Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

HHS Public Access

Author manuscript

Stat Surv. Author manuscript; available in PMC 2016 April 27.

Published in final edited form as:

Stat Surv. 2014 ; 8: 45–114. doi:10.1214/14-SS107.

Log-Concavity and Strong Log-Concavity: a review

Adrien Saumard[*] and

Departamento de Estadística, CIMFAV, Universidad de Valparaíso, Chile

Jon A. Wellner[†]

Department of Statistics, Box 354322, University of Washington, Seattle, WA 98195-4322

Adrien Saumard: asaumard@gmail.com; Jon A. Wellner: jaw@stat.washington.edu

Abstract

We review and formulate results concerning log-concavity and strong-log-concavity in both discrete and continuous settings. We show how preservation of log-concavity and strongly log-concavity on $\mathbb{R}$ under convolution follows from a fundamental monotonicity result of Efron (1969). We provide a new proof of Efron's theorem using the recent asymmetric Brascamp-Lieb inequality due to Otto and Menz (2013). Along the way we review connections between log-concavity and other areas of mathematics and statistics, including concentration of measure, log-Sobolev inequalities, convex geometry, MCMC algorithms, Laplace approximations, and machine learning.

Keywords and phrases

concave; convex; convolution; inequalities; log-concave; monotone; preservation; strong log-concave

1. Introduction: log-concavity

Log-concave distributions and various properties related to log-concavity play an increasingly important role in probability, statistics, optimization theory, econometrics and other areas of applied mathematics. In view of these developments, the basic properties and facts concerning log-concavity deserve to be more widely known in both the probability and statistics communities. Our goal in this survey is to review and summarize the basic preservation properties which make the classes of log-concave densities, measures, and functions so important and useful. In particular we review preservation of log-concavity and “strong log-concavity” (to be defined carefully in section 2) under marginalization, convolution, formation of products, and limits in distribution. The corresponding notions for discrete distributions (log-concavity and ultra log-concavity) are also reviewed in section 4.

[*]Research supported in part by NI-AID grant 2R01 AI29168-04, a PIMS post-doctoral fellowship and post-doctoral Fondecyt Grant 3140600

[†]Research supported in part by NSF Grant DMS-1104832, NI-AID grant 2R01 AI291968-04, and the Alexander von Humboldt Foundation

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

A second goal is to acquaint our readers with a useful monotonicity theorem for log-concave distributions on $\mathbb{R}$ due to Efron [1965], and to briefly discuss connections with recent progress concerning "asymmetric" Brascamp-Lieb inequalities. Efron's theorem is reviewed in Section 6.1, and further applications are given in the rest of Section 6.

There have been several reviews of developments connected to log-concavity in the mathematics literature, most notably Das Gupta [1980] and Gardner [2002]. We are not aware of any comprehensive review of log-concavity in the statistics literature, although there have been some review type papers in econometrics, in particular An [1998] and Bagnoli and Bergstrom [2005]. Given the pace of recent advances, it seems that a review from a statistical perspective is warranted.

Several books deal with various aspects of log-concavity: the classic books by Marshall and Olkin [1979] (see also Marshall, Olkin and Arnold [2011]) and Dharmadhikari and Joag-Dev [1988] both cover aspects of log-concavity theory, but from the perspective of majorization in the first case, and a perspective dominated by unimodality in the second case. Neither treats the important notion of strong log-concavity. The recent book by Simon [2011] perhaps comes closest to our current perspective with interesting previously unpublished material from the papers of Brascamp and Lieb in the 1970's and a proof of the Brascamp and Lieb result to the effect that strong log-concavity is preserved by marginalization. Unfortunately Simon does not connect with recent terminology and other developments in this regard and focuses on convexity theory more broadly. Villani [2003] (chapter 6) gives a nice treatment of the Brunn-Minkowski inequality and related results for log-concave distributions and densities with interesting connections to optimal transportation theory. His chapter 9 also gives a nice treatment of the connections between log-Sobolev inequalities and strong log-concavity, albeit with somewhat different terminology. Ledoux [2001] is, of course, a prime source for material on log-Sobolev inequalities and strong log concavity. The nice book on stochastic programming by Prékopa [1995] has its chapter 4 devoted to log-concavity and s—concavity, but has no treatment of strong log-concavity or inequalities related to log-concavity and strong log-concavity. In this review we will give proofs some key results in the body of the review, while proofs of supporting results are postponed to Section 11 (Appendix B).

2. Log-concavity and strong log-concavity: definitions and basic results

We begin with some basic definitions of log-concave densities and measures on $\mathbb{R}^d$.

Definition 2.1. (0-d): A density function p with respect to Lebesgue measure λ on $(\mathbb{R}^d, \mathscr{B}^d)$ is log-concave if $p = e^{-\phi}$ where ϕ is a convex function from $\mathbb{R}^d$ to $(-\infty, \infty]$. Equivalently, p is log-concave if $p = \exp(\tilde{\phi})$ where $\tilde{\phi} = -\phi$ is a concave function from $\mathbb{R}^d$ to $[-\infty, \infty)$.

We will usually adopt the convention that p is lower semi-continuous and $\phi = -\log p$ is upper semi-continuous. Thus $\{x \in \mathbb{R}^d : p(x) > t\}$ is open, while $\{x \in \mathbb{R}^d : \phi(x) \le t\}$ is closed. We will also say that a non-negative and integrable function f from $\mathbb{R}^d$ to $[0, \infty)$ is log-concave if $f = e^{-\phi}$ where ϕ is convex even though f may not be a density; that is $\int_{\mathbb{R}^d} f d\lambda \in (0, \infty)$.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Many common densities are log-concave; in particular all Gaussian densities

$$p_{\mu,\sum}(x)=(2\pi|\sum|)^{-d/2}\exp\left(-\frac{1}{2}(x-\mu)^T\sum^{-1}(x-\mu)\right)$$

with $\mu\in\mathbb{R}^d$ and Σ positive definite are log-concave, and

$$p_C(x)=1_C(x)/\lambda(C)$$

is log-concave for any non-empty, open and bounded convex subset $C\subset\mathbb{R}^d$. With C open, p is lower semi-continuous in agreement with our convention noted above; of course taking C closed leads to upper semi-continuity of p.

In the case $d = 1$, log-concave functions and densities are related to several other important classes. The following definition goes back to the work of Pólya and Schoenberg.

Definition 2.2. Let p be a function on $\mathbb{R}$ (or some subset of $\mathbb{R}$), and let $x_1 < \cdots < x_k$, $y_1 < \cdots < y_k$. Then p is said to be a Pólya frequency function of order k (or $p \in PF_k$) if $\det(p(x_i - y_j)) \geq 0$ for all such choices of the x's and y's. If p is PF_k for every k, then $p \in PF_\infty$, the class of Pólya frequency functions of order ∞.

A connecting link to Pólya frequency functions and to the notion of monotone likelihood ratios, which is of some importance in statistics, is given by the following proposition:

Proposition 2.3.

- **a.** The class of log-concave functions on $\mathbb{R}$ coincides with the class of Pólya frequency functions of order 2.
- **b.** A density function p on $\mathbb{R}$ is log-concave if and only if the translation family $\{p(\cdot - \theta) : \theta \in \mathbb{R}\}$ has monotone likelihood ratio: i.e. for every $\theta_1 < \theta_2$ the ratio $p(x - \theta_2)/p(x - \theta_1)$ is a monotone nondecreasing function of x.

Proof. See Section 11.

Definition 2.4. (0-m): A probability measure P on $(\mathbb{R}^d, \mathscr{B}^d)$ is log-concave if for all non-empty sets A, $B \in \mathscr{B}^d$ and for all $0 < \theta < 1$ we have

$$P(\theta A+(1-\theta)B)\geq P(A)^\theta P(B)^{1-\theta}.$$

It is well-known that log-concave measures have sub-exponential tails, see Borell [1983] and Section 5.1 below. To accommodate densities having tails heavier than exponential, the classes of s—concave densities and measures are of interest.

Definition 2.5. (s-d): A density function p with respect to Lebesgue measure λ on an convex set $C\subset\mathbb{R}^d$ is s—concave if

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

$$p(\theta x+(1-\theta)y) \geq M_s(p(x), p(y);\theta)$$

where the generalized mean $M_s(u, v; \theta)$ is defined for $u, v \geq 0$ by

$$M_s(u,v;\theta) \equiv \begin{cases} (\theta u^s+(1-\theta)v^s)^{1/s}, & s \neq 0, \\ u^\theta v^{1-\theta}, & s=0, \\ \min\{u,v\}, & s=-\infty, \\ \max\{u,v\}, & s=+\infty. \end{cases}$$

Definition 2.6. (s-m): A probability measure P on $(\mathbb{R}^d, \mathscr{B}^d)$ is s—concave if for all non-empty sets A, B in $\mathscr{B}^d$ and for all $\theta \in (0, 1)$,

$$P(\theta A+(1-\theta)B) \geq M_s(P(A), P(B);\theta)$$

where $M_s(u, v; \theta)$ is as defined above.

These classes of measures and densities were studied by Prékopa [1973] in the case $s = 0$ and for all $s \in \mathbb{R}$ by Brascamp and Lieb [1976], Borell [1975], Borell [1974], and Rinott [1976]. The main results concerning these classes are nicely summarized by Dharmadhikari and Joag-Dev [1988]; see especially sections 2.3-2.8 (pages 46-66) and section 3.3 (pages 84-99). In particular we will review some of the key results for these classes in the next section. For bounds on densities of s—concave distributions on $\mathbb{R}$ see Doss and Wellner [2013]; for probability tail bounds for s—concave measures on $\mathbb{R}^d$, see Bobkov and Ledoux [2009]. For moment bounds and concentration inequalities for s—concave distributions with $s < 0$ see Adamczak et al. [2012] and Guédon [2012], section 3.

A key theorem connecting probability measures to densities is as follows:

Theorem 2.7. Suppose that P is a probability measure on $(\mathbb{R}^d, \mathscr{B}^d)$ such that the affine hull of supp(P) has dimension d. Then P is a log-concave measure if and only if it has a log-concave density function p on $\mathbb{R}^d$; that is $p = e^\phi$ with ϕ concave satisfies

$$P(A)\int_A p d\lambda \text{ for } A \in \mathscr{B}^d.$$

For the correspondence between s—concave measures and t—concave densities, see Borell [1975], Brascamp and Lieb [1976] section 3, Rinott [1976], and Dharmadhikari and Joag-Dev [1988].

One of our main goals here is to review and summarize what is known concerning the (smaller) classes of (what we call) *strongly log-concave* densities. This terminology is not completely standard. Other terms used for the same or essentially the same notion include:

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

- *Log-concave perturbation of Gaussian;* Villani [2003], Caffarelli [2000]. pages 290-291.
- *Gaussian weighted log-concave;* Brascamp and Lieb [1976] pages 379, 381.
- *Uniformly convex potential:* Bobkov and Ledoux [2000], abstract and page 1034, Gozlan and Léonard [2010], Section 7.
- Strongly convex potential: Caffarelli [2000].

In the case of real-valued discrete variables the comparable notion is called *ultra log-concavity;* see e.g. Liggett [1997], Johnson, Kontoyiannis and Madiman [2013], and Johnson [2007]. We will re-visit the notion of ultra log-concavity in Section 4.

Our choice of terminology is motivated in part by the following definition from convexity theory: following Rockafellar and Wets [1998], page 565, we say that a proper convex function $h:\mathbb{R}^d\to\bar{\mathbb{R}}$ is *strongly convex* if there exists a positive number c such that

$$h(\theta x+(1-\theta)y)\leq\theta h(x)+(1-\theta)h(y)-\frac{1}{2}c\theta(1-\theta)\|x-y\|^2$$

for all $x, y\in\mathbb{R}^d$ and $\theta\in(0,1)$. It is easily seen that this is equivalent to convexity of $h(x)-(1/2)c\|x\|^2$ (see Rockafellar and Wets [1998], Exercise12.59, page 565).

Thus our first definition of strong log-concavity of a density function p on $\mathbb{R}^d$ is as follows:

Definition 2.8. For any $\sigma^2>0$ define the class of strongly log-concave densities with variance parameter σ^2, or $SLC_1(\sigma^2, d)$ to be the collection of density functions p of the form

$$p(x)=g(x)\phi_{\sigma^2 I}(x)$$

for some log-concave function g where, for a positive definite matrix Σ and $\mu\in\mathbb{R}^d$, $\varphi_\Sigma(\cdot-\mu)$ denotes the $N_d(\mu,\Sigma)$ density given by

$$\phi_{\sum}(x-\mu)=(2\pi|\sum|)^{-d/2}\exp\left(-\frac{1}{2}(x-\mu)^T\sum^{-1}(x-\mu)\right). \quad (2.1)$$

If a random vector X has a density p of this form, then we also say that X is strongly log-concave.

Note that this agrees with the definition of strong convexity given above since,

$$h(x)\equiv-\log p(x)=-\log g(x)+d\log(\sigma\sqrt{2\pi})+\frac{|x|^2}{2\sigma^2},$$

so that

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

$$-\log p(x) - \frac{|x|^2}{2\sigma^2} = -\log g(x) + d\log(\sigma\sqrt{2\pi})$$

is convex; i.e. $-\log p(x)$ is strongly convex with $c = 1/\sigma^2$. Notice however that if $p \in SLC_1(\sigma^2, d)$ then larger values of σ^2 corresp to smaller values of $c = 1/\sigma^2$, and hence p becomes *less* strongly log-concave as σ^2 increases. Thus in our definition of strong log-concavity the coefficient σ^2 measures the "flatness" of the convex potential

It will be useful to relax this definition in two directions: by allowing the Gaussian distribution to have a non-singular covariance matrix Σ other than the identity matrix and perhaps a non-zero mean vector μ. Thus our second definition is as follows.

Definition 2.9. Let Σ be a $d \times d$ positive definite matrix and let $\mu \in \mathbb{R}^d$. We say that a random vector X and its density function p are strongly log-concave and write $p \in SLC_2(\mu, \Sigma, d)$ if

$$p(x) = g(x)\phi_{\sum}(x-\mu) \text{ for } x \in \mathbb{R}^d$$

for some log-concave function g where $\varphi_\Sigma(\cdot - \mu)$ denotes the $N_d(\mu, \Sigma)$ density given by (2.1).

Note that $SLC_2(0, \sigma^2 I, d) = SLC_1(\sigma^2, d)$ as in Definition 2.8. Furthermore, if $p \in SLC_2(\mu, \Sigma, d)$ with Σ non-singular, then we can write

$$\begin{aligned} p(x) &= g(x)\frac{\phi_{\sum}(x-\mu)}{\phi_{\sum}(x)} \\ &\cdot \frac{\phi_{\sum}(x)}{\phi_{\sigma^2 I}(x)}\phi_{\sigma^2 I}(x) \\ &= g(x)\exp(\mu^T \sum^{-1} x \\ &\quad - (1/2)\mu^T \sum^{-1} \mu T) \\ &\cdot \exp\left(-\frac{1}{2}x^T(\sum^{-1} \right. \\ &\quad \left. - \frac{1}{\sigma^2}I)x\right) \\ &\cdot \phi_{\sigma^2 I}(x) \equiv h(x)\phi_{\sigma^2 I}(x), \end{aligned}$$

where $\Sigma^{-1} - I/\sigma^2$ is positive definite if $1/\sigma^2$ is smaller than the smallest eigenvalue of Σ^{-1}. In this case, h is log-concave, so $p \in SLC_1(\sigma^2, d)$.

Example 2.10. (Gaussian densities) If $X \sim p$ where p is the $N_d(0, \Sigma)$ density with Σ positive definite, then X (and p) is strongly log-concave $SLC_2(0, \Sigma, d)$ and hence also log-concave.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

In particular for $d = 1$, if $X \sim p$ where p is the $N_1(0, \sigma^2)$ density, then X (and p) is $SLC_1(\sigma^2, 1) = SLC_2(0, \sigma^2, 1)$ and hence is also log-concave. Note that $\varphi_X''(x) \equiv (-\log p)'(x) = 1/\sigma^2$ is constant in this latter case.

Example 2.11. (Logistic density) If $X \sim p$ where $p(x) = e^{-x}/(1 + e^{-x})^2 = (1/4)/(\cosh(x/2))^2$, then X (and p) is log-concave and even strictly log-concave since $\varphi_X''(x) = (-\log p)''(x) = 2p(x) > 0$ for all $x \in \mathbb{R}$, but X is not strongly log-concave.

Example 2.12. (Bridge densities) If $X \sim p_\theta$ where, for $\theta \in (0, 1)$,

$$p_\theta(x) = \frac{\sin(\pi\theta)}{2\pi(\cosh(\theta x) + \cos(\pi\theta))},$$

then X (and p_θ) is log-concave for $\theta \in (0, 1/2]$, but fails to be log-concave for $\theta \in (1/2, 1)$. For $\theta \in (1/2, 1)$, $\varphi_\theta''(x) = (-\log p_\theta)''(x)$ is bounded below, by some negative value depending on θ, and hence these densities are semi-log-concave in the terminology of Cattiaux and Guillin [2013] who introduce this further generalization of log-concave densities by allowing the constant in the definition of a class of strongly log-concave densities to be negative as well as positive. This particular family of densities on $\mathbb{R}$ was introduced in the context of binary mixed effects models by Wang and Louis [2003].

Example 2.13. (Subbotin density) If $X \sim p_r$ where $p_r(x) = C_r \exp(-|x|^r/r)$ for $x \in \mathbb{R}$ and $r > 0$ where $C_r = 1/[2\Gamma(1/r)r^{1/r-1}]$, then X (and p_r) is log-concave for all $r \geq 1$. Note that this family includes the Laplace (or double exponential) density for $r = 1$ and the Gaussian (or standard normal) density for $r = 2$. The only member of this family that is strongly log-concave is p_2, the standard Gaussian density, since $(-\log p)''(x) = (r - 1)|x|^{r-2}$ for $x \neq 0$.

Example 2.14. (Supremum of Brownian bridge) If $\mathbb{U}$ is a standard Brownian bridge process on [0, 1], Then $P(\sup_{0 \leq t \leq 1} \mathbb{U}t) > x) = \exp(-2x^2)$ for $x > 0$, so the density is $f(x) = 4x \exp(-2x^2)1_{(0,\infty)}(x)$, which is strongly log concave since $(-\log f)''((x) = 4 + x^{-2} \geq 4$. This is a special case of the Weibull densities $f_\beta(x) = \beta x^{\beta-1} \exp(-x^\beta)$ which are log-concave if $\beta \geq 1$ and strongly log-concave for $\beta \geq 2$. For more about suprema of Gaussian processes, see Section 9.3 below.

For further interesting examples, see Dharmadhikari and Joag-Dev [1988] and Prékopa [1995].

There exist a priori many ways to strengthen the property of log-concavity An very interesting notion is for instance the log-concavity of order p. This is a one-dimensional notion, and even if it can be easily stated for one-dimensional measures on $\mathbb{R}^d$, see Bobkov and Madiman [2011] Section 4, we state it in its classical way on $\mathbb{R}$.

Definition 2.15. A random variable $\xi > 0$ is said to have a log-concave distribution of order $p \geq 1$, if it has a density of the form $f(x) = x^{p-1}g(x)$, $x > 0$, where the function g is log-concave on $(0, \infty)$.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Notice that the notion of log-concavity of order 1 coincides with the notion of log-concavity for positive random variables. Furthermore, it is easily seen that log-concave variables of order $p > 1$ are more concentrated than log-concave variables. Indeed, with the notations of Definition 2.15 and setting moreover $f = \exp(-\phi_f)$ and $g = \exp(-\phi_g)$, assuming that f is $\mathscr{C}^2$ we get,

$$\mathrm{Hess}\varphi_f = \mathrm{Hess}\varphi_g + \frac{p-1}{x^2}.$$

As a matter of fact, the exponent p strengthens the Hessian of the potential of g, which is already a log-concave density. Here are some example of log-concave variables of order p.

Example 2.16. The Gamma distribution with $\alpha \geq 1$ degrees of freedom, which has the density $f(x) = \Gamma(\alpha)^{-1} x^{\alpha-1} e^{-x} 1_{(0,\infty)}(x)$ is log-concave of order α.

Example 2.17. The Beta distribution $B_{\alpha,\beta}$ with parameters $\alpha \geq 1$ and $\beta \geq 1$ is log-concave of order α. We recall that its density g is given by $g(x) = B(\alpha,\beta)^{-1} x^{\alpha-1} (1-x)^{\beta-1} 1_{(0,1)}(x)$.

Example 2.18. The Weibull density of parameter $\beta \geq 1$, given by $h_\beta(x) = \beta x^{\beta-1} \exp(-x^\beta) 1_{(0,\infty)}(x)$ is log-concave of order β.

It is worth noticing that when X is a log-concave vector in $\mathbb{R}^d$ with spherically invariant distribution, then the Euclidian norm of X, denoted $\|X\|$, follows a log-concave distribution of order $d-1$ (this is easily seen by transforming to polar coordinates; see Bobkov [2003] for instance). The notion of log-concavity of order p is also of interest when dealing with problems in greater dimension. Indeed, a general way to reduce a problem defined by d-dimensional integrals to a problem involving one-dimensional integrals is given by the "localization lemma" of Lovász and Simonovits [1993]; see also Kannan, Lovász and Simonovits [1997]. We will not further review this notion and we refer to Bobkov [2003], Bobkov [2010] and Bobkov and Madiman [2011] for nice results related in particular to concentration of log-concave variables of order p.

The following sets of equivalences for log-concavity and strong log-concavity will be useful and important. To state these equivalences we need the following definitions from Simon [2011], page 199. First, a subset A of $\mathbb{R}^d$ is *balanced* (Simon [2011]) or *centrally symmetric* (Dharmadhikari and Joag-Dev [1988]) if $x \in A$ implies $-x \in A$.

Definition 2.19. A nonnegative function f on $\mathbb{R}^d$ is convexly layered if $\{x : f(x) > \alpha\}$ is a balanced convex set for all $\alpha > 0$. It is called even, radial monotone if (*i*) $f(-x) = f(x)$ and (ii) $f(rx) \geq f(x)$ for all $0 \leq r \leq 1$ and all $x \in \mathbb{R}^d$.

Proposition 2.20. (Equivalences for log-concavity). Let $p = e^{-\phi}$ be a density function with respect to Lebesgue measure λ on $\mathbb{R}^d$; that is, $p \geq 0$ and $\int_{\mathbb{R}^d} p d\lambda = 1$. Suppose that $\phi \in C^2$. Then the following are equivalent:

a. $\phi = -\log p$ is convex; i.e. p is log-concave.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

b. $\nabla\phi = -\nabla p/p : \mathbb{R}^d \to \mathbb{R}^d$ is monotone:

$$\langle \nabla\varphi(x_2) - \nabla\varphi(x_1), x_2 - x_1 \rangle \geq 0 \text{ for } \text{ all } x_1, x_2 \in \mathbb{R}^d.$$

c. $\nabla^2\phi = \nabla^2(\phi) \geq 0$.

d. $J_a(x; p) = p(a + x)p(a - x)$ is convexly layered for each $a \in \mathbb{R}^d$.

e. $J_a(x;p)$ is even and radially monotone.

f. p is mid-point log-concave: for all $x_1, x_2 \in \mathbb{R}^d$,

$$p\left(\frac{1}{2}x_1 + \frac{1}{2}x_2\right) \geq p(x_1)^{1/2} p(x_2)^{1/2}.$$

The equivalence of (a), (d), (e), and (f) is proved by Simon [2011], page 199, without assuming that $p \in C^2$. The equivalence of (a), (b), and (c) under the assumption $\phi \in C^2$ is classical and well-known. This set of equivalences generalizes naturally to handle $\phi \notin C^2$, but ϕ proper and upper semicontinuous so that p is lower semicontinuous; see Section 5.2 below for the adequate tools of convex regularization.

In dimension 1, Bobkov [1996] proved the following further characterizations of log-concavity on $\mathbb{R}$.

Proposition 2.21 (Bobkov [1996]). Let μ be a nonatomic probability measure with distribution function $F = \mu((-\infty, x])$, $x \in \mathbb{R}$. Set $a = \inf\{x \in \mathbb{R} : F(x) > 0\}$ and $b = \sup\{x \in \mathbb{R} : F(x) < 1\}$. Assume that F strictly increases on (a, b), and let $F^{-1} : (0, 1) \to (a, b)$ denote the inverse of F restricted to (a, b). Then the following properties are equivalent:

a. μ is log-concave;

b. for all $h > 0$, the function $R_h(p) = F(F^{-1}(p) + h)$ is concave on (a, b);

c. μ has a continuous, positive density f on (a, b) and, moreover, the function $I(p) = f(F^{-1}(p))$ is concave on $(0, 1)$.

Properties (b) and (c) of Proposition 2.21 were first used in Bobkov [1996] along the proofs of his description of the extremal properties of half-planes for the isoperimetric problem for log-concave product measures on $\mathbb{R}^d$. In Bobkov and Madiman [2011] the concavity of the function $I(p) = f(F^{-1}(p))$ defined in point (c) of Proposition 2.21, plays a role in the proof of concentration and moment inequalities for the following information quantity: $-\log f(X)$ where X is a random vector with log-concave density f. Recently, Bobkov and Ledoux [2014] used the concavity of I to prove upper and lower bounds on the variance of the order statistics associated to an i.i.d. sample drawn from a log-concave measure on $\mathbb{R}$. The latter results allow then the authors to prove refined bounds on some Kantorovich transport distances between the empirical measure associated to the i.i.d. sample and the log-concave measure on $\mathbb{R}$. For more facts about the function I for general measures on $\mathbb{R}$ and in particular, its relationship to isoperimetric profiles, see Appendix A. 4-6 of Bobkov and Ledoux [2014].

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Example 2.22. If μ is the standard Gaussian measure on the real line, then I is symmetric around 1/2 and there exist constants $0 < c_0 \le c_1 < \infty$ such that

$$c_0 t\sqrt{\log(1/t)} \le I(t) \le c_1 t\sqrt{\log(1/t)},$$

for $t \in (0, 1/2]$ (*see* Bobkov and Ledoux [2014] p.73).

We turn now to similar characterizations of strong log-concavity.

Proposition 2.23. (Equivalences for strong log-concavity, SLC_1). Let $p = e^{-\phi}$ be a density function with respect to Lebesgue measure λ on $\mathbb{R}^d$; that is, $p \ge 0$ and $\int_{\mathbb{R}^d} p d\lambda = 1$. Suppose that $\phi \in C^2$. Then the following are equivalent:

- **a.** p is strongly log-concave; $p \in \mathrm{SLC}_1(\sigma^2, d)$.
- **b.** $\rho(x) = \nabla\phi(x) - x/\sigma^2 : \mathbb{R}^d \to \mathbb{R}^d$ is monotone:

$$\langle \rho(x_2) - \rho(x_1), x_2 - x_1 \rangle \ge 0 \text{ for all } x_1, x_2 \in \mathbb{R}^d.$$

- **c.** $\nabla\rho(x) = \nabla^2\phi - I/\sigma^2 \ge 0$.
- **d.** For each $a \in \mathbb{R}^d$ the function

$$J_a^\phi(x;p) \equiv \frac{p(a+x)p(a-x)}{\phi_{\sigma^2 I/2}(x)}$$

  is convexly layered.
- **e.** The function $J_a^\phi(x;p)$ in (*d*) is even and radially monotone for all $a \in \mathbb{R}^d$.
- **f.** For all $x, y \in \mathbb{R}^d$,

$$p\left(\frac{1}{2}x + \frac{1}{2}y\right) \ge p(x)^{1/2} p(y)^{1/2} \exp\left(\frac{1}{8}|x-y|^2\right).$$

Proof. See Section 11.

We investigate the extension of Proposition 2.21 concerning log-concavity on R, to the case of strong log-concavity. (The following result is apparently new.) Recall that a function h is strongly concave on (a, b) with parameter $c > 0$ (or c-strongly concave), if for any $x, y \in (a, b)$, any $\theta \in (0, 1)$,

$$h(\theta x + (1-\theta)y) \ge \theta h(x) + (1-\theta)h(y) + \frac{1}{2}c\theta(1-\theta)\|x-y\|^2.$$

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Proposition 2.24. Let μ be a nonatomic probability measure with distribution function $F = \mu((-\infty, x])$, $x \in \mathbb{R}$. Set $a = \inf\{x \in \mathbb{R} : F(x) > 0\}$ and $b = \sup\{x \in \mathbb{R} : F(x) < 1\}$, possibly infinite. Assume that F strictly increases on (a, b), and let $F^{-1} : (0,1) \to (a, b)$ denote the inverse of F restricted to (a, b). Suppose that X is a random variable with distribution μ. Then the following properties hold:

- **i.** If $X \in SLC_1(c, 1)$, $c > 0$, then $I(p) = f(F^{-1}(p))$ is $(c\|f\|_\infty)^{-1}$ -strongly concave and $\left(c^{-1}\sqrt{\mathrm{Var}(X)}\right)$-strongly concave on (0, 1).
- **ii.** The converse of point (***i***) is false: there exists a log-concave variable X which is not strongly concave (for any parameter $c > 0$) such that the associated I function is strongly log-concave on (0, 1).
- **iii.** There exist a strongly log-concave random variable $X \in SLC(c, 1)$ and $h_0 > 0$ such that the function $R_{h_0}(p) = F(F^{-1}(p) + h_0)$ is concave but not strongly concave on (a, b).
- **iv.** There exists a log-concave random variable X which is not strongly log-concave (for any positive parameter), such that for all $h > 0$, the function $R_{h_0}(p) = F(F^{-1}(p) + h)$ is strongly concave on (a, b).

From (**i**) and (**ii**) in Proposition 2.24, we see that the strong concavity of the function I is a necessary but not sufficient condition for the strong log-concavity of X. Points (**iii**) and (**iv**) state that no relations exist in general between the strong log-concavity of X and strong concavity of its associated function R_h.

Proof. See Section 11.

The following proposition gives a similar set of equivalences for our second definition of strong log-concavity, Definition 2.9.

Proposition 2.25. (Equivalences for strong log-concavity, SLC_2). Let $p = e^{-\phi}$ be a density function with respect to Lebesgue measure λ on $\mathbb{R}^d$; that is, $p \geq 0$ and $\int_{\mathbb{R}^d} p d\lambda = 1$. Suppose that $\phi \in C^2$. Then the following are equivalent:

- **a.** p is strongly log-concave; $p \in SLC_2(\mu, \Sigma, d)$ with $\Sigma > 0$, $\mu \in \mathbb{R}^d$.
- **b.** $\rho(x) = \nabla\phi(x) - \Sigma^{-1}(x - \mu) : \mathbb{R}^d \to \mathbb{R}^d$ is monotone:

$$\langle \rho(x_2) - \rho(x_1), x_2 - x_1 \rangle \geq 0 \text{ for all } x_1, x_2 \in \mathbb{R}^d.$$

- **c.** $\nabla\rho(x) = \nabla^2\phi - \Sigma^{-1} \geq 0$.
- **d.** For each $a \in \mathbb{R}^d$, the function

$$J_a^\phi(x;p) = p(a+x)p(a-x)/\phi_{\sum/2}(x)$$

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

is convexly layered. (*e*) For each $a \in \mathbb{R}^d$ the function $J_a^{\phi}(x;p)$ in (*d*) is even and radially monotone.

e. For all $x, y \in \mathbb{R}^d$,

$$p\left(\frac{1}{2}x+\frac{1}{2}y\right) \geq p(x)^{1/2}p(y)^{1/2}\exp\left(\frac{1}{8}(x-y)^T\sum^{-1}(x-y)\right).$$

Proof. To prove Proposition 2.25 it suffices to note the log-concavity of $g(x) = p(x)/\varphi_{\Sigma/2}(x)$ and to apply Proposition 2.20 (which holds as well for log-concave functions). The claims then follow by straightforward calculations; see Section 11 for more details.

3. Log-concavity and strong log-concavity: preservation theorems

Both log-concavity and strong log-concavity are preserved by a number of operations. Our purpose in this section is to review these preservation results and the methods used to prove such results, with primary emphasis on: (a) affine transformations, (b) marginalization, (c) convolution. The main tools used in the proofs will be: (i) the Brunn-Minkowski inequality; (ii) the Brascamp-Lieb Poincaré type inequality; (iii) Prékopa's theorem; (iv) Efron's monotonicity theorem.

3.1. Preservation of log-concavity

3.1.1. Preservation by affine transformations—Suppose that X has a log-concave distribution P on $(\mathbb{R}^d, \mathscr{B}^d)$, and let A be a non-zero real matrix of order $m \times d$. Then consider the distribution Q *of* $Y = AX$ on $\mathbb{R}^m$.

Proposition 3.1. (log-concavity is preserved by affine transformations). The probability measure Q on $\mathbb{R}^m$ defined by $Q(B) = P(AX \in B)$ for $B \in \mathscr{B}^m$ is a log-concave probability measure. If P is non-degenerate log-concave on $\mathbb{R}^d$ with density p and $m = d$ with A of rank d, then Q is non-degenerate with log-concave density q.

Proof. See Dharmadhikari and Joag-Dev [1988], Lemma 2.1, page 47.

3.1.2. Preservation by products—Now let P_1 and P_2 be log-concave probability measures on $(\mathbb{R}^{d_1}, \mathscr{B}^{d_1})$ and $(\mathbb{R}^{d_2}, \mathscr{B}^{d_2})$ respectively. Then we have the following preservation result for the product measure $P_1 \times P_2$ on $(\mathbb{R}^{d_1} \times \mathbb{R}^{d_2}, \mathscr{B}^{d_1} \times \mathscr{B}^{d_2})$:

Proposition 3.2. (log-concavity is preserved by products) If P_1 and P_2 are log-concave probability measures then the product measure $P_1 \times P_2$ is a log-concave probability measure.

Proof. See Dharmadhikari and Joag-Dev [1988], Theorem 2.7, page 50. A key fact used in this proof is that if a probability measure P on $(\mathbb{R}^d, \mathscr{B}^d)$ assigns zero mass to every hyperplane in $\mathbb{R}^d$, then log-concavity of P holds if and only if $P(\theta A + (1-\theta)B) \geq P(A)^{\theta} P(B)^{1-\theta}$ for all rectangles A, B with sides parallel to the coordinate axes; see Dharmadhikari and Joag-Dev [1988], Theorem 2.6, page 49.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

3.1.3. Preservation by marginalization—Now suppose that p is a log-concave density on $\mathbb{R}^{m+n}$ and consider the marginal density $q(y) = \int_{\mathbb{R}^m} p(x, y)dx$. The following result due to Prékopa [1973] concerning preservation of log-concavity was given a simple proof by Brascamp and Lieb [1976] (Corollary 3.5, page 374). In fact they also proved the whole family of such results for s—concave densities.

Theorem 3.3. (log-concavity is preserved by marginalization; Prékopa's theorem). Suppose that p is log-concave on $\mathbb{R}^{m+n}$ and let $q(y) = \int_{\mathbb{R}^m} p(x, y)dx$. Then q is log-concave.

This theorem is a center piece of the entire theory. It was proved independently by a number of mathematicians at about the same time: these include Prékopa [1973], building on Dinghas [1957], Prékopa [1971], Brascamp and Lieb [1974], Brascamp and Lieb [1975], Brascamp and Lieb [1976], Borell [1975], Borell [1974], and Rinott [1976]. Simon [2011], page 310, gives a brief discussion of the history, including an unpublished proof of Theorem 3.3 given in Brascamp and Lieb [1974]. Many of the proofs (including the proofs in Brascamp and Lieb [1975], Borell [1975], and Rinott [1976]) are based fundamentally on the Brunn-Minkowski inequality; see Das Gupta [1980], Gardner [2002], and Maurey [2005] for useful surveys.

We give two proofs here. The first proof is a *transportation* argument from Ball, Barthe and Naor [2003]; the second is a proof from Brascamp and Lieb [1974] which has recently appeared in Simon [2011].

Proof. (Via *transportation*). We can reduce to the case $n = 1$ since it suffices to show that the restriction of q to a line is log-concave. Next note that an inductive argument shows that the claimed log-concavity holds for $m + 1$ if it holds for m, and hence it suffices to prove the claim for $m = n = 1$.

Since log-concavity is equivalent to mid-point log concavity (by the equivalence of (a) and (e) in Proposition 2.20), we only need to show that

$$q\left(\frac{u+v}{2}\right) \geq q(u)^{1/2} q(v)^{1/2} \quad (3.2)$$

for all $u, v \in \mathbb{R}$. Now define

$$f(x)=p(x,u),\ g(x)=p(x,v),\ h(x)=p(x,(u+v)/2).$$

Then (3.2) can be rewritten as

$$\int h(x)\mathrm{dx} \geq \left(\int f(x)\mathrm{dx}\right)^{1/2}\left(\int g(x)\mathrm{dx}\right)^{1/2}.$$

From log-concavity of p we know that

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

$$h\left(\frac{z+w}{2}\right)=p\left(\frac{z+w}{2},\frac{u+v}{2}\right)\geq p(z,u)^{1/2}p(w,v)^{1/2}=f(z)^{1/2}g(w)^{1/2}. \quad (3.3)$$

By homogeneity we can arrange f, g, and h so that $\int f(x)dx = \int g(x)dx = 1$; if not, replace f and g with $\tilde{f}$ and $\tilde{g}$ defined by $\tilde{f}(x) = f(x)/\int f(x')dx' = f(x)/q(u)$ and $\tilde{g}(x) = g(x)/\int g(x')dx' = g(x)/q(v)$.

Now for the transportation part of the argument: let Z be a real-valued random variable with distribution function K having smooth density k. Then define maps S and T by $K(z) = F(S(z))$ and $K(z) = G(T(z))$ where F and G are the distribution functions corresponding to f and g. Then

$$k(z)=f(S(z))S'(z)=g(T(z))T'(z)$$

where S', $T' \geq 0$ since the same is true for k, f, and g, and it follows that

$$1=\int k(z)\mathrm{dz}=\int f(S(z))^{1/2}g(T(z))^{1/2}(S'(z))^{1/2}(T'(z))^{1/2}\mathrm{dz}\leq\int h\left(\frac{S(z)+T(z)}{2}\right)(S'(z))^{1/2}(T'(z))^{1/2}\mathrm{dz}\leq\int h\left(\frac{S(z)+T(z)}{2}\right)\cdot\frac{S'(z)+T'(z)}{2}\mathrm{dz}=\int h(x)\mathrm{dx}$$

by the inequality (3.3) in the first inequality and by the arithmetic - geometric mean inequality in the second inequality.

Proof. (Via *symmetrization*). By the same induction argument as in the first proof we can suppose that $m = 1$. By an approximation argument we may assume, without loss of generality that p has compact support and is bounded.

Now let $a \in \mathbb{R}^n$ and note that

$$J_a(y;q)=q(a+y)q(a-y)=\int\int p(x,a+y)p(z,a-y)\mathrm{dxdz}=2\int\int p(u+v,a+y)p(u-v,a-y)\mathrm{dudv}=2\int\int J_{u,a}(v,y;p)\mathrm{dudv}$$

where, for (u, a) fixed, the integrand is convexly layered by Proposition 2.20 (d). Thus by the following Lemma 3.4, the integral over v is an even lower semi-continuous function of y for each fixed u, a. Since this class of functions is closed under integration over an indexing parameter (such as u), the integration over u also yields an even radially monotone function, and by Fatou's lemma $J_a(y; g)$ is also lower semicontinuous. It then follows from Proposition 2.20 again that g is log-concave.

Lemma 3.4. Let f be a lower semicontinuous convexly layered function on $\mathbb{R}^{n+1}$ written as $f(x, t)$, $x \in \mathbb{R}^n$, $t \in \mathbb{R}$. Suppose that f is bounded and has compact support. Let

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

$$g(x)=\int_{\mathbb{R}} f(x,t)\mathrm{d}t.$$

Then g is an even, radially monotone, lower semicontinuous function.

Proof. First note that sums and integrals of even radially monotone functions are again even and radially monotone. By the wedding cake representation

$$f(x)=\int_0^\infty 1\{f(x)>t\}\mathrm{d}t,$$

it suffices to prove the result when f is the indicator function of an open balanced convex set K. Thus we define

$$K(x)=\{t\in\mathbb{R}:(x,t)\in K\},\ \text{for}\, x\in\mathbb{R}^n.$$

Thus $K(x) = (c(x), d(x))$, an open interval in $\mathbb{R}$ and we see that

$$g(x)=d(x)-c(x).$$

But convexity of K implies that $c(x)$ is convex and $d(x)$ is concave,and hence $g(x)$ is concave. Since K is balanced, it follows that $c(-x) = -d(x)$, or $d(-x) = -c(x)$, so g is even. Since an even concave function is even radially monotone, and lower semicontinuity of g holds by Fatou's lemma, the conclusion follows.

3.1.4. Preservation under convolution—Suppose that X, Y are independent with log-concave distributions P and Q on $(\mathbb{R}^d, \mathscr{B}^d)$, and let R denote the distribution of $X+Y$. The following result asserts that R is log-concave as a measure on $\mathbb{R}^d$.

Proposition 3.5. (log-concavity is preserved by convolution). Let P and Q be two log-concave distributions on $(\mathbb{R}^d, \mathscr{B}^d)$ and let R be the convolution defined by $R(B) = \int_{\mathbb{R}^d} P(B-y)dQ(y)$ for $B\in\mathscr{B}^d$. Then R is log-concave.

Proof. It suffices to prove the proposition when P and Q are absolutely continuous with densities p and q on $\mathbb{R}^{\mathrm{d}}$. Now $h(x, y) = p(x-y)q(y)$ is log-concave on $\mathbb{R}^{2d}$, and hence by Proposition 3.3 it follows that

$$r(y)=\int_{\mathbb{R}^d} h(x,y)\mathrm{d}y=\int_{\mathbb{R}^d} p(x-y)q(y)\mathrm{d}y$$

is log-concave.

Proposition 3.5 was proved when $d = 1$ by Schoenberg [1951] who used the PF_2 terminology of Pólya frequency functions. In fact all the Pólya frequency classes PF_k, $k\geq 2$, are closed

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

under convolution as shown by Karlin [1968]; see Marshall, Olkin and Arnold [2011], Lemma A.4 (page 758) and Proposition B.1, page 763. The first proof of Proposition 3.5 when $d \geq 2$ is apparently due to Davidovič, Korenbljum and Hacet [1969]. While the proof given above using Prékopa's theorem is simple and quite basic, there are at least two other proofs according as to whether we use:

a. the equivalence between log-concavity and monotonicity of the scores of f, or

b. the equivalence between log-concavity and non-negativity of the matrix of second derivatives (or Hessian) of $-\log f$, assuming that the second derivatives exist.

The proof in (a) relies on Efron's inequality when $d = 1$, and was noted by Wellner [2013] in parallel to the corresponding proof of ultra log-concavity in the discrete case given by Johnson [2007]; see Theorem 4.1. We will return to this in Section 6. For $d > 1$ this approach breaks down because Efron's theorem does not extend to the multivariate setting without further hypotheses. Possible generalizations of Efron's theorem will be discussed in Section 7. The proof in (b) relies on a Poincaré type inequality of Brascamp and Lieb [1976]. These three different methods are of some interest since they all have analogues in the case of proving that strong log-concavity is preserved under convolution.

It is also worth noting the following difference between the situation in one dimension and the result for preservation of convolution in higher dimensions: as we note following Theorems 29 and 33, Ibragimov [1956a] and Keilson and Gerber [1971] showed that in the one-dimensional continuous and discrete settings respectively that if $p \star q$ is unimodal for every unimodal q, then p is log-concave. The analogue of this for $d > 1$ is more complicated in part because of the great variety of possible definitions of "unimodal" in this case; see Dharmadhikari and Joag-Dev [1988] chapters 2 and 3 for a thorough discussion. In particular Sherman [1955] provided the following counterexample when the notion of unimodality is taken to be *centrally symmetric convex* unimodality; that is, the sets $S_c(p) = \{x \in \mathbb{R}^d : p(x) \geq c\}$ are symmetric and convex for each $c \geq 0$. Let p be the uniform density on $[-1, 1]^2$ (so that $p(x) = (1/4)1_{[-1,1]^2}(x)$); then p is log-concave. Let q be the density given by 1/12 on $[-1, 1]^2$ and 1/24 on $([-1, 1] \times (1, 5]) \bigcup ([-1, 1] \times [-5, -1))$. Thus q is centrally symmetric convex (and hence also *quasi-concave*, $q \in \mathscr{P}_{-\infty}$ as in Definition 2.5. But $h = p \star q$ is not centrally symmetric convex (and also is not quasi-concave), since the sets $S_c(h)$ are not convex: see Figure 1.

3.1.5. Preservation by (weak) limits—Now we consider preservation of log-concavity under convergence in distribution.

Proposition 3.6. (log-concavity is preserved under convergence in distribution). Suppose that $\{P_n\}$ is a sequence of log-concave probability measures on $\mathbb{R}^d$, and suppose that $P_n \to P_0$. Then P_0 is a log-concave probability measure.

Proof. See Dharmadhikari and Joag-Dev [1988], Theorem 2.10, page 53.

Note that the limit measure in Proposition 3.6 might be concentrated on a proper subspace of $\mathbb{R}^d$. If we have a sequence of log-concave densities p_n which converge pointwise to a density function p_0, then by Scheffe's theorem we have $p_n \to p_0$ in $L_1(\lambda)$ and hence $d_{TV}(P_n, P_0) \to$

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

0. Since convergence in total variation implies convergence in distribution we conclude that P_0 is a log-concave measure where the affine hull of supp(P_0) has dimension d and hence P_0 is the measure corresponding to p_0 which is necessarily log-concave by Theorem 2.7.

Recall that the class of normal distributions on $\mathbb{R}^d$ is closed under all the operations discussed above: affine transformation, formation of products, marginalization, convolution, and weak limits. Since the larger class of log-concave distributions on $\mathbb{R}^d$ is also preserved under these operations, the preservation results of this section suggest that the class of log-concave distributions is a very natural nonparametric class which can be viewed naturally as an enlargement of the class of all normal distributions. This has stimulated much recent work on nonparametric estimation for the class of log-concave distributions on $\mathbb{R}$ and $\mathbb{R}^d$: for example, see Dümbgen and Rufibach [2009], Cule and Samworth [2010], Cule, Samworth and Stewart [2010], Walther [2009], Balabdaoui, Rufibach and Wellner [2009], and Henningsson and Astrom [2006], and see Section 9.13 for further details.

3.2. Preservation of strong log-concavity

Here is a theorem summarizing several preservation results for strong log-concavity Parts (a), (b), and (d) were obtained by Henningsson and Astrom [2006].

Theorem 3.7. (Preservation of strong log-concavity)

- **a.** (Linear transformations) Suppose that X has density $p \in SLC_2(0, \Sigma, d)$ and let A be a $d \times d$ nonsingular matrix. Then $Y = AX$ has density $q \in SLC_2\,(0, A\Sigma A^T, d)$ given by $q(y) = p(A^{-1}\,y)\,\det(A^{-1})$.
- **b.** (Convolution) If $Z = X + Y$ where $X \sim p \in SLC_2(0, \Sigma, d)$ and $Y \sim q \in SLC_2\,(0, \Gamma, d)$ are independent, then $Z = X + Y \sim p \star q \in SLC_2\,(0, \Sigma + \Gamma, d)$.
- **c.** (c) (Product distribution) If $X \sim p \in SLC_2\,(0, \Sigma, m)$ and $Y \sim q \in SLC_2\,(0, \Gamma, n)$, then

$$(X, Y) \sim p \cdot q \in \mathrm{SLC}_2\left(0, \begin{pmatrix} \Sigma & 0 \\ 0 & \Gamma \end{pmatrix}, m+n\right).$$

- **d.** (Product function) If $p \in SLC_2(0, \Sigma, d)$ and $q \in SLC_2(0, \Gamma, d)$, then h given by $h(x) = p(x)q(x)$ (which is typically not a probability density function) satisfies $h \in SLC_2(0, (\Sigma^{-1} + \Gamma^{-1})^{-1})$.

Part (b) of Theorem 3.7 is closely related to the following result which builds upon and strengthens Prékopa's Theorem 3.3. It is due to Brascamp and Lieb [1976] (Theorem 4.3, page 380); see also Simon [2011], Theorem 13.13, page 204.

Theorem 3.8. (Preservation of strong log-concavity under marginalization). Suppose that $p \in SLC_2(0, \Sigma, m + n)$. Then the marginal density q on $\mathbb{R}^n$ given by

$$q(x) = \int_{\mathbb{R}^m} p(x, y)\mathrm{d}y$$

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

is strongly log-concave: $q \in SLC_2(0, \Sigma_{11}, m)$ where

$$\sum = \begin{pmatrix} \sum_{11} & \sum_{12} \\ \sum_{21} & \sum_{22} \end{pmatrix}. \quad (3.4)$$

Proof. Since $p \in SLC_2(0, \Sigma, m+n)$ we can write

$$p(x,y) = g(x,y)\phi_{\sum}(x,y) = g(x,y)\frac{1}{(2\pi|\sum|)^{(m+n)/2}}\exp\left(-\frac{1}{2}(x^T, y^T)\begin{pmatrix} \sum_{11} & \sum_{12} \\ \sum_{21} & \sum_{22} \end{pmatrix}^{-1}\begin{pmatrix} x \\ y \end{pmatrix}\right)$$

where g is log-concave. Now the Gaussian term in the last display can be written as

$$\begin{aligned} &\phi_{Y|X}(y|x) \\ &\quad \cdot \phi_X(x) \\ &\quad = \frac{1}{(2\pi|\sum_{22.1}|)^{n/2}}\exp\left(-\frac{1}{2}(y - \sum_{21}\sum_{11}^{-1}x)^T\sum_{22.1}^{-1}(y \right. \\ &\qquad \left. - \sum_{21}\sum_{11}^{-1}x)\right) \\ &\quad \cdot \frac{1}{(2\pi|\sum_{11}|)^{m/2}}\exp\left(\right. \\ &\qquad \left. -\frac{1}{2}x^T\sum_{11}^{-1}x\right) \end{aligned}$$

where $\sum_{22.1} \equiv \sum_{22} - \sum_{21}\sum_{11}^{-1}\sum_{12}$, and hence

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

$$
\begin{aligned}
q(x) = & \int_{\mathbb{R}^n} g(x,y) \frac{1}{(2\pi|\sum_{22.1}|)^{n/2}} \exp\Big(\\
& -\frac{1}{2}(y - \sum_{21}\sum_{11}^{-1}x)^T \sum_{22.1}^{-1}(y \\
& \quad - \sum_{21}\sum_{11}^{-1}x)\Big)\,\mathrm{dy} \\
& \cdot \frac{1}{(2\pi|\sum_{11}|)^{m/2}} \exp\Big(\\
& \quad -\frac{1}{2}x^T \sum_{11}^{-1} x\Big) \\
= & \int_{\mathbb{R}^n} g(x, \tilde{y} + \sum_{21}\sum_{11}^{-1}x) \\
& \cdot \frac{1}{(2\pi|\sum_{22.1}|)^{n/2}} \exp\Big(\\
& \quad -\frac{1}{2}\tilde{y}^T \sum_{22.1}^{-1}\tilde{y}\Big)\, d\tilde{y} \\
& \cdot \frac{1}{(2\pi|\sum_{11}|)^{m/2}} \exp(\\
& \quad -(1/2)x^T \sum_{11}^{-1} x) \equiv h(x)\phi_{\sum_{11}}(x)
\end{aligned}
$$

where

$$
h(x) \equiv \int_{\mathbb{R}^n} g(x, \tilde{y} + \sum_{21}\sum_{11}^{-1}x) \cdot \frac{1}{(2\pi|\sum_{22.1}|)^{n/2}} \exp\left(-\frac{1}{2}\tilde{y}^T \sum_{22.1}^{-1}\tilde{y}\right) d\tilde{y}
$$

is log-concave: g is log-concave, and hence $\tilde{g}(x,\tilde{y}) \equiv g(x, \tilde{y} + \sum_{21}\sum_{11}^{-1}x)$ is log-concave; the product $\tilde{g}(x,\tilde{y}) \cdot \exp(-(1/2)\tilde{y}^T \sum_{22.1}^{-1}\tilde{y})$ is (jointly) log-concave; and hence h is log-concave by Prékopa's Theorem 3.3.

Proof. (Theorem 3.7):

a. The density q is given by $q(y) = p(A^{-1}y)\det(A^{-1})$ by a standard computation. Then since $p \in SLC_2(0, \Sigma, d)$ we can write

$$
q(y) = g(A^{-1}y)\det(A^{-1})\phi_{\sum}(A^{-1}y) = g(A^{-1}y)\phi_{A\sum A^T}(y)
$$

where $g(A^{-1}y)$ is log-concave by Proposition 3.1.

b. If $p \in SLC_2(0, \Sigma, d)$ and $q \in SLC_2(0, \Gamma, d)$, then the function

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

$$h(z,x)=p(x)q(z-x)$$

is strongly log-concave jointly in x and z: since

$$x^T\sum^{-1}x+(z-x)^T\Gamma^{-1}(z-x)=z^T(\sum+\Gamma)^{-1}z+(x-C_z)^T(\sum^{-1}+\Gamma^{-1})(x-C_z)$$

where $C\equiv(\Sigma^{-1}+\Gamma^{-1})^{-1}\Gamma^{-1}$, it follows that

$$h(z,x)=g_p(x)g_q(z-x)\phi_{\sum}(x)\phi_{\Gamma}(z-x)=g(z,x)\phi_{\sum+\Gamma}(z)\cdot\phi_{\sum^{-1}+\Gamma^{-1}}(x-C_z)$$

is jointly log-concave. Hence it follows that

$$p*q(z)=\int_{\mathbb{R}^d}h(z,x)\mathrm{dx}=\phi_{\sum+\Gamma}(z)\int_{\mathbb{R}^d}g(z,x)\phi_{\sum^{-1}+\Gamma^{-1}}(x-C_z)\mathrm{dx}\equiv\phi_{\sum+\Gamma}(z)g_0(z)$$

where $g_0(z)$ is log-concave by Prékopa's theorem, Theorem 3.3.

c. This is easy since

$$p(x)q(y)=g_p(x)g_q(y)\phi_{\sum}(x)\phi_{\Gamma}(y)=g(x,y)\phi_{\tilde{\sum}}(x,y)$$

where $\tilde{\Sigma}$ is the given $2d\times 2d$ block diagonal matrix and g is jointly log-concave (by Proposition 3.2).

d. Note that

$$p(x)q(x)=g_p(x)g_q(x)\phi_{\sum}(x)\cdot\phi_{\Gamma}(x)\equiv g_0(x)\phi_{(\sum^{-1}+\Gamma^{-1})^{-1}}(x)$$

where g_0 is log-concave.

4. Log-concavity and ultra-log-concavity for discrete distributions

We now consider log-concavity and ultra-log-concavity in the setting of discrete random variables. Some of this material is from Johnson, Kontoyiannis and Madiman [2013] and Johnson [2007].

An integer-valued random variable X with probability mass function $\{p_x : x\in\mathbb{Z}\}$ is *log-concave* if

$$p_x^2\geq p_{x+1}p_{x-1}\text{ for all }x\in\mathbb{Z}. \quad (4.5)$$

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

If we define the *score function* ϕ by $\phi(x) = p_{x+1}/p_x$, then log-concavity of $\{p_x\}$ is equivalent to ϕ being decreasing (nonincreasing).

A stronger notion, analogous to strong log-concavity in the case of continuous random variables, is that of *ultra-log-concavity:* for any $\lambda > 0$ define **ULC**(λ) to be the class of integer-valued random variables X with mean $EX = \lambda$ such that the probability mass function p_x satisfies

$$xp_x^2 \geq (x+1)p_{x+1}p_{x-1} \text{ for all } x \geq 1. \quad (4.6)$$

Then the class of ultra log-concave random variables is **ULC** = $\bigcup_{\lambda>0}$**ULC**(λ). Note that (4.6) is equivalent to log-concavity of $x \to p_x/\pi_{\lambda,x}$ where $\pi_{\lambda,x} = e^{-\lambda}\lambda^x/x!$ is the Poisson distribution on $\mathbb{N}$, and hence ultra-log-concavity corresponds to p being *log-concave relative to* π_λ (or $p \leq_{1c} \pi_\lambda$) in the sense defined by Whitt [1985]. Equivalently, $p_x = h_x\pi_{\lambda,x}$ where h is log-concave. When we want to emphasize that the mass function $\{p_x\}$ corresponds to X, we also write $p_X(x)$ instead of p_x.

If we define the *relative score function* ρ by

$$\rho(x) \equiv \frac{(x+1)p_{x+1}}{\lambda p_x} - 1,$$

then $X \sim p \in$ **ULC**(λ) if and only if ρ is decreasing (nonincreasing). Note that

$$\rho(x) = \frac{(x+1)\varphi(x)}{\lambda} - 1 = \frac{(x+1)\varphi(x)}{\lambda} - \frac{(x+1)\pi_{\lambda,x+1}}{\lambda\pi_{\lambda,x}}.$$

Our main interest here is the preservation of log-concavity and ultra-log-concavity under convolution.

Theorem 4.1. (a) (Keilson and Gerber [1971]) The class of log-concave distributions on $\mathbb{Z}$ is closed under convolution. If $U \sim p$ and $V \sim q$ are independent and p and q are log-concave, then $U + V \sim p \star q$ is log-concave. (b) (Walkup [1976], Liggett [1997]) The class of ultra-log-concave distributions on $\mathbb{Z}$ is closed under convolution. More precisely, these classes are closed under convolution in the following sense: if $U \in$ **ULC**(λ) and $V \in$ **ULC**(μ) are independent, then $U + V \in$ **ULC**($\lambda + \mu$).

Actually, Keilson and Gerber [1971] proved more: analogously to Ibragimov [1956a] they showed that p is strongly unimodal (i.e. $X + Y \sim p \star q$ with X, Y independent is unimodal for every unimodal q on $\mathbb{Z}$) if and only if $X \sim p$ is log-concave. Liggett's proof of (b) proceeds by direct calculation; see also Walkup [1976]. For recent alternative proofs of this property of ultra-log-concave distributions, see Gurvits [2009] and Kahn and Neiman [2011]. A relatively simple proof is given by Johnson [2007] using results from Kontoyiannis, Harremoës and Johnson [2005] and Efron [1965], and that is the proof we

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

will summarize here. See Nayar and Oleszkiewicz [2012] for an application of ultra log-concavity and Theorem 4.1 to finding optimal constants in Khinchine inequalities.

Before proving Theorem 4.1 we need the following lemma giving the score and the relative score of a sum of independent integer-valued random variables.

Lemma 4.2. If X, Y are independent non-negative integer-valued random variables with mass functions $p = p_X$ and $q = p_Y$ then:

a. $\phi_{X+Y}(z) = E\{\phi_X(X)|X+Y=z\}$.

b. If moreover, X and Y have means μ and ν respectively, then with $\alpha = \mu/(\mu+\nu)$,

$$\rho_{X+Y}(z)=E\{\alpha\rho_X(X)+(1-\alpha)\rho_Y(Y)|X+Y=z\}.$$

Proof. For (a), note that with $F_z \equiv p_{X+Y}(z)$ we have

$$\varphi_{X+Y}(z)=\frac{p_{X+Y}(z+1)}{p_{X+Y}(z)}=\sum_x\frac{p(x)q(z+1-x)}{F_z}=\sum_x\frac{p(x)}{p(x-1)}\cdot\frac{p(x-1)q(z+1-x)}{F_z}=\sum_x\frac{p(x+1)}{p(x)}\cdot\frac{p(x)q(z-x)}{F_z}.$$

To prove (b) we follow Kontoyiannis, Harremoës and Johnson [2005], page 471: using the same notation as in (a),

$$\begin{aligned}\rho_{X+Y}(z)=&\frac{(z+1)p_{X+Y}(z+1)}{(\mu+\nu)p_{X+Y}(z)}\\ &-1=\sum_x\frac{(z+1)p(x)q(z+1-x)}{(\mu+\nu)F_z}\mathscr{L}1\\ &=\sum_x\left\{\frac{xp(x)q(z+1-x)}{(\mu+\nu)F_z}\right.\\ &+\frac{(z-x+1)p(x)q(z+1-x)}{(\mu+\nu)F_z}-1=\alpha\left\{\sum_x\frac{xp_X(x)}{\mu p(x-1)}\right.\\ &\cdot\frac{p(x-1)q(z-x+1)}{F_z}\\ &-1+(1-\alpha)\left\{\sum_x\frac{z-x+1}{\nu}\frac{q(z-x+1)}{q(z-x)}\right.\\ &\cdot\frac{p(x)q(z-x)}{F_z}\\ &-1=\sum_x\frac{p(x)q(z-x)}{F_z}\{\alpha\rho_X(x)\\ &+(1-\alpha)\rho_Y(z-x)\}.\end{aligned}$$

Proof. Theorem 4.1: (b) This follows from (b) of Lemma 4.2 and Theorem 1 of Efron [1965], upon noting Efron's remark 1, page 278, concerning the discrete case of his theorem:

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

for independent log-concave random variables X and Y and a measurable function Φ monotone (decreasing here) in each argument, $E\{\Phi(X, Y)|X+Y=z\}$ is a monotone decreasing function of z: note that log-concavity of X and Y implies that

$$\Phi(x,y)=\frac{\mu}{\mu+\nu}\rho_X(x)+\frac{\nu}{\mu+\nu}\rho_Y(y)$$

is a monotone decreasing function of x and y (separately) by since the relative scores ρ_X and ρ_Y are decreasing. Thus ρ_{X+Y} is a decreasing function of z, and hence $X+Y \in \mathbf{ULC}(\mu+\nu)$.

(a) Much as in part (b), this follows from (a) of Lemma 4.2 and Theorem 1 of Efron [1965], upon replacing the relative scores ρ_X and ρ_Y by scores ϕ_X and ϕ_Y and by taking $\Phi(x, y) = \phi_X(x)$.

For interesting results concerning the entropy of discrete random variables, Bernoulli sums, log-concavity, and ultra-log-concavity, see Johnson, Kontoyiannis and Madiman [2013], Ehm [1991], and Johnson [2007]. For recent results concerning nonparametric estimation of a discrete log-concave distribution, see Balabdaoui et al. [2013] and Balabdaoui [2014]. It follows from Ehm [1991] that the hypergeometric distribution (sampling without replacement count of "successes") is equal in distribution to a Bernoulli sum; hence the hypergeometric distribution is ultra-log-concave.

5. Regularity and approximations of log-concave functions

5.1. Regularity

The regularity of a log-concave function $f = \exp(-\phi)$ depends on the regularity of its convex potential ϕ. Consequently, log-concave functions inherit the special regularity properties of convex functions.

Any log-concave function is nonnegative. When the function f is a log-concave *density* (with respect to the Lebesgue measure), which means that f integrates to 1, then it is automatically bounded. More precisely, it has exponentially decreasing tails and hence, it has finite Ψ_1 Orlicz norms; for example, see Borell [1983] and Ledoux [2001]. The following lemma gives a pointwise estimate of the density.

Theorem 5.1 (Cule and Samworth [2010], Lemma 1). Let f be a log-concave density on $\mathbb{R}^d$. Then there exist $a_f = a > 0$ and $b_f = b \in \mathbb{R}$ such that $f(x) \le e^{-a\|x\|+b}$ for all $x \in \mathbb{R}^d$.

Similarly, strong log-concavity implies a finite Ψ_2 Orlicz norm (see Ledoux [2001] Theorem 2.15, page 36, Villani [2003], Theorem 9.9, page 280), Bobkov [1999], and Bobkov and Götze [1999].

For other pointwise bounds on log-concave densities themselves, see Devroye [1984], Dümbgen and Rufibach [2009] and Lovász and Vempala [2007].

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

As noticed in Cule and Samworth [2010], Theorem 5.1 implies that if a random vector X has density f, then the moment generating function of X is finite in an open neighborhood of the origin. Bounds can also be obtained for the supremum of a log-concave density as well as for its values on some special points in the case where $d = 1$.

Proposition 5.2. Let X be a log-concave random variable, with density f on $\mathbb{R}$ and median m. Then

$$\frac{1}{12\mathrm{Var}(X)} \leq f(m)^2 \leq \frac{1}{2\mathrm{Var}(X)}, \quad (5.7)$$

$$\frac{1}{12\mathrm{Var}(X)} \leq \sup_{x\in\mathbb{R}} f(x)^2 \leq \frac{1}{\mathrm{Var}(X)}, \quad (5.8)$$

$$\frac{1}{3e^2\mathrm{Var}(X)} \leq f(\mathbb{E}[X])^2 \leq \frac{1}{\mathrm{Var}(X)}. \quad (5.9)$$

Proposition 5.2 can be found in Bobkov and Ledoux [2014], Proposition B.2. See references therein for historical remarks concerning these inequalities. Proposition 5.2 can also be seen as providing bounds for the variance of a log-concave variable. See Kim and Samworth [2014], section 3.2, for some further results of this type.

Notice that combining (5.7) and (5.8) we obtain the inequality $\sup_{x\in\mathbb{R}} f(x) \leq 2\sqrt{3}\, f(m)$. In fact, the concavity of the function I defined in Proposition 2.21 allows to prove the stronger inequality $\sup_{x\in\mathbb{R}} f(x) < 2f(m)$. Indeed, with the notations of Proposition 2.21, we have $I(1/2) = f(m)$ and for any $x \in (a, b)$, there exists $t \in (0, 1)$ such that $x = F^{-1}(t)$. Hence,

$$2f(m)=2I\left(\frac{1}{2}\right)=2I\left(\frac{t}{2}+\frac{1-t}{2}\right) \geq 2\left(\frac{1}{2}I(t)+\frac{1}{2}I(1-t)\right) \geq I(t)=f(x).$$

A classical result on continuity of convex functions is that any real-valued convex function ϕ defined on an open set $U \subset \mathbb{R}^d$ is locally Lipschitz and in particular, ϕ is continuous on U. For more on continuity of convex functions see Section 3.5 of Niculescu and Persson [2006]. Of course, any continuity of ϕ (local or global) corresponds to the same continuity of f.

For an exposé on differentiability of convex functions, see Niculescu and Persson [2006] (in particular sections 3.8 and 3.11; see also Alberti and Ambrosio [1999] section 7). A deep result of Alexandroff [1939] is the following (we reproduce here Theorem 3.11.2 of Niculescu and Persson [2006]).

Theorem 5.3 (Alexandroff [1939]). Every convex function ϕ on $\mathbb{R}^d$ is twice differentiable almost everywhere in the following sense: f is twice differentiable at a, with Alexandrov Hessian $\nabla^2 f(a)$ in $\mathrm{Sym}^+(d, \mathbb{R})$ (the space of real symmetric $d \times d$ matrices), if ∇f(a) exists, and if for every $\varepsilon > 0$ there exists $\delta > 0$ such that

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

$$\|x-a\|<\delta \text{ implies } \sup_{y\in\partial f(x)} \|y-\nabla f(a)-\nabla^2 f(a)(x-a)\| \leq \varepsilon\|x-a\|.$$

Here $\partial f(x)$ is the subgradient of f at x (see Definition 8.3 in Rockafellar and Wets [1998]). Moreover, if a is such a point, then

$$\lim_{h\to 0}\frac{f(a+h)-f(a)-\langle\nabla f(a),h\rangle-\frac{1}{2}\langle\nabla^2 f(a)h,h\rangle}{\|h\|^2}=0.$$

We immediately see by Theorem 5.3, that since ϕ is convex and $f=\exp(-\phi)$, it follows that f is almost everywhere twice differentiable. For further results in the direction of Alexandrov's theorem see Dudley [1977, 1980].

5.2. Approximations

Again, if one wants to approximate a non-smooth log-concave function $f=\exp(-\phi)$ by a sequence of smooth log-concave functions, then convexity of the potential ϕ can be used to advantage. For an account about approximation of convex functions see Niculescu and Persson [2006], section 3.8.

On the one hand, if $\varphi\in L^1_{\text{loc}}(\mathbb{R}^d)$ the space of locally integrable functions, then the standard use of a regularization kernel (i.e. a one-parameter family of functions associated with a mollifier) to approximate ϕ preserves the convexity as soon as the mollifier is nonnegative. A classical result is that this gives in particular approximations of ϕ in L^p spaces, $p\geq 1$, as soon as $\phi\in L^p(\mathbb{R}^d)$.

On the other hand, *infimal convolution* (also called *epi-addition*, see Rockafellar and Wets [1998]) is a nonlinear analogue of mollification that gives a way to approximate a lower semicontinuous proper convex function from below (section 3.8, Niculescu and Persson [2006]). More precisely, take two proper convex functions f and g from $\mathbb{R}^d$ to $\mathbb{R}\bigcup\{\infty\}$, which means that the functions are convex and finite for at least one point. The infimal convolution between f and g, possibly taking the value $-\infty$, is

$$(f\odot g)(x)=\inf_{y\in\mathbb{R}^n}\{f(x-y)+g(y)\}.$$

Then, $f\odot g$ is a proper convex function as soon as $f\odot g(x)>-\infty$ for all $x\in\mathbb{R}^d$. Now, if f is a lower semicontinuous proper convex function on $\mathbb{R}^d$, the Moreau-Yosida approximation f_ε of f is given by

$$f_\varepsilon(x)=\left(f\odot\frac{1}{2\varepsilon}\|\cdot\|^2\right)(x)=\inf_{y\in\mathbb{R}^n}\left\{f(y)+\frac{1}{2\varepsilon}\|x-y\|^2\right\}.$$

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

for any $x \in \mathbb{R}^d$ and $\varepsilon > 0$. The following theorem can be found in Alberti and Ambrosio [1999] (Proposition 7.13), see also Barbu and Precupanu [1986], Brézis [1973] or Niculescu and Persson [2006].

Theorem 5.4. The Moreau-Yosida approximates f_ε are $\mathscr{C}^{1,1}$ (i.e. differentiable with Lipschitz derivative) convex functions on $\mathbb{R}^d$ and $f_\varepsilon \to f$ as $\varepsilon \to 0$. Moreover, $\partial f_\varepsilon = (\varepsilon I + (\partial f)^{-1})^{-1}$ as set-valued maps.

An interesting consequence of Theorem 5.4 is that if two convex and proper lower semicontinuous functions agree on their subgradients, then they are equal up to a constant (corollary 2.10 in Brézis [1973]).

Approximation by a regularization kernel and Moreau-Yosida approximation have different benefits. While a regularization kernel gives the most differentiability the Moreau-Yosida approximation provides an approximation of a convex function from below (and so, a log-concave function from above). It is thus possible to combine these two kinds of approximations and obtain the advantages of both. For an example of such a combination in the context of a (multivalued) stochastic differential equation and the study of the so-called Kolmogorov operator, see Barbu and Da Prato [2008].

When considering a log-concave random vector, the following simple convolution by Gaussian vectors gives an approximation by log-concave vectors that have $\mathscr{C}^\infty$ densities and finite Fisher information matrices. In the context of Fisher information, regularization by Gaussians was used for instance in Port and Stone [1974] to study the Pitman estimator of a location parameter.

Proposition 5.5 (convolution by Gaussians). Let X be a random vector in $\mathbb{R}^d$ with density p w.r.t. the Lebesgue measure and Gad-dimensional standard Gaussian variable, independent of X. Set $Z = X + \sigma G$, $\sigma > 0$ and $p_Z = \exp(-\phi_Z)$ the density of Z. Then:

- **i.** If X is log-concave, then Z is also log-concave.
- **ii.** If X is strongly log-concave, $Z \in SLC_1\,[\tau^2, d)$ then Z is also strongly log-concave; $Z \in SLC_1\,(\tau^2 + \sigma^2, d)$.
- **iii.** Z has a positive density p_Z on $\mathbb{R}^d$. Furthermore, ϕ_Z is C^∞ on $\mathbb{R}^d$ and

$$\nabla\varphi_Z(z) = \sigma^{-2}\mathbb{E}[\sigma G | X + \sigma G = z] = \mathbb{E}[\rho_{\sigma G}(\sigma G) | X + \sigma G = z], \quad (5.10)$$

  where $\rho_a G\,(x) = \sigma^{-2} x$ is the score of σG.
- **iv.** The Fisher information matrix for location $J(Z) = \mathbb{E}[\nabla\phi_Z \otimes \nabla\phi_Z(Z)]$, is finite and we have $J(Z) \le J(\sigma G) = \sigma^{-4} I_d$ as symmetric matrices.

Proof. See Section 11.

We now give a second approximation tool, that allows to approximate any log-concave density by strongly log-concave densities.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Proposition 5.6. Let f be a log-concave density on $\mathbb{R}^d$. Then for any $c > 0$, the density

$$h_c(x) = \frac{f(x)e^{-c\|x\|^2/2}}{\int_{\mathbb{R}^d} f(v)e^{-c\|v\|^2/2}\mathrm{dv}},\ x \in \mathbb{R}^d,$$

is SLC_1 (c^{-1}, d) and $h_c \to f$ as $c \to 0$ in L_p, $p \in [1, \infty]$. More precisely, there exists a constant $A_f > 0$ depending only on f, such that for any $\varepsilon > 0$,

$$\sup\left\{\sup_{x\in\mathbb{R}^d}|h_c(x) - f(x)|; \left(\int_{\mathbb{R}^d}|h_c(x) - f(x)|^p\mathrm{dx}\right)^{1/p}\right\} \leq A_f c^{1-\varepsilon}.$$

Proof. See Section 11.

Finally, by combining Proposition 5.6 and 5.5, we obtain the following approximation lemma.

Proposition 5.7. For any log-concave density on $\mathbb{R}^d$, there exists a sequence of strongly log-concave densities that are $\mathscr{C}^\infty$, have finite Fisher information matrices and that converge to f in L_p (Leb), $p \in [1, \infty]$.

Proof. Approximate f by a strongly log-concave density h as in Proposition 5.6. Then approximate h by convolving with a Gaussian density. In the two steps the approximations can be as tight as desired in L_p, for any $p \in [1, \infty]$. The fact that the convolution with Gaussians for a (strongly) log-concave density (that thus belongs to any L_p (Leb), $p \in [1, \infty]$) gives approximations in L_p, $p \in [1, \infty]$ is a simple application of general classical theorems about convolution in L_p (see for instance Rudin [1987], p. 148).

6. Efron's theorem and more on preservation of log-concavity and strong log-concavity under convolution in 1-dimension

Another way of proving that strong log-concavity is preserved by convolution in the one-dimensional case is by use of a result of Efron [1965]. This has already been used by Johnson, Kontoyiannis and Madiman [2013] and Johnson [2007] to prove preservation of ultra log-concavity under convolution (for discrete random variables), and by Wellner [2013] to give a proof that strong log-concavity is preserved by convolution in the one-dimensional continuous setting. These proofs operate at the level of scores or relative scores and hence rely on the equivalences between (a) and (b) in Propositions 2.20 and 2.23. Our goal in this section is to re-examine Efron's theorem, briefly revisit the results of Johnson, Kontoyiannis and Madiman [2013] and Wellner [2013], give alternative proofs using second derivative methods via symmetrization arguments, and to provide a new proof of Efron's theorem using some recent results concerning asymmetric Brascamp-Lieb inequalities due to Menz and Otto [2013] and Carlen, Cordero-Erausquin and Lieb [2013].

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

6.1. Efron's monotonicity theorem

The following monotonicity result is due to Efron [1965].

Theorem 6.1 (Efron). Suppose that $\Phi : \mathbb{R}^m \to \mathbb{R}$ where Φ is coordinatewise non-decreasing and let

$$g(z) \equiv E\left\{\Phi(X_1, \cdots, X_m) | \sum_{j=1}^{m} X_j = z\right\},$$

where $X_1, \ldots, X_m$ are independent and log-concave. Then g is non-decreasing.

Remark 6.2. As noted by Efron [1965], Theorem 6.1 continues to hold for integer valued random variables which are log-concave in the sense that $p_x = P(X = x)$ for $x \in \mathbb{Z}$ satisfies $p_x^2 \geq p_{x+1}p_{x-1}$ for all $x \in \mathbb{Z}$.

In what follows, we will focus on Efron's theorem for $m = 2$. As it is shown in Efron [1965], the case of a pair of variables ($m = 2$) indeed implies the general case with $m \geq 2$. Let us recall the argument behind this fact, which involves preservation of log-concavity under convolution. In fact, stability under convolution for log-concave variables is not needed to prove Efron's theorem for $m = 2$ as will be seen from the new proof of Efron's theorem given here in Section 6.4, so it is consistent to prove the preservation of log-concavity under convolution via Efron's theorem for $m = 2$.

Proposition 6.3. If Theorem 6.1 is satisfied for $m = 2$, then it is satisfied for $m \geq 2$.

Proof. We proceed as in Efron [1965] by induction on $m \geq 2$. Let $(X_1, \ldots, X_m)$ be a m–tuple of log-concave variables, let $S = \sum_{i=1}^{m} X_i$ be their sum and set

$$\Lambda(t, u) = \mathbb{E}\left[\Phi(X_1, \cdots, X_m) | \sum_{i=1}^{m-1} X_i = t, X_m = u\right].$$

Then

$$\mathbb{E}[\Phi(X_1, \cdots, X_m) | S = s] = \mathbb{E}[\Lambda(T, X_m) | T + X_m = s],$$

where $T = \sum_{i=1}^{m-1} X_i$. Hence, by the induction hypothesis or functions of two variables, it suffices to prove that Λ is coordinatewise non-decreasing. As T is a log-concave variable (by preservation of log-concavity by convolution), $\Lambda(t, u)$ is non-decreasing in t by the induction hypothesis for functions of $m - 1$ variables. Also $\Lambda(t, u)$ is non-decreasing in u since Φ is non-decreasing in its last argument.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Efron [1965] also gives the following corollaries of Theorem 6.1 above.

Corollary 6.4. Let $\{\Phi_t(x_1, \ldots, x_m) : t \in T\}$ be a family of measurable functions increasing in every argument for each fixed value oft, and increasing in t for each fixed value of $x_1, x_2, \ldots, x_m$. Let $X_1, \ldots, X_m$ be independent and log-concave and write $S \equiv \sum_{j=1}^{m} X_j$. Then

$$g(a,b)=E\left\{\Phi_{a+b-S}(X_1,\cdots,X_m)\middle|a \leq S \leq a+b\right\}$$

is increasing in both a and b.

Corollary 6.5. Suppose that the hypotheses of Theorem 6.1 hold and that $A = \{x = (x_1, \ldots, x_m) \in \mathbb{R}^m : a_j \leq x_j \leq b_j\}$ with $-\infty \leq a_j \leq b_j \leq \infty$ is a rectangle in $\mathbb{R}^m$. Then

$$g(z)=E\left\{\Phi(X_1,\cdots,X_m)\middle|\sum_{j=1}^{m}X_j=z,(X_1,\cdots,X_m)\in A\right\}$$

is a non-decreasing function of z.

The following section will give applications of Efron's theorem to preservation of log-concavity and strong log-concavity in the case of real-valued variables.

6.2. First use of Efron's theorem: strong log-concavity is preserved by convolution via scores

Theorem 6.6. (log-concavity and strong log-concavity preserved by convolution via scores)

(a) (Ibragimov [1956b]) If X and Y are independent and log-concave with densities p and q respectively, then $X + Y \sim p \star q$ is also log-concave.

(b) If $X \in SLC_1(\sigma^2, 1)$ and $Y \in SLC_1(\tau^2, 1)$ are independent, then $X + Y \in SLC_1(\sigma^2 + \tau^2, 1)$

Actually Ibragimov [1956b] proved more: he showed that p is strongly unimodal (i.e. $X + Y \sim p \star q$ with X, Y independent is unimodal for every unimodal q on $\mathbb{R}$) if and only if X is log-concave.

Proof. (a) From Proposition 2.20 log-concavity of p and q is equivalent to monotonicity of their score functions $\varphi_p'=(-\log p)'=-p'/p\ a.e.$ and $\varphi_q'=(-\log q)'=-q'/q\ a.e.$ respectively. From the approximation scheme described in Proposition 5.5 above, we can assume that both p and q are absolutely continuous. Indeed, Efron's theorem applied to formula (5.10) of Proposition 5.5 with $m = 2$ and $\Phi(x, y) = \rho_\sigma G(x)$, gives that the convolution with a Gaussian variable preserves log-concavity Then, from Lemma 3.1 of Johnson and Barron [2004],

$$E\left\{\rho_X(X)\middle|X+Y=z\right\}=\rho_{X+Y}(z).$$

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Thus by Efron's theorem with $m = 2$ and

$$\Phi(x,y)=\rho_Y(y),$$

we see that $E\{\Phi(X,Y)|X+Y=z\}=\varphi'_{p*q}(z)$ is a monotone function of z, and hence by Proposition 2.20, (a) if and only if (b), log-concavity of the convolution $p \star q = p_{X+Y}$ holds.

(b) The proof of preservation of strong log-concavity under convolution for p and q strong log-concave on $\mathbb{R}$ is similar to the proof of (a), but with scores replaced by relative scores, but it is interesting to note that a symmetry argument is needed. From Proposition 2.23 strong log-concavity of p and q is equivalent to monotonicity of their relative score functions $\rho_p(x) \equiv \varphi'_p(x) - x/\sigma^2$ and $\rho_q(x) \equiv \varphi'_q(x) - x/\tau^2$ respectively. Now we take $m = 2$, $\lambda = \sigma^2/(\sigma^2 + \tau^2)$, and define

$$\Phi(x,y)=\lambda\rho_p(x)+(1-\lambda)\rho_q(y).$$

Thus Φ is coordinate-wise monotone and by using Lemma 7.2 with $d = 1$ we find that

$$E\{\Phi(X,Y)|X+Y=z\}=\varphi'_{p*q}(z)-\frac{z}{\sigma^2+\tau^2}=\rho_{p*q}(z).$$

Hence it follows from Efron's theorem that the relative score $\rho_{p\star q}$ of the convolution $p \star q$, is a monotone function of z. By Proposition 2.23(b) it follows that $p \star q \in SLC_1(\sigma^2 + \tau^2, 1)$.

6.3. A special case of Efron's theorem via symmetrization

We now consider a particular case of Efron's theorem. Our motivation is as follows: in order to prove that strong log-concavity is preserved under convolution, recall that we need to show monotonicity in z of

$$\rho_{X+Y}(z)=E\left\{\frac{\sigma^2}{\sigma^2+\tau^2}\rho_X(X)+\frac{\tau^2}{\sigma^2+\tau^2}\rho_Y(Y)\middle|X+Y=z\right\}.$$

Thus we only need to consider functions Φ of the form

$$\Phi(X,Y)=\Psi(X)+\Gamma(Y),$$

where Ψ and Γ are non-decreasing, and show the monotonicity of

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

$$E\left\{\Phi(X,Y)\middle|X+Y=z\right\}.$$

for functions Φ of this special form. By symmetry between X and Y, this reduces to the study of the monotonicity of

$$E\left\{\Psi(X)\middle|X+Y=z\right\}.$$

We now give a simple proof of this monotonicity in dimension 1.

Proposition 6.7. Let $\Psi:\mathbb{R}\to\mathbb{R}$ be non-decreasing and suppose that $X\sim f_X$, $Y\sim f_Y$ are independent and that f_X, f_Y are log-concave. If the function $\eta:\mathbb{R}\to\mathbb{R}$ given by

$$\eta(z)\equiv E\left\{\Psi(X)\middle|X+Y=z\right\}$$

is well-defined (Ψ integrable with respect to the conditional law of $X+Y$), then it is non-decreasing.

Proof. First notice that by truncating the values of Ψ and using the monotone convergence theorem, we assume that Ψ is bounded. Moreover, by Proposition 5.5, we may assume that f_Y is $\mathscr{C}^1$ with finite Fisher information, thus justifying the following computations. We write

$$E\left\{\Psi(X)\middle|X+Y=z\right\}=\int_{\mathbb{R}}\Psi(x)\frac{f_X(x)f_Y(z-x)}{F_z}dx,$$

where

$$F_z=\int_{\mathbb{R}}f_X(x)f_Y(z-x)dx>0.$$

Moreover, with $f_X=\exp(-\phi_X)$ and $f_Y=\exp(-\phi_Y)$,

$$\begin{aligned}\frac{\partial}{\partial z}\left(\Psi(x)\frac{f_X(x)f_Y(z-x)}{F_z}\right)&=-\Psi(x)\varphi_Y'(z-x)\frac{f_X(x)f_Y(z-x)}{F_z}\\&\quad+\Psi(x)\frac{f_X(x)f_Y(z-x)}{F_z}\int_{\mathbb{R}}\varphi_Y'(z-x)\frac{f_X(x)f_Y(z-x)}{F_z}dx,\end{aligned}$$

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

where $\varphi_Y'(y) = -f_Y'(y)/f_Y(y)$. As f_X is bounded (see Section 5.1) and Y has finite Fisher information, we deduce that $\int_{\mathbb{R}} |\varphi_Y'(z-x)| f_X(x) f_Y(z-x)dx$ is finite. Then,

$$\begin{aligned}\frac{\partial}{\partial z}\left(E\left\{\Psi(X)\,|X\right.\right.&\\ &\left.\left.+Y=z\right\}\right)\\ &=-E\left\{\Psi(X)\cdot\varphi_Y'(Y)|X+Y=z\right\}+E\left\{\Psi(X)|X+Y=z\right\}E\left\{\varphi_Y'(Y)|X\right.\\ &\left.+Y=z\right\}=-\operatorname{Cov}\left\{\Psi(X),\varphi_Y'(Y)\right|X+Y=z.\end{aligned}$$

If we show that the latter covariance is negative, the result will follow. Let $(\tilde{X}, \tilde{Y})$ be an independent copy of (X, Y). Then

$$E\left\{\left(\Psi(X)-\Psi\left(\tilde{X}\right)\right)\left(\varphi_Y'(Y)-\varphi_Y'\left(\tilde{Y}\right)\right)\left|\tilde{X}+\tilde{Y}=z, X+Y=z\right.\right\}=2\operatorname{Cov}\left\{\Psi(X),\varphi_Y'(Y)\right|X+Y=z\}.$$

Furthermore, since $X \geq \tilde{X}$ implies $Y \leq \tilde{Y}$ under the given condition $[X+Y=z, \tilde{X}+\tilde{Y}=z]$,

$$\begin{aligned}E\left\{\left(\Psi(X)-\Psi\left(\tilde{X}\right)\right)\left(\varphi_Y'(Y)\right.\right.&\\ \left.-\varphi_Y'\left(\tilde{Y}\right)\right)\left|\tilde{X}\right.&\\ +\tilde{Y}=z, X&\\ +Y=z\}=2E\Bigg\{&\underbrace{\left(\Psi(X)-\Psi\left(\tilde{X}\right)\right)}_{\geq 0}\underbrace{\left(\varphi_Y'(Y)-\varphi_Y'\left(\tilde{Y}\right)\right)}_{\leq 0}1_{\{X\geq\tilde{X}\}}\Bigg|X\\ &+Y=z, \tilde{X}+\tilde{Y}=z\Bigg\}\leq 0.\end{aligned}$$

This proves Proposition 6.7.

6.4. Alternative proof of Efron's theorem via asymmetric Brascamp-Lieb inequalities

Now our goal is to give a new proof of Efron's Theorem 6.1 in the case $m = 2$ using results related to recent asymmetric Brascamp-Lieb inequalities and covariance formulas due to Menz and Otto [2013].

Theorem 6.8 (Efron). Suppose that $\Phi : \mathbb{R}^2 \to \mathbb{R}$, such that Φ is coordinatewise non-decreasing and let

$$g(z) \equiv E\left\{\Phi(X,Y)\middle|X+Y=z\right\},$$

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

where X and Y are independent and log-concave. Then g is non-decreasing.

Proof. Notice that by truncating the values of Φ and using the monotone convergence theorem, we may assume that Φ is bounded. Moreover, by convolving Φ with a positive kernel, we preserve coordinatewise monotonicity of Φ and we may assume that Φ is $\mathscr{C}^1$. As Φ is taken to be bounded, choosing for instance a Gaussian kernel, it is easily seen that we can ensure that $\nabla\Phi$ is uniformly bounded on $\mathbb{R}^2$. Indeed, if

$$\Psi_{\sigma^2}(a,b)=\int_{\mathbb{R}^2}\Phi(x,y)\frac{1}{2\pi\sigma^2}e^{-\|(a,b)-(x,y)\|^2/2\sigma^2}dxdy,$$

then

$$\nabla\Psi_{\sigma^2}(a,b)=-\int_{\mathbb{R}^2}\Phi(x,y)\frac{\|(a,b)-(x,y)\|}{2\pi\sigma^4}e^{-\|(a,b)-(x,y)\|^2/2\sigma^2}dxdy,$$

which is uniformly bounded in (a, b) whenever Φ is bounded. Notice also that by Lemma 5.7, it suffices to prove the result for strictly (or strongly) log-concave variables that have $\mathscr{C}^\infty$ densities and finite Fisher information. We write

$$N(z)=\int_{\mathbb{R}}f_X(z-y)f_Y(y)dy$$

and

$$g(z)=\int_{\mathbb{R}}\Phi(z-y,y)\frac{f_X(z-y)f_Y(y)}{N(z)}dy,$$

with $f_X = \exp(-\phi_X)$ and $f_Y = \exp(-\phi_Y)$ the respective strictly log-concave densities of X and Y. We note μ_X and μ_Y respectively the distributions of X and Y. Since $\varphi_X^{'}$ is L_2 (μ_X) (which means that μ_X has finite Fisher information) and f_Y is bounded (see Theorem 5.1), we get that $f_X^{'}(z-y)f_Y(y)=-\varphi_X^{'}(z-y)f_X(z-y)f_Y(y)$ is integrable and so N is differentiable with gradient given by

$$N^{'}(z)=-\int_{\mathbb{R}}\varphi_X^{'}(z-y)f_X(z-y)f_Y(Y)dy.$$

By differentiating with respect to z inside the integral defining g we get

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

$$\begin{aligned}
&\frac{d}{dz}\left(\Phi(z-y,y)\frac{f_X(z-y)f_Y(y)}{\int_{\mathbb{R}} f_X(z-y')f_Y(y')dy'}\right)\\
&=(\partial_1\Phi)(z\\
&-y,y)\frac{f_X(z-y)f_Y(y)}{N(z)}-\Phi(z\\
&-y,y)\varphi_X^{'}(z\\
&\quad -y)\frac{f_X(z-y)f_Y(y)}{N(z)}\\
&+\Phi(z\\
&-y,y)\frac{f_X(z-y)f_Y(y)}{N(z)}\int_{\mathbb{R}}\varphi_X^{'}(z\\
&\quad -y)\frac{f_X(z-y)f_Y(y)}{N(z)}dy.
\end{aligned} \tag{6.11}$$

We thus see that the quantity in (6.11) is integrable (with respect to Lebesgue measure) and we get

$$g^{'}(z)=\mathbb{E}[(\partial_1\Phi)(X,Y)|X+Y=z]-\mathrm{Cov}[\Phi(X,Y),\varphi_X^{'}(X)|X+Y=z]. \tag{6.12}$$

Now, by symmetrization we have

$$\begin{aligned}
&\mathrm{Cov}[\Phi(X,Y),\varphi_X^{'}(X)|X+Y=z]=\mathbb{E}\left[\left(\Phi(X,Y)-\Phi\left(\tilde{X},\tilde{Y}\right)\right)\left(\varphi_X^{'}(X)-\varphi_X^{'}\left(\tilde{X}\right)\right)1_{\{X\geq\tilde{X}\}}|X+Y=z,\tilde{X}+\tilde{Y}=z\right]\\
&=\mathbb{E}\left[\left(\int_{\tilde{X}}^{X}(\partial_1\Phi-\partial_2\Phi)(u,z-u)du\right)\underbrace{\left(\varphi_X^{'}(X)-\varphi_X^{'}\left(\tilde{X}\right)\right)}_{\geq 0}1_{\{X\geq\tilde{X}\}}\middle|X+Y=z,\tilde{X}+\tilde{Y}=z\right.\\
&\leq\mathbb{E}\left[\left(\int_{\tilde{X}}^{X}(\partial_1\Phi)(u,z-u)du\right)\left(\varphi_X^{'}(X)-\varphi_X^{'}\left(\tilde{X}\right)\right)1_{\{X\geq\tilde{X}\}}|X+Y=z,\tilde{X}+\tilde{Y}=z\right]=\mathrm{Cov}[\Phi_1(X),\varphi_X^{'}(X)|X+Y=z],
\end{aligned}$$

where $\Phi_1(x)=\int_0^x(\partial_1\Phi)(u,z-u)du$. We denote η the distribution of X given $X+Y=z$. The measure η has density $h_z(x)=N^{-1}(z)\,f_X(x)\,f_Y(z-x)$, $y\in\mathbb{R}$. Notice that h_z is strictly log-concave on $\mathbb{R}$ and that for all $x\in\mathbb{R}$,

$$(-\log h_z)^{''}(x)=\varphi_X^{''}(x)+\varphi_Y^{''}(z-x).$$

Now we are able to use the asymmetric Brascamp and Lieb inequality of Menz and Otto [2013] (Lemma 2.11, page 2190, with their $\delta\psi\equiv 0$ so their $\psi=\psi_c$ with $\psi''>0$) or Carlen, Cordero-Erausquin and Lieb [2013] ((1.2), page 2); see Proposition 10.3 below. This yields

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

$$\begin{aligned}\mathrm{Cov}[\Phi_1(X),\varphi_X'(X)|X+Y=z] &= \int_{\mathbb{R}}(\Phi_1(X)-\mathbb{E}[\Phi_1(X,Y)|X+Y=z])(\varphi_X'(x)-\mathbb{E}[\varphi_X'(X)|X+Y=z])h_z(x)dx \\ &\le \sup_{x\in\mathbb{R}}\left\{\frac{\varphi_X''(x)}{(-\log h_z)''(x)}\right\}\int_{\mathbb{R}}\Phi_1'(x)h_z(x)dx \\ &= \sup_{x\in\mathbb{R}}\left\{\frac{\varphi_X''(x)}{\varphi_X''(x)+\varphi_Y''(z-x)}\right\}\mathbb{E}[(\partial_1\Phi)(X,Y)|X+Y=z] \\ &\le \mathbb{E}[(\partial_1\Phi)(X,Y)|X+Y=z].\end{aligned}$$

Using the latter bound in (6.12) then gives the result.

7. Preservation of log-concavity and strong log-concavity under convolution in $\mathbb{R}^d$ via Brascamp-Lieb inequalities and towards a proof via scores

In Sections 6 and 4, we used Efron's monotonicity theorem 6.1 to give alternative proofs of the preservation of log-concavity and strong log-concavity under convolution in the cases of continuous or discrete random variables on $\mathbb{R}$ or $\mathbb{Z}$ respectively. In the former case, we also used asymmetric Brascamp-Lieb inequalities to give a new proof of Efron's monotonicity theorem. In this section we look at preservation of log-concavity and strong log-concavity under convolution in $\mathbb{R}^d$ via:

- **a.** the variance inequality due to Brascamp and Lieb [1976];
- **b.** scores and potential (partial) generalizations of Efron's monotonicity theorem to $\mathbb{R}^d$.

While point (a) gives a complete answer (Section 7.1), the aim of point (b) is to give an interesting link between preservation of (strong) log-concavity in $\mathbb{R}^d$ and a (guessed) monotonicity property in $\mathbb{R}^d$ (Section 7.2). This latter property would be a partial generalization of Efron's monotonicity theorem to the multidimensional case and further investigations remain to be accomplished in order to prove such a result.

We refer to Section 10 (Appendix A) for further comments about the Brascamp-Lieb inequalities and related issues, as well as a recall of various functional inequalities.

7.1. Strong log-concavity is preserved by convolution (again): proof via second derivatives and a Brascamp-Lieb inequality

We begin with a different proof of the version of Theorem 3.7(b) corresponding to our first definition of strong log-concavity, Definition 2.8, which proceeds via the Brascamp-Lieb variance inequality as given in part (a) of Proposition 10.1:

Proposition 7.1. If $X \sim p \in SLC_1(\sigma^2, d)$ and $Y \sim q \in SLC_1(\tau^2, d)$ are independent, then $Z = X+Y \sim p\star q \in SLC_1(\sigma^2+\tau^2, d)$.

Proof. Now $p_Z = p_{X+Y} = p\star q$ is given by

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

$$p_Z(z)=\int p(x)q(z-x)dx=\int p(z-y)q(y)dy. \quad (7.13)$$

Now $p = \exp(-\phi_p)$ and $q = \exp(-\phi_p)$ where we may assume (by (b) of Proposition 5.5) that $\phi_p, \phi_q \in C^2$ and that p and q have finite Fisher information. Then, by Proposition 2.23,

$$\nabla^2(\varphi_p)(x) \geq \frac{1}{\sigma^2}I, \text{ and } \nabla^2(\varphi_q)(x) \geq \frac{1}{\tau^2}I.$$

As we can interchange differentiation and integration in (7.13) (see for instance the detailed arguments for a similar situation in the proof of Proposition 6.7), we find that

$$\nabla(-\log p_Z)(z)=-\frac{\nabla p_Z}{p_Z}(z)=E\{\nabla\varphi_q(Y)|X+Y=z\}=E\{\nabla\varphi_p(X)|X+Y=z\}.$$

Then

$$\begin{aligned}
&\nabla^2(-\log p_Z)(z)\\
&\quad=\nabla\,\{E[q(z-X)\nabla(-\log q)(z-X)]\\
&\quad\cdot\frac{1}{p_Z(z)}=E\{-\nabla_q(Y)(\nabla\log q(Y))^T|X\\
&\qquad+Y=z\}+E\{\nabla^2(-\log q)(Y)|X+Y=z\}+(E\{\nabla\log q(Y)|X+Y=z\})^{\otimes 2}\\
&\quad=-\operatorname{Var}(\nabla\varphi_q(Y)|X\\
&\qquad+Y=z)+E\{\nabla^2\varphi_q(Y)|X\\
&\qquad+Y=z\}=-\operatorname{Var}(\nabla\varphi_q(X)|X\\
&\qquad+Y=z)+E\{\nabla^2\varphi_q(X)|X+Y=z\}.
\end{aligned}$$

Now we apply Brascamp and Lieb [1976] Theorem 4.1 (see Proposition 10.1(a)) with

$$h(x)=\nabla_z\varphi_q(z-x), \quad (7.14)$$

$$F(x)=p(x)q(z-x), \quad (7.15)$$

to obtain

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

$$\mathrm{Var}(\nabla_z\varphi_q(Y)|Z+Y=z) \le \int_{\mathbb{R}^d}\nabla^2\varphi_g(z-x)\{\nabla^2\varphi_p(x)+\nabla^2\varphi_q(z-x)\}^{-1}\cdot\nabla^2\varphi_q(z-x)\frac{F(x)}{\int_{\mathbb{R}^d}F(x')dx'}dx. \quad (7.13)$$

This in turn yields

$$\nabla^2(-\log p_Z)(z) \ge E\left\{\nabla^2\varphi_q(Y)-\nabla^2\varphi_q(Y)[\nabla^2\varphi_p(X)+\nabla^2\varphi_q(Y)]^{-1}\nabla^2\varphi_q(Y)|X+Y=z\right\}. \quad (7.16)$$

By symmetry between X and Y we also have

$$\nabla^2(-\log p_Z)(z) \ge E\left\{\nabla^2\varphi_p(X)-\nabla^2\varphi_p(X)[\nabla^2\varphi_p(X)+\nabla^2\varphi_q(Y)]^{-1}\nabla^2\varphi_p(X)|X+Y=z\right\}. \quad (7.17)$$

In proving the inequalities in the last two displays we have in fact reproved Theorem 4.2 of Brascamp and Lieb [1976] in our special case given by (7.15). Indeed, Inequality (4.7) of Theorem 4.2 in Brascamp and Lieb [1976] applied to our special case is the first of the two inequalities displayed above.

Now we combine (7.16) and (7.17). We set

$$\begin{gathered}\alpha \equiv \tfrac{\sigma^2}{\sigma^2+\tau^2},\ \beta \equiv 1-\alpha=\tfrac{\tau^2}{\sigma^2+\tau^2},\\ A \equiv [\nabla^2\varphi_p(X)+\nabla^2\varphi_q(Y)]^{-1},\\ s=s(X) \equiv \nabla^2\varphi_p(X),\ t=t(X) \equiv \nabla^2\varphi_q(Y).\end{gathered}$$

We get from (7.16) and (7.17):

$$\begin{aligned}\nabla^2(-\log p_Z)(z) \ge E\{&\alpha s\\ &+\beta t-\alpha sAs\\ &-\beta tAt|X\\ &+Y=z\}=E\{(\alpha s+\beta t)A(s+t)-\alpha sAs-\beta tAt|X+Y=z\} \text{ since } A(s\\ &+t)=I \equiv \text{identity}\\ =E\{&\alpha sAt\\ &+\beta tAs|X+Y=z\}.\end{aligned}$$

Now

$$\alpha sAt=\frac{\sigma^2}{\sigma^2+\tau^2}\nabla^2\varphi_p[\nabla^2\varphi_p(X)+\nabla^2\varphi_q(Y)]^{-1}\nabla^2\varphi_q(Y)=\frac{\sigma^2}{\sigma^2+\tau^2}[(\nabla^2\varphi_p)^{-1}(X)+(\nabla^2\varphi_q)^{-1}(Y)]^{-1}.$$

By symmetry

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

$$\beta tAs = \frac{\tau^2}{\sigma^2+\tau^2}\left[(\nabla^2\varphi_p)^{-1}(X)+(\nabla^2\varphi_q)^{-1}(Y)\right]^{-1}$$

and we therefore conclude that

$$\nabla^2(-\log p_Z)(z) \geq \frac{\sigma^2+\tau^2}{\sigma^2+\tau^2}E\left\{\left[(\nabla^2\varphi_p)^{-1}(X)+(\nabla^2\varphi_q)^{-1}(Y)\right]^{-1}|X+Y=z\right\} \geq \frac{1}{\sigma^2+\tau^2}I.$$

Note that the resulting inequality

$$\nabla^2(-\log p_Z)(z) \geq E\left\{\left[(\nabla^2\varphi_p)^{-1}(X)+(\nabla^2\varphi_q)^{-1}(Y)\right]^{-1}|X+Y=z\right\}$$

also gives the right lower bound for convolution of strongly log-concave densities in the definition of $SLC_2(\mu, \Sigma, d)$, namely

$$\nabla^2(-\log p_Z)(z) \geq \left(\sum\nolimits_X + \sum\nolimits_Y\right)^{-1}.$$

7.2. Strong log-concavity is preserved by convolution (again): towards a proof via scores and a multivariate Efron inequality

We saw in the previous sections that Efron's monotonicity theorem allows to prove stability under convolution for (strongly) log-concave measures on $\mathbb{R}$. However, the stability holds also in $\mathbb{R}^d$, $d > 1$. This gives rise to the two following natural questions: does a generalization of Efron's theorem in higher dimensions exist? Does it allow recovery stability under convolution for log-concave measures in $\mathbb{R}^d$?

Let us begin with a projection formula for scores in dimension d.

Lemma 7.2. (Projection) Suppose that X and Y are d–dimensional independent random vectors with log-concave densities p_X and q_Y respectively on $\mathbb{R}^d$. Then $\nabla\phi_{X+Y}$ and ρ_{X+Y}: $\mathbb{R}^d \to \mathbb{R}^d$ are given by

$$\nabla\varphi_{X+Y}(z) = E\{\lambda\nabla\varphi_X(X)+(1-\lambda)\nabla\varphi_Y(Y)|X+Y=z\}$$

for each $\lambda \in [0, 1]$, and, if $p_X \in SLC_1(\sigma^2, d)$ and $p_Y \in SLC_1(\tau^2, d)$, then

$$\rho_{X+Y}(z) = E\left\{\frac{\sigma^2}{\sigma^2+\tau^2}\rho_X(X)+\frac{\tau^2}{\sigma^2+\tau^2}\rho_Y(Y)\big|X+Y=z\right\}.$$

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Proof. This can be proved just as in the one-dimensional case, much as in Brown [1982], but proceeding coordinate by coordinate.

Since we know from Propositions 2.20 and 2.23 that the scores $\nabla\phi_X$ and $\nabla\phi_Y$ and the relative scores ρ_X and ρ_Y are multivariate monotone, the projection Lemma 7.2 suggests that proofs of preservation of multivariate log-concavity and strong log-concavity might be possible via a multivariate generalization of Efron's monotonicity Theorem 6.1 to $d \geq 2$ along the following lines: Suppose that $\Phi : (\mathbb{R}^d)^n \to \mathbb{R}^d$ is coordinatewise multivariate monotone: for each fixed $j \in \{1,\ldots, n\}$ the function $\Phi_j : \mathbb{R}^d \to \mathbb{R}^d$ defined by

$$\Phi_j(x) = \Phi(x_1, \ldots, x_{j-1}, x, x_{j+1}, \ldots, x_n)$$

is multivariate monotone: that is

$$\langle \Phi_j(x_1) - \Phi_j(x_2), x_1 - x_2 \rangle \geq 0 \text{ for } \text{ all } x_1, x_2 \in \mathbb{R}^d.$$

If $X_1,\ldots, X_m$ are independent with $X_j \sim f_j$ log-concave on $\mathbb{R}^d$, then it might seem natural to conjecture that the function g defined by

$$g(z) \equiv E\left\{\Phi(X_1, \ldots, X_n) \middle| X_1 + \cdots + X_m = z\right\}$$

is a monotone function of $z \in \mathbb{R}^d$:

$$\langle g(z_1) - g(z_2), z_1 - z_2 \rangle \geq 0 \text{ for } \text{ all } z_1, z_2 \in \mathbb{R}^d.$$

Unfortunately, this seemingly natural generalization of Efron's theorem does not hold without further assumptions. In fact, it fails for $m = 2$ and $\mathbf{X}_1$, $\mathbf{X}_2$ Gaussian with covariances Σ_1 and Σ_2 sufficiently different. For an explicit example see Saumard and Wellner [2014].

Again, the result holds for m random vectors if it holds for $2,\ldots, m-1$ random vectors. It suffices to prove the theorem for $m = 2$ random vectors. Since everything reduces to the case where Φ is a function of two variables (either for Efron's theorem or for a multivariate generalization), we will restrict ourselves to this situation.

Thus if we define

$$g(s) \equiv E\left\{\Phi(X_1, X_2) \middle| X_1 + X_2 = s\right\},$$

then we want to show that

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

$$\langle g(s_1) - g(s_2), s_1 - s_2 \rangle \geq 0 \text{ for } \text{ all } s_1, s_2 \in \mathbb{R}^d.$$

Finally, our approach to Efron's monotonicity theorem in dimension $d \geq 2$ is based on the following remark.

Remark 7.3. For suitable regularity of $\Phi : (\mathbb{R}^d)^2 \to \mathbb{R}^d$ and $\rho_X : \mathbb{R}^d \to \mathbb{R}$, we have

$$(\nabla g)(z) = E\left\{(\nabla_1 \Phi)(X,Y) \,|X+Y=z\right\} - \mathrm{Cov}\left\{\Phi(X,Y), \rho_X(X) |\, X+Y=z\right\} (\in \mathbb{R}^{d\times d}).$$

Recall that $\nabla_1 \Phi \equiv \nabla \Phi_1 : (\mathbb{R}^d)^2 \to \mathbb{R}^{d\times d}$. Furthermore, the matrix $(\nabla g)(z)$ is positive semi-definite if for all $a \in \mathbb{R}^d$, $a^T \nabla g(z) a^T \geq 0$, which leads to leads to asking if the following covariance inequality holds:

$$\mathrm{Cov}\left\{a^T \Phi(X,Y), \rho_X^T(X) a \,|X+Y=z\right\} \leq E\left\{a^T \nabla_1 \Phi(X,Y) a \middle| \, X+Y=z\right\} ? \quad (7.18)$$

Covariance inequality (7.18) would imply a multivariate generalization of Efron's theorem (under sufficient regularity).

8. Peakedness and log-concavity

Here is a summary of the results of Proschan [1965], Olkin and Tong [1988], Hargé [2004], and Kelly [1989].

First Hargé [2004]. Let f be log-concave, and let g be convex. Then if $X \sim N_d(\mu, \Sigma) \equiv \gamma$,

$$E\{g(X+\mu-\nu) f(X)\} \leq Ef(X) \cdot Eg(X) \quad (8.19)$$

where $\mu = E(X)$, $\nu = E(Xf(X))/E(f(X))$. Assuming that $f \geq 0$, and writing $f d\tilde{\gamma} \equiv f d\gamma / \int f d\gamma$, $g(\tilde{x} - \mu) \equiv g(x)$, and $\tilde{X} \sim f d\tilde{\gamma}$ so that $\tilde{X}$ is strongly log-concave, this can be rewritten as

$$E\tilde{g}(\tilde{X} - E(\tilde{X})) \leq E\tilde{g}(X - \mu). \quad (8.20)$$

In particular, for $g(\tilde{x}) = |x|^r$ with $r \geq 1$,

$$E|\tilde{X} - \tilde{\mu}|^r \leq E|X - \mu|^r,$$

and for $g(\tilde{x}) = |a^T x|^r$ with $a \in \mathbb{R}^d$, $r \geq 1$,

$$E|a^T(\tilde{X} - \tilde{\mu})|^r \leq E|a^T(X - \mu)|^r,$$

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

which is Theorem 5.1 of Brascamp and Lieb [1976]. Writing (8.19) as (8.20) makes it seem more related to the "peakedness" results of Olkin and Tong [1988] to which we now turn.

An n-dimensional random vector Y is said to be *more peaked* than a vector X if they have densities and if

$$P(Y \in A) \geq P(X \in A)$$

holds for all $A \in \mathscr{A}_n$, the class of compact, convex, symmetric (about the origin) Borel sets in $\mathbb{R}^n$. When this holds we will write $Y \overset{p}{\geq} X$. A vector a majorizes the vector b (and we write $a > b$) if $\sum_{i=1}^{k} b_{[i]} \leq \sum_{i=1}^{k} a_{[i]}$ for $k = 1,\ldots, n-1$ and $\sum_{i=1}^{n} b_{[i]} = \sum_{i=1}^{n} a_{[i]}$ where $a_{[1]} \geq a_{[2]} \geq \cdots \geq a_{[n]}$ and similarly for b. (In particular $b = (1,\ldots,1)/n < (1,0,\ldots,0) = a$.)

Proposition 8.1. (Sherman, 1955; see Olkin and Tong [1988]) Suppose that f_1, f_2, g_1, g_2 are log-concave densities on $\mathbb{R}^n$ which are symmetric about 0. Suppose that $X_j \sim f_j$ and $Y_j \sim g_j$ for $j = 1, 2$ are independent. Suppose that $Y_1 \overset{p}{\geq} X_1$ and $Y_2 \overset{p}{\geq} X_2$. Then $Y_1+Y_2 \overset{p}{\geq} X_1+X_2$.

Proposition 8.2. If $X_1,\ldots X_n$ are independent random variables with log-concave densities symmetric about 0, and $Y_1,\ldots, Y_n$ are independent with log-concave densities symmetric about 0, and $Y_j \overset{p}{>} X_j$ for $j = 1,\ldots, n$, then

$$\sum_{j=1}^{n} c_j Y_j \overset{p}{\geq} \sum_{j=1}^{n} c_j X_j$$

for all real numbers $\{c_j\}$.

Proposition 8.3. If $\{X_m\}$ and $\{Y_m\}$ are two sequences of n–dimensional random vectors with $Y_m \overset{p}{>} X_m$ for each m and $X_m \to_d X$, $Y_m \to_d Y$, then $Y \overset{p}{>} X$.

Proposition 8.4. $Y \overset{p}{\geq} X$ if and only if $CY \overset{p}{\geq} CX$ for all $k \times n$ matrices C with $k \leq n$.

Proposition 8.5. (Proschan [1965]) Suppose that $Z_1,\ldots, Z_n$ are i.i.d. random variables with log-concave density symmetric about zero. Then if $a, b \in \mathbb{R}_+^n$ with $a > b$ (a majorizes b), then

$$\sum_{j=1}^{n} b_j Z_j \overset{p}{\geq} \sum_{j=1}^{n} a_j Z_j \text{ in } \mathbb{R}$$

Proposition 8.6. (Olkin and Tong [1988]) Suppose that $Z_1,\ldots, Z_n$ are i.i.d. d–dimensional random vectors with log-concave density symmetric about zero. Then if a_j, $b_j \in \mathbb{R}^1$ with $a > b$ (a majorizes b), then

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

$$\sum_{j=1}^{n} b_j Z_j \overset{p}{>} \sum_{j=1}^{n} a_j Z_j \text{ in } \mathbb{R}^d.$$

Now let $\mathscr{K}_n \equiv \{x \in \mathbb{R}^n : x_1 \le x_2 \le \ldots \le x_n\}$. For any $y \in \mathbb{R}^n$, let $\hat{y} = (\hat{y}_1, \ldots, \hat{y}_n)$ denote the projection of y onto $\mathscr{K}_n$. Thus $|y - \hat{y}|^2 = \min_{x \in \mathscr{K}} |y - x|^2$.

Proposition 8.7. (Kelly [1989]). Suppose that $\underline{Y} = (Y_1, \ldots, Y_n)$ where $Y_j \sim N(\mu_j, \sigma^2)$ are independent and $\mu_1 \le \mu_2 \le \ldots \le \mu_n$. Thus $\underline{\mu} \in \mathscr{K}_n$ and $\underline{\mu} \hat{\equiv} \hat{\underline{Y}} \in \mathscr{K}_n$. Then $\hat{\mu}_k - \mu_k \overset{p}{\ge} Y_k - \mu_k$ for each $k \in \{1, \ldots, n\}$; i.e.

$$P(|\hat{\mu}_k - \mu_k| \le t) \ge P(|Y_k - \mu_k| \le t) \text{ for all } t > 0, k \in \{1, \ldots, n\}.$$

9. Some open problems and further connections with log-concavity

9.1. Two questions

Question 1: Does Kelly's Proposition 8.7 continue to hold if the normal distributions of the Y_i's is replaced some other centrally-symmetric log-concave distribution, for example Chernoff's distribution (see Balabdaoui and Wellner [2014])?

Question 2: Balabdaoui and Wellner [2014] show that Chernoff's distribution is log-concave. Is it strongly log-concave? A proof would probably give a way of proving strong log-concavity for a large class of functions of the form $f(x) = g(x)g(-x)$ where $g \in PF_\infty$ is the density of the sum $\sum_{j=1}^{\infty}(Y_j - \mu_j)$ where Y_j's are independent exponential random variables with means μ_j satisfying $\sum_{j=1}^{\infty} \mu_j = \infty$ and $\sum_{j=1}^{\infty} \mu_j^2 <$.

9.2. Cross-connections with the families of hyperbolically monotone densities

A theory of hyperbolically monotone and completely monotone densities has been developed by Bondesson [1992], Bondesson [1997].

Definition 9.1. A density f on $\mathbb{R}^+$ is hyperbolically completely monotone if $H(w) = f(uv)f(u/v)$ is a completely monotone function of $w = (v + 1/v)/2$. A density f on $\mathbb{R}^+$ is hyperbolically monotone of order k, or $f \in HM_k$ if the function H satisfies $(-1)^j H^{(j)}(w) \ge 0$ for $j = 0, \ldots, k-1$ and $(-1)^{k-1} H^{(k-1)}(w)$ is right-continuous and decreasing.

For example, the exponential density $f(x) = e^{-x} 1_{(0,\infty)}(x)$ is hyperbolically completely monotone, while the half-normal density $f(x) = \sqrt{2/\pi} \exp(-x^2/2) 1_{(0,\infty)(x)}$ is HM_1 but not HM_2.

Bondesson [1997] page 305 shows that if $X \sim f \in HM_1$, then $\log X \sim e^x f(e^x)$ is log-concave. Thus HM_1 is closed under the formation of products: if $X_1, \ldots, X_n \in HM_1$, then $Y \equiv X_1 \cdots X_n \in HM_1$.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

9.3. Suprema of Gaussian processes

Gaenssler, Molnár and Rost [2007] use log-concavity of Gaussian measures to show that the supremum of an arbitrary non degenerate Gaussian process has a continuous and strictly increasing distribution function. This is useful for bootstrap theory in statistics. The methods used by Gaenssler, Molnár and Rost [2007] originate in Borell [1974] and Ehrhard [1983]; see Bogachev [1998] chapters 1 and 4 for an exposition.

Furthermore, in relation to Example 2.14 above, one can wonder what is the form of the density of the maximum of a Gaussian process in general? Bobkov [2008] actually gives a complete characterization of the distribution of suprema of Gaussian processes. Indeed, the author proves that F is the distribution of the supremum of a general Gaussian process if and only if $\Phi^{-1}(F)$ is concave, where Φ^{-1} is the inverse of the standard normal distribution function on the real line. Interestingly, the "only if" part is a direct consequence the Brunn-Minkowski type inequality for the standard Gaussian measure γ_d on $\mathbb{R}^d$ due to Ehrhard [1983]: for any A and $B \in \mathbb{R}^d$ of positive measure and for all $\lambda \in (0, 1)$,

$$\Phi^{-1}(\gamma_d(\lambda A+(1-\lambda)B)) \geq \lambda\Phi^{-1}(\gamma_d(A))+(1-\lambda)\Phi^{-1}(\gamma_d(B)).$$

9.4. Gaussian correlation conjecture

The Gaussian correlation conjecture, first stated by Das Gupta et al. [1972], is as follows. Let A and B be two symmetric convex sets. If μ is a centered, Gaussian measure on $\mathbb{R}^n$, then

$$\mu(A \cap B) \geq \mu(A)\mu(B). \quad (9.21)$$

In other words, the correlation between the sets A and B under the Gaussian measure μ is conjectured to be nonnegative. As the indicator of a convex set is log-concave, the Gaussian correlation conjecture intimately related to log-concavity.

In Hargé [1999], the author gives an elegant partial answer to Problem (9.21), using semigroup techniques. The Gaussian correlation conjecture has indeed been proved to hold when $d = 2$ by Pitt [1977] and by Hargé [1999] when one of the sets is a symmetric ellipsoid and the other is convex symmetric. Cordero-Erausquin [2002] gave another proof of Hargé's result, as a consequence of Caffarelli's Contraction Theorem (for more on the latter theorem, see Section 9.7 below). Extending Caffarelli's Contraction Theorem, Kim and Milman [2012] also extended the result of Hargé and Cordero-Erausquin, but without proving the full Gaussian correlation conjecture.

Hargé [1999] gives some hints towards a complete solution of Problem (9.21). Interestingly, a sufficient property would be the preservation of log-concavity along a particular family of semigroups. More precisely, let $A(x)$ be a positive definite matrix for each $x \in \mathbb{R}^d$ and define

$$Lf(x)=(1/2)(\mathrm{div}(A(x)^{-1}\nabla f) - (\nabla f(x)^T A^{-1}(x)x).$$

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

The operator L is the infinitesimal generator of an associated semigroup. The question is: does L preserve log-concavity? See Hargé [1999] and Kolesnikov [2001]. For further connections involving the semi-group approach to correlation inequalities, see Bakry [1994], Ledoux [1995], Hargé [2008], and Cattiaux and Guillin [2013].

Further connections in this direction involve the theory of parabolic and heat-type partial differential equations; see e.g. Keady [1990], Kolesnikov [2001], Andreu, Caselles and Mazón [2008], Korevaar [1983a], Korevaar [1983b].

9.5. Further connections with Poincaré, Sobolev, and log-Sobolev inequalities

For a very nice paper with interesting historical and expository passages, see Bobkov and Ledoux [2000]. Among other things, these authors establish an entropic or log-Sobolev version of the Brascamp-Lieb type inequality under a concavity assumption on $h^T\phi''(x)h$ for every h. The methods in the latter paper build on Maurey [1991]. See Bakry, Gentil and Ledoux [2014] for a general introduction to these analytic inequalities from a Markov diffusion operator viewpoint.

9.6. Further connections with entropic central limit theorems

This subject has its beginnings in the work of Linnik [1959], Brown [1982], and Barron [1986], but has interesting cross-connections with log-concavity in the more recent papers of Johnson and Barron [2004], Carlen and Soffer [1991], Ball, Barthe and Naor [2003], Artstein et al. [2004a], and Artstein et al. [2004b]. More recently still, further results have been obtained by: Carlen, Lieb and Loss [2004], Carlen and Cordero-Erausquin [2009], and Cordero-Erausquin and Ledoux [2010].

9.7. Connections with optimal transport and Caffarelli's contraction theorem

Gozlan and Léonard [2010] give a nice survey about advances in transport inequalities, with Section 7 devoted to strongly log-concave measures (called measures with “uniform convex potentials” there). The theory of optimal transport is developed in Villani [2003] and Villani [2009]. See also Caffarelli [1991], Caffarelli [1992], Caffarelli [2000], and Kim and Milman [2012] for results on (strongly) log-concave measures. The latter authors extend the results of Caffarelli [2000] under a third derivative hypothesis on the “potential” ϕ.

In the following, we state the celebrated Caffarelli's Contraction Theorem (Caffarelli [2000]). Let us recall some related notions. A Borel map T is said to *push-forward* μ onto ν, for two Borel probability measures μ and ν, denoted $T_*(\mu) = \nu$, if for all Borel sets A, $\nu(A) = \mu(T^{-1}(A))$. Then the Monge-Kantorovich problem (with respect to the quadratic cost) is to find a map T_{opt} such that

$$T_{\mathrm{opt}} \in \arg \min_{T \text{ s.t.} T_*(\mu)=v} \left\{ \int_{\mathbb{R}^d} |T(x) - x|^2 d\mu(x) \right\}.$$

The map T_{opt} (when it exists) is called the Brenier map and it is μ-a.e. unique. Moreover, Brenier [1991] showed that Brenier maps are characterized to be gradients of convex

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

functions (see also McCann [1995]). See Ball [2004] for a very nice elementary introduction to monotone transportation. We are now able to state Caffarelli's Contraction Theorem.

Theorem 9.2 (Caffarelli [2000]). Let $b \in \mathbb{R}^d$, $c \in \mathbb{R}$ and V a convex function on $\mathbb{R}^d$. Let A be a positive definite matrix in $\mathbb{R}^{d\times d}$ and Q be the following quadratic function,

$$Q(x)=\langle Ax, x\rangle+\langle b, x\rangle+c,\ x \in \mathbb{R}^d.$$

Let μ and ν denote two probability measures on $\mathbb{R}^d$ with respective densities exp $(-Q)$ and exp $(-(Q+V))$ with respect to Lebesgue measure. Then the Brenier map T_{opt} pushing μ forward onto ν is a contraction:

$$|T(x)-T(y)| \leq |x-y| \text{ for all } x, y \in \mathbb{R}^d.$$

Notice that Caffarelli's Contraction Theorem is in particular valid when μ is a Gaussian measure and that case, ν is a strongly log-concave measure.

9.8. Concentration and convex geometry

Guédon [2012] gives a nice survey, explaining the connections between the Hyperplane conjecture, the KLS conjecture, the Thin Shell conjecture, the Variance conjecture and the Weak and Strong moments conjecture. Related papers include Guédon and Milman [2011] and Fradelizi, Guédon and Pajor [2013].

It is well-known that concentration properties are linked the behavior of moments. Bobkov and Madiman [2011] prove that if $\eta > 0$ is log-concave then the function

$$\overline{\lambda}_p=\frac{1}{\Gamma(p+1)}\mathbb{E}[\eta^p],\ p \geq 0,$$

is also log-concave, where Γ is the classical Gamma function. This is equivalent to having a so-called “reverse Lyapunov's inequality”,

$$\overline{\lambda}_a^{b-c}\overline{\lambda}_c^{a-b} \leq \overline{\lambda}_b^{a-c},\ a \geq b \geq c \geq 0.$$

Also, Bobkov [2003] proves that log-concavity of $\tilde{\lambda_p} = \mathbb{E}[(\eta/p)^p]$ holds (this is a consequence of the Prékopa-Leindler inequality). These results allow for instance Bobkov and Madiman [2011] to prove sharp concentration results for the information of a log-concave vector.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

9.9. Sampling from log concave distributions; convergence of Markov chain Monte Carlo algorithms

Sampling from log-concave distributions has been studied by Devroye [1984], Devroye [2012] for log-concave densities on $\mathbb{R}$, and by Frieze, Kannan and Polson [1994a,b], Frieze and Kannan [1999], and Lovász and Vempala [2007] for log-concave densities on $\mathbb{R}^d$; see also Lovász and Vempala [2003], Lovasz and Vempala [2006], Kannan, Lovász and Simonovits [1995], and Kannan, Lovász and Simonovits [1997].

Several different types of algorithms have been proposed: the rejection sampling algorithm of Devroye [1984] requires knowledge of the mode; see Devroye [2012] for some improvements. The algorithms proposed by Gilks and Wild [1992] are based on adaptive rejection sampling. The algorithms of Neal [2003] and Roberts and Rosenthal [2002] involve "slice sampling"; and the algorithms of Lovász and Vempala [2003], Lovasz and Vempala [2006], Lovász and Vempala [2007] are based on random walk methods.

Log-concavity and bounds for log-concave densities play an important role in the convergence properties of MCMC algorithms. For entry points to this literature, see Gilks and Wild [1992], Polson [1996], Brooks [1998], Roberts and Rosenthal [2002], Fort et al. [2003], Jylänki, Vanhatalo and Vehtari [2011], and Rudolf [2012].

9.10. Laplace approximations

Let $X_1, \ldots, X_n$ be i.i.d. real-valued random variables with density q and Laplace transform

$$\phi(s)=\mathbb{E}[\exp(\theta X_1)].$$

Let x^* be the upper limit of the support of q and let $\tau > 0$ be the upper limit of finiteness of φ. Let us assume that q is almost log-concave (see Jensen [1995] p155) on (x_0, x^*) for some $x_0 < x^*$. This means that there exist two constants $c_1 > c_2 > 0$ and two functions c and h on $\mathbb{R}$ such that

$$q(x)=c(x)\exp(-h(x)),\ x<x^*,$$

where $c_2 < c(x) < c_1$ whenever $x > x_0$ and h is convex. In particular, log-concave functions are almost log-concave for $x_0 = -\infty$. Now, fix $y \in \mathbb{R}$. The saddlepoint s associated to y is defined by

$$\left(\frac{d}{dt}\log\phi\right)(s)=y$$

and the variance $\sigma^2(s)$ is defined to be

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

$$\sigma^2(s)=\left(\frac{d^2}{dt^2}\log\phi\right)(s).$$

Let us write f_n the density of the empirical mean $\overline{X}=\sum_{i=1}^{n}X_i/n$. By Theorem 1 of Jensen [1991], if $q\in L^{\zeta}(\lambda)$ for $1<\zeta<2$, then the following saddlepoint approximations hold uniformly for $s_0<s<\tau$ for any $s_0>0$:

$$f_n(y)=\sqrt{\frac{n}{2\pi\sigma^2(s)}}\phi(s)^n\exp(-nsy)\left\{1+O\left(\frac{1}{n}\right)\right\}$$

and

$$\mathbb{P}(\overline{X}>y)=\frac{\phi(s)^n\exp(-nsy)}{s\sigma(s)\sqrt{n}}\left\{B_0(s\sigma(s)\sqrt{n})+O\left(\frac{1}{n}\right)\right\}$$

where $B_0(z)=z\exp(z^2/2)(1-\Phi(z))$ with Φ the standard normal distribution function. According to Jensen [1991], this result extends to the multidimensional setting where almost log-concavity is required on the entire space (and not just on some directional tails). As detailed in Jensen [1995], saddlepoint approximations have many applications in statistics, such as in testing or Markov chain related estimation problems.

As Bayesian methods are usually expensive in practice, approximations of quantities linked to the prior/posterior densities are usually needed. In connection with Laplace's method, log-quadratic approximation of densities are especially suited when considering log-concave functions, see Jensen [1995], Barber and Williams [1997], Minka [2001], and references therein.

9.11. Machine learning algorithms and Gaussian process methods

Boughorbel, Tarel and Boujemaa [2005] used the radius margin bound of Vapnik [2000] on the performance of a Support Vector Machine (SVM) in order to tune hyper-parameters of the kernel. More precisely they proved that for a weighted L^1 -distance kernel the radius is log-convex while the margin is log-concave. Then they used this fact to efficiently tune the multi-parameter of the kernel through a direct application of the Convex ConCave Procedure (or CCCP) due to Yuille and Rangarajan [2003]. In contrast to the gradient descent technique (Chapelle et al. [2002]), Boughorbel, Tarel and Boujemaa [2005] show that a variant of the CCCP which they call Log Convex ConCave Procedure (or LCCP) ensures that the radius margin bound decreases monotonically and converges to a local minimum without a search for the size step.

Bayesian methods based on Gaussian process priors have become popular in statistics and machine learning: see, for example Seeger [2004], Zhang, Dai and Jordan [2011], van der

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Vaart and van Zanten [2008], and van der Vaart and van Zanten [2011]. These methods require efficient computational techniques in order to be scalable, or even tractable in practice. Thus, log-concavity of the quantities of interest becomes important in this area, since it allows efficient optimization schemes.

In this context, Paninski [2004] shows that the predictive densities corresponding to either classification, regression, density estimation or point process intensity estimation models, are log-concave given any observed data. Furthermore, in the density and point process intensity estimation, the likelihood is log-concave in the hyperparameters controlling the mean function of the Gaussian prior. In the classification and regression settings, the mean, covariance and observation noise parameters are log-concave. As noted in Paninski [2004], the results still hold for much more general prior distributions than Gaussian: it suffices that the prior and the noise (in models where a noise appears) are jointly log-concave. The proofs are based on preservation properties for log-concave functions such as pointwise limit or preservation by marginalization.

9.12. Compressed sensing and random matrices

Compressed sensing, aiming at reconstructing sparse signals from incomplete measurement, is extensively studied since the seminal works of Donoho [2006], Candès, Romberg and Tao [2006] and Candès and Tao [2006]. As detailed in Chafaï et al. [2012], compressed sensing is intimately linked to the theory of random matrices. The matrices ensembles that are most frequently used and studied are those linked to Gaussian matrices, Bernoulli matrices and Fourier (sub-)matrices.

By analogy with the Wishart Ensemble, the Log-concave Ensemble is defined in Adamczak et al. [2010] to be the set of squared $n \times n$ matrices equipped with the distribution of AA^*, where A is a $n \times N$ matrix with i.i.d. columns that have an isotropic log-concave distribution. Adamczak et al. [2010] show that the Log-concave Ensemble satisfies a sharp Restricted Isometry Property (RIP), see also Chafaï et al. [2012] Chapter 2.

9.13. Log-concave and s-concave as nonparametric function classes in statistics

Nonparametric estimation of log-concave densities was initiated by Walther [2002] in the context of testing for unimodality. For log-concave densities on $\mathbb{R}$ it has been explored in more detail by Dümbgen and Rufibach [2009], Balabdaoui, Rufibach and Wellner [2009], and recent results for estimation of log-concave densities on $\mathbb{R}^d$ have been obtained by Cule and Samworth [2010], Cule, Samworth and Stewart [2010], Dümbgen, Samworth and Schuhmacher [2011]. Cule, Samworth and Stewart [2010] formulate the problem of computing the maximum likelihood estimator of a multidimensional log-concave density as a non-differentiable convex optimization problem and propose an algorithm that combines techniques of computational geometry with Shor's r-algorithm to produce a sequence that converges to the estimator. An R version of the algorithm is available in the package LogConcDEAD: *Log-Concave Density Estimation in Arbitrary Dimensions*, with further description of the algorithm given in Cule, Gramacy and Samworth [2009]. Nonparametric estimation of s–concave densities has been studied by Seregin and Wellner [2010]. They show that the MLE exists and is Hellinger consistent. Doss and Wellner [2013] have

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

obtained Hellinger rates of convergence for the maximum likelihood estimators of log-concave and s–concave densities on $\mathbb{R}$, while Kim and Samworth [2014] study Hellinger rates of convergence for the MLEs of log-concave densities on $\mathbb{R}^d$. Henningsson and Astrom [2006] consider replacement of Gaussian errors by log-concave error distributions in the context of the Kalman filter.

Walther [2009] gives a review of some of the recent progress.

Acknowledgments

We owe thanks to Michel Ledoux for a number of pointers to the literature and for sending us a pre-print version of Bobkov and Ledoux [2014].

10. Appendix A: Brascamp-Lieb inequalities and more

Let $\mathbf{X}$ have distribution P with density $p = \exp(-\phi)$ on $\mathbb{R}^d$ where ϕ is strictly convex and $\phi \in C^2(\mathbb{R}^d)$; thus $\nabla^2(\phi)(x) = \phi''(x) > 0$, $x \in \mathbb{R}^d$ as symmetric matrices. Let G, H be real-valued functions on $\mathbb{R}^d$ with $G, H \in C^1(\mathbb{R}^d) \bigcap L_2(P)$. We let $H_1(P)$ denote the set of functions f in $L_2(P)$ such that ∇f (in the distribution sense) is in $L_2(P)$.

Let $\mathbf{Y}$ have distribution Q with density $q = \psi^{-\beta}$ on an open, convex set $\Omega \subset \mathbb{R}^d$ where $\beta > d$ and ψ is a positive, strictly convex and twice continuously differentiable function on Ω. In particular, Q is $s = -1/(\beta - d)$-concave (see Definition 2.5 and Borell [1974], Borell [1975]). Let T be a real-valued function on $\mathbb{R}^d$ with $T \in C^1(\Omega) \bigcap L_2(Q)$. The following Proposition summarizes a number of analytic inequalities related to a Poincaré-type inequality from Brascamp and Lieb [1976]. Such inequalities are deeply connected to concentration of measure and isoperimetry, as exposed in Bakry, Gentil and Ledoux [2014]. Concerning log-concave measures, these inequalities are also intimately linked to the geometry of convex bodies. Indeed, as noted by Carlen, Cordero-Erausquin and Lieb [2013] page 9,

> "The Brascamp-Lieb inequality (1.3), as well as inequality (1.8), have connections with the geometry of convex bodies. It was observed in [2] (Bobkov and Ledoux [2000]) that (1.3) (*see Proposition 10.1, (a)*) can be deduced from the Prékopa-Leindler inequality (which is a functional form of the Brunn-Minkowski inequality). But the converse is also true: the Prékopa theorem follows, by a local computation, from the Brascamp-Lieb inequality (see [5] (Cordero-Erausquin [2005]) where the procedure is explained in the more general complex setting). To sum up, the Brascamp-Lieb inequality (1.3) can be seen as the *local* form of the Brunn-Minkowski inequality for convex bodies."

Proposition 10.1.

a. Brascamp and Lieb [1976]: If p is strictly log-concave, then

$$\mathrm{Var}(G(\mathbf{X})) \le E[\nabla G(\mathbf{X})^T (\varphi''(\mathbf{X}))^{-1} \nabla G(\mathbf{X})].$$

b. If $p = \exp(-\phi)$ where $\phi'' > cI$ with $c > 0$, then

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

$$\mathrm{Var}(G(\mathbf{X})) \leq \frac{1}{c} E|\nabla G(\mathbf{X})|^2.$$

c. Hargé [2008]: If $\phi \in L_2(P)$, then for all $f \in H_1(P)$,

$$\mathrm{Var}(f(\mathbf{X})) \leq E[\nabla f(\mathbf{X})^T (\varphi''(\mathbf{X}))^{-1} \nabla f(\mathbf{X})] - \frac{1+a/b}{d} (\mathrm{Cov}(\varphi(\mathbf{X}), f(\mathbf{X})))^2,$$

Where

$$a = \inf_{x \in \mathbb{R}^d} \min\{\lambda \text{ eigenvalue of } \varphi''(x)\}$$

and

$$b = \sup_{x \in \mathbb{R}^d} \max\{\lambda \text{ eigenvalue of } \varphi''(x)\}.$$

Notice that $0 \leq a \leq b \leq +\infty$ and $b > 0$.

d. Bobkov and Ledoux [2009]: If $U = \psi T$, then

$$(\beta+1)\mathrm{Var}(T(\boldsymbol{Y})) \leq E\left[\frac{1}{V(\boldsymbol{Y})} \nabla U(\boldsymbol{Y})^T (\varphi''(\boldsymbol{Y}))^{-1} \nabla U(\boldsymbol{Y})\right] + \frac{n}{\beta - n} E[T(\boldsymbol{Y})]^2.$$

Taking $\psi = \exp(\phi/\beta)$ and setting $R_{\phi,\beta} \equiv \phi'' + \beta^{-1} \nabla\phi \otimes \nabla\phi$, this implies that for any $\beta \geq d$,

$$\mathrm{Var}(G(\mathbf{X})) \leq C_\beta E[\nabla G(\mathbf{X})^T (R_{\varphi,\beta}(\mathbf{X}))^{-1} \nabla G(\mathbf{X})],$$

where $C_\beta = (1 + \sqrt{\beta+1})^2/\beta$. Notice that $1 \leq C_\beta \leq 6$.

e. Bakry [1994]: If $p = \exp(-\phi)$ where $\phi'' \geq cI$ with $c > 0$, then

$$\mathrm{Ent}_P(G^2(\mathbf{X})) \leq \frac{1}{c} E|\nabla G(\mathbf{X})|^2.$$

where

$$\mathrm{Ent}_P(Y^2) = E[Y^2 \log(Y^2)] - E[Y^2] \log(E[Y^2]).$$

f. Ledoux [1996], Ledoux [2001]: If the conclusion of (*e*) holds for all smooth G, then $E_P \exp(a|\mathbf{X}|^2) < \infty$ for every $a < 1/(2c)$.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

g. Bobkov [1999]: If $E_P \exp(a|\mathbf{X}|^2) < \infty$ for a log-concave measure P and some $a > 0$, then the conclusion of (*e*) holds for some $c = c_d$.

h. Bobkov and Ledoux [2000]: If ϕ is strongly convex with respect to a norm $\|\cdot\|$ (so p is strongly log-concave with respect to $\|\cdot\|$), then

$$\mathrm{Ent}_P(G^2(\mathbf{X})) \leq \frac{2}{c} E_P \|\nabla G(\mathbf{X})\|_*^2$$

for the dual norm $\|\cdot\|_*$.

Inequality (a) originated in Brascamp and Lieb [1976] and the original proof of the authors is based on a dimensional induction. For more details about the induction argument used by Brascamp and Lieb [1976], see Carlen, Cordero-Erausquin and Lieb [2013]. Building on Maurey [1991], Bobkov and Ledoux [2000] give a non-inductive proof of (a) based on the Prékopa-Leindler theorem Prékopa [1971], Prékopa [1973], Leindler [1972] which is the functional form of the celebrated Brunn-Minkowski inequality. The converse is also true in the sense that the Brascamp-Lieb inequality (a) implies the Prékopa-Leindler inequality, see Cordero-Erausquin [2005]. Inequality (b) is an easy consequence of (a) and is referred to as a Poincaré inequality for strongly log-concave measures.

Inequality (c) is a reinforcement of the Brascamp-Lieb inequality (a) due to Hargé [2008]. The proof is based on (Marvovian) semi-group techniques, see Bakry, Gentil and Ledoux [2014] for a comprehensive introduction to these tools. In particular, Hargé [2008], Lemma 7, gives a variance representation for strictly log-concave measures that directly implies the Brascamp-Lieb inequality (a).

The first inequality in (d) is referred in Bobkov and Ledoux [2009] as a "weighted Poincaré-type inequality" for convex (or s–concave with negative parameter s) measures. It implies the second inequality of (d) which is a quantitative refinement of the Brascamp-Lieb inequality (a). Indeed, Inequality (a) may be viewed as the limiting case in the second inequality of (d) for $\beta \to +\infty$ (as in this case $C_\beta \to 1$ and $R_{\phi,\beta} \to \phi''$). As noted in Bobkov and Ledoux [2009], for finite β the second inequality of (d) may improve the Brascamp-Lieb inequality in terms of the decay of the weight. For example, when Y is a random variable with exponential distribution with parameter $\lambda > 0$ ($q(y) = \lambda e^{-\lambda y}$ on $\Omega = (0, \infty)$), the second inequality in (d) gives the usual Poincaré-type inequality,

$$\mathrm{Var}(G(Y)) \leq \frac{6}{\lambda^2} E[(G'(Y))^2],$$

which cannot be proved as an direct application of the Brascamp-Lieb inequality (a). Klaassen [1985] shows that the inequality in the last display holds (in the exponential case) with 6 replaced by 4, and establishes similar results for other distributions. The exponential and two-sided exponential (or Laplace) distributions are also treated by Bobkov and Ledoux [1997].

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Points (e) to (h) deal, in the case of (strongly) log-concave measures, with the so-called logarithmic-Sobolev inequality, which is known to strengthen the Poincaré inequality (also called spectral gap inequality) (see for instance Chapter 5 of Bakry, Gentil and Ledoux [2014]). Particularly, Bobkov and Ledoux [2000] proved their result in point (d), *via* the use of the Prékopa-Leindler inequality. In their survey on optimal transport, Gozlan and Léonard [2010] show how to obtain the result of Bobkov and Ledoux from some transport inequalities.

We give now a simple application of the Brascamp-Lieb inequality (a), that exhibits its relation with the Fisher information for location.

Example 10.2. Let $G(x) = a^T x$ for $a \in \mathbb{R}^d$. Then the inequality in (a) becomes

$$a^T \mathrm{Cov}(\mathbf{X}) a \leq a^T E\{[\varphi''(\mathbf{X})]^{-1}\} a \quad (10.22)$$

or equivalently

$$\mathrm{Cov}(\mathbf{X}) \leq E\{[\varphi''(\mathbf{X})]^{-1}\}$$

with equality if $\mathbf{X} \sim N_d(\mu, \Sigma)$ with Σ positive definite. When $d = 1$ (10.22) becomes

$$\mathrm{Cov}(\mathbf{X}) \leq E[(\varphi'')^{-1}(\mathbf{X})] = E[((-\log p)'')^{-1}(\mathbf{X})] \quad (10.23)$$

while on the other hand

$$\mathrm{Cov}(\mathbf{X}) \geq [E(\varphi'')(\mathbf{X})]^{-1} \equiv I_{\mathrm{loc}}^{-1}(\mathbf{X}) \quad (10.24)$$

where $I_{loc}(X) = E(\phi'')$ denotes the Fisher information for location (in X or p); in fact for $d \geq 1$

$$\mathrm{Cov}(\mathbf{X}) \geq [E(\varphi'')(\mathbf{X})]^{-1} \equiv I_{\mathrm{loc}}^{-1}(\mathbf{X})$$

where $I_{loc}(X) \equiv E(\phi'')$ is the Fisher information matrix (for location). If $X \sim N_d(\mu, \Sigma)$ then equality holds (again). On the other hand, when $d = 1$ and p is the logistic density given in Example 2.11, then $\phi'' = 2p$ so the right side in (10.23) becomes $E\{(2p(\mathbf{X}))^{-1}\} = \int_{\mathbb{R}} (1/2) dx = \infty$ while V ar$(\mathbf{X}) = \pi^2/3$ and $I_{loc}(\mathbf{X}) = 1/3$ so the inequality (10.23) holds trivially, while the inequality (10.24) holds with strict inequality:

$$3 = I_{\mathrm{loc}}^{-1}(\mathbf{X}) < \frac{\pi^2}{3} = \mathrm{Var}(\mathbf{X}) < E[(\varphi'')^{-1}(\mathbf{X})] = \infty.$$

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

(Thus while X is slightly inefficient as an estimator of location for p, it is not drastically inefficient.)

Now we briefly summarize the asymmetric Brascamp - Lieb inequalities of Menz and Otto [2013] and Carlen, Cordero-Erausquin and Lieb [2013].

Proposition 10.3.

a. Menz and Otto [2013]: Suppose that $d = 1$ and $G, H \in C^1(\mathbb{R}) \cap L^2(P)$. If p is strictly log-concave and $1/r + 1/s = 1$ with $r \geq 2$, then

$$|\mathrm{Cov}(G(X), H(X))| \leq \sup_x \left\{ \frac{|H'(x)|}{\varphi''(X)} \right\} E\{|G'(X)|\}.$$

b. Carlen, Cordero-Erausquin and Lieb [2013]: If p is strictly log-concave on $\mathbb{R}^d$ and $\lambda_{min}(x)$ denotes the smallest eigenvalue of ϕ'', then

$$|\mathrm{Cov}(G(X), H(X))| \leq \|(\varphi'')^{-1/r} \nabla G\|_s \cdot \|\lambda_{\min}^{(2-r)/r} (\varphi'')^{-1/r} \nabla H\|_r.$$

Remark 10.4.

i. When $r = 2$, the inequality in (b) yields

$$(\mathrm{Cov}(G(X), H(X)))^2 \leq E\{\nabla G^T (\varphi'')^{-1} \nabla G\} \cdot E\{\nabla G^T (\varphi'')^{-1} \nabla G\}$$

which can also be obtained from the Cauchy-Schwarz inequality and the Brascamp-Lieb inequality (a) of Proposition 10.1.

ii. (ii) The inequality (b) also implies that

$$|\mathrm{Cov}(G(X), H(X))| \leq \|(\lambda_{\min}^{-1/r} \nabla G\|_s \cdot \|\lambda_{\min}^{-1/s} \nabla H\|_r;$$

taking $r = \infty$ and $s = 1$ yields

$$|\mathrm{Cov}(G(X), H(X))| \leq \|\nabla G\|_1 \cdot \|\lambda_{\min}^{-1} \nabla H\|_\infty$$

which reduces to the inequality in (a) when $d = 1$.

11. Appendix B: some further proofs

Proof. Proposition 2.3: **(b):** $p_\theta(x) = f(x - \theta)$ has MLR if and only if

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

$$\frac{f(x-\theta')}{f(x-\theta)} \leq \frac{f(x'-\theta')}{f(x-\theta)} \text{ for all } x<x', \theta<\theta'$$

This holds if and only if

$$\log f(x-\theta')+\log f(x'-\theta) \leq \log f(x'-\theta')+\log f(x-\theta). \quad (11.25)$$

Let $t = (x' - x)/(x' - x + \theta' - \theta)$ and note that

$$\begin{aligned} x-\theta &= t(x-\theta')+(1-t)(x'-\theta), \\ x'-\theta' &= (1-t)(x-\theta')+t(x'-\theta) \end{aligned}$$

Hence log-concavity of f implies that

$$\begin{aligned} \log f(x-\theta) &\geq t\log f(x-\theta')+(1-t)\log f(x'-\theta), \\ \log f(x'-\theta') &\geq (1-t)\log f(x-\theta')+t\log f(x'-\theta). \end{aligned}$$

Adding these yields (11.25); i.e. f log-concave implies $p_\theta(x)$ has MLR in x.

Now suppose that $p_\theta(x)$ has MLR so that (11.25) holds. In particular that holds if x, x', θ, θ' satisfy $x - \theta' = a < b = x' - \theta$ and $t = (x' - x)/(x'-x+\theta' - \theta) = 1/2$, so that $x - \theta = (a + b)/2 = x' - \theta'$. Then (11.25) becomes

$$\log f(a)+\log f(b) \leq 2\log f((a+b)/2).$$

This together with measurability of f implies that f is log-concave.

(a): Suppose f is PF_2. Then for $x < x'$, $y < y'$,

$$\det\begin{pmatrix} f(x-y) & f(x-y') \\ f(x'-y) & f(x'-y') \end{pmatrix} = f(x-y)f(x'-y')-f(x-y')f(x'-y) \geq 0$$

if and only if

$$f(x-y')f(x'-y) \leq f(x-y)f(x'-y'),$$

or, if and only if

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

$$\frac{f(x-y')}{f(x-y)} \le \frac{f(x'-y')}{f(x'-y)}.$$

That is, $p_Y(x)$ has MLR in x. By (b) this is equivalent to f log-concave.

Proof. Proposition 2.23: To prove Proposition 2.23 it suffices to note the log-concavity of $g(x)=p(x)/\prod_{j=1}^{d}\phi(x_j/\sigma)$ and to apply Proposition 2.20 (which holds as well for log-concave functions). The claims then follow by basic calculations.

Here are the details. Under the assumption that $\phi \in C^2$ (and even more generally) the equivalence between (a) and (b) follows from Rockafellar and Wets [1998], Exercise 12.59, page 565. The equivalence between (a) and (c) follows from the corresponding proof concerning the equivalence of (a) and (c) in Proposition 2.20; see e.g. Boyd and Vandenberghe [2004], page 71.

(a) implies (d): this follows from the corresponding implication in Proposition 2.20. Also note that for $x_1, x_2 \in \mathbb{R}^d$ we have

$$\begin{aligned}
&\langle \nabla\varphi_{J_a}(x_2) - x_2/c - (\nabla\varphi_{J_a}(x_1) - x_1/c), x_2 - x_1\rangle \\
&= \langle \nabla\varphi(a+x_2) - \nabla\varphi(a-x_2) - x_2/c - (\nabla\varphi(a+x_1) - \nabla\varphi(a-x_1) - x_1/c), x_2 - x_1\rangle \\
&= \langle \nabla\varphi(a+x_2) - \nabla\varphi(a+x_1) - (x_2-x_1)/(2c), x_2 - x_1\rangle - \langle \nabla\varphi(a-x_2) - \nabla\varphi(a-x_1) + (x_2-x_1)/(2c), x_2 - x_1\rangle \\
&= \langle \nabla\varphi(a+x_2) - \nabla\varphi(a+x_1) - (a+x_2-(a+x_1))/(2c), x_2 - x_1\rangle + \langle \nabla\varphi(a-x_1) - \nabla\varphi(a-x_2) - (a-x_1-(a-x_2))/(2c), a - x_1
\end{aligned}$$

if $c = \sigma^2/2$.

(d) implies (e): this also follows from the corresponding implication in Proposition 2.20. Also note that when $\phi \in C^2$ so that $\nabla^2\phi$ exists,

$$\nabla^2\varphi_{J_a}(x) - 2I/\sigma^2 = \nabla^2\varphi(a+x) + \nabla^2\varphi(a-x) - 2I/\sigma^2 = \nabla^2\varphi(a+x) - I/\sigma^2 + \nabla^2\varphi(a-x) - I/\sigma^2 \ge 0+0=0.$$

To complete the proof when $\phi \in C^2$ we show that (e) implies (c). Choosing $a = x_0$ and $x = 0$ yields

$$0 \le \nabla^2\varphi_{J_a}(0) - 2I/\sigma^2 = \nabla^2\varphi(x_0) + \nabla^2\varphi(x_0) - 2I/\sigma^2 = 2(\nabla^2\varphi(x_0) - I/\sigma^2),$$

and hence (c) holds.

To complete the proof more generally, we proceed as in Simon [2011], page 199: to see that (e) implies (f), let $a = (x_1 + x_2)/2$, $x = (x_1 - x_2)/2$. Since $J_a(\cdot; g)$ is even and radially monotone, $J_a(0; g)^{1/2} \ge J_a(x; g)^{1/2}$; that is,

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

$$\{g(a+0)g(a-0)\}^{1/2} \geq \{g(a+x)g(a-x)\}^{1/2},$$

or

$$g((x_1+x_2)/2) \geq g(x_1)^{1/2}g(x_2)^{1/2}.$$

Finally (f) implies (a): as in Simon [2011], page 199 (with "convex" changed to "concave" three times in the last three lines there): midpoint log-concavity of g together with lower semicontinuity implies that g is log-concave, and hence p is strongly log-concave, so (a) holds.

Proof. Proposition 2.24: First notice that by Proposition 5.5, we may assume that f is $\mathscr{C}^\infty$ (so ϕ is also $\mathscr{C}^\infty$).

i. As I is $\mathscr{C}^\infty$, we differentiate I twice. We have $I'(p) = f'(F^{-1}(p))/I(p) = -\phi'(F^{-1}(p))$ and

$$I''(p) = -\varphi''(F^{-1}(p))/I(p) \leq -c^{-1}\|f\|_\infty^{-1}. \quad (11.26)$$

This gives the first part of **(i).** The second part comes from the fact that $\|f\|_\infty^{-1} \geq \sqrt{\mathrm{var}(X)}$ by Proposition 5.2 below.

ii. It suffices to exhibit an example. We take $X \geq 0$, with density

$$f(x) = xe^{-x}\mathbf{1}_{(0,\infty)}(x).$$

Then $f = e^{-\phi}$ is log-concave (in fact, f log-concave of order 2, see Definition 2.15) and not strongly log-concave as, on the support of f, $\phi''(x) = x^{-2} \to 0$ as $x \to \infty 0$. By the equality in (11.26) we have

$$I''(p) = -\frac{\varphi''}{f}(F^{-1}(p)).$$

Hence, to conclude it suffices to show that $\inf_{x>0}\{\phi''/f\} > 0$. By simple calculations, we have

$$\varphi''(x)/f(x) = x^{-3}e^x\mathbf{1}_{(0,\infty)}(x) \geq e^3/27 > 0,$$

so **(ii)** is proved.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

iii. We take $f(x) = \exp(-\phi) = \alpha^{-1} \exp(-\exp(x))\, 1_{\{x>0\}}$ where $\alpha = \int_0^\infty \exp(-\exp(x))dx$. Then the function R_h is $\mathscr{C}^\infty$ on (0, 1) and we have by basic calculations, for any $p \in (0, 1)$,

$$R_h'(p) = f(F^{-1}(p)+h)/f(F^{-1}(p))$$

and

$$R_h''(p) = \frac{f(F^{-1}(p)+h)}{f(F^{-1}(p))^2}(\varphi'(F^{-1}(p)) - \varphi'(F^{-1}(p)+h)).$$

Now, for any $x > 0$, taking $p = F(x)$ in the previous identity gives

$$R_h''(F(x)) = \frac{f(x+h)}{f(x)^2}(\varphi'(x) - \varphi'(x+h)) = \alpha^{-1}\exp(\exp(x)(2-\exp(h)))\cdot\exp(x)(1-\exp(h)). \quad (11.27)$$

We deduce that if $h > \log 2$ then $R_h''(F(x)) \to 0$ whenever $x \to +\infty$. Taking $h_0 = 1$ gives point **(iii)**.

iv. For X of density $f(x) = xe^{-x}\mathbf{1}_{(0,+\infty)}(x)$, we have $\inf R_h'' \leq -he^{-h} < 0$. Our proof of the previous fact is based on identity (11.27) and left to the reader.

Proof. Proposition 2.25: Here are the details. Under the assumption that $\phi \in C^2$ (and even more generally) the equivalence of (a) and (b) follows from Rockafellar and Wets [1998], Exercise 12.59, page 565. The equivalence of (a) and (c) follows from the corresponding proof concerning the equivalence of (a) and (c) in Proposition 2.20; see e.g. Boyd and Vandenberghe [2004], page 71.

That (a) implies (d): this follows from the corresponding implication in Proposition 2.20. Also note that for $x_1, x_2 \in \mathbb{R}^d$ we have

$$\begin{aligned}
&\langle \nabla\varphi_{J_a}(x_2) - x_2/c - (\nabla\varphi_{J_a}(x_1) - x_1/c), x_2 - x_1\rangle \\
&= \langle \nabla\varphi(a+x_2) - \nabla\varphi(a-x_2) - x_2/c - (\nabla\varphi(a+x_1) - \nabla\varphi(a-x_1) - x_1/c), x_2 - x_1\rangle \\
&= \langle \nabla\varphi(a+x_2) - \nabla\varphi(a+x_1) - (x_2-x_1)/(2c), x_2 - x_1\rangle - \langle \nabla\varphi(a-x_2) - \nabla\varphi(a-x_1) + (x_2-x_1)/(2c), x_2 - x_1\rangle \\
&= \langle \nabla\varphi(a+x_2) - \nabla\varphi(a+x_1) - (a+x_2-(a+x_1))/(2c), x_2 - x_1\rangle + \langle \nabla\varphi(a-x_1) - \nabla\varphi(a-x_2) - (a-x_1-(a-x_2)) \\
&/(2c), a - x_1 - (a-x_2)\rangle \geq 0
\end{aligned}$$

if $c = \sigma^2/2$.

(d) implies (e): this also follows from the corresponding implication in Proposition 2.20. Also note that when $\phi \in C^2$ so that $\nabla^2\phi$ exists,

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

$$\nabla^2\varphi_{J_a}(x)-2I/\sigma^2=\nabla^2\varphi(a+x)+\nabla^2\varphi(a-x)-2I/\sigma^2=\nabla^2\varphi(a+x)-I/\sigma^2+\nabla^2\varphi(a-x)-I/\sigma^2\geq 0+0=0.$$

To complete the proof when $\phi\in C^2$ we show that (e) implies (c). Choosing $a = x_0$ and $x = 0$ yields

$$0\leq\nabla^2\varphi_{J_a}(0)-2I/\sigma^2=\nabla^2\varphi(x_0)+\nabla^2\varphi(x_0)-2I/\sigma^2=2(\nabla^2\varphi(x_0)-I/\sigma^2),$$

and hence (c) holds.

To complete the proof more generally, we proceed as in Simon [2011], page 199: to see that (e) implies (f), let $a = (x_1 + x_2)/2$, $x = (x_1 - x_2)/2$. Since $J_a(\cdot; g)$ is even and radially monotone, $J_a(0; g)^{1/2} \geq J_a(x; g)^{1/2}$; that is,

$$\{g(a+0)g(a-0)\}^{1/2}\geq\{g(a+x)g(a-x)\}^{1/2},$$

or

$$g((x_1+x_2)/2)\geq g(x_1)^{1/2}g(x_2)^{1/2}.$$

Finally (f) implies (a): as in Simon [2011], page 199 (with “convex” changed to“concave” three times in the last three lines there): midpoint log-concavity of g together with lower semicontinuity implies that g is log-concave, and hence p is strongly log-concave, so (a) holds.

Proof. Proposition 5.5:

i. This is given by the stability of log-concavity through convolution.

ii. This is point **(b)** of Theorem 3.7.

iii. We have

$$\varphi_Z(z)=-\log\int_{y\in\mathbb{R}^d}p(y)q(z-y)dy$$

and

$$\int_{y\in\mathbb{R}^d}\|\nabla_q(z-y)\|p(y)dy=\int_{y\in\mathbb{R}^d}\|z-y\|q(z-y)p(y)dy<\infty$$

since $y\mapsto\|z-y\|\,q(z-y)$ is bounded. This implies that $p_Z > 0$ on $\mathbb{R}^d$ and

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

$$\nabla\varphi_Z(z)=\int_{y\in\mathbb{R}^d}\frac{z-y}{\sigma^2}\frac{p(y)q(z-y)}{\int_{y\in\mathbb{R}^d}p(u)q(z-u)du}dy=\sigma^{-2}\mathbb{E}[\sigma G|X+\sigma G=z]=\mathbb{E}[\rho_{\sigma G}(\sigma G)|X+\sigma G=z].$$

In the same manner, successive differentiation inside the integral shows that ϕ_Z is $\mathscr{C}^\infty$, which gives **(iii)**.

iv. Notice that

$$|||\int_{z,y\in\mathbb{R}^d}\sigma^{-4}(z-y)(z-y)^Tp(y)q(z-y)dydx|||\leq\sigma^{-4}\int_{y\in\mathbb{R}^d}\int_{z\in\mathbb{R}^d}\|z-y\|^2q(z-y)p(y)dy<\infty$$

as $y\mapsto\|z-y\|^2\,q(z-y)$ is bounded. Hence the Fisher information $J(Z)$ of Z is finite and we have

$$\begin{aligned}J(Z)&=\sigma^{-4}\int_{z,y\in\mathbb{R}^d}\mathbb{E}[\sigma G|X+\sigma G=z]\mathbb{E}[(\sigma G)^T|X+\sigma G=z]p(y)q(z-y)dydz\\&\leq\sigma^{-4}\int_{z,y\in\mathbb{R}^d}\mathbb{E}\left[\sigma G(\sigma G)^T|X+\sigma G=z\right]p(y)q(z-y)dydz\\&=\sigma^{-4}\int_{z,y\in\mathbb{R}^d}\left(\int_{u\in\mathbb{R}^d}(z-u)(z-u)^T\frac{p(u)q(z-u)}{\int_{y\in\mathbb{R}^d}p(v)q(z-v)dv}du\right)p(y)q(z-y)dydz\\&=\sigma^{-4}\int_{y\in\mathbb{R}^d}\left(\int_{u\in\mathbb{R}^d}(z-u)(z-u)^Tp(u)q(z-u)du\right)\frac{p(y)q(z-y)}{\int_{v\in\mathbb{R}^d}p(v)q(z-v)dv}dydz\\&=\sigma^{-4}\int_{z\in\mathbb{R}^d}\int_{u\in\mathbb{R}^d}(z-u)(z-u)^Tp(u)q(z-u)dudz\\&=\int_{u\in\mathbb{R}^d}\left(\underbrace{\sigma^{-4}\int_{z\in\mathbb{R}^d}(z-u)(z-u)^Tq(z-u)dz}_{J(\sigma G)}\right)p(u)du\\&=J(\sigma G),\end{aligned}$$

which is **(iv)**.

Proof. Proposition 5.6: The fact that $h_c\in SLC_1(c^{-1},d)$ is obvious due to Definition 2.8. By Theorem 5.1 above, there exist $a>0$ and $b\in\mathbb{R}$ such that

$$f(x)\leq e^{-a\|x\|+b},x\in\mathbb{R}^d.$$

We deduce that if X is a random vector with density f on $\mathbb{R}^d$, then $\mathbb{E}[e^{(a/2)\|X\|}]<\infty$ and so, for any $\beta>0$,

$$\mathbb{P}(\|X\|>2\beta)\leq Ae^{-a\beta},$$

where $A=\mathbb{E}e^{(a/2)\|X\|}]>0$ Take $\varepsilon\in(0,1)$. We have

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

$$\begin{aligned}\left|\int_{\mathbb{R}^d} f(v)e^{-c\|v\|^2/2}dv-1\right| &= \int_{\mathbb{R}^d} f(v)\left(1-e^{-c\|v\|^2/2}\right)dv\\ &= \int_{\mathbb{R}^d} f(v)\left(1-e^{-c\|v\|^2/2}\right)\mathbf{1}_{\left\{\|v\|\le 2c^{-\varepsilon/2}\right\}}dv+\int_{\mathbb{R}^d} f(v)\left(1-e^{-c\|v\|^2/2}\right)\mathbf{1}_{\left\{\|v\|>2c^{-\varepsilon/2}\right\}}dv\\ &\le \left(1-e^{-2c^{1-\varepsilon}}\right)\int_{\mathbb{R}_d} f(v)\mathbf{1}_{\left\{\|v\|\le\sqrt{2}c^{-\varepsilon/2}\right\}}dv+\mathbb{P}(\|X\|>2c^{-\varepsilon/2})\\ &\le \left(1-e^{-2c^{1-\varepsilon}}\right)+Ae^{-ac^{-\varepsilon/2}}.\end{aligned}$$

We set $B_a = (1 - e^{-2a^{1-\varepsilon}}) + Ae^{-aa^{-\varepsilon/2}}$ and we then have

$$\left|\int_{\mathbb{R}^d} f(v)e^{-c\|v\|^2/2}dv-1\right| \le B_c = O_{c\to 0}\left(c^{1-\varepsilon}\right)\to_{c\to 0} 0.$$

Now, for $x \in \mathbb{R}^d$, we have, for all $c > 0$ such that $B_c < 1$,

$$\begin{aligned}|h_c(x)-f(x)| &= \left|\frac{f(x)e^{-c\|x\|^2/2}}{\int_{\mathbb{R}^d}(v)e^{-c\|v\|^2/2}dv}-f(x)\right.\\ &\le \left|\frac{f(x)e^{-c\|x\|^2/2}}{\int_{\mathbb{R}^d} f(v)e^{-c\|v\|^2/2}dv}-f(x)e^{-c\|x\|^2/2}\right|+\left|f(x)e^{-c\|x\|^2/2}-f(x)\right|\\ &\le f(x)\left|\left(\int_{\mathbb{R}^d} f(v)e^{-c\|v\|^2/2}dv\right)^{-1}-1\right|+f(x)\left(1-e^{-c\|x\|^2/2}\right)\\ &\le f(x)\left(\frac{B_c}{1-B_c}+1-e^{-c\|x\|^2/2}\right).\end{aligned}$$

Hence, for all $c > 0$ such that $B_c < 1$,

$$\begin{aligned}\sup_{x\in\mathbb{R}^d}|h_c(x) &- f(x)| \le \sup_{\left\{x;\|x\|\le 2c^{-\varepsilon/2}\right\}}|h_c(x)\\ &- f(x)|\\ &+\sup_{\left\{x;\|x\|\le 2c^{-\varepsilon/2}\right\}}|h_c(x)\\ &- f(x)| \le e^b\left(\frac{B_c}{1-B_c}\right.\\ &+1-e^{-2c^{1-\varepsilon}}+e^{-2ac^{-\varepsilon/2}+b}\left(\frac{B_c}{1-B_c}\right.\\ &\left.\left.+1\right)\right)=O(c^{1-\varepsilon})\text{ as } c\to 0.\end{aligned}$$

Furthermore, for $p \in [1, \infty)$,

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

$$\begin{aligned}\int_{\mathbb{R}^d}|h_c(x)-f(x)|^p dx &= \int_{\mathbb{R}^d}|h_c(x)-f(x)|^p \mathbf{1}_{\left\{\|x\|\le 2c^{-\varepsilon/2}\right\}}dx + \int_{\mathbb{R}^d}|h_c(x)-f(x)|^p \mathbf{1}_{\left\{\|x\|>2c^{-\varepsilon/2}\right\}}dx \\ &\le \sup_{\left\{x;\|x\|>2c^{-\varepsilon/2}\right\}}|h_c(x)-f(x)|^p + \int_{\mathbb{R}^d} f(x)^p\left(\frac{B_c}{1-B_c}+1\right)\mathbf{1}_{\left\{\|x\|>2c^{-\varepsilon/2}\right\}}dx \\ &\le e^{pb}\left(\frac{B_c}{1-B_c}+1-e^{-2c^{1-\varepsilon}}\right)^p + \left(\frac{B_c}{1-B_c}+1\right)e^{(p-1)b}\mathbb{P}\left(\|X\|>2c^{-\varepsilon/2}\right) \\ &\le e^{pb}\left(\frac{B_c}{1-B_c}+1-e^{-2c^{1-\varepsilon}}\right)^p + A\left(\frac{B_c}{1-B_c}+1\right)e^{(p-1)b}e^{-ac^{-\varepsilon/2}} \\ &= O\left(c^{p(1-\varepsilon)}\right) \text{ as } c\to 0.\end{aligned}$$

The proof is now complete.

References

Adamczak R, Litvak AE, Pajor A, Tomczak-Jaegermann N. Quantitative estimates of the convergence of the empirical covariance matrix in log-concave ensembles. J Amer Math Soc. 2010; 23:535–561. MR2601042 (2011c:60019).

Adamczak R, Guédon O, Latala R, Litvak AE, Oleszkiewicz K, Pajor A, Tomczak-Jaegermann N. Moment estimates for convex measures. Electron J Probab. 2012; 17(101):19. MR3005719.

Alberti G, Ambrosio L. A geometrical approach to monotone functions in $\mathbf{R}^n$. Math Z. 1999; 230:259–316. MR1676726 (2000d:49033).

Alexandroff AD. Almost everywhere existence of the second differential of a convex function and some properties of convex surfaces connected with it. Leningrad State Univ Annals [Uchenye Zapiski] Math Ser. 1939; 6:3–35.

An MY. Logconcavity versus logconvexity: a complete characterization. J Economic Theory. 1998; 80:350–369.

Andreu F, Caselles V, Mazón JM. Some regularity results on the 'relativistic' heat equation. J Differential Equations. 2008; 245:3639–3663. MR2462697 (2009m:35515).

Artstein S, Ball KM, Barthe F, Naor A. Solution of Shannon's problem on the monotonicity of entropy. J Amer Math Soc. 2004a; 17:975–982. (electronic). MR2083473 (2005e:94047).

Artstein S, Ball KM, Barthe F, Naor A. On the rate of convergence in the entropic central limit theorem. Probab Theory Related Fields. 2004b; 129:381–390. MR2128238 (2006a:94023).

Bagnoli M, Bergstrom T. Log-concave probability and its applications. Econom Theory. 2005; 26:445–469. MR2213177.

Bakry, D. Lectures on probability theory (Saint-Flour, 1992) Lecture Notes in Math. Vol. 1581. Springer; Berlin: 1994. L'hypercontractivité et son utilisation en théorie des semigroupes; p. 1-114.MR1307413 (95m:47075)

Bakry, D.; Gentil, I.; Ledoux, M. Analysis and geometry of Markov diffusion operators Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]. Vol. 348. Springer; Berlin: 2014.

Balabdaoui F. Global convergence of the log-concave MLE when the true distribution is geometric. J Nonparametr Stat. 2014; 26:21–59. MR3174308.

Balabdaoui F, Rufibach K, Wellner JA. Limit distribution theory for maximum likelihood estimation of a log-concave density. Ann Statist. 2009; 37:1299–1331. MR2509075 (2010h:62290).

Balabdaoui F, Wellner JA. Chernoff's density is log-concave. Bernoulli. 2014; 20:231–244. [PubMed: 24761134]

Balabdaoui F, Jankowski H, Rufibach K, Pavlides M. Asymptotics of the discrete log-concave maximum likelihood estimator and related applications. J R Stat Soc Ser B Stat Methodol. 2013; 75:769–790. MR3091658.

Ball, K. Geometric aspects of functional analysis Lecture Notes in Math. Vol. 1850. Springer; Berlin: 2004. An elementary introduction to monotone transportation; p. 41-52.MR2087149 (2005h: 49102)

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Ball K, Barthe F, Naor A. Entropy jumps in the presence of a spectral gap. Duke Math J. 2003; 119:41–63. MR1991646 (2006k:94035).

Barber, D.; Williams, CKI. Advances in Neural Information Processing Systems (NIPS). Vol. 9. MIT Press; 1997. Gaussian Processes for Bayesian Classification via Hybrid Monte Carlo.

Barbu V, Da Prato G. The Kolmogorov operator associated with a stochastic variational inequality in $\mathbb{R}^n$ with convex potential. Rev Roumaine Math Pures Appl. 2008; 53:377–388. MR2474491 (2010d:60176).

Barbu, V.; Precupanu, T. Convexity and optimization in Banach spaces, second ed Mathematics and its Applications (East European Series). Vol. 10. D. Reidel Publishing Co.; Dordrecht: 1986.

Barron AR. Entropy and the central limit theorem. Ann Probab. 1986; 14:336–342. MR815975 (87h: 60048).

Bobkov S. Extremal properties of half-spaces for log-concave distributions. Ann Probab. 1996; 24:35–48. MR1387625 (97e:60027).

Bobkov SG. Isoperimetric and analytic inequalities for log-concave probability measures. Ann Probab. 1999; 27:1903–1921. MR1742893 (2001h:60026).

Bobkov, SG. Geometric aspects of functional analysis Lecture Notes in Math. Vol. 1807. Springer; Berlin: 2003. Spectral gap and concentration for some spherically symmetric probability measures; p. 37-43.MR2083386 (2005i:60028)

Bobkov SG. A note on the distributions of the maximum of linear Bernoulli processes. Electron Commun Probab. 2008; 13:266–271. MR2415134 (2009g:62010).

Bobkov SG. Gaussian concentration for a class of spherically invariant measures. J Math Sci (N Y). 2010; 167:326–339. Problems in mathematical analysis. No. 46. MR2839024 (2012i:60038).

Bobkov SG, Götze F. Exponential integrability and transportation cost related to logarithmic Sobolev inequalities. J Funct Anal. 1999; 163:1–28. MR1682772 (2000b:46059).

Bobkov S, Ledoux M. Poincaré's inequalities and Talagrand's concentration phenomenon for the exponential distribution. Probab Theory Related Fields. 1997; 107:383–400. MR1440138 (98e: 60026).

Bobkov SG, Ledoux M. From Brunn-Minkowski to Brascamp-Lieb and to logarithmic Sobolev inequalities. Geom Funct Anal. 2000; 10:1028–1052. MR1800062 (2002k:26028).

Bobkov SG, Ledoux M. Weighted Poincaré-type inequalities for Cauchy and other convex measures. Ann Probab. 2009; 37:403–427. MR2510011 (2011c:46095).

Bobkov S, Ledoux M. One-dimensional empirical measures, order statistics, and Kantorovich transport distances. 2014

Bobkov S, Madiman M. Concentration of the information in data with log-concave distributions. Ann Probab. 2011; 39:1528–1543. MR2857249 (2012h:60106).

Bogachev, VI. Gaussian measures Mathematical Surveys and Monographs. Vol. 62. American Mathematical Society; Providence, RI: 1998. MR1642391 (2000a:60004)

Bondesson, L. Generalized gamma convolutions and related classes of distributions and densities Lecture Notes in Statistics. Vol. 76. Springer-Verlag; New York: 1992. MR1224674 (94g:60031)

Bondesson, L. Advances in the theory and practice of statistics Wiley Ser Probab Statist Appl Probab Statist. Wiley; New York: 1997. On hyperbolically monotone densities; p. 299-313.MR1481175 (99c:60020)

Borell C. Convex measures on locally convex spaces. Ark Mat. 1974; 12:239–252. MR0388475 (52 ##9311).

Borell C. Convex set functions in d-space. Period Math Hungar. 1975; 6:111–136. MR0404559 (53 ##8359).

Borell C. Convexity of measures in certain convex cones in vector space σ-algebras. Math Scand. 1983; 53:125–144. MR733944 (86f:60010).

Boughorbel S, Tarel JP, Boujemaa N. The LCCP for Optimizing Kernel Parameters for SVM. Proceedings of International Conference on Artificial Neural Networks (ICANN'05). 2005; II: 589–594. http://perso.lcpc.fr/tarel.jean-philippe/publis/icann05a.html.

Boyd, S.; Vandenberghe, L. Convex optimization. Cambridge University Press; Cambridge: 2004. MR2061575 (2005d:90002)

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Brascamp, HJ.; Lieb, EH. A logarithmic concavity theorem with some applications Technical Report. Princeton University; 1974.

Brascamp, HJ.; Lieb, EH. Some inequalities for Gaussian measures. In: Arthurs, AM., editor. Functional Integration and its Applications. Clarendon Press; Oxford: 1975.

Brascamp HJ, Lieb EH. On extensions of the Brunn-Minkowski and Prékopa-Leindler theorems, including inequalities for log concave functions, and with an application to the diffusion equation. J Functional Analysis. 1976; 22:366–389. MR0450480 (56 ##8774).

Brenier Y. Polar factorization and monotone rearrangement of vector-valued functions. Comm Pure Appl Math. 1991; 44:375–417. MR1100809 (92d:46088).

Brézis, H. Opérateurs maximaux monotones et semi-groupes de contractions dans les espaces de Hilbert. North-Holland Publishing Co.; Amsterdam: 1973. North-Holland Mathematics Studies, No. 5. Notas de Matemática (50). MR0348562 (50 ##1060)

Brooks SP. MCMC convergence diagnosis via multivariate bounds on log-concave densities. Ann Statist. 1998; 26:398–433.

Brown, LD. Statistics and probability: essays in honor of C R Rao. North-Holland; Amsterdam: 1982. A proof of the central limit theorem motivated by the Cramér-Rao inequality; p. 141-148.MR659464 (83h:60021)

Caffarelli LA. Some regularity properties of solutions of Monge Ampère equation. Comm Pure Appl Math. 1991; 44:965–969. MR1127042 (92h:35088).

Caffarelli LA. The regularity of mappings with a convex potential. J Amer Math Soc. 1992; 5:99–104. MR1124980 (92j:35018).

Caffarelli LA. Monotonicity properties of optimal transportation and the FKG and related inequalities. Comm Math Phys. 2000; 214:547–563. MR1800860 (2002c:60029).

Candès EJ, Romberg JK, Tao T. Stable signal recovery from incomplete and inaccurate measurements. Comm Pure Appl Math. 2006; 59:1207–1223. MR2230846 (2007f:94007).

Candès EJ, Tao T. Near-optimal signal recovery from random projections: universal encoding strategies? IEEE Trans Inform Theory. 2006; 52:5406–5425. MR2300700 (2008c:94009).

Carlen EA, Cordero-Erausquin D. Subadditivity of the entropy and its relation to Brascamp-Lieb type inequalities. Geom Funct Anal. 2009; 19:373–405. MR2545242 (2010i:26032).

Carlen EA, Cordero-Erausquin D, Lieb EH. Asymmetric covariance estimates of Brascamp-Lieb type and related inequalities for log-concave measures. Ann Inst Henri Poincaré Probab Stat. 2013; 49:1–12. MR3060145.

Carlen EA, Lieb EH, Loss M. A sharp analog of Young's inequality on S^N and related entropy inequalities. J Geom Anal. 2004; 14:487–520. MR2077162 (2005k:82046).

Carlen EA, Soffer A. Entropy production by block variable summation and central limit theorems. Comm Math Phys. 1991; 140:339–371. MR1124273 (92m:60020).

Cattiaux P, Guillin A. Semi log-concave Markov diffusions. arXiv:0812.3141. 2013

Chafaï, D.; Guédon, O.; Lecué, G.; Pajor, A. Interactions between compressed sensing random matrices and high dimensional geometry Panoramas et Synthèses [Panoramas and Syntheses]. Vol. 37. Société Mathématique de France; Paris: 2012. MR3113826

Chapelle O, Vapnik V, Bousquet O, Mukherjee S. Choosing multiple parameters for support vector machines. Machine Learning. 2002; 46:131–159.

Cordero-Erausquin D. Some applications of mass transport to Gaussian-type inequalities. Arch Ration Mech Anal. 2002; 161:257–269. MR1894593 (2003h:49076).

Cordero-Erausquin D. On Berndtsson's generalization of Prékopa's theorem. Math Z. 2005; 249:401–410. MR2115450 (2005m:26016).

Cordero-Erausquin D, Ledoux M. The geometry of Euclidean convolution inequalities and entropy. Proc Amer Math Soc. 2010; 138:2755–2769. MR2644890 (2011b:42021).

Cule M, Gramacy RB, Samworth R. LogConcDEAD: An R Package for Maximum Likelihood Estimation of a Multivariate Log-Concave Density. Journal of Statistical Software. 2009; 29:1–20.

Cule M, Samworth R. Theoretical properties of the log-concave maximum likelihood estimator of a multidimensional density. Electron J Stat. 2010; 4:254–270. MR2645484 (2012f:62071).

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Cule M, Samworth R, Stewart M. Maximum likelihood estimation of a multi-dimensional log-concave density. J R Stat Soc Ser B Stat Methodol. 2010; 72:545–607. MR2758237.

Das Gupta S. Brunn-Minkowski inequality and its aftermath. J Multivariate Anal. 1980; 10:296–318. MR588074 (81m:26011).

Das Gupta, S.; Eaton, ML.; Olkin, I.; Perlman, M.; Savage, LJ.; Sobel, M. Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability (Univ California, Berkeley, Calif, 1970/1971), Vol II: Probability theory. Univ. California Press; Berkeley, Calif: 1972. Inequalitites on the probability content of convex regions for elliptically contoured distributions; p. 241-265.

Davidovič JS, Korenbljum BI, Hacet BI. A certain property of logarithmically concave functions. Dokl Akad Nauk SSSR. 1969; 185:1215–1218. MR0241584 (39 ##2924).

Devroye L. A simple algorithm for generating random variates with a log-concave density. Computing. 1984; 33:247–257. MR773927 (86d:65019).

Devroye L. A note on generating random variables with log-concave densities. Statist Probab Lett. 2012; 82:1035–1039. MR2910053.

Dharmadhikari, S.; Joag-Dev, K. Unimodality, convexity, and applications Probability and Mathematical Statistics. Academic Press Inc.; Boston, MA: 1988. MR954608 (89k:60020)

Dinghas A. Über eine Klasse superadditiver Mengenfunktionale von Brunn-Minkowski-Lusternikschem Typus. Math Z. 1957; 68:111–125. MR0096173 (20 ##2668).

Donoho DL. Compressed sensing. IEEE Trans Inform Theory. 2006; 52:1289–1306. MR2241189 (2007e:94013).

Doss, CH.; Wellner, JA. arXiv:1306.1438. Department of Statistics, University of Washington; 2013. Global rates of convergence of the MLEs of log-concave and s-concave densities Technical Report No. 614.

Dudley RM. On second derivatives of convex functions. Math Scand. 1977; 41:159–174. MR0482164 (58 ##2250).

Dudley RM. Acknowledgment of priority: “On second derivatives of convex functions”. Math Scand. 1980; 46:61. Math. Scand. **41** (1977), no. 1, 159–174; MR **58** #2250. MR585231 (82d:46061).

Dümbgen L, Rufibach K. Maximum likelihood estimation of a log-concave density and its distribution function: basic properties and uniform consistency. Bernoulli. 2009; 15:40–68. MR2546798 (2011b:62096).

Dümbgen L, Samworth R, Schuhmacher D. Approximation by log-concave distributions, with applications to regression. Ann Statist. 2011; 39:702–730. MR2816336 (2012e:62039).

Efron B. Increasing properties of Pólya frequency functions. Ann Math Statist. 1965; 36:272–279. MR0171335 (30 ##1566).

Ehm W. Binomial approximation to the Poisson binomial distribution. Statist Probab Lett. 1991; 11:7–16. MR1093412 (92b:60024).

Ehrhard A. Symétrisation dans l'espace de Gauss. Math Scand. 1983; 53:281–301. MR745081 (85f:60058).

Fort G, Moulines E, Roberts GO, Rosenthal JS. On the geometric ergodicity of hybrid samplers. J Appl Probab. 2003; 40:123–146. MR1953771 (2004b:60180).

Fradelizi, M.; Guédon, O.; Pajor, A. arXiv:1306.6794v1. Université Paris-Est; 2013. Spherical thin-shell concentration for convex measures Technical Report.

Frieze A, Kannan R, Polson N. Sampling from log-concave distributions. Ann Appl Probab. 1994a; 4:812–837. MR1284987 (95h:60105a).

Frieze A, Kannan R, Polson N. Correction:“Sampling from log-concave distributions”. Ann Appl Probab. 1994b; 4:1255. MR1304785 (95h:60105b).

Frieze A, Kannan R. Log-Sobolev inequalities and sampling from log-concave distributions. Ann Appl Probab. 1999; 9:14–26. MR1682608 (2000f:68052).

Gaenssler P, Molnár P, Rost D. On continuity and strict increase of the CDF for the sup-functional of a Gaussian process with applications to statistics. Results Math. 2007; 51:51–60. MR2375638 (2008h:60131).

Gardner RJ. The Brunn-Minkowski inequality. Bull Amer Math Soc (N S). 2002; 39:355–405. MR1898210 (2003f:26035).

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Gilks WR, Wild P. Adaptive Rejection Sampling for Gibbs Sampling. Applied Statistics. 1992; 41:337–348.

Gozlan N, Léonard C. Transport inequalities. A survey. Markov Process Related Fields. 2010; 16:635–736. MR2895086.

Guédon O. Concentration phenomena in high dimensional geometry. Proceedings of the Journées MAS. 2012; 2012

Guédon O, Milman E. Interpolating thin-shell and sharp largedeviation estimates for isotropic log-concave measures. Geom Funct Anal. 2011; 21:1043–1068. MR2846382.

Gurvits L. A short proof, based on mixed volumes, of Liggett's theorem on the convolution of ultra-logconcave sequences. Electron J Combin. 2009; 16 Note 5:5. MR2482101 (2010g:05400).

Hargé G. A particular case of correlation inequality for the Gaussian measure. Ann Probab. 1999; 27:1939–1951. MR1742895 (2001a:60043).

Hargé G. A convex/log-concave correlation inequality for Gaussian measure and an application to abstract Wiener spaces. Probab Theory Related Fields. 2004; 130:415–440. MR2095937 (2005k: 28029).

Hargé G. Reinforcement of an inequality due to Brascamp and Lieb. J Funct Anal. 2008; 254:267–300. MR2376572 (2009f:60025).

Henningsson T, Astrom KJ. Log-concave Observers. Proceedings of the 17th International Symposium on Mathematical Theory of Networks and Systems. 2006

Ibragimov IA. On the composition of unimodal distributions. Theor Probability Appl. 1956a; 1:255–266.

Ibragimov IA. On the composition of unimodal distributions. Teor Veroyatnost i Primenen. 1956b; 1:283–288. MR0087249 (19,326f).

Jensen JL. Uniform saddlepoint approximations and log-concave densities. J Roy Statist Soc Ser B. 1991; 53:157–172.

Jensen, JL. Saddlepoint approximations Oxford Statistical Science Series. Vol. 16. The Clarendon Press Oxford University Press, New York: Oxford Science Publications: 1995.

Johnson O. Log-concavity and the maximum entropy property of the Poisson distribution. Stochastic Process Appl. 2007; 117:791–802. MR2327839 (2008k:62012).

Johnson O, Barron A. Fisher information inequalities and the central limit theorem. Probab Theory Related Fields. 2004; 129:391–409. MR2128239 (2006b:62018).

Johnson O, Kontoyiannis I, Madiman M. Log-concavity, ultra-log-concavity, and a maximum entropy property of discrete compound Poisson measures. Discrete Appl Math. 2013; 161:1232–1250. MR3030616.

Jylänki P, Vanhatalo J, Vehtari A. Robust Gaussian process regression with a Student-t likelihood. J Mach Learn Res. 2011; 12:3227–3257. MR2877599 (2012k:62136).

Kahn J, Neiman M. A strong log-concavity property for measures on Boolean algebras. J Combin Theory Ser A. 2011; 118:1749–1760. MR2793607 (2012c:60009).

Kannan R, Lovász L, Simonovits M. Isoperimetric problems for convex bodies and a localization lemma. Discrete Comput Geom. 1995; 13:541–559. MR1318794 (96e:52018).

Kannan R, Lovász L, Simonovits M. Random walks and an $O^*(n^5)$ volume algorithm for convex bodies. Random Structures Algorithms. 1997; 11:1–50. MR1608200 (99h:68078).

Karlin, S. Total positivity. Vol. I. Stanford University Press; Stanford, Calif: 1968. MR0230102 (37 ##5667)

Keady G. The persistence of logconcavity for positive solutions of the one-dimensional heat equation. J Austral Math Soc Ser A. 1990; 48:246–263. MR1033179 (91b:35010).

Keilson J, Gerber H. Some results for discrete unimodality. J Amer Statist Assoc. 1971; 66:386–389.

Kelly RE. Stochastic reduction of loss in estimating normal means by isotonic regression. Ann Statist. 1989; 17:937–940. MR994278 (90c:62032).

Kim YH, Milman E. A generalization of Caffarelli's contraction theorem via (reverse) heat flow. Math Ann. 2012; 354:827–862. MR2983070.

Kim, AKH.; Samworth, RJ. Global rates of convergence in log-concave density estimation Technical Report. Statistical Laboratory, Cambridge University; 2014. available as arXiv:1404.2298v1

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Klaassen CAJ. On an inequality of Chernoff. Ann Probab. 1985; 13:966–974. MR799431 (86k:60029).

Kolesnikov AV. On diffusion semigroups preserving the log-concavity. J Funct Anal. 2001; 186:196–205. MR1863297 (2002g:47089).

Kontoyiannis I, Harremoës P, Johnson O. Entropy and the law of small numbers. IEEE Trans Inform Theory. 2005; 51:466–472. MR2236061 (2007j:94025).

Korevaar N. Capillary surface convexity above convex domains. Indiana Univ Math J. 1983a; 32:73–81. MR684757 (84b:35054).

Korevaar NJ. Capillary surface continuity above irregular domains. Comm Partial Differential Equations. 1983b; 8:213–245. MR692284 (84f:53005).

Ledoux M. L'algèbre de Lie des gradients itérés d'un générateur markovien—développements de moyennes et entropies. Ann Sci École Norm Sup. 1995; 28(4):435–460. MR1334608 (96f:60130).

Ledoux, M. Lectures on probability theory and statistics (Saint-Flour, 1994) Lecture Notes in Math. Vol. 1648. Springer; Berlin: 1996. Isoperimetry and Gaussian analysis; p. 165-294.MR1600888 (99h:60002)

Ledoux, M. The concentration of measure phenomenon Mathematical Surveys and Monographs. Vol. 89. American Mathematical Society; Providence, RI: 2001. MR1849347 (2003k:28019)

Leindler L. On a certain converse of Hölder's inequality. II. Acta Sci Math (Szeged). 1972; 33:217–223. MR2199372.

Liggett TM. Ultra logconcave sequences and negative dependence. J Combin Theory Ser A. 1997; 79:315–325. MR1462561 (98j:60018).

Linnik YV. An information theoretic proof of the central limit theorem with the Lindeberg condition. Theory Probab Appl. 1959; 4:288–299.

Lovász L, Simonovits M. Random walks in a convex body and an improved volume algorithm. Random Structures Algorithms. 1993; 4:359–412. MR1238906 (94m:90091).

Lovász, L.; Vempala, S. The geometry of logconcave functions and an O*(n3) sampling algorithm Technical Report, Technical Report MSR-TR-2003-04. Microsoft Research; 2003.

Lovasz L, Vempala S. Fast Algorithms for Logconcave Functions: Sampling, Rounding, Integration and Optimization. Foundations of Computer Science, 2006 FOCS '06 47th Annual IEEE Symposium on 57-68. 2006

Lovász L, Vempala S. The geometry of logconcave functions and sampling algorithms. Random Structures Algorithms. 2007; 30:307–358. MR2309621 (2008h:94032).

Marshall, AW.; Olkin, I. Inequalities: theory of majorization and its applications Mathematics in Science and Engineering. Vol. 143. Academic Press Inc. [Harcourt Brace JovanovichPublishers]; New York: 1979. MR552278 (81b:00002)

Marshall, AW.; Olkin, I.; Arnold, BC. Springe Series in Statistics. second. Springer; New York: 2011. Inequalities: theory of majorization and its applications. MR2759813

Maurey B. Some deviation inequalities. Geom Funct Anal. 1991; 1:188–197. MR1097258 (92g:60024).

Maurey B. Inégalité de Brunn-Minkowski-Lusternik, et autres inégalités géométriques et fonctionnelles. Astérisque. 2005; 299 Exp. 928 vii, 95–113. Séminaire Bourbaki. Vol. 2003/2004. MR2167203 (2006g:52006).

McCann RJ. Existence and uniqueness of monotone measure-preserving maps. Duke Math J. 1995; 80:309–323. MR1369395 (97d:49045).

Menz G, Otto F. Uniform logarithmic Sobolev inequalities for conservative spin systems with super-quadratic single-site potential. Annals of Probability. 2013; 41:2182–2224.

Minka, TP. PhD thesis. Massachusetts Institute of Technology; 2001. A family of algorithms for approximate Bayesian inference.

Nayar P, Oleszkiewicz K. Khinchine type inequalities with optimal constants via ultra log-concavity. Positivity. 2012; 16:359–371. MR2929095.

Neal RM. Slice sampling. Ann Statist. 2003; 31:705–767. With discussions and a rejoinder by the author.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Niculescu, CP.; Persson, LE. Convex functions and their applications CMS Books in Mathematics/ Ouvrages de Mathématiques de la SMC. Vol. 23. Springer; New York: 2006. A contemporary approach. MR2178902 (2006m:26001)

Olkin, I.; Tong, YL. Statistical decision theory and related topics, IV, Vol 2 (West Lafayette, Ind, 1986). Springer; New York: 1988. Peakedness in multivariate distributions; p. 373-383.MR927147 (89c:62086)

Paninski L. Log-concavity results on Gaussian process methods for supervised and unsupervised learning. Advances in Neural Information Processing Systems. 2004:1025–1032.

Pitt LD. A Gaussian correlation inequality for symmetric convex sets. Ann Probability. 1977; 5:470–474. MR0448705 (56 ##7010).

Polson, NG. Bayesian statistics. Vol. 5. Oxford Univ. Press; New York: 1996. Convergence of Markov chain Monte Carlo algorithms. (Alicante, 1994). Oxford Sci. Publ. 297–321MR1425412 (97i: 60086)

Port SC, Stone CJ. Fisher information and the Pitman estimator of a location parameter. Ann Statist. 1974; 2:225–247. MR0362665 (50 ##15105).

Prékopa A. Logarithmic concave measures with application to stochastic programming. Acta Sci Math (Szeged). 1971; 32:301–316. MR0315079 (47 ##3628).

Prékopa A. On logarithmic concave measures and functions. Acta Sci Math (Szeged). 1973; 34:335–343. MR0404557 (53 ##8357).

Prékopa, A. Stochastic programming Mathematics and its Applications. Vol. 324. Kluwer Academic Publishers Group; Dordrecht: 1995. MR1375234 (97f:90001)

Proschan F. Peakedness of distributions of convex combinations. Ann Math Statist. 1965; 36:1703–1706. MR0187269 (32 ##4722).

Rinott Y. On convexity of measures. Ann Probability. 1976; 4:1020–1026. MR0428540 (55 ##1561).

Roberts GO, Rosenthal JS. The polar slice sampler. Stoch Models. 2002; 18:257–280. MR1904830 (2003f:65011).

Rockafellar, RT.; Wets, RJB. Variational analysis Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]. Vol. 317. Springer-Verlag; Berlin: 1998. MR1491362 (98m:49001)

Rudin, W. Real and complex analysis. Third. McGraw-Hill Book Co.; New York: 1987. MR924157 (88k:00002)

Rudolf D. Explicit error bounds for Markov chain Monte Carlo. Dissertationes Math (Rozprawy Mat). 2012; 485:1–93. MR2977521.

Saumard, A.; Wellner, JA. Efron's monotonicity theorem and asymmetric Brascamp-Lieb type inequalities Technical Report. Department of Statistics, University of Washington; 2014. In preparation

Schoenberg IJ. On Pólya frequency functions. I. The totally positive functions and their Laplace transforms. J Analyse Math. 1951; 1:331–374. MR0047732 (13,923c).

Seeger M. Gaussian processes for machine learning. International Journal of Neural Systems. 2004; 14:69–106. [PubMed: 15112367]

Seregin A, Wellner JA. Nonparametric estimation of multivariate convex-transformed densities. Ann Statist. 2010; 38:3751–3781. With supplementary material available online. MR2766867 (2012b: 62126).

Sherman S. A theorem on convex sets with applications. Ann Math Statist. 1955; 26:763–767. MR0074845 (17,655l).

Simon, B. Convexity Cambridge Tracts in Mathematics. Vol. 187. Cambridge University Press; Cambridge: 2011. An analytic viewpoint. MR2814377

van der Vaart AW, van Zanten JH. Rates of contraction of posterior distributions based on Gaussian process priors. Ann Statist. 2008; 36:1435–1463. MR2418663 (2009i:62068).

van der Vaart A, van Zanten H. Information rates of non-parametric Gaussian process methods. J Mach Learn Res. 2011; 12:2095–2119. MR2819028 (2012g:62176).

Vapnik, VN. The nature of statistical learning theory, second ed Statistics for Engineering and Information Science. Springer-Verlag; New York: 2000. MR1719582 (2001c:68110)

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Villani, C. Topics in optimal transportation Graduate Studies in Mathematics. Vol. 58. American Mathematical Society; Providence, RI: 2003. MR1964483 (2004e:90003)

Villani, C. Optimal transport Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]. Vol. 338. Springer-Verlag; Berlin. Old and new: 2009. MR2459454 (2010f:49001)

Walkup DW. Pólya sequences, binomial convolution and the union of random sets. J Appl Probability. 1976; 13:76–85. MR0494391 (58 ##13258).

Walther G. Detecting the presence of mixing with multiscale maximum likelihood. J Amer Statist Assoc. 2002; 97:508–513. MR1941467 (2003k:62152).

Walther G. Inference and modeling with log-concave distributions. Statist Sci. 2009; 24:319–327. MR2757433 (2011j:62110).

Wang Z, Louis TA. Matching conditional and marginal shapes in binary random intercept models using a bridge distribution function. Biometrika. 2003; 90:765–776. MR2024756 (2004j:62135).

Wellner, JA. High dimensional probability VI: the Banff volume Progresss in Probability. Vol. 66. Birkhauser; Basel: 2013. Strong log-concavity is preserved by convolution; p. 95-102.

Whitt W. Uniform conditional variability ordering of probability distributions. J Appl Probab. 1985; 22:619–633. MR799285 (86j:60062).

Yuille A, Rangarajan A. The concave-convex procedure (CCCP). Neural Computation. 2003; 15:915–936. [PubMed: 12689392]

Zhang Z, Dai G, Jordan MI. Bayesian generalized kernel mixed models. J Mach Learn Res. 2011; 12:111–139. MR2773550 (2012e:62136).

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

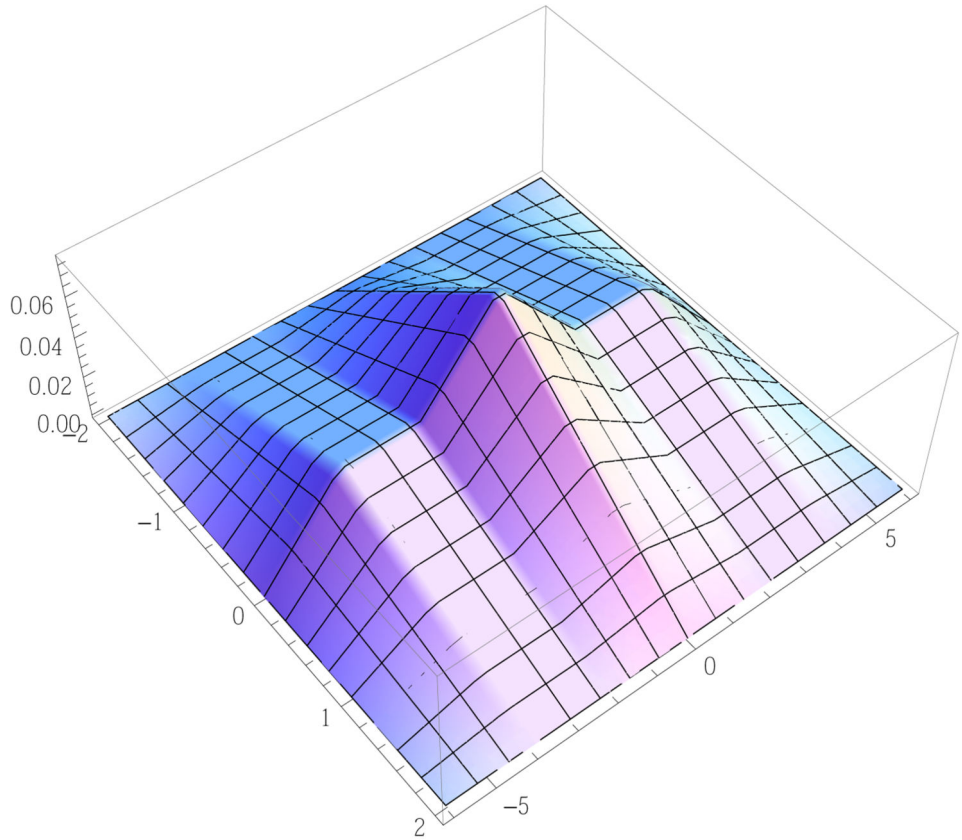

Fig 1. Sherman's example, h = p * q