SIAM J. CONTROL OPTIM.
Vol. 42, No. 2, pp. 576–595
© 2003 Society for Industrial and Applied Mathematics

Downloaded 02/24/16 to 128.95.104.109. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

VARIATIONAL ANALYSIS APPLIED TO THE PROBLEM OF OPTICAL PHASE RETRIEVAL*

JAMES V. BURKE† AND D. RUSSELL LUKE‡

Abstract. We apply nonsmooth analysis to a well-known optical inverse problem, phase retrieval. The phase retrieval problem arises in many different modalities of electromagnetic imaging and has been studied in the optics literature for over forty years. The state of the art for this problem in two dimensions involves iterated projections for solving a nonconvex feasibility problem. Despite widespread use of these algorithms, current mathematical theory cannot explain their success. At the heart of projection algorithms is a nonconvex, nonsmooth optimization problem. We obtain some insight into these algorithms by applying techniques from nonsmooth analysis. In particular, we show that the weak closure of the set of directions toward the projection generate the subdifferential of the corresponding squared set distance function. Following a pattern of proof described in [F. H. Clarke, Yu. S. Ledyaev, R. J. Stern, and P. R. Wolenski, *Nonsmooth Analysis and Control Theory*, Springer-Verlag, New York, 1998], this result is generalized to provide conditions under which the subdifferential of an integral function equals the integral of the subdifferential.

Key words. phase retrieval, least squares, nonsmooth analysis, variational analysis

AMS subject classifications. 78A45, 93E24, 49J52, 49J53

PII. S0363012902406436

1. Introduction.

The phase retrieval problem arises frequently in a number of different optical imaging modalities including diffraction imaging and interferometry. While the imaging models differ slightly, the feature common to these techniques is the problem of recovering the phase of a complex-valued function from measurements of the amplitude of that function, as well as other a priori constraints. There are many unsolved mathematical problems surrounding wavefront reconstruction and phase retrieval in general. Nevertheless, engineers and physicists have been solving this problem in some sense for over thirty years. The most famous application of phase retrieval came with NASA's Hubble Space Telescope (HST). Optical wavefront reconstruction played a central role in the effort to identify gross manufacturing errors in the HST and to design, in effect, a pair of glasses for the near-sighted telescope. We refer the reader to [16] for a review and tutorial of wavefront reconstruction. Here we present only the abstract setting.

The *forward imaging* model is formulated on the space $L^2[\mathbb{R}^2, \mathbb{R}^2]$ of square integrable functions mapping $\mathbb{R}^2$ to $\mathbb{R}^2$. The model input $u : \mathbb{R}^2 \to \mathbb{R}^2$ is an optical field generated by the object we are trying to observe. The optical device is characterized by a unitary bounded linear operator $\mathcal{F}_m : L^2[\mathbb{R}^2, \mathbb{R}^2] \to L^2[\mathbb{R}^2, \mathbb{R}^2]$. The subscript m indicates certain parameter settings in the optical device that constitute a particular known "tuning" such as focus. Let $\mathbb{R}_+$ denote the nonnegative orthant. The model output, or data, corresponding to the mth tuning of the device, $\psi_m : \mathbb{R}^2 \to \mathbb{R}_+$, is

*Received by the editors April 26, 2002; accepted for publication (in revised form) November 22, 2002; published electronically May 12, 2003.
http://www.siam.org/journals/sicon/42-2/40643.html

†Department of Mathematics, University of Washington, Seattle, WA 98195 (burke@math.washington.edu). This author's work was partially supported by NSF grant DMS-0203175.

‡Corresponding author. Institut für Numerische und Angewandte Mathematik, Universität Göttingen, Germany (luke@math.uni-goettingen.de). This author's work was partially supported by NASA grant NGT 5-66 and a Postdoctoral Fellowship from the Pacific Institute for the Mathematical Sciences through Simon Fraser University.

Downloaded 02/24/16 to 128.95.104.109. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

amplitude measurements. The imaging model is given by

$$|\mathcal{F}_m(u(\cdot))| = \psi_m(\cdot), \qquad m = 0, 1, \ldots, M, \tag{1}$$

where the modulus $|\cdot|$ is the *pointwise Euclidean* magnitude. Our discussion switches frequently between the finite- and infinite-dimensional settings. Whenever there is chance for confusion, we indicate a mapping F on the function space explicitly as $F(u(\cdot))$.

Wavefront reconstruction is an *inverse problem*: given $\mathcal{F}_m$ and ψ_m, $m = 0, 1, \ldots, M$, determine u satisfying (1). For a more detailed review of the existence and uniqueness theory behind this problem we refer to [16] and references therein. For our purposes it suffices to note that there is no known closed-form solution to this inverse problem. Moreover, in the presence of noise it is likely that a solution does not exist, thus solution techniques involve minimizing a performance measure. Even though the performance measure that we consider is smooth, the modulus in (1) leads to a nonsmooth objective (see Theorem 3.1 in section 3). At first glance, it would seem that one could easily handle nonsmoothness by squaring both sides of (1). It turns out, however, that objectives based on the *modulus* function, or a nearby smooth approximation, perform better than objectives built upon the modulus squared [16]. Therefore, it can be advantageous to exploit nonsmoothness rather than to avoid it.

Since noise in the data is most often modeled as additive white noise, the least squares error metric is used to find the best fit to (1). For $m = 0, 1, \ldots, M$ and ψ_m not equal to zero a.e., define

$$\mathbb{Q}_m := \left\{ u \in L^2[\mathbb{R}^2, \mathbb{R}^2] \;\middle|\; |\mathcal{F}_m(u)| = \psi_m \text{ a.e.} \right\}. \tag{2}$$

The phase retrieval problem is given by

$$\begin{array}{ll} \text{minimize} & J(u) \\ \text{over} & u \in L^2[\mathbb{R}^2, \mathbb{R}^2], \end{array} \tag{3}$$

where

$$J(u) = \sum_{m=0}^{M} \frac{\beta_m}{2} \operatorname{dist}^2(u; \mathbb{Q}_m) \tag{4}$$

is the weighted ($\beta_m > 0$ for $m = 0, \ldots, M$) squared set distance error for the phase retrieval problem and

$$\operatorname{dist}(u; \mathbb{Q}_m) := \inf_{w \in \mathbb{Q}_m} \|u - w\|. \tag{5}$$

The error metric (4) has a long tradition in the optics literature [9, 10]. It has also been studied in the convex setting where each of the sets $\mathbb{Q}_m$ is assumed to be convex (e.g., see [2, 7]).

Problem (3) is often reformulated as a feasibility problem: the function u must lie in the intersection of the sets $\mathbb{Q}_0 \cap \mathbb{Q}_1 \cap \cdots \mathbb{Q}_m$, assuming that this intersection is nonempty. Projection algorithms are often used to find a point in the intersection of such a collection of sets. Independent of the mathematical literature on projections (and in some cases *before* these algorithms appeared in the mathematical literature) optical scientists developed image processing algorithms for recovering the phase from amplitude measurements known in the optics literature as *iterative transform methods*.

Downloaded 02/24/16 to 128.95.104.109. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

Here one adjusts the phase of the current estimate, $u^{(\nu)}$, at iteration ν by replacing the magnitude of the image $\mathcal{F}_m(u^{(\nu)}(\cdot))$ with the known pointwise magnitude $\psi_m(\cdot)$ and then inverse transforming the result, $\mathcal{F}_m^*(\psi_m(\cdot)\exp(\sqrt{-1}\ \arg(\mathcal{F}_m(u^{(\nu)}(\cdot)))))$. It is straightforward to show that this operation is a projection [16]. The Gerchberg–Saxton algorithm [10] is a classical example of this type of algorithm. When the sets $\mathbb{Q}_m$ are convex and the intersection is nonempty, then this approach is perfectly reasonable since cyclic projections onto such a finite collection of convex sets converges to the intersection (e.g., see [3] and the references therein). In the setting of phase retrieval, however, the sets $\mathbb{Q}_m$ are not even weakly closed, let alone convex [16, Property 4.1]. This poses serious challenges to any convergence theory for algorithms based on projections. Not surprisingly, many have noted that iterative transform algorithms often stagnate. There are some well-known strategies for dealing with these problems [9], but it has recently been observed that these too are applications of convex operator splitting strategies in nonconvex, nonlinear settings [4], so convergence is still problematic.

To overcome some of the problems inherent in treating the leading algorithms as nonconvex instances of projection algorithms, we approach the problem in its variational form (3) using the tools of nonsmooth analysis. We show that, for the squared set distance error metric (4), some projection algorithms can be viewed as *subgradient descent algorithms*. Thus, the critical object for our analysis is the *subdifferential*, or generalized derivative of the squared set distance error metric $J(u)$. In this analysis, the space to which the data $\{\psi_m : m = 0, 1, \ldots M\}$ belongs is of critical importance. We require these functions to be nonnegative and finite-valued with their value tending to zero as their argument diverges to infinity in norm. Specifically, we assume that the data belongs to the set $\mathbb{U}$ where

$$\mathbb{U} = \left\{ v \in L^1 \cap L^2 \cap L^\infty[\mathbb{R}^2, \mathbb{R}] \text{ such that } v(x) \geq 0 \text{ a.e. and } |v(x)| \to 0 \text{ as } |x| \to \infty \right\}. \tag{6}$$

In section 2 we review the theory of projections applied to this problem. The most common projection algorithms, stated in general form in section 2.3, are central to current numerical techniques for this problem. In section 3, we look at the problem from the perspective of nonsmooth least squares, beginning first with finite-dimensional nonsmooth analysis in section 3.2 and building toward the infinite-dimensional analysis in section 3.5. We then apply these results to the problem of wavefront reconstruction in section 3.6. In the final section of the paper we present a result on the exchange of subdifferentiation and integration. Such results have a long history, beginning with Rockafellar's result [20] for convex normal integrands. Our result is in the spirit of [6, Theorem 3.5.18]. Indeed, our method of proof parallels that given by Clarke, Ledyaev, Stern, and Wolenski. The key difference between our result and [6, Theorem 3.5.18] is that our domain of integration is all of $\mathbb{R}^2$ as opposed to an interval in $\mathbb{R}$.

2. Geometric approaches.

2.1. Projections. In general, it may be difficult to prove that the projection of a given point onto a given set exists, much less to identify it with a formula. Much of the general theory of projections [24] does not apply since the sets in question are neither weakly closed nor convex [16, Property 4.1]. However, in the application to phase retrieval there is a very simple characterization in terms of pointwise, finite-dimensional projections.

Downloaded 02/24/16 to 128.95.104.109. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

Our focus is on sets of the form

$$\mathbb{Q}(b) := \left\{ u \in L^2[\mathbb{R}^2, \mathbb{R}^2] \ \mid |u| = b \text{ a.e.} \right\}. \tag{7}$$

Here the set $\mathbb{Q}(b)$ is parameterized by the *function* $b : \mathbb{R}^2 \to \mathbb{R}_+$. Alternatively, one can think of this set as being parameterized *pointwise* by $x \in \mathbb{R}^2$, that is, at each point x, the set $\mathbb{Q}(b(x)) \subset \mathbb{R}^2$ is simply the sphere of radius $b(x)$, denoted $b(x)\mathbb{S}$, where $\mathbb{S}$ is the unit sphere in $\mathbb{R}^2$. For the closed set $\mathbb{Q}$ in the Hilbert space $\mathbb{X}$, we define the projection operator $\Pi_{\mathbb{Q}}(v)$ as the multivalued mapping, or multifunction, given as the set of all solutions to the minimum distance problem for the set $\mathbb{Q}$:

$$\Pi_{\mathbb{Q}}(v) := \arg\min_{u \in \mathbb{Q}} \|v - u\| = \{\bar{u} \in \mathbb{Q} : \|v - \bar{u}\| = \inf_{u \in \mathbb{Q}} \|v - u\|\}. \tag{8}$$

It is a simple matter to characterize the pointwise projection $\Pi_{b(x)\mathbb{S}} : \mathbb{R}^2 \rightrightarrows \mathbb{R}^2$:

$$\Pi_{b(x)\mathbb{S}}(v) = b(x)\Pi_{\mathbb{S}}(v) = b(x) \times \begin{cases} \frac{v}{|v|} & \text{for } v \neq 0, \\ \mathbb{S} & \text{for } v = 0, \end{cases} \qquad v \in \mathbb{R}^2. \tag{9}$$

Note that the projection is multivalued at $v = 0$. In the following sections we construct the infinite-dimensional projection $\Pi_{\mathbb{Q}(b)} : L^2[\mathbb{R}^2, \mathbb{R}^2] \rightrightarrows L^2[\mathbb{R}^2, \mathbb{R}^2]$ onto $\mathbb{Q}(b)$ from the corresponding pointwise projection at the point x, $\Pi_{b(x)\mathbb{S}} : \mathbb{R}^2 \rightrightarrows \mathbb{R}^2$ onto $b(x)\mathbb{S}$.

2.2. Measurable multifunctions. We now review some of the properties of measurable multifunctions used in this study [1, 6, 11, 21]. In section 3.3 we extend this review to include the integration theory of measurable multivalued mappings. For more information on this and related topics, we refer the interested reader to [21, chapter 14].

Let $\Omega \neq \emptyset$ and let $\mathcal{A}$ be a σ-field of subsets of Ω, called the *measurable* subsets of Ω or the $\mathcal{A}$-*measurable* subsets. The corresponding measure space is denoted $(\Omega, \mathcal{A})$. Our discussion is limited to *complete nonatomic* measure spaces.

The multifunction $F : \Omega \rightrightarrows \mathbb{R}^n$ is said to be $\mathcal{A}$-measurable, or simply measurable, if for all open sets $\mathbb{V}$ the set $\{x \mid \mathbb{V} \cap F(x) \neq \emptyset\}$ is in $\mathcal{A}$. The multifunction F is said to be $\mathcal{A} \otimes \mathcal{B}^n$-*measurable* if $\text{gph}(F) = \{(x, v) \mid v \in F(x)\} \in \mathcal{A} \otimes \mathcal{B}^n$. Here $\mathcal{B}^n$ denotes the Borel σ-field on $\mathbb{R}^n$ and $\mathcal{A} \otimes \mathcal{B}^n$ is the σ-field on $\Omega \times \mathbb{R}^n$ generated by all sets $A \times D$ with $A \in \mathcal{A}$ and $D \in \mathcal{B}^n$. If $F(x)$ is closed for each x, then F is *closed*. Similarly, F is said to be *convex* if $F(x)$ is convex for each x. Finally, we note that the completeness of the measure space guarantees the measurability of subsets of Ω obtained as the projections of measurable subsets $\mathbb{G}$ of $\Omega \times \mathbb{R}^n$:

$$\mathbb{G} \in \mathcal{A} \otimes \mathcal{B}^n \quad \Longrightarrow \quad \{\omega \in \Omega \mid \exists\, x \in \mathbb{R}^n \text{ with } (\omega, x) \in \mathbb{G}\} \in \mathcal{A},$$

and thus F is $\mathcal{A}$-measurable if and only if F is $\mathcal{A} \otimes \mathcal{B}^n$-measurable [21, Theorem 14.8].

Let $F : \Omega \rightrightarrows \mathbb{R}^n$. Denote by $\mathcal{S}(F)$ the set of μ-measurable functions $f : \Omega \to \mathbb{R}^n$ that satisfy $f(x) \in F(x)$ a.e. in Ω $(x \in \Omega)$. We call $\mathcal{S}(F)$ the *set of measurable selections* of F.

THEOREM 2.1 (measurable selections [21, Corollary 14.6]). *A closed-valued measurable map $F : \Omega \rightrightarrows \mathbb{R}^n$ always admits a measurable selection.*

For a measurable function $f = (f_1, \ldots, f_n)$, $f_i : \Omega \to \mathbb{R}$, for $i = 1, \ldots, n$, the integral $\int f d\mu$ is defined to be the vector

$$\left(\int f_1, d\mu, \ldots, \int f_n d\mu \right).$$

Downloaded 02/24/16 to 128.95.104.109. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

The set

$$\left\{ \int f d\mu \mid f \in \mathcal{S}(F) \right\}$$

is the *integral* of the multivalued mapping $F : \Omega \rightrightarrows \mathbb{R}^n$ and is denoted by $\int F d\mu$ or $\int F$. We say that $F : \Omega \rightrightarrows \mathbb{R}^n$ is *integrably bounded*, or for emphasis μ-integrably bounded, if there is a μ-integrable $a : \Omega \to \mathbb{R}^n_+$ such that

$$(|v_1|, \ldots, |v_n|) \leq a(x)$$

for all pairs $(x, v) \in (\Omega, \mathbb{R}^n)$ satisfying $v \in F(x)$. Here and elsewhere we interpret vector inequalities as elementwise inequalities. If $a(x)$ in the above inequality is *square-integrable* with respect to the measure μ on the measure space $(\Omega, \mathcal{A}, \mu)$, then the multifunction F is said to be L^2-bounded. When $\Omega = \mathbb{R}^n$, we let $L^2_m(\mathbb{R}^n, \mathcal{A}, \mu)$ denote the Hilbert space of functions mapping $\mathbb{R}^n$ to $\mathbb{R}^m$ with inner product on the measure space $(\mathbb{R}^n, \mathcal{A}, \mu)$ given by

$$\langle f, g \rangle = \int_{\mathbb{R}^n} (f(x), g(x)) \mu(dx), \tag{10}$$

where $(\cdot, \cdot)$ denotes the usual finite-dimensional vector inner product.

The next property is a generalization of [6, Exercise 3.5.14].

Proposition 2.2 (weak compactness of measurable selections). *Let the multifunction $F : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ be closed, convex-valued, and L^2-bounded on $L^2_m(\mathbb{R}^n, \mathcal{M}^n, \nu_n)$, where $\mathcal{M}^n$ is the Lebesgue field on $\mathbb{R}^n$ and ν_n is the n-dimensional Lebesgue measure. Then the set of measurable selections $\mathcal{S}(F)$ is a weakly compact, convex set in $L^2_m(\mathbb{R}^n, \mathcal{M}^n, \nu_n)$.*

Proof. This set is clearly convex since F is pointwise convex-valued. Thus, by [8, Theorem 1, p. 58] we need only show that $\mathcal{S}(F)$ is weakly sequentially compact. Consider any sequence $\{f_i\} \subset \mathcal{S}(F)$. We must show that $\{f_i\}$ has a weakly convergent subsequence with limit $f_* \in \mathcal{S}(F)$. Since the sequence is L^2-bounded, reflexivity, separability, and Alaoglu's theorem [23, Exercise 18(b), p. 269] imply that there exists a weakly convergent subsequence whose limit belongs to the weak closure of $\mathcal{S}(F)$. Since $\mathcal{S}(F)$ is convex, the strong and weak closures of $\mathcal{S}(F)$ coincide. Hence the result follows if $\mathcal{S}(F)$ is strongly closed. Since strong convergence implies the existence of a subsequence that is almost everywhere pointwise convergent [23, Theorem 3.12], and $F(x)$ is pointwise closed, we have that $\mathcal{S}(F)$ is strongly closed. ∎

2.3. Application to wavefront reconstruction: Projection algorithms. We now characterize the projections associated with the problem of phase retrieval in terms of the corresponding pointwise projections. This allows us to describe a general algorithmic framework that includes many of the currently used phase retrieval algorithms. Let $b \in L^2[\mathbb{R}^2, \mathbb{R}]$ with $b(x) \geq 0$ a.e., let the pointwise projection $b(x)\Pi_{\mathbb{S}}$ be defined by (9), and let $\mathbb{Q}(b)$ be defined by (7). For $u, v \in L^2[\mathbb{R}^2, \mathbb{R}^2]$, it is shown in [16, Theorem 4.2] that the projection $\Pi_{\mathbb{Q}(b)} : L^2[\mathbb{R}^2, \mathbb{R}^2] \rightrightarrows L^2[\mathbb{R}^2, \mathbb{R}^2]$ onto $\mathbb{Q}(b)$ is characterized as the collection of measurable selections from the pointwise projection mapping (9):

$$\Pi_{\mathbb{Q}(b)}(u) = \mathcal{S}\left(b(\cdot)\Pi_{\mathbb{S}}(u(\cdot))\right) \quad \text{and} \quad \text{dist}\,(u; \mathbb{Q}(b)) = \|\, |u| - b \,\|. \tag{11}$$

One can characterize the projection onto the sets $\mathbb{Q}_m$ defined in (2) in a similar fashion. The $\mathcal{F}_m$-transform of $\Pi_{\mathbb{Q}(b)}(u)$ is the $\mathcal{F}_m$-transform of all $v \in \Pi_{\mathbb{Q}(b)}(u)$ and is

Downloaded 02/24/16 to 128.95.104.109. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

written $\mathcal{F}_m\left(\Pi_{\mathbb{Q}(b)}(u)\right)$. For each of the unitary operators $\mathcal{F}_m$ and all $u \in L^2[\mathbb{R}^2, \mathbb{R}^2]$, we know from [16, Corollary 4.3] that

$$\Pi_{\mathbb{Q}_m}(u) = \mathcal{F}_m^*\left(\Pi_{\mathbb{Q}(\psi_m)}(\mathcal{F}_m(u))\right) \quad \text{and} \quad \operatorname{dist}(u; \mathbb{Q}_m) = \| \, |\mathcal{F}_m(u)| - \psi_m \| . \tag{12}$$

A general framework for projection algorithms can be found in [3], which considers sequences of weighted relaxed projections of the form

$$u^{(\nu+1)} \in \left(\sum_{m=0}^{M} \gamma_m^{(\nu)} \left[(1 - \alpha_m^{(\nu)})\mathcal{I} + \alpha_m^{(\nu)} \Pi_{\mathbb{Q}_m} \right] \right) (u^{(\nu)}). \tag{13}$$

Here $\mathcal{I}$ is the identity mapping, $\alpha_m^{(\nu)}$ is a relaxation parameter usually in the interval $[0, 2]$, and the weights $\gamma_m^{(\nu)}$ are nonnegative scalars summing to one. General results for these types of algorithms apply only to convex sets. In the convex setting the inclusion in (13) is an equality since projections onto convex sets are single-valued. In the nonconvex setting this is not the case.

It is shown in [16] that the Gerchberg–Saxton algorithm [10] and its variants can be viewed as an instance of (13). As in [16] we use the change of variables $\lambda^{(\nu)}\beta_m^{(\nu)} = \gamma_m^{(\nu)}\alpha_m^{(\nu)}$ to rewrite (13) as

$$u^{(\nu+1)} \in \left(\mathcal{I} - \lambda^{(\nu)}\mathcal{G}^{(\nu)}\right)\left(u^{(\nu)}\right), \tag{14}$$

where for all ν the operators $\mathcal{G}^{(\nu)} : L^2 \to L^2$ are given by

$$\mathcal{G}^{(\nu)} := \sum_{m=0}^{M} \mathcal{G}_m^{(\nu)} \qquad \text{with} \qquad \mathcal{G}_m^{(\nu)} := \beta_m^{(\nu)} \left(\mathcal{I} - \Pi_{\mathbb{Q}_m}\right). \tag{15}$$

In (14) the nonnegative weights $\beta_m^{(\nu)}$ do not necessarily sum to 1, and the parameters $\lambda^{(\nu)}$ are to be interpreted as *step lengths*. This formulation of the projection algorithm is shown in the next section to correspond to a steepest descent algorithm for a weighted squared distance function.

3. Nonsmooth analysis.

Convergence results for projection methods applied to the phase retrieval problem are not possible in general due to the nonconvexity of the constraint sets. The nonconvexity of the constraint sets is associated with the nonsmoothness of the square of the set distance error $\operatorname{dist}(u; \mathbb{Q}_m)$ defined in (5). This is fundamentally different from the convex setting in a Hilbert space where the squared distance function is smooth.

3.1. Least squares.

In general the optimal value of the weighted squared set distance error $J(u)$ defined by (4) is nonzero. Classical techniques for solving the problem numerically are based on satisfying a first-order necessary condition for optimality. For smooth functions this condition simply states that the gradient takes the value zero at any local solution to the optimization problem. However, the functions $\operatorname{dist}^2(u; \mathbb{Q}_m)$ are not differentiable. The easiest way to see this is to consider the one-dimensional function $a(x) = |\,|x| - b|^2$, where $b > 0$. This function is not differentiable at $x = 0$. (Indeed, it is not even subdifferentiably regular at $x = 0$—see (19)). It is precisely at these points that the finite-dimensional projection operator $\Pi_{b\mathbb{S}}$ is multivalued. Similarly, $\operatorname{dist}^2(u; \mathbb{Q}_m)$ is not differentiable at functions u for which there exists a set $\Omega \subset \operatorname{supp}(\psi_m)$ of positive measure on which u vanishes.

Downloaded 02/24/16 to 128.95.104.109. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

In the nonsmooth setting the usual first-order necessary condition for optimality is replaced by a first-order variational principle of the form $0 \in \partial J(u_*)$, where ∂ denotes a *subdifferential* operator such as those studied in [5, 6, 12, 13, 15, 18]. In this paper, the phrase *the subdifferential* refers to the nonconvex subdifferential introduced by Kruger and Mordukhovich [15]. This subdifferential is precisely described in Definition 3.12, and its calculus is extensively developed in [18]. The main result of this paper is the characterization of the subdifferential of the distance functions $\operatorname{dist}^2(\cdot;\mathbb{Q}_m)$ and the objective function J (equation (4)). We do this by following the pattern of proof used by Clarke, Ledyaev, Stern, and Wolenski in [6, Theorem 3.5.18]. A consequence of this approach is that we also establish the subdifferential regularity of the functions $\operatorname{dist}^2(\cdot;\mathbb{Q}_m)$ and J. This in turn implies that for these functions the Clarke subdifferential [5, 6] and the nonconvex subdifferential [15] are equivalent. The statement of the main result now follows.

THEOREM 3.1 (projections and subdifferentials). *Let $\psi_m : \mathbb{R}^2 \to \mathbb{R}_+$ belong to $\mathbb{U}$ where the set $\mathbb{U}$ is defined in* (6)*, and let $\Pi_{\mathbb{Q}_m} : L^2 \rightrightarrows \mathbb{Q}_m$ be defined by* (8)*. Then the functions $\operatorname{dist}^2(\cdot;\mathbb{Q}_m)$ and J are everywhere subdifferentially regular and for $u \in L^2[\mathbb{R}^2,\mathbb{R}^2]$ we have*

$$\partial\left(\operatorname{dist}^2(u;\mathbb{Q}_m)\right) = 2\mathrm{cl}^*\left(\mathcal{I} - \Pi_{\mathbb{Q}_m}(u)\right) \tag{16}$$

and

$$\partial J(u) = \sum_{m=0}^{M} \mathrm{cl}^*\left(\mathcal{G}_m(u)\right), \tag{17}$$

where $\mathcal{G}_m$ is defined by (15)*, J is defined in* (4)*, and $\mathrm{cl}^*(\cdot)$ denotes the weak-star closure.*

Note that in a Hilbert-space setting $\mathrm{cl}^*(\cdot) = w\text{–}\mathrm{cl}(\cdot)$, where $w\text{–}\mathrm{cl}(\cdot)$ denotes the weak closure. The proof is given at the end of this section. In passing, we note that in the convex case Theorem 3.1 is an elementary consequence of a much more general result for convex functions given in [19, Theorem 20]. For further results along these lines we refer the reader to [5, Proposition 2.5.4] and [21, Example 8.53].

3.2. Finite-dimensional nonsmooth analysis. In [16, Theorem 4.2] it is shown that the squared set distance error $\operatorname{dist}^2(u;\mathbb{Q}(b))$ defined in (7) is given as the integral of the *pointwise* distance function defined by (11). In Theorem 3.1 we extend this correspondence to the subdifferentials of the associated infinite- and finite-dimensional functions. We begin this analysis by introducing the necessary tools from finite-dimensional variational analysis.

Recall that

$$\operatorname{dist}^2(u;\mathbb{Q}(b)) \;=\; \int_{\mathbb{R}^2} r^2(u(x);b(x))dx \;=\; \|u\|^2 + \|b\|^2 + 2h(u;b),$$

where the pointwise residual $r : \mathbb{R}^2 \times \mathbb{R}_+ \to \mathbb{R}$ and the mapping $h : L^2[\mathbb{R}^2,\mathbb{R}^2] \to \mathbb{R}$ are given by

$$r(u(x);b(x)) = |u(x)| - b(x) \qquad \text{and} \qquad h(u;b) := \int_{\mathbb{R}^2} -|u(x)|\, b(x)dx, \tag{18}$$

respectively. While $\operatorname{dist}^2(u;\mathbb{Q}(b))$ is not smooth, it is straightforward to show that it is Lipschitz continuous on bounded subsets of $L^2[\mathbb{R}^2,\mathbb{R}^2]$.

Downloaded 02/24/16 to 128.95.104.109. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

A function $f : \mathbb{X} \to \mathbb{R}$ is *locally Lipschitz near* x if there exists a constant $K \geq 0$ and a neighborhood $\mathbb{V}(x) \subset \mathbb{X}$ of x such that

$$|f(z) - f(y)| \leq K\|z - y\| \qquad \forall\, z, y \in \mathbb{V}(x).$$

For any set $\mathbb{V} \subset X$ over which f is finite-valued, f is said to be *locally Lipschitz on* $\mathbb{V}$ if it is locally Lipschitz at every $x \in \mathbb{V}$. The function is said to be *(globally) Lipschitz on* $\mathbb{V}$ if

$$|f(x) - f(y)| \leq K\|x - y\| \qquad \forall\, x, y \in \mathbb{V}.$$

PROPOSITION 3.2 (Lipschitz constants). *If* $b \in L^2[\mathbb{R}^2, \mathbb{R}]$ *with* $b(x) \geq 0$ *a.e., then the mapping* $\operatorname{dist}^2(\cdot; \mathbb{Q}(b)) : L^2[\mathbb{R}^2, \mathbb{R}^2] \to \mathbb{R}_+$ *is finite-valued and Lipschitz on any bounded subset* $\mathbb{V} \subset L^2[\mathbb{R}^2, \mathbb{R}^2]$ *with Lipschitz constant*

$$K = K_{\|\cdot\|^2} + K_{2h(\cdot;b)},$$

where $K_{\|\cdot\|^2} = 2\sup_{u \in \mathbb{V}} \|u\|$ *is a Lipschitz constant for* $\|u\|^2$ *on* $\mathbb{V}$ *and* $K_{2h(\cdot;b)} = 2\|b\|$ *is a Lipschitz constant for* $h(\cdot; b)$*, independent of* $\mathbb{V}$*.*

Proof. This follows from the proof of [16, Lemma B.2]. □

Lipschitz continuity of the squared set distance error J is a straightforward consequence of Proposition 3.2 and the fact the mappings $\mathcal{F}_m$ are unitary.

We now introduce some basic definitions from nonsmooth analysis. In our discussion we allow mappings to have infinite values; thus it is convenient to define the extended reals $\overline{\mathbb{R}}$, where $\overline{\mathbb{R}} = \mathbb{R} \cup \{+\infty\}$. The *effective domain* of $f : \mathbb{R}^n \to \overline{\mathbb{R}}$, denoted $\operatorname{dom} f \subset \mathbb{R}^n$, is the set on which f is finite. To avoid certain pathological mappings the discussion is restricted to *proper*, i.e., not everywhere infinite, *lower semicontinuous* (l.s.c.) functions.

DEFINITION 3.3 (subderivatives [21]). *For a Lipschitz function* $f : \mathbb{R}^m \to \overline{\mathbb{R}}$ *and a point* $u_* \in \mathbb{R}^m$ *with* $f(u_*)$ *finite,*

(i) *the* subderivative *function* $df(u_*) : \mathbb{R}^m \to \overline{\mathbb{R}}$ *is defined by*

$$df(u_*)(w) := \liminf_{\tau \searrow 0} \frac{f(u_* + \tau w) - f(u_*)}{\tau};$$

(ii) *the* regular subderivative *function (or the Clarke generalized directional derivative when* f *is Lipschitz)* $\widehat{d}f(u_*) : \mathbb{R}^m \to \overline{\mathbb{R}}$ *is defined by*

$$\widehat{d}f(u_*)(w) := \limsup_{u \to u_*,\, \tau \searrow 0} \frac{f(u + \tau w) - f(u)}{\tau}.$$

DEFINITION 3.4 (subgradients: finite-dimensions [21]). *Consider a function* $f : \mathbb{R}^m \to \overline{\mathbb{R}}$*, a point* $v \in \mathbb{R}^m$*, and a point* $u_* \in \mathbb{R}^m$ *with* $f(u_*)$ *finite.*

(i) v *is a* regular subgradient *of* f *at* u_* *if*

$$\liminf_{\substack{u \to u_* \\ u \neq u_*}} \frac{f(u) - f(u_*) - \langle v,\ u - u_* \rangle}{|u - u_*|} \geq 0.$$

We call the set of regular subgradients v *the* regular subdifferential *of* f *at* u_* *and denote this set by* $\widehat{\partial} f(u_*)$*.*

Downloaded 02/24/16 to 128.95.104.109. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

(ii) *v is a* (general) subgradient *of f at u_* if there are sequences $u^{(\nu)} \to u_*$ and $v^{(\nu)} \in \widehat{\partial} f(u^{(\nu)})$ with $f(u^{(\nu)}) \to f(u_*)$ and $v^{(\nu)} \to v$. We call the set of (general) subgradients v the* (general) subdifferential *of f at u_* and denote this set by $\partial f(u_*)$.*

(iii) *v is a* Clarke subgradient *of f at u_* if f is l.s.c. on a neighborhood of u_* and v satisfies*

$$\langle v,\ w\rangle \le \widehat{d} f(u_*)(w)\ \forall w \in \mathbb{R}^m.$$

We call the set of Clarke subgradients v the Clarke subdifferential *of f at u_* and denote this set by $\overline{\partial} f(u_*)$.*

(iv) *A Lipschitz function $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ is said to be* (subdifferentially) regular *at $u_* \in dom\, f$ with $\partial f(u_*) \neq \emptyset$ if*

$$\partial f(u_*) = \widehat{\partial} f(u_*). \tag{19}$$

Remark 3.5 (subdifferentials with closed graphs). From the definitions it can be shown that if $f : \mathbb{R}^n \to \mathbb{R}$ is continuous, then the subgradients ∂f and $\widehat{\partial} f$ are closed with $\widehat{\partial} f$ convex and $\widehat{\partial} f \subset \partial f$. Moreover, the mapping ∂f is outer semicontinuous [21, Definition 5.4]. Therefore, by [21, Theorem 5.7] the graph of ∂f is closed.

Remark 3.6 (subdifferentials of compositions). If $g : \mathbb{X} \to \overline{\mathbb{R}}$ is given as the composition of two functions $f : \mathbb{Y} \to \overline{\mathbb{R}}$ and $h : \mathbb{X} \to \mathbb{Y}$, i.e., $g(x) = (f \circ h)(x) = f(h(x))$, then we write $\partial g(x) = \partial(f \circ h)(x)$. On the other hand, we write $\partial f(h(x))$ to denote the subdifferential of f evaluated at $h(x)$.

The subdifferential definitions are illustrated with the following important example.

Example 3.7 (subdifferential of the modulus). Let $b \in (0, \infty)$. Since the function $b|u|$ is convex it is subdifferentially regular for all u, and

$$\partial\left(b|u|\right) = b\partial\left(|u|\right) = \begin{cases} b\frac{u}{|u|} & \text{if } u \neq 0, \\ b\mathbb{B} & \text{if } u = 0, \end{cases}$$

where $b\mathbb{B}$ is the ball of radius b: $\mathbb{B} = \{u\ :\ |u| \le 1\}$.

In contrast, the function $-b|u|$ for $b \in (0, \infty)$ is not regular at 0. Nevertheless for all u

$$\partial\left(-b|u|\right) = b\partial\left(-|u|\right) = \begin{cases} -b\frac{u}{|u|} & \text{if } u \neq 0, \\ b\mathbb{S} & \text{if } u = 0, \end{cases}$$

where $b\mathbb{S}$ is the sphere of radius b: $\mathbb{S} = \{u\ :\ |u| = 1\}$. The Clarke subdifferential of $-b|u|$ is the convex hull, denoted conv$\,(\cdot\,)$, of the generalized subdifferential:

$$\overline{\partial}\left(-b|u|\right) \;=\; \text{conv}\ \partial\left(-b|u|\right) \;=\; -\partial(b|u|\,).$$

Proof. The first part of the statement is a trivial modification of [21, Exercise 8.27]. The last statement follows from [21, Theorem 8.49]. □

This example yields the following correspondence between finite-dimensional projections $\Pi_{b\mathbb{S}}$ and the subdifferential $\partial(-b|u|\,)$.

PROPOSITION 3.8 (pointwise projections and subdifferentials). *Let $\Pi_{b\mathbb{S}}(u)$ be the projection defined in* (9). *For $u \in \mathbb{R}^2$, $b \in \mathbb{R}_+$, and $r^2 : \mathbb{R}^2 \to \mathbb{R}_+$ defined in* (18) *we have*

$$\partial(-b|u|\,) = -\Pi_{b\mathbb{S}}(u), \quad \overline{\partial}(-b|u|\,) = -\,\text{conv}\,(\Pi_{b\mathbb{S}}(u)), \quad \textit{and} \quad \partial r^2(u;b) = 2(I - \Pi_{b\mathbb{S}}(u)),$$

Downloaded 02/24/16 to 128.95.104.109. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

where I is the finite-dimensional identity operator. Moreover,

$$\overline{\partial} r^2(u;b) = \operatorname{conv}\,[2(I - \Pi_{b\mathbb{S}}(u))].$$

As with the finite-dimensional projection $\Pi_{b\mathbb{S}}$ and the infinite-dimensional projection $\Pi_{\mathbb{Q}(b)} : L^2[\mathbb{R}^2, \mathbb{R}^2] \rightrightarrows L^2[\mathbb{R}^2, \mathbb{R}^2]$ defined in (11), there is a relationship between the finite-dimensional Clarke subdifferential $\overline{\partial} r^2(u(x); b(x))$ (x fixed) and the "subdifferential" of the square distance function, $\partial(\operatorname{dist}^2(u; \mathbb{Q}(b)))$. In infinite-dimensional spaces there are several possible definitions for the subdifferential depending on the underlying geometry and topology of the space. Fortunately, in the separable Hilbert-space setting of phase retrieval many of these definitions coincide [18, Theorem 9.2]. Thus we can choose the characterization that is most convenient. The following development parallels that of Clarke, Ledyaev, Stern, and Wolenski in [6, chapter 3, section 5]. We begin by recalling the definitions and theorems necessary for the analysis.

3.3. Integrals of multivalued functions. We now develop some properties of integrals of multivalued mappings. The next theorem, due to Hildenbrand [11], is a restatement of Theorems 3 and 4 of Aumann [1] for multifunctions on the nonatomic measure space $(\Omega, \mathcal{A}, \mu)$. These results are central to the theory of integrals of multivalued functions.

Theorem 3.9 (integrals of multifunctions [11, Theorem 4 and Proposition 7]). *The following properties hold for integrably bounded multifunctions $F : \Omega \rightrightarrows \mathbb{R}^n$ on nonatomic measure spaces $(\Omega, \mathcal{A}, \mu)$:*

(i) *if F is $\mathcal{A} \otimes \mathcal{B}^n$-measurable, then $\int F = \int \operatorname{conv} F$;*

(ii) *if F is closed (not necessarily $\mathcal{A} \otimes \mathcal{B}^n$-measurable), then $\int F$ is compact.*

The following result is instrumental in the proof of our main result. It is a generalization of [6, Exercise 3.5.17].

Proposition 3.10 (weak closure of nonconvex multivalued integrands). *Let v be chosen from the set of selections $\mathcal{S}(\operatorname{conv} F)$, where $F : \mathbb{R}^2 \rightrightarrows \mathbb{R}^2$ is a nonempty, closed, $\mathcal{M}^2 \otimes \mathcal{B}^2$-measurable, L^2-bounded multifunction on $L_2^2(\mathbb{R}^2, \mathcal{M}^2, P)$ for the probability measure $P(dx) = b(x)dx$ defined by the density $b : \mathbb{R}^2 \to \mathbb{R}_+$. Then there exists a sequence $\{f_i\}$ of measurable selections of F which converges weakly to v. Consequently,*

$$\mathcal{S}(\operatorname{conv} F) \subset \operatorname{cl}^* (\mathcal{S}(F)). \tag{20}$$

Proof. Consider the box $\mathbb{I}_n = [-n, n] \times [-n, n]$ for $n = 1, 2, 3, \ldots$. Suppose each box $\mathbb{I}_n$ is partitioned into $(2n^2)^2$ pixels of width $1/n$. Set

$$t_k^n = \frac{k}{n} - n \quad \text{for } k = 0, 1, \ldots, 2n^2,$$

and for each $t \in [-n, n]$ define

$$\underline{(t)}_n = \max\{t_k^n : t_k^n \le t, k = 0, \ldots, 2n^2\} \quad \text{and} \quad \overline{(t)}_n = \min\{t_k^n : t_k^n \ge t, k = 0, \ldots, 2n^2\}.$$

Note that $0 < \max\{t - \underline{(t)}_n, \overline{(t)}_n - t\} \le 1/n$ whenever $t \in [-n, n]$. By Theorem 3.9 there exists a selection $f_n \in F$ on $(\mathbb{R}^2, \mathcal{M}^2, P)$ corresponding to the partition of the box $\mathbb{I}_n$ such that

$$\int_{\mathbb{R}^2} f_n(x) b(x) dx = \int_{\mathbb{R}^2} v(x) b(x) dx$$

with

$$\int_{t_j^n}^{t_{j+1}^n}\int_{t_k^n}^{t_{k+1}^n} f_n(x)b(x)dx = \int_{t_j^n}^{t_{j+1}^n}\int_{t_k^n}^{t_{k+1}^n} v(x)b(x)dx, \quad n=1,2,3,\dots,\ j,k=0,\dots,2n^2.$$

We show that the sequence f_n converges weakly to v. Let $g \in C^\infty[\mathbb{R}^2,\mathbb{R}^2]$ and $\mathcal{X}_{\mathbb{M}}$ be the indicator of the box $\mathbb{M} = [\alpha,\beta]\times[\gamma,\eta]$. Given $\epsilon > 0$ we will show that there exists n' such that $|\langle g\mathcal{X}_{\mathbb{M}},\ f_n - v\rangle| \le \epsilon$ for all $n \ge n'$, i.e., $\langle g\mathcal{X}_{\mathbb{M}},\ f_n - v\rangle \to 0$.

Let n_1 be such that $\mathbb{M} \subset \mathbb{I}_{n_1}$ for all $n \ge n_1$. Choose $n \ge n_1$. Integration by parts yields

$$\langle g\mathcal{X}_{\mathbb{M}},\ f_n - v\rangle = \left(g(\beta,\eta),\ \int_\gamma^\eta \int_\alpha^\beta [f_n(s,t)-v(s,t)]b(s,t)ds\,dt \right) \tag{21}$$
$$- \int_\gamma^\eta \left(g_y(\beta,y),\ \int_\gamma^y \int_\alpha^\beta [f_n(s,t)-v(s,t)]b(s,t)ds\,dt \right) dy \tag{22}$$
$$- \int_\alpha^\beta \left(g_x(x,\eta),\ \int_\gamma^\eta \int_\alpha^x [f_n(s,t)-v(s,t)]b(s,t)ds\,dt \right) dx \tag{23}$$
$$+ \int_\gamma^\eta \int_\alpha^\beta \left(g_{xy}(x,y), \int_\gamma^y \int_\alpha^x [f_n(s,t)-v(s,t)]b(s,t)ds\,dt \right) dx\,dy. \tag{24}$$

Note that each of these terms contains an expression of the form

$$\begin{aligned}\int_{\hat\gamma}^{\hat\eta}\int_{\hat\alpha}^{\hat\beta}(f_n(s,t)-v(s,t))b(s,t)ds\,dt = &\int_{\underline{(\hat\eta)}_n}^{\hat\eta}\int_{\hat\alpha}^{\hat\beta}(f_n(s,t)-v(s,t))b(s,t)ds\,dt\\ &+\int_{\hat\gamma}^{\overline{(\hat\gamma)}_n}\int_{\hat\alpha}^{\hat\beta}(f_n(s,t)-v(s,t))b(s,t)ds\,dt\\ &+\int_{\overline{(\hat\gamma)}_n}^{\underline{(\hat\eta)}_n}\int_{\hat\alpha}^{\overline{(\hat\alpha)}_n}(f_n(s,t)-v(s,t))b(s,t)ds\,dt\\ &+\int_{\overline{(\hat\gamma)}_n}^{\underline{(\hat\eta)}_n}\int_{\underline{(\hat\beta)}_n}^{\hat\beta}(f_n(s,t)-v(s,t))b(s,t)ds\,dt,\end{aligned} \tag{25}$$

where $[\hat\gamma,\hat\eta]\times[\hat\alpha,\hat\beta] \subset [\gamma,\eta]\times[\alpha,\beta] \subset [-n,n]\times[-n,n]$. Let $a \in L_2^2(\mathbb{R}^2,\mathcal{M}^2,P)$ be an L^2-bound for conv F. For any box of the form $[\alpha',\beta']\times[\gamma',\eta']$, we have the bound

$$\begin{aligned}\left|\int_{\gamma'}^{\eta'}\int_{\alpha'}^{\beta'}(f_n(s,t)-v(s,t))b(s,t)ds\,dt\right| &\le \int_{\gamma'}^{\eta'}\int_{\alpha'}^{\beta'}|f_n(s,t)-v(s,t)|b(s,t)ds\,dt\\ &\le \int_{\gamma'}^{\eta'}\int_{\alpha'}^{\beta'} 2|a(s,t)|b(s,t)ds\,dt\\ &= 2\int_{\mathbb{R}^2}|a(x)|\mathcal{X}_{[\alpha',\beta']\times[\gamma',\eta']}(x)b(x)dx\\ &\le 2\|a\|\int_{\mathbb{R}^2}\mathcal{X}_{[\alpha',\beta']\times[\gamma',\eta']}(x)b(x)dx\\ &= 2\|a\|\int_{[\alpha',\beta']\times[\gamma',\eta']} b(x)dx.\end{aligned}$$

Downloaded 02/24/16 to 128.95.104.109. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

Downloaded 02/24/16 to 128.95.104.109. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

Next note that the Lebesgue measure of each of the sets $[\underline{(\hat\eta)}_n, \hat\eta] \times [\hat\alpha, \hat\beta]$, $[\hat\gamma, \overline{(\hat\gamma)}_n] \times [\hat\alpha, \hat\beta]$, $[\overline{(\hat\gamma)}_n, \hat\eta] \times [\hat\alpha, \overline{(\hat\alpha)}_n]$, and $[\overline{(\hat\gamma)}_n, \underline{(\hat\eta)}_n] \times [\underline{(\hat\beta)}_n, \hat\beta]$ appearing in (25) is bounded by

$$\frac{1}{n}\max\{(\eta-\gamma), (\beta-\alpha)\},$$

which can be made arbitrarily small. By [23, Exercise 12, p. 33], for every $\bar\epsilon > 0$ there is an $\delta(\bar\epsilon) > 0$ such that

$$\int_{\mathbb{E}} b(x)dx \le \bar\epsilon \quad \text{whenever } \mathcal{M}(\mathbb{E}) \le \delta(\bar\epsilon),$$

where $\mathcal{M}(\mathbb{E})$ is the Lebesgue measure of the set $\mathbb{E}$. Therefore, given $\bar\epsilon > 0$, we can choose n so that $\frac{1}{n}\max\{(\eta-\gamma), (\beta-\alpha)\} < \delta(\bar\epsilon)$. By combining this with (25), we obtain the bound

$$\left|\int_{\hat\gamma}^{\hat\eta}\int_{\hat\alpha}^{\hat\beta} (f_n(s,t) - v(s,t)) b(s,t) ds\, dt\right| \le 8\|a\|\bar\epsilon. \tag{26}$$

If we set

$$\Gamma = \max\left\{|g(s,t)|,\, |g_y(s,t)|,\, |g_x(s,t)|,\, |g_{xy}(s,t)| \,:\, (s,t) \in [\alpha,\beta]\times[\gamma,\eta]\right\},$$

the bound (26) yields the following bound for the sum of the four integrands (21)–(24):

$$|\langle g\mathcal{X}_{\mathbb{M}},\ f_n - v\rangle| \le \Gamma[1 + (\eta-\gamma) + (\beta-\alpha) + (\eta-\gamma)(\beta-\alpha)]\,[8\|a\|\bar\epsilon].$$

Given any $\epsilon > 0$ there exists an $\bar\epsilon > 0$ such that the left-hand side, and so also the right-hand side, of this inequality is less than ϵ; moreover, for this $\bar\epsilon$ there is an n' such that

$$\frac{1}{n}\max\{(\eta-\gamma), (\beta-\alpha)\} < \delta(\bar\epsilon) \qquad \forall\, n \ge n'.$$

Therefore, for all $n \ge n'$ we have $|\langle g\mathcal{X}_{\mathbb{M}},\, f_n - v\rangle| \le \epsilon$, which is what we set out to show. Since functions of the form $g\mathcal{X}_{\mathbb{M}}$, where $g \in C^\infty[\mathbb{R}^2, \mathbb{R}^2]$ and $\mathbb{M} \subset \mathbb{R}^2$ is a box, are dense in $L^2_2(\mathbb{R}^2, \mathcal{M}^2, P)$ we have that the sequence f_n converges weakly to v. ∎

3.4. Application to wavefront reconstruction. We now apply the above results to the weighted negative modulus mapping $-b(\cdot)|u(\cdot)|$.

Proposition 3.11 (integrals of projections and subgradients). *Let $b \in \mathbb{U}$ be a density function for the probability measure $P(dx) = b(x)dx$ on $(\mathbb{R}^2, \mathcal{M}^2)$ and let $u \in L^2[\mathbb{R}^2, \mathbb{R}^2]$. The negative modulus function $-|u(x)|$ has the following properties:*

(i) $\mathcal{S}\left(b(\cdot)\overline{\partial}(-|u(\cdot)|)\right)$ *is a weakly compact, convex set in $L^2_2(\mathbb{R}^2, \mathcal{M}^2, \nu_2)$;*

(ii) $\int \partial(-|u(x)|)b(x)dx = \int -\operatorname{conv}\left(\Pi_{\mathbb{S}}(u(x))\right) b(x)dx$, *and $\int \partial(-|u(x)|)b(x)dx$ is a compact subset of $\mathbb{R}^2$;*

(iii) $\mathcal{S}\left(b(\cdot)\overline{\partial}(-|u(\cdot)|)\right) \subset -\mathrm{cl}^*\left(\Pi_{\mathbb{Q}(b)}(u)\right)$ *for all $u \in L^2[\mathbb{R}^2, \mathbb{R}^2]$, where $\mathbb{Q}(b)$ is defined by* (7) *and $\Pi_{\mathbb{Q}(b)}(u)$ by* (8).

Proof. (i) At each x, $b(x)\overline{\partial}(-|u(x)|)$ is closed and convex-valued. In addition, by Example 3.7 every element of the set $\overline{\partial}(-|u(x)|)$ has magnitude less than or equal to 1 and so the multifunction $b(\cdot)\overline{\partial}(-|u(\cdot)|)$ is L^2-bounded in $(\mathbb{R}^2, \mathcal{M}^2, \nu_2)$. Hence, by Proposition 2.2, the multifunction $\mathcal{S}\left(b(x)\overline{\partial}(-|u(x)|)\right)$ is weakly compact in $L^2_2(\mathbb{R}^2, \mathcal{M}^2, \nu_2)$.

Downloaded 02/24/16 to 128.95.104.109. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

(ii) We wish to apply Theorem 3.9, so we must show that the multifunction F written as the composition of a multifunction with a measurable function

$$F(x) = [\partial(-|\cdot|) \circ u](x) = \partial(-|u(x)|)$$

is P-integrably bounded and $\mathcal{M}^2 \otimes \mathcal{B}^2$-measurable. By Example 3.7, the multifunction $F : \mathbb{R}^2 \rightrightarrows \mathbb{R}^2$ is P-integrably bounded with bound equal to 1. By Remark 3.5 $\partial(-|\cdot|) : \mathbb{R}^2 \rightrightarrows \mathbb{R}^2$ has closed graph and is therefore $\mathcal{M}^2 \otimes \mathcal{B}^2$-measurable. By hypothesis, the function u is a Lebesgue measurable mapping from $(\mathbb{R}^2, \mathcal{M}^2)$ into $(\mathbb{R}^2, \mathcal{M}^2)$. Thus, by [11, Proposition 1.b, p. 59] the composite multifunction F defined above is $\mathcal{M}^2 \otimes \mathcal{B}^2$-measurable. Therefore Theorem 3.9 applies to give the result.

(iii) By Proposition 3.10 every $v(\cdot) \in \mathcal{S}\left(b(\cdot)\overline{\partial}(-|u(\cdot)|)\right)$ is the weak limit of a sequence of functions in $\mathcal{S}\left(b(\cdot)\partial(-|u(\cdot)|)\right)$, since $\operatorname{conv}\left(\partial(-|u(\cdot)|)\right) = \overline{\partial}(-|u(\cdot)|)$ (see Example 3.7). If $v \in \mathcal{S}\left(b(\cdot)\partial(-|u(\cdot)|)\right)$, then by [16, Theorem 4.2] and Proposition 3.8 $-v \in \mathcal{S}(b(\cdot)\Pi_{\mathbb{S}}(u(\cdot)))$. Hence, by (11),

$$\mathcal{S}\left(b(\cdot)\partial(-|u(\cdot)|)\right) \subset -\Pi_{\mathbb{Q}(b)}(u),$$

from which the result follows. ◻

3.5. Infinite-dimensional nonsmooth analysis. The next step is to relate the subdifferential of the integral to the integral of the subdifferential. We begin with a brief review of infinite-dimensional nonsmooth analysis. For a complete discussion see [5, 6, 12, 13, 14, 15, 17, 18] and the references therein. To begin with, let $df(u)$ and $\widehat{d}f(u)$ be defined in exactly the same way that they were defined in the finite-dimensional setting in Definition 3.3.

DEFINITION 3.12 (subgradients: infinite-dimensions). *Let $\mathbb{X}$ be a separable Hilbert space, let $f : \mathbb{X} \to \overline{\mathbb{R}}$ be locally Lipschitz continuous, and let $u_* \in \operatorname{dom} f$.*

(i) *A vector $v \in \mathbb{X}^*$ is a* Dini ϵ-subgradient *of f at u_* if*

$$\langle v, w\rangle \le df(u_*)(w) + \epsilon\|w\| \qquad \forall\, w \in \mathbb{X},$$

where $df(u_*)(w)$ is the infinite-dimensional version of the subderivative defined in Definition 3.3(i). *We call the set of Dini ϵ-subgradients v the* Dini ϵ-subdifferential *of f at u_* and denote this set by $\partial_\epsilon^- f(u_*)$. When $\epsilon = 0$, we write $\partial^- f(u_*)$ instead of $\partial_0^- f(u_*)$. By the definition of the subderivative function Definition* 3.3(i) *and the regular subgradient Definition* 3.4(i) *it can be shown that for Lipschitz f the Dini* 0*-subdifferential is simply the infinite-dimensional version of the regular subgradient, $\partial^- f(u_*) = \widehat{\partial} f(u_*)$.*

(ii) *A vector $v \in \mathbb{X}^*$ is a* subgradient *of f at u_* if there are sequences $\epsilon^{(\nu)} \searrow 0$, $u^{(\nu)} \to u_*$, and $v^{(\nu)} \in \partial_\epsilon^- f(u^{(\nu)})$ with $v^{(\nu)} \overset{w^*}{\to} v$, where $\overset{w^*}{\to}$ denotes weak-star convergence. We call the set of subgradients v the* subdifferential *of f at u_* and denote this set by $\partial f(u_*)$.*

(iii) *We define the Clarke generalized subdifferential, $\overline{\partial} f(u_*)$ of f at u_*, as in the finite-dimensional case, Definition* 3.4(iii).

(iv) *The function f is said to be* subdifferentially regular *at u_* if $\partial f(u_*) \ne \emptyset$ and*

$$\partial f(u_*) = \widehat{\partial} f(u_*).$$

Remark 3.13. This construction of the subdifferential comes from [14] where it is used the to construct the A-subdifferential, or approximate subdifferential. However,

Downloaded 02/24/16 to 128.95.104.109. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

due to the equivalence theorem of Mordukhovich and Shao [18, Theorem 9.2] it can also be used in the separable Hilbert space setting to define the subdifferential given in [15]. From Mordukhovich and Shao [18, Theorem 8.11], we also obtain the relation

$$\overline{\partial} f(u_*) = \mathrm{cl}^* \left(\mathrm{conv}\ \partial f(u_*)\right). \tag{27}$$

In particular, this implies that f is subdifferentiably regular at u_* if and only if

$$\overline{\partial} f(u_*) = \partial f(u_*).$$

In addition, when f is strictly differentiable, then $\partial f(u)$ coincides with the Fréchet derivative. Finally, we note that the sets $\partial f(u)$ are weakly closed.

Until now we have been concerned with the issue of when a subset of $\mathbb{R}^n$ depends measurably on the parameter $x \in \Omega$. It is equally important for us to consider the properties of measurable real-valued functions on $\mathbb{R}^n$. For this we make use of *normal* integrands as defined in [21, Definition 14.27]. A function $f : \Omega \times \mathbb{R}^n \to \overline{\mathbb{R}}$ is called a normal integrand if its epigraphical mapping $\mathrm{epi}\, f(x, \cdot)$, $x \in \Omega$, is closed-valued and measurable. Any autonomous, Lipschitz continuous mapping, i.e., $f(x, u) := g(u)$, where $g : \mathbb{R}^n \to \mathbb{R}$ is Lipschitz, is a normal integrand [21, Example 14.30]. For example, the mapping $|u|$ is a normal integrand. We use normal integrands to prove the measurability of the following important mappings.

LEMMA 3.14 (measurability of exposed faces). *Consider a closed-valued Lebesgue measurable multifunction $F : \mathbb{R}^m \rightrightarrows \mathbb{R}^n$. For $x \in \mathbb{R}^m$ and $w \in \mathbb{R}^n$ define $F_* : \mathbb{R}^m \times \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ by*

$$F_*(x, w) = \mathrm{argmax}\left\{\langle v\,,\, w\rangle \mid v \in F(x)\right\}.$$

Then F_* is closed-valued and Lebesgue measurable.

Remark 3.15. Whenever the set $F_*(x, w)$ is nonempty it is called an exposed face of the convex set $F(x)$ [22, section 18]. It is easily shown that these sets are indeed *faces* of $F(x)$ in the sense of [22, section 18]. Here we have focused on Lebesgue measure, but other σ-finite complete measures are possible.

Proof. Since F is closed-valued and measurable, [21, Example 14.32] implies that the function $f : (\mathbb{R}^m \times \mathbb{R}^n) \times \mathbb{R}^n \to \overline{\mathbb{R}}$ given by

$$f(x, w, v) = \langle v\,,\, -w\rangle + \delta_{F(x)}(v)$$

is a normal integrand. Hence the result follows from [21, Theorem 14.37] since

$$F_*(x, w) = \mathrm{argmin}\, f(x, w, v). \qquad \square$$

We remark that if, in addition, F is compact-valued, then so is F_*.

LEMMA 3.16 (subgradients of normal integrands [21, Theorem 14.56]). *Let $(\Omega, \mathcal{A}, \mu)$ be a complete measure space. For the proper normal integrand $f : \Omega \times \mathbb{R}^n \to \mathbb{R}$, and any $u(x) \in \mathrm{dom}\, f(x, \cdot)$ depending measurably on $x \in \Omega$, the subderivative functions*

$$(x, w) \mapsto \widehat{d} f(x, u(x))(w), \qquad (x, w) \mapsto df(x, u(x))(w)$$

are normal integrands and the subdifferential mappings

$$x \mapsto \widehat{\partial} f(x, u(x)), \qquad x \mapsto \partial f(x, u(x))$$

are closed-valued and measurable.

Downloaded 02/24/16 to 128.95.104.109. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

In the remainder of this section, whenever we speak of measure we will be referring to Lebesgue measure.

LEMMA 3.17 (measurable selections for the regular subderivative). *Let $f : \mathbb{R}^n \to \mathbb{R}$ be locally Lipschitz and let $u : \mathbb{R}^m \to \mathbb{R}^n$ and $w : \mathbb{R}^m \to \mathbb{R}^n$ be measurable mappings. Then the subdifferential mapping $\overline{\partial} f(u(\cdot))$ is measurable and possesses a measurable selection $v : \mathbb{R}^m \to \mathbb{R}^n$ such that*

$$\langle v(x)\,,\, w(x)\rangle = \widehat{d} f(u(x))(w(x)) \qquad \text{a.e. } x \in \mathbb{R}^m. \tag{28}$$

Proof. By [21, Theorem 14.56] the mapping ∂f is measurable. Since $\overline{\partial} f(u)$ is simply the convex hull of $\partial f(u)$ for all $u \in \mathbb{R}^n$, [21, Exercise 14.12] implies that $\overline{\partial} f$ is compact convex-valued and measurable. Hence, by [21, Theorem 14.13], the mapping $\overline{\partial} f(u(\cdot))$ is also compact convex-valued and measurable. It remains to establish the existence of a measurable selection satisfying (28).

By [21, Theorem 8.49], we have $\widehat{d} f(u)(w) = \sup \left\langle \overline{\partial} f(u)\,,\, w\right\rangle$ for all $w \in \mathbb{R}^n$, and we have shown that the mapping $\overline{\partial} f$ is compact convex-valued and measurable. Therefore, by Lemma 3.14, the mapping

$$F_*(u, w) = \text{argmax}\, \left\{ \langle v\,,\, w\rangle \;\middle|\; v \in \overline{\partial} f(u) \right\}$$

is also compact convex-valued and measurable with

$$\text{dom}\,(F_*) = \{(u, w) \mid F_*(u, w) \neq \emptyset\} = \mathbb{R}^n \times \mathbb{R}^n.$$

Again, by [21, Theorem 14.13], the mapping $F_*(u(\cdot), w(\cdot))$ is also compact convex-valued and measurable. The measurable selection theorem Theorem 2.1 now implies the existence of a measurable function $v(\cdot)$ such that $v(x) \in F_*(u(x), w(x))$ a.e., which proves the lemma. ∎

We now have our first general result on the interchange of integration and subdifferentiation.

LEMMA 3.18 (interchange of subdifferentiation and integration. I). *Let $\mathcal{H} = L^2_m(\mathbb{R}^n, \mathcal{M}^n, \nu_n)$ be the Hilbert space of square integrable functions mapping from $\mathbb{R}^n$ to $\mathbb{R}^m$ defined in section 2.2, where $\mathcal{M}^n$ is the σ-field of Lebesgue measurable sets on $\mathbb{R}^n$ and ν_n is Lebesgue measure. For simplicity, we write $dx = \nu_n(dx)$. Let $f : \mathbb{R}^m \to \mathbb{R}$ be globally Lipschitz continuous with Lipschitz constant K, and suppose there exists $\hat{u} \in \mathcal{H}$ such that $f \circ \hat{u}$ is an L^2-bounded function on the space $(\mathbb{R}^n, \mathcal{M}^n, \mu)$ where $\mu = b\nu_n$, where $b : \mathbb{R}^n \to \mathbb{R}_+$ with $b \in L^1 \cap L^2 \cap L^\infty[\mathbb{R}^n, \mathbb{R}]$. Define the integral functional $J : \mathcal{H} \to \overline{\mathbb{R}}$ by*

$$J(u) = \int f(u(x))b(x)dx.$$

Then J is globally Lipschitz with Lipschitz constant $K\|b\|_2$, and for every $u \in \mathcal{H}$ the mapping $f \circ u$ is L^2-bounded and

$$\overline{\partial} J(u) \subset \mathcal{S}(b(\cdot)\overline{\partial} f(u(\cdot))). \tag{29}$$

Proof. Let $u \in \mathcal{H}$. The fact that $f \circ u$ is L^2-bounded follows immediately from the inequality

$$|f(u(x))| \leq |f(\hat{u}(x))| + K|u(x) - \hat{u}(x)|.$$

Downloaded 02/24/16 to 128.95.104.109. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

The global Lipschitz continuity of J is a consequence of the following derivation:

$$\begin{aligned}|J(u)-J(v)| &\leq \int K|u(x)-v(x)|b(x)dx\\ &= K\langle |u-v|,\, b\rangle\\ &\leq K\|b\|_2\|u-v\|_2.\end{aligned}$$

Remark 3.13 tells us that $\overline{\partial} J(u)$ is a weakly compact convex subset of $\mathcal{H}$ for all $u\in\mathcal{H}$. We also have from Proposition 2.2 that the set $\mathcal{S}(b(\cdot)\overline{\partial} f(u(\cdot)))$ is also a weakly compact convex subset of $\mathcal{H}$ for all $u\in\mathcal{H}$. Hence the inclusion (29) follows if it can be shown that

$$\sup\left\{\langle v\,,\,w\rangle \;\middle|\; v\in\mathcal{S}(b(\cdot)\overline{\partial} f(u(\cdot)))\right\}\geq \widehat{d}J(u)(w)$$

for all $w\in\mathcal{H}$.

Let $w\in\mathcal{H}$ and let $\{u_i\}\subset\mathcal{H}$ and $\{\tau_i\}\subset\mathbb{R}_+$ be such that $\{u_i\}$ strongly converges to u and $\tau_i\downarrow 0$ with

$$\widehat{d}J(u)(w)=\lim_{i\to\infty}\frac{J(u_i+\tau_i w)-J(u_i)}{\tau_i}.$$

Then, by Fatou's lemma,

$$\begin{aligned}\widehat{d}J(u)(w) &= \lim_{i\to\infty}\int \frac{f(u_i(x)+\tau_i w(x))-f(u_i(x))}{\tau_i}b(x)dx\\ &\leq \int \limsup_{i\to\infty}\frac{f(u_i(x)+\tau_i w(x))-f(u_i(x))}{\tau_i}b(x)dx\\ &\leq \int \widehat{d}f(u(x))(w(x))b(x)dx.\end{aligned}\tag{30}$$

By Lemma 3.17, the multifunction $\overline{\partial} f(u(\cdot))$ possesses a measurable selection v such that $\widehat{d}f(u(x))(w(x))=\langle v(x)\,,\,w(x)\rangle$ a.e. on $\mathbb{R}^n$. Therefore, by (30) we have

$$\begin{aligned}\widehat{d}J(u)(w) &\leq \int \langle v(x)\,,\,w(x)\rangle\, b(x)dx\\ &\leq \sup\left\{\langle v\,,\,w\rangle \;\middle|\; v\in\mathcal{S}(b(\cdot)\overline{\partial} f(u(\cdot)))\right\},\end{aligned}$$

proving the result. □

3.6. Application to wavefront reconstruction. In the next proposition we establish the connection between the projection $\Pi_{\mathbb{Q}(b)}$ defined by (8) and the subdifferential of $h:\ L^2[\mathbb{R}^2,\mathbb{R}^2]\to\mathbb{R}$ defined by (18), where $b\in\mathbb{U}$ with $\mathbb{U}$ defined in (6). Proposition 3.19 is a special case of a more general result to be proved in the final section (Theorem 4.2). However, here we provide a separate and fundamentally different proof which provides the motivation for the perturbation methods studied in [16].

PROPOSITION 3.19 (projection-subdifferential equivalence). *Let $b\in\mathbb{U}$ and let $\Pi_{\mathbb{Q}(b)}:\ L^2\rightrightarrows\mathbb{Q}(b)$ be as defined in* (8) *with $\mathbb{Q}(b)$ defined by* (7)*, and $h:\ L^2[\mathbb{R}^2,\mathbb{R}^2]\to\overline{\mathbb{R}}$ be as defined by* (18)*. Then for all $u\in L^2[\mathbb{R}^2,\mathbb{R}^2]$*

$$\overline{\partial}\left(h(u;b)\right)=\mathcal{S}\left(b(\cdot)\overline{\partial}\left(-|u(\cdot)|\right)\right)=\mathrm{cl}^*\left(-\Pi_{\mathbb{Q}(b)}(u)\right)=\partial\left(h(u;b)\right).\tag{31}$$

Thus, in particular, $h(\cdot;b)$ is everywhere subdifferentiably regular.

Downloaded 02/24/16 to 128.95.104.109. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

Proof. Note that the equivalences in (31) are scale invariant in the sense that if they are shown to be true for a given function b, then they must be true with b replaced by αb for any choice of $\alpha > 0$ since

$$\alpha\partial(h(u;b)) = \partial(h(u;\alpha b)), \qquad \alpha\mathcal{S}(b(\cdot)\bar{\partial}(-|u(\cdot)|)) = \mathcal{S}(\alpha b(\cdot)\bar{\partial}(-|u(\cdot)|)),$$

and

$$\alpha\,\mathrm{cl}^*\left(-\Pi_{Q(b)}(u)\right) = \mathrm{cl}^*\left(-\Pi_{Q(\alpha b)}(u)\right).$$

Since b is nonnegative and integrable, we may therefore assume with no loss in generality that b is a probability density function for some probability measure $P(dx) = b(x)dx$.

If (31) holds, then the subdifferential regularity of $h(\cdot;b)$ follows immediately from Proposition 3.11(i) and (27). By Lemma 3.18 and part (iii) of Proposition 3.11,

$$\overline{\partial}h(u;b) \subset \mathcal{S}\left(b(\cdot)\overline{\partial}(-|u(\cdot)|)\right) \subset \mathrm{cl}^*\left(-\Pi_{\mathbb{Q}(b)}(u)\right).$$

Since $\partial h(u;b) \subset \overline{\partial}h(u;b)$, the result follows once it is shown that

$$\mathrm{cl}^*\left(-\Pi_{\mathbb{Q}(b)}(u)\right) \subset \partial h(u;b). \tag{32}$$

By Proposition 3.2 the mapping h is globally Lipschitz continuous with Lipschitz constant $K = \|b\|$, and by Remark 3.13 $\partial h(u;b)$ is weakly closed. Therefore, if $-\Pi_{\mathbb{Q}(b)}(u) \subset \partial h(u;b)$, then $\mathrm{cl}^*\left(-\Pi_{\mathbb{Q}(b)}(u)\right) \subset \partial h(u;b)$. We now show that $-\Pi_{\mathbb{Q}(b)}(u) \subset \partial h(u;b)$.

Let $v \in -\Pi_{\mathbb{Q}(b)}(u)$ and for all $\epsilon > 0$ define $\tilde{u}_\epsilon := u\mathcal{X}_{\mathrm{supp}\,(u)} + \epsilon v(1 - \mathcal{X}_{\mathrm{supp}\,(u)})$. Then, by [16, Theorem 4.1],

$$\|u - \tilde{u}_\epsilon\| = \epsilon\|v(1 - \mathcal{X}_{\mathrm{supp}\,(u)})\| \le \epsilon\|b\|,$$

and $|\cdot|$ is differentiable at $\tilde{u}_\epsilon(x)$ for every $x \in \mathrm{supp}\,(b)$ with

$$v(x) = -\nabla|\tilde{u}_\epsilon(x)|b(x) \qquad \forall\,\epsilon > 0.$$

For every $w \in L^2[\mathbb{R}^2, \mathbb{R}^2]$ and $x \in \mathrm{supp}\,(b)$, we have

$$\frac{|\tilde{u}_\epsilon(x) + tw(x)| - |\tilde{u}_\epsilon(x)|}{t} \to (\nabla|\tilde{u}_\epsilon(x)|, w(x)),$$

and, since $|\cdot|$ is Lipschitz with Lipschitz constant 1,

$$\left|\frac{|\tilde{u}_\epsilon(x) + tw(x)| - |\tilde{u}_\epsilon(x)|}{t}\right| \le |w(x)| \qquad \forall\, x \in \mathrm{supp}\,(b).$$

Therefore, by the Lebesgue dominated convergence theorem, the function $h(\cdot;b)$ is Gâteaux differentiable at $\tilde{u}_\epsilon$ with Gâteaux derivative $-\nabla|\tilde{u}_\epsilon|b = v$. Hence, since $|\cdot|$ is Lipschitz continuous the $\liminf$ in Definition 3.3(ii) is attained as a limit yielding $dh(\tilde{u}_\epsilon;b)(w) = \langle v, w\rangle$. Consequently

$$v \in \partial^- h(\tilde{u}_\epsilon;b) \qquad \forall\,\epsilon > 0.$$

Taking the limit as $\epsilon \downarrow 0$, we find that $v \in \partial h(u;b)$. Therefore, $-\Pi_{\mathbb{Q}(b)}(u) \subset \partial h(u;b)$. □

Downloaded 02/24/16 to 128.95.104.109. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

The proof of Theorem 3.1 now follows easily from the calculus of subdifferentials.

Proof of Theorem 3.1. [16, Corollary 4.3] gives the representation

$$\begin{aligned}\operatorname{dist}^2(u,\mathbb{Q}_m) &= \operatorname{dist}^2(\mathcal{F}_m(u),\mathbb{Q}(\psi_m))\\ &= \|\mathcal{F}_m(u)\|^2+\|\psi_m\|^2+2h[\mathcal{F}_m(u);\psi_m].\end{aligned}$$

By applying [18, Theorem 6.7] together with Proposition 3.19 and [16, Corollary 4.3], we obtain

$$\begin{aligned}\partial\operatorname{dist}^2(\mathcal{F}_m(u),\mathbb{Q}(\psi_m)) &= 2\partial\left(\left(\frac{1}{2}\|\cdot\|^2+h(\cdot;\psi_m)\right)\circ\mathcal{F}_m\right)(u)\\ &= 2\mathcal{F}_m^*\left[\mathcal{F}_m(u)+\operatorname{cl}^*\left(-\Pi_{\mathbb{Q}_m}(\mathcal{F}_m(u))\right)\right]\\ &= 2\operatorname{cl}^*\left(\mathcal{I}-\Pi_{\mathbb{Q}_m}(u)\right).\end{aligned}$$

Hence the subdifferential regularity of all the functions involved in conjunction with [18, Theorem 4.1] yields the result. $\square$

4. Concluding remarks. We conclude with a generalization of Theorem 3.19. Theorem 4.2 establishes the equivalence of the infinite-dimensional subdifferential objects in the setting relevant to phase retrieval and establishes their relation to the finite-dimensional Clarke subdifferential. The result, and its proof, closely parallels that given in [6, Theorem 3.5.18].

LEMMA 4.1 (interchange of subdifferentiation and integration. II). *Let the hypotheses of Lemma* 3.18 *hold. Then*

$$\mathcal{S}(b(\cdot)\partial f(u(\cdot)))\subset\partial J(u). \tag{33}$$

Proof. Let $z\in\mathcal{S}(b(\cdot)\partial f(u(\cdot)))$. Since $\partial f(u(\cdot))$ is closed-valued and measurable, there exists $v\in\mathcal{S}(\partial f(u(\cdot)))$ for which $z=bv$. We show that $z\in\partial J(u)$. For this purpose, let C be a countably dense subset of $\operatorname{gph}\widehat{\partial}f$. Observe that

$$\partial f(u)=\left\{\lim_{j\to\infty}v^j\;\middle|\;\{(u^j,\,v^j)\}\subset C,\,u^j\to u\right\}.$$

Let $\{(u^k,\,v^k)\}$ be an enumeration of C. Then for each $x\in\mathbb{R}^n$ and each integer $i\in\{1,2,\dots\}$, define $k_i(x)$ be the first integer k for which

$$|u^k-u(x)|\le\frac{1}{i}\quad\text{and}\quad|v^k-v(x)|\le\frac{1}{i}.$$

For each $i=1,2,\dots$, define $u^i:\mathbb{R}^n\to\mathbb{R}^m$ and $v^i:\mathbb{R}^n\to\mathbb{R}^m$ by

$$u^i(x)=u^{k_i(x)}\quad\text{and}\quad v^i(x)=v^{k_i(x)}.$$

We claim that the functions u^i and v^i are measurable with

$$v^i(x)\in\widehat{\partial}f(u^i(x))\ \text{ a.e.} \tag{34}$$

for $i=1,2,\dots$. Indeed, the range of both u^i and v^i is contained in the set C and so is countable. Moreover, for a given integer k,

$$\begin{aligned}&\left\{x\;\middle|\;(u^i(x),v^i(x))=(u^k,\,v^k)\right\}\\ &=\left[\bigcap_{j=1}^{k-1}\left\{x\;\middle|\;\max\{|u^j-u(x)|,|v^j-v(x)|\}>\frac{1}{i}\right\}\right]\\ &\quad\cap\left\{x\;\middle|\;\max\{|u^k-u(x)|,|v^k-v(x)|\}\le\frac{1}{i}\right\},\end{aligned}$$

where each of the sets on the left-hand side is measurable.

Downloaded 02/24/16 to 128.95.104.109. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

Next observe that for all $w \in \mathcal{H}$, we have from Fatou's lemma that

$$dJ(u^i)(w) = \liminf_{\tau \searrow 0} \frac{J(u^i + \tau w) - J(u^i)}{\tau} \geq \int_{\mathbb{R}^2} df(u^i(x))(w(x))b(x)dx \geq \langle bv^i \, , \, w \rangle \, ,$$

where the last inequality follows from (34). Hence $bv^i \in \widehat{\partial} J(u^i)$ for $i = 1, 2, \ldots$. Finally, since $u^i \to u$ and $v^i \to v$ by construction, we have $bv \in \partial J(u)$. ∎

Theorem 4.2 (interchange of subdifferentiation and integration). *Let the hypotheses of Lemma* 3.18 *hold with $n = m = 2$. Then, for all $u \in \mathcal{H} = L^2[\mathbb{R}^2, \mathbb{R}^2]$,*

$$\partial J(u) = \mathrm{cl}^* \mathcal{S}(b(\cdot\,)\partial f(\cdot\,)) = \mathcal{S}(b(\cdot\,)\overline{\partial} f(\cdot\,)) = \overline{\partial} J(u).$$

In particular, this implies that J is everywhere subdifferentially regular.

Proof. By Proposition 3.10 we have

$$\mathcal{S}(b(\cdot\,)\overline{\partial} f(u(\cdot\,))) \subset \mathrm{cl}^*\, \mathcal{S}(b(\cdot\,)\partial f(u(\cdot\,))).$$

Since the set $\partial J(u)$ is weakly closed, Lemma 4.1 implies that

$$\mathrm{cl}^*\, \mathcal{S}(b(\cdot\,)\partial f(u(\cdot\,))) \subset \partial J(u).$$

Combining these facts with Lemma 3.18 yields

$$\begin{aligned} \overline{\partial} J(u) &\subset \mathcal{S}(b(\cdot\,)\overline{\partial} f(u(\cdot\,))) \\ &\subset \mathrm{cl}^*\, \mathcal{S}(b(\cdot\,)\partial f(u(\cdot\,))) \\ &\subset \partial J(u) \\ &\subset \overline{\partial} J(u), \end{aligned}$$

which proves the result. ∎

The restriction in Theorem 4.2 to the case $n = m = 2$ follows from the use of this hypothesis in Proposition 3.10. However, we believe that it is possible to extend this proposition to the general case, which would allow us to remove the restriction $n = m = 2$ from Theorem 4.2.

REFERENCES

[1] R. J. Aumann, *Integrals of set-valued functions*, J. Math. Anal. Appl., 12 (1965), pp. 1–12.

[2] H. Bauschke and J. Borwein, *Dykstra's alternating projection algorithm for two sets*, J. Approx. Theory, 79 (1994), pp. 418–443.

[3] H. H. Bauschke and J. M. Borwein, *On projection algorithms for solving convex feasibility problems*, SIAM Rev., 38 (1996), pp. 367–426.

[4] H. H. Bauschke, P. L. Combettes, and D. R. Luke, *Phase retrieval, error reduction algorithm, and Fienup variants: A view from convex optimization*, J. Opt. Soc. Amer. A, 19 (2002), pp. 1334–1345.

[5] F. H. Clarke, *Optimization and Nonsmooth Analysis*, Classics Appl. Math. 5, SIAM, Philadelphia, 1990.

[6] F. H. Clarke, Yu. S. Ledyaev, R. J. Stern, and P. R Wolenski, *Nonsmooth Analysis and Control Theory*, Springer-Verlag, New York, 1998.

[7] P. L. Combettes, *Inconsistent signal feasibility problems: Least-squares solutions in a product space*, IEEE Trans. Signal Process., 42 (1994), pp. 2955–2966.

[8] M. M. Day, *Normed Linear Spaces*, 3rd ed., Springer-Verlag, New York, 1973.

[9] J. Fienup, *Phase retrieval algorithms: A comparison.* Appl. Optim., 21 (1982), pp. 2758–2769.

[10] R. Gerchberg and W. Saxton, *A practical algorithm for the determination of phase from image and diffraction plane pictures*, Optik, 35 (1972), pp. 237–246.

[11] W. Hildenbrand, *Core and Equilibria of a Large Economy*, Princeton University Press, Princeton, NJ, 1974.
[12] A. D. Ioffe, *Approximate subdifferentials and applications:* II, Mathematika, 33 (1986), 111–128.
[13] A. D. Ioffe, *Approximate subdifferentials and applications:* III, Mathematika, 36 (1989), pp. 1–38.
[14] A. D. Ioffe, *Proximal analysis and approximate subdifferentials*, J. London Math. Soc., 41 (1990), pp. 175–192.
[15] A. Y. Kruger and B. S. Mordukhovich, *Extremal points and the Euler equation in nonsmooth optimization*, Dokl. Akad. Nauk BSSR, 24 (1980), pp. 684–687.
[16] D. R. Luke, J. V. Burke, and R. G. Lyon, *Optical wavefront reconstruction: Theory and numerical methods*, SIAM Rev., 44 (2002), pp. 169–224.
[17] B. S. Mordukhovich, *Maximum principle in problems of time optimal control with nonsmooth constraints*, J. Appl. Math. Mech., 40 (1976), pp. 960–969.
[18] B. S. Mordukhovich and Y. Shao, *Nonsmooth sequential analysis in Asplund spaces*, Trans. Amer. Math. Soc., 328 (1996), pp. 1235–1280.
[19] R. T. Rockafellar, *Conjugate Duality and Optimization*, SIAM, Philadelphia, 1974.
[20] R. T. Rockafellar, *Integral functionals, normal integrands and measurable selections*, in Nonlinear Operators in the Calculus of Variations, Lecture Notes in Math. 543, Springer-Verlag, New York, 1976, pp. 157–207.
[21] R. T. Rockafellar and R. J. Wets, *Variational Analysis*, Springer-Verlag, New York, 1998.
[22] R. T. Rockafellar, *Convex Analysis*, Princeton University Press, Princeton, NJ, 1970.
[23] W. Rudin, *Real and Complex Analysis*, 2nd ed., McGraw–Hill, San Francisco, CA, 1974.
[24] L. P. Vlasov, *Approximative properties of sets in normed linear spaces*, Russian Math. Surveys, 28 (1973), pp. 1–66.

Downloaded 02/24/16 to 128.95.104.109. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php