Volume 75 Number 4 March 2012 ISSN 0362-546X

ELSEVIER

Nonlinear Analysis
Theory, Methods & Applications
An International Multidisciplinary Journal

Series A: Theory and Methods

Nonlinear Analysis 75 (2012) 2398–2415

Contents lists available at SciVerse ScienceDirect

Nonlinear Analysis

journal homepage: www.elsevier.com/locate/na

Convex regularization of local volatility models from option prices: Convergence analysis and rates

A. De Cezaro [a], O. Scherzer [b,c], J.P. Zubelli [a,*]

[a] IMPA, Estr. D. Castorina 110, Rio de Janeiro RJ 22460-320, Brazil

[b] Computational Science Center, University of Vienna, Nordbergstr. 15, A-1090 Vienna, Austria

[c] Radon Institute of Computational and Applied Mathematics, Austrian Academy of Sciences, Altenbergerstr. 69, A-4040 Linz, Austria

ARTICLE INFO

Article history:
Received 26 April 2011
Accepted 25 October 2011
Communicated by Enzo Mitidieri

MSC:
primary 47A52
65M32
91B28

Keywords:
Local volatility surface identification
Convex regularization
Convergence rates
Source condition interpretation
Convex risk measures

ABSTRACT

We study a convex regularization of the local volatility surface identification problem for the Black–Scholes partial differential equation from prices of European call options. This is a highly nonlinear ill-posed problem which in practice is subject to different noise levels associated to bid–ask spreads and sampling errors. We analyze, in appropriate function spaces, different properties of the parameter-to-solution map that assigns to a given volatility surface the corresponding option prices. Using such properties, we show stability and convergence of the regularized solutions in terms of the Bregman distance with respect to a class of convex regularization functionals when the noise level goes to zero.

We improve convergence rates available in the literature for the volatility identification problem. Furthermore, in the present context, we relate convex regularization with the notion of exponential families in Statistics. Finally, we connect convex regularization functionals with convex risk measures through Fenchel conjugation. We do this by showing that if the source condition for the regularization functional is satisfied, then convex risk measures can be constructed.

1. Introduction

In financial markets a number of contracts are negotiated in such a way that their values are derived from other underlying assets or equities. Such derivative contracts play a fundamental role in risk management and corporate strategies. Their presence became so widespread that currently, the volume of many derivative markets surpasses the value of the corresponding underlying markets.

The development of mathematical methods for pricing derivatives has been a major reason for the expansion of derivative markets. Such theoretical achievement was recognized by the Nobel prize in Economics award to Merton and Scholes. The corresponding methods involve the solution of the Black–Scholes partial differential equation, which in turn depends on the risk-free interest rate prevalent in the market, the dividend rate, and the volatility of the underlying asset. There are many models to describe the volatility. Among those, one that is very popular with practitioners is to assume that such volatilities are functions of the form $\sigma = \sigma(t, S)$, where t is the time and S is the asset price. It is usually referred to as Dupire's local volatility model [1] and σ is called the volatility surface.

This paper is concerned with theoretical aspects of the practical problem of determining the volatility from market observed prices of European call options. This is a nonlinear ill-posed problem whose solution calls for regularization

* Corresponding author. Tel.: +55 21 2529 5102; fax: +55 21 2529 5129.
E-mail address: zubelli@impa.br (J.P. Zubelli).

0362-546X/$ – see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.na.2011.10.037

techniques. We propose Tikhonov regularization by means of a convex regularizing functional as an extension to the quadratic regularization that has been used previously in the inverse problem literature for this specific problem [2–4].

We address the regularization problem from the perspective of convex analysis methods and Bregman distances. On the theoretical side, our result is that this yields better convergence rates and allows for convergence in spaces different from those in the quadratic regularization setting. In fact, in some cases, the convergence of certain convex regularization expressions implies convergence in the L^1-norm. Besides those results, our approach connects with central topics in different areas of current research. Such topics include *exponential families* of probability distributions, which is an important subject in Statistics and *convex risk measures* in Risk Management and Quantitative Finance [5,6].

The connection between Bregman distances and exponential families is well established in some context [7,8], albeit in the present context our motivation in Section 5 is heuristic. From the financial intuition, it can be understood as follows: each volatility surface leads to a corresponding risk neutral measure whose expectation of the payoff are the observed derivative prices. Thus, if we are given the problem of inferring the volatility surface from market observed option prices, the use of Bregman distances leads to the choice of certain exponential families of probability distributions. The latter, can be thought of as optimal (in an appropriate sense) *a posteriori* distributions for the class of models under consideration. Indeed, under some circumstances, exponential families are connected to minimal entropy measures. This hints to yet another connection with the now classical work developed by Avellaneda et al. See [9] and references therein.

The passage of the regularized volatility to the market probability measures allows us to also connect the results to convex risk measures. In fact, in Section 6, we exhibit procedures to produce such risk measures which depend on the regularization functional. This in turn relates to Malliavin calculus results and the determination of the so-called Greeks of option prices [10].

The setting and the inverse problem: We consider a complete financial market, where cash can be borrowed at a constant *interest rate* r, and a risky stock of value $S = S(t)$ that yields a continuously compounded dividend at a constant rate q, satisfying the diffusion price process

$$dS(t) = S(t)(\nu(t, S(t))dt + \sigma(t, S(t))dW(t)), \quad t > 0, \qquad S(0) = S_0, \tag{1}$$

where $W(t)$ denotes the standard Wiener process [11]. The parameters ν and σ are called drift rate and the volatility of the underlying asset, respectively.

A *European call option* with maturity date T and strike K, on the underlying asset S, consists of the right, but not the obligation, to buy, at a price K, a unit of S at time T. In the context of complete and arbitrage-free markets, the theoretical fair price, for the European call on S, has the probabilistic representation

$$U(0, S_0; T, K, r, q, \sigma^2) = \exp(-rT)\mathbb{E}^{0,S_0}_{\mathbb{Q}}(S(T) - K)^+, \tag{2}$$

where $\mathbb{E}^{0,S_0}_{\mathbb{Q}}$ is the expected value with respect to the *risk-neutral* probability measure $\mathbb{Q}$ given that, at $t = 0$, we have $S(0) = S_0$. Here, as usual, we define

$$(S - K)^+ := \max\{S - K, 0\}.$$

The interpretation of Eq. (2) is that for each realization ω of the market, the payoff $(S(T, \omega) - K)^+$ should be brought to its present value $e^{-rT}(S(T, \omega) - K)^+$ by means of discounting by the interest rate r. Then, we average over all the possible realizations with respect to the risk neutral measure $\mathbb{Q}$. The risk neutral measure differs from the so-called subjective one in the sense that it is the one for which the discounted process $S(t)/e^{rt}$ is a martingale. For more details see [12].

In this framework the fair price for an European call option is given by the solution to the Black–Scholes equation [13]

$$U_t + \frac{1}{2}\sigma^2(t, S)S^2U_{SS} + (r - q)SU_S - rU = 0, \quad t < T, \tag{3}$$

with final condition

$$U(t = T, S) = (S - K)^+. \tag{4}$$

An important consequence of the Black–Scholes–Merton theory is that the drift rate ν in Eq. (1) does not enter into (3). Indeed, this is at the root of the concept of the risk-neutral measure $\mathbb{Q}$.

In the case where σ is a deterministic function of time only, explicit formulas for the price U are well known. See the seminal paper [13]. In this context, a careful analysis of the theoretical volatility calibration problem was carried out in [14,15].

We note that the option price U depends also on the maturity T and strike K. It satisfies the, by now classical, *Dupire* forward equation [1]

$$-U_T + \frac{1}{2}\sigma^2(T, K)K^2U_{KK} - (r - q)KU_K - qU = 0, \quad T > 0, \tag{5}$$

with the initial value

$$U(T = 0, K) = (S_0 - K)^+, \quad \text{for } K > 0. \tag{6}$$

Dupire's equation is the starting point of our inverse problem analysis. As usual, the dividend and interest rates are known during the option life. However, the crucial parameter in the initial value problem determined by (5) and (6) is the volatility.

Thus, the nonlinear inverse problem of option pricing under consideration is the identification (or calibration) of a local volatility surface $\sigma(T, K)$ by observations of the solutions

$$U(t, S; T, K, r, q, \sigma) = U_*^{t,S}(T, K) \tag{7}$$

of (5) and (6) to match quoted market prices $U_*^{t,S}(T, K)$. Each observation is linked to the solution of (5) and (6) with different values of T and K.

Organization of the article: In Section 2 we define and review some facts about the inverse problem under consideration as well as the Tikhonov regularization theory with convex regularization functionals. Properties of the forward operator that guarantee the well-posedness and regularization analysis of the proposed Tikhonov functional for the inverse problem under consideration are described at Sections 3 and 4. Section 4.1 is dedicated to the analysis of the source condition assumptions needed to obtain convergence rates. In Section 5 we motivate the general regularization theory with convex penalization by making use of a statistical point of view. In Section 6, we relate the convex penalization on the Tikhonov functional and the respective source condition with convex risk measures. We conclude in Section 7 with some final comments and directions for further investigations.

2. Convex regularization for calibration

We start our analysis by reformulating the inverse problem in more convenient variables. More precisely, we perform the usual change of variables

$$K = S_0 e^y, \qquad \tau = T - t, \qquad b = q - r, \qquad u(\tau, y) = e^{q\tau} U^{t,S}(T, K) \tag{8}$$

and

$$a(\tau, y) = \frac{1}{2}\sigma^2(T - \tau; S_0 e^y), \tag{9}$$

in (5) and (6). This yields the Dupire equation with forward variables (τ, y)

$$-u_\tau + a(\tau, y)(u_{yy} - u_y) + bu_y = 0 \tag{10}$$

and initial condition

$$u(0, y) = S_0(1 - e^y)^+. \tag{11}$$

Existence and uniqueness results for the solution of the parabolic equations (10) and (11) in Sobolev spaces can be found in [2,4,16].

Volatility calibration in extended Black–Scholes models has been investigated by many authors. See [17,9,18,2,4,19,20,15] as some references. The stable identification of local volatility surfaces in the Black–Scholes equation from market prices using standard Tikhonov regularization with $\|\cdot\|^2_{H^1(\Omega)}$ penalization was investigated by Crépey [2] and later by Egger and Engl [4]. In [15] the inverse problem of identification of a time-dependent volatility function of a European call option with a fixed strike $K > 0$ was considered. In [20], Hofmann et al. analyzed the same financial problem of [15] with general source conditions for the regularization functional $f(\cdot) = \|\cdot\|^2_{L_2(0,T)}$. In [2,4,20,15], the ill-posedness of the inverse problem is proved, convergence and convergence rates of a regularized solution are derived.

The idea of convex regularization for inverse problems has been suggested by different authors. An early reference on Bregman distance regularization is [21]. See also [22,20,23] and references therein.

In the initial part of this work, we consider the following admissible class of calibration parameters:

Definition 1. Let $\varepsilon \geq 0$ be fixed. We consider the Hilbert space $H^{1+\varepsilon}(\Omega)$ with the $H^{1+\varepsilon}$-inner product $\langle \cdot, \cdot \rangle$.

Moreover, let $\overline{a} > \underline{a} > 0$ and let a_0 be a function defined on $\Omega = (0, T) \times \mathbb{R}$ that satisfies $\underline{a} \leq a_0 \leq \overline{a}$ with $\nabla a_0 \in (L_2(\Omega))^2$. We define the admissible parameter class by

$$\mathcal{D}(F) := \{a \in a_0 + H^{1+\varepsilon}(\Omega) : \underline{a} \leq a \leq \overline{a}\}. \tag{12}$$

We emphasize that by definition $\mathcal{D}(F)$ is a convex set.

We apply convex regularization as discussed in [22,20,23] to solve the ill-posed operator equation

$$F(a) = u(a), \tag{13}$$

where $F : \mathcal{D}(F) \subset H^{1+\varepsilon}(\Omega) \to L^2(\Omega)$ is the parameter-to-solution operator between Hilbert spaces $H^{1+\varepsilon}(\Omega)$ and $L_2(\Omega)$. Here $u(a)$ is the solution of (10) and (11), where $a \in \mathcal{D}(F)$.

The novelty of the present article *vis-a-vis* [2,4,20,15] is that we consider a regularization method for solving the calibration problem for a general class of convex functionals. For given convex f the proposed methods consists in minimizing the Tikhonov functional

$$\mathcal{F}_{\beta,u^\delta}(a) := \|F(a) - u^\delta\|^2_{L_2(\Omega)} + \beta f(a) \tag{14}$$

over $\mathcal{D}(F)$, where, $\beta > 0$ is the regularization parameter.

In this paper we only make the following assumptions on f:

Assumption 2. Let $\varepsilon \geq 0$ be fixed. $f : \mathcal{D}(f) \subset H^{1+\varepsilon}(\Omega) \longrightarrow [0, \infty]$ is a convex, proper and sequentially weakly lower semi-continuous functional with domain $\mathcal{D}(f)$ containing $\mathcal{D}(F)$.

The use of convex functionals in the present calibration context is motivated by both mathematical and financial reasons. From the mathematical point of view, we are able to extend known results to take into account different features of the surfaces. In particular convergence rate results known in the literature are extended [4,19,14,2]. From the financial side, it is known that practitioners have for a while used different regularization functionals to pin down particular aspects of volatility surfaces that suit their needs. Yet, not many analytical results are known. See [9] and references therein. As a byproduct of the theory, we obtain the following interpretation in Section 6: the existence of a source condition for the regularized inverse problem leads to a convex risk measure. The latter is a way of quantifying risk associated to the different replication portfolios in the market.

In practical situations, the price $U^{t,S}(T,K)$ is only known for a discrete set of maturities and strikes. Since we are interested in continuous observations of the price $U^{t,S}(T,K)$, this leads to an interpolation or an approximation that introduces noisy data u^δ, whose level δ is assumed to be known *a priori* and satisfies the inequality

$$\|\bar{u} - u^\delta\|_{L_2(\Omega)} \leq \delta, \tag{15}$$

where $\bar{u}$ is the data associated to the actual value $\hat{a} \in \mathcal{D}(F)$.

An important tool in the studies of Tikhonov type regularization [22,20,24,23] is the Bregman distance with respect to f.

Definition 3. Let f be as in Assumption 2. For given $a \in \mathcal{D}(f)$, let $\partial f(a) \subset H^{1+\varepsilon}(\Omega)$ denote the sub-differential of the functional f at a, which we define and denote by

$$\mathcal{D}(\partial f) = \{\tilde{a} : \partial f(\tilde{a}) \neq \emptyset\}$$

the domain of the sub-differential [25]. The Bregman distance with respect to $\zeta \in \partial f(a_1)$ is defined on $\mathcal{D}(f) \times \mathcal{D}(\partial f)$ by

$$D_\zeta(a_2, a_1) = f(a_2) - f(a_1) - \langle \zeta, a_2 - a_1 \rangle.$$

Concerning the definition of the sub-differential and the Bregman distance, we emphasize that the sub-differential is a subset of the dual of $H^{1+\varepsilon}(\Omega)$. However, in Hilbert spaces there exists an isomorphism between the space $H^{1+\varepsilon}(\Omega)$ and its dual $(H^{1+\varepsilon}(\Omega))^*$. This justifies Definition 3 where $\partial f(a)$ is considered a subset of $H^{1+\varepsilon}(\Omega)$ and the Bregman distance, which is considered with respect to the $H^{1+\varepsilon}(\Omega)$-inner product.

Notation 4. Throughout this paper we use the following notation: $I \subset \mathbb{R}$ denotes an open (possibly unbounded) interval and $1 \leq p < \infty$. We assume that $T > 0$ and we use the notation $\Omega := (0, T) \times I$. Moreover, $W^{1,2}_p(\Omega)$ denotes the space of functions $u(\cdot, \cdot)$ satisfying

$$\|u\|_{W^{1,2}_p(\Omega)} := \|u\|_{L_p(\Omega)} + \|u_t\|_{L_p(\Omega)} + \|u_y\|_{L_p(\Omega)} + \|u_{yy}\|_{L_p(\Omega)} < \infty.$$

We now summarize the convergence-rate results of regularization methods to the proposed problem available in the literature. In all the examples, presented below, the regularization parameter is chosen by $\beta = \beta(\delta) \sim \delta$.

(i) Egger and Engl [4] applied the standard results for nonlinear Tikhonov regularization in a Hilbert space setting, and obtained convergence rates of

$$\left\|a^\delta_\beta - a^\dagger\right\|_{H^1(\Omega)} = \mathcal{O}(\sqrt{\delta}) \quad \text{and} \quad \left\|F(a^\delta_\beta) - u^\delta\right\|_{L_2(\Omega)} = \mathcal{O}(\delta) \tag{16}$$

to $a^\delta_\beta, a^\dagger \in \mathcal{D}(F) \subset H^1(\Omega)$ under the assumption of the source condition

$$a_0 - a^\dagger = F'(a^\dagger)^* w$$

with $\|w\|$ sufficiently small. Moreover, the above convergence rates are proved for time-independent volatilities in a more regular set and with a variational source condition. See [4, Theorem 4.1].

(ii) Focusing on the time dependent case only, Hofmann and Krämer [15] studied the maximum entropy regularization functional $f(\cdot)$ in the setting of $\mathcal{D}(F) \subset L_1[0,T]$ and data in $L_2[0,T]$. Under the source condition $\log(a^\dagger/\hat{a}) = F'(a^\dagger)^* w$, for some $w \in L_2[0,T]$, the convergence rates of

$$\|a_\beta^\delta - a^\dagger\|_{L_1[0,T]} = \mathcal{O}(\sqrt{\delta}) \tag{17}$$

was proven. In addition the authors had to assume the nonlinear estimate

$$\|F(a) - F(a^\dagger) - F'(a - a^\dagger)\|_{L_2[0,T]} \le C\|a - a^\dagger\|^2_{L_1[0,T]}. \tag{18}$$

We will return to maximum entropy regularization in Section 5 and, more generally, to Bregman distance regularization in Section 4.1.

(iii) Hofmann et al. [20] improved the convergence rates of [15] for the regularization functional $f(\cdot) = \|\cdot\|_{L_2[0,T]}$. Also here, as in [20,15], the volatility parameter is considered to be time-dependent only.

One of the goals of the present work is to generalize the above mentioned convergence rate results for local volatility (volatility that is time and space dependent) estimation by using recent abstract convergence results for Tikhonov regularization [26], in the $H^1(\Omega)$ norm.

3. Properties of the forward operator

Below we summarize some properties of the operator F defined in (13).

Theorem 5. *Assume that $\varepsilon \ge 0$ and consider the operator $F : \mathcal{D}(F) \subset H^{1+\varepsilon}(\Omega) \to L_2(\Omega)$, with $H^{1+\varepsilon}(\Omega)$ as in Definition 1. Then,*

(i) *F is continuous and (sequentially) compact. Moreover, F is sequentially weakly continuous and weakly closed.*

(ii) *F is differentiable at $a \in \mathcal{D}(F)$ in every direction h such that $a + h \in \mathcal{D}(F)$. The derivative $F'(a)$ is extensible to a bounded linear operator on $H^{1+\varepsilon}(\Omega)$. Moreover, $F'(a)$ satisfies the Lipschitz condition*

$$\left\|F'(a) - F'(a+h)\right\|_{\mathcal{L}(H^{1+\varepsilon}(\Omega);L^2(\Omega))} \le c\,\|h\|_{H^{1+\varepsilon}(\Omega)}\,, \tag{19}$$

for $a + h \in \mathcal{D}(F)$.

Proof. The proof is sketched in the Appendix. See also [4], Propositions A.3 and 4.3 respectively. □

The (sequentially) compactness and weak closedness of the operator F, concluded in Theorem 5, imply the local ill-posedness of the inverse problem of identification of the local volatility surface $\sigma(T,K)$. In fact, for every $H^{1+\varepsilon}(\Omega)$-bounded sequence $\{a_n\}_{n\in\mathbb{N}}$ in $\mathcal{D}(F)$, that has no strong convergent subsequences, we can extract an $H^{1+\varepsilon}(\Omega)$-weakly-convergent subsequence, say $\{a_{n_k}\}_{k\in\mathbb{N}}$. Since $\mathcal{D}(F)$ is weakly closed with respect to the $H^{1+\varepsilon}$-norm, the weak limit of $\{a_{n_k}\}_{k\in\mathbb{N}}$ belongs to $\mathcal{D}(F)$. Thus, since F is (sequentially) compact, $\{F(a_{n_k})\}$ has a convergent subsequence. So, similar option prices may correspond to completely different volatilities.

As observed in [4, Remark 4.1], $\mathcal{D}(F)$ has no interior points when equipped with the $H^1(\Omega)$ norm. Because of that, $F'(a)$ is not necessarily differentiable in every direction $h \in H^1(\Omega)$. In other words, $F'(a)$ is not Gateaux differentiable. This will not affect the convergence analysis that follows. In fact, for such analysis we only need that the operator F attains a one-sided directional derivative at $a^\dagger$ in the directions $a - a^\dagger$ for all $a \in \mathcal{D}(F)$. The sufficient condition for this to happen is $\mathcal{D}(F)$ to be starlike with respect to $a^\dagger$. That is, for every $a \in \mathcal{D}(F)$ there exists $t_0 > 0$ such that

$$a^\dagger + t(a - a^\dagger) = ta + (1-t)a^\dagger \in \mathcal{D}(F) \quad \forall 0 \le t \le t_0.$$

Because $\mathcal{D}(F)$ is convex, the requirement above follows. Moreover, the bounded linear operator $F'(a^\dagger)$ has properties that mimic the Gateaux derivative.

In particular, there exists an adjoint operator

$$F'(a^\dagger)^* : L_2(\Omega) \longrightarrow H^{1+\varepsilon}(\Omega)$$

defined by

$$\langle F'(a^\dagger)^* v, a\rangle_{L_2(\Omega)} = \langle v, F'(a^\dagger)a\rangle_{H^{1+\varepsilon}(\Omega)}, \quad a \in H^{1+\varepsilon}(\Omega),\ v \in L_2(\Omega).$$

We emphasize that Theorem 5 holds true if we restrict our attention to

$$\mathcal{D}(F) := \{a \in a_0 + H^{1+\varepsilon}(\Omega) : \underline{a} \le a \le \overline{a}\} \tag{20}$$

and a convex, weakly lower semi-continuous functional f on $H^{1+\varepsilon}(\Omega)$ with $\mathcal{D}(F) \subseteq \mathcal{D}(f)$. Moreover, for $\varepsilon > 0$, by the Sobolev embedding theorem, each function of $\mathcal{D}(F) \subset H^{1+\varepsilon}(\Omega)$ is an interior point, for which Fréchet-differentiability holds, as Theorem 5 shows.

Lemma 6. *Let $\varepsilon > 0$ be fixed. For $a \in \mathcal{D}(F)$ the Frèchet derivative of F exists and is injective.*

Proof. Let $h \in \mathcal{N}(F'(a)) \subset H^{1+\varepsilon}(\Omega)$. Because of Eq. (A.2) we have that

$$h \cdot (u_{yy} - u_y) = 0. \tag{21}$$

However, $G(\tau, y) = (u_{yy} - u_y)$ is the distributional solution of the initial value problem

$$\partial_\tau G(\tau, y) = \frac{1}{2}(\partial_{yy}^2 - \partial_y)(a(t, y)G(\tau, y)) + b\,G(\tau, y) \tag{22}$$
$$G(0, y) = \delta(y),$$

where $\delta(y)$ is the Dirac's delta. In others words, $G(\tau, y)$ is the Green's function of the Cauchy problem (22). Hence, $G(\tau, y) \neq 0$ for every y and $\tau \geq 0$ (See, for example, [27] or [2, Theorem 4.3]). Thus, it follows from (21) that $h = 0$ a.e. □

The above lemma states that for every $a \in \mathcal{D}(F)$ the operator $F'(a)$ has a trivial null-space, and thus the range of $F'(a)^*$ is dense in $H^{1+\varepsilon}(\Omega)$. Interestingly, $(F'(a)^\dagger)^*$ shares the same properties, and consequently the range of $F'(a^\dagger)$ is dense in $L_2(\Omega)$. This property will be used later on to characterize source conditions in the inverse problem theory.

Lemma 7. *Let $\varepsilon > 0$. The operator $F'(a^\dagger)^*$ has a trivial kernel.*

Proof. As before, we take $b = 0$ for simplicity. Denote by

$$\mathcal{L}_u := -\partial_\tau + a(\partial_{yy} - \partial_y)$$

and by $G_{u_{yy}-u_y}$, the parabolic partial differential operator on the left hand side of Eq. (A.2) with homogeneous boundary condition and the multiplication operator by the function $u_{yy} - u_y$, respectively. Hence, the solution of (A.2) has a functional form $u'(a) := F'(a) = \left(\mathcal{L}_u\right)^{-1} G_{u_{yy}-u_y}$, where by $\left(\mathcal{L}_u\right)^{-1}$ we mean the left-inverse of the operator $\mathcal{L}_u$ with vanishing boundary and initial conditions.

Since $F'(a^\dagger)^* : L_2(\Omega) \longrightarrow H^{1+\varepsilon}(\Omega)$, we have

$$\langle F'(a^\dagger)h, z\rangle_{L_2(\Omega)} = \langle h, \varphi\rangle_{H^{1+\varepsilon}(\Omega)}, \quad \forall h \in H^{1+\varepsilon}(\Omega),\ \forall z \in L_2(\Omega)$$

and $F'(a^\dagger)^* z = \varphi$. Now, let $z \in \mathcal{N}(F'(a^\dagger)^*)$. Then,

$$0 = \left\langle F'(a^\dagger)h, z\right\rangle_{L_2(\Omega)} = \left\langle \left(\mathcal{L}_u\right)^{-1} G_{u_{yy}-u_y} h, z\right\rangle_{L_2(\Omega)} = \left\langle G_{u_{yy}-u_y} h, \left(\left(\mathcal{L}_u\right)^{-1}\right)^* z\right\rangle_{L_2(\Omega)}$$
$$= \left\langle G_{u_{yy}-u_y} h, g\right\rangle_{L_2(\Omega)} = \int_\Omega (u_{yy} - u_y) h\, g\, d\tau\, dy \quad \forall h \in H^{1+\varepsilon}(\Omega)$$

where g is a solution of the adjoint equation

$$g_\tau + (a^\dagger g)_{yy} + (a^\dagger g)_y = z,$$

with homogeneous final and boundary conditions. Since $z \in L^2(\Omega)$, $g \in H^{1+\varepsilon}(\Omega)$. See [27]. In particular

$$\int_\Omega (u_{yy} - u_y) h\, g d\tau\, dy = 0,$$

holds true for $h = g$. Since $G_{u_{yy}-u_y} > 0$ (see the end of the proof of Lemma 6) it follows that $g = 0$. Consequently, $z = 0$ and $\mathcal{N}(F'(a^\dagger)^*) = \{0\}$. □

Remark 8. The range of $F'(a^\dagger)^*$ is dense in $H^{1+\varepsilon}(\Omega)$. Indeed,

$$H^{1+\varepsilon}(\Omega) = \overline{\mathcal{R}(F'(a^\dagger)^*)}^{H^{1+\varepsilon}(\Omega)} \oplus \mathcal{N}(F(a^\dagger))$$

and the claim follows from Lemma 6.

4. Stability and convergence of regularized solutions

Given the properties of F on Theorem 5, the general result of [26, Chapter 3] implies the well-posedness, stability and convergence of the minimizers of the Tikhonov functional $\mathcal{F}_{\beta,u^\delta}$. We refer to [26, Chapter 3, Theorems 3.22–3.26], for details of proofs. These results are summarized below with the following abstract assumptions:

Assumption 9.

1. The spaces B and V are Banach spaces endowed with topologies τ_B and τ_V that are weaker than the norm topologies and $\|\cdot\|_V$ is sequentially lower semi-continuous with respect to τ_V.
2. There exists a solution of (13) on $\mathcal{D}(F) \subset B$.

3. The functional $f : \mathcal{D}(f) \subset B \to [0, \infty]$ is convex and sequentially lower semi-continuous with respect to τ_B and $\mathcal{D} := \mathcal{D}(F) \cap \mathcal{D}(f) \neq \emptyset$.
4. $F : \mathcal{D}(F) \subset B \longrightarrow V$ is continuous with respect to τ_B and τ_V.
5. Let $\mathcal{F}_{\beta,\bar{u}}$ the Tikhonov functional defined in (14). Then,

$$\mathcal{M}_\beta(M) := \text{level}_M(\mathcal{F}_{\beta,\bar{u}}) = \{a : \mathcal{F}_{\beta,\bar{u}}(a) \leq M\}$$

is sequentially pre-compact and closed with respect to τ_B. The restrictions of F to $\mathcal{M}_\beta(M)$ are sequentially continuous with respect to the topologies τ_B and τ_V.

The general result of [26, Chapter 3, Theorems 3.22-3.26], then implies well-posedness, stability, convergence. These results are summarized below.

Theorem 10 (*Existence, Stability, Convergence*). *Suppose that $F, f, \mathcal{D}, B$, and V satisfy Assumption 9. Furthermore, assume that $\beta > 0$ and $u^\delta \in V$. Then, we have that:*

- *There exists a minimizer of $\mathcal{F}_{\beta,u^\delta}$. Moreover, there exists an f-minimizing solution of (13).*
- *If (u_k) is a sequence converging to u in V with respect to the norm topology, then every sequence (a_k) with*

$$a_k \in \text{argmin}\left\{\mathcal{F}_{\beta,u_k}(a) : a \in \mathcal{D}\right\}$$

has a subsequence which converges with respect to τ_B. The limit of every τ_B-convergent subsequence $(a_{k'})$ of (a_k) is a minimizer $\tilde{a}$ of $\mathcal{F}_{\beta,u}$, and $\left(f(a_{k'})\right)$ converges to $f(\tilde{a})$.
- *Let $\beta : (0, \infty) \to (0, \infty)$ satisfies $\beta(\delta) \to 0$ and $\delta^2/\beta(\delta) \to 0$, as $\delta \to 0$. Moreover, assume that the sequence (δ_k) converges to 0, and that $u_k := u^{\delta_k}$ satisfies $\|\bar{u} - u_k\| \leq \delta_k$. Set $\beta_k := \beta(\delta_k)$. Then, every sequence (a_k) of elements minimizing $\mathcal{F}_{\beta_k,u_k}$, has a subsequence $(a_{k'})$ that converges with respect to τ_B. The limit $a^\dagger$ of any τ_B convergent subsequence $(a_{k'})$ is an f-minimizing solution of (13), and $f(a_k) \to f(a^\dagger)$. In addition, if the f-minimizing solution $a^\dagger$ is unique, then $a_k \to a^\dagger$ with respect to τ_B.*

Note that, for the special setting of the calibration problem, the spaces $B := H^{1+\varepsilon}(\Omega)$ and $V := L_2(\Omega)$ are Hilbert spaces with their weak topologies, respectively. Moreover, the functional f satisfies Assumption 2. Therefore, the first three conditions of Assumption 9 are satisfied for our particular problem. The last condition of Assumption 9 is a consequence of Theorem 5.

Convergence rate results will be based on the following theorem which requires further assumptions.

Assumption 11. Besides Assumption 9, we assume that

1. There exists an f-minimizing solution $a^\dagger$ of (13), which is an element of the Bregman domain $\mathcal{D}_B(f)$.
2. There exist $\beta_1 \in [0, 1)$, $\beta_2 \geq 0$, and $\zeta^\dagger \in \partial f(a^\dagger)$ such that

$$\langle \zeta^\dagger, a^\dagger - a\rangle \leq \beta_1 D_{\zeta^\dagger}(a, a^\dagger) + \beta_2 \left\|F(a) - F(a^\dagger)\right\|_V, \tag{23}$$

for $a \in \mathcal{M}_{\beta_{\max}}(\rho)$, where $\beta_{\max}, \rho > 0$ satisfy the relation $\rho > \beta_{\max} f(a^\dagger)$.

Under this assumption we have the following:

Theorem 12 (*Convergence Rates [26, Theorem 3.42]*). *Let $F, f, \mathcal{D}, B$, and V satisfy Assumption 11. Moreover, let $\beta : (0, \infty) \to (0, \infty)$ satisfy $\beta(\delta) \sim \delta$. Then*

$$D_{\zeta^\dagger}(a_\beta^\delta, a^\dagger) = O(\delta), \qquad \left\|F(a_\beta^\delta) - u^\delta\right\|_V = O(\delta),$$

and there exists $c > 0$, such that $f(a_\beta^\delta) \leq f(a^\dagger) + \delta/c$ for every δ with $\beta(\delta) \leq \beta_{\max}$.

The following proposition reveals that the technical conditions in Assumption 9 can be obtained from rather classical ones:

Proposition 13 (*[26, Proposition 3.35]*). *Let $F, f, \mathcal{D}, B$, and V satisfy Assumption 9. Assume that there exists an f-minimizing solution $a^\dagger$ of (13), and that F is Gateaux differentiable at $a^\dagger$.*

Moreover, assume that there exist $\gamma \geq 0$ and $\omega^\dagger \in V^*$ with $\gamma \left\|\omega^\dagger\right\| < 1$, such that

$$\zeta^\dagger := F'(a^\dagger)^*\omega^\dagger \in \partial f(a^\dagger) \tag{24}$$

and there exists $\beta_{\max} > 0$ satisfying $\rho > \beta_{\max} f(a^\dagger)$ such that

$$\left\|F(a) - F(a^\dagger) - F'(a^\dagger)(a - a^\dagger)\right\| \leq \gamma\, D_{\zeta^\dagger}(a, a^\dagger), \quad \text{for } a \in \mathcal{M}_{\beta_{\max}}(\rho). \tag{25}$$

Then, Assumption 11 holds.

We emphasize that $B = H^{1+\varepsilon}(\Omega)$ is a Hilbert space and thus we can use the inner product on B and the adjoint operator $F'(a^\dagger)$ instead of the duality pairing of $F'(a^\dagger)$ and the dual adjoint of $F'(a^\dagger)^\#$, respectively, as in [26, Proposition 3.35].

The next section is devoted to verifying the assumptions of the previous results in convergence rates. In particular, it require us to investigate (23), or alternatively (24) and (25), respectively.

4.1. Convergence rates for the calibration inverse problem

It turns out that, for the specific problem under consideration, we are not able to characterize the source condition (24). However, we can guarantee (23) under mildly restrictive conditions. The first step in order to guarantee (23) is the following simple lemma:

Lemma 14. *Let $\zeta^\dagger \in \partial f(a^\dagger)$. Then, there exists a function $w^\dagger \in L_2(\Omega)$ and a function $r \in H^{1+\varepsilon}(\Omega)$ such that*

$$\zeta^\dagger = F'(a^\dagger)^* w^\dagger + r \tag{26}$$

holds. Furthermore, $\|r\|_{H^{1+\varepsilon}(\Omega)}$ can to be taken arbitrarily small.

Proof. Indeed, Lemma 6 implies that $\mathcal{R}(F'(a^\dagger)^*)$ is dense in $H^{1+\varepsilon}(\Omega)$. In particular, from Lemma 7, if $\zeta^\dagger = 0$ we can take $w^\dagger = r = 0$. □

In this subsection we exhibit a class of functionals for which we are able to prove that condition (23) holds provided the variational source condition (26) is satisfied. For that we shall make use of the following concept:

Definition 15. Let $1 \le q < \infty$ and $\tilde{H}^{1+\varepsilon}(\Omega)$ be a subset of $H^{1+\varepsilon}(\Omega)$. The Bregman distance $D_\zeta(\cdot, \tilde{a})$ of $f : H^{1+\varepsilon}(\Omega) \to \mathbb{R} \cup \{+\infty\}$ at $\tilde{a} \in \mathcal{D}_B(f)$ and $\zeta \in \partial f$ is said to be q-coercive with constant $\underline{c} > 0$ if

$$D_\zeta(a, \tilde{a}) \ge \underline{c} \left\| a - \tilde{a} \right\|^q_{\tilde{H}^{1+\varepsilon}(\Omega)}, \quad \forall a \in \mathcal{D}(f). \tag{27}$$

In the next lemma we prove the existence of an approximate source condition as (26) and f satisfying Definition 15 is sufficient for convergence rates:

Lemma 16. *Let $\zeta^\dagger \in \partial f(a^\dagger)$ satisfy (26) with $w^\dagger$ and r such that*

$$\left(C\|w^\dagger\|_{L_2(\Omega)} + \|r\|_{L^2(\Omega)} \right) := \beta_1 \in [0, 1),$$

and the Bregman distance with respect to f be q-coercive with $1 < q < \infty$, with constant $\underline{c} \ge e^{-2}$ and with $\tilde{H}^{1+\varepsilon}(\Omega) := H^{1+\varepsilon}(\Omega)$. Then, Eq. (23) holds. In particular, the convergence rates of Theorem 12 hold.

Proof. Using the Sobolev Embedding Theorem [28], Eq. (26) and the Lipschitz condition of $F'(\cdot)$ of Eq. (19), we have that

$$\begin{aligned} |\langle \zeta^\dagger, a - a^\dagger \rangle| &\le |\langle \zeta^\dagger - r, a - a^\dagger \rangle + \langle r, a - a^\dagger \rangle| \\ &\le \|w^\dagger\|_{L_2(\Omega)} \|F(a) - F(a^\dagger)\|_{L_2(\Omega)} + \left(C\|w^\dagger\|_{L_2(\Omega)} + \|r\|_{H^{1+\varepsilon}(\Omega)} \right) \|a - a^\dagger\|_{H^{1+\varepsilon}(\Omega)}. \end{aligned}$$

Therefore, from here, we obtain, for $p, q > 1$ such that $p^{-1} + q^{-1} = 1$, that

$$|\langle \zeta^\dagger, a - a^\dagger \rangle| \le \|w^\dagger\|_{L_2(\Omega)} \|F(a) - F(a^\dagger)\|_{L_2(\Omega)} + \left(C\|w^\dagger\|_{L_2(\Omega)} + \|r\|_{H^{1+\varepsilon}(\Omega)} \right) \|a - a^\dagger\|^{\frac{1}{p}}_{H^{1+\varepsilon}(\Omega)} \|a - a^\dagger\|^{\frac{1}{q}}_{H^{1+\varepsilon}(\Omega)}, \tag{28}$$

Applying to the second term on the right hand side of Eq. (28) the following variant of Young's inequality

$$ab \le \tilde{\varepsilon} a^{p_1} + \frac{b^{p_2}}{(\tilde{\varepsilon} p_1)^{p_1/p_2} p_2} \quad a, b > 0 \text{ and } \tilde{\varepsilon} > 0,$$

where $a = \|a - a^\dagger\|^{\frac{1}{p}}_{H^{1+\varepsilon}(\Omega)}$, $b = \|a - a^\dagger\|^{\frac{1}{q}}_{H^{1+\varepsilon}(\Omega)}$, $p_1 = q^2/(q-1)$ and $p_2 = q^2$, it follows that,

$$\left(\tilde{\varepsilon} \|a - a^\dagger\|^q_{H^{1+\varepsilon}(\Omega)} + \frac{\|a - a^\dagger\|^q_{H^{1+\varepsilon}(\Omega)}}{\left(\tilde{\varepsilon} \frac{q^2}{q-1} \right)^{1/(q-1)} q^2} \right) \le \left(\tilde{\varepsilon} + \frac{1}{\left(\tilde{\varepsilon} \frac{q^2}{q-1} \right)^{1/(q-1)} q^2} \right) \|a - a^\dagger\|^q_{H^{1+\varepsilon}(\Omega)}.$$

Now, we take $\tilde{\varepsilon} > 0$ such that

$$\tilde{\varepsilon} + \frac{1}{(\tilde{\varepsilon} \frac{q^2}{q-1})^{1/(q-1)} q^2} \le \underline{c}. \tag{29}$$

To check that the Eq. (29) holds true for some $\varepsilon > 0$, it is enough taking $\tilde{\varepsilon} = q - 1 > 0$ and verify that the limit when $q \to 1^+$, in Eq. (29), is e^{-2}.

Therefore, from the assumption that f satisfies Definition 15 and the definition of β_1 we have

$$|\langle \zeta^\dagger, a - a^\dagger \rangle| \le \beta_1 D_{\zeta^\dagger}(a, a^\dagger) + \beta_2 \left\| F(a) - F(a^\dagger) \right\|_{L_2(\Omega)}.$$

The convergence rates now follow from Theorem 12. □

The condition $\left(C\|w^\dagger\|_{L_2(\Omega)} + \|r\|_{L^2(\Omega)}\right) := \beta_1 \in [0, 1)$, is a standard condition that is used frequently in inverse problems. See for example [29]. We note that one does not have an explicit description of $w^\dagger$. Therefore, the size of $w^\dagger$ is not totally controllable for the specific setting of the calibration inverse problem. Hence, in this paper, we assume that $\left(C\|w^\dagger\|_{L_2(\Omega)} + \|r\|_{L^2(\Omega)}\right) := \beta_1 \in [0, 1)$, holds. However in Section 6 we have a financial interpretation of the source condition (26), when $r = 0$.

Under the assumption of Lemma 16, if in addition f is q-coercive, a convergence rate in the norm holds:

$$\left\|a_\beta^\delta - a^\dagger\right\|_{H^{1+\varepsilon}(\Omega)} = \mathcal{O}((\delta)^{\frac{1}{q}}). \tag{30}$$

In the sequel we present possible choices for q-coercive Bregman distance.

Example 17 (*q-Coercive Bregman Distance*)**.** Let $H^{1+\tilde\varepsilon}(\Omega)$ be a Hilbert space and $\mathcal{D}(f) \subset H^{1+\tilde\varepsilon}(\Omega)$ and $f(a) := q^{-1}\left\|a - a^\dagger\right\|^q_{H^{1+\tilde\varepsilon}(\Omega)}$. Then, the Bregman distance associated to f is q-coercive. See [30] and references in there. In particular, if $H^{1+\tilde\varepsilon}(\Omega)$ is a uniformly convex Banach space and continuous embedding in $H^1(\Omega)$, then $f(a) := p^{-1}\left\|a - a^\dagger\right\|^p_{H^{1+\tilde\varepsilon}(\Omega)}$, with $1 < p \le q$ has a q-coercive Bregman distance.

Example 18. Let $1 < q \le 2$ and $\varepsilon > 0$. We consider the functional

$$f(a) = \sum_{n=1}^{\infty} |\langle a, \phi_n\rangle|^q,$$

where $\{\phi_n\}$ is an orthonormal basis in $H^{1+\varepsilon}(\Omega)$. The functional is convex, proper and sequentially weakly lower semi-continuous. Moreover, the Bregman distance of the functional f satisfies

$$f(a) - f(a^\dagger) - \langle\partial f(a^\dagger), a - a^\dagger\rangle \ge C\sum_{n=1}^{\infty} |\langle a - a^\dagger, \phi_n\rangle|^2 = C\|a - a^\dagger\|^2_{H^{1+\varepsilon}(\Omega)}.$$

Hence, f is 2-coercive. Therefore, according to Lemma 16 and Equation (30) the rate of $\mathcal{O}(\sqrt{\delta})$ holds for the $H^{1+\varepsilon}$-norm. This method is usually considered in the case of sparsity regularization [31]. The case $p = 1$, which refers to the original sparsity regularization is not taken into account here, since we aim at convergence rates in the Hilbert space norm.

5. Exponential families

In this section, we will motivate the use of Bregman distances for regularization from a statistical perspective and then connect it to the general theory developed earlier.

"The Darmois–Koopman–Pitman theorem states that under certain regularity conditions on the probability density, a necessary and sufficient condition for the existence of a sufficient statistic of fixed dimension is that the probability density belongs to the exponential family" [32]. We start with the definition of an *exponential family* in dimension 1, which is used later on to define appropriate priors.

Definition 19 (*Regular Exponential Family*)**.** Let $\psi : \mathbb{R} \to \mathbb{R} \cup \{+\infty\}$ be convex and $p_0 : \mathbb{R} \to \mathbb{R}_+$ by continuous. The family of probability distribution functions $p_{\psi,\theta} : \mathbb{R} \to \mathbb{R}_+$ defined by

$$p_{\psi,\theta}(s) := \exp(s \cdot \theta - \psi(\theta))p_0(s)$$

is called a regular exponential family. In this context the function ψ is called *log-partition* or *circulant* function. The *expectation number* $a(\theta)$ is defined by

$$a(\theta) := \int_{\mathbb{R}} s p_{\psi,\theta}(s)\, ds.$$

This definition calls for an example, namely:

Example 20. We consider the exponential family of normal distributions on $\mathbb{R}$ with known variance $\varpi^2 = 1$. The density is

$$p_{\psi,\theta}(s) = \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{(s-\theta)^2}{2}\right), \quad s > 0.$$

This is a one parameter exponential family with

$$p_0(s) = \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{s^2}{2}\right) \quad \text{and} \quad \psi(\theta) = \frac{\theta^2}{2},$$

The expectation number is

$$\begin{aligned} a(\theta) &= \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} s \exp\left(-\frac{(s-\theta)^2}{2}\right) ds \\ &= \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} (s-\theta) \exp\left(-\frac{(s-\theta)^2}{2}\right) ds + \frac{\theta}{\sqrt{2\pi}} \int_{\mathbb{R}} \exp\left(-\frac{s^2}{2}\right) ds \\ &= 0 + \theta. \end{aligned}$$

We have the following result from [8] which relates exponential families with Bregman distances.

Theorem 21 (*Banerjee et al. [8]*). *Let ψ^* denote the Fenchel transform of ψ, which we assume to be differentiable. Then, the Bregman distance with respect to ψ^* is given by*

$$D_{\psi^*}(\hat{a}, \tilde{a}) = \psi^*(\hat{a}) - \psi^*(\tilde{a}) - \psi^{*\prime}(\tilde{a})(\hat{a} - \tilde{a}).$$

If we assume that $a(\theta) \in \operatorname{int}(\operatorname{dom}(\psi^*))$, then

$$p_{\psi,\theta}(a) = \exp\big(-D_{\psi^*}(a, a(\theta))\big) \exp\big(\psi^*(a)\big) p_0(a). \tag{31}$$

We now present some interesting Exponential Families and respective Fenchel conjugate.

Example 22 (*Exponential Families and their Fenchel Conjugates*). For a Gaussian distribution $\psi(\theta) = \frac{\varpi^2}{2}\theta^2$, then $\psi^*(a) = \frac{a^2}{2\varpi^2}$. For Poisson distribution $\psi(\theta) = \exp(\theta)$ we have $\psi^*(a) = a\log(a) - a$.

We shall now motivate Bregman distance regularization as a log-maximum a-posteriori estimator for an exponential family. For the time being and for motivation purposes, we consider a discrete statistical setting. As usual, we consider $(\mathcal{X}, \mathcal{F}, \mathbb{P})$ a probability space. We let $\vec{x} := (x_i)_i$ be a sequence of elements in $\mathcal{X}$ and $\vec{a} = (a_i)_i$, where $a_i = a(x_i) \in \mathbb{R}$. We assume that the conditional probability density for observable data $u_i^\delta := u^\delta(x_i)$ from $u_i := F(a)(x_i)$ are normally and identically distributed with mean zero and variance ϖ^2. That is, the probability of observing u_i^δ given u_i is given by

$$p(u_i^\delta | u_i) = \frac{1}{\varpi\sqrt{2\pi}} \exp\left(-\frac{|u_i^\delta - u_i|^2}{2\varpi^2}\right).$$

Now, for $a \in \mathbb{R}_{>0}$ denote $\theta := \theta(a)$. With this notation, for some prior $\hat{a}$, the *a priori* distribution is defined by

$$p(a) := p_{\psi,\theta}(\hat{a}) = \exp(\hat{a}\theta - \psi(\theta))p_0(\hat{a}).$$

In order to clarify this formula, recall that θ depends on a and this is the only a dependence, which shows up on the right hand side.

This in turn, according to Theorem 21, can be rewritten as

$$p(a) = \exp(-D_{\psi^*}(\hat{a}, a)) \exp(\psi^*(\hat{a}))p_0(\hat{a}).$$

The advantage of this representation is that it does not involve any parametrization of the exponential family (that is, with respect to θ). In this context the Log-maximum estimation then consists in minimizing the functional

$$\vec{a} \longmapsto \sum_i \big(-\log(p(u_i^\delta|u_i)) - \log(p(a_i))\big),$$

which is equivalent to minimizing the functional

$$\vec{a} \longmapsto \sum_i (u_i - u_i^\delta)^2 + \beta \sum_i D_{\psi^*}(\hat{a}_i, a_i),$$

where $\beta = 2\varpi^2$. Note that the Bregman distance is in general not symmetric, and we minimize with respect to the second component of the Bregman distance.

In summary, we have shown that Bregman distance regularization can be considered a log maximum a-posteriori estimator for the expectation number, in our case for the expected variance.

In this model, we shall introduce some regularization techniques. For notational simplicity we formulate them in an infinite dimensional framework. Hereafter, we shall assume again that Ω is a bounded sub-domain of $\mathbb{R}^2$. With this framework, we remark that $\mathcal{D}(F) \subset H^{1+\varepsilon}(\Omega) \cap L^\infty_{>0}(\Omega) \subset L^1(\Omega)$, where $L^\infty_{>0}(\Omega)$ is the set of functions that are (essentially) bounded from below and above by some positive constants.

Example 23. According to Example 22, if we use the exponential family associated to Poisson distributions, we obtain Kullback–Leibler regularization, consisting in minimization of

$$a \longmapsto \mathcal{F}_{\beta,u^\delta}(a) := \left\|F(a) - u^\delta\right\|^2_{L^2(\Omega)} + \beta \mathrm{KL}(\hat{a}, a), \tag{32}$$

where

$$\mathrm{KL}(\hat{a}, a) = \int_\Omega a \log(\hat{a}/a) - (\hat{a} - a)\, dx.$$

We note that the Kullback–Leibler distance is the Bregman distance associated to the Boltzmann–Shannon entropy

$$\mathcal{G}(a) := \int_\Omega a \log(a)\, dx. \tag{33}$$

We also note that the standard Kullback–Leibler regularization [33], and more generally, the Bregman distance regularization, is in general considered with respect to the first component. However, the modeling with exponential families results in Bregman distances with respect to the second component.

Remark 24. The domains of $\mathcal{G}$, $\mathcal{D}(G)$, and of the sub-gradient of $\mathcal{G}$, $\mathcal{D}(\partial\mathcal{G})$, are $L^\infty_{\geq 0}(\Omega)$ (the set of bounded non-negative functions) and $L^\infty_{>0}(\Omega)$, respectively.

The Kullback–Leibler distance, which is the Bregman distance of the Boltzmann–Shannon entropy, is defined on the Bregman domain $\mathcal{D}_B(\mathcal{G})$, that is a subset of $L^\infty_{>0}$. Moreover, the Kullback–Leibler distance is lower semi-continuous with respect to the L^1-norm [33]. Based on this property we extend the Kullback–Leibler distance, to take value $+\infty$ if either $a \notin \mathcal{D}(G)$ or $b \notin \mathcal{D}_B(\mathcal{G})$.

Note that there are exceptional cases, when the integral

$$\int_\Omega a \log(a/\hat{a}) - (a - \hat{a})\, dx$$

is actually finite, but $\mathrm{KL}(a, \hat{a}) = \infty$. This can be seen by taking for instance $a \in L^1_{>0}(\Omega)$ which is not in $L^\infty(\Omega)$ and $\hat{a} = Ca$, where C is a constant. The reason here, is that a is not an element of the sub-gradient of the Boltzmann–Shannon entropy. This follows directly from the definition of the domains of the convex functionals and sub-gradients.

To prove that minimization of $\mathcal{F}_{\beta,u^\delta}$ in (32) is well-posed we have to first prove that our problem of interest has suitable properties in appropriate topologies.

Lemma 25. *Let Ω be a bounded subset of $\mathbb{R}^2$ with Lipschitz boundary. Moreover, let $a_n \in \mathcal{D}(F)$ with $a_n \rightharpoonup a$ in $L^2(\Omega)$. Then $F(a_n) \rightharpoonup F(a)$ in $L^2(\Omega)$.*

Proof. Since $\mathcal{D}(F)$ is convex and closed, $a \in \mathcal{D}(F)$. Let u^n, u be the respectively $W_2^{1,2}(\Omega)$ solutions of (10) and (11), for $a_n, a \in \mathcal{D}(F)$. By linearity, $v^n := u^n - u$ satisfies

$$-v^n_\tau + a_n(v^n_{yy} - v^n_y) + bv^n_y = -(a - a_n)(u_{yy} - u_y), \tag{34}$$

with homogeneous initial and boundary conditions. Standard parabolic regularity estimates implies that $\|v_n\|_{W_2^{1,2}(\Omega)} \leq C\,\|a - a_n\|_{L^2(\Omega)}\,\|u_{yy} - u_y\|_{L^2(\Omega)}$. As, $\|u_{yy} - u_y\|_{L^2(\Omega)} \leq \tilde{C}$ [2, Proposition 4.4(i)], $a_n \rightharpoonup a$ and from the continuous embedding of $W_2^{1,2}(\Omega)$ in $L^2(\Omega)$, it follows that $v^n \rightharpoonup \tilde{v}$ in $L^2(\Omega)$. Moreover

$$\int_\Omega (-v^n_\tau + a_n(v^n_{yy} - v^n_y) + bv^n_y)\varphi dx = -\int_\Omega (a - a_n)(u_{yy} - u_y)\varphi dx \longrightarrow 0, \quad \text{as } n \to \infty,$$

for any $\varphi \in C_0^\infty(\Omega)$. Since $a_n \in L^\infty(\Omega)$, the weak limit $\tilde{v}$ of v^n satisfies

$$-\tilde{v}_\tau + a(\tilde{v}_{yy} - \tilde{v}_y) + b\tilde{v}_y = 0, \quad \text{in } D'(\Omega), \tag{35}$$

with homogeneous initial and boundary conditions. Using another time the parabolic regularity estimate [27], we have that $\tilde{v} = 0$. □

Lemma 26. *Let the assumptions in Lemma 25 be satisfied.*

1. *Let $a, b \in \mathcal{D}(G)$. Set $0 \cdot (+\infty) = 0$. Then,*

$$\|a - b\|^2_{L^1(\Omega)} \leq \left(\frac{2}{3}\|a\|_{L^1(\Omega)} + \frac{4}{3}\|b\|_{L^1(\Omega)}\right)\mathrm{KL}(a, b). \tag{36}$$

2. *For $a, b \in \mathcal{D}(G)$, we have*

$$\frac{1}{2}\|a - b\|^2_{L^2(\Omega)} \leq \mathrm{KL}(a, b). \tag{37}$$

3. *Let the Kullback–Leibler distance as in Remark 24. For sequences $(a_k)_k$ and $(b_k)_k$ in $L^1(\Omega)$, such that one of them is bounded: If $\mathrm{KL}(a_k, b_k) \to 0$, then $\|a_k - b_k\|_{L^1(\Omega)} \to 0$.*

4. *Let* $0 \neq \hat{a} \in \mathcal{D}_B(\mathcal{G})$*, then the sets*

$$\mathcal{M}_{\beta,u^\delta}(M) := \{a \in \mathcal{D}_B(\mathcal{G}) : \mathcal{F}_{\beta,u^\delta}(a) \leq M\}$$

are $\tau_{\tilde{H}^{1+\varepsilon}(\Omega)}$ *sequentially compact in* $L^2(\Omega)$.

Proof. For the proofs of Item 1, Item 2 and Item 3 see [33] or [34, Lemma 2.2 and Proposition 2.3]. To prove Item 4, we use (37). Let $(a_k)_{k=1}^\infty$ be a sequence in $\mathcal{M}_{\beta,u^\delta}(M)$, then according to (37), it is uniformly bounded in $L^2(\Omega)$. Therefore $a_k \rightharpoonup a$ in $L^2(\Omega)$. Since $L^2(\Omega) \subset L^1(\Omega)$, $a_k \rightharpoonup a$ in $L^1(\Omega)$. Furthermore, according to [33] the *KL* functional satisfies

$$\mathrm{KL}(\hat{a}, \tilde{a}) \leq \liminf \mathrm{KL}(\hat{a}, a^k).$$

Now, assumption on F and the weak lower semi-continuity of the L^2 norm implies that

$$\left\| F(\tilde{a}) - u^\delta \right\|_{L^2(\Omega)}^2 + \beta \mathrm{KL}(\hat{a}, \tilde{a}) \leq M. \quad \square$$

Using standard results on variational regularization (see for instance [26]), we have:

Theorem 27. *There exists a minimizer of* $\mathcal{F}_{\beta,u^\delta}$ *in* (32)*. The minimizers are stable and convergent for* $\beta(\delta) \to 0$ *and* $\delta^2/\beta(\delta) \to 0$*. Stable means that* $\operatorname{argmin} \mathcal{F}_{\beta,u^{\delta_k}} \to \operatorname{argmin} \mathcal{F}_{\beta,u^0}$ *for* $\delta_k \to 0$ *and that* $\operatorname{argmin} \mathcal{F}_{\beta(\delta_k),u^{\delta_k}}$ *converges to a solution of* (13) *with minimal energy.*

A consequence of Lemma 26 Item 3 and the continuous embedding of $L^2(\Omega)$ in $L^1(\Omega)$, is the following: Let δ_k be a sequence converging to zero and $a_k = a_{\beta_k}^{\delta_k}$ the respective minimizers of the Tikhonov functional (14). Then, for $b_k = a^\dagger$ for all $k \in \mathbb{N}$, we have

$$\left\| a_k - a^\dagger \right\|_{L^1(\Omega)} \to 0, \quad \text{as } \delta_k \to 0.$$

Moreover, a consequence of (36) and Theorem 12 is that

$$\left\| a_\beta^\delta - a^\dagger \right\|_{L^1(\Omega)} = \mathcal{O}(\sqrt{\delta}). \tag{38}$$

6. Relation with convex risk measures

A very natural question is how to interpret the source condition given by (24) in financial terms. In order to answer it, in this section, we relate the convex regularization functional f and the theory of coherent (convex) risk measures [5,35–37] by assuming that the source condition in (24) is satisfied. The upshot will be that the existence of a source condition allows us to introduce different convex risk measures. We start this section by reviewing the latter.

In financial practice, a number of ways have been proposed to assess the risk of a given portfolio or investment choice [6]. Perhaps the most well-known is the so-called *value at risk* (VaR). It is defined as follows: For a given portfolio, probability level and time period, the VaR is defined as the threshold value such that the probability of loss on the portfolio over the given time period exceeds this value is the given probability level. A minute's thought indicates that the higher the VaR the higher the risk, and, in principle, the more undesirable such investment would be. It turns out that the VaR has a serious pitfall, namely, it does not encourage diversification. This is related to the fact that it is not in general a convex function of the portfolio choice.

Several authors have developed a theory of desirable properties for risk measures. See [6] and references therein. One of the most popular is the concept of a convex risk measure. It represents a quantitative assessment of the risk involved by the investor's preference on a financial position. Usually a position is described by the resulting discounted net worth at the end of a given period. Thus, it is represented by a random variable in a suitable probability space. More precisely, we denote by $\mathcal{X}$ a convex set of real valued random variables over all possible scenarios. Following [5,35–37] we shall now introduce the definition of convex risk measure and postpone to the next paragraph a brief explanation of its meaning.

Definition 28. A map $\rho : \mathcal{X} \longrightarrow \mathbb{R}$ will be called a convex measure of risk if it satisfies the following conditions:

- Convexity.
- Non-increasing monotonicity, i.e., if the random variable ν_2 is dominated by the random variable ν_1 a.e., then $\rho(\nu_2) \geq \rho(\nu_1)$.
- Translation invariance, i.e., if $m \in \mathbb{R}$ is a deterministic variable in the sense that it takes the value m a.e., then

$$\rho(\nu + m) = \rho(\nu) - m. \tag{39}$$

We now digress to give an intuitive interpretation of the different requirements above. The condition of convexity is related to risk aversion and it is important in diversifying risk. See [6] for details. The translation invariance condition, is natural since adding a deterministic quantity to a portfolio must decrease its risk of that amount. The monotonicity says that if two portfolios ν_1 and ν_2 are such that for almost all events the return of ν_1 is greater than, or equal to, the return of ν_2, then the risk associated to ν_1 is smaller than the corresponding risk associated to ν_2.

In the sequel, we present a connection between such convex risk measures and the interpretation of source condition (24). The main point is that we present a construction that allows us to associate the convex regularization functional f involved in the source condition to a convex risk measure. This circle of ideas is novel, to the best of our knowledge, and deserves careful further investigations.

The first assumption is that Ω is a bounded set. This is the same to assuming that the strikes K are bounded below and above by some positive constants. Moreover, we define the functional $f(a) = +\infty$ if $a \notin \mathcal{D}(F)$. Using the assumption of existence of a source function $w^\dagger \in L_2(\Omega)$ that satisfies (24) and the definition of $\partial f(a^\dagger)$ we have that

$$f(a) - \langle w^\dagger, F'(a^\dagger)a\rangle \geq f(a^\dagger) - \langle w^\dagger, F'(a^\dagger)a^\dagger\rangle, \quad \forall a \in H^{1+\varepsilon}(\Omega) \text{ and } \forall w^\dagger \text{ s.t. } F'(a^\dagger)^* w^\dagger \in \partial f(a^\dagger). \tag{40}$$

Let us set $g(-F'(a^\dagger)a) := \langle w, -F'(a^\dagger)a\rangle$. The existence of $w^\dagger$ satisfying (40) implies that it is the Lagrangian multiplier of

$$\begin{aligned} &L : \mathcal{D}(F) \times L_2(\Omega) \longrightarrow \mathbb{R} \\ &(a, w) \rightarrow f(a) + g(-F'(a^\dagger)a), \end{aligned}$$

i.e., it satisfies

$$L(a^\dagger, w) \leq L(a^\dagger, w^\dagger) \leq L(a, w^\dagger).$$

However, it is not clear whether we have more than one $w^\dagger \in \mathcal{R}(F'(a^\dagger))$ satisfying (40). Indeed, it depends on the choice of f. For example, if f is differentiable on $a^\dagger$, then $\partial f(a^\dagger)$ is a single element. Then, from Lemma 7 it follows that $w^\dagger$ satisfies Eq. (24) and therefore it is unique.

We define a family of separately convex functions (meaning that for a fixed w it is convex in a and vice versa) by

$$\begin{aligned} &L_2(\Omega) \ni w \longmapsto h_w : \mathcal{D}(F) \longrightarrow \mathbb{R} \cup \{+\infty\} \\ &a \longmapsto L(a, w) = f(a) + g(-F'(a^\dagger)a). \end{aligned} \tag{41}$$

Observe that $h_w(a)$ is a family of functions of the variable a depending on the parameter w.

Remark 29. A particular property of $h_{w^\dagger}$ is that

$$h_{w^\dagger}(a) - h_{w^\dagger}(a^\dagger) = L(a, w^\dagger) - L(a^\dagger, w^\dagger) = D_{\zeta^\dagger}(a, a^\dagger).$$

However, this property holds only in the special case when $w^\dagger$ satisfies (40).

Remark 30. Note, that the source condition (24) together with the existence of an f-minimum norm solution for (13) is equivalent to the Karush–Kuhn–Tucker condition in convex optimization [38].

Now, from the theory of Fenchel conjugation [39,40] we obtain a unique Fenchel conjugate function of h_w given by

$$\begin{aligned} &\hat{h}_w^* : L_2(\Omega) \longrightarrow \mathbb{R} \\ &v \longmapsto g^*(v) + f^*(-F'(a^\dagger)^* v). \end{aligned} \tag{42}$$

If it happens that

$$g^*(v) = \begin{cases} 0 & \text{if } v = w \\ +\infty & \text{otherwise,} \end{cases}$$

then we would have difficulties in the above definition of $\hat{h}_w^*$. Hence, we focus on the related function h_w^* defined as

$$\begin{aligned} &h_w^* : \mathbb{X} \subset L_2(\Omega) \longrightarrow \mathbb{R} \\ &v \longmapsto h_w^*(v) := f^*(-F'(a^\dagger)^* v), \end{aligned} \tag{43}$$

where $\mathbb{X} := \{v \in L^2(\Omega) : f^*(-F'(a^\dagger)^* v) \text{ is finite}\}$.

We note that since $\{0\} = \mathcal{N}(F'(a^\dagger)^*)$, then $h_w^*(0) = f^*(0) = 0$.

Lemma 31. *The functional h_w^* satisfies the convexity and monotonicity axioms.*

Proof. The convexity follows directly from the properties of the Fenchel conjugate function [40, Theorem 2.3.1]. To prove the monotonicity: let $v_1, v_2 \in \mathbb{X}$ satisfy $v_1 \geq v_2$. From the definition of the Fenchel conjugate we have $h_w^*(v) = f^*(-F'(a^\dagger)^* v) \geq \langle a, -F'(a^\dagger)^* v\rangle - f(a)$. Positivity of $F'(a^\dagger)a$ (see [2, Theorem 4.2]) implies that

$$\begin{aligned} 0 \leq \langle F'(a^\dagger)a, v_1 - v_2\rangle &= \langle F'(a^\dagger)a, v_1\rangle + f(a) - (\langle F'(a^\dagger)a, v_2\rangle + f(a)) \\ &\leq -h_w^*(v_1) + h_w^*(v_2). \quad \square \end{aligned}$$

In the sequel we give a construction of a convex risk measure ρ in the present context. This will be achieved using the properties of h_w^* and an interesting probabilistic representation of $v \in \mathbb{X}$ coming from Malliavin Calculus [10].

We start by relating our notation with that of [10]. Eq. (10) is associated to the diffusion process $\{y_t : 0 \leq t \leq T\}$ that satisfies the dynamics

$$dy_t = \left(r - q - \frac{\sigma(t, y_t)^2}{2}\right) dt + \sigma(t, y_t)dW_t, \qquad y_{t_0} = y_0, \tag{44}$$

in the risk neutral probability measure $\mathbb{Q}$.

We recall that the process (44) is the diffusion (1) in a logarithmic variables where $\sigma \longmapsto a \in \mathcal{D}(F)$ by (9).

For the sake of simplicity, we assume that the process (44) has no dividend and interest rates, i.e., $b = 0$.

Following [10], denote by $\{Y_t : 0 \leq t \leq T\}$ the first variation process associated to $\{y_t : 0 \leq t \leq T\}$ and defined by the stochastic differential equation

$$dY_t = (\sigma^2(Y_t))' Y_t dt + \sigma'(Y_t)dW_t, \qquad Y_{t_0} = 1.$$

Remark 32. We now identify $\sigma^\dagger \longmapsto \sqrt{2a^\dagger}$ and $\tilde{\sigma} \longmapsto \sqrt{2\tilde{a}}$ given by (9) with $a^\dagger, \tilde{a} \in \mathcal{D}(F)$. Then, for sufficiently small $\varepsilon > 0$, the diffusion coefficient $\sigma^\dagger + \varepsilon\tilde{\sigma}$ satisfies the uniform ellipticity condition

$$\exists \eta > 0 : \zeta^T(\sigma^\dagger + \varepsilon\tilde{\sigma})^T(x)(\sigma^\dagger + \varepsilon\tilde{\sigma})(x)\zeta \geq \eta|\zeta|^2,$$

for all $\zeta \in \mathbb{R}^2$ and for all $x \in \Omega$.

We introduce the auxiliary set

$$\Gamma := \left\{\Theta \in L^2[0, T] \mid \int_0^T \Theta(t)dt = 1\right\},$$

which contains for example the constant function $\Theta(t) = 1/T$.

Our first result is a representation lemma.

Lemma 33. *Let $v \in \mathcal{R}(F'(a^\dagger))$. Then, there exists a random variable $\pi_{a^\dagger}$ such that*

$$v = \mathbb{E}^{y_0}_{\mathbb{Q}}[\Phi(y_t)\pi_{a^\dagger}], \tag{45}$$

where $\mathbb{Q}$ is the risk neutral probability measure.

Proof. Let

$$\tilde{\beta}_\Theta = \Theta(t)(\beta(T) - \beta(0))\chi_{0\leq t\leq T}$$

where $\{\beta(t) : 0 \leq t \leq T\}$ is the process given in [10, Lemma 3.1].

Since $\sigma^\dagger + \varepsilon\tilde{\sigma}$ satisfies the uniform ellipticity condition (see Remark 32) we have from [10, Proposition 3.3] that the Gateaux derivative at $\sigma^\dagger$ in the direction $\tilde{\sigma}$ is given by

$$\mathbb{E}^{y_0}_{\mathbb{Q}}[\Phi(y_t)D^*_t((\sigma^\dagger)^{-1}(y_t)Y_t\tilde{\beta}_\Theta(T))]$$

where $D^*_t((\sigma^\dagger)^{-1}(y_t)Y_t\tilde{\beta}_\Theta(T))$ is the Skorohod integral [41] of the possibly anticipative process

$$\{(\sigma^\dagger)^{-1}(y_t)Y_t\tilde{\beta}_\Theta(T) : 0 \leq t \leq T\},$$

for any $\Theta \in \Gamma$. □

We remark that the linearity of D^*_t with respect to $\tilde{\sigma}$ arises through the process β_t. See Proposition 3.3 of [10].

Lemma 34. *The constants do not belong to $\mathcal{R}(F'(a^\dagger))$.*

Proof. If $1 \in \mathcal{R}(F'(a^\dagger))$, then there exist $h \in \mathcal{D}(F'(a^\dagger))$ such that $F'(a^\dagger)h = 1$. Thus, 1 would satisfy (A.2), i.e.,

$$0 = 1_\tau + a^\dagger(1_{yy} - 1_y) = h(u_{yy} - u_y).$$

Using the same argument in the proof of Lemma 6 we have that $(u_{yy} - u_y)$ cannot vanish in a set of positive measure. Thus $h = 0$ a.e. This is a contradiction with the fact that $F'(a^\dagger)h = 1$ since $F'(a^\dagger)$ is linear. □

At this point, we have two interesting sets of random variables for our convex risk measure construction. Firstly,

$$\mathcal{X} := \{v + m : v = \Phi(y_t) \text{ and } m \in \mathcal{C}\}$$

and secondly,

$$\mathcal{X}_1 := \{\pi_{a^\dagger} + m : \pi_{a^\dagger} = D^*_t((\sigma^\dagger)^{-1}(y_t)Y_t\tilde{\beta}_\Theta(T)) \text{ and } m \in \mathcal{C}\},$$

where $\mathcal{C}$ is the set of all constants.

Remark 35. It follows from Lemma 33 that we have a representation of $\mathbb{X}$ by $\mathcal{X}$ and $\mathcal{X}_1$ given by the weighted expectation $\mathbb{E}_{\mathbb{Q}}^{y_0}[\cdot]$ with weight $D_t^*((\sigma^\dagger)^{-1}(y_t)Y_t\tilde{\beta}_\Theta(T))$ and $\Phi(y_t)$ respectively. We remark that the terminology weight here is used in a loose sense, since indeed $D_t^*((\sigma^\dagger)^{-1}(y_t)Y_t\tilde{\beta}_\Theta(T))$ may take negative values.

The following lemma plays a central part in our analysis below.

Lemma 36. *If* $\nu \equiv 1$, *then*

$$\mathbb{E}_{\mathbb{Q}}^{y_0}[\nu D_t^*((\sigma^\dagger)^{-1}(y_t)Y_t\tilde{\beta}_\Theta(T))] = 0.$$

Proof. This follows directly by the duality between the Skorohod integral and the Malliavin derivative [41], and the fact that $D_t 1 = 0$. □

We are now ready to state the mains results of this section.

Proposition 37 (*First Alternative for a Convex Risk Measure*)**.** *The functional*

$$\rho : \mathcal{X} \longrightarrow \mathbb{R} \qquad \nu \longmapsto \rho(\nu) := h_w^*(\mathbb{E}_{\mathbb{Q}}^{y_0}[\nu \cdot \pi_{a^\dagger}]) - \mathbb{E}_{\mathbb{Q}}^{y_0}[\nu] \tag{46}$$

satisfies the convex risk measure axioms.

Proof. By the linearity of the expectation operator and the properties of the functional h_w^* in Lemma 31, the convexity and monotonicity axioms follows.

In order to prove the translation axiom, we write

$$\tilde{\rho} : \mathcal{X} \longrightarrow \mathbb{R} \qquad \nu \longmapsto \tilde{\rho}(\nu) := h_w^*(\mathbb{E}_{\mathbb{Q}}^{y_0}[(\nu - \mathbb{E}_{\mathbb{Q}}^{y_0}[\nu]) \cdot \pi_{a^\dagger}]) - \mathbb{E}_{\mathbb{Q}}^{y_0}[\nu].$$

Let $\nu + m \in \mathcal{X}$. By the linearity of the expected value

$$\begin{aligned}\tilde{\rho}(\nu + m) &= h_w^*(\mathbb{E}_{\mathbb{Q}}^{y_0}[(\nu + m - \mathbb{E}_{\mathbb{Q}}^{y_0}[\nu + m]) \cdot \pi_{a^\dagger}]) - \mathbb{E}_{\mathbb{Q}}^{y_0}[\nu + m] \\ &= h_w^*(\mathbb{E}_{\mathbb{Q}}^{y_0}[(\nu - \mathbb{E}_{\mathbb{Q}}^{y_0}[\nu]) \cdot \pi_{a^\dagger}]) - \mathbb{E}_{\mathbb{Q}}^{y_0}[\nu] - m = \tilde{\rho}(\nu) - m.\end{aligned}$$

Hence $\tilde{\rho}$ satisfies the translation axiom.

Now we show that $\tilde{\rho} = \rho$. Indeed, by definition, $\mathcal{X} = \mathcal{D}(\tilde{\rho}) = \mathcal{D}(\rho)$. Let us take now $\nu \in \mathcal{X}$. Then, by definition of expectation $\mathbb{E}_{\mathbb{Q}}^{y_0}[\nu] = c$ where c is a constant. It follows from Lemma 36 that

$$\begin{aligned}\tilde{\rho}(\nu) &= h_w^*(\mathbb{E}_{\mathbb{Q}}^{y_0}[(\nu - \mathbb{E}_{\mathbb{Q}}^{y_0}[\nu]) \cdot \pi_{a^\dagger}]) - \mathbb{E}_{\mathbb{Q}}^{y_0}[\nu] \\ &= h_w^*(\mathbb{E}_{\mathbb{Q}}^{y_0}[\nu \cdot \pi_{a^\dagger}] - \mathbb{E}_{\mathbb{Q}}^{y_0}[c \cdot \pi_{a^\dagger}]) - \mathbb{E}_{\mathbb{Q}}^{y_0}[\nu] = \rho(\nu) \quad \text{for all } \nu \in \mathcal{X}.\end{aligned}$$

Thus $\tilde{\rho} = \rho$. □

Proposition 38 (*Second Alternative for a Convex Measure of Risk*)**.** *The functional*

$$\rho_1 : \mathcal{X}_1 \longrightarrow \mathbb{R} \qquad \pi \longmapsto \rho_1(\pi) := h_w^*(\mathbb{E}_{\mathbb{Q}}^{y_0}[\nu \cdot \pi]), \tag{47}$$

satisfies the convex risk measure axioms.

Proof. Using the same argument of Proposition 37, the convexity and monotonicity axioms follow.

In order to prove the translation axiom, we write

$$\tilde{\rho}_1 : \mathcal{X}_1 \longrightarrow \mathbb{R} \qquad \pi \longmapsto \tilde{\rho}_1(\pi) := h_w^*(\mathbb{E}_{\mathbb{Q}}^{y_0}[\nu \cdot (\pi - \mathbb{E}_{\mathbb{Q}}^{y_0}[\pi])]) - \mathbb{E}_{\mathbb{Q}}^{y_0}[\pi].$$

Then, for $\pi + m \in \mathcal{X}_1$, by the linearity of the expectation operator we have that

$$\begin{aligned}\tilde{\rho}_1(\pi + m) &= h_w^*(\mathbb{E}_{\mathbb{Q}}^{y_0}[\nu \cdot (\pi + m - \mathbb{E}_{\mathbb{Q}}^{y_0}[\pi + m])]) - \mathbb{E}_{\mathbb{Q}}^{y_0}[\pi + m] \\ &= h_w^*(\mathbb{E}_{\mathbb{Q}}^{y_0}[\nu \cdot (\pi - \mathbb{E}_{\mathbb{Q}}^{y_0}[\pi])]) - \mathbb{E}_{\mathbb{Q}}^{y_0}[\pi] - m = \tilde{\rho}_1(\pi) - m.\end{aligned}$$

Hence, $\tilde{\rho}_1$ satisfies the translation axiom.

By definition, $\mathcal{X}_1 = \mathcal{D}(\tilde{\rho}_1) = \mathcal{D}(\rho_1)$. Let us take $\pi \in \mathcal{X}_1$. From Lemma 36 we conclude that $\mathbb{E}_{\mathbb{Q}}^{y_0}[\pi] = \mathbb{E}_{\mathbb{Q}}^{y_0}[1 \cdot \pi] = 0$.

Thus, $\tilde{\rho}_1(\pi) = \rho(\pi)$ for all $\pi \in \mathcal{X}_1$. □

We note that the choice of $\sigma^\dagger$ enters in a crucial and nonlinear way in the convex risk measure. Furthermore, the source condition (24) allows us to construct convex risk measures in the spaces of random variables associated to the diffusion process (44).

Example 39 (*Example of a Convex Risk Measure Associated with the Boltzmann–Shannon Entropy*)**.** We now illustrate the construction of the convex risk measure by considering the process (44) under constant volatility with vanishing interest and dividend rates. For this particular case, the representation (45) (or the *vega* in financial terms) is given by the formula (see [10])

$$\begin{aligned}&\mathbb{E}^{y_0}_{\mathbb{Q}}\left[\Phi\left(y\exp\left(\sigma^\dagger W_\tau-\frac{(\sigma^\dagger)^2}{2}\tau\right)\right)\cdot\left(\frac{W_\tau^2}{\sigma^\dagger\tau}-W_\tau-\frac{1}{\sigma^\dagger}\right)\right]\\&\quad=\int_\Omega dz\,d\tau\;p(z,\tau)\Phi\left(y\exp\left(\sigma^\dagger z-\frac{(\sigma^\dagger)^2}{2}\tau\right)\right)\cdot\left(\frac{z^2}{\sigma^\dagger\tau}-z-\frac{1}{\sigma^\dagger}\right),\end{aligned}\tag{48}$$

where $p(z,\tau)=e^{-\frac{z^2}{2\tau}}/\sqrt{2\pi\tau}$ is the Gaussian probability density function.

Let us take $v\in\mathbb{X}$ and compute $F'(a^\dagger)^*v$. By Fubini's Theorem

$$\begin{aligned}\langle F'(a^\dagger)a,v\rangle&=\int_\Omega d\tau'\,dy v(\tau',y)\int_\Omega d\tau\,dzp(z,\tau)\Phi\left(y\exp\left(\sigma^\dagger z-\frac{(\sigma^\dagger)^2}{2}\tau\right)\right)\cdot\left(\frac{z^2}{\sigma^\dagger\tau}-z-\frac{1}{\sigma^\dagger}\right)\\&=\int_\Omega d\tau\,dzp(z,\tau)\left(\frac{z^2}{\sigma^\dagger\tau}-z-\frac{1}{\sigma^\dagger}\right)\int_\Omega d\tau'\,dy v(\tau',y)\Phi\left(y\exp\left(\sigma^\dagger z-\frac{(\sigma^\dagger)^2}{2}\tau\right)\right).\end{aligned}$$

Thus,

$$-F'(a^\dagger)^*v=\left(\frac{z^2}{\sigma^\dagger\tau}-z-\frac{1}{\sigma^\dagger}\right)\langle -v,\Phi(\cdot)\rangle.\tag{49}$$

We now consider the regularization functional f as the Boltzmann–Shannon entropy

$$f(a)=\int_\Omega a\log(a)\,dx,\quad a\in\mathcal{D}(F),$$

whose Fenchel conjugate is given by

$$f^*(\mu)=\int_\Omega e^{\mu-1}\,d\tilde{x}.$$

Since we are in a Gaussian model, applying [7, Lemma 11] and (49) to the definition of ρ with $\nu=\Phi\left(y\exp\left(\sigma^\dagger(z)-(\sigma^\dagger)^2\tau/2\right)\right)$ we get

$$\rho(\nu)=-\log\left(\mathbb{E}^{y_0}_{\mathbb{Q}}\left[\exp\left(\frac{z^2}{\sigma^\dagger\tau}-z-\frac{1}{\sigma^\dagger}\right)\langle -v,\nu\rangle\right]\right)-\mathbb{E}^{y_0}_{\mathbb{Q}}[\nu].\tag{50}$$

7. Conclusions and future directions

In this work, we have established existence and convergence results for a convex Tikhonov regularization of the inverse problem associated to the calibration of the local volatility surface from Black–Scholes prices.

The main novelty is the use of a regularization term that only requires convexity properties and weak lower-semi-continuity. Thus, the present regularization applies to a large class of regularization functionals. In particular, in Section 5 we connect with the statistical viewpoint through the concept of exponential families. This is turn, allows the use of a Kullback–Leibler regularization of the calibration problem.

We establish for Bregman distances better convergence rates than those available in the literature to the calibration problem. This analysis also allows us to obtain convergence of the regularized solution with respect to the noise level in $L^1(\Omega)$ by means of a Kullback–Leibler regularization functional. See Eq. (38). Another advantage of the current approach is the requirement of weaker conditions than those previously required in the literature. Namely, we only require q-coerciveness of (27).

The convergence results also hold true if we measure the misfit at the Tikhonov functional (14) in $W^{1,2}_p(\Omega)$. The intuition behind the use of the $W^{1,2}_p(\Omega)$ norm is that we have continuous dependence of the Tikhonov functional with respect to information not only about the prices but also with respect to the sensitivities u_τ, u_{yy}, and u_y. Those are the so called Greeks. On the other hand, we need more information on the measurement data u^δ.

We prove the validity of an approximate source condition of the form (26) for the regularization problem under consideration. In particular, if the regularization functional is $f(\cdot)=\|\cdot\|^2_{H^{1+\varepsilon}(\Omega)}$, then the source condition (24) coincides with the representation that remained an open problem in [2,4].

A heuristic financial interpretation of the source condition (24) is that we have a restriction that allows us to quantify the risk associated to a given volatility level. By this we mean that upon computing the corresponding Black–Scholes solution as a function of the volatility, we are quantifying how much risk one has in the space of random variables associated to such volatility. This is done with the help of the source condition (24). Indeed, we constructed a functional that, through the Fenchel duality, defines different convex risk measures. The availability of such risk measures permits quantifying the risk associated to random variables and portfolios of the underlying model. We remark that convex risk measures are a sub-class of the coherent risk measures. A natural continuation of the present work would be to explore further such connection to risk measures [35,36].

Another direction of future research would be the numerical implementation of the present results with actual market data. An implementation for the case of the standard quadratic Tikhonov regularization can be found in [4,42].

Acknowledgments

The authors gratefully thank Prof. Dr. Bernd Hofmann (TU Chemnitz) for a stimulating discussion associated to this work.

JPZ was supported by CNPq under grants 302161/2003-1 and 474085/2003-1. Part of this work was conducted during the Special Semester on Finance Analyzed by Stochastic Methods, September 1st, 2008–December 5th, 2008, organized by RICAM, Austrian Academy of Sciences. The work of ADC was supported by the ALFA network PDEINET under contract AML/37-311-97/0666/II-0143-FCD. The work of OS was supported by the Austrian Science Fund (FWF) within the national research networks Industrial Geometry, project 9203-N12.

Appendix. Technical appendix

In this appendix we collect a few technical definitions and proofs. Although they are known in the literature, we feel it would be useful to have them collected in the present section since they are used extensively throughout the article.

We start with the concept of one-sided directional derivative:

Definition 40. Let $F : \mathcal{D}(F) \subset B \to V$ be an operator between B and V.

1. The operator F admits a one-sided directional derivative $F'(a; h) \in V$ at $a \in \mathcal{D}(F)$ in the direction $h \in B$, if $a + th \in \mathcal{D}(F)$ for all $t > 0$ and
$$F'(a; h) = \lim_{t \to 0^+} \frac{F(a + th) - F(a)}{t}. \tag{A.1}$$
2. If $F'(a; h)$ is a bounded linear operator with respect to h, we shall write $F'(a; h) := F'(a)h$.

We conclude with the proof of Theorem 5.

Proof of Theorem 5.

(i) The proof follows from [4, Theorem 2.1] or [2, Proposition 4.4 and 5.1], where it is proven that $F : \mathcal{D}(F) \subset H^{1+\varepsilon}(\Omega) \to W_p^{1,2}(\Omega)$ satisfies the property for all $2 \le p < \bar{p}$ with an appropriate $\bar{p} > 2$ and the continuous embedding of $W_p^{1,2}(\Omega)$ into $L_2(\Omega)$.

(ii) Let $a \in \mathcal{D}(F)$ and the direction $h \in H^{1+\varepsilon}(\Omega)$ be such that $a + h \in \mathcal{D}(F)$. For simplicity of exposition, let us assume that $b = 0$ in (10) and (11). By the linearity of equation (10) the directional derivative $u' \cdot h$ in the direction h satisfies
$$-(u' \cdot h)_\tau + a((u' \cdot h)_{yy} - (u' \cdot h)_y) = -h(u_{yy} - u_y) \tag{A.2}$$
with homogeneous initial conditions. From [4, Proposition A.1] there exists a single solution $u' \cdot h \in W_p^{1,2}(\Omega)$ of (A.2), for all $2 \le p < \bar{p}$ and some $\bar{p} > 2$.
Using regularity estimates to parabolic problems (see for example [27]) we have
$$\|u' \cdot h\|_{W_p^{1,2}(\Omega)} \le c\|h(u_{yy} - u_y)\|_{L_p(\Omega)} \le c\|h\|_{L_{p_2}(\Omega)}\|(u_{yy} - u_y)\|_{L_{p_1}(\Omega)}, \tag{A.3}$$
where $p_1 \in (p, \bar{p})$ and p_2 satisfies $1/p = 1/p_1 + 1/p_2$. Note that, $p_2 = \frac{p_1 p}{p_1 - p}$. From [4, Corollary A.1] it follows that $\|u_{yy} - u_y\|_{L_{p_1}(\Omega)} \le C$ for all $a \in \mathcal{D}(F)$. Moreover, from the Sobolev Embedding Theorem [28, Theorem 4.12, case B, pg 85] it follows that there exists a constant $c > 0$ such that $\|h\|_{L_{p_2}(\Omega)} \le c\|h\|_{H^{1+\varepsilon}(\Omega)}$, for all $h \in H^{1+\varepsilon}(\Omega)$. From (A.3)
$$\|u' \cdot h\|_{W_p^{1,2}(\Omega)} \le C\|h\|_{H^{1+\varepsilon}(\Omega)}. \tag{A.4}$$
Thus, the derivative $u'(a) = F'(a)$ can be extended as a bounded linear operator to $H^{1+\varepsilon}(\Omega)$. The next step is to obtain the Lipschitz condition (19). To do this, denote by $\tilde{u}(\tilde{a})$ the solution of (10) and (11) with a replaced by $\tilde{a} = a + h$ and $h \in H^{1+\varepsilon}(\Omega)$. Setting $v := (F'(\tilde{a}) - F'(a)) \cdot q = (\tilde{u'} - u') \cdot q$ with $q \in H^{1+\varepsilon}(\Omega)$. Then, from the linearity of (10), v is a solution of
$$(v)_\tau + a((v)_y - (v)_{yy}) = q((\tilde{u} - u)_{yy} - (\tilde{u} - u)_y) + (\tilde{a} - a)((\tilde{u'} \cdot q)_{yy} - (\tilde{u'} \cdot q)_y).$$

Using an estimates analogous to (A.4) we find

$$\|v\|_{W_p^{1,2}(\Omega)} \le (\tilde{c}\|q\|_{H^{1+\varepsilon}(\Omega)}\|\tilde{u}-u\|_{W_{\tilde{p}}^{1,2}(\Omega)} + \bar{c}\|\tilde{a}-a\|_{H^{1+\varepsilon}(\Omega)}\|\tilde{u'}\cdot q\|_{W_{\tilde{p}}^{1,2}(\Omega)})$$
$$\le C\|q\|_{H^{1+\varepsilon}(\Omega)}\|\tilde{a}-a\|_{H^{1+\varepsilon}(\Omega)}.$$

Taking the sup over all $q \in H^{1+\varepsilon}(\Omega)$ satisfying $\|q\|_{H^{1+\varepsilon}(\Omega)} \le 1$, on both sides of the above inequalities we have the Lipschitz condition (19). □

References

[1] B. Dupire, Pricing with a smile, Risk 7 (1994) 18–20.
[2] S. Crépey, Calibration of the local volatility in a generalized Black–Scholes model using Tikhonov regularization, SIAM J. Matrix Anal. Appl. 34 (5) (2003) 1183–1206 (electronic).
[3] S. Crépey, Calibration of the local volatility in a trinomial tree using Tikhonov regularization, Inverse Problems 19 (1) (2003) 91–127.
[4] H. Egger, H.W. Engl, Tikhonov regularization applied to the inverse problem of option pricing: convergence analysis and rates, Inverse Problems 21 (3) (2005) 1027–1045.
[5] P. Artzner, F. Delbaen, J.-M. Eber, D. Heath, Coherent measures of risk, Math. Finance 9 (3) (1999) 203–228.
[6] A. Meucci, Risk and Asset Allocation, Springer Finance, Springer-Verlag, Berlin, 2005.
[7] Y. Altun, A. Smola, Unifying divergence minimization and statistical inference via convex duality, in: Learning Theory, in: Lecture Notes in Comput. Sci., vol. 4005, Springer, Berlin, 2006, pp. 139–153.
[8] A. Banerjee, S. Merugu, I. Dhillon, J. Ghosh, Clustering with Bregman divergences, J. Mach. Learn. Res. 6 (2005) 1705–1749.
[9] M. Avellaneda, C. Friedman, R. Holmes, D. Samperi, Calibrating volatility surfaces via relative-entropy minimization, Appl. Math. Finance 4 (1) (1997) 37–64.
[10] E. Fournié, J.-M. Lasry, J. Lebuchoux, P.-L. Lions, N. Touzi, Applications of Malliavin calculus to Monte Carlo methods in finance, Finance Stoch. 3 (4) (1999) 391–412.
[11] I. Karatzas, S.E. Shreve, Methods of Mathematical Finance, in: Applications of Mathematics (New York), vol. 39, Springer-Verlag, New York, 1998.
[12] R. Korn, E. Korn, Option Pricing and Portfolio Optimization, in: Graduate Studies in Mathematics, vol. 31, American Mathematical Society, Providence, RI, 2001, Modern Methods of Financial Mathematics, Translated from the 1999 German original by the authors.
[13] F. Black, M. Scholes, The pricing of options and corporate liabilities, J. Polit. Econ. 81 (1973) 637–659.
[14] T. Hein, B. Hofmann, On the nature of ill-posedness of an inverse problem arising in option pricing, Inverse Problems 19 (6) (2003) 1319–1338.
[15] B. Hofmann, R. Krämer, On maximum entropy regularization for a specific inverse problem of option pricing, J. Inverse Ill-Posed Probl. 13 (1) (2005) 41–63.
[16] V. Isakov, Inverse Problems for Partial Differential Equations, 2nd ed., in: Applied Mathematical Sciences, vol. 127, Springer, New York, 2006.
[17] M. Avellaneda, R. Buff, C. Friedman, N. Grandechamp, L. Kruk, J. Newman, Weighted Monte Carlo: a new technique for calibrating asset-pricing models, in: Marco Avellaneda (Ed.), Quantitative Analysis in Financial Markets, in: Collected Papers of the New York University Mathematical Finance Seminar, vol. II, World Scientific, Singapore, 2001, pp. 239–265.
[18] I. Bouchouev, V. Isakov, The inverse problem of option pricing, Inverse Problems 13 (5) (1997) L11–L17.
[19] H. Egger, T. Hein, B. Hofmann, On decoupling of volatility smile and term structure in inverse option pricing, Inverse Problems 22 (4) (2006) 1247–1259.
[20] B. Hofmann, B. Kaltenbacher, C. Pöschl, O. Scherzer, A convergence rates result for Tikhonov regularization in Banach spaces with non-smooth operators, Inverse Problems 23 (3) (2007) 987–1010.
[21] G.A.G. Cidade, C. Anteneodo, N.C. Roberty, A.J.S. Neto, A generalized approach for atomic force microscopy image restoration with Bregman distances as Tikhonov regularization terms, Inverse Probl. Eng. 8 (5) (2000) 457–472.
[22] M. Burger, S. Osher, Convergence rates of convex variational regularization, Inverse Problems 20 (5) (2004) 1411–1421.
[23] E. Resmerita, O. Scherzer, Error estimates for non-quadratic regularization and the relation to enhancement, Inverse Problems 22 (3) (2006) 801–814.
[24] E. Resmerita, Regularization of ill-posed problems in Banach spaces: convergence rates, Inverse Problems 21 (4) (2005) 1303–1314.
[25] F. Clarke, Optimization and Nonsmooth Analysis, 1th ed., Wiley-Interscience, Vancouver, 1983
[26] O. Scherzer, M. Grasmair, H. Grossauer, M. Haltmeier, F. Lenzen, Variational Methods in Imaging, in: Applied Mathematical Sciences, vol. 167, Springer, New York, 2008.
[27] O.A. Ladyzenskaya, V.A. Solonikov, N.N. Urealceva, Linear and Quasilinear Equations of Parabolic Type, in: Translations of Mathematical Monographs, vol. 23, AMS, Providence, RI, 1967.
[28] R.A. Adams, Sobolev Spaces, in: Pure and Applied Mathematics, vol. 65, Academic Press, New York, London, 1975, [A subsidiary of Harcourt Brace Jovanovich, Publishers].
[29] H.W. Engl, M. Hanke, A. Neubauer, Regularization of Inverse Problems, in: Mathematics and its Applications, vol. 375, Kluwer Academic Publishers Group, Dordrecht, 1996.
[30] T. Hein, B. Hofmann, Approximate source conditions for nonlinear ill-posed problems-chances and limitations, Inverse Problems 25 (3) (2009) 035003.
[31] M. Grasmair, M. Haltmeier, O. Scherzer, Sparse regularization with l^q penalty term, Inverse Problems 24 (5) (2009) 055020. 13.
[32] E.B. Andersen, Sufficiency and exponential families for discrete sample spaces, 65 (1970), 1248–1255.
[33] E. Resmerita, R.S. Anderssen, Joint additive Kullback–Leibler residual minimization and regularization for linear inverse problems, 30 (13) (2007), 1527–1544.
[34] J.M. Borwein, A.S. Lewis, Convergence of best entropy estimates, SIAM J. Optim. 1 (2) (1991) 191–205. doi:10.1137/0801014.
[35] T. Fischer, Risk capital allocation by coherent risk measures based on one-sided moments, Insurance Math. Econom. 32 (1) (2003) 135–146.
[36] H. Föllmer, A. Schied, Convex measures of risk and trading constraints, Finance Stoch. 6 (4) (2002) 429–447.
[37] H. Föllmer, A. Schied, Robust preferences and convex measures of risk, in: Advances in Finance and Stochastics, Springer, Berlin, 2002, pp. 39–56.
[38] I. Ekeland, R. Temam, Convex Analysis and Variational Problems, North-Holland, Amsterdam, 1976.
[39] R.T. Rockafellar, Conjugate Duality and Optimization, in: Conference Board of the Mathematical Sciences Regional Conference Series in Applied Mathematics, vol. 16, Society for Industrial and Applied Mathematics, Philadelphia, Pa., 1974, Lectures Given at the Johns Hopkins University, Baltimore, MD, June, 1973.
[40] C. Zălinescu, Convex Analysis in General Vector Spaces, World Scientific Publishing Co. Inc., River Edge, NJ, 2002.
[41] D. Nualart, The Malliavin Calculus and Related Topics, in: Probability and its Applications (New York), Springer-Verlag, New York, 1995.
[42] J. Jackson, E. Suli, S. Howison, Computation of deterministic volatility surfaces, J. Comput. Finance 2 (2) (1998) 5–32.