A Convex-Regularization Framework for Local-Volatility Calibration in Derivative Markets: The Connection with Convex Risk Measures and Exponential Families

A. De Cezaro
IMPA,
Estr. D. Castorina 110,
22460-320 Rio de Janeiro, Brazil
decezaro@impa.br

O. Scherzer*
Computational Science Center, University of Vienna,
Nordbergstr. 15,
A-1090 Vienna, Austria
Otmar.Scherzer@univie.ac.at

Jorge P. Zubelli†
IMPA,
Estr. D. Castorina 110,
22460-320 Rio de Janeiro, Brazil
http://www.impa.br/~zubelli/MATHFIN

Abstract

We present a unified framework for the calibration of local volatility models that makes use of recent tools of convex regularization of ill-posed Inverse Problems.

The unique aspect of the present approach is that it address in a general and rigorous way the key issue of convergence and sensitivity of the regularized solution when the noise level of the observed prices goes to zero. In particular, we present convergence results that include convergence rates with respect to noise level in fairly general contexts and go well beyond the classical quadratic regularization.

Our approach directly relates to many of the different techniques that have been used in volatility surface estimation. In particular, it directly connects with the Statistical concept of exponential families and entropy-based estimation. Finally, we also show that our framework connects with the Financial concept of Convex Risk Measures.

1 Introduction

Good model selection is crucial for modern sound financial practice. In this context, a well accepted class of models consists of the local-volatility ones that were pioneered by Dupire [Dup94]. Once one agrees to work within the context of local-volatility models, a central problem in option pricing is the determination of the local volatility surface from quoted prices. This question has been investigated by a long list of authors and has been the subject of intense scrutiny. See, for example, [ABF+00, Ave98c, Ave98b, Ave98a, AFHS97, BI97, Cré03a, DKZ96, EE05, HKPS07, HK05] and references therein. It is well known that such calibration problem, as many important ones in Finance, is highly ill-posed. In particular, small changes and noises in the data may lead to substantial changes in the results. Yet, good volatility surface calibration is crucial in a plethora of applications, including risk management, hedging, and the evaluation of exotic derivatives.

One is naturally tempted to ask: How could such given data change since it is composed by well defined and fixed market prices? The answer is that the underlying set up of the calibration problem is an infinite dimensional inverse problem which is intrinsically ill-posed. Although this is intuitively well

G. Sutcliffe, A. Voronkov (eds.): easychair 1.0, 2008, volume 1, issue: 1, pp. 1-19

*RICAM, Austrian Academy of Sciences, Altenbergerstr. 69, A-4040 Linz, Austria
†Fields Institute of Mathematical Sciences, Toronto, Canada

known it will be confirmed under fairly general conditions in the sequel. Thus, any finite dimensional observation set will naturally correspond to a (non-unique) element in a function space and such representation alone is already a substantial source of noise. On a more mundane level, the simple fact that one increases the data set by adding a few observation points might completely ruin the reconstruction. The only practical way to deal with such instability issue is by assuming some *a priori* information and regularizing the corresponding volatility surface.

The introduction of regularization techniques in order to the stabilize the problem leads to a crucial question: If the noise in the data goes to zero, does the corresponding regularized solution converge to the true volatility? If this is the case, it would be also natural to inquire about the rate of convergence. If convergence is slow, the regularization is bound to be useless in practical applications.

In this work, we present a general theoretical framework based on Tikhonov regularization by means of a convex penalizing functions. This is an extension of the quadratic regularization that has been previously studied in the Inverse Problem literature [Cré03a, Cré03b, EE05, BJ99]. On the theoretical side, the strength is that this yields better convergence rates with respect to the noise level in the measurements. Furthermore, it 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 [ADEH99, Meu05, FS06].

We recall the concept of a *regular exponential family* of probability distributions. It consists of family of probability distribution functions $p_{\psi,\theta} : \mathbb{R} \to \mathbb{R}_+$ defined by

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

where $\psi : \mathbb{R} \to \mathbb{R} \cup \{+\infty\}$ is convex and $p_0 : \mathbb{R} \to \mathbb{R}_+$ is continuous. 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 [And70]. From our perspective, it is particularly relevant that there is an equivalence between exponential families and Bregman distances [BMDG05]. We stress that an exponential family is a parametric family of probability functions. Fixing the element of the family $\psi(\theta)$ corresponds to the *expectation value*.

The concept of exponential family arises naturally in order to answer the question: What is the maximum entropy distribution consistent with given constraints on expected values? Given the interpretation of option prices as expected values, with respect to appropriate measures which depend on the local volatility surface, a minute's thought leads us to naturally associate the problem of volatility estimation from observed option prices to exponential measures. Financially, 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. This hints to yet another connection with the now classical work developed by Avellaneda *et al*. See [AFHS97] and references therein.

Another component in this picture turns out to be the passage of the regularized volatility to the market probability measures. Indeed, we exhibit procedures to produce convex risk measures which depend on the regularization functional. This in turn, requires the use of Malliavin calculus results which appear in the computation of the greeks [FLL+99a].

The first main contribution of this work consists in establishing existence, stability, and convergence with respect to the noise level of the regularized solutions to the local volatility surface calibration problem under a fairly general class of convex penalizing functionals. We also present, under suitable assumptions, convergence rates. This is the subject of Section 2.

The generality of the convex regularization allows the use of Bregman distances and permits a fairly unified treatment of calibration techniques associated to exponential families of probability distributions. This is presented in Section 3.

Finally the third main point of the present work is the connection of the convex regularization framework. We show that it naturally yields convex coherent risk measures. On the theoretical side, such risk measures can be used to interpret the so-called *source-condition* that appears often in the assumptions of the regularization results. On the economic side, it underscores the role that model selection, which is done by choosing the volatility surface, plays in defining risk aversion or indifference. This is the topic of Section 4

2 Convex Regularization and Bregman Distances

We shall describe our framework on the setting of a single risky-asset market. In such market, besides the risk-free bond we have an asset $S = S_t$ satisfying the dynamics

$$dS_t = S_t(\nu(t,S_t)dt + \sigma(t,S_t)dW_t)\,, \quad t > 0\,,, \tag{1}$$

where W_t denotes the standard Wiener process. The price of, say, a *vanilla European call* is given by the solution of by the solution to the Black-Scholes equation [BS73]

$$\partial_t U + \frac{1}{2}\sigma^2(t,S)S^2\partial_S^2 U + (r-q)S\partial_S U - rU \;=0\,, \quad t < T\,, \tag{2}$$

with final condition

$$U(t=T,S) = (S-K)^+\,. \tag{3}$$

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 [Dup94]

$$-\partial_T U + \frac{1}{2}\sigma^2(T,K)K^2\partial_K^2 U - (r-q)K\partial_K U - qU = 0\,, \quad T > 0\,, \tag{4}$$

with the initial value

$$U(T=0,K) = (S_0 - K)^+, \quad \text{for} \quad K > 0\,. \tag{5}$$

Dupire's equation 4 is the starting point of our inverse problem analysis. Here one might get tempted to say that if we know U at sufficiently many points, then it is enough to compute (numerically) the derivative in Equation 4. However, such computation is extremely ill-posed. In fact the financial literature possesses a vast number of approaches to overcome such difficulties.

For simplicity and notational convenience we take $q = r = 0$. We stress that such choice is immaterial in the upcoming results. We also perform the usual change of variables to the log-price and the time-to-maturity variables. This corresponds to

$$K = S_0 e^y\,,\; \tau = T - t\,,\; b = q - r\,, \quad u(\tau,y) = e^{q\tau}U^{t,S}(T,K) \tag{6}$$

and

$$a(\tau,y) = \frac{1}{2}\sigma^2(T-\tau; S_0 e^y)\,, \tag{7}$$

in (4) and (5). This yields the Dupire equation with forward variables (τ, y)

$$u_\tau = a(\tau,y)(\partial_y^2 u - \partial_y u) \tag{8}$$

and initial condition

$$u(0,y) = S_0(1-e^y)^+ \tag{9}$$

Existence and uniqueness results for the solution of the parabolic equation (8) and (9) in Sobolev spaces can be found in [Cré03a, EE05, Isa06]. This defines the direct problem that associates the model to the data through the map: $F : a \longmapsto u$ in appropriate spaces that will be specified below.

Volatility calibration in extended Black - Scholes models has been investigated by many authors. See [ABF$^+$00, ABF$^+$01, AFHS97, BI97, Cré03a, EE05, HKPS07, HK05] 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 [Cré03a] and later by Egger & Engl [EE05]. In [HK05] 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 [HKPS07], Hofmann *et al.* analyzed the same financial problem of [HK05] with general source conditions for the regularization functional $f(\cdot) = \|\cdot\|^2_{L^2(0,T)}$. In [Cré03a, EE05, HKPS07, HK05], the ill-posedness of the inverse problem is proved, convergence and convergence rates of a regularized solution are derived.

We now define the admissible *convex* class of calibration parameters. For $0 \leq \varepsilon$ fixed, take $U := H^{1+\varepsilon}(\Omega)$ with the standard $H^{1+\varepsilon}$–inner product $\langle \cdot, \cdot \rangle$ and let $\overline{a} > \underline{a} > 0$. Fix the *a priori* model a_0 as 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

Definition 1.

$$\mathscr{D}(F) := \{a \in a_0 + U : \underline{a} \leq a \leq \overline{a}\}. \tag{10}$$

We consider convex regularization as discussed in [BO04, HKPS07, RS06] to solve the ill-posed operator equation

$$F(a) = u(a), \tag{11}$$

where $F : \mathscr{D}(F) \subset U \to V$ is the parameter-to-solution operator between Hilbert spaces U and $V := L^2(\Omega)$. Here $u(a)$ is the solution of (8) and (9).

For given convex f the proposed methods consists in minimizing the Tikhonov functional

$$\mathscr{F}_{\beta,u^\delta}(a) := ||F(a) - u^\delta||^2_{L^2(\Omega)} + \beta f(a) \tag{12}$$

over $\mathscr{D}(F)$, where, $\beta > 0$ is the regularization parameter. Remark that f incorporates the *a priori* info on a.

The idea of convex regularization for inverse problems has been suggested by different authors. An early reference on Bregman distance regularization is [CARN00]. See also [BO04, HKPS07, RS06] and references therein.

In order to exemplify the power of the theory, we shall now present a number of assumptions that will specify the class of admissible functionals. Needless to say that one can conceive various alternatives to such assumptions as well as generalizations thereof.

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

As alluded in the Introduction, in practice 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. The level δ is assumed to be known *a priori* and satisfies the inequality

$$||\bar{u} - u^\delta||_{L^2(\Omega)} \leq \delta\,, \tag{13}$$

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

An important tool in the studies of Tikhonov type regularization [BO04, HKPS07, Res05, RS06] is the Bregman distance with respect to f.

Definition 3. *Let f as in Assumption 2. For given $a \in \mathscr{D}(f)$, let $\partial f(a) \subset U$ denote the subdifferential of the functional f at a. We denote by*

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

the domain of the subdifferential [Cla83]. The Bregman distance with respect to $\zeta \in \partial f(a_1)$ is defined on $\mathscr{D}(f) \times \mathscr{D}(\partial f)$ by

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

An important issue, both from the theoretical as well as from the practical point of view if the domain of definition where we the volatility is defined. In this work we will let $I \subset \mathbb{R}$ denote an open (possibly unbounded) interval. We assume that $T > 0$ and use the notation $\Omega := (0,T) \times I$.

As usual, for $1 \le p < \infty$, $W_p^{1,2}(\Omega)$ denotes the space of functions $u(\cdot,\cdot)$ satisfying

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

The next assumption is motivated by recent abstract convergence results for Tikhonov regularization [SGG$^+$08].

Assumption 4.

1. *The Banach spaces U and V are endowed with topologies τ_U and τ_V that are weaker than the norm topologies.*

2. *The norm $\|\cdot\|_V$ is sequentially lower semi-continuous with respect to τ_V.*

3. *The functional $f : \mathscr{D}(f) \subseteq U \to [0,\infty]$ is convex and sequentially lower semi-continuous with respect to τ_U and $\mathscr{D} := \mathscr{D}(F) \cap \mathscr{D}(f) \neq \emptyset$. In the context of this paper we have $\mathscr{D}(F) \neq \emptyset$ and $\mathscr{D}(F) \subseteq \mathscr{D}(f)$ and thus the assumption is satisfied.*

4. *Let $\mathscr{F}_{\beta,\bar{u}}$ the Tikhonov functional defined in* (12). *Then,*

$$\mathscr{M}_\beta(M) := level_M(\mathscr{F}_{\beta,\bar{u}}) = \{a : \mathscr{F}_{\beta,\bar{u}}(a) \le M\}$$

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

The first three conditions of Assumption 4 are satisfied for our particular problem. In Section A.1 we shall analyze the last condition of Assumption 4. The general result of [SGG$^+$08] then implies well-posedness, stability, convergence. These results are summarized below.

Theorem 5 (Existence, Stability, Convergence)**.** *Suppose that F, f, $\mathscr{D}$, U, and V satisfy Assumption 4. Furthermore, assume that $\beta > 0$ and $u^\delta \in V$. Then, we have that*

- *There exists a minimizer of $\mathscr{F}_{\beta,u^\delta}$.*

- *If (u_k) is a sequence converging to u in V with respect to the norm topology, then every sequence (a_k) with*

$$a_k \in argmin\big\{\mathscr{F}_{\beta,u_k}(a) : a \in \mathscr{D}\big\}$$

has a subsequence which converges with respect to τ_U. The limit of every τ_U-convergent subsequence $(a_{k'})$ of (a_k) is a minimizer $\tilde{a}$ of $\mathscr{F}_{\beta,u}$, and $\big(f(a_{k'})\big)$ converges to $f(\tilde{a})$.

- *If there exists a solution of* (11) *in $\mathscr{D}$, then there exists an f-minimizing solution of* (11).
- *Assume that* (11) *has a solution in $\mathscr{D}$ (which implies the existence of an f-minimizing solution) and that $\beta : (0,\infty) \to (0,\infty)$ satisfies*

$$\beta(\delta) \to 0 \text{ and } \frac{\delta^2}{\beta(\delta)} \to 0\,, \text{ as } \delta \to 0\,. \tag{14}$$

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 $\mathscr{F}_{\beta_k,u_k}$, has a subsequence $(a_{k'})$ that converges with respect to τ_U. The limit $a^\dagger$ of any τ_U convergent subsequence $(a_{k'})$ is an f-minimizing solution of (11)*, 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_U.*

The proof can be found in [CSZ10]. In order to obtain convergence rate results will we require a further assumption. Namely,

Assumption 6. *Besides Assumption 4, we assume that*

1. *There exists an f-minimizing solution $a^\dagger$ of* (11)*, which is an element of the Bregman domain $\mathscr{D}_B(f)$.*
2. *There exist $\beta_1 \in [0,1)$, $\beta_2 \geq 0$, and $\zeta^\dagger \in \partial f(a^\dagger)$ such that*

$$\langle \zeta^\dagger, a^\dagger - a \rangle \leq \beta_1 D_{\zeta^\dagger}(a, a^\dagger) + \beta_2 \left\| F(a) - F(a^\dagger) \right\|_V\,, \tag{15}$$

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

Under this assumption we have the following:

Theorem 7 (Convergence rates [SGG+08])**.** *Let F, f, $\mathscr{D}$, U, and V satisfy Assumption 6. Moreover, let $\beta : (0,\infty) \to (0,\infty)$ satisfy $\beta(\delta) \sim \delta$. Then*

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

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

Although Assumption 4 may seem too restrictive, the next result reveals that it can be obtained from rather classical ones:

Proposition 8. *Let F, f, $\mathscr{D}$, U, and V satisfy Assumption 4. Assume that there exists an f-minimizing solution $a^\dagger$ of* (11)*, and that F is Gâteaux 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

$$\zeta^\dagger := F'(a^\dagger)^* \omega^\dagger \in \partial f(a^\dagger) \tag{16}$$

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

$$\left\| F(a) - F(a^\dagger) - F'(a^\dagger)(a - a^\dagger) \right\| \leq \gamma D_{\zeta^\dagger}(a, a^\dagger)\,, \text{ for } a \in \mathscr{M}_{\beta_{max}}(\rho)\,. \tag{17}$$

Then, Assumption 6 holds.

Once again, for the the application of Proposition 8 to the local volatility problem see [CSZ10]. Some of the results are outlined in the Technical Appendix at the end of the paper. We note that in [CSZ10] $U = H^{1+\varepsilon}(\Omega)$, with $\varepsilon > 0$ and $V = L^2(\Omega)$ have been used. Since both spaces are Hilbert spaces we can use the inner product on U and the adjoint operator $F'(a^\dagger)$ instead of the duality pairing of $F'(a^\dagger)$, $F'(a^\dagger)^\#$, respectively, as in [SGG+08].

3 Exponential Families and the Statistical Approach

We start by recalling the notion of Fenchel conjugate which plays an important role in convex analysis. Let X be a normed space and $X^\star$ be its dual endowed with the duality pairing $\langle\cdot,\cdot\cdot\rangle$. Given a function $f : X \to \mathbb{R} \cap +\infty$, the *Fenchel* dual $f^* X^* \to \mathbb{R} \cap +\infty$ is defined by

$$f^*(x^*) := \sup\{\langle x^*, x\rangle - f(x) \mid x \in X\}$$

The following result connects, through Fenchel duality, the concept of Bregman distances with Exponential Families of distributions.

Theorem 9 (Banerjee et al. [BMDG05]). *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 $a(\theta) \in int(dom(\psi^*))$, then

$$p_{\psi,\theta}(a) = \exp\big(-D_{\psi^*}(a, a(\theta))\big) \exp\big(\psi^*(a)\big) p_0(a) \,. \tag{18}$$

Example 10 (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 $\psi^*(a) = a\log(a) - a$.*

In order to illustrate, the strength of the convex regularization theory in connection with Statistics, we shall now motivate Bregman distance regularization as a log-maximum a-posteriori estimator for an exponential family. We do this in a discrete statistical setting. As usual, we consider $(\mathscr{X}, \mathscr{F}, \mathbb{P})$ a probability space. We let $\vec{x} := (x_i)_i$ be a sequence of elements in $\mathscr{X}$ 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 $a \in \mathbb{R}_{>0}$ assume that $\theta := \theta(a)$. 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 9, 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.

Summing up, we have shown that Bregman distance regularization can be considered a log maximum a-posteriori estimator for the expectation number.

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 subdomain of $\mathbb{R}^2$. With this framework, we remark that $\mathscr{D}(F) \subset U \cap L^\infty_{>0}(\Omega) \subset L^1(\Omega)$, where $L^\infty_{>0}(\Omega)$ is the set of functions that are (essentially) bounded from below and above by some positive constants.

Example 11. *According to Example 10, if we use the exponential family associated to Poisson distributions, we obtain Kullback-Leibler regularization, consisting in minimization of*

$$a \longmapsto \mathscr{F}_{\beta,u^\delta}(a) := \left\| F(a) - u^\delta \right\|^2_{L^2(\Omega)} + \beta KL(\hat{a}, a) \;, \tag{19}$$

where

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

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

$$\mathscr{G}(a) := \int_\Omega a \log(a)\, dx \;. \tag{20}$$

We also note that the standard Kullback-Leibler regularization has been analyzed for instance in [RA07]. Since the Bregman distance is not symmetric, the component, which one is optimizing for makes a difference. Here, modelling with exponential families results in Bregman distances with respect to the second component.

The domains of $\mathscr{G}$, $\mathscr{D}(\mathscr{G})$, and of the subgradient of $\mathscr{G}$, $\mathscr{D}(\partial\mathscr{G})$, 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 in the Bregman domain $\mathscr{D}_B(\mathscr{G})$, that is a subset of $L^\infty_{>0}$. Moreover, the Kullback-Leibler distance is lower semi-continuous with respect to the L^1-norm [RA07]. Based on this property we extend the Kullback-Leibler distance, to take value $+\infty$ if either $a \notin \mathscr{D}(\mathscr{G})$ or $b \notin \mathscr{D}_B(\mathscr{G})$.

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 subgradient of the Boltzmann-Shannon entropy. This follows directly from the definition of the domains of the convex functionals and subgradients.

To prove that minimization of $\mathscr{F}_{\beta,u^\delta}$ in (19) is well–posed we have to choose appropriate spaces and topologies first. We choose $\tau_{\tilde{U}}$, $\tau_{\tilde{V}}$ the weak topologies on $L^1(\Omega)$ and $L^2(\Omega)$, respectively

Lemma 12. *Let Ω be a bounded subset of $\mathbb{R}^2$ with Lipschitz boundary. Moreover, assume that F is continuous with respect to the weak topologies on $L^1(\Omega)$ and $L^2(\Omega)$, respectively.*

1. *Let $a, b \in \mathscr{D}(\mathscr{G})$. Then*

$$\|a - b\|^2_{L^1(\Omega)} \leq \left(\frac{2}{3} \|a\|_{L^1(\Omega)} + \frac{4}{3} \|b\|_{L^1(\Omega)} \right) KL(a, b) \;. \tag{21}$$

Here, we set $0 \cdot (+\infty) = 0$.

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

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

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

are $\tau_{\tilde{U}}$ *sequentially compact.*

For a proof, see [CSZ10].

Using standard results on variational regularization (see for instance [SGG+08]), we have:

Theorem 13. *There exists a minimizer of* $\mathscr{F}_{\beta,u^\delta}$ *in* (19)*. The minimizers are stable and convergent for* $\beta(\delta) \to 0$ *and* $\delta^2/\beta(\delta) \to 0$*. Stable means that* $\operatorname{argmin} \mathscr{F}_{\beta,u^{\delta_k}} \to \operatorname{argmin} \mathscr{F}_{\beta,u^0}$ *for* $\delta_k \to 0$ *and that* $\operatorname{argmin} \mathscr{F}_{\beta(\delta_k),u^{\delta_k}}$ *converges to a solution of* (11) *with minimal energy.*

Remark 14. *Note that since* $\mathscr{D}(F) \subset U$ *we have* $\mathscr{D}(F) \subset \mathscr{D}_B(\mathscr{G})$*. Moreover, from Theorem 26,* $F : \mathscr{D}(F) \subset U \to W_2^{1,2}(\Omega)$ *is weakly continuous. If* Ω *is bounded, we have* $\mathscr{D}(F) \subset U \subset L^2(\Omega) \subset L^1(\Omega)$ *and* $W_2^{1,2}(\Omega) \subset L^2(\Omega)$*, with continuous embedding. It follows that* $F : \mathscr{D}(F) \subset L^1(\Omega) \to L^2(\Omega)$ *is weakly continuous, i.e., satisfies the assumptions on the Lemma 12.*

An important consequence of (21) and Theorem 7 is that

$$\left\| a_\beta^\delta - a^\dagger \right\|_{L^1(\Omega)} = \mathscr{O}(\sqrt{\delta})\,. \tag{22}$$

Now, let δ_k be a sequence converging to zero and $a_k = a_{\beta_k}^{\delta_k}$ the respective minimizers of the Tikhonov functional (12). Take $b_k = a^\dagger$ for all $k \in \mathbb{N}$. Then, from Lemma 12 Item 2 we have

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

4 Convex Risk Measures

In this section we connect the convex regularization framework described in the previous sections with the so-called (coherent) convex risk measures. The latter have been the subject of intense attention for the past few years. See [ADEH99, Fis03, FS02a, FS02b] and references therein.

We recall that a *convex measure of risk* consists of a map $\rho : \mathscr{X} \longrightarrow \mathbb{R}$ satisfying the following properties:

- 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
$$\rho(\nu + m) = \rho(\nu) - m\,. \tag{23}$$

Throughout this section, for technical reasons, we assume that the domain $\Omega = [0,T]\times I$ where our local volatility is unknown is bounded. Furthermore, we define the functional $f(a) = +\infty$ if $a \notin \mathscr{D}(F)$.

Using the assumption of existence of a source function $w^\dagger \in L^2(\Omega)$ that satisfies (16) and the definition of $\partial f(a^\dagger)$ we have that

$$f(a) - \langle w^\dagger, F'(a^\dagger)a\rangle \geq f(a^\dagger) - \langle w^\dagger, F'(a^\dagger)a^\dagger\rangle\,, \qquad \forall a \in U \text{ and } \forall w^\dagger \text{ s.t.} F'(a^\dagger)^* w^\dagger \in \partial f(a^\dagger)\,. \tag{24}$$

Let us set $g(-F'(a^\dagger)a) := \langle w, -F'(a^\dagger)a\rangle$. The existence of $w^\dagger$ satisfying (24) implies that it is the Lagrangian multiplier of

$$\begin{aligned} L : \mathscr{D}(F)\times L^2(\Omega) &\longrightarrow \mathbb{R} \\ (a,w) &\to f(a) + g(-F'(a^\dagger)a)\,, \end{aligned}$$

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 $w^\dagger \in \mathscr{R}(F'(a^\dagger))$ satisfying (24). 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, it can be shown [CSZ10] that $w^\dagger$ satisfies Equation (16) 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

$$\begin{aligned} L^2(\Omega) \ni w \longmapsto h_w : \mathscr{D}(F) &\longrightarrow \mathbb{R}\cup\{+\infty\} \\ a &\longmapsto L(a,w) = f(a) + g(-F'(a^\dagger)a)\,. \end{aligned} \tag{25}$$

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

A few comments are in order.

Remark 15. *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 (24).

Remark 16. *Note, that the source condition* (16) *together with the existence of an f-minimum norm solution for* (11) *is equivalent to the Karush-Kuhn-Tucker condition in convex optimization [ET76].*

Now, from the theory of Fenchel conjugation [Roc74, Zăl02], we obtain a unique Fenchel conjugate function of h_w given by

$$\begin{aligned} \hat{h}^*_w : L^2(\Omega) &\longrightarrow \mathbb{R} \\ v &\longmapsto g^*(v) + f^*(-F'(a^\dagger)^* v)\,. \end{aligned} \tag{26}$$

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

$$\begin{aligned} h^*_w : \mathbb{X} \subset L^2(\Omega) &\longrightarrow \mathbb{R} \\ v &\longmapsto h^*_w(v) := f^*(-F'(a^\dagger)^* v)\,, \end{aligned} \tag{27}$$

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

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

Lemma 17. *The functional h_w^* satisfies the convexity and monotonicity axioms.*

Proof. The convexity follows directly from the properties of the Fenchel conjugate function [Zăl02, 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 [Cré03a, Theorem 4.2]) implies that

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

□

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 [FLL+99a].

We start by relating our notation with that of [FLL+99a]. Equation (8) is associated to the diffusion process $\{y_t \,:\, 0 \leq t \leq T\}$ that satisfies the dynamics

$$dy_t = \left(r - q - \frac{\sigma(t,y_t)^2}{2}\right) dt + \sigma(t,y_t) dW_t\,, \qquad y_{t_0} = y_0\,, \tag{28}$$

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

We recall that the process (28) is the diffusion (1) in a logarithmic variables where is defined by (7).

As before, for the sake of simplicity, we assume that the process (28) has no dividend and interest rates, i.e., $b = 0$.

Following [FLL+99a], 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 18. *We now identify $\sigma^\dagger \longmapsto \sqrt{2a^\dagger}$ and $\tilde{\sigma} \longmapsto \sqrt{2\tilde{a}}$ given by* (7) *with $a^\dagger, \tilde{a} \in \mathscr{D}(F)$. 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

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

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

Our first result is a representation lemma.

Lemma 19. *Let $v \in \mathscr{R}(F'(a^\dagger))$. Then, there exists a random variable $\pi_{a^\dagger}$ such that*

$$v = \mathbb{E}_{\mathbb{Q}}^{y_0}[\Phi(y_t)\pi_{a^\dagger}]\,, \tag{29}$$

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

Proof. Let

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

where $\{\beta(t) : 0 \le t \le T\}$ is the process given in [FLL+99a, Lemma 3.1].

Since $\sigma^\dagger + \varepsilon\tilde{\sigma}$ satisfies the uniform ellipticity condition (see Remark 18) we have from [FLL+99a, Proposition 3.3] that the Gâteaux derivative at $\sigma^\dagger$ in the direction $\tilde{\sigma}$ is given by

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

where $D_t^*((\sigma^\dagger)^{-1}(y_t)Y_t\tilde{\beta}_\Theta(T))$ is the Skorohod integral [Nua95] of the possibly anticipative process

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

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 [FLL+99a].

Lemma 20. *The constants do not belong to $\mathscr{R}(F'(a^\dagger))$.*

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

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

However, $(u_{yy} - u_y)$ is the Green's function of a parabolic well-posed problem [CSZ10] and thus 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,

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

and secondly,

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

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

Remark 21. *It follows from Lemma 19 that we have a representation of $\mathbb{X}$ by $\mathscr{X}$ and $\mathscr{X}_1$ given by the weighted expectation $\mathbb{E}_{\mathbb{Q}}^{y_0}[\cdot]$ with weight factor $D_t^*((\sigma^\dagger)^{-1}(y_t)Y_t\tilde{\beta}_\Theta(T))$ and $\Phi(y_t)$ respectively.*

We now bring in the following useful result:

Lemma 22. *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 [Nua95], and the fact that $D_t 1 = 0$. □

This allows us to construct convex risk measures. More precisely, we have:

Proposition 23. *[First alternative for a convex risk measure] The functional*

$$\rho : \mathscr{X} \longrightarrow \mathbb{R} \qquad \nu \longmapsto \rho(\nu) := h_w^*(\mathbb{E}_{\mathbb{Q}}^{y_0}[\nu \cdot \pi_{a^\dagger}]) - \mathbb{E}_{\mathbb{Q}}^{y_0}[\nu] \tag{30}$$

satisfies the convex risk measure axioms.

Proof. By the linearity of the expectation operator and the properties of the functional h_w^* in Lemma 17, the convexity and monotonicity axioms follows.

In order to prove the translation axiom, we write

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

Let $\nu + m \in \mathscr{X}$. 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, $\mathscr{X} = \mathscr{D}(\tilde{\rho}) = \mathscr{D}(\rho)$. Let us take now $\nu \in \mathscr{X}$. Then, by definition of expectation $\mathbb{E}_{\mathbb{Q}}^{y_0}[\nu] = c$ where c is a constant. It follows from Lemma 22 that

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

Thus $\tilde{\rho} = \rho$. □

Proposition 24. *[Second alternative for a convex measure of risk] The functional*

$$\rho_1 : \mathscr{X}_1 \longrightarrow \mathbb{R} \qquad \pi \longmapsto \rho_1(\pi) := h_w^*(\mathbb{E}_{\mathbb{Q}}^{y_0}[\nu \cdot \pi])\,, \tag{31}$$

satisfies the convex risk measure axioms.

Proof. Using the same argument of Proposition 23, the convexity and monotonicity axioms follow. In order to prove the translation axiom, we write

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

Then, for $\pi + m \in \mathscr{X}_1$, 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, $\mathscr{X}_1 = \mathscr{D}(\tilde{\rho}_1) = \mathscr{D}(\rho_1)$. Let us take $\pi \in \mathscr{X}_1$. From Lemma 22 we conclude that $\mathbb{E}_{\mathbb{Q}}^{y_0}[\pi] = \mathbb{E}_{\mathbb{Q}}^{y_0}[1 \cdot \pi] = 0$. Thus, $\tilde{\rho}_1(\pi) = \rho(\pi) \quad \text{for all} \quad \pi \in \mathscr{X}_1\,.$ □

We note that the choice of $\sigma^\dagger$ enters in a crucial and nonlinear way in the convex risk measure. Furthermore, the source condition (16) allows us to construct convex risk measures in the spaces of random variables associated to the diffusion process (28).

We now illustrate the construction of the convex risk measure by considering the process (28) (and as before under constant volatility with vanishing interest and dividend rates). For this particular case, the representation (29) (or the *vega* in financial terms) is given by the formula (see [FLL+99a])

$$\begin{aligned}&\mathbb{E}_{\mathbb{Q}}^{y_0}\left[\Phi\left(y\exp\left(\sigma^\dagger W_\tau - \frac{(\sigma^\dagger)^2}{2}\tau\right)\right) \cdot \left(\frac{W_\tau^2}{\sigma^\dagger \tau} - W_\tau - \frac{1}{\sigma^\dagger}\right)\right] \\ &\quad = \int_\Omega dz\, d\tau\, p(z,\tau) \Phi\left(y\exp\left(\sigma^\dagger z - \frac{(\sigma^\dagger)^2}{2}\tau\right)\right) \cdot \left(\frac{z^2}{\sigma^\dagger \tau} - z - \frac{1}{\sigma^\dagger}\right)\,,\end{aligned} \tag{32}$$

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,

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

Thus,

$$
-F'(a^\dagger)^* v = \left(\frac{z^2}{\sigma^\dagger\tau} - z - \frac{1}{\sigma^\dagger}\right)\langle -v, \Phi(\cdot)\rangle\,. \tag{33}
$$

We now exemplify the regularization functional f as the Boltzmann-Shannon entropy

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

whose Fenchel conjugate is given by

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

Since we are in a Gaussian model, applying [AS06, Lemma 11] and (33) to the definition of ρ with $\nu = \Phi\left(y\exp\left(\sigma^\dagger(z) - (\sigma^\dagger)^2\tau/2\right)\right)$ we get

$$
\rho(\nu) = -\log\left(\mathbb{E}_{\mathbb{Q}}^{y_0}\left[\exp\left(\frac{z^2}{\sigma^\dagger\tau} - z - \frac{1}{\sigma^\dagger}\right)\langle -v, \nu\rangle\right]\right) - \mathbb{E}_{\mathbb{Q}}^{y_0}[\nu]\,. \tag{34}
$$

5 Conclusions and Future Work

The problem of volatility surface calibration is a classical and fundamental one in Quantitative Finance. See [BJ99, DKZ96] for an early survey of the literature. We address this problem by presenting a unifying framework for the regularization that makes use of tools from Inverse Problem theory and Convex Analysis.

We illustrate our approach's theoretical strength by establishing convergence and convergence rate results. In this respect, the main novelty is the use of a regularization term that only requires convexity properties and weak lower-semicontinuity. Thus, the present regularization applies to a large class of regularization functionals. In particular, we connect with the statistical viewpoint through the concept of exponential families. This is turn, allows the derivation 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 Equation (22). Another advantage of the current framework is the requirement of weaker conditions than those previously required in the literature. Namely, we only require (36).

The convergence results also hold true if we measure the misfit of the Tikhonov functional (12) in $W_p^{1,2}(\Omega)$. See [CSZ10]. This may have implications at the level of stabilization by means of less discrepancy of the greeks, since the $W_p^{1,2}(\Omega)$ norm involves information associated to Θ, Δ, and Γ.

The connection with exponential families opens the door to recent works on entropy-based estimation methods. It also connects to the work developed by Avellaneda and collaborators on minimum-relative-entropy regularization [Ave98a, Ave98c, Ave98b, AFHS97]. In particular, it might allow more information on the convergence rates of such approach.

The connection with convex risk measures required the use of techniques from Malliavin calculus. In fact, such techniques have been attracting substantial attention from the Quantitative Finance community since the pioneering work [FLL$^+$99b]. See also [BET04] and references therein. It would be interesting to investigate further the relation between the convex risk measures we obtain and the regularization properties of the corresponding functionals. This might even allow to incorporate risk aversion or indifference in the choice of such functional.

Finally, a natural continuation of the present work and of that described in [CSZ10] would be to consider applications to vol-surface estimation in the multi-asset case as well to fixed-income markets.

Acknowledgments

The final version of this manuscript was written while JPZ was at the Fields Institute, Toronto, whose hospitality is gratefully acknowledged. JPZ was supported by CNPq under grants 302161 and 474085. 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.

References

[ABF$^+$00] M. Avellaneda, R. Buff, C. Friedman, N. Grandchamp, L. Kruk, and J. Newman. Weighted Monte Carlo: A new technique for calibrating asset-pricing models. Spigler, Renato (ed.), Applied and industrial mathematics, Venice-2, 1998. Selected papers from the 'Venice-2/Symposium', Venice, Italy, June 11-16, 1998. Dordrecht: Kluwer Academic Publishers. 1-31 (2000)., 2000.

[ABF$^+$01] M. Avellaneda, R. Buff, C. Friedman, N. Grandechamp, L. Kruk, and J. Newman. Weighted Monte Carlo: A new technique for calibrating asset-pricing models. Avellaneda, Marco (ed.), Quantitative analysis in financial markets. Collected papers of the New York University mathematical finance seminar. Vol. II. Singapore: World Scientific. 239-265 (2001)., 2001.

[Ada75] R. A. Adams. *Sobolev spaces*. Academic Press [A subsidiary of Harcourt Brace Jovanovich, Publishers], New York-London, 1975. Pure and Applied Mathematics, Vol. 65.

[ADEH99] P. Artzner, F. Delbaen, J-M. Eber, and D. Heath. Coherent measures of risk. *Math. Finance*, 9(3):203–228, 1999.

[AFHS97] M. Avellaneda, C. Friedman, R. Holmes, and D. Samperi. Calibrating volatility surfaces via relative-entropy minimization. *Appl. Math. Finance*, 4(1):37–64, 1997.

[And70] E. B. Andersen. Sufficiency and exponential families for discrete sample spaces. 65:1248–1255, 1970.

[AS06] Y. Altun and A. Smola. Unifying divergence minimization and statistical inference via convex duality. In *Learning theory*, volume 4005 of *Lecture Notes in Comput. Sci.*, pages 139–153. Springer, Berlin, 2006.

[Ave98a] M. Avellaneda. Minimum-relative-entropy calibration of asset-pricing models. *International Journal of Theoretical and Applied Finance*, 1(4):447–472, 1998.

[Ave98b] Marco Avellaneda. The minimum-entropy algorithm and related methods for calibrating asset-pricing model. In *Trois applications des mathématiques*, volume 1998 of *SMF Journ. Annu.*, pages 51–86. Soc. Math. France, Paris, 1998.

[Ave98c] Marco Avellaneda. The minimum-entropy algorithm and related methods for calibrating asset-pricing models. In *Proceedings of the International Congress of Mathematicians, Vol. III (Berlin, 1998)*, number Extra Vol. III, pages 545–563 (electronic), 1998.

[BET04] Bruno Bouchard, Ivar Ekeland, and Nizar Touzi. On the Malliavin approach to Monte Carlo approximation of conditional expectations. *Finance Stoch.*, 8(1):45–71, 2004.

[BI97] I. Bouchouev and V. Isakov. The inverse problem of option pricing. *Inverse Problems*, 13(5):L11–L17, 1997.

[BJ99] James N. Bodurtha, Jr. and Martin Jermakyan. Nonparametric estimation of an implied volatility surface. *Journal of Computational Finance*, 2(4), Summer 1999.

[BMDG05] A. Banerjee, S. Merugu, I.S. Dhillon, and J. Ghosh. Clustering with bregman divergences. *Journal of Machine Learning Research*, 6:1705–1749, 2005.

[BO04] M. Burger and S. Osher. Convergence rates of convex variational regularization. *Inverse Problems*, 20(5):1411–1421, 2004.

[BS73] F. Black and M. Scholes. The pricing of options and corporate liabilities. *J. Political Economy*, 81:637–659, 1973.

[CARN00] G. A. G. Cidade, C. Anteneodo, N. C. Roberty, and A. J. S. Neto. A generalized approach for atomic force microscopy image restoration with Bregman distances as Tikhonov regularization terms. *Inverse Problems in Engineering*, 8(5):457–472, 2000.

[Cla83] F.H. Clarke. *Optimization and Nonsmooth Analysis*. Wiley-Interscience, Vancouver, 1983. 1th edition.

[Cré03a] S. Crépey. Calibration of the local volatility in a generalized Black-Scholes model using Tikhonov regularization. *SIAM J. Math. Anal.*, 34(5):1183–1206 (electronic), 2003.

[Cré03b] S. Crépey. Calibration of the local volatility in a trinomial tree using Tikhonov regularization. *Inverse Problems*, 19(1):91–127, 2003.

[CSZ10] A. De Cezaro, O. Schertzer, and J.P Zubelli. Convex Regularization of Local Volatility Models from Option Prices: Convergence Analysis and Rates. Submitted, 2010.

[DKZ96] Emanuel Derman, Iraj Kani, and Joseph Z. Zou. The local volatility surface: Unlocking the information in index option prices. *Financial Analysts Journal*, 52(4):25–36, 1996.

[Dup94] B. Dupire. Pricing with a smile. *Risk*, 7:18– 20, 1994.

[EE05] H. Egger and H. W. Engl. Tikhonov regularization applied to the inverse problem of option pricing: convergence analysis and rates. *Inverse Problems*, 21(3):1027–1045, 2005.

[ET76] I. Ekeland and R. Temam. *Convex Analysis and Variational Problems*. North-Holland, Amsterdam, 1976.

[Fis03] T. Fischer. Risk capital allocation by coherent risk measures based on one-sided moments. *Insurance Math. Econom.*, 32(1):135–146, 2003.

[FLL+99a] E. Fournié, J.-M. Lasry, J. Lebuchoux, P.-L. Lions, and N. Touzi. Applications of Malliavin calculus to Monte Carlo methods in finance. *Finance Stoch.*, 3(4):391–412, 1999.

[FLL+99b] Eric Fournié, Jean-Michel Lasry, Jérôme Lebuchoux, Pierre-Louis Lions, and Nizar Touzi. Applications of Malliavin calculus to Monte Carlo methods in finance. *Finance Stoch.*, 3(4):391–412, 1999.

[FS02a] H. Föllmer and A. Schied. Convex measures of risk and trading constraints. *Finance Stoch.*, 6(4):429–447, 2002.

[FS02b] H. Föllmer and A. Schied. Robust preferences and convex measures of risk. In *Advances in finance and stochastics*, pages 39–56. Springer, Berlin, 2002.

[FS06] Marco Frittelli and Giacomo Scandolo. Risk measures and capital requirements for processes. *Math. Finance*, 16(4):589–612, 2006.

[HH09] T. Hein and B. Hofmann. Approximate source conditions for nonlinear ill-posed problems—chances and limitations. *Inverse Problems*, 25, 2009.

[HK05] B. Hofmann and R. Krämer. On maximum entropy regularization for a specific inverse problem of option pricing. *J. Inverse Ill-Posed Probl.*, 13(1):41–63, 2005.

[HKPS07] B. Hofmann, B. Kaltenbacher, C. Pöschl, and O. Scherzer. A convergence rates result for Tikhonov regularization in Banach spaces with non-smooth operators. *Inverse Problems*, 23(3):987–1010, 2007.

[Isa06] V. Isakov. *Inverse problems for partial differential equations*, volume 127 of *Applied Mathematical Sciences*. Springer, New York, second edition, 2006.

[Meu05] A. Meucci. *Risk and asset allocation*. Springer Finance. Springer-Verlag, Berlin, 2005.

[Nua95] D. Nualart. *The Malliavin calculus and related topics*. Probability and its Applications (New York). Springer-Verlag, New York, 1995.

[RA07] E. Resmerita and R. S. Anderssen. Joint additive Kullback–Leibler residual minimization and regularization for linear inverse problems. 30(13):1527–1544, 2007.

[Res05] E. Resmerita. Regularization of ill-posed problems in Banach spaces: convergence rates. *Inverse Problems*, 21(4):1303–1314, 2005.

[Roc74] R. T. Rockafellar. *Conjugate duality and optimization*. Society for Industrial and Applied Mathematics, Philadelphia, Pa., 1974. Lectures given at the Johns Hopkins University, Baltimore, Md., June, 1973, Conference Board of the Mathematical Sciences Regional Conference Series in Applied Mathematics, No. 16.

[RS06] E. Resmerita and O. Scherzer. Error estimates for non-quadratic regularization and the relation to enhancement. *Inverse Problems*, 22(3):801–814, 2006.

[SGG+08] O. Scherzer, M. Grasmair, H. Grossauer, M. Haltmeier, and F. Lenzen. *Variational Methods in Imaging*, volume 167 of *Applied Mathematical Sciences*. Springer, New York, 2008.

[Zăl02] C. Zălinescu. *Convex analysis in general vector spaces*. World Scientific Publishing Co. Inc., River Edge, NJ, 2002.

A Technical Appendix

A.1 Properties of the forward operator and ill-posedness of the inverse problem

In this section we verify Item 4 of Assumption 4. In this Appendix, we denote by $U = H^{1+\varepsilon}(\Omega)$ and $V = L^2(\Omega)$. This allows us to apply the Theorem 5 in order to guarantee well-posedness, stability, and convergence of the regularized solutions of the Tikhonov functional (12). We use the following definition of compactness:

Definition 25. *$F : \mathscr{D}(F) \subset U \to V$ is compact if for every bounded sequence (x_k) in $\mathscr{D}(F)$ $(F(x_k))$ has a convergent subsequence.*

In particular the composition of a compact linear operator and a sequentially continuous non-linear operator is compact.

Theorem 26. *Let $\varepsilon \geq 0$. Then $F : \mathscr{D}(F) \subset U \to V$ is continuous and compact. Moreover, F is sequentially weakly continuous and weakly closed.*

Proof. The proof follows from [EE05, Theorem 2.1] or [Cré03a, Proposition 4.4 and 5.1], where it is proven that $F : \mathscr{D}(F) \subset U \to W_p^{1,2}(\Omega)$ satisfies the property for all $2 \leq p < \bar{p}$ with an appropriate $\bar{p} > 2$. The result then follows by using that the embedding from $W_p^{1,2}(\Omega)$ into $L^2(\Omega)$ is bounded. □

The compactness and weak closedness of the operator F, concluded in Theorem 26, imply the local ill-posedness of the inverse problem of identification of the local volatility surface $\sigma(T,K)$. In fact, for any U-bounded sequence $\{a_n\}_{n\in\mathbb{N}}$ in $\mathscr{D}(F)$, that has no strong convergent subsequences, we can extract an U-weakly-convergent subsequence, say $\{a_{n_k}\}_{k\in\mathbb{N}}$. Since $\mathscr{D}(F)$ is weakly closed with respect to the H^1-norm, the weak limit of $\{a_{n_k}\}_{k\in\mathbb{N}}$ belongs to $\mathscr{D}(F)$. Thus, since F is compact, $\{F(a_{n_k})\}$ has a convergent subsequence. So, similar option prices may correspond to completely different volatilities.

Remark 27. *Theorem 26 and the continuity of the embedding of $W_p^{1,2}(\Omega)$ in $L^2(\Omega)$ ensures that Item 4 of Assumption 4 holds. Therefore, Theorem 5 is applicable for the functional $\mathscr{F}_{\beta,u^\delta}$ defined in* (12).

A.2 Attainment of source conditions

The convergence result of Theorem 7 is directly connected to the existence of a source function w that satisfies the source condition (16).

Theorem 28. *Let $\varepsilon > 0$. Assume that $\hat{a} \in \mathscr{D}(F) \subset U$ is a minimizer of (12) with u^δ substituted by $\tilde{u}$. Then, there exists $\tilde{w} := \lambda(\tilde{u} - F(\hat{a}))$ such that*

$$\zeta = \lambda F'(\hat{a})^* \tilde{w} \in \partial f(\hat{a})$$

In particular, if $\hat{a} = a^\dagger$, then (16) holds.

For a proof, see [CSZ10].

It turns out that, for the specific problem under consideration, we are not able to characterize the source condition (16). However, we can guarantee (15). See section A.3. The first step in order to guarantee (15) is the following simple Lemma, whose proof can be found in [CSZ10].

Lemma 29. *Let $\zeta^\dagger \in \partial f(a^\dagger)$. Then, there exists a function $w^\dagger \in V$ and a function $r \in U$ such that*

$$\zeta^\dagger = F'(a^\dagger)^* w^\dagger + r \tag{35}$$

holds. Furthermore, $\|r\|_U$ can to be taken arbitrarily small.

A.3 Convergence rates

In this subsection we exhibit a class of functionals such that we are able to prove that condition (15) holds provided the variational source condition (35) is satisfied. For that we shall make use of the following concept:

Definition 30. *Let $1 \le q < \infty$ and $\tilde{U}$ be a subset of U. The Bregman distance $D_\zeta(\cdot, \tilde{a})$ of $f : U \to \mathbb{R} \cup \{+\infty\}$ at $\tilde{a} \in \mathscr{D}_B(f)$ and $\zeta \in \partial f$ is said to be q-coercive with constant $\underline{c} > 0$ if*

$$D_\zeta(a, \tilde{a}) \ge \underline{c} \|a - \tilde{a}\|_{\tilde{U}}^q, \quad \forall a \in \mathscr{D}(f). \tag{36}$$

In the next lemma we prove that the existence of an approximated source condition as (35) and f satisfying Definition 30 is sufficient for convergence rates:

Lemma 31. *Let $\zeta^\dagger \in \partial f(a^\dagger)$ satisfy (35) with $w^\dagger$ and r such that*

$$\underline{c}\big(C\|w^\dagger\|_V + \|r\|_{L^2(\Omega)}\big) := \beta_1 \in [0, 1),$$

and the Bregman distance with respect to f is $1-$coercive with $\tilde{U} := U$. Then, equation (15) holds. In particular, the convergence rates of Theorem 7 hold.

Proof. Using the continuously Sobolev embedding theorem [Ada75], Equations (35) and the estimate

$$\|u' \cdot h\|_{W_p^{1,2}(\Omega)} \le C\|h\|_U\,. \tag{37}$$

we have that

$$\begin{aligned} |\langle \zeta^\dagger, a - a^\dagger \rangle| &\le |\langle \zeta^\dagger - r, a - a^\dagger \rangle + \langle r, a - a^\dagger \rangle| \\ &\le |\langle w^\dagger, F'(a^\dagger)(a - a^\dagger) \rangle| + \|r\|_U \, \|a - a^\dagger\|_U \\ &\le \|w^\dagger\|_V \, \|F'(a^\dagger)(a - a^\dagger)\|_V + \|r\|_U \, \|a - a^\dagger\|_U \\ &\le \big(C\|w^\dagger\|_V + \|r\|_U\big) \|a - a^\dagger\|_U\,. \end{aligned}$$

From the assumption that f satisfies Definition 30 and the definition of β_1 we have

$$\begin{aligned}|\langle \zeta^\dagger, a-a^\dagger\rangle| &\le \left(C\|w^\dagger\|_V + \|r\|_U\right)\|a-a^\dagger\|_U \\ &\le \beta_1 D_{\zeta^\dagger}(a,a^\dagger) \le \beta_1 D_{\zeta^\dagger}(a,a^\dagger) + \beta_2 \left\|F(a)-F(a^\dagger)\right\|_V .\end{aligned}$$

The convergence rates now follow from Theorem 7. □

Under the assumption of Lemma 31 if in addition f is q-coercive a convergence rate in the norm holds:

$$\left\|a_\beta^\delta - a^\dagger\right\|_U \le D_{\zeta^\dagger}(a_\beta^\delta, a^\dagger) = \mathscr{O}(\delta)\,. \tag{38}$$

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

Example 32 (q-coercive Bregman distance)**.** *Let $\tilde{U}$ be a Hilbert space and $\mathscr{D}(f) \subset \tilde{U}$ and $f(a) := q^{-1}\left\|a-a^\dagger\right\|_{\tilde{U}}^q$. Then, the Bregman distance associated to f is q-coercive. See [HH09] and references in there.*

Example 33. *Let $1 < q \le 2$ and $\varepsilon > 0$. We consider the functional*

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

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

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

Hence, f is 2-coercive. Therefore, according to Lemma 31 and equation (38) the rate of $\mathscr{O}(\delta)$ holds for the $H^{1+\varepsilon}$-norm.