$G_g^{(l)}(y_1, \ldots, y_K) = \int_{\mathbf{R}^M} \prod_{i=1}^K g_i^{(l)}(B_i x - y_i) g(x) \mathrm{d}x\,. \tag{7}$$

Similarly, (6) and (7) hold for I in place of I_g by deleting the g. (Note: Because W is positive definite, the integral in (7) is always finite.)

The number C_g' defined in (5) is either finite or infinite. In either case, there is some finite integer k such that $C_g'' \equiv I_g(f_1^{(k)}, \ldots, f_K^{(k)}) > D_g$. However, by (6) we see that C_g'' is a multilinear form as in 6.1 (1). Such a form has the property, as we have seen in Section 6.1, that its supremum over f_i's with $\|f_i\|_{p_i} = 1$ is equal to its supremum over real, centered Gaussian functions. But if we set all the f_i's equal to Gaussian functions we have that $f_i^{(k)}$'s are also Gaussian functions and $\|f_i^{(k)}\|_{p_i} \leqq 1$.

This means that $C_g'' \leqq D_g$, and this is a contradiction. The same proof holds for I in place of I_g. □

References

[BA] Babenko, K.I.: An inequality in the theory of Fourier integrals. Izv. Akad. Nauk SSR Ser. Mat. **25**, 531–542; (1961) English transl. Am. Math. Soc. Transl. (2) **44**, 115–128 (1965)

[B1] Beckner, W.: Inequalities in Fourier analysis. Ann. Math. **102**, 159–182 (1975)

[B2] Beckner, W.: Inequalities in Fourier analysis on $\mathbf{R}^n$. Proc. Natl. Acad. Sci. USA **72**, 638–641 (1975)

[BL] Brascamp, H.J., Lieb, E.H.: Best constants in Young's inequality, its converse, and its generalization to more than three functions. Adv. Math. **20**, 151–173 (1976)

[CA] Carlen, E.: Superadditivity of Fisher's information and logarithmic Sobolev inequalities. J. Funct. Anal. (in press)

[CL] Carlen, E., Loss, M.: Extremals of functionals with competing symmetries. J. Funct. Anal. **88**, 437–456 (1990)

[C] Coifman, R., Cwikel, M., Rochberg, R., Sagher, Y., Weiss, G.: Complex interpolation for families of Banach spaces. Am. Math. Soc. Proc. Symp. Pure Math. **35**, 269–282 (1979)

[DGS] Davies, E.B., Gross, L., Simon, B.: Hypercontractivity: a bibliographic review. Proceedings of the Hoegh-Krohn memorial conference. Albeverio, S. (ed.) Cambridge: Cambridge University Press, 1990

[E] Epperson, Jr., J.B.: The hypercontractive approach to exactly bounding an operator with complex Gaussian kernel. J. Funct. Anal. **87**, 1–30 (1989)

[GL] Glimm, J.: Boson fields with nonlinear self-interaction in two dimensions. Commun. Math. Phys. **8**, 12–25 (1968)

[G] Gross, L.: Logarithmic Sobolev inequalities. Am. J. Math. **97**, 1061–1083 (1975)

[HLP] Hardy, G.H., Littlewood, J.E., Pólya, G.: Inequalities. See Theorem 202 on p. 148. Cambridge: Cambridge University Press 1959

[J] Janson, S.: On hypercontractivity for multipliers on orthogonal polynomials. Ark. Mat. **21**, 97–110 (1983)

[L1] Lieb, E.H.: Proof of an entropy conjecture of Wehrl. Commun. Math. Phys. **62**, 35–41 (1978)

[L2] Lieb, E.H.: Integral bounds for radar ambiguity functions and Wigner distributions. J. Math. Phys. **31**, 594–599 (1990)

[N1] Nelson, E.: A quartic interaction in two dimensions in: Mathematical theory of elementary particles. Goodman, R., Segal, I. (eds.), pp. 69–73. Cambridge: M.I.T. Press 1966

[N2] Nelson, E.: The free Markov field. J. Funct. Anal. **12**, 211–227 (1973)

[NE] Neveu, J.: Sur l'esperance conditionelle par rapport à un mouvement Brownien. Ann. Inst. H. Poincaré Sect. B. (N.S.) **12**, 105–109 (1976)

[S1] Simon, B.: A remark on Nelson's best hypercontractive estimates. Proc. Am. Math. Soc. **55**, 376–378 (1976)

[S] Segal, I.: Construction of non-linear local quantum processes: I. Ann. Math. **92**, 462–481 (1970)

[T] Titchmarsh, E.C.: A contribution to the theory of Fourier transforms. Proc. London Math. Soc. Ser. 2, **23**, 279–289 (1924)

[W] Weissler, F.B.: Two-point inequalities, the Hermite semigroup, and the Gauss-Weierstrass semigroup. J. Funct. Anal. **32**, 102–121 (1979)