



On Convex Envelopes and Regularization of Non-convex Functionals Without Moving Global Minima

Marcus Carlsson¹

Received: 28 May 2018 / Accepted: 9 May 2019 / Published online: 29 May 2019
© The Author(s) 2019

Abstract

We provide theory for the computation of convex envelopes of non-convex functionals including an ℓ^2 -term and use these to suggest a method for regularizing a more general set of problems. The applications are particularly aimed at compressed sensing and low-rank recovery problems, but the theory relies on results which potentially could be useful also for other types of non-convex problems. For optimization problems where the ℓ^2 -term contains a singular matrix, we prove that the regularizations never move the global minima. This result in turn relies on a theorem concerning the structure of convex envelopes, which is interesting in its own right. It says that at any point where the convex envelope does not touch the non-convex functional, we necessarily have a direction in which the convex envelope is affine.

Keywords Fenchel conjugate · Convex envelope · Regularization · Non-convex/non-smooth optimization · Proximal hull

Mathematics Subject Classification 49M20 · 65K10 · 90C26

1 Introduction

The present work is the extension of a chain of ideas with its roots in compressed sensing. $\ell^1 - \ell^2$ -minimization tricks have a long history and got renewed attention with the work of Donoho, Candés and Tao among others [1–3]. In the same spirit, the nuclear norm minimization strategy was investigated by Fazel and coworkers [4,5], and in both cases, it was shown that these methods yield perfect reconstructions in the case of no noise. However, in realistic scenarios, these results often do not apply and additionally there is of course noise, in which case the methods come with a (sometimes

Communicated by Jérôme Bolte.

✉ Marcus Carlsson
marcus.carlsson@math.lu.se

¹ Centre for Mathematical Sciences, University of Lund, Lund, Sweden

severe) bias. Moreover, they are slow since one needs to find an appropriate value of involved penalty parameters.

Due to such issues, there is a wealth of non-convex variations to replace ℓ^1 /nuclear norm in the area of compressed sensing, we refer to [6] for a survey. Two fairly recent contributions in this vein are the work by Carl Olsson and coworkers [7] as well as by Gilles Aubert and coworkers [8]. The former paper deals with non-convex matrix minimization problems with subspace constraints, the latter with sparse reconstructions, and in particular, the latter shows that the concrete regularizer considered there has the desirable property of not moving global minima. In this paper, we find a unifying framework and show that all these penalties are particular cases of the so-called proximal hull or quadratic envelope. We systematically study this as a regularizer, and in particular, we lift the result of Aubert et al. to a general context. In order to do so, we provide new results on the structure of lower semi-continuous (abbreviated l.s.c.) convex envelopes which are interesting in their own right. More precisely, we show that whenever a l.s.c. convex envelope is not in touch with the function that generates it, then it necessarily has a direction in which it is affine linear.

2 Outline and Motivation

We develop methods to compute the lower semi-continuous convex envelope of functionals of the form

$$f(x) + \frac{1}{2} \|x - d\|_2^2, \quad (1)$$

and show that this is of the form $\mathcal{Q}(f)(x) + \frac{1}{2} \|x - d\|_2^2$, where $\mathcal{Q}(f)$ is the proximal hull or “quadratic envelope”, as we shall call it. Here, x can be in any separable Hilbert space, but f needs to be such that the global minimization of (1) is computable. The practical applications of $\mathcal{Q}(f)$ pertain to optimization of (1) with additional constraints, as well as unconstrained optimization of

$$f(x) + \frac{1}{2} \|Ax - d\|_2^2, \quad (2)$$

where A is a linear operator.

To introduce the main ideas behind this work, we consider two concrete problems. A multitude of applications can be posed mathematically as finding the lowest rank matrix X satisfying some equation $A(X) = d$, where A is a linear operator and d is a measurement (see, e.g., [5,9]). Usually, the measurement d is not perfect, so in practice, one wishes to find the minimum rank given some accepted error: $\|A(X) - d\| \leq \rho$. The dual formulation of this problem is

$$\arg \min_X \left[\lambda \text{rank}(X) + \|A(X) - d\|^2 \right], \quad (3)$$

where λ is a parameter. However, the functional $\text{rank}(X)$ is non-convex and highly discontinuous, so the problem can not be solved as stated (in general). It can be solved

for the case $A = I$, but the problem is still hard when combined with additional priors, see, e.g., Section 1.1 in [7] for an overview and applications in signal processing and imaging.

Due to the problematic nature of $\text{rank}(X)$, it has become popular to replace $\text{rank}(X)$ with the nuclear norm of X . However, $\text{rank}(X)$ and the nuclear norm are quite far apart and the method leads to a bias in the solution, which led the authors of [7] to suggest working instead with the convex envelope of $\text{rank}(X) + \frac{1}{2}\|X - D\|_F^2$ for which they obtained an explicit expression. They also provided the convex envelope when $\text{rank}(X)$ is replaced by the indicator functional of the set $\{X : \text{rank}(X) \leq K\}$, in order to treat problems where a matrix of a fixed rank is sought, and this convex envelope was further studied in [10].

Independently, convex envelopes were used in [8] to suggest a regularizer to functionals of the type

$$\|x\|_0 + \frac{1}{2}\|Ax - d\|_2^2, \quad x \in \mathbb{R}^n, \quad (4)$$

which is usually dealt with by replacing $\|x\|_0$ by $\lambda\|x\|_1$. The main contribution of their work is to show that their regularizer does not move global minima. A common misconception is that the same holds for ℓ^1 -methods, which is true only if there is no noise [11]. In the presence of noise, the estimates for ℓ^1 -methods are rather poor and [8] are the first framework which allows for regularization without moving minima in a more realistic scenario.

This paper presents a unified approach to this circle of ideas by connecting them with the “quadratic envelope” $\mathcal{Q}(f)$. We also extend the findings of [8] to any problem of the form (2) as long as $\mathcal{Q}(f)$ is computable. An expanded version of this article is found in [12], which contains a long list of instances where $\mathcal{Q}(f)$ is computable.

In particular, $\mathcal{Q}(f)$ is computable for ι_K ; the indicator functional for

$$\{x \in \mathbb{C}^n : \|x\|_0 \leq K\}.$$

As a proof of concept, we compare performance of (4) with $\|x\|_0$ replaced by $\lambda\|x\|_1$, with $\mathcal{Q}(\text{card})$ and with $\mathcal{Q}(\iota_K)$ (where $\text{card}(x) = \|x\|_0$). We use a 100×200 matrix A and minimize the regularized version of (4) for d of the form $Ax_0 + \epsilon$, where x_0 has cardinality 10 and ϵ takes on various levels of noise. As noted in [13], the best one can hope for is then to recover the so-called oracle solution x_S , (obtained if an oracle a priori revealed the correct support). As Fig. 1 shows, both $\mathcal{Q}(\text{card})$ and $\mathcal{Q}(\iota_{10})$ outperform ℓ^1 and find the oracle solution for fairly large levels of noise. Also, $\mathcal{Q}(\iota_{10})$ beats $\mathcal{Q}(\text{card})$, which is no surprise since it contains additional information about the problem built into it and demonstrates the versatility of the new \mathcal{Q} -transform. The article [6] contains much more information about this particular case.

We now outline the main contributions of this paper in greater detail. Consider any functional of the form

$$f(x) + \frac{\gamma}{2}\|x - d\|_{\mathcal{V}}^2 \quad (5)$$

where $\gamma > 0$ is a parameter, \mathcal{V} is an arbitrary separable Hilbert space and f a non-negative functional on \mathcal{V} . In Sect. 3, we introduce the transform \mathcal{Q}_{γ} and show that the l.s.c. convex envelope of the functional in (5) is

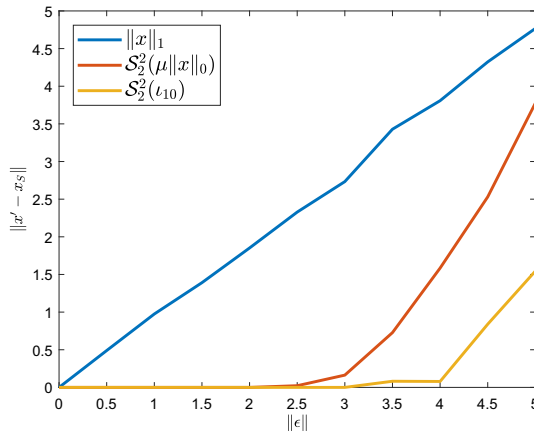


Fig. 1 Regularizing (17) by $\mathcal{Q}(\text{card})$ and $\mathcal{Q}(t_{10})$ finds the oracle solution up to noise levels of around $\|\epsilon\| = 3$ and 4 (roughly 30% of $\|d\|$), whereas ℓ^1 -regularization only finds this solution with no noise

$$\mathcal{Q}_\gamma(f)(x) + \frac{\gamma}{2} \|x - d\|_{\mathcal{V}}^2. \tag{6}$$

In order for $\mathcal{Q}_\gamma(f)$ to be computable, the global minimization of (5) needs to be solvable, and hence, the problem of minimizing (5) in itself is not an instance where the \mathcal{Q}_γ -transform is useful. However, it is useful for finding global minimizers of (5) in combination with additional prior restrictions. To illustrate, consider the problem

$$\arg \min_{x \in \mathcal{H}} \left[f(x) + \frac{1}{2} \|x - d\|^2 \right], \tag{7}$$

where \mathcal{H} is a closed and convex subset of \mathcal{V} , and suppose we are unable to find a closed form solution. Upon replacing (7) with

$$\arg \min_{x \in \mathcal{H}} \left[\mathcal{Q}_\gamma(f)(x) + \frac{1}{2} \|x - d\|^2 \right], \tag{8}$$

for some fixed $\gamma \leq 1$, we obtain a convex problem which can be solved. However, even for $\gamma = 1$, it is possible that (7) and (8) have different solutions, despite the functional in (8) being the l.s.c. convex envelope of the one in (7). The rationale behind replacing (7) by (8) is pragmatical; since the latter is convex, the solution may be found using convex optimization routines. This may seem ad hoc, but we remind the reader that replacing, e.g., $\|x\|_0$ by $\|x\|_1$ or $\text{rank}(X)$ by the nuclear norm $\|X\|_1$ has had a substantial impact, and that for these concrete cases the modification $\mathcal{Q}_\gamma(f)$ is much closer to the original functional f , (which leads to a better performance as an estimator, see the numerical sections of [6,7]). A reason for this is that $\mathcal{Q}_\gamma(f)$ has the desirable feature that $\mathcal{Q}_\gamma(f)(x) = f(x)$ often holds, and since (8) is a convex problem below the original problem (7), it is easy to see that a minimum \hat{x} to (8) is the solution to (7) whenever $\mathcal{Q}_\gamma(f)(\hat{x}) = f(\hat{x})$. This is highlighted in Fig. 2 where the

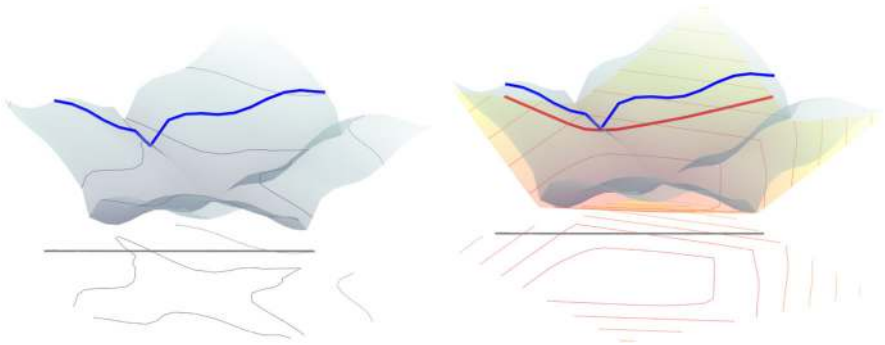


Fig. 2 Illustration of a non-convex optimization problem with linear constraints. The left panel shows a non-convex functional along with its level sets. The gray line represents the subspace we are interested in, and the blue curve represents the values of the functional restricted to the subspace. The right panel shows the same setup, but here the convex envelope is shown as well in orange/yellow. The values of the convex envelope over the subspace are shown in the red curve. In this case, the minima of blue and red function coincide

two problems have the same solution. More information and examples on this type of problems are found in Part II of [12].

In Sect. 5, we consider regularization of functionals like (3) and (4), or more generally

$$f(x) + \frac{1}{2} \|Ax - d\|_{\mathcal{W}}^2 \quad (9)$$

for arbitrary non-negative f , where $A : \mathcal{V} \rightarrow \mathcal{W}$ is a linear operator between separable Hilbert spaces. We assume that \mathcal{V} is such that $\mathcal{Q}_\gamma(f)$ is computable and that the convex envelope of (9) is untractable. We propose to use as regularizer the function $\mathcal{Q}_\gamma(f)$, i.e., we will study the relationship between minimizers of (9) and those of

$$\mathcal{Q}_\gamma(f)(x) + \frac{1}{2} \|Ax - d\|_{\mathcal{W}}^2. \quad (10)$$

Since it often holds that $\mathcal{Q}_\gamma(f)(x) = f(x)$, we again see that a global minimizer of (10) for which this is the case must also be a global minimizer of (9), in view of the inequality $\mathcal{Q}_\gamma(f) \leq f$ (shown in Sect. 3). The parameter γ now becomes a useful tool as it tunes the curvature of $\mathcal{Q}_\gamma(f)$, and we pause to illustrate its role by considering a toy problem in one variable; see Fig. 3. We let $|x|_0$ be the function equalling 1 on $\mathbb{R} \setminus \{0\}$ and zero at $x = 0$. In red, we see the functional $|x|_0 + \frac{1}{2}|x - 1|^2$ (which is a particular case of both (3) and (4) in one dimension, the matrix A is here the number 1), in blue its convex envelope and in pink the ℓ^1 convex relaxation $|x| + \frac{1}{2}|x - 1|^2$. Clearly, the global minimum of the red and blue coincides, but the global minimum of the ℓ^1 -relaxation is different. For (10), we have two options, either $|A|^2 > \gamma$ or $|A|^2 < \gamma$. The regularizer (10) is illustrated in black for these two cases in Fig. 3. The circles represent global minima of the respective functions. In the case $|A|^2 > \gamma$, we see that (10) is a *convex* minorant of (9) whose global minima (for this choice of parameters) is equal to that of (9). In the case $|A|^2 < \gamma$, (10) is no longer convex but the local minima of (10) are also minima of (9), and (10) has fewer local minima. In

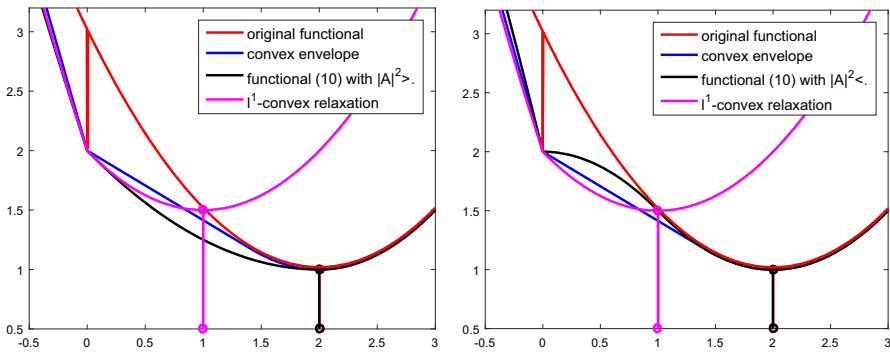


Fig. 3 The black curve shows two regularizations of the red curve, for different levels of γ

particular, the global minima coincide. The main point of the paper is loosely that the general behavior is the same.

In Sect. 5.1, we generalize the situation in Fig. 3 (left) and assume that γ satisfies $A^*A \succcurlyeq \gamma I$, i.e., that

$$\|Ax\|^2 \geq \gamma \|x\|^2. \tag{11}$$

For such choice of γ , we prove that the functional (10) is a convex functional below (9) and hence minimization of (10) will produce a minimizer which, although not necessarily equal to the minimizer of the original problem, potentially is closer than that obtained by other convex relaxation methods.

For the problem (4), A is usually a matrix with a large kernel which rules out the above approach. In Sect. 5.2, we consider the case

$$\|A\|^2 \leq \gamma, \tag{12}$$

generalizing the situation in the right picture of Fig. 3. We can then show that (10) is a continuous (but not convex) functional with the following desirable properties:

- (i) Equation (10) lies between (9) and its l.s.c. convex envelope,
- (ii) any local minimizer of (10) is a local minimizer of (9),
- (iii) the global minimizers of (10) and (9) coincide.

These findings in turn rely on general results about l.s.c. convex envelopes which we provide in Sect. 4. The computation of the l.s.c. convex envelope of $f(x) + \frac{\gamma}{2} \|x\|^2$ can be thought of as stretching a plastic foil from below onto the graph of $f(x) + \frac{\gamma}{2} \|x\|^2$ (see Fig. 2). Consider a point x where the plastic foil is not in contact with the graph, i.e., where $\mathcal{Q}_\gamma(f)(x) < f(x)$. It is intuitively obvious that the plastic foil, i.e., the graph of $\mathcal{Q}_\gamma(f)(x) + \frac{\gamma}{2} \|x\|^2$ has some direction in which it is affine linear and thus $\mathcal{Q}_\gamma(f)$ should have some direction in which the curvature is $-\gamma$. This is surprisingly difficult to show and despite the wealth of results on l.s.c. convex envelopes it is not found in any standard reference on the topic. The statement is shown in the Ph.D. thesis [14] for the finite- dimensional case. Here, we provide a proof based on an extension of Milman’s theorem due to Arne Brøndsted [15] in a short note from 1966.

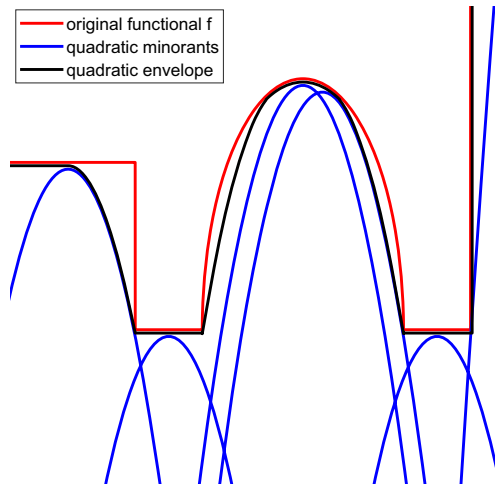


Fig. 4 Illustration of a non-convex function f (red) and its quadratic envelope $Q_2(f)$ (black). The black graph lies slightly below for illustration only

The final Sect. 6 is more practical in nature. Critical points of (10) can be found using the forward–backward splitting method (FBS), given that $Q_\gamma(f)$ is “semi-algebraic”, as was shown in [16]. To simplify verification of when $Q_\gamma(f)$ is semi-algebraic, we show in Sect. 6 that this is true as long as f itself is semi-algebraic. Further tools to compute $Q_\gamma(f)$ as well as related proximal operators are found in [12].

3 The Quadratic Envelope

Let \mathcal{V} be a separable Hilbert space over \mathbb{R} or \mathbb{C} , such as \mathbb{C}^n or the set matrices of a certain size equipped with the Frobenius norm. All Hilbert spaces over \mathbb{C} are also Hilbert spaces over \mathbb{R} with the scalar product $\langle x, y \rangle_{\mathbb{R}} = \operatorname{Re} \langle x, y \rangle$, and hence, it is no restriction to assume that \mathcal{V} is a real Hilbert space wherever needed. Even if \mathcal{V} is a Hilbert space over \mathbb{C} , we will implicitly assume that the scalar product is $\langle x, y \rangle_{\mathbb{R}}$.

Given any functional $f : \mathcal{V} \rightarrow \mathbb{R} \cup \{\infty\}$ and parameter $\gamma > 0$, we introduce the “quadratic envelope” Q_γ as the supremum of all minimizers of the form $\alpha - \frac{\gamma}{2} \|x - y\|^2$ for $\alpha \in \mathbb{R}$ and $y \in \mathcal{V}$;

$$Q_\gamma(f)(x) = \sup_{\alpha \in \mathbb{R}, y \in \mathcal{V}} \left\{ \alpha - \frac{\gamma}{2} \|x - y\|^2 : \alpha - \frac{\gamma}{2} \|\cdot - y\|^2 \leq f \right\}. \quad (13)$$

The quadratic envelope has appeared previously, e.g., in [17] under the name “proximal hull”, denoted $h_{\gamma^{-1}}$ (Example 1.44), but it seems that the term is not widespread (see the discussion in Sect. 7), and it seems that its connection with convex envelopes has not been noted, or at least not systematically studied. We prefer the term quadratic envelope since it is more suggestive and prefer the notation Q_γ since it would be messy to always have to invert γ which in this context has a concrete meaning; the parameter γ basically tunes the maximum negative curvature of $Q_\gamma(f)$, as we shall

see in Sect. 4 (Corollary 4.2). When $\gamma = 1$, we simply write \mathcal{Q} as opposed to \mathcal{Q}_γ . In this section, we first provide some tools to compute \mathcal{Q}_γ and then prove the connection with l.s.c. convex envelopes and end with some auxiliary results.

The Legendre transform (or Fenchel conjugate) is defined as $g^*(y) := \sup_x \langle x, y \rangle - g(x)$. We remind the reader that g^* is l.s.c convex and that g^{**} equals the l.s.c. convex envelope of g by the Fenchel–Moreau theorem (see, e.g. Proposition 13.11 and 13.39 in [18]). We now introduce the transform \mathcal{S}_γ defined as follows:

$$\mathcal{S}_\gamma(f)(y) := \left(f(\cdot) + \frac{\gamma}{2} \|\cdot\|^2 \right)^* (\gamma y) - \frac{\gamma}{2} \|y\|^2 = \sup_x -f(x) - \frac{\gamma}{2} \|x - y\|^2. \tag{14}$$

\mathcal{S}_γ is simply the negative of the Moreau envelope computed with constant γ^{-1} . If we set $q_\gamma(x, y) = -\frac{\gamma}{2} \|x - y\|^2$ then, in the terminology of [17] Sec. 11.L, $\mathcal{S}_\gamma(f)$ is the q_γ -conjugate of f and $\mathcal{Q}_\gamma(f)$ the q_γ -envelope of f (reinforcing our choice of terminology “quadratic envelope” for \mathcal{Q}_γ). We introduce the symbol \mathcal{S}_γ mainly since we believe the notation $-e_{\gamma^{-1}}(f)$ or ${}^{q_\gamma}f$ (c.f. [17]) or $-\gamma^{-1}f$ (c.f. [18]) would be confusing for our present purposes. Its connection to the quadratic envelope is described by the following proposition (Fig. 4);

Proposition 3.1 *Let $\gamma > 0$ and let f be a $[0, \infty]$ -valued l.s.c. functional on a separable Hilbert space \mathcal{V} . We have $\mathcal{Q}_\gamma = \mathcal{S}_\gamma \circ \mathcal{S}_\gamma := \mathcal{S}_\gamma^2$, i.e.,*

$$\mathcal{Q}_\gamma(f)(x) = \sup_y \left(\inf_w f(w) + \frac{\gamma}{2} \|w - y\|^2 \right) - \frac{\gamma}{2} \|x - y\|^2 \tag{15}$$

Proof The argument is a replica of Example 1.44 of [17] but is included for completeness. We have $\alpha - \frac{\gamma}{2} \|\cdot - y\|^2 \leq f$ iff $\alpha \leq f + \frac{\gamma}{2} \|\cdot - y\|^2$, so the maximal α for fixed y is given by $\alpha = -\mathcal{S}_\gamma(f)(y)$. Thus, $\mathcal{Q}_\gamma(f)(x) = \sup_{y \in \mathcal{V}} -\mathcal{S}_\gamma(f)(y) - \frac{\gamma}{2} \|x - y\|^2$ as desired. \square

The next proposition contains some basic observations about the behavior of \mathcal{S}_γ and \mathcal{Q}_γ .

Proposition 3.2 *Let $\gamma > 0$ and let f be a $[0, \infty]$ -valued l.s.c. functional on a separable Hilbert space \mathcal{V} . Then, $\mathcal{S}_\gamma(f)$ takes values in $]-\infty, 0]$ and is continuous, whereas $\mathcal{Q}_\gamma(f)$ is lower semi-continuous, takes values in $[0, \infty]$ and is continuous in the interior of $\text{dom}(\mathcal{Q}_\gamma(f))$.*

Proof The statement of the interchanging signs follows easily by the last line of (14) which also shows that $\mathcal{S}_\gamma(f)$ avoids $-\infty$. By (14), it follows that $\mathcal{S}_\gamma(f)$ (and $\mathcal{Q}_\gamma(f)$) by Proposition 3.1) is the difference of an l.s.c. convex functional and a quadratic term. With this in mind, the continuity statements follow by the standard properties of l.s.c. convex functionals (see, e.g., Corollary 8.30 [18]). \square

The following result is the key result of this section connecting the \mathcal{Q}_γ -transform with l.s.c. convex envelopes.

Theorem 3.1 Let $\gamma > 0$ and let f be a $[0, \infty]$ -valued functional on a separable Hilbert space \mathcal{V} . Then, $(f + \frac{\gamma}{2} \|\cdot - d\|^2)^*(y) = \mathcal{S}_\gamma(f) \left(\frac{y}{\gamma} + d \right) + \frac{\gamma}{2} \left\| \frac{y}{\gamma} + d \right\|^2 - \frac{\gamma}{2} \|d\|^2$ and

$$\left(f + \frac{\gamma}{2} \|\cdot - d\|^2 \right)^{**}(x) = \mathcal{Q}_\gamma(f)(x) + \frac{\gamma}{2} \|x - d\|^2.$$

In particular, $\mathcal{Q}_\gamma(f)(x) + \frac{\gamma}{2} \|x - d\|^2$ is the l.s.c. convex envelope of $f(x) + \frac{\gamma}{2} \|x - d\|^2$ and

$$0 \leq \mathcal{Q}_\gamma(f) \leq f.$$

Proof We have

$$\begin{aligned} \left(f(\cdot) + \frac{\gamma}{2} \|\cdot - d\|^2 \right)^*(y) &= \sup_x \langle x, y \rangle - f(x) - \frac{\gamma}{2} \|x - d\|^2 \\ &= \sup_x -f(x) - \frac{\gamma}{2} \left\| x - \left(\frac{y}{\gamma} + d \right) \right\|^2 \\ &\quad + \frac{\gamma}{2} \left\| \frac{y}{\gamma} + d \right\|^2 - \frac{\gamma}{2} \|d\|^2 \end{aligned}$$

from which the first identity follows. Similarly,

$$\begin{aligned} \left(f(\cdot) + \frac{\gamma}{2} \|\cdot - d\|^2 \right)^{**}(x) &= \left(\mathcal{S}_\gamma(f) \left(\frac{\cdot}{\gamma} + d \right) + \frac{\gamma}{2} \left\| \frac{\cdot}{\gamma} + d \right\|^2 - \frac{\gamma}{2} \|d\|^2 \right)^*(x) \\ &= \sup_y \langle x, y \rangle - \mathcal{S}_\gamma(f) \left(\frac{y}{\gamma} + d \right) - \frac{\gamma}{2} \left\| \frac{y}{\gamma} + d \right\|^2 + \frac{\gamma}{2} \|d\|^2 \\ &= \sup_y -\mathcal{S}_\gamma(f) \left(\frac{y}{\gamma} + d \right) - \frac{\gamma}{2} \left\| \frac{y}{\gamma} + d - x \right\|^2 \\ &\quad + \frac{\gamma}{2} \|x - d\|^2 = \mathcal{S}_\gamma^2(f)(x) + \frac{\gamma}{2} \|x - d\|^2. \end{aligned}$$

The statement about the convex envelope follows by the Fenchel–Moreau theorem and Proposition 3.1, which also gives $\mathcal{Q}_\gamma(f)(x) + \frac{\gamma}{2} \|x - d\|^2 \leq f(x) + \frac{\gamma}{2} \|x - d\|^2$. This implies the latter part of the inequality $0 \leq \mathcal{Q}_\gamma(f) \leq f$, whereas the former has already been noticed in Proposition 3.2. \square

We end this section with some observations about the behavior of $\mathcal{Q}_\gamma(f)$ as a function of γ .

Proposition 3.3 Let f be an l.s.c. $[0, \infty]$ -valued functional. Then, $\mathcal{Q}_\gamma(f)(x)$ is increasing as a function of γ . Moreover,

$$\lim_{\gamma \rightarrow \infty} \mathcal{Q}_\gamma(f)(x) = f(x) \tag{16}$$

whereas the limit as $\gamma \searrow 0$ equals a convex minimizer of f above the l.s.c. convex envelope of f .

We remark that (16) is shown in [17], whereas nothing is said about the case $\gamma \searrow 0$. In fact, $\lim_{\gamma \searrow 0} Q_\gamma(f)$ usually equals the l.s.c. convex envelope of f , but this is not necessarily the case in general, which is a surprise at least for the author. To see this, consider $P = \{x \in \mathbb{R}^2 : x_1 > 0, x_2 = \sqrt{x_1}\}$, $Q = \{x \in \mathbb{R}^2 : x_1 > 0, 0 < x_2 \leq \sqrt{x_1}\} \cup \{0\}$ and $f = \iota_P$, where ι_P is the indicator functional of P . It is easy to see that the l.s.c. convex envelope of ι_P equals $\iota_{cl(Q)}$ (where cl denotes closure), whereas some thinking reveals that $\lim_{\gamma \searrow 0} Q_\gamma(f) = \iota_Q$. However, if \mathcal{V} is finite dimensional and $\lim_{\gamma \searrow 0} Q_\gamma(f)$ is everywhere finite, then it is automatically continuous (Corollary 8.30 in [18]), and hence it must equal the l.s.c. convex envelope of f .

Proof If $\gamma_1 > \gamma_2$, then $Q_{\gamma_2}(f)(x) + \frac{\gamma_1}{2}\|x\|^2$ equals the l.s.c. convex functional $Q_{\gamma_2}(f)(x) + \frac{\gamma_2}{2}\|x\|^2$ plus the term $\frac{\gamma_1 - \gamma_2}{2}\|x\|^2$ so it is l.s.c. and convex. In view of $Q_{\gamma_2}(f) \leq f$, it also lies below $f + \frac{\gamma_1}{2}\|x\|^2$, and so, we conclude that

$$Q_{\gamma_2}(f)(x) + \frac{\gamma_1}{2}\|x\|^2 \leq (f + \frac{\gamma_1}{2}\|x\|^2)^{**} = Q_{\gamma_1}(f)(x) + \frac{\gamma_1}{2}\|x\|^2.$$

The first claim follows. To see (16), let $\alpha < f(x)$ be arbitrary. Since f is l.s.c., the set $\{y : f(y) > \alpha\}$ is open and, as $f \geq 0$, it follows that for any γ large enough we have $\alpha - \frac{\gamma}{2}\| \cdot -x \|^2 \leq f$. For such γ , we thus have $\alpha \leq Q_\gamma(f)(x) \leq f(x)$ by (13) and Theorem 3.1, so (16) follows.

Concerning the limit as $\gamma \searrow 0$ set $g(x) = \lim_{\gamma \searrow 0} Q_\gamma(f)(x)$, which exist by the first part of this proposition. Now note that

$$g(x) = \lim_{\gamma \searrow 0} Q_\gamma(f)(x) = \lim_{\gamma \searrow 0} Q_\gamma(f)(x) + \frac{\gamma}{2}\|x\|^2 = \lim_{\gamma \searrow 0} (f + \frac{\gamma}{2}\| \cdot \|^2)^{**}(x) \geq f^{**}.$$

We also see that g is the limit of a decreasing sequence of convex functions, hence it is also convex (Proposition 8.16 [18]). Finally, $g \leq f$ by Theorem 3.1. □

4 Finer Properties of Convex and Quadratic Envelopes

In this section, we prove a result about the structure of l.s.c. convex envelopes which seems relatively unknown. For this, we need the concept of weak lower semi-continuity, which is nothing but semi-continuity with respect to the weak topology of the underlying separable Hilbert space \mathcal{V} . We remind the reader that for convex proper functionals there is no difference (Theorem 9.1 [18]) between weakly l.s.c. functionals and the standard l.s.c. functionals. Also, if \mathcal{V} is finite dimensional and the topology is Hausdorff, the two topologies are the same so there is no difference in this case either. However, we wish to underline that the difficulty in proving the coming results is present also in the finite-dimensional setting.

We begin with a neat fact concerning weakly l.s.c. convex envelopes, which does not seem to have made its way into the modern literature on the subject. It is a reformulation of Arne Brøndsted's extension of Milman's theorem [15]. To state it, we remind the reader that a functional g is *coercive* if and only if its (lower) level sets are bounded, (see, e.g., Proposition 11.11 [18]). Note that l.s.c. convex envelopes of the type $\mathcal{Q}_\gamma(f)(x) + \frac{\gamma}{2}\|x - d\|^2$ (for non-negative f) always are coercive, by virtue of Proposition 3.2 and the quadratic term. A function f on \mathbb{R} is called *affine* if it is of the form $f(t) = at + b$ with $a, b \in \mathbb{R}$.

Theorem 4.1 *Let g be a weakly l.s.c. functional on a separable Hilbert space \mathcal{V} such that g^{**} is coercive. Given any $x \in \mathcal{V}$ such that $g(x) \neq g^{**}(x)$ there exists a unit vector v and $t_0 > 0$ such that the function $h(t) = g^{**}(x_0 + tv)$ is affine on $] - t_0, t_0[$.*

To prove Theorem 4.1, we recall some concepts from [15]. Given a convex function f , a point x is called *extremal* if and only if $(x, f(x))$ is extremal for the epigraph of f , denoted $[f]$. Equivalently, x is extremal if and only if $x \in \text{dom } f$ and f is not affine on any relatively open segment containing x . Moreover, f_{ext} denotes the functional which equals $f(x)$ for all extremal points x and ∞ else.

To illustrate, consider the convex functional $g : \mathbb{R}^2 \rightarrow] - \infty, \infty]$ given by $g(x, \pm 1) = -\sqrt{1 - x^2}$ for (x, y) in $[-1, 1] \times \{\pm 1\}$, and which equals ∞ elsewhere. Then, $g^{**}(x, y) = -\sqrt{1 - x^2}$ on $[-1, 1]^2$ and $g^{**} = \infty$ elsewhere. The extremal points then equal $[-1, 1] \times \{\pm 1\}$, which should not be confused with the exposed points which are $] - 1, 1[\times \{\pm 1\}$ (see, e.g., Ch. 25 of [19]). Note that on the other two edges $\{\pm 1\} \times] - 1, 1[$, the graph of g^{**} is indeed affine, but that at these points, g^{**} is not sub-differentiable. Also note that $(g^{**})_{ext} = g$ for this particular function. The result we need from [15] reads as follows.

Theorem 4.2 *Let g be a weakly l.s.c. functional on a separable Hilbert space \mathcal{V} such that g^{**} is coercive, then*

$$[(g^{**})_{ext}] \subset [g].$$

Proof In the setting of [15], we let E be the separable Hilbert space \mathcal{V} with the weak topology. Since, convex functionals are l.s.c. with respect to the weak topology if and only if they are with respect to the norm topology, it follows that the l.s.c. convex envelope of g equals the weakly l.s.c. convex envelope. In the notation of Theorem 1 of [15], we can then take $f = g^{**}$ and the theorem states that $[f_{ext}] \subset [g_{cl}]$, where g_{cl} is the greatest l.s.c. minorant of g . Since, g is assumed to be l.s.c. we have $g = g_{cl}$, and the desired inclusion follows. It remains to check that the conditions of Theorem 1 are fulfilled, which is that “ g is inf-compact in some direction” (with respect to the weak topology, referring to the terminology of [15]). For this, it suffices to check that g^{**} is inf-compact, i.e., that all level sets are compact. The level sets of g^{**} are closed and convex, and since g^{**} is assumed coercive they are also bounded. It follows that such level sets are compact in the weak topology and the proof is complete. \square

Based on this, we can now easily prove Theorem 4.1.

Proof of Theorem 4.1 Since $g \geq g^{**}$, Theorem 4.2 clearly implies that $g(x) = g^{**}(x)$ for all extremal points x for g^{**} . Consequently, if $g(x) = g^{**}(x)$ does not hold, then x is not extremal for g^{**} , and the existence of ν follows by the definition of an extremal point for g^{**} . \square

Next, we discuss what the theorem implies about minimizers of g versus g^{**} . Denote by G the set of global minimizers of g , and by G^{**} the set of global minimizers of g^{**} .

Corollary 4.1 *Let g be a weakly l.s.c. functional on a separable Hilbert space \mathcal{V} such that g^{**} is coercive. Then, G^{**} is a closed bounded convex set containing G . Letting G_{ext}^{**} denote the extremal points of G^{**} , we also have that $G_{ext}^{**} \subset G$. Finally, the closed convex hull of G_{ext}^{**} equals G^{**} .*

Proof The convexity of G^{**} and the inclusion $G \subset G^{**}$ are immediate. The boundedness of G^{**} follows since g^{**} is coercive. Let x be in the closure of G^{**} , and let c be the value of the global minimum. Then, $g^{**}(x) \leq c$ follows by l.s.c., and the reverse inequality is obvious from the fact that c is a global minimum. It follows that $x \in G^{**}$, and hence G^{**} is closed.

The existence of points in G_{ext}^{**} , and the statement concerning the closed convex hull, are now immediate consequences of the Krein-Milman theorem and the fact that bounded closed convex sets are weakly compact in separable Hilbert spaces (Theorem 3.33, [18]). It remains to prove that $G_{ext}^{**} \subset G$. Let $x_0 \in G_{ext}^{**}$ and suppose $x_0 \notin G$. Then, Theorem 4.1 implies the existence of a direction ν on which g^{**} is constant near x_0 , contradicting that x_0 is an extremal point. \square

We end by noting that Theorem 4.1 implies that γ tunes the maximum negative curvature in the \mathcal{Q}_γ -transform, as discussed in the introduction.

Corollary 4.2 *Let f be a weakly l.s.c. $[0, \infty]$ -valued functional on a separable Hilbert space \mathcal{V} . For each $x_0 \in \mathcal{V}$ with $f(x_0) > \mathcal{Q}_\gamma(f)(x_0)$, there exists a unit vector ν such that*

$$\mathcal{Q}_\gamma(f)(x_0 + t\nu) = a + bt - \frac{\gamma}{2}t^2$$

for t near 0 and some $a, b \in \mathbb{R}$.

Proof Set $g(x) = f(x) + \frac{\gamma}{2}\|x\|^2$. By Theorem 3.1, we have $\mathcal{Q}_\gamma(f)(x) + \frac{\gamma}{2}\|x\|^2 = g^{**}(x)$, by which it is immediate that g^{**} is coercive, (since $\mathcal{Q}_\gamma(f) \geq 0$ by Proposition 3.2). It also follows that $g(x_0) > g^{**}(x_0)$, and hence, Theorem 4.1 implies that a unit vector ν exists such that $t \mapsto \mathcal{Q}_\gamma(f)(x_0 + t\nu)$ equals an affine function minus $\frac{\gamma}{2}\|(x_0 + t\nu)\|^2$ in a neighborhood of $t = 0$. \square

5 The Quadratic Envelope as a Regularizer

We now let $A : \mathcal{V} \rightarrow \mathcal{W}$ be a bounded linear operator, where \mathcal{V}, \mathcal{W} are possibly different (separable) Hilbert spaces, and consider functionals of the type

$$\mathcal{J}(x) = f(x) + \frac{1}{2} \|Ax - d\|_{\mathcal{W}}^2, \quad x \in \mathcal{V}. \quad (17)$$

Our aim is to develop strategies to deal with the general problem of minimizing (17), in the case when f is an $[0, \infty]$ -valued functional such that $\mathcal{Q}_\gamma(f)$ is computable and focus on computing (explicit) approximations of the l.s.c convex envelope of \mathcal{J} . The theory is split in two cases, either we approximate the convex envelope from below by a convex functional, or we approximate it from above with a non-convex functional having a number of desirable properties, most notably the fact that local minimizers do not change. More precisely, we will study the relationship between the original functional (17) and the modified functional

$$\mathcal{J}_\gamma(x) = \mathcal{Q}_\gamma(f)(x) + \frac{1}{2} \|Ax - d\|_{\mathcal{W}}^2, \quad x \in \mathcal{V}, \quad (18)$$

under the assumption that either $\gamma I \preceq A^*A$ or $\gamma I \succeq A^*A$ (c.f. (11)–(12) and recall Fig. 3). Note that $\gamma I \succeq A^*A$ if and only if $\gamma \geq \|A\|^2$.

5.1 Case $A^*A \succeq \gamma I$

Let f be a $[0, \infty]$ -valued functional and $A : \mathcal{V} \rightarrow \mathcal{W}$ a bounded linear operator. The main result of this section states that \mathcal{J}_γ is a convex minorant of the l.s.c. convex envelope \mathcal{J}^{**} .

Theorem 5.1 *For $\gamma > 0$ such that $A^*A \succeq \gamma I$, \mathcal{J}_γ is convex and $\mathcal{J}_\gamma \leq \mathcal{J}^{**}$. Moreover, if $A^*A \succ \gamma I$ then it is strongly convex, in which case it has a unique minimizer. Finally, a minimizer \hat{x} of \mathcal{J}_γ is a minimizer of \mathcal{J} whenever $f(\hat{x}) = \mathcal{Q}_\gamma(f)(\hat{x})$.*

Proof Upon expanding $\|Ax - d\|^2 = \|Ax\|^2 - 2\langle Ax, d \rangle + \|d\|^2$ and noting that the latter two terms are affine linear, it is easily seen that it suffices to prove the first part of the statement for $d = 0$. By Theorem 3.1, it is clear that \mathcal{J}_γ is l.s.c. and that $\mathcal{J}_\gamma \leq \mathcal{J}$, and thus $\mathcal{J}_\gamma \leq \mathcal{J}^{**}$ follows immediately upon showing that \mathcal{J}_γ is convex. Define

$$\langle x, y \rangle_{\mathcal{U}} = \langle Ax, Ay \rangle_{\mathcal{W}} - \gamma \langle x, y \rangle_{\mathcal{V}}$$

and note that this is a semi-inner product, as long as $A^*A \succeq \gamma I$. It is also an inner product if the inequality is strict. In either case, $\|x\|_{\mathcal{U}}^2 := \langle x, x \rangle_{\mathcal{U}}$ is convex. It follows that

$$\mathcal{Q}_\gamma(f)(x) + \frac{1}{2} \|Ax\|_{\mathcal{W}}^2 = \left(\mathcal{Q}_\gamma(f)(x) + \frac{\gamma}{2} \|x\|_{\mathcal{V}}^2 \right) + \frac{1}{2} \|x\|_{\mathcal{U}}^2,$$

which by Theorem 3.1 implies that \mathcal{J}_γ equals the l.s.c. convex envelope of $f(x) + \frac{\gamma}{2} \|x\|_{\mathcal{V}}^2$, plus the term $\frac{1}{2} \|x\|_{\mathcal{U}}^2$. We conclude that \mathcal{J}_γ is a convex functional, which is strongly convex when $A^*A \succ \gamma I$. In the latter case, the existence of a unique minimizer follows by Corollary 11.15 in [18], (supercoercivity of \mathcal{J}_γ is obvious by the term $\frac{1}{2} \|x\|_{\mathcal{U}}^2$). Finally, let d be fixed and let \hat{x} be a minimizer of \mathcal{J}_γ . Suppose that $f(\hat{x}) = \mathcal{Q}_\gamma(f)(\hat{x})$, and let $y \in \mathcal{V}$ be arbitrary. Then

$$\mathcal{J}(y) \geq \mathcal{J}_\gamma(y) \geq \mathcal{J}_\gamma(\hat{x}) = \mathcal{J}(\hat{x}),$$

showing that \hat{x} is a global minimizer of \mathcal{J} . □

5.2 Case $A^*A \preceq \gamma I$

Let f be a $[0, \infty]$ -valued functional, and $A : \mathcal{V} \rightarrow \mathcal{W}$ a bounded linear operator. Again, we are interested in the relationship between \mathcal{J} and \mathcal{J}_γ , defined in (17) and (18), respectively. The main result of this section is that \mathcal{J}_γ does not move minima for γ in the stated range. We begin by noting the following inequalities, the first one being reverse of the one proved in Theorem 5.1.

Proposition 5.1 *For γ such that $\|A\|^2 \leq \gamma$ we have $\mathcal{J}^{**} \leq \mathcal{J}_\gamma \leq \mathcal{J}$.*

Proof The right inequality is immediate since $\mathcal{Q}_\gamma(f) \leq f$ by Theorem 3.1. As in Theorem 5.1, we moreover see that it suffices to prove the left inequality for $d = 0$. To this end, set $h(x) = \mathcal{J}^{**}(x) - \frac{1}{2} \|Ax\|^2$. Since $\mathcal{J}^{**} \leq f + \frac{1}{2} \|Ax\|^2$, we have $h \leq f$, and moreover

$$h(x) + \frac{\gamma}{2} \|x\|^2 = \mathcal{J}^{**} + \left(\frac{\gamma}{2} \|x\|^2 - \frac{1}{2} \|Ax\|^2 \right).$$

The right hand side is convex and l.s.c. by which we conclude that

$$h(x) + \frac{\gamma}{2} \|x\|^2 \leq (f + \frac{\gamma}{2} \|\cdot\|^2)^{**}(x) = \mathcal{Q}_\gamma(f)(x) + \frac{\gamma}{2} \|x\|^2,$$

(where the last identity follows by Theorem 3.1), which gives $h(x) \leq \mathcal{Q}_\gamma(f)(x)$. In other words, $\mathcal{J}^{**}(x) \leq \mathcal{Q}_\gamma(f)(x) + \frac{1}{2} \|Ax\|^2$, which is the desired inequality (for $d = 0$). □

We now come to the main theorem of this section, inspired by Theorems 4.5 and 4.8 in [8]. We say that x is a *local minimizer* of \mathcal{J} if there exists a neighborhood U of x in \mathcal{V} such that $\mathcal{J}(y) \geq \mathcal{J}(x)$ for all $y \in U$, and we say that x is a *strict local minimizer* of \mathcal{J} if the inequality is strict for $y \neq x$.

Theorem 5.2 *Suppose that $\|A\|^2 < \gamma$. If x is a local minimizer (resp. strict local minimizer) of \mathcal{J}_γ , then it is also a local minimizer (resp. strict local minimizer) of \mathcal{J} , and $\mathcal{J}_\gamma(x) = \mathcal{J}(x)$. In addition, the global minimizers coincide.*

Proof Let x be a local minimizer of \mathcal{J}_γ . If $\mathcal{Q}_\gamma(f)(x) = f(x)$ does not hold, then Corollary 4.2 implies that there exists a unit vector v such that

$$\begin{aligned} & \frac{d^2}{dt^2} \mathcal{J}_\gamma(x + tv)(0) \\ &= \frac{d^2}{dt^2} \left(\mathcal{Q}_\gamma(f)(x + tv) + \frac{1}{2} \|A(x + tv) - d\|_{\mathcal{V}}^2 \right) (0) = \|Av\|^2 - \gamma < 0. \end{aligned} \quad (19)$$

We thus conclude that $\mathcal{Q}_\gamma(f)(x) = f(x)$ holds, which immediately gives that $\mathcal{J}_\gamma(x) = \mathcal{J}(x)$. In view of Proposition 5.1, it follows that x is a local minimizer also for \mathcal{J} . The same argument applies to strict local minimizers.

We now prove that the global minimizers coincide. Note that global minimizers of \mathcal{J} are global minimizers of \mathcal{J}_γ , in view of Proposition 5.1 and the fact that $\mathcal{J}(x) = \mathcal{J}^{**}(x)$ for all global minimizers x . From this, we also see that the global minimum of \mathcal{J} and \mathcal{J}_γ coincides, let us denote this value by c . Conversely, suppose that x is a global minimizer of \mathcal{J}_γ (i.e., $\mathcal{J}_\gamma(x) = c$). Then, it is a local minimizer of \mathcal{J} by the first part, which automatically is global for \mathcal{J} since we otherwise would have $\mathcal{J}(y) < c$ for some other value y . The proof is complete. \square

The situation when $\gamma = \|A\|^2$ is a bit more involved, so we content ourselves with the following statement concerning the global minimizers.

Theorem 5.3 *Set $\gamma = \|A\|^2$, let G be the global minimizers of \mathcal{J} , and G_γ the global minimizers of \mathcal{J}_γ . Then, $G \subset G_\gamma$, and each connected component of G_γ contains points of G .*

Proof The statement $G \subset G_\gamma$ follows as in the above proof, as well as the fact that the global minimum of \mathcal{J} and \mathcal{J}_γ coincide; we denote it by c . If $x \in G_\gamma$ and $\mathcal{J}(x) > c$, then it follows by (19) that there exists a unit vector v such that $\frac{d^2}{dt^2} \mathcal{J}_\gamma(x + tv) \leq 0$ in a neighborhood of $t = 0$. Strict inequality contradicts the assumption of global minima, so we deduce that $\gamma \|v\|^2 = \|Av\|^2$. Introducing the semi-norm $\|x\|_{\mathcal{U}}^2 = \gamma \|x\|_{\mathcal{V}}^2 - \|Ax\|_{\mathcal{W}}^2$, this means that $\|v\|_{\mathcal{U}} = 0$, i.e., that v lies in the kernel of the semi-norm $\|\cdot\|_{\mathcal{U}}$, (which is a linear subspace by bilinearity of the semi-norm). Let P be the affine hyperplane $P = x + \ker \|\cdot\|_{\mathcal{U}}$, and set $S = P \cap G_\gamma$. For $y \in \ker \|\cdot\|_{\mathcal{U}}$, we have

$$\begin{aligned} \mathcal{J}_\gamma(x + y) &= \left(\mathcal{Q}_\gamma(f)(x + y) + \frac{\gamma}{2} \|x + y\|_{\mathcal{V}}^2 \right) \\ &\quad - \frac{1}{2} \|x\|_{\mathcal{U}}^2 - \langle A(x + y), d \rangle_{\mathcal{W}} + \frac{1}{2} \|d\|_{\mathcal{W}}^2, \end{aligned} \quad (20)$$

so Theorem 3.1 implies that \mathcal{J}_γ is convex on P . In particular, S is convex. Since \mathcal{J}_γ is l.s.c., S is also closed. Moreover, S is bounded due to the quadratic term $\|x + y\|_{\mathcal{V}}^2$ in (20). S is therefore weakly closed, and hence it equals the closed convex hull of its extremal points by the Krein-Milman theorem. If x now is one of these extremal points,

then we can argue as in the beginning of this proof and conclude that $\mathcal{J}_\gamma(x) = \mathcal{J}(x)$, since the existence of a v with the properties stated initially would contradict that x is an extremal point of S . \square

6 The Quadratic Envelope and Semi-algebraicity

We briefly treat semi-algebraicity of $\mathcal{Q}_\gamma(f)$, since it was shown in [16] that this is connected with the convergence of the forward–backward splitting method, applied to functionals of the type (18). We remind the reader that a function on a finite-dimensional space is semi-algebraic if its graph is a semi-algebraic set [20].

Theorem 6.1 *If \mathcal{V} is finite dimensional and f is semi-algebraic then so are $\mathcal{S}_\gamma(f)$ and $\mathcal{Q}_\gamma(f)$.*

Proof We assume for simplicity that $\gamma = 1$. It is a consequence of the Tarski-Seidenberg theorem that the set of semi-algebraic functions is closed under addition, (see, e.g., Prop. 2.2.6 in [20]), and similarly, one can prove that the epigraph of a semi-algebraic function is a semi-algebraic set. If f is semi-algebraic on \mathbb{R}^n it follows that $g(x, y) = \langle x, y \rangle - (f(x) + \frac{1}{2} \|x\|^2)$ is semi-algebraic on \mathbb{R}^{2n} , and by the argument following Theorem 2.2 in [16], it follows that the Legendre transform of $f + \frac{1}{2} \|x\|^2$ is semi-algebraic. The first result now follows since this function minus $\frac{\gamma}{2} \|y\|^2$ equals $\mathcal{S}_\gamma(f)(y)$ by (14), and the second is immediate by Proposition 3.1. \square

7 Related Works

The operations $\mathcal{S}_\gamma(f)$ and $\mathcal{Q}_\gamma(f)$ were introduced around 1970 in greater generality by Moreau [21] and (seemingly independently) Weiss [22] and were further studied around 1990 by Poliquin [23] with a focus on smoothness properties. Variations of Propositions 3.1 and 3.2 date back to these early articles, and are also found, e.g., in Rockafellar–Wets [17] Section 11.L. The transforms \mathcal{S}_γ and \mathcal{Q}_γ go under names like “ Φ -conjugate”/“proximal transform” and “ Φ -biconjugate”/“ Φ -convex envelope,” and arise by the concrete choice $\Phi(x, y) = q_\gamma(x, y) = -\frac{\gamma}{2} \|x - y\|^2$. Following Rockafellar–Wets [17], $\mathcal{Q}_\gamma(f)$ should be called “proximal hull” or “ q_γ -envelope”. We believe that the “quadratic envelope”, which is closer to the latter, is more suggestive. Functions that satisfy $\mathcal{Q}_\gamma(f) = f$ have been called, e.g., γ^{-1} -proximal or quadratically convex.

However they are called, it seems that the connection with convex envelopes à la Theorem 3.1 has not been investigated, which is the main novelty of this publication along with the structural result Corollary 4.2 and its applications to regularization in Sect. 5. Apart from the already mentioned works by Aubert, Blanc-Feraud, Soubies and Larsson, Olsson, we have not found any similar result in the literature. The fairly recent survey paper [24] is about the closely related concept of computing Fenchel conjugates and also mentions proximal hulls, yet it has no overlap with the present paper. It primarily deals with numeric computation of convex envelopes in cases when symbolic formulas are not available, and as such it is an interesting alternative to

the methods developed here. The same goes for the papers [25,26]. The importance of computing convex envelopes is stressed in [27], where techniques for computing convex envelopes of so-called convex polyhedral functions are developed. Convex approximations from below are considered in [28], which should be compared with the results in Sect. 5.1. An alternative to approximating the convex envelope is to numerically try to compute the proximal operator of the original functional directly, which is pursued in [29]. The papers [30,31] deal with Lasry–Lions approximants in Hilbert space, but do not make the connection with the convex envelopes. For parameters $s < t$, the Lasry–Lions approximation of f [32] is defined by

$$\begin{aligned} \mathcal{S}_{1/s}\mathcal{S}_{1/t}(f)(x) &= -\left(\inf_y -\left(\inf_w f(w) + \frac{1}{2t}\|w - y\|^2\right) + \frac{1}{2s}\|x - y\|^2\right) \\ &= \sup_y \left(\inf_w f(w) + \frac{1}{2t}\|w - y\|^2\right) - \frac{1}{2s}\|x - y\|^2, \end{aligned} \quad (21)$$

which for $s = t$ gives \mathcal{Q}_{s-1} . This regularization is also studied in Sect. 6 of the more recent publication [31], (with the notation $C(1)f$), mainly with focus on differentiability-results. It is also closely connected to the more general “proximal average,” see, e.g., [33,34] and regularization by “self-dual smoothing” [35]. However, these techniques have been used mainly for modification of convex functions f , whereas $\mathcal{Q}_\gamma(f) = f$ for any l.s.c. convex function.

8 Conclusions

We have provided theory for computing l.s.c. convex envelopes of certain functionals and shown a connection with quadratic envelopes (a.k.a. proximal hulls), which was then used to regularize more intricate problems. We showed that, for sufficiently small values of the parameter γ , this yields convex functionals below the original functional, which coincide with the original functional on a large part of the underlying Hilbert space. For γ sufficiently large, on the other hand, we lose convexity, but gain the desirable feature that the modified functional has the same global minimizers as the original one and fewer local ones. This in turn was based on results regarding the structure of l.s.c. convex envelopes. The results are inspired by prior work from Carl Olsson and Viktor Larsson as well as Emmanuel Soubies, Laure Blanc-Féraud and Gilles Aubert.

Particular cases of these ideas have already been applied to compressed sensing [6], computer vision [36], signal processing and frequency estimation [37,38], as well as structured low-rank approximation [39]. Currently, we are working on more concrete results regarding low-rank approximation, total variation denoising, as well as an application to the classical phase retrieval problem. We hope that other researchers will try these methods on their problems and find that the method is a valuable tool. To aid with this task, an expanded version of this article is available on arXiv [12] with many more examples and useful details.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

1. Chen, S.S., Donoho, D.L., Saunders, M.A.: Atomic decomposition by basis pursuit. *SIAM Rev.* **43**(1), 129–159 (2001)
2. Donoho, D.L.: For most large underdetermined systems of linear equations the minimal ℓ^1 -norm solution is also the sparsest solution. *Commun. Pure Appl. Math.* **59**(6), 797–829 (2006)
3. Candes, E.J., Tao, T.: Decoding by linear programming. *IEEE Trans. Inf. Theory* **51**(12), 4203–4215 (2005)
4. Fazel, M.: Matrix rank minimization with applications. Ph.D. thesis, Stanford University (2002)
5. Recht, B., Fazel, M., Parrilo, P.A.: Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization. *SIAM Rev.* **52**(3), 471–501 (2010)
6. Carlsson, M., Gerosa, D., Olsson, C.: A un-biased approach to compressed sensing. arXiv preprint [arXiv:1806.05283](https://arxiv.org/abs/1806.05283) (2018)
7. Larsson, V., Olsson, C.: Convex low rank approximation. *Int. J. Comput. Vis.* **120**, 194–214 (2016)
8. Soubies, E., Blanc-Féraud, L., Aubert, G.: A continuous exact ℓ_0 penalty (CEL0) for least squares regularized problem. *SIAM J. Imaging Sci.* **8**(3), 1607–1639 (2015)
9. Tseng, P.: Approximation accuracy, gradient methods, and error bound for structured convex optimization. *Math. Program.* **125**(2), 263–295 (2010)
10. Andersson, F., Carlsson, M., Olsson, C.: Convex envelopes for fixed rank approximation. *Optim. Lett.* **11**(8), 1783–1795 (2017)
11. Candès, E.J., Romberg, J., Tao, T.: Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information. *IEEE Trans. Inf. Theory* **52**(2), 489–509 (2006)
12. Carlsson, M.: On convexification/optimization of functionals including an L2-misfit term. arXiv preprint [arXiv:1609.09378](https://arxiv.org/abs/1609.09378) (2016)
13. Candès, E.J., Romberg, J.K., Tao, T.: Stable signal recovery from incomplete and inaccurate measurements. *Commun. Pure Appl. Math.* **59**(8), 1207–1223 (2006)
14. Lucet, Y.: The Legendre–Fenchel transform and the convex Hull of a function: fast computational algorithms. Second-order smoothness and analysis, Ph.D. thesis (1997)
15. Brøndsted, A.: Milman’s theorem for convex functions. *Math. Scand.* **19**, 5–10 (1966)
16. Attouch, H., Bolte, J., Svaiter, B.F.: Convergence of descent methods for semi-algebraic and tame problems: proximal algorithms, forward–backward splitting, and regularized Gauss–Seidel methods. *Math. Program.* **137**(1–2), 91–129 (2013)
17. Rockafellar, R.T., Wets, R.J.B.: *Variational Analysis*, vol. 317. Springer, Berlin (2009)
18. Bauschke, H.H., Combettes, P.L.: *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*. Springer, Berlin (2011)
19. Rockafellar, R.T.: *Convex Analysis*. Princeton University Press, Princeton (2015)
20. Bochnak, J., Coste, M., Roy, M.F.: *Real Algebraic Geometry*, vol. 36. Springer, Berlin (2013)
21. Moreau, J.J.: Inf-convolutions, sous-additivé, convexité des fonctions numériques. *J. Math. Pures Appl.* **49**, 109–154 (1970)
22. Weiss, E.A.: Konjugierte funktionen. *Archiv Math.* **20**(5), 538–545 (1969)
23. Poliquin, R.: Subgradient monotonicity and convex functions. *Nonlinear Anal. Theory Methods Appl.* **14**(4), 305–317 (1990)
24. Lucet, Y.: What shape is your conjugate? A survey of computational convex analysis and its applications. *SIAM Rev.* **52**(3), 505–542 (2010)
25. McCormick, G.P.: Computability of global solutions to factorable nonconvex programs: part I convex underestimating problems. *Math. Program.* **10**(1), 147–175 (1976)
26. Borwein, J.M., Hamilton, C.H.: Symbolic Fenchel conjugation. *Math. Program.* **116**(1–2), 17–35 (2009)
27. Meyer, C.A., Floudas, C.A.: Convex envelopes for edge-concave functions. *Math. Program.* **103**(2), 207–224 (2005)

28. Brighi, B., Chipot, M.: Approximated convex envelope of a function. *SIAM J. Numer. Anal.* **31**(1), 128–148 (1994)
29. Hare, W., Sagastizábal, C.: Computing proximal points of nonconvex functions. *Math. Program.* **116**(1–2), 221–258 (2009)
30. Attouch, H., Azé, D.: Approximation and regularization of arbitrary functions in Hilbert spaces by the Lasry–Lions method. *Ann. l’IHP Anal. Non linéaire* **10**, 289–312 (1993)
31. Strömberg, T.: On regularization in Banach spaces. *Ark. Mat.* **34**(2), 383–406 (1996)
32. Lasry, J.M., Lions, P.L.: A remark on regularization in Hilbert spaces. *Isr. J. Math.* **55**(3), 257–266 (1986)
33. Bauschke, H.H., Goebel, R., Lucet, Y., Wang, X.: The proximal average: basic theory. *SIAM J. Optim.* **19**(2), 766–785 (2008)
34. Hare, W.L.: A proximal average for nonconvex functions: a proximal stability perspective. *SIAM J. Optim.* **20**(2), 650–666 (2009)
35. Goebel, R.: Self-dual smoothing of convex and saddle functions. *J. Convex Anal.* **15**(1), 179 (2008)
36. Olsson, C., Carlsson, M., Andersson, F., Larsson, V.: Non-convex rank/sparsity regularization and local minima. In: *Proceedings of the IEEE International Conference on Computer Vision*, pp. 332–340 (2017)
37. Caprani, J., Carlsson, M.: Quadratic envelope regularization for structured low rank approximation. In: *ICASSP Proceedings* (2019)
38. Andersson, F., Carlsson, M., Wendt, H.: On a fixed-point algorithm for structured low-rank approximation and estimation of half-life parameters. In: *2016 24th European Signal Processing Conference (EUSIPCO)*, pp. 326–330. IEEE (2016)
39. Andersson, F., Carlsson, M.: Fixed-point algorithms for frequency estimation and structured low rank approximation. *Appl. Comput. Harmon. Anal.* **46**, 40–65 (2017)

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.