

Supplementary Material for “Estimating a discrete log-concave distribution in higher dimensions”

Hanna Jankowski and Amanda Tian

November 16, 2017

Appendices

A Appendix: some background from convex analysis

A.1 Discrete

A density function f is said to be log-concave if $(-\log f)(x)$ is a convex function on \mathbb{R}^d . Recall that on \mathbb{R}^d , a function h is convex if for all $x, x' \in \mathbb{R}^d$ and for all $\alpha \in [0, 1]$ it satisfies $h(\alpha x + (1 - \alpha)x') \leq \alpha h(x) + (1 - \alpha)h(x')$. Furthermore, if h is twice differentiable, then h is concave if and only if $h''(x) \geq 0$ for all $x \in \mathbb{R}$ and if and only if the Hessian matrix of h is positive semi-definite for all $x \in C$, where C is an open convex set on \mathbb{R}^d for $d > 1$ (Rockafellar, 1970, Theorem 4.5, page 27).

Similarly, one can define convex functions in the one-dimensional discrete setting, which naturally leads to a definition of log-concave probability mass functions. That is, let $p(z) : \mathbb{Z} \rightarrow [0, 1]$ denote a probability mass function (PMF), where \mathbb{Z} denotes the integers

$\{\dots, -2, -1, 0, 1, 2, \dots\}$. The PMF p is said to be log-concave if for any $z \in \mathbb{Z}$

$$(\Delta h)(z) = h(z-1) - 2h(z) + h(z+1) \geq 0, \quad (1)$$

where $h(z) = (-\log p)(z)$ (Balabdaoui et al., 2013, Proposition 1). In the notation above (Δh) denotes the discrete Laplacian operator, which can also be expressed as $(\Delta h)(z) = \{h(z+1) - h(z)\} - \{h(z) - h(z-1)\}$. This is the second difference of the function h , and hence this definition matches well that of the continuous setting.

Perhaps surprisingly at first, in higher dimensions, the definition of a discrete convex (equivalently, concave) function is not so straightforward. For a discrete function defined on \mathbb{Z}^d for $d > 1$ there are multiple definitions of convexity. Murota and Shioura (2001) provide a detailed survey of convex functions and sets in the higher-dimensional discrete setting, including a summary of the relationships between the various definitions. Among these definitions there are three which are relevant to our initial considerations: discretely-convex, separable-convex, and convex-extendible. To this end, consider a function $h : \mathbb{Z}^d \rightarrow \mathbb{R} \cup \{+\infty\}$ and define the domain $\text{dom}(h) = \{z \in \mathbb{Z}^d \mid h(z) < \infty\}$.

- The function h is said to be **separable-convex** if $h(z) = \sum_{i=1}^d h_i(z_i)$ ($z \in \mathbb{Z}^d$) for some family of convex functions $h_i : \mathbb{Z} \rightarrow \mathbb{R} \cup \{+\infty\}$, $i \in \{1, \dots, d\}$. That is, $(\Delta h_i)(z) \geq 0$ for all $z \in \mathbb{Z}$ and all $i \in \{1, \dots, d\}$.
- For $x \in \mathbb{R}^d$, let $\lfloor x \rfloor$ (respectively, $\lceil x \rceil$) denote the floor (respectively, the ceiling) of the vector x , obtained by rounding down (respectively, up) each component of x to its nearest integer. Next, define the set $N_0(x) = \{z \in \mathbb{Z}^d \mid \lfloor x \rfloor \leq z \leq \lceil x \rceil\}$. The function h is said to be **discretely-convex** if, for any $z', z'' \in \text{dom}(h)$ and any $\alpha \in [0, 1]$, it holds

that

$$\min\{h(z) \mid z \in N_0(\alpha z' + (1 - \alpha)z'')\} \leq \alpha h(z') + (1 - \alpha)h(z'').$$

Similarly, a set $S \subseteq \mathbb{Z}^d$ is said to be discretely-convex if, for any $z', z'' \in S$ and any $\alpha \in [0, 1]$, it holds that $N_0(\alpha z' + (1 - \alpha)z'') \cap S$ is non-empty.

– Define the convex closure of $h(z)$

$$\bar{h}(x) = \sup_{\alpha \in \mathbb{R}, \beta \in \mathbb{R}^d} \{\alpha + \beta^T x : \alpha + \beta^T z \leq h(z) \text{ for all } z \in \mathbb{Z}^d\}, \quad x \in \mathbb{R}^d.$$

The function h is **convex-extendible** if $\bar{h}(z) = h(z)$ for all $z \in \mathbb{Z}^d$. Similarly, a set $S \subseteq \mathbb{Z}^d$ is said to be convex-extendible if $\bar{S} \cap \mathbb{Z}^d = S$, where $\bar{S} \subseteq \mathbb{R}^d$ is the convex closure of S , that is, it is the smallest closed convex set (in \mathbb{R}^d) containing S . A related notion which will be useful later is that of a convex extension: A convex function $h^R : \mathbb{R}^d \rightarrow \mathbb{R} \cup \{+\infty\}$ is called a convex extension of h if $h^R(z) = h(z)$ for all $z \in \mathbb{Z}^d$. Clearly, a convex closure is a convex extension, but not vice versa.

Murota and Shioura (2001) summarize the relationships between the various definitions of convexity. In particular, some but not all discretely-convex functions are convex-extendible functions and vice versa, while separable-convex functions are both discrete-convex and convex-extendible. Indeed, consider the set

$$\begin{aligned} S &= \{z \in \mathbb{Z}^3 \mid z_1 + z_2 + z_3 = 2, z_i \geq 0, i = 1, 2, 3\} \cup \{(1, 2, 0), (0, 1, 2), (2, 0, 1)\} \\ &= \{(0, 1, 1), (1, 0, 1), (1, 1, 0), (0, 0, 2), (0, 2, 0), (2, 0, 0), (1, 2, 0), (0, 1, 2), (2, 0, 1)\}. \end{aligned}$$

This set, as well as the function h equal to zero on S and $+\infty$ on $\mathbb{Z} \setminus S$, are discrete-convex.

However,

$$\frac{1}{3}(1, 2, 0) + \frac{1}{3}(0, 1, 2) + \frac{1}{3}(2, 0, 1) = (1, 1, 1)$$

is an element of $\bar{S} \cap \mathbb{Z}^d$, but $(1, 1, 1) \notin S$, and hence h is not convex-extendible. On the other hand, let $S = \{(0, 0), (2, 1)\}$ and again define the function h equal to zero on S and $+\infty$ on $\mathbb{Z} \setminus S$. The convex closure of S is the segment between points $(0, 0)$ and $(2, 1)$, hence $\bar{S} \cap \mathbb{Z}^d = \{(0, 0), (2, 1)\} = \{z_1, z_2\} = S$, we conclude that h is convex-extendible. On the other hand, $N_0(0.5z_1 + 0.5z_2) = N_0((1, 0.5)) = \{(1, 0), (1, 1)\}$, whence $N_0(\alpha x' + (1 - \alpha)x'') \cap S = \emptyset$ and h is not discrete-convex. Both examples appear in Murota and Shioura (2001).

A.2 Continuous

Theorem A.1. (Rockafellar, 1970, Theorem 10.6, page 88) *Let \mathcal{C} be a relatively open convex set, and let $\{f_i | i \in \mathcal{I}\}$ be an arbitrary collection of convex functions finite and pointwise bounded on \mathcal{C} . Let \mathcal{S} be any closed bounded subset of \mathcal{C} . Then $\{f_i | i \in \mathcal{I}\}$ is uniformly bounded on \mathcal{S} and equi-Lipschitzian relative to \mathcal{S} .*

The conclusion remains valid if the pointwise boundedness assumption is weakened to the following pair of assumptions:

- (a). *There exists a subset \mathcal{C}' of \mathcal{C} such that $\text{conv}(\text{cl } \mathcal{C}') \supset \mathcal{C}$ and $\sup\{f_i(x) | i \in \mathbb{I}\}$ is finite for every $x \in \mathcal{C}'$;*
- (b). *There exists at least one $x \in \mathcal{C}$ such that $\inf\{f_i(x) | i \in \mathbb{I}\}$ is finite.*

Theorem A.2. (Rockafellar, 1970, Theorem 32.2, page 343) *Let f be a convex function, and let $\mathcal{C} = \text{conv } \mathcal{S}$, where \mathcal{S} is an arbitrary set of points. Then*

$$\sup\{f(x) | x \in \mathcal{C}\} = \sup\{f(x) | x \in \mathcal{S}\},$$

where the first supremum is attained only when the second (more restrictive) supremum is attained.

B Appendix: Proofs

B.1 Proofs from Sections 2.1 and 2.2

Proof of Proposition 2.1. By definition, for any $z \in \mathcal{S}$, we have that

$$h(z) = -\log p(z) = \sum_{i=1}^d \{-\log p_i(z_i)\} = \sum_{i=1}^d h_i(z_i),$$

where each $h_i(z_i)$ is convex. Note that this depends on the support \mathcal{S} . For each i , let $\mathcal{S}_i = \{k \in \mathbb{Z} : (k)_i \in \mathcal{S}\}$, where $(k)_i$ denotes any point of \mathbb{Z}^d with its i th element equal to k . Let $\text{ext}(\mathcal{S}) = \bar{\mathcal{S}}_1 \times \cdots \times \bar{\mathcal{S}}_d$. Each function h_i is defined on \mathcal{S}_i , and let h_i^R denote its convex extension on $\bar{\mathcal{S}}_i$. The function $h^R = \sum_{i=1}^d h_i^R$ is convex extendible on $\text{ext}(\mathcal{S})$, since it is convex separable. Furthermore, if \mathcal{S} is itself convex extendible, the restriction of h^R to \mathcal{S} is also convex and closed, and satisfies $h^R(z) = h(z)$ on \mathcal{S} . This shows that $p \in \mathcal{P}_0$. \square

Proof of Proposition 2.2. Let f denote a log-concave density on \mathbb{R}^d . For $A = [-1/2, 1/2]^d$, consider the function $q(x) = \int_{x+A} f(y) dy = P(Y \in A + x)$, letting Y denote the random variable with density f . Then, by the property of log-concave distributions (see e.g. Dharmadhikari and Joag-Dev (1988, (2.6) on page 47)), for any $\alpha \in (0, 1)$ and any $x, y \in \mathbb{R}^d$ we have that

$$q(\alpha x + (1 - \alpha)y) \geq q(x)^\alpha q(y)^{1-\alpha}$$

which implies that the function $h^R(x) = -\log q(x)$ is convex. The function $q(x)$ is continuous

by properties of integrals (applying, for example, the dominated convergence theorem and the fact that f must be bounded). In fact, letting B denote an upper bound on f , we have that

$$|q(x) - q(y)| \leq B\lambda\{(A+x)\Delta(A+y)\} \leq 4d\lambda\{A\}B\|x-y\|_\infty,$$

where $\lambda\{A\}$ denotes the Lebesgue measure of the set A . It follows that $-\log q(x)$ is continuous on its effective domain, and therefore it is lower semi-continuous. Therefore, it is closed (Rockafellar, 1970, Theorem 7.1, page 51) on its effective domain. Lastly, $h^R(z) = -\log p(z)$ by definition on \mathbb{Z}^d . It follows that the restriction of h^R to \bar{S} is a closed convex extension of $-\log p(z)$, and hence $p \in \mathcal{P}_0$ by Murota and Shioura (2001, Lemma 2.3) (Lemma 2.1). \square

B.2 Proofs from Section 2.3

Lemma B.1. *Suppose $p_1, p_2 \in \mathcal{P}_0$. Then a PMF $p \propto (p_1 p_2)^\alpha$ for any $\alpha \in (0, 1)$ also satisfies $p \in \mathcal{P}_0$.*

Proof. Let $h_1 = -\log p_1, h_2 = -\log p_2$ and $h = \alpha(h_1 + h_2) + c$ (defined on \mathbb{Z}^d) for some appropriate constant $c \in \mathbb{R}$. Let h_1^R and h_2^R denote the closed convex extensions of h_1, h_2 (respectively), which exist by assumption. Then $h^R = \alpha(h_1^R + h_2^R) + c$ is closed, convex, and by definition satisfies $h^R(z) = h(z)$ on \mathbb{Z}^d . Therefore, $p \in \mathcal{P}_0$. \square

Proof of Proposition 2.3. 1. Let $h(z) = -\log p(z)$, then $h(z)$ is convex-extendible by assumption, and $\mathcal{S} = \{z \mid h(z) < \infty\}$. Hence, by Lemma 2.1 (Murota and Shioura, 2001), there exists a convex extension $h^R(x)$ of $h(z)$, which is a closed convex function on \mathbb{R}^d . Therefore, the effective domain of $h^R, \{x \mid h^R(x) < +\infty\}$, is a closed convex set in \mathbb{R}^d (Rockafellar, 1970, page 23 and Theorem 7.1 on page 51). The latter follows since for a closed function, its epigraph must be closed (Rockafellar, 1970, Theorem 7.1 on page

51) and the effective domain is the projection of the epigraph onto \mathbb{R}^d , (Rockafellar, 1970, page 23). Since such a projection of a closed set must be closed (appealing to the characterization of closed sets via Cauchy sequences), it follows that the effective domain is closed. Therefore, $\mathcal{S} \subset \bar{\mathcal{S}} \subset \{x \mid h^R(x) < +\infty\}$. Furthermore, we have that $\mathcal{S} = \mathbb{Z}^d \cap \{x \mid h^R(x) < +\infty\}$. Therefore, it follows that $\bar{\mathcal{S}} \cap \mathbb{Z}^d = \mathcal{S}$, and hence \mathcal{S} is convex extendible.

2. Let $h(z) = -\log p(z)$, then $h(z)$ is convex-extendible by assumption. By Lemma 2.1 (Murota and Shioura, 2001), there exist a convex extension $h^R(x)$ of $h(z)$, which is a closed convex function on \mathbb{R}^d . We define a function

$$\tilde{h}^R(x) = \begin{cases} h^R(x) - \log c, & x \in \bar{\mathcal{A}} \\ +\infty, & x \notin \bar{\mathcal{A}}, \end{cases}$$

for $c^{-1} = \sum_{z \in \mathcal{A}} p(z)$, and where $\bar{\mathcal{A}}$ denotes the convex closure of \mathcal{A} . It is obvious that \tilde{h}^R is also a closed convex function. Also, $-\log \tilde{p}(z) = \tilde{h}^R(z)$, for $z \in \mathcal{A} \subset \text{conv } \mathcal{A} \subset \bar{\mathcal{A}}$, and hence \tilde{h}^R is a convex extension of $-\log \tilde{p}$. Therefore $\tilde{p} \in \mathcal{P}_0$.

3. Letting $h_1(z_1) = -\log p_1(z_1)$, $h_2(z_2) = -\log p_2(z_2)$ then $h_1(z), h_2(z)$ are both convex-extendible by assumption. By Lemma 2.1 (Murota and Shioura, 2001), there exist convex extensions $h_1^R(x_1), h_2^R(x_2)$ respectively, of $h_1(z_1), h_2(z_2)$. These are closed convex functions on $\mathbb{R}^{d_1}, \mathbb{R}^{d_2}$. Next, $h^R(x_1, x_2) = h_1^R(x_1) + h_2^R(x_2)$ is also a convex function on $\mathbb{R}^{d_1+d_2}$. Furthermore, it is closed, since it is the sum of lower semi-continuous functions, and hence lower semi-continuous (Rockafellar, 1970, Theorem 7.1, page 51). Finally, $h^R(z) = -\log(p_1(z_1)p_2(z_2)) = -\log p(z)$, where $z = (z_1, z_2)$. Therefore $p \in \mathcal{P}_0$.
4. Let $h(z) = -\log p(z)$, then $h(z)$ is convex-extendible by assumption and fix $z_2 \in \mathbb{Z}^{d_2}$. By Lemma 2.1, there exists a convex extension $h^R(x)$ of $h(z)$, which is a closed convex

function on \mathbb{R}^d . Let p_2 denote the marginal of p : $p_2(z_2) = \sum_{z_1 \in \mathbb{Z}^{d_1}} p(z_1, z_2)$. We then define $\tilde{h}^R(x_1) = h^R(x_1, x_2 = z_2) + \log p_2(z_2)$, where $x_1 \in \mathbb{R}^{d_1}$, and $z_2 \in \mathbb{Z}^{d_2} \subset \mathbb{R}^{d_2}$ is fixed. We will show that \tilde{h}^R is the convex extension of $-\log p(z_1|z_2)$, and therefore $p(z_1|z_2)$ is eLC.

Firstly, we have that for any $z_1 \in \mathbb{Z}^{d_1}$

$$\begin{aligned} \tilde{h}^R(z_1) &= h^R(z_1, z_2) + \log p_2(z_2) = -\log p(z_1, z_2) + \log p_2(z_2) \\ &= -\log p(z_1|z_2). \end{aligned}$$

Secondly, $\tilde{h}^R(x_1)$ is convex since $h^R(x)$ is convex in x_1 and $\log p_2(z_2)$ is a constant. Finally, we need that \tilde{h}^R is closed. This follows from Rockafellar (1970, Theorem 7.1, page 51) by appealing to the definition of closed sets via Cauchy sequences.

5. Let $h(z) = -\log p(z)$, then $h(z)$ is convex-extendible by assumption. Hence, by Lemma 2.1 (Murota and Shioura, 2001), there exists a convex extension $h^R(x)$ of $h(z)$, which is a closed convex function. Note that $\tilde{p}(z) = p(A^{-1}(z - b))$ for any $z \in \mathbb{Z}^d$. We then construct $\tilde{h}^R(x) = h^R(A^{-1}(x - b))$, for any $x \in \mathbb{R}^d$. Clearly, \tilde{h}^R is also convex and closed. Moreover, $\tilde{h}^R(z) = h^R(A^{-1}(z - b)) = h(A^{-1}(z - b)) = -\log p(A^{-1}(z - b)) = -\log \tilde{p}(z)$, for any $z \in \mathbb{Z}$. Hence \tilde{h}^R is the convex extension of \tilde{p} , and therefore $\tilde{p} \in \mathcal{P}_0$. □

Proof of Theorem 2.1. Define $\mathcal{S}_0 = \{z \in \mathbb{Z}^d \mid p(z) > 0\}$ and assume (for the moment) that $\mathcal{S}_0 = \mathbb{Z}^d$. Define also $h_n(z) = -\log p_n(z)$, for each $n \geq 1$ and $h(z) = -\log p(z)$. By assumption, h_n is convex-extendible, and converges to h pointwise on \mathcal{S}_0 . To prove that p is eLC, we need to show that h is convex-extendible. To do this, we will use Lemma 2.1 (Murota and Shioura, 2001), and find a closed convex extension of h .

By Lemma 2.1 (Murota and Shioura, 2001), there exists a closed convex extension of h_n , for each n . We denote this by $h_n^R : \mathbb{R}^d \rightarrow \mathbb{R} \cup \{+\infty\}$. By definition, h_n^R is a closed convex function, and $h_n^R(z) = h_n(z)$ for any $z \in \mathbb{Z}^d$.

Fix $K \in \mathbb{Z}^+$ to be a large, positive integer, and let $\mathcal{B}_K = \{x \in \mathbb{R}^d : \|x\|_\infty \leq K\}$, a closed (in \mathbb{R}^d) and bounded set. Since $p_n \rightarrow p$ for all $z \in \mathbb{Z}^d$, there exists an n_0 such that for all $n \geq n_0$, $p_n(z) > 0$, and hence $h_n(z) < \infty$ for all $z \in \mathcal{B}_K$.

Note that \mathcal{B}_K is a subset of \mathbb{R}^d , and also the convex hull of $\mathcal{B}_K \cap \mathbb{Z}^d$ (in \mathbb{R}^d). Since each h_n^R is closed and convex, we can apply Theorem D.2 (Rockafellar, 1970) in the Appendix, and conclude that for each n ,

$$\sup_{x \in \mathcal{B}_K} h_n^R(x) \leq \sup_{z \in \mathcal{B}_K \cap \mathbb{Z}^d} h_n^R(z) = \sup_{z \in \mathcal{B}_K \cap \mathbb{Z}^d} h_n(z).$$

Therefore,

$$\sup_{n \geq n_0} \sup_{x \in \mathcal{B}_K} h_n^R(x) \leq \sup_{n \geq n_0} \sup_{z \in \mathcal{B}_K \cap \mathbb{Z}^d} h_n(z) = M_{K, n_0}, \quad (2)$$

where M_{K, n_0} is finite because $S_0 = \mathbb{Z}^d$. Therefore the sequence $\{h_n^R(x)\}_{n \geq n_0}$ is finite and pointwise bounded (uniformly) for all $x \in \mathcal{B}_K$. The statement continues to hold on the relative interior of \mathcal{B}_K (again, in \mathbb{R}^d), which we denote $\text{rl } \mathcal{B}_K$. By Rockafellar (1970, Theorem 10.6, page 88), Theorem D.1 in the Appendix, we conclude that $h_n^R(x)$ is uniformly bounded and equi-Lipschitzian relative to, say, \mathcal{B}_{K-1} . By the Arzelà-Ascoli theorem, we conclude that h_n^R is compact and hence there is a subsequence of h_n^R that converges uniformly on $\mathcal{B}_{(K-1)}$. We denote this subsequence as $h_{n_K}^R$, and its limit as h^R .

We now argue that $h^R(x)$ is a convex extension of $h(z)$ on $\mathcal{B}_{(K-1)}$:

- Since h^R is the limit of a sequence of convex functions defined on $\mathcal{B}_{(K-1)}$, it follows that h^R is convex on $\mathcal{B}_{(K-1)}$.

- By definition, $h^R(z) = \lim_{n_K \rightarrow \infty} h_{n_K}^R(z) = \lim_{n_K \rightarrow \infty} h_{n_K}(z) = h(z)$, for any $z \in \mathcal{B}_{(K-1)}$.
- For any K , $h_{n_K}^R(x)$ is finite by inequality (2). We also know that it is continuous and uniformly converges to $h^R(x)$ on $\mathcal{B}_{(K-1)}$. Hence $h^R(x)$ is also finite, and continuous on $\mathcal{B}_{(K-1)}$ by Krantz (1991, Theorem 9.1, page 201), and therefore $h^R(x)$ is closed on $\mathcal{B}_{(K-1)}$ by the definition of continuous functions (Krantz, 1991, Theorem 6.9, page 142).

Hence we can conclude that $h^R(x)$ is a closed convex extension of $h(z)$ on $\mathcal{B}_{(K-1)}$. Therefore $h(z)$ is convex-extendible by Murota and Shioura (2001, Lemma 2.3). Recall that $h(z) = -\log p(z)$, we conclude that $p(z)$ is also eLC for $z \in \mathcal{B}_{(K-1)}$. Since the above conclusion is true for any $K \in \mathbb{Z}^+$, therefore $p(z)$ is eLC for $z \in \mathbb{Z}^d$.

Now we consider the situation that $\mathcal{S}_0 \subset \mathbb{Z}^d$. Let $\bar{\mathcal{S}}_0$ denote the convex closure in \mathbb{R}^d of \mathcal{S}_0 . We may repeat the above proof, but considering $\mathcal{B}_K \cap \bar{\mathcal{S}}_0$ instead of \mathcal{B}_K throughout. On each $z \in \mathcal{B}_K \cap \bar{\mathcal{S}}_0$, we will eventually have that $p_n(z) > 0$ for sufficiently large n , and $\mathcal{B}_K \cap \bar{\mathcal{S}}_0$ is closed and convex by definition. The proof may now be repeated as above, and h will be convex-extendible on \mathcal{S}_0 . \square

Proof of Theorem 2.2. We first prove the existence part of the theorem. Without loss of generality, we assume that the support of p_0 , $\mathcal{S}_0 = \mathbb{Z}^d$. Let $\tilde{q} \propto e^{-\|z\|_\infty}$, where $z \in \mathbb{Z}^d$, such that $\tilde{q} \neq \hat{p}_0$ (if $\hat{p}_0 \propto e^{-\|z\|_\infty}$, then we can put $\tilde{q} \propto e^{-0.5\|z\|_\infty}$ instead, say). Note that $-\log \tilde{q}(z) = \|z\|_\infty$, and since all norms on \mathbb{R}^d are closed convex functions, $\|x\|_\infty$, $x \in \mathbb{R}^d$ is a convex extension of $\|z\|_\infty$. Hence \tilde{q} is eLC by Murota and Shioura (2001, Lemma 2.3).

We can also show that $\rho_{KL}(\tilde{q} \parallel p_0) < \infty$.

$$\begin{aligned}
\rho_{KL}(\tilde{q} \parallel p_0) &= \sum_{z \in \mathbb{Z}^d} p_0 \log p_0(z) - \sum_{z \in \mathbb{Z}^d} p_0 \log \tilde{q}(z) \\
&= \sum_{z \in \mathbb{Z}^d} p_0 \log p_0(z) - \sum_{z \in \mathbb{Z}^d} \|z\|_\infty p_0(z) < \infty.
\end{aligned}$$

Hence, $\inf_{q \in eLC} \rho_{KL}(q \parallel p_0) < \infty$.

Therefore, there exists a sequence of eLC PMFs $\{q_n\}$, such that

$$\rho_{KL}(q_n \parallel p_0) \rightarrow \inf_{q \in eLC} \rho_{KL}(q \parallel p_0).$$

Because $\inf_{q \in \mathcal{LC}} \rho_{KL}(q \parallel p_0) < \rho_{KL}(\tilde{q} \parallel p_0)$, there exists an $N > 0$, such that for all $n > N$, we have

$$\rho_{KL}(q_n \parallel p_0) \leq \rho_{KL}(\tilde{q} \parallel p_0).$$

Hence,

$$\sup_{n > N} \sum_{z \in \mathbb{Z}^d} |-\log q_n(z)| p_0(z) \leq \sum_{z \in \mathbb{Z}^d} |-\log \tilde{q}(z)| p_0(z) = \sum_{z \in \mathbb{Z}^d} \|z\|_\infty p_0(z) < \infty.$$

Let $M > 0$ and consider $\mathcal{S}_M = \{z : \|z\|_\infty \leq M\}$. Let $\alpha_M = \min_{z \in \mathcal{S}_M} p_0(z)$, and note that as $M \rightarrow \infty$, we have that $\alpha_M \rightarrow 0$, since p_0 is summable. It follows that

$$\begin{aligned} \sup_{n > N} \sum_{z \in \mathcal{S}_M} |-\log q_n(z)| &\leq \left\{ \max_{z \in \mathcal{S}_M} \frac{1}{p_0(z)} \right\} \left\{ \sup_{n > N} \sum_{z \in \mathcal{S}_M} |-\log q_n(z)| p_0(z) \right\} \\ &\leq \left\{ \max_{z \in \mathcal{S}_M} \frac{1}{p_0(z)} \right\} \left\{ \sup_{n > N} \sum_{z \in \mathbb{Z}^d} |-\log q_n(z)| p_0(z) \right\} \leq \frac{B}{\alpha_M}, \end{aligned}$$

where $B = E_{p_0}[\|Z\|_\infty] < \infty$. Hence, $\sup_{n > N} \sup_{z \in \mathcal{S}_M} |-\log q_n(z)| < B/\alpha_M$, and hence $\inf_{n > N} \min_{x \in \mathcal{S}_M} q_n(z) > e^{-B/\alpha_M} = \delta_M$.

Furthermore, we can find an integer $M_1 > M$ large enough so that

$$\sup_{n > N} \sup_{z \in \mathcal{S}_{M_1}^c} q_n(z) \leq \delta_M/2.$$

Therefore, we can find an envelope function $e^{l(z)}$, where $l(z) = -\alpha\|z\|_\infty + \beta$ with $\alpha, \beta \in \mathbb{R}^+$, such that $\sup_{n>N} q_n(z) \leq e^{l(z)}$.

Let X_n be a sequence of random vectors with PMF q_n . Since $e^{l(z)}$ is summable it follows that X_n is tight. Hence, there exists a convergent subsequence q_{n_i} , and a limit point q_0 . As q_n is eLC, by Theorem 2.1 q_0 is also eLC.

By Fatou's lemma, we have

$$\rho_{KL}(q_0 \parallel p_0) = \sum_{z \in \mathbb{Z}^d} p_0 \log \frac{p_0}{q_0} \leq \liminf_{n_i} \sum_{z \in \mathbb{Z}^d} p_0 \log \frac{p_0}{q_{n_i}} = \liminf_{n_i} \rho_{KL}(q_{n_i} \parallel p_0).$$

Since $\rho_{KL}(q_{n_i} \parallel p_0) \rightarrow \inf_{q \in eLC} \rho_{KL}(q \parallel p_0)$, we have $\rho_{KL}(q_0 \parallel p_0) \leq \inf_{q \in eLC} \rho_{KL}(q \parallel p_0)$. That is, a minimizer \hat{p}_0 exists, and the proof of existence is done.

We now prove uniqueness. Let's assume that \hat{p}_1, \hat{p}_2 are both eLC and minimize $\rho_{KL}(\cdot \parallel p_0)$. Let $\tilde{p} \propto (\hat{p}_1 \hat{p}_2)^{1/2}$ is a proper PMF. Note that by Lemma B.1, \tilde{p} is also eLC. Now,

$$\begin{aligned} \rho_{KL}(\tilde{p} \parallel p_0) &= (1/2) \sum p_0 \log \frac{p_0}{\hat{p}_1} + (1/2) \sum p_0 \log \frac{p_0}{\hat{p}_2} + \log \sum (\hat{p}_1 \hat{p}_2)^{1/2} \\ &= \rho_{KL}(\hat{p}_1 \parallel p_0) + \log \sum (\hat{p}_1 \hat{p}_2)^{1/2} \leq \rho_{KL}(\hat{p}_1 \parallel p_0). \end{aligned}$$

The last inequality follows that $\log \sum (\hat{p}_1 \hat{p}_2)^{1/2} \leq 0$ by Cauchy-Schwarz. However, since $\rho_{KL}(\tilde{p} \parallel p_0) \geq \rho_{KL}(\hat{p}_1 \parallel p_0)$, we find that $\sum (\hat{p}_1 \hat{p}_2)^{1/2} = \sum \hat{p}_1 \sum \hat{p}_2$. Therefore $\hat{p}_1 = \hat{p}_2$, again by Cauchy-Schwarz. This completes the proof. \square

Proof of Lemma 2.2. Let $\hat{\mathcal{S}}_0$ denote the support of \hat{p}_0 . Our goal is to show that $\hat{\mathcal{S}}_0 = \tilde{\mathcal{S}}_0 = \text{conv}(\mathcal{S}_0) \cap \mathbb{Z}^d$. Here, we denote the convex hull of \mathcal{S}_0 as $\text{conv } \mathcal{S}_0$, and note that by assumption, this is closed.

Note first that if $p_0(z_0) > 0$, then $\hat{p}_0(z_0) > 0$ (we call this fact one). This follows directly from the form of the KL divergence, as PMFs with support strictly smaller than that of p_0

have an infinite KL divergence, and can therefore not act as minimizers. We thus have that $\mathcal{S}_0 \subset \widehat{\mathcal{S}}_0$.

Next, consider $z_0 \in \widetilde{\mathcal{S}}_0$, such that $p_0(z_0) = 0$. Then by Carathéodory's Theorem (Rockafellar, 1970, Theorem 17.1 page 155) we can write $z_0 = \sum_{i=1}^{d+1} \lambda_i z_i$, where $\lambda_i > 0$, $\sum_{i=1}^{d+1} \lambda_i = 1$ and $z_i \in \mathcal{S}_0$ for each $i = 1, \dots, d+1$. Since \widehat{p}_0 is eLC and therefore $\log \widehat{p}_0$ has a concave extension equal to $\log \widehat{p}_0$ on \mathbb{Z}^d , we find that

$$\log \widehat{p}_0(z_0) \geq \sum_{i=1}^{d+1} \lambda_i \log \widehat{p}_0(z_i).$$

But then $\widehat{p}_0(z_0) = 0$ implies that $\log \widehat{p}_0(z_0) = -\infty$ and hence $\widehat{p}_0(z_i) = 0$ for at least one $1 \leq i \leq d+1$, a direct contradiction with fact one above. It follows that $\widetilde{\mathcal{S}}_0 \setminus \mathcal{S}_0 \subset \widehat{\mathcal{S}}_0$. Together with fact one, this yields $\widetilde{\mathcal{S}}_0 \subset \widehat{\mathcal{S}}_0$.

Finally, consider a $z_0 \in \mathbb{Z}^d$ such that $\widehat{p}_0(z_0) > 0$ for some $z_0 \notin \widetilde{\mathcal{S}}_0$. Construct a PMF

$$\widetilde{p}(z) = \begin{cases} c \widehat{p}_0(z) & z \in \widetilde{\mathcal{S}}_0 \\ 0 & z \notin \widetilde{\mathcal{S}}_0 \end{cases}$$

where c denotes an appropriate normalizing constant. By Proposition 2.3, \widetilde{p} is also eLC. Also, note that $c > 1$ by assumption. Then,

$$\begin{aligned} \rho_{KL}(\widehat{p}_0 \parallel p_0) &= \rho_{KL}(\widetilde{p} \parallel p_0) + \sum_{z \in \mathcal{S}_0} p_0(z) \{\log \widetilde{p}(z) - \log \widehat{p}_0(z)\} \\ &> \rho_{KL}(\widetilde{p} \parallel p_0). \end{aligned}$$

Therefore, \widehat{p}_0 cannot minimize the KL divergence. Therefore, $\widehat{\mathcal{S}}_0 \subset \widetilde{\mathcal{S}}_0$. □

B.3 Proofs from Section 3

Proof of Lemma 3.1. First, note that \hat{p}_n is obtained by maximizing the following functional

$$\Phi(\varphi) = \sum_{i=1}^n \varphi(z_i) \bar{p}_n - \sum_{z \in \mathbb{Z}^d} \exp\{\varphi(z)\}$$

over all concave-extendible functions, see Lemma B.2. Letting $\hat{\varphi}_n = \operatorname{argmax} \Phi(\varphi)$, we then have $\hat{p}_n(z) = \exp\{\hat{\varphi}_n(z)\}$.

Let $g(z) : \mathbb{Z}^d \mapsto \mathbb{R}$ be any concave-extendible function, and hence for any $\epsilon > 0$, $\varphi + \epsilon g$ is also concave-extendible (Murota and Shioura, 2001, Theorem 4). Therefore, $\Phi(\hat{\varphi}_n + \epsilon g) \leq \Phi(\hat{\varphi}_n)$. This implies that

$$\lim_{\epsilon \rightarrow 0} \frac{\Phi(\hat{\varphi}_n + \epsilon g) - \Phi(\hat{\varphi}_n)}{\epsilon} = \sum_{i=1}^n g(z_i) \bar{p}_n - \sum_{z \in \mathbb{Z}^d} g(z) \hat{p}_n \leq 0.$$

Similarly, for any convex-extendible function h , we have

$$\sum_{z \in \mathbb{Z}^d} h(z) \hat{p}_n \leq \sum_{z \in \mathbb{Z}^d} h(z) \bar{p}_n.$$

□

B.4 Proofs from Section 3.1

To simplify notation slightly, in this section we let $w_j = \bar{p}_n(z_j)$, $j = 1, \dots, m$. The first result shows that we can minimize the criterion function $\Phi(\varphi)$ to obtain the MLE \hat{p}_n .

Lemma B.2. *When the criterion function*

$$\Phi(\varphi) = - \sum_{j=1}^m w_j \varphi(z_j) + \sum_{z \in \mathbb{Z}^d} e^{\varphi(z)}$$

is minimized over all concave extendible functions φ , the minimizer satisfies $\sum_{z \in \mathbb{Z}^d} e^{\varphi(z)} = 1$.

Proof. Consider any concave extendible φ_0 and $p_0 = \exp\{\varphi_0\}$ such that $\sum_{z \in \mathbb{Z}^d} \exp\{\varphi_0(z)\} = c \neq 1$. Let $\tilde{\varphi}_0 = \varphi_0 - \log c$. Then $\sum_{z \in \mathbb{Z}^d} \exp\{\tilde{\varphi}_0(z)\} = 1$. Now,

$$\begin{aligned} -\sum_{j=1}^m w_j \tilde{\varphi}_0(z_j) + \sum_{z \in \mathbb{Z}^d} e^{\tilde{\varphi}_0(z)} &= -\sum_{j=1}^m w_j \varphi_0(z_j) + \sum_{z \in \mathbb{Z}^d} e^{\varphi_0(z)} - c + \log c + 1 \\ &\leq -\sum_{j=1}^m w_j \varphi_0(z_j) + \sum_{z \in \mathbb{Z}^d} e^{\varphi_0(z)}, \end{aligned}$$

since $\log c \leq c - 1$ for any $c > 0$. □

Lemma B.3. *Consider the function*

$$\tau(y_1, \dots, y_m) = -\sum_{i=1}^m w_i t_y(z_i) + \sum_{z \in \hat{S}_n} \exp\{t_y(z)\}.$$

Then τ has a minimum over $y \in \mathbb{R}^m$, \hat{y} , and $\hat{p}_n(z) = \exp\{t_{\hat{y}}(z)\}$. Furthermore, $t_{\hat{y}}$ is a concave extension of $\log \hat{p}_n$.

Proof. Let $\hat{\varphi}_n = \log \hat{p}_n$, which minimizes $\Phi(\varphi)$. Let $\hat{y}_i = \hat{\varphi}_n(z_i)$, for $i = 1, \dots, m$, and consider

$$t_{\hat{y}}(x) = \inf\{g(x) : \mathbb{R}^d \mapsto \mathbb{R} \mid g \text{ is concave, and } g(z_i) \geq \hat{y}_i \text{ for } i = 1, \dots, m\}.$$

Let $\hat{\varphi}_n^R : \mathbb{R}^d \mapsto \mathbb{R}$ denote the concave extension of $\hat{\varphi}_n$, and note that $\hat{\varphi}_n^R(z_i) = \hat{\varphi}_n(z_i) = \hat{y}_i$, $i = 1, \dots, m$. Therefore $\hat{\varphi}_n^R$ belongs to the set

$$\{g(x) : \mathbb{R}^d \rightarrow \mathbb{R} \mid g \text{ is concave, and } g(z_i) \geq \hat{y}_i \text{ for } i = 1, \dots, m\}.$$

As $t_{\hat{y}}$ is the infimum of the above class of functions, we have $t_{\hat{y}}(z) \leq \hat{\varphi}_n^R(z)$, $z \in \mathbb{Z}^d$.

Assume then that for some $z_0 \in \mathbb{Z}^d$, $\widehat{\varphi}_n(z_0) > t_{\widehat{y}}(z_0)$. Then $\sum_{z \in \mathbb{Z}^d} \exp \widehat{\varphi}_n(z) > \sum_{z \in \mathbb{Z}^d} \exp \{t_{\widehat{y}}(z)\}$.

Hence

$$\begin{aligned} \Phi(\widehat{\varphi}_n) &= - \sum_{i=1}^m w_i \widehat{\varphi}_n(z_i) + \sum_{z \in \mathbb{Z}^d} \exp \widehat{\varphi}_n(z) \\ &= - \sum_{i=1}^m w_i \widehat{y}_i + \sum_{z \in \mathbb{Z}^d} \exp \widehat{\varphi}_n(z) \\ &> - \sum_{i=1}^m w_i t_{\widehat{y}}(z_i) + \sum_{z \in \mathbb{Z}^d} \exp \{t_{\widehat{y}}(z)\} = \Phi(t_{\widehat{y}}). \end{aligned}$$

However, this creates a contradiction since $\widehat{\varphi}_n$ minimizes Φ . Therefore, $\widehat{\varphi}_n(z) = t_{\widehat{y}}(z)$, for any $z \in \mathbb{Z}^d$. This also implies that $t_{\widehat{y}}$ is a concave extension of $\log \widehat{p}_n$. \square

Proof of Theorem 3.1. We first prove that σ is convex. For $u, v \in \mathbb{R}^m$, $\lambda \in (0, 1)$, we have

$$\begin{aligned} &\lambda t_u(x) + (1 - \lambda)t_v(x) \\ &= \lambda \inf\{g_1(x) \mid g_1 \text{ concave, and } g_1(z_i) \geq u_i, i = 1, \dots, m\} \\ &\quad + (1 - \lambda) \inf\{g_2(x) \mid g_2 \text{ concave, and } g_2(z_i) \geq v_i, i = 1, \dots, m\} \\ &= \inf\{g_1(x) \mid g_1 \text{ concave, and } g_1(z_i) \geq \lambda u_i, i = 1, \dots, m\} \\ &\quad + \inf\{g_2(x) \mid g_2 \text{ concave, and } g_2(z_i) \geq (1 - \lambda)v_i, i = 1, \dots, m\} \\ &\geq \inf\{g_1(x) + g_2(x) \mid g_1, g_2 \text{ are concave, } g_1(z_i) \geq \lambda u_i, g_2(z_i) \geq (1 - \lambda)v_i, i = 1, \dots, m\}. \end{aligned}$$

Since $\{g_1(x) + g_2(x) \mid g_1, g_2 \text{ concave, } g_1(z_i) \geq \lambda u_i, g_2(z_i) \geq (1 - \lambda)v_i, i = 1, \dots, m\}$ is a subset of $\{g(x) \mid g \text{ concave, } g(z_i) \geq \lambda u_i + (1 - \lambda)v_i, i = 1, \dots, m\}$, we have

$$\lambda t_u(x) + (1 - \lambda)t_v(x) \geq t_{\lambda u + (1 - \lambda)v}(x), x \in \mathbb{R}^d.$$

Finally, by convexity of e^x ,

$$\begin{aligned}
\sigma(\lambda u + (1 - \lambda)v) &= -\sum_{i=1}^m w_i \{\lambda u_i + (1 - \lambda)v_i\} + \sum_{z \in \mathbb{Z}^d} \exp \{t_{\lambda u + (1 - \lambda)v}(z)\} \\
&\leq -\sum_{i=1}^m w_i \{\lambda u_i + (1 - \lambda)v_i\} + \sum_{z \in \mathbb{Z}^d} \exp \{\lambda t_u(z) + (1 - \lambda)t_v(z)\} \\
&\leq -\sum_{i=1}^m w_i \{\lambda u_i + (1 - \lambda)v_i\} + \lambda \sum_{z \in \mathbb{Z}^d} e^{t_u(z)} + (1 - \lambda) \sum_{z \in \mathbb{Z}^d} e^{t_v(z)} \\
&= \lambda \sigma(u) + (1 - \lambda) \sigma(v).
\end{aligned}$$

Hence, $\sigma(y)$ is convex.

Next, for any $y \in \mathbb{R}^m$,

$$\sigma(y) = \tau(y) + \sum_{i=1}^m w_i (t_y(z_i) - y_i) \geq \tau(y),$$

by definition of the tent function t_y . Furthermore, recall from Lemma B.3, that $\hat{y} = \log \hat{p}_n$ minimizes τ and satisfies $t_{\hat{y}}(z_i) = \hat{y}_i, i = 1, \dots, m$. Therefore,

$$\sigma(\hat{y}) = \tau(\hat{y}) = \min_y \tau(y),$$

and hence $\min_y \sigma(y) = \sigma(\hat{y})$. □

B.5 Proofs from Section 4

The following lemma states that when a sequence of PMFs on \mathbb{Z}^d converges pointwise, then it also converges in the l_k -distance and Hellinger distances. This lemma is a slight generalization of a similar result appearing in Balabdaoui et al. (2013).

Lemma B.4. *Let p_n, p be discrete probability mass functions on \mathbb{Z}^d , and $p_n \rightarrow p$ for all*

$z \in \mathbb{Z}^d$, then $l_k(p_n, p) \rightarrow 0$ for $1 \leq k \leq \infty$, and $h^2(p_n, p) \rightarrow 0$, as $n \rightarrow \infty$.

Proof. Clearly, it is sufficient to show that pointwise convergence implies the other, stronger, types of convergence. To this end, fix $\varepsilon > 0$. Then, there exists a K such that $\sum_{\|z\|_\infty \leq K} p(z) \geq 1 - \varepsilon/4$. Furthermore, since $p_n \rightarrow p$, for all $z \in \mathbb{Z}^d$, there exists large enough N , such that

$$\sup_{\|z\|_\infty \leq K} |p_n(z) - p(z)| \leq \frac{\varepsilon}{4(2K+1)^d},$$

for all $n \geq N$. From the above, it also follows that for all $n \geq N$,

$$\sum_{\|z\|_\infty \leq K} p_n(z) \geq \sum_{\|z\|_\infty \leq K} p(z) - \frac{\varepsilon}{4(2K+1)^d} * (2K+1)^d \geq 1 - \varepsilon/4 - \varepsilon/4 = 1 - \varepsilon/2.$$

Putting these facts together, we find that

$$\begin{aligned} \sum_{z \in \mathbb{Z}^d} |p_n(z) - p(z)| &\leq \sum_{\|z\|_\infty \leq K} |p_n(z) - p(z)| + \sum_{\|z\|_\infty > K} p_n(z) + \sum_{\|z\|_\infty > K} p(z) \\ &\leq \varepsilon/4 + \varepsilon/2 + \varepsilon/4 = \varepsilon. \end{aligned}$$

We have thus shown that pointwise convergence implies $l_1(p_n, p) \rightarrow 0$.

Note that for any fixed z_0 , we have $|p_n(z_0) - p(z_0)| \leq \sum_{z \in \mathbb{Z}^d} |p_n(z) - p(z)|$, and hence $\sup_{z \in \mathbb{Z}^d} |p_n(z) - p(z)| \leq \sum_{z \in \mathbb{Z}^d} |p_n(z) - p(z)|$. Moreover, $0 \leq p_n(z), p(z) \leq 1$ implies that $|p_n(z) - p(z)| \leq 1$. Hence $|p_n(z) - p(z)|^k \leq |p_n(z) - p(z)|$, for any $1 < k < \infty$. Therefore, for any $1 < k \leq \infty$,

$$l_k^k(p_n, p) = \sum_{z \in \mathbb{Z}^d} |p_n(z) - p(z)|^k \leq \sum_{z \in \mathbb{Z}^d} |p_n(z) - p(z)| = l_1(p_n, p)$$

Lastly, recall that $2h^2(p, q) \leq l_1(p, q)$. We conclude that pointwise convergence implies all other types of convergence as well. \square

Proof of Theorem 4.1. Let X_n be a random vector with PMF \hat{p}_n . Then by Markov's inequality and Lemma 3.1, we have that

$$P(\|X_n\|_\infty \geq m) \leq m^{-1} \sum_{\|z\|_\infty \geq m} \hat{p}_n(z) \leq m^{-1} \sum_{z \in \mathbb{Z}^d} \|z\|_\infty \bar{p}_n(z),$$

since the norm $\|\cdot\|_\infty$ is convex-extendible. By strong law of large number and the finite mean assumption of the Theorem, $\sum_{z \in \mathbb{Z}^d} \|z\|_\infty \bar{p}_n(z) \leq 2 \sum_{z \in \mathbb{Z}^d} \|z\|_\infty p_0(z)$, say, almost surely for all n sufficiently large. It follows that the sequence X_n is tight. Therefore, there exists a subsequence of \hat{p}_n and a \tilde{p} , which we denote again by n , such that $\hat{p}_n \rightarrow \tilde{p}$. By Theorem 2.1 we conclude that \tilde{p} is also eLC. It remains to show that $\tilde{p} = \hat{p}_0$ to finish the proof.

Since the MLE maximizes the likelihood and since \log is a strictly increasing function, we have that for any $b > 0$

$$\sum_{z \in \mathbb{Z}^d} \bar{p}_n(z) \log(\hat{p}_n(z) + b) \geq \sum_{z \in \mathbb{Z}^d} \bar{p}_n(z) \log(\hat{p}_0(z)).$$

Therefore,

$$\begin{aligned} & \sum_{z \in \mathbb{Z}^d} \bar{p}_n(z) \log(\hat{p}_n(z) + b) - \sum_{z \in \mathbb{Z}^d} \bar{p}_n(z) \log(\hat{p}_0(z)) \\ &= \sum_{z \in \mathbb{Z}^d} (\bar{p}_n(z) - p_0(z)) \log(\hat{p}_n(z) + b) + \sum_{z \in \mathbb{Z}^d} (p_0(z) - \bar{p}_n(z)) \log(\hat{p}_0(z)) \\ & \quad + \sum_{z \in \mathbb{Z}^d} p_0(z) \log \frac{\hat{p}_n(z) + b}{\hat{p}_0(z) + b} + \sum_{z \in \mathbb{Z}^d} p_0(z) \log \frac{\hat{p}_0(z) + b}{\hat{p}_0(z)} \geq 0. \end{aligned} \quad (3)$$

We next get rid of the first two terms on the right-hand side. For the first, note that for

$$M = \max\{|\log(b)|, |\log(b+1)|\} < \infty,$$

$$\begin{aligned} \left| \sum_{z \in \mathbb{Z}^d} (\bar{p}_n(z) - p_0(z)) \log(\hat{p}_n(z) + b) \right| &\leq \sum_{z \in \mathbb{Z}^d} |\bar{p}_n(z) - p_0(z)| |\log(\hat{p}_n(z) + b)| \\ &\leq M \sum_{z \in \mathbb{Z}^d} |\bar{p}_n(z) - p_0(z)| \xrightarrow{a.s.} 0, \quad \text{as } n \rightarrow \infty. \end{aligned}$$

We now show that the 2nd term converges to zero. Since \hat{p}_0 minimizes KL divergence,

$$E_{p_0}[\log \hat{p}_0(Z)] = - \sum_{z \in \mathbb{Z}^d} p_0(z) \log \hat{p}_0(z) \leq - \sum_{z \in \mathbb{Z}^d} p_0(z) \log p_0(z) = E_{p_0}[\log p_0(Z)] < \infty.$$

Therefore, by the strong law of large numbers,

$$\begin{aligned} \sum_{z \in \mathbb{Z}^d} (p_0(z) - \bar{p}_n(z)) \log \hat{p}_0(z) &= \sum_{z \in \mathbb{Z}^d} \bar{p}_n(z) \left(-\log \hat{p}_0(z) \right) - \sum_{z \in \mathbb{Z}^d} p_0(z) \left(-\log \hat{p}_0(z) \right) \\ &= E_{\bar{p}_n}[\log \hat{p}_0(Z)] - E_{p_0}[\log \hat{p}_0(Z)] \xrightarrow{a.s.} 0, \quad \text{as } n \rightarrow \infty. \end{aligned}$$

Thus, (3) yields that

$$\limsup_n \sum_{z \in \mathbb{Z}^d} p_0(z) \log \frac{\hat{p}_0(z) + b}{\hat{p}_n(z) + b} \leq \sum_{z \in \mathbb{Z}^d} p_0(z) \log \frac{\hat{p}_0(z)}{\hat{p}_0(z) + b}.$$

By Fatou's Lemma, we have

$$\liminf_{b \rightarrow 0} \sum_{z \in \mathbb{Z}^d} \left\{ -p_0(z) \log \frac{\hat{p}_0(z)}{\hat{p}_0(z) + b} \right\} \geq \sum_{z \in \mathbb{Z}^d} \liminf_{b \rightarrow 0} \left\{ -p_0(z) \log \frac{\hat{p}_0(z)}{\hat{p}_0(z) + b} \right\} = 0,$$

and therefore,

$$\limsup_{b \rightarrow 0} \limsup_n \sum_{z \in \mathbb{Z}^d} p_0(z) \log \frac{\hat{p}_0(z) + b}{\hat{p}_n(z) + b} \leq \limsup_{b \rightarrow 0} \sum_{z \in \mathbb{Z}^d} p_0(z) \log \frac{\hat{p}_0(z)}{\hat{p}_0(z) + b} = 0.$$

Next, since

$$\left| p_0(z) \log \frac{\widehat{p}_0(z) + b}{\widehat{p}_n(z) + b} \right| \leq p_0(z) \max \left\{ \left| \log \frac{b+1}{b} \right|, \left| \log \frac{b}{b+1} \right| \right\},$$

by dominated convergence theorem, we have

$$\limsup_{b \rightarrow 0} \limsup_n \sum_{z \in \mathbb{Z}^d} p_0(z) \log \frac{\widehat{p}_0(z) + b}{\widehat{p}_n(z) + b} = \limsup_{b \rightarrow 0} \sum_{z \in \mathbb{Z}^d} p_0(z) \log \frac{\widehat{p}_0(z) + b}{\widehat{p}(z) + b}.$$

Without loss of generality, we can restrict $0 < b \leq 1$, and hence $-\log(\widehat{p}_0 + b) \geq -\log 2$, which implies that $-\log(\widehat{p}_0 + b)$ is bounded below and increases as $b \rightarrow 0$. Therefore, by monotone convergence theorem, we have

$$\limsup_{b \rightarrow 0} \sum_{z \in \mathbb{Z}^d} p_0(z) \{-\log(\widehat{p}_0(z) + b)\} = - \sum_{z \in \mathbb{Z}^d} p_0(z) \log \widehat{p}_0(z),$$

and similarly when \widehat{p}_0 is replaced by \widetilde{p} . Putting together the above arguments, we thus arrive at

$$\limsup_{b \rightarrow 0} \sum_{z \in \mathbb{Z}^d} p_0(z) \log \frac{\widehat{p}_0(z) + b}{\widetilde{p}(z) + b} = \sum_{z \in \mathbb{Z}^d} p_0(z) \log \frac{\widehat{p}_0(z)}{\widetilde{p}(z)} \leq 0.$$

Rearranging, we find that

$$\sum_{z \in \mathbb{Z}^d} p_0(z) \log \frac{p_0(z)}{\widetilde{p}(z)} \leq \sum_{z \in \mathbb{Z}^d} p_0(z) \log \frac{p_0(z)}{\widehat{p}_0(z)}.$$

However, as \widehat{p}_0 is the unique minimizer of the quantity on the right hand side, we obtain that $\widetilde{p} = \widehat{p}_0$. Therefore $\widehat{p}_n \rightarrow \widehat{p}_0$, and by Lemma B.4 we have $d(\widehat{p}_n, \widehat{p}_0) \rightarrow 0$, as $n \rightarrow \infty$. \square

C Appendix: Algorithm

We begin by deriving an explicit formula for $\sigma(y)$. To this end, a few definitions are necessary. For our observations z_1, \dots, z_m and $y \in \mathbb{R}^m$, consider the set $\mathcal{Z} = \{(z_1, y_1), \dots, (z_m, y_m)\}$. The convex hull of the set $\mathcal{Z} \subset \mathbb{R}^{d+1}$ is made up of the upper hull and the lower hull. Projecting the upper hull on the first dimensional subspace \mathbb{Z}^d , the faces of the upper hull create a subdivision of the points z_1, \dots, z_m . We denote the subdivision as $\mathcal{S}(y)$, to emphasize its dependence on the vector y . This notion is best illustrated with examples. Consider the observations $\{z_i\}_{i=1}^4 = \{(0, 0), (2, 0), (3, 1), (1, 1)\}$ with $m = 4$. We compute, for three different y vectors, the associated $\mathcal{S}(y)$:

- Let $y_1 = (1, 1.9, 2, 1)$, then $\mathcal{Z} = \{(0, 0, 1), (2, 0, 1.9), (3, 1, 2), (1, 1, 1)\}$, and $\mathcal{S}(y_1)$ has two subdivisions: $\{(0, 0), (2, 0), (3, 1)\}$, $\{(0, 0), (1, 1), (3, 1)\}$.
- Let $y_2 = (1, 2, 2, 1)$, then $\mathcal{Z} = \{(0, 0, 1), (2, 0, 2), (3, 1, 2), (1, 1, 1)\}$, and $\mathcal{S}(y_2)$ has only one subdivision: $\{(0, 0), (2, 0), (3, 1), (1, 1)\}$.
- Let $y_3 = (1, 2.1, 2, 1)$, then $\mathcal{Z} = \{(0, 0, 1), (2, 0, 2.1), (3, 1, 2), (1, 1, 1)\}$, and $\mathcal{S}(y_3)$ has two subdivisions: $\{(0, 0), (2, 0), (1, 1)\}$, $\{(0, 0), (2, 0), (3, 1)\}$.

These three examples are illustrated in Figure 1. We can refine each subdivision into a triangulation (a partition into simplices). Note that $\mathcal{S}(y_1)$ and $\mathcal{S}(y_3)$ are both triangulations, while $\mathcal{S}(y_2)$ needs further partitioning. Let $\mathcal{T}(y) = \{S_j, j \in \mathcal{J}\}$ denote the triangulation, where each S_j is a simplex (given by $d+1$ vertices, $\{z_{j_0}, \dots, z_{j_d}\}$). Let $J_j = \{j_0, \dots, j_d\}$ denote the indices. Finally, let C_j denote the convex hull of $S_j, j = 1, \dots, |\mathcal{J}|$.

For finitely many points, the tent functions can be written explicitly via the triangulations

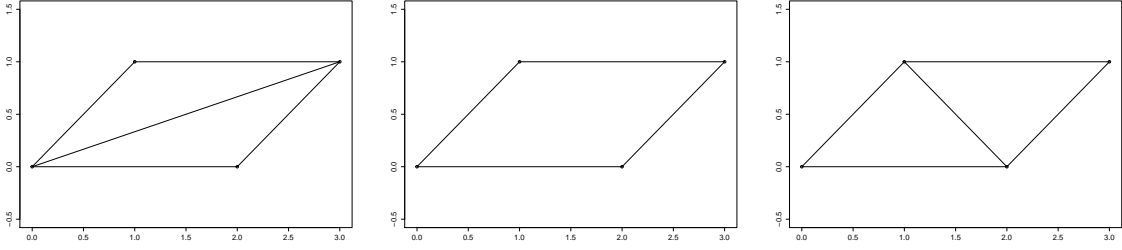


Figure 1: Subdivisions $\mathcal{S}(y)$ for $y = y_1$ (left), $y = y_2$ (centre), and $y = y_3$ (right).

(Cule, 2009, Equation 3.6, page 26)

$$t_y(z) = \sum_{j \in \mathcal{J}} (b_j^T z - \beta_j) \mathbb{I}_{C_j}(z) + \delta_{\widehat{\mathcal{S}}_n}(z),$$

for some b_j, β_j . Here, $\mathbb{I}_{C_j}(z)$ is an indicator function and \mathcal{J} indicates a triangulation by y .

Finally,

$$\delta_{\widehat{\mathcal{S}}_n}(z) = \begin{cases} 0 & \text{if } z \in \widehat{\mathcal{S}}_n, \\ -\infty & \text{if } z \notin \widehat{\mathcal{S}}_n. \end{cases}$$

Let θ denote an element in a unit d -simplex: $\theta \in [0, \infty)^d$, and $\sum_{i=1}^d \theta_i \leq 1$. Following Cule (2009, page 27), we perform a translation to re-write the above formulas over the unit simplex. Define $A_j = (z_{j_1} - z_{j_0} \mid \dots \mid z_{j_d} - z_{j_0})$ to be a $d \times d$ matrix and let $a_j = z_{j_0}$. Then for $z \in C_j$, $\theta = (A_j)^{-1}(z - a_j)$ is in the unit simplex. Next, let $\tilde{y}_j \in \mathbb{R}^d$ have components $(y_{j_1} - y_{j_0}, \dots, y_{j_d} - y_{j_0})$. Then we can write, $b_j = (A_j^T)^{-1} \tilde{y}_j$ and $\beta_j = a_j^T b_j - y_{j_0}$. Thus,

$$\begin{aligned} b_j^T z - \beta_j &= [(A_j^T)^{-1} \tilde{y}_j]^T (A_j \theta + a_j) - a_j^T (A_j^T)^{-1} \tilde{y}_j + y_{j_0} \\ &= \tilde{y}_j^T \theta + y_{j_0} \\ &= y_{j_0} \theta_0 + y_{j_1} \theta_1 + \dots + y_{j_d} \theta_d \equiv \theta^T y_{J_j}, \end{aligned}$$

where $\theta_0 = 1 - \theta_1 - \dots - \theta_d$. Therefore,

$$\begin{aligned}\sigma(y) &= -\sum_{i=1}^m w_i y_i + \sum_{z \in \widehat{\mathcal{S}}_n} \exp \left\{ \sum_{j \in \mathcal{J}} (b_j^T z - \beta_j) \mathbb{I}_{C_j}(z) \right\} \\ &= -\sum_{i=1}^m w_i y_i + \sum_{j \in \mathcal{J}} \sum_{z \in C_j} \exp \left\{ (b_j^T z - \beta_j) \right\},\end{aligned}$$

We then obtain

$$\sigma(y) = -\sum_{i=1}^m w_i y_i + \sum_{j \in \mathcal{J}} \sum_{z \in C_j, z \notin \cup_{k=1}^{j-1} C_k} \exp \{y_{j_0} \theta_0 + y_{j_1} \theta_1 + \dots + y_{j_d} \theta_d\},$$

where $\theta = A_j^{-1}(z - a_j)$ for $z \in C_j$. Note that some z may belong to more than one simplex, and hence the need to exclude these cases in the second summand above.

We also need to compute the derivatives, or when not differentiable, the directional derivative of $\sigma(y)$. As in Cule (2009, Section 3.4.2, page 34), $\sigma(y)$ is differentiable if $\mathcal{S}(y)$ is a triangulation, while if $\mathcal{S}(y)$ is not a triangulation, it is not differentiable. This is relatively straightforward to see from Figure 1, as small changes to the second element of y yield very different subdivisions.

When $\sigma(y)$ is differentiable, we easily obtain that

$$\partial_i \sigma(y) = -w_i + \sum_{j \in \mathcal{J}} \mathbb{I}_{C_j}(y_i) \sum_{z \in C_j, z \notin \cup_{k=1}^{j-1} C_k} \partial_i \{ \theta^T y_{J_j} \} \exp \{ \theta^T y_{J_j} \},$$

Note that when we compute the i th partial derivative, we only need to consider those simplices which involve y_i , so the indicator function above ensures that only the simplex involving y_i will be counted.

Proposition C.1. *The function $\sigma(y) = -\sum_{i=1}^m w_i y_i + \sum_{z \in \widehat{\mathcal{S}}_n} \exp\{t_y(z)\}$ is not differentiable everywhere.*

Proof. Denote the directional derivatives as

$$\partial\sigma(y; u) = \lim_{t \rightarrow 0} \frac{\sigma(y + tu) - \sigma(y)}{t}.$$

Since σ is convex, the directional derivatives exist (Rockafellar, 1970, Theorem 23.1 page 213). Furthermore, the function is differentiable if $\partial\sigma(y; u) = -\partial\sigma(y; -u)$. We will show that $\partial\sigma(y; e_i) + \partial\sigma(y; -e_i) > 0$ occurs when $\mathcal{S}(y)$ is not a triangulation, where $e_i \in \{0, 1\}^m$ is the i th row of the m dimensional identity matrix. For simplicity, consider the case when there are $m = d + 2$ elements in general position, as the more complex case is similar.

For each i and $\varepsilon_0 > 0$ sufficiently small, we have that $\mathcal{S}(y + \varepsilon_0 e_i), \mathcal{S}(y - \varepsilon_0 e_i)$ both form triangulations. Following Cule (2009, Section 3.4.2, page 35), we may write

$$t_{y+\varepsilon e_i}(x) = t_y(x) + \varepsilon g_{e_i, \mathcal{S}(y+\varepsilon_0 e_i)}(x) \quad t_{y-\varepsilon e_i}(x) = t_y(x) + \varepsilon g_{-e_i, \mathcal{S}(y-\varepsilon_0 e_i)}(x),$$

where $g_{e_i, \mathcal{S}(y+\varepsilon_0 e_i)}, -g_{-e_i, \mathcal{S}(y-\varepsilon_0 e_i)}(x)$ are the upper and lower hulls of the points

$$\{(z_1, 0), \dots, (z_{i-1}, 0), (z_i, 1), (z_{i+1}, 0), \dots, (z_m, 0)\},$$

respectively.

Letting e_{ij} denote the (i, j) -element of the $n \times n$ identity matrix, we can write

$$\begin{aligned} \partial\sigma(y; e_i) &= -w_i + \lim_{\varepsilon \rightarrow 0} \varepsilon^{-1} \left(\sum_{z \in \widehat{\mathcal{S}}_n} \exp \{t_y(z) + \varepsilon g_{e_i, \mathcal{S}(y+\varepsilon e_i)}(z)\} - \sum_{z \in \widehat{\mathcal{S}}_n} \exp \{t_y(z)\} \right) \\ &= -w_i + \sum_{z \in \widehat{\mathcal{S}}_n} \exp \{t_y(z)\} g_{e_i, \mathcal{S}(y+\varepsilon_0 e_i)}(z). \end{aligned}$$

Similarly, we find

$$\partial\sigma(y; -e_i) = w_i + \sum_{z \in \widehat{\mathcal{S}}_n} \exp\{t_y(z)\} g_{-e_i, \mathcal{S}(y - \varepsilon_0 e_i)}(z).$$

Hence,

$$\partial\sigma(y; e_i) + \partial\sigma(y; -e_i) = \sum_{z \in \widehat{\mathcal{S}}_n} \exp\{t_y(z)\} \{g_{e_i, \mathcal{S}(y + \varepsilon_0 e_i)}(z) + g_{-e_i, \mathcal{S}(y - \varepsilon_0 e_i)}(z)\}.$$

From the arguments above, it follows that $g_{e_i, \mathcal{S}(y + te_i)} + g_{-e_i, \mathcal{S}(y - te_i)} > 0$ and hence $\partial\sigma(y; e_i) + \partial\sigma(y; -e_i) > 0$. □

C.1 Subgradient algorithm

Since the function $\sigma(y)$ is not differentiable, following Cule et al. (2010), we apply a well-known subgradient-based method, known as Shor’s r -algorithm, to compute our MLE.

The general idea of subgradient algorithms is to proceed iteratively as follows:

Theorem C.1 (Shor (1985)). *Let (h_i) be a positive sequence with $h_i \rightarrow 0$ as $i \rightarrow \infty$ and $\sum_{i=0}^{\infty} h_i = \infty$. Then, for any convex function σ , the sequence generated by the formula*

$$y_{i+1} = y_i - h_i \frac{\partial\sigma(y_i)}{\|\partial\sigma(y_i)\|}$$

has the property that either there exists an i_0 and y^ such that $y_{i_0} = y^*$, or $y_i \rightarrow y^*$ and $\sigma(y_i) \rightarrow \sigma(y^*)$ as $i \rightarrow \infty$.*

Shor’s r -algorithm is a modification of the above with the goal of improving convergence rates, see e.g. Kappel and Kuntsevich (2000) for a description. The idea here is to “make steps in the direction opposite to a sub-gradient” (Kappel and Kuntsevich, 2000, page 193).

These steps are made in a transformed, “dilated”, space. Kappel and Kuntsevich (2000) describe further improvements to the method via modified stopping criteria. As in Cule et al. (2010), we use this latter modification with stopping criteria

$$\begin{aligned} |y_i^{k+1} - y_i^k| &\leq \delta |y_i^k| \text{ for } i = 1, \dots, n \\ |\sigma(y^{k+1}) - \sigma(y^k)| &\leq \varepsilon |\sigma(y^k)| \\ \left| 1 - \sum_{z \in \mathbb{Z}^d} \exp \{t_{y^k}(z)\} \right| &\leq \eta \end{aligned}$$

for fixed tolerances δ, ε and η . The last criterion above is not one suggested by Kappel and Kuntsevich (2000), but is there to ensure that the algorithm returns close to a proper probability mass function. In our current implementation, the tolerances are set to $\delta = \varepsilon = \eta = 10^{-4}$.

C.2 Algorithm to calculate optimization function and gradient

To compute $\sigma(y)$, we refine the projection of \mathcal{Z} into simplices. We then work on each simplex, and find all lattice points inside or on the boundary of the simplex. If the discrete point has not been counted, we compute the corresponding θ , and add in the exponential term. The quickhull algorithm is applied to compute convex hulls and triangulations. Details of these calculations, as well as gradient and subgradient calculations are given in Algorithm 1.

D Appendix: some background from convex analysis

Theorem D.1. *(Rockafellar, 1970, Theorem 10.6, page 88) Let \mathcal{C} be a relatively open convex set, and let $\{f_i | i \in \mathcal{I}\}$ be an arbitrary collection of convex functions finite and pointwise bounded on \mathcal{C} . let \mathcal{S} be any closed bounded subset of \mathcal{C} . Then $\{f_i | i \in \mathcal{I}\}$ is uniformly bounded*

Algorithm 1 Calculate $\sigma(y)$ and gradient of $\sigma(y)$, input z_{obs}, y

- 1: Compute convex hull of observations: $C = \text{conv } z_{obs}$
 - 2: Compute extreme/out points of $C : z_{out}$, corresponding subset of $y: y_{out}$
 - 3: Compute inner points of $C : z_{in} = z_{obj} \setminus z_{out}$, corresponding subset of $y: y_{in}$
 - 4: $y_{max} = \max\{y_1, \dots, y_m\}$
 - 5: $y_{min} = \min\{y_1, \dots, y_m\}$
 - 6: Combine $\frac{y_{out}}{y_{max}-y_{min}}$ and z_{out} to get $d + 1$ dimensional data set: zz_{out}
 - 7: Combine $\frac{y_{in}}{y_{max}-y_{min}}$ and z_{in} to get $d + 1$ dimensional data set: zz_{in}
 - 8: Combine $\frac{y_{min}-1}{y_{max}-y_{min}}$ and z_{out} to get $d + 1$ dimensional data set: zz_{xtr}
 - 9: All points set: $zz_{all} = zz_{out} \cup zz_{in} \cup zz_{xtr}$
 - 10: Compute convex hull of All points set: $C_{all} = \text{conv } zz_{all}$
 - 11: Compute facet set of $C_{all} : fct = \{fct_1, \dots, fct_k\}$
 - 12: Initial $\sigma(y) = -(\bar{p}_1 * y_1 + \dots + \bar{p}_m * y_m)$
 - 13: Initial $\partial_i \sigma(y) = -\bar{p}_i, i = 1, \dots, m$
 - 14: Initial E_{all} as an empty list ▷ Used to check duplication
 - 15: **for** each facet $fct_j, 1 \leq j \leq k$ **do**
 - 16: **if** fct_j is a true facet **then**
 - 17: The extreme (out) points set of $fct_j : p_j = \{z_{j_0}, \dots, z_{j_d}\}$
 - 18: Matrix $A = [z_{j_1} - z_{j_0} | \dots | z_{j_d} - z_{j_0}]$
 - 19: Vector $a_j = z_{j_0}$
 - 20: Inverse matrix of $A : A^{-1}$
 - 21: Vector $y_{tmp} = \{y_{j_0}, \dots, y_{j_d}\}$
 - 22: Generate a rectangle of $p_j : rec = \{r \in \mathbb{Z}^d\}$, such that $r_i = \{z \in \mathbb{Z} | \min\{z_{j_0}^i, \dots, z_{j_d}^i\} \leq z \leq \max\{z_{j_0}^i, \dots, z_{j_d}^i\}\}$, for $1 \leq i \leq d$
 - 23: **for** each point r in rec **do**
 - 24: **if** r is inside convex hull of p_j **then**
 - 25: Add r to enumerate list: E_j
 - 26: **end if**
 - 27: **end for**
 - 28: **for** each point of $E_j : e$ **do**
 - 29: **if** e is not duplicated with any points of E_{all} **then**
 - 30: Vector $w = A^{-1}(e - a_j), w_0 = 1 - w_1 - \dots - w_d$
 - 31: Sigma function: $\sigma(y) += \exp\{y_{j_0} w_0 + \dots + y_{j_d} w_d\}$
 - 32: **for** $i \in \{j_0, \dots, j_d\}$ **do**
 - 33: Gradient: $\partial_i \sigma(y) += w_i \exp\{y_{j_0} w_0 + \dots + y_{j_d} w_d\}$
 - 34: **end for**
 - 35: Add e to enumerate list E_{all}
 - 36: **end if**
 - 37: **end for**
 - 38: **end if**
 - 39: **end for**
 - 40: Return $\sigma(y), \partial_i \sigma(y)$ for $i = 1, \dots, m$
-

on \mathcal{S} and equi-Lipschitzian relative to \mathcal{S} .

The conclusion remains valid if the pointwise boundedness assumption is weakened to the following pair of assumptions:

(a). There exists a subset \mathcal{C}' of \mathcal{C} such that $\text{conv}(\text{cl } \mathcal{C}') \supset \mathcal{C}$ and $\sup\{f_i(x) | i \in \mathbb{I}\}$ is finite for every $x \in \mathcal{C}'$;

(b). There exists at least one $x \in \mathcal{C}$ such that $\inf\{f_i(x) | i \in \mathbb{I}\}$ is finite.

Theorem D.2. (Rockafellar, 1970, Theorem 32.2, page 343) Let f be a convex function, and let $\mathcal{C} = \text{conv } \mathcal{S}$, where \mathcal{S} is an arbitrary set of points. Then

$$\sup\{f(x) | x \in \mathcal{C}\} = \sup\{f(x) | x \in \mathcal{S}\},$$

where the first supremum is attained only when the second (more restrictive) supremum is attained.

References

- Balabdaoui, F., Jankowski, H., Rufibach, K. and Pavlides, M. (2013). Asymptotics of the discrete log-concave maximum likelihood estimator and related applications. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **75** 769–790.
- Cule, M. (2009). *Maximum likelihood estimation of a multivariate log-concave density*. Ph.D. thesis, University of Cambridge, UK.
- Cule, M., Samworth, R. and Stewart, M. (2010). Maximum likelihood estimation of a multi-dimensional log-concave density. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **72** 545–607.

- Dharmadhikari, S. and Joag-Dev, K. (1988). *Unimodality, convexity, and applications*. Probability and Mathematical Statistics, Academic Press, Inc., Boston, MA.
- Kappel, F. and Kuntsevich, A. V. (2000). An implementation of Shor's r -algorithm. *Comput. Optim. Appl.* **15** 193–205.
- Krantz, S. G. (1991). *Real analysis and foundations*. 3rd ed. Textbooks in Mathematics, CRC Press, Boca Raton, FL.
- Murota, K. and Shioura, A. (2001). Relationship of M -/ L -convex functions with discrete convex functions by Miller and Favati-Tardella. *Discrete Appl. Math.* **115** 151–176. 1st Japanese-Hungarian Symposium for Discrete Mathematics and its Applications (Kyoto, 1999).
- Rockafellar, R. T. (1970). *Convex analysis*. Princeton Mathematical Series, No. 28, Princeton University Press, Princeton, N.J.
- Shor, N. Z. (1985). *Minimization methods for nondifferentiable functions*, vol. 3 of *Springer Series in Computational Mathematics*. Springer-Verlag, Berlin. Translated from the Russian by K. C. Kiwiel and A. Ruszczyński.