SIAM J. OPTIMIZATION
Vol. 4, No. 1, pp. 146–158, February 1994

008

STRONG ROTUNDITY AND OPTIMIZATION*

J. M. BORWEIN† AND A. S. LEWIS†

Abstract. Standard techniques from the study of well-posedness show that if a fixed convex objective function is minimized in turn over a sequence of convex feasible regions converging Mosco to a limiting feasible region, then the optimal solutions converge in norm to the optimal solution of the limiting problem. Certain conditions on the objective function are needed as is a constraint qualification. If, as may easily occur in practice, the constraint qualification fails, stronger set convergence is required, together with stronger analytic/geometric properties of the objective function: strict convexity (to ensure uniqueness), weakly compact level sets (to ensure existence and weak convergence), and the Kadec property (to deduce norm convergence). By analogy with the L_p norms, such properties are termed "strong rotundity." A very simple characterization of strongly rotund integral functionals on L_1 is presented that shows, for example, that the Boltzmann–Shannon entropy $\int x \log x$ is strongly rotund. Examples are discussed, and the existence of everywhere- and densely-defined strongly rotund functions is investigated.

Key words. strongly rotund, well-posed, set-convergence, maximum entropy method, Kadec

AMS subject classifications. primary 46B20, 49D37; secondary 90C48

1. Introduction. Consider the sequence of optimization problems

$$(P_n) \qquad \inf\{f(x) \mid x \in C_n\},$$

together with an associated limiting problem

$$(P_\infty) \qquad \inf\{f(x) \mid x \in C_\infty\},$$

and let us ask the question "under what conditions will optimal solutions x_n for (P_n) converge to an optimal solution x_∞ for (P_∞)?" Such questions have of course been studied extensively in the literature of well-posedness and variational convergence (see, for example, [17] and [24]). Indeed, as we shall see, it is straightforward to establish by standard techniques that when f is a strictly convex function on a Banach space, with weakly compact level sets, and the sets C_n and C_∞ are closed and convex, with C_n tending Mosco to C_∞, then, providing a suitable constraint qualification holds for (P_∞), x_n will tend weakly to x_∞. With the extra assumption that the conjugate function f^* is Fréchet differentiable whenever it has a subgradient, we obtain our goal: x_n converges to x_∞ in norm. In the language of well-posedness, this derivation is closely related to the well-known fact that the function f is *Tikhonov well posed;* in other words, every minimizing sequence converges in norm to the unique minimizer $\bar{x}$, if and only if f^* has Fréchet derivative $\bar{x}$ at 0 (see Proposition 1 in [1], and [10]).

The natural constraint qualification required is that C_∞ intersect the interior of the domain of f. We use it at two points in the argument: first, to ensure the convergence of the values of (P_n) to that of (P), and second, to deduce norm convergence of x_n from weak convergence.

Our aim in this paper is in no sense to extend this general theory, but rather to cope with a *practical* difficulty, namely, that in concrete problems the constraint qualification may easily fail. Specifically, we have in mind cases where the limit set C_∞ reduces to a single point at which the objective function f has no subgradient. For example, in best entropy estimation (see, for example, [5]), the set C_n may consist of those functions whose first n Fourier coefficients agree with those of an unknown function x_∞, and the limit set C_∞ is then simply $\{x_\infty\}$. In general, the objective function f, a measure of entropy, is a convex

* Received by the editors January 6, 1992; accepted for publication (in revised form) December 23, 1992. This work was supported in part by the Natural Sciences and Engineering Research Council of Canada.

† Department of Combinatorics and Optimization, University of Waterloo, Waterloo, Ontario, Canada N2L 3G1.

integral functional on L_1, with domain some subset of the positive cone; typically f will have no subgradients at x_∞ unless, at the very least, x_∞ is almost everywhere strictly positive. Thus strong conditions on x_∞ are required if we wish to apply the above theory to deduce convergence of the best entropy estimates; if this unknown density can take the value 0 on regions of its domain, then we need a different approach to guarantee L_1-norm convergence.

Two refinements are necessary if we wish to remove the requirement of the constraint qualification. The first step is to ensure the weak convergence of solutions by strengthening the type of convergence of C_n to C_∞. In the next section we show how this may be accomplished by replacing the strong lim inf in the definition of Mosco convergence by the discrete lim inf. We illustrate by considering the types of moment problems that appear in maximum entropy estimation.

The section that follows contains the second step: moving from weak to norm convergence. What is needed here is the *strong rotundity* of f, in other words, the fact that f is strictly convex on its domain and has weakly compact level sets, and the fact that the *Kadec* property holds: if x_n tends weakly to x_∞ in the domain of f, and $f(x_n)$ tends to $f(x_\infty)$, then x_n tends to x_∞ in norm. The central result shows that strong rotundity of the integral functional on $L_1, f(x) := \int \phi(x(s))$, is characterized under very general conditions by the simple condition that the conjugate integrand ϕ^* is everywhere differentiable on $\mathbb{R}$ (see also [25]). Examples are given, including the Boltzmann–Shannon entropy and various other natural choices of entropy (see, for example, [7]).

2. Sequences of optimization problems. We shall suppose throughout this section that X is a Banach space, the function $f : X \to (-\infty, +\infty]$ is a proper convex function with weakly compact lower level sets, and $C_1, C_2, \ldots$ and C_∞ are closed convex subsets of X. We shall use the following limiting notions:

$$C_\infty \supset w - \overline{\lim}\, C_n := \{x \mid \exists x_{n_r} \in C_{n_r} \text{ with } x_{n_r} \to x \text{ weakly}\}. \tag{2.1}$$

$$C_\infty \subset s - \underline{\lim}\, C_n := \{x \mid \exists x_n \in C_n \text{ with } \|x_n - x\| \to 0\}. \tag{2.2}$$

$$C_\infty \subset d - \underline{\lim}\, C_n := \bigcup_{m=1}^{\infty} \bigcap_{n=m}^{\infty} C_n. \tag{2.3}$$

Conditions (2.1) and (2.2) together say that C_n converges Mosco to C_∞ [18]. Condition (2.3) is generally more restrictive than (2.2); in general,

$$\mathrm{cl}(d - \underline{\lim}\, C_n) \subset s - \underline{\lim}\, C_n$$

with equality if $d - \underline{\lim}\, C_n$ has nonempty interior [16].

We will denote the value of the problem (P_n) by $V(P_n) \in [-\infty, +\infty]$, for $n = 1, 2, \ldots, \infty$. The following easy result is standard.

LEMMA 2.4. *If* (2.1) *holds, then* $\underline{\lim}\, V(P_n) \geq V(P_\infty)$ (*finite or infinite*).

Proof. Suppose not. Then for some $M < V(P_\infty)$, there is a subsequence $x_{n_r} \in C_{n_r}$, with $f(x_{n_r}) \leq M$ for all r. Since f has compact level sets, there is a weakly convergent subsequence $x_{n_{r'}}$, with limit x in $w - \overline{\lim}\, C_n$, and therefore in C_∞ by (2.1), such that $f(x) \leq M$. But then $V(P_\infty) \leq M$, which is a contradiction. □

LEMMA 2.5. *If* (2.3) *holds, then* $\overline{\lim}\, V(P_n) \leq V(P_\infty)$.

Proof. For any $x \in C_\infty$, (2.3) implies that $x \in C_n$ for all n sufficiently large, so $f(x) \geq V(P_n)$, and the result follows. □

By contrast, if we wish to weaken (2.3) to (2.2), we need to add a regularity condition, again using standard techniques.

Constraint Qualification. $\|\cdot\| - \text{int}(\text{dom}\, f) \cap C_\infty \neq \emptyset$. (The domain of f, dom f, is the set on which f is finite.) Clearly, if the Constraint Qualification and (2.2) hold, then for large n, a constraint qualification also holds for (P_n):

$$\|\cdot\| - \text{int}(\text{dom}\, f) \cap C_n \neq \emptyset.$$

LEMMA 2.6. *If* (2.2) *and the Constraint Qualification hold, then* $\overline{\lim}\, V(P_n) \leq V(P_\infty) < +\infty$.

Proof. By Proposition 3.3 in [19], f is continuous on the interior of its domain. Suppose x is arbitrary in int(dom f) $\cap\, C$. By (2.2) there exist $x_n \in C_n$ with $\|x_n - x\| \to 0$, and hence $V(P_n) \leq f(x_n) \to f(x)$. Thus $\overline{\lim}\, V(P_n) \leq f(x)$.

Now by the Constraint Qualification there is an x_0 in int(dom f) $\cap\, C_\infty$. Suppose y is arbitrary in (dom f) $\cap\, C_\infty$. For any λ in $(0, 1]$,

$$\lambda x_0 + (1 - \lambda) y \in C_\infty \cap \ \text{int}(\text{dom}\, f),$$

so by the previous paragraph, $\overline{\lim}\, V(P_n) \leq f(\lambda x_0 + (1-\lambda) y)$. But by Theorem 10.2 in [20], $f(\lambda x_0 + (1-\lambda)y) \to f(y)$ as $\lambda \downarrow 0$, so $\overline{\lim}\, V(P_n) \leq f(y)$. Since y was arbitrary, the result now follows. □

Very simple examples demonstrate that the above result may fail without the Constraint Qualification. The next lemma shows how we obtain weak convergence of the optimal solutions.

LEMMA 2.7. *Suppose* (2.1) *holds,* $V(P_n) \to V(P_\infty)$, *and* (P_∞) *has a unique optimum,* x_∞. *Then for any sequence of optimal solutions* x_n *for* (P_n), $x_n \to x_\infty$ *weakly.*

Proof. Suppose the result fails, so for some ϕ in X^* and some subsequence, $\langle x_{n_r} - x_\infty, \phi\rangle \geq 1$ for all r. Now since $f(x_{n_r}) \to f(x_\infty) < +\infty$ and f has compact level sets, there is a convergent subsequence $x_{n_{r_i}} \to \bar{x}$ weakly with $f(\bar{x}) \leq f(x_\infty)$ and $\bar{x} \in C_\infty$ by (2.1). Hence, by uniqueness, $\bar{x} = x_\infty$, which contradicts $\langle x_{n_{r_i}} - x_\infty, \phi\rangle \geq 1$ for all i. □

The final lemma will allow us to move from weak to strong convergence. As usual, $f^* : X^* \to (-\infty, +\infty]$ denotes the conjugate function

$$f^*(\phi) := \sup_{x \in X} \{\langle x, \phi\rangle - f(x)\}.$$

LEMMA 2.8. *Suppose* f^* *is Fréchet differentiable on the domain of* ∂f^*. *If* $x_n \to x_\infty$ *weakly,* $f(x_n) \to f(x_\infty) < +\infty$, *and* $\partial f(x_\infty) \neq \emptyset$, *then it follows that* $\|x_n - x_\infty\| \to 0$.

Proof. Pick $\phi \in \partial f(x_\infty)$, so $x_\infty \in \partial f^*(\phi) = \{\nabla f^*(\phi)\}$ the Fréchet derivative of f^* at ϕ. Now

$$\langle x_n, \phi\rangle - f(x_n) \to \langle x_\infty, \phi\rangle - f(x_\infty) = f^*(\phi),$$

so by Proposition 1 in [1], $\|x_n - x_\infty\| \to 0$. □

The result above shows that f^* being Fréchet differentiable at any point where a subgradient exists is almost enough to guarantee the Kadec property of f; the missing condition is $\partial f(x_\infty) \neq \emptyset$. It is easily checked that the differentiability condition also implies that f is strictly convex on any convex subset of dom ∂f (see [2]; cf. Chap. 26 in [20]). Note that the converse of the above result is not true. For example, if $f(x) := \|x\|_2$ on $L_2[0, 1]$, the conclusion of the lemma is true by the Kadec property but f^* is the indicator function of the unit ball, which clearly is not Fréchet differentiable at boundary points.

We can now compare two contrasting convergence results. The first is standard, using Mosco convergence and the Constraint Qualification, whereas the second uses the stronger convergence involving the discrete lim inf and strong rotundity.

THEOREM 2.9 (i) (Mosco version). *Suppose* (2.1) *and* (2.2) *hold* (*in other words,* $C_n \to C_\infty$ *Mosco*) *and the Constraint Qualification holds. Then* $V(P_n) \to V(P_\infty) < +\infty$. *Furthermore, if* f^* *is Fréchet differentiable on the domain of* ∂f^*, *then* (P_n) *and* (P_∞) *have unique optimal solutions* x_n *and* x_∞, *respectively* (*for all* n *sufficiently large*), *and* $\|x_n - x_\infty\| \to 0$.

(ii) (Discrete version). *Suppose* (2.1) *and* (2.3) *hold. Then* $V(P_n) \to V(P_\infty)$ (*finite or infinite*). *Furthermore, if* $V(P_\infty) < +\infty$ *and* f *is strongly rotund, then* (P_n) *and* (P_∞) *have unique optimal solutions* x_n *and* x_∞, *respectively* (*for all* n *sufficiently large*), *and* $\|x_n - x_\infty\| \to 0$.

Proof. (i) $V(P_n) \to V(P_\infty) < +\infty$ follows from Lemmas 2.4 and 2.6. If we write δ_{C_∞} for the indicator function of C_∞, when x_∞ is optimal for (P_∞),

$$0 \in \partial(f + \delta_{C_\infty})(x_\infty) = \partial f(x_\infty) + \partial\delta_{C_\infty}(x_\infty),$$

using the Constraint Qualification and Theorem 20 in [22]. Thus, we have $\partial f(x_\infty) \neq \emptyset$, so the set of optimal solutions of (P_∞) is a convex subset of dom ∂f. Since f must be strictly convex on this set, (P_∞) has a unique optimal solution, x_∞. A similar argument using the constraint qualification for (P_n) shows that for large n (P_n) has a unique optimal solution x_n, and $x_n \to x_\infty$ weakly by Lemma 2.7. Finally, we apply Lemma 2.8 to deduce $\|x_n - x_\infty\| \to 0$.

(ii) $V(P_n) \to V(P)$ follows from Lemmas 2.4 and 2.5. Strict convexity again ensures the uniqueness of x_n and x_∞, so Lemma 2.7 once more shows weak convergence, and then strong rotundity implies convergence in norm. □

Note. In fact, due to Lemma 2.4, we could relax the strong rotundity assumption in (ii) by replacing the Kadec property by the assumption that x_n tends weakly to x_∞ and $f(x_n) \uparrow f(x_\infty)$ implies that x_n tends to x_∞ in norm.

Example. If $C_1 \supset C_2 \supset C_3 \supset \cdots$ and $C_\infty = \bigcap_1^\infty C_n$, then (2.1) and (2.3) hold. Indeed, it is almost immediate that

$$w - \overline{\lim}\, C_n = \bigcap_{n=1}^{\infty} C_n = d - \underline{\lim}\, C_n.$$

Our second example is modeled on sequences of moment problems. In best entropy estimation we seek to estimate an unknown density given some of its moments by choosing that density (perhaps within some tolerance) satisfying the known moment conditions which minimizes a certain objective function (a measure of entropy). We seek conditions ensuring that the estimates converge in norm to the unknown density as the amount of moment information increases.

Example. Suppose the functions $a_1, a_2 \ldots$ are weak-star densely spanning in $L_\infty(S, \mu)$ and $\bar{x} \in L_1$. Suppose that for each i, n, F_i^n is a closed subset of $\mathbb{R}$, containing $\langle a_i, \bar{x}\rangle$, whose diameter tends to 0 as $n \to \infty$. Suppose finally that

$$C_n := \{x \in L_1 \mid \langle a_i, x\rangle \in F_i^n \text{ for all } i\}.$$

(Thus, C_n consists of those densities whose moments agree with those of $\bar{x}$ within certain tolerances.) Then $w - \overline{\lim}\, C_n = \{\bar{x}\} = d - \underline{\lim}\, C_n$.

To see this, note that certainly, since $\bar{x} \in C_n$ for all $n, \bar{x} \in w - \overline{\lim}\, C_n$ and $\bar{x} \in d - \underline{\lim}\, C_n$. On the other hand, if $x_{n_r} \in C_{n_r}$ and $x_{n_r} \to x$ weakly, for any i, $\langle a_i, x_{n_r} - \bar{x}\rangle \leq \text{diam } F_i^{n_r} \to 0$ as $r \to \infty$, so $\langle a_i, x - \bar{x}\rangle = 0$. Thus, $x = \bar{x}$, which demonstrates the first equality.

Finally, if $x \in d - \underline{\lim}\, C_n$, then $x \in C_n$ for all n sufficiently large, so $\langle a_i, x - \bar{x}\rangle \leq \text{diam } F_i^n \to 0$ as $n \to \infty$, which shows that $x = \bar{x}$. Hence, the second equality.

3. Strong rotundity. We shall suppose throughout that (S, μ) is a complete finite measure space (with nonzero μ), and $\phi : \mathbb{R} \to (-\infty, +\infty]$ is a proper, closed, convex function. We denote the interior of the domain of ϕ (assumed nonempty) by (α, β), where $-\infty \leq \alpha < \beta \leq +\infty$. Since ϕ is a normal convex integrand, we can define the proper weakly lower semicontinuous functional $I_\phi : L_1(S, \mu) \to (-\infty, +\infty]$ by $I_\phi(x) := \int_S \phi(x(s))d\mu(s)$, with conjugate $I_\phi^* : L_\infty(S, \mu) \to (-\infty, +\infty]$ given by $I_\phi^* = I_{\phi^*}$, where ϕ^* is the conjugate of ϕ (see [22]).

LEMMA 3.1. *I_ϕ is strictly convex on its domain if and only if ϕ is strictly convex on its domain.*

Proof. It is well known and straightforward that I_ϕ is strictly convex if ϕ is. The converse follows by considering constant functions. □

LEMMA 3.2. *I_{ϕ^*} is Fréchet differentiable everywhere on $L_\infty(S, \mu)$ if and only if ϕ^* is differentiable everywhere on $\mathbb{R}$.*

Proof. Suppose ϕ^* is differentiable on $\mathbb{R}$, so by Theorem 25.5 in [20] it is continuously differentiable. Given any $y \in L_\infty(S, \mu)$, pick m and M in $\mathbb{R}$ with $m \leq y(s) \leq M$ almost everywhere. Since $(\phi^*)'$ is uniformly continuous on $[m-1, M+1]$, for almost every s, given any $\epsilon > 0$, there is a $\delta > 0$ such that $|(\phi^*)'(y(s) + v) - (\phi^*)'(y(s))| < \epsilon$ whenever $|v| < \delta$. Thus, by the Mean Value Theorem, for some $v' \in (-\delta, \delta)$

$$\begin{aligned}
&|\phi^*(y(s) + v) - \phi^*(y(s)) - v(\phi^*)'(y(s))| \\
&\quad = |v|\, |(\phi^*)'(y(s) + v') - (\phi^*)'(y(s))| \\
&\quad \leq \epsilon |v|,
\end{aligned}$$

so if $\|h\|_\infty \leq \delta$,

$$\left| I_{\phi^*}(y + h) - I_{\phi^*}(y) - \int_S h(s)(\phi^*)'(y(s))d\mu(s) \right| \leq \epsilon \|h\|_\infty.$$

This demonstrates that I_{ϕ^*} has Fréchet derivative $\nabla I_{\phi^*}(y) = (\phi^*)'(y(\cdot))$ at y. The converse follows by considering constant functions. □

LEMMA 3.3. *If $x \in L_\infty(S, \mu)$ with* ess inf $x > \alpha$ *and* ess sup $x < \beta$, *then $\partial I_\phi(x) \neq \emptyset$.*

Proof. Denote the essential infimum and supremum of x by m and M, respectively. The multifunction $\partial\phi : [m, M] \to \mathbb{R}$ is upper semicontinuous [15]. In other words, for any closed set F in $\mathbb{R}$,

$$(\partial\phi)_F := \{u \in [m, M] \mid \partial\phi(u) \cap F \neq \emptyset\}$$

is closed. Thus, the multifunction $s \mapsto \partial\phi(x(s))$ is measurable, since

$$\{s \in S \mid \partial\phi(x(s)) \cap F \neq \emptyset\} = \{s \in S \mid x(s) \in (\partial\phi)_F\}$$

is measurable for any closed F. Since this multifunction has nonempty closed (actually compact) values, there is a measurable selection $y(s) \in \partial\phi(x(s))$ almost everywhere, by Theorem 14.2.2 in [15]. Since $-\infty < \min \partial\phi(m) \leq y(s) \leq \max \partial\phi(M) < +\infty$ almost everywhere, it follows that $y \in L_\infty(S, \mu)$; so, directly from the definition, $y \in \partial I_\phi(x)$. □

For any m, define (for a fixed $x \in L_1(S,\mu)$)

$$S_m := \left\{ s \in S \mid (-m) \vee \left(\alpha + \frac{1}{m}\right) \le x(s) \le \left(\beta - \frac{1}{m}\right) \wedge m \right\}.$$

LEMMA 3.4. *If $\alpha < x(s) < \beta$ almost everywhere, then $\mu(S_m^c) \downarrow 0$ as $m \to \infty$.*

Proof. The sets S_m are nested and increasing with

$$\bigcup_{m=1}^{\infty} S_m = \{s \mid \alpha < x(s) < \beta\}.$$

Thus, $\mu(S_m) \uparrow \mu(S)$ as $m \to \infty$. □

For any measurable subset T of S, we denote the restrictions of μ and x to T by $\mu|_T$ and $x|_T$, and define $I_\phi^T : L_1(T, d\mu|_T) \to (-\infty, +\infty]$ by $I_\phi^T(z) := \int_T \phi(z(s))d\mu(s)$.

LEMMA 3.5. *Suppose $x_n \to x$ weakly in $L_1(S,\mu)$ and $I_\phi(x_n) \to I_\phi(x) < +\infty$. Then for any measurable subset T of $S, x_n|_T \to x|_T$ weakly in $L_1(T,\mu|_T)$ and $I_\phi^T(x_n|_T) \to I_\phi^T(x|_T) < +\infty$.*

Proof. The weak convergence of $x_n|_T$ to $x|_T$ follows by integrating with respect to functions zero on T^c. Since I_ϕ^T and $I_\phi^{T^c}$ are both weakly lower semicontinuous, we have $\underline{\lim} I_\phi^T(x_n|_T) \ge I_\phi^T(x|_T)$ and $\underline{\lim} I_\phi^{T^c}(x_n|_{T^c}) \ge I_\phi^{T^c}(x|_{T^c})$ (since we also have $x_n|_{T^c} \to x|_{T^c}$ weakly in $L_1(T^c, \mu|_{T^c})$). On the other hand, since $I_\phi(z) = I_\phi^T(z|_T) + I_\phi^{T^c}(z|_{T^c})$ for any z,

$$\begin{aligned}
\limsup I_\phi^T(x_n|_T) &= \limsup(I_\phi(x_n) - I_\phi^{T^c}(x_n|_{T^c})) \\
&= I_\phi(x) - \liminf I_\phi^{T^c}(x_n|_{T^c}) \\
&\le I_\phi(x) - I_\phi^{T^c}(x|_{T^c}) \\
&= I_\phi^T(x|_T),
\end{aligned}$$

and the result follows. □

LEMMA 3.6. *Suppose ϕ^* is differentiable everywhere on $\mathbb{R}$. If $x_n \to x$ weakly in $L_1(s,\mu)$, $I_\phi(x_n) \to I_\phi(x) < +\infty$, and $\alpha < x(s) < \beta$ almost everywhere, then it follows that $\|x_n - x\|_1 \to 0$.*

Proof. Since $x_n \to x$ weakly, $(x_n)_1^\infty \cup \{x\}$ is weakly compact in $L_1(S,\mu)$. Given any $\epsilon > 0$, by the Dunford–Pettis criterion [9] there exists a $\delta > 0$ such that if $\mu(T) \le \delta$, $\int_T |x_n(s)|d\mu(s) < \epsilon$ for all n and $\int_T |x(s)|d\mu(s) < \epsilon$. By Lemma 3.4 there is an m with $\mu(S_m^c) \le \delta$, so $\int_{S_m^c} |x_n(s) - x(s)|d\mu(s) < 2\epsilon$ for all n.

Now by Lemma 3.5, as $n \to \infty, x_n|_{S_m} \to x|_{S_m}$ weakly in $L_1(S_m, \mu|_{S_m})$ and $I_\phi^{S_m}(x_n|_{S_m}) \to I_\phi^{S_m}(x|_{S_m}) < +\infty$, and certainly $x|_{S_m} \in L_\infty(S_m, \mu|_{S_m})$ with

$$\text{ess inf } x|_{S_m} > \alpha \quad \text{and} \quad \text{ess sup } x|_{S_m} < \beta.$$

Thus, by Lemma 3.3, $\partial I_\phi^{S_m}(x|_{S_m}) \ne \emptyset$, so we can apply Lemma 2.8 (on $L_1(S|_m, \mu|_{S_m})$) to deduce that $\int_{S_m} |x_n(s) - x(s)|d\mu(s) \to 0$ as $n \to \infty$.

Finally, since for all n,

$$\begin{aligned}
\|x_n - x\|_1 &= \int_{S_m} |x_n(s) - x(s)|d\mu(s) + \int_{S_m^c} |x_n(s) - x(s)|d\mu(s) \\
&< \int_{S_m} |x_n(s) - x(s)|d\mu(s) + 2\epsilon,
\end{aligned}$$

we obtain lim sup $\|x_n - x\|_1 \le 2\epsilon$. As ϵ was arbitrary, the result follows. □

THEOREM 3.7. *Suppose ϕ^* is differentiable everywhere on $\mathbb{R}$. If $x_n \to x$ weakly in $L_1(S,\mu)$ and $I_\phi(x_n) \to I_\phi(x) < +\infty$, then $\|x_n - x\|_1 \to 0$.*

Proof. Since $I_\phi(x) < +\infty, \alpha \le x(s) \le \beta$ almost everywhere and for all n sufficiently large $\alpha \le x_n(s) \le \beta$ almost everywhere. Define

$$S^\gamma := \{s \in S \mid \alpha < x(s) < \beta\},$$
$$S^\alpha := \{s \in S \mid x(s) = \alpha\},$$
$$S^\beta := \{s \in S \mid x(s) = \beta\}.$$

By Lemma 3.5, $x_n|_{S^\gamma} \to x|_{S^\gamma}$ weakly in $L_1(S^\gamma, \mu|_{S^\gamma})$, and

$$I_\phi^{S^\gamma}(x_n|_{S^\gamma}) \to I_\phi^{S^\gamma}(x|_{S^\gamma}) < +\infty,$$

so applying Lemma 3.6, $\int_{S^\gamma} |x_n(s) - x(s)| d\mu(s) \to 0$. But now, for all n sufficiently large,

$$\begin{aligned}\|x_n - x\|_1 &= \int_{S^\gamma} |x_n(s) - x(s)| d\mu(s) \\ &\quad + \int_{S^\alpha} |x_n(s) - x(s)| d\mu(s) + \int_{S^\beta} |x_n(s) - x(s)| d\mu(s) \\ &= \int_{S^\gamma} |x_n(s) - x(s)| d\mu(s) + \int_{S^\alpha} (x_n(s) - \alpha) d\mu(s) \\ &\quad + \int_{S^\beta} (\beta - x_n(s)) d\mu(s) \\ &\to 0,\end{aligned}$$

as $n \to \infty$. □

We can now prove the central result.

THEOREM 3.8. *If ϕ^* is differentiable everywhere on $\mathbb{R}$, then $I_\phi(\cdot)$ is strongly rotund on $L_1(S,\mu)$. The converse is also true if (S,μ) is not purely atomic.*

Proof. If ϕ^* is differentiable everywhere on $\mathbb{R}$, it is essentially smooth, so ϕ is essentially strictly convex by Theorem 26.3 in [20]. On $\mathbb{R}$ this is equivalent to ϕ being strictly convex on its domain, and hence I_ϕ is strictly convex on its domain (Lemma 3.1). Since ϕ is everywhere finite, I_ϕ has weakly compact level sets by Corollary 2B in [21]. Finally, Theorem 3.7 shows the Kadec property.

Conversely, using Lemma 3.1, strict convexity of ϕ and hence essential smoothness of ϕ^* follows from the strict convexity of I_ϕ, whereas I_ϕ having compact level sets implies that ϕ^* is everywhere finite by Theorem 2.10 in [5]. □

Note. In fact the above argument shows that if ϕ^* is differentiable everywhere on $\mathbb{R}$, then for any y in $L_\infty(S,\mu)$, the functional $I_\phi(\cdot) + \langle y, \cdot\rangle$ is strongly rotund.

The approach we have taken has been to work as far as possible via Fréchet differentiability of the conjugate. A different, more geometric technique may be found in [25], and another independent approach in [23]. Theorem 3 in [25] shows that the hypotheses of Theorem 3.7 also give that $\|\phi(x_n(\cdot)) - \phi(x(\cdot))\|_1 \to 0$.

As suggested in the previous section, our interest in strongly rotund functions arose from the study of best entropy estimation for moment problems. Many of the functions ϕ that are used in practice may be described in a unified framework based on a Bayesian statistical interpretation (see [7]). Each such choice of ϕ is associated with a unique "prior" nonnegative measure ρ on $\mathbb{R}$, via conjugation and (two-sided) Laplace transformation:

$$\phi^*(v) = \log \int_{-\infty}^{\infty} e^{uv} d\rho(u). \tag{3.9}$$

It may be easily checked from Hölder's inequality that ϕ^* is strictly convex (providing ρ is not supported on a single point) and is differentiable on the interior of its domain. In particular, when ϕ^* is everywhere finite our results show that I_ϕ is strongly rotund, implying convergence in L_1 norm of the associated estimates. Some examples follow:

(i) *Normal distribution* $d\rho = \frac{1}{\sqrt{2\pi}}e^{-u^2/2}du$. We obtain $\phi^*(v) = \frac{1}{2}v^2$, so $I_\phi(x) = \frac{1}{2}\|x\|_2^2$.

(ii) *Poisson distribution* $\rho = \frac{1}{e}\sum_{r=0}^{\infty}\frac{1}{r!}\delta_{\{r\}}$. We obtain $\phi^*(v) = e^{v-1}$, so $I_\phi(x) = \int x\log x$, which is (minus) the Boltzmann–Shannon entropy. The strong rotundity in this case was shown directly in [4].

(iii) *Binomial distribution* $\rho = \frac{1}{2}(\delta_0 + \delta_1)$. We obtain $\phi^*(v) = \log\big(\frac{1}{2}(1 + e^v)\big)$, and

$$\phi(u) = \begin{cases} u\log u + (1-u)\log(1-u) + \log 2 & \text{if } u \in (0,1), \\ +\infty & \text{otherwise,} \end{cases}$$

which is, up to a constant, (minus) the Fermi–Dirac entropy. It is curious that the general binomial distribution leads to the same entropy.

(iv) *Lebesgue measure.*

$$d\rho = \begin{cases} du & \text{on } \mathbb{R}_+, \\ 0 & \text{on } \mathbb{R}_-. \end{cases}$$

We obtain

$$\phi^*(v) = \begin{cases} -\log(-v) & \text{if } v < 0, \\ +\infty & \text{if } v \geq 0, \end{cases}$$

so

$$\phi(u) = \begin{cases} -\log(u) - 1 & \text{if } u > 0, \\ +\infty & \text{if } u \leq 0, \end{cases}$$

and I_ϕ is the Burg entropy [6]. In this case, since ϕ^* is not everywhere finite, I_ϕ does not have weakly compact level sets.

Not all of the entropies appearing in the literature may be encompassed in this framework. For example, the choice

$$\phi(u) := \begin{cases} \dfrac{1}{p}u^p & \text{if } u \geq 0, \\ +\infty & \text{if } u < 0 \end{cases} \tag{3.10}$$

is made in "L_p spectral estimation" $(1 < p < +\infty)$ (see [3] and [11]) and then $\phi^*(v) := \frac{1}{q}v_+^q$, where $\frac{1}{p} + \frac{1}{q} = 1$. Since ϕ^* is not strictly convex, it cannot be written in the form (3.9). Our results show that I_ϕ is strongly rotund in L_1, and in fact more is clearly true: since the unit ball in L_p is weakly compact and $\|\cdot\|_p$ is Kadec [9], I_ϕ is actually strongly rotund in L_p. Notice that when $p = 1$ in (3.10), we obtain

$$\phi^*(v) = \begin{cases} 0 & \text{if } v \leq 1, \\ +\infty & \text{if } v > 1, \end{cases}$$

which is not everywhere finite, and thus I_ϕ does not have weakly compact level sets (not surprisingly, since the unit ball in L_1 is not weakly compact).

4. Existence of strongly rotund functions. On any measure space (S, μ), if $1 < p < +\infty$, then $\|\cdot\|_p^p$ is a strongly rotund functional on $L_p(S)$ (see, for example, [9]), which is everywhere finite. Thus, as we have seen, sequences of optimization problems whose objective function is an L_p norm are particularly well behaved. Of course, the L_1 norm will not have this property; the results below show that the existence of everywhere finite, strongly rotund functions is characterized by reflexivity. The Boltzmann–Shannon entropy, $\int x \log x$, may be thought of as a surrogate reflexive norm on the positive cone in L_1 due to its strong rotundity.

LEMMA 4.1. *If V is a vector space, $f : V \to (-\infty, +\infty]$ is a convex function, and $\alpha \in (\inf f, +\infty)$, then* $\operatorname{dom} f \subset \operatorname{aff}\{v \in V \mid f(v) \leq \alpha\}$.

Proof. Pick any v_0 with $f(v_0) < \alpha$, and suppose $v_1 \in \operatorname{dom} f$. Providing $\lambda > 0$ is sufficiently small,

$$\begin{aligned} f((1-\lambda)v_0 + \lambda v_1) &\leq (1-\lambda)f(v_0) + \lambda f(v_1) \\ &= f(v_0) + \lambda(f(v_1) - f(v_0)) \\ &\leq \alpha, \end{aligned}$$

and since $v_1 = (1/\lambda)((1-\lambda)v_0 + \lambda v_1) + (1 - (1/\lambda))v_0$, the result follows. □

THEOREM 4.2. *Suppose that X is a Banach space and $f : X \to (-\infty, +\infty]$ is a convex function with weakly compact level sets, whose domain is spanning. Then X is reflexive.*

Proof. Choose any $\alpha \in (\inf f, +\infty)$ and define $L := \{x \in X \mid f(x) \leq \alpha\}$, which, by assumption, is weakly compact. Thus, the closed, absolutely convex hull of L, $\overline{\mathrm{aco}}\, L$, is also weakly compact [13, p. 162]. But now we have

$$X = \operatorname{span}(\operatorname{dom} f) \subset \operatorname{span}(\operatorname{aff} L) = \operatorname{span} L \subset \operatorname{span}(\overline{\mathrm{aco}}\, L) = \bigcup_{n=1}^{\infty} n(\overline{\mathrm{aco}}\, L)$$

using Lemma 4.1. Now by Baire category, $\overline{\mathrm{aco}}\, L$ has nonempty interior, so for some $r > 0$ and $\bar{x}$ in X the closed ball $\bar{B}(\bar{x}; r) \subset \overline{\mathrm{aco}}(L)$, and so is weakly compact. Thus X is reflexive [13, p. 126]. □

Suppose that we wished to select an objective function for solving sequences of moment problems that would guarantee convergence in the norm $\|\cdot\|_1$ of estimates to an arbitrary unknown, nonnegative, L_1 density. To ensure convergence of the estimates using our methods, we would require a strongly rotund functional on the space $L_1(S, \mu)$ whose domain contained the positive cone and hence was spanning. Since $L_1(S, \mu)$ is not reflexive unless the measure space (S, μ) decomposes as a finite set of atoms, the above result shows this to be impossible: since the level sets of the objective function cannot be weakly compact, we cannot even ensure attainment.

Example. Suppose X is a reflexive Banach space. Then X has a renorming $\|\cdot\|$ that is Fréchet differentiable and locally uniformly convex (and whose dual norm has the same properties) [8, p. 167]. It is then easy to check that $f(x) := \|x\|^2$ is everywhere finite and strongly rotund.

COROLLARY 4.3. *Suppose X is a Banach space. Then there exists an everywhere finite strongly rotund function $f : X \to \mathbb{R}$ if and only if X is reflexive.*

Suppose instead that we ask just for a strongly rotund functional on $L_1(S, \mu)$ whose domain is *dense*. Assuming (S, μ) is finite, we could use $f(x) := \int x(s)^2 ds$. This is a consequence of the following general construction.

THEOREM 4.4. *Suppose X and Y are normed spaces, $T : X \to Y$ is continuous and linear, and $f : X \to (-\infty, +\infty]$ is strongly rotund. Define $g : Y \to (-\infty, +\infty]$ by $g(y) := \inf\{f(x) \mid Tx = y\}$. Then g is strongly rotund.*

Proof. Because the level sets of f are weakly compact, for any x in $\mathbb{R}$, $g(y) \leq \alpha$ if and only if there exists an x with $Tx = y$ and $f(x) \leq \alpha$, so

$$\{y \in Y \mid g(y) \leq \alpha\} = T\{x \in X \mid f(x) \leq \alpha\}.$$

Thus the level sets of g are weakly compact. A similar calculation shows that dom $g =$ T(dom f).

Now suppose that $y_1 \neq y_2$ in dom g and $\lambda \in (0,1)$. By compactness, $g(y_i) = f(x_i)$ for some x_i with $Tx_i = y_i, i = 1,2$, and because

$$T(\lambda x_1 + (1-\lambda)x_2) = \lambda y_1 + (1-\lambda)y_2,$$

we have

$$\begin{aligned} g(\lambda y_1 + (1-\lambda)y_2) &\leq f(\lambda x_1 + (1-\lambda)x_2) \\ &< \lambda f(x_1) + (1-\lambda)f(x_2) \\ &= \lambda g(y_1) + (1-\lambda)g(y_2). \end{aligned}$$

Thus g is strictly convex.

It remains to demonstrate the Kadec property, so suppose $y_n \to y_\infty$ weakly in Y and $g(y_n) \to g(y_\infty) < +\infty$. By compactness, $g(y_n) = f(x_n)$ for some x_n with $Tx_n = y_n, n = 1,2,\ldots,\infty$. Pick an arbitrary subsequence (y_{n_i}). Since f has weakly compact level sets, there is a subsequence $(x_{n_{i_j}})$ converging weakly to some $\bar{x}$ in X. Now note that

$$T\bar{x} = \text{w-lim}\, Tx_{n_{i_j}} = \text{w-lim}\, y_{n_{i_j}} = y_\infty,$$

and so, by the lower semicontinuity of f,

$$g(y_\infty) \leq f(\bar{x}) \leq \underline{\lim}\, f(x_{n_{i_j}}) = \underline{\lim}\, g(y_{n_{i_j}}) = g(y_\infty).$$

Thus $g(y_\infty) = f(\bar{x})$, and since f is strictly convex, $\bar{x} = x_\infty$. Now we have $x_{n_{i_j}} \to x_\infty$ weakly in X and $f(x_{n_{i_j}}) \to f(x_\infty) < +\infty$, so by the strong rotundity of f, $\|x_{n_{i_j}} - x_\infty\| \to 0$, and thus $\|y_{n_{i_j}} - y_\infty\| \to 0$ by continuity. Since (y_{n_i}) was an arbitrary subsequence, we deduce that $\|y_n - y_\infty\| \to 0$ as required. □

COROLLARY 4.5. *Suppose (S,μ) is a finite measure space,$1 < r < +\infty, 1 \leq p \leq r$, and $f : L_r(S,\mu) \to (-\infty,+\infty]$ is strongly rotund. Then the functional $g : L_p(S,\mu) \to (-\infty,+\infty]$ defined by*

$$g(y) := \begin{cases} f(y) & \text{if } y \in L_r, \\ +\infty & \text{otherwise}, \end{cases}$$

is strongly rotund.

Proof. Let $T : L_r \to L_p$ be the (continuous) embedding, and apply Theorem 4.4. □

Example. For any $1 < r < +\infty$, $\|\cdot\|_r^r$ is strongly rotund on $L_r(S,\mu)$, so providing (S,μ) is finite, $g(y) := \int y(s)^r ds$ will be strongly convex on $L_p(S,\mu)$ for any $1 \leq p \leq r$.

Example. Let $X := L_1[0,1] \times \mathbb{R}, Y := C[0,1]$, and then proceed to define $T : X \to Y$ by $(T(x;r))(t) := r + \int_0^t x(s)ds$. If we define $f : X \to (-\infty,+\infty]$ by

$$f(x;r) := \int_0^1 \phi(x(s))ds + \phi(r),$$

where ϕ is (minus) the Boltzmann–Shannon entropy,

$$\phi(u) := \begin{cases} u \log u & \text{if } u > 0, \\ 0 & \text{if } u = 0, \\ +\infty & \text{if } u < 0, \end{cases}$$

then f is strongly rotund by Theorem 3.8.

It is now easy to verify that the functional $g : C[0,1] \to (-\infty, +\infty]$ of Theorem 4.4 is given by

$$g(y) := \begin{cases} \displaystyle\int_0^1 \phi\left(\frac{dy}{ds}\right) ds + \phi(y(0)), & \text{if } \dfrac{dy}{ds} \in L_1[0,1], \\ +\infty, & \text{otherwise,} \end{cases}$$

and so g is strongly rotund.

Now consider the sequence of moment problems

$$(M_n) \begin{cases} \text{inf} & g(y) \\ \text{subject to} & \langle a_i, y \rangle = \langle a_i, \bar{y} \rangle, \quad i = 1, \ldots, n, \\ & y \in C[0,1], \end{cases}$$

where the functions $a_1, a_2, \ldots,$ are weak-star densely spanning in $M[0,1]$, the regular Borel measures, and $\bar{y}$ is a fixed, monotonically increasing, nonnegative function. Then, providing $g(\bar{y}) < +\infty$ (which will be the case, for example, if $\bar{y}$ is Lipschitz on [0,1]), we can apply Theorem 2.9(ii) to deduce the uniform convergence of the optimal solutions of (M_n) to $\bar{y}$.

The construction in Theorem 4.4 also allows us to characterize those spaces on which there exist strongly rotund functions with dense domain. We say a normed space is *weakly compactly generated* (WCG) if there exists a weakly compact convex set whose linear span is dense (see [8]). Any reflexive space is WCG, as is any separable space. For any σ-finite measure space (S, μ), $L_1(S, \mu)$ is WCG [8, p. 143].

THEOREM 4.6. *Suppose Y is a Banach space and $f : Y \to (-\infty, +\infty]$ is a convex function with weakly compact level sets, whose domain is densely spanning. Then Y is WCG.*

Proof. Choose any α in $(\inf f, +\infty)$ and define $L := \{y \in Y \mid f(y) \leq \alpha\}$. By Lemma 4.1, span(dom f) $\subset$ span L, so the weakly compact, convex set L is densely spanning, as required. □

Suppose we wished to select an objective function for solving sequences of moment problems that would guarantee convergence in $\|\cdot\|_\infty$ of estimates to an unknown density. To ensure convergence using our methods to even a dense subset of the possible nonnegative, essentially bounded, unknown densities, we would require a strongly convex functional on $L_\infty(S, \mu)$ whose domain is dense in the positive cone. The above result demonstrates that this is impossible, since $L_\infty(S, \mu)$ is not WCG unless (S, μ) decomposes into a finite number of atoms. To see this, note that in any other case $L_1(S, \mu)$ contains a copy of l_1 as a subspace. Since any subspace of an Asplund space is Asplund, and l_1 is not Asplund, $L_1(S, \mu)$ is not an Asplund space, and hence its dual, $L_\infty(S, \mu)$, is not WCG [19].

By contrast, if we choose

$$\phi(u) = \begin{cases} e^u, & u \geq 0, \\ +\infty, & u < 0, \end{cases}$$

then using I_ϕ as the objective functional will ensure convergence in $\|\cdot\|_p$ to any nonnegative, essentially bounded, unknown density for every $p < +\infty$, provided (S, μ) is finite (see [5]).

THEOREM 4.7. *Suppose Y is a Banach space. Then there is a strongly rotund function on Y with dense domain if and only if Y is WCG.*

Proof. By the Davis–Figiel–Johnson–Pekzynski factorization theorem (see, for example, [9]), if Y is WCG there is a reflexive Banach space X and a continuous linear map $T : X \to Y$ with dense range. By Corollary 4.3 there is an everywhere finite strongly rotund function f on X. The function g of Theorem 4.4 is then strongly rotund on Y and has domain $T(\text{dom } f)$, which is dense. The converse follows by Theorem 4.6. $\square$

We end with examples of weak-star (weak*) strongly rotund functions: strictly convex functions $f : Y^* \to (-\infty, +\infty]$ on a dual space Y^* whose level sets are weak*-compact and such that $\theta_n \to \theta$ weak* and $f(\theta_n) \to f(\theta)$ implies $\|\theta_n - \theta\|_* \to 0$.

Let Γ be any nonempty set. Then $c_0(\Gamma)$ is WCG [8, p. 143], and so has an equivalent norm $|||\cdot|||$ whose dual norm $|||\cdot|||_*$ on $l_1(\Gamma)$ is strictly convex [8, p. 167]. The standard norm $\|\cdot\|_1$ on $l_1(\Gamma)$ has the weak* Kadec property, by Theorem 13.47 in [12]: if $x_n \to x$ weak* and $\|x_n\|_1 \to \|x\|_1$, then $\|x_n - x\|_1 \to 0$.

Now define a new norm on $l_1(\Gamma)$ by $|x| := |||x|||_* + \|x\|_1$. This norm is strictly convex, since $|||\cdot|||_*$ is, and it has weak*-compact level sets, since both $|||\cdot|||_*$ and $\|\cdot\|_1$ are dual norms and have weak*-compact balls. If we now define $f : l_1(\Gamma) \to \mathbb{R}$ by $f(x) = \frac{1}{2}|x|^2$, then it is clear that f is strictly convex with weak*-compact level sets. To see that it is actually weak* strongly rotund, suppose $x_n \to x$ weak* and $f(x_n) \to f(x)$. By weak* lower-semicontinuity, $\underline{\lim}\|x_n\|_1 \geq \|x\|_1$, while $\underline{\lim}|||x_n|||_* \geq |||x|||_*$, so

$$\overline{\lim}\|x_n\|_1 = \overline{\lim}(|x_n| - |||x_n|||_*) = |x| - \underline{\lim}|||x_n|||_* \leq |x| - |||x|||_* = \|x\|_1.$$

Thus, $\|x_n\|_1 \to \|x\|_1$, so $\|x_n - x\|_1 \to 0$ as required.

Notice that $f^* : l_\infty(\Gamma) \to \mathbb{R}$ is given by $f^*(z) = \frac{1}{2}|z|_*^2$, where $|\cdot|_*$ is the dual norm. Thus, f^* is certainly not Fréchet differentiable everywhere, since this would imply that $|\cdot|_*$ was Fréchet and so $l_1(\Gamma)$ would have to be reflexive [8, p. 34], which fails whenever Γ is infinite.

In general, whenever X is a separable Banach space with a separable dual, X can be renormed (by $\|\cdot\|$, say) in such a way that the dual norm $\|\cdot\|_*$ is strictly convex and has the weak* Kadec property [14]. It then follows that $\frac{1}{2}\|\cdot\|_*^2$ is weak* strongly rotund.

REFERENCES

[1] E. Asplund, *Fréchet differentiability of convex functions*, Acta Math., 121 (1968), pp. 31–47.

[2] E. Asplund and R. T. Rockafellar, *Gradients of convex functions*, Trans. Amer. Math. Soc., 139 (1969), pp. 443–467.

[3] A. Ben-Tal, J. M. Borwein, and M. Teboulle, *A dual approach to multidimensional L_p spectral estimation problems*, SIAM J. Control Optim., 26 (1988), pp. 985–996.

[4] J. M. Borwein and A. S. Lewis, *Convergence of best entropy estimates*. SIAM J. Optim., 1 (1991), pp. 191–205.

[5] ———, *On the convergence of moment problems*, Trans. Amer. Math. Soc., 325 (1991), pp. 249–271.

[6] J. P. Burg, *Maximum entropy spectral analysis*, paper presented at 37th Meeting of the Society of Exploration Geophysicists, Oklahoma City, 1967.

[7] D. Dacunha-Castelle and F. Gamboa, *Maximum d'entropie et problème des moments*, Ann. Inst. H. Poincaré, 26 (1990), pp. 567–596.

[8] J. Diestel, *Geometry of Banach Spaces—Selected Topics*, Lecture Notes in Mathematics 485, Springer-Verlag, New York, 1975.

[9] ———, *Sequences and Series in Banach Spaces*, Springer-Verlag, New York, 1984.

[10] A. Dontchev and T. Zolezzi, *Well-Posed Optimization Problems*, Springer-Verlag, New York, 1993.

[11] R. K. Goodrich and A. Steinhardt, *L_2 spectral estimation*, SIAM J. Appl. Math., 46 (1986), pp. 417–428.

[12] E. Hewitt and K. Stromberg, *Real and Abstract Analysis*, Springer-Verlag, New York, 1975.

[13] R. B. Holmes, *Geometric Functional Analysis and Its Applications*, Springer-Verlag, New York, 1975.

[14] V. L. Klee, *Mappings into normed linear spaces*, Fund. Math., 49 (1960), pp. 25–34.

[15] E. Klein and A. C. Thompson, *Theory of Correspondences*, Wiley, New York, 1984.

[16] A. S. Lewis and R. Lucchetti, Private communication, 1991.
[17] R. Lucchetti and F. Patrone, *Hadamard and Tyhonov well-posedness of a certain class of convex functions*, J. Math. Anal. Appl., 88 (1982), pp. 204–215.
[18] U. Mosco, *Convergence of convex sets and solutions of variational inequalities*, Adv. Math., 3 (1969), pp. 510–585.
[19] R. R. Phelps, *Convex Functions, Monotone Operators and Differentiability*, Lecture Notes in Mathematics 1364, Springer-Verlag, New York, 1989.
[20] R. T. Rockafellar, *Convex Analysis*, Princeton University Press, Princeton, NJ, 1970.
[21] ———, *Integrals which are convex functionals*, II, Pacific J. Math., 39 (1971), pp. 439–469.
[22] ———, *Conjugate Duality and Optimization*, Society for Industrial and Applied Mathematics, Philadelphia, PA, 1974.
[23] M. Teboulle and I. Vajda, *Convergence of best ϕ-entropy estimates*, IEEE Trans. Inform. Theory, 39 (1993), pp. 297–301.
[24] A. N. Tyhonov, *On the stability of the functional optimization problem*, U.S.S.R. Comput. Math. and Phys., 6 (1966), pp. 28–33.
[25] A. Visintin, *Strong convergence results related to strict convexity*, Comm. Partial Differential Equations, 9 (1984), pp. 439–466.