Penalty Functions Derived From Monotone Mappings

İlker Bayram

***Abstract*—We consider the problem of constructing a penalty function associated with a given monotone function. We provide a construction that allows the monotone function to be discontinuous or bounded. We show that, although the penalty may be non-convex, it is weakly convex. We briefly discuss the implications for iterative solutions of linear inverse problems.**

***Index Terms*—Monotone denoising operator, nonconvex penalty, weakly convex, hard threshold**

I. Introduction

We consider functions $T : \mathbb{R} \to \mathbb{R}$ associated with denoising problems of the form,

$$T(y) = \arg\min_{x} \frac{1}{2}(x-y)^2 + P(x), \tag{1}$$

where P is a penalty function. Although T is derived from the penalty function P in this description, monotone functions with desirable properties can be defined independently and utilized in iterative algorithms that solve more complicated problems [1]–[4]. Therefore, it is also of interest to start from a monotone function and find the associated penalty function and/or its derivative (since some iterative algorithms [5] require the derivative of the cost). In this letter, we consider such a construction.

This problem has been considered in [3], [6]. However, both papers impose certain conditions on the monotone mappings involved. Thm. 1 in [3] requires that the monotone mapping be continuous and Prop. 3.2 in [6] requires that the monotone mapping increase indefinitely (for the positive reals), while being bounded by the identity mappping. Nevertheless, both papers contain complementary properties of the penalty functions. Here, we do not impose any condition other than monotonicity. By extending the monotone functions to maximal monotone mappings, we show that the constructions in [3], [6] (which actually obtain equivalent penalties through alternative routes) can be generalized to handle arbitrary monotone functions.

In Section II, we describe a construction that proves the following proposition.

Proposition 1. Let $T : \mathbb{R} \to \mathbb{R}$ be a non-decreasing, non-constant function with at most a countable number of discontinuities. Then, there exists a function $P(x)$ such that

(i) $Q(x) = \frac{1}{2}x^2 + P(x)$ is convex,
(ii) $T(y) \in \arg\min_x \frac{1}{2}(x-y)^2 + P(x)$. $\square$

The author is with Istanbul Technical University, Istanbul, Turkey. Email: ibayram@itu.edu.tr

Note that the proposition does not require T to be continuous. If it is, then Q in (i) can be shown to have a unique minimum and the inclusion in (ii) becomes an equality.

When P is convex, T is also called the proximity operator (or proximal mapping) of P [7], [8]. Proximal mappings are of interest not merely in denoising problems. They also arise in iterative algorithms that solve more complicated linear inverse problems [7]. However, we would like to point out that the induced penalty function P constructed in this letter may be non-convex. Non-convexity of P becomes crucial [2], [3], [6], [9], [10] if one wants a monotone mapping that converges to the identity (which implies unbiasedness in a certain context). Although P is not convex, if we add a quadratic as $x^2/2$, the sum becomes convex. Such functions are called weakly-convex (in fact, ρ-convex with $\rho = -1/2$, according to [11]). Therefore, Prop. 1 implies that, to come up with a given monotone denoising function, one may restrict attention to weakly convex penalty functions. This is an interesting remark because usually, the hard threshold is associated with the ℓ_0 count [1], [3], which is not weakly convex. Another recent example is the $\log$-threshold introduced in [4] derived from the local minima of a denoising problem using a non-convex symmetric penalty function. According to our proposition, there exist weakly-convex penalty functions that give the same threshold functions.

II. Construction of the Penalty

In the following, we give a constructive proof for Prop. 1 along with some discussion. We start with an outline. Then, the main proof is provided in Sec. II-B, followed by supplementary discussion subsections.

A. Outline

To set the stage, let us start with a definition from convex analysis. For detailed discussions of concepts from convex analysis and monotone operator theory we refer to [8], [12], [13].

Definition 1. Let $f : \mathbb{R} \to \mathbb{R}$ be a convex function. The subdifferential of f at x is denoted by $\partial f(x)$ and is the set of $y \in \mathbb{R}$ that satisfy

$$f(x) + (z-x)\, y \leq f(z), \tag{2}$$

for all $z \in \mathbb{R}$. $\square$

Note that we can regard ∂f as a set valued mapping (i.e., $\partial f : \mathbb{R} \to 2^{\mathbb{R}}$). It follows from this definition that x^* is a minimizer of f if and only if $0 \in \partial f(x^*)$. Consider now the cost function in (1). Note that we can write it as,

$$h(x) = \underbrace{\frac{1}{2}x^2 + P(x)}_{Q(x)} - x\,y + \frac{1}{2}\,y^2. \tag{3}$$

If $Q(x)$ is convex, we can therefore write

$$0 \in -y + \partial Q(x^*), \tag{4}$$

where $x^* = T(y)$ is a minimizer. Therefore, we have, formally,

$$T(y) \in (\partial Q)^{-1}(y). \tag{5}$$

If Q were strictly convex and differentiable, then (5) would actually be an equality of the form $T = (\partial Q)^{-1}$. This suggests that Q can be obtained from T by considering the anti-derivative of T^{-1}. Then, P can be obtained by subtracting $x^2/2$ from Q. However, if Q is not strictly convex and differentiable, the plan above does not work. In such a case, T might have discontinuities, or may be bounded. As we demonstrate below, discontinuities in T will lead to gaps in T^{-1}, preventing integration. When T is bounded as in a $\tanh$ function, T^{-1} is again not defined everywhere on $\mathbb{R}$. In order to handle such monotone mappings as well, our plan is to first extend the given monotone function to a maximal monotone mapping. Then we will take the inverse of the resulting mapping and integrate to obtain the sum of our penalty function and $x^2/2$.

B. Inverting the Monotone Mapping (Proof of Prop. 1)

In this subsection, we provide a constructive proof for Prop. 1. Proofs of Prop. 2 and Prop. 5 can be found in [8] (Chp. 12) or [13] (Chp. 20). The proofs of the remaining propositions are provided in the supplementary material.

A concept we will rely on is the graph of a set-valued mapping. In the following, let $M : \mathbb{R} \to 2^{\mathbb{R}}$ be a generic set-valued mapping.

Definition 2. The *graph* of M, denoted as $\operatorname{gra} M$, is the set of pairs (x, y) such that $y \in M(x)$. □

Note that the graph of M can also be used to define M. We can therefore think of M and its graph as describing the same object.

Definition 3. M is said to be *monotone* if for any $(x, y) \in \operatorname{gra} M$, $(x', y') \in \operatorname{gra} M$ with $x < x'$, we have $y \le y'$. □

According to this definition, any monotone function on $\mathbb{R}$ can be associated with a monotone set-valued mapping. In the following, with an abuse of language, we will therefore talk about the graph of a monotone function. An example of a monotone mapping is the hard threshold, defined as,

$$T(x) = \begin{cases} 0, & \text{if } |x| \le \tau, \\ x, & \text{if } \tau < |x|, \end{cases} \tag{6}$$

where τ is a constant.

In going from the threshold function to the penalty, we need some sort of an inversion, per (5). In [3], this is achieved by a convex conjugation operation. Here, we will make use of the inverse monotone mapping.

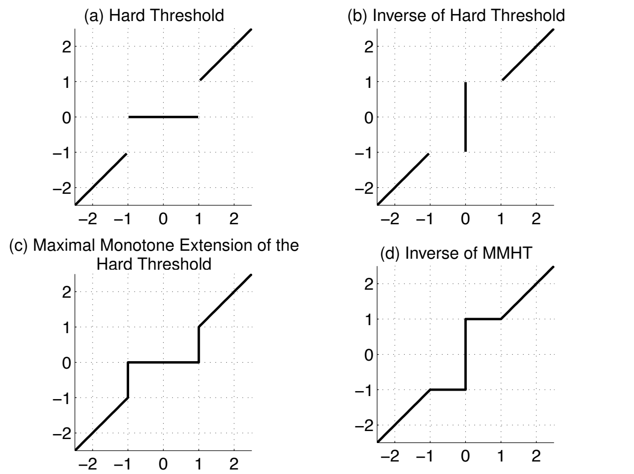

Fig. 1. (a) The hard threshold function with $\tau = 1$. (b) The inverse of the hard threshold function as a set valued mapping. (c) Maximal monotone extension of the hard threshold function. (d) The inverse of the mapping in (c).

Definition 4. The *inverse* of M, denoted as M^{-1} is the set valued mapping whose graph consist of pairs (x, y) such that $x \in M(y)$. □

Proposition 2. M is monotone if and only if M^{-1} is monotone. □

The graphs of the hard threshold and its inverse are shown in Fig.1a,b. Observe that, regarded as set valued mappings, both the hard threshold and its inverse are monotone. However, the inverse of the hard threshold is not defined for $0 < |x| \le \tau$, although it is defined for $\tau < |x|$. This prevents integration, which is required to pass to the function Q. We will see that this issue does not arise if the monotone mapping is maximal.

Definition 5. A monotone mapping M is said to be *maximally monotone* if there exists no monotone mapping M' whose graph is a proper superset of $\operatorname{gra} M$. □

For the monotone functions we are interested in, an equivalent condition for maximality is given in the following.

Proposition 3. Let $M : \mathbb{R} \to 2^{\mathbb{R}}$ be monotone, non-constant and $M(x)$ be non-empty for all $x \in \mathbb{R}$. Also, let M^-, M^+ denote the extremal values of M, defined as,

$$M^- = \lim_{x \to -\infty} \big(\inf M(x)\big), \quad M^+ = \lim_{x \to \infty} \big(\sup M(x)\big). \tag{7}$$

In this setting, M is maximal if and only if

(i) for any x, $M(x)$ is a finite closed interval, and
(ii) for any y with $M^- < y < M^+$, there exists x such that $y \in M(x)$. □

Condition (ii) in Prop. 3 essentially implies that if a maximal monotone M has values below and above y then we can find an x with $y \in M(x)$ – moreover the converse (given the further condition in (i)) also holds. This result is useful for finding maximal monotone extensions of monotone functions considered in the letter and also checking maximality.

The hard threshold is not maximally monotone, because it violates condition (ii) of Prop. 3. T assumes the values $-\tau$ and τ, but there is no x such that $\tau/2 \in T(x)$. Note that, in this case, the discontinuity at τ prevents maximality. In fact, we obtain maximal monotone extensions by filling in these jumps.

In general, there may exist more than one maximal extension of a monotone mapping [13]. However, in our setup, where we consider maximal extensions of monotone (single-valued) functions, the extension is unique.

Proposition 4. Suppose M is a single valued monotone function, defined everywhere and discontinuous on at most a countable number of points, collected in the set J. For $x \in J$, let I_x^- and I_x^+ denote the limit inferior and superior at x defined as,

$$I_x^- = \liminf_{z \to x} M(z), \quad I_x^+ = \limsup_{z \to x} M(z), \tag{8}$$

and set V_x to be the set of pairs (x, y) such that $I_x^- \leq y \leq I_x^+$. Now let $\widetilde{M}$ be defined such that

$$\text{gra } \widetilde{M} = \big(\text{gra } M\big) \cup \left(\bigcup_{x \in J} V_x \right). \tag{9}$$

Then $\widetilde{M}$ is the unique maximal monotone mapping whose graph contains the graph of M. □

If we apply the construction in Prop. (4) to the hard threshold, we end up with the maximal monotone mapping shown in Fig. 1c. Let us now consider the properties of the inverse of a maximal monotone operator.

Proposition 5. M is maximal monotone if and only if M^{-1} is maximal monotone. □

Proposition 6. Suppose M is maximal monotone and $M(x)$ is non-empty for all x. Then there exists an interval I such that $M^{-1}(x)$ is non-empty if $x \in I$ and empty if $x \notin I$. □

This proposition ensures that for $x < t < x'$, if $M^{-1}(x)$ and $M^{-1}(x')$ are both non-empty, then $M^{-1}(t)$ is non-empty too. This is demonstrated by the inverse of the maximal extension of the hard threshold (see Fig. 1d). Observe that the region $0 < |x| < \tau$ is now filled.

It is well known that if M is a maximal monotone mapping on $\mathbb{R}$, then there is a convex lower semicontinuous proper function f such that $M = \partial f$ [8], [13]. Further, we can obtain f from ∂f up to a constant. Given a pair (x_0, y_0) with $y_0 \in \partial f(x_0)$, $f(x)$ can be written (up to a constant) as [8], [13]

$$f(x) = \sup_m \sup_{\substack{x_1, \ldots, x_m \\ y_1, \ldots y_m}} y_m\,(x - x_m) + \sum_{i=0}^{m-1} y_i (x_{i+1} - x_i), \tag{10}$$

where $y_i \in \partial f(x_i)$. This definition chooses the unknown constant in the definition of f such that $f(x_0) = 0$.

Recalling the discussion in Sec. II-A, suppose now that $\widetilde{T}^{-1}$ denotes the inverse of the maximal extension of T. By Prop. 5, $\widetilde{T}^{-1}$ is maximal monotone. So we can find a convex function Q whose subdifferential is $\widetilde{T}^{-1}$. Also, for a given y, let $z = T(y)$. Then, we will have that $z \in \widetilde{T}(y)$ since $\widetilde{T}$ is an extension of T. But this means that $y \in \widetilde{T}^{-1}(z) = \partial Q(z)$. Equivalently, $0 \in \partial Q(z) - y$, which in turn is equivalent to

$$z \in \arg\min_x \; Q(x) - x\,y \tag{11a}$$

$$= \arg\min_x \frac{1}{2}(x - y)^2 + \big(Q(x) - x^2/2\big). \tag{11b}$$

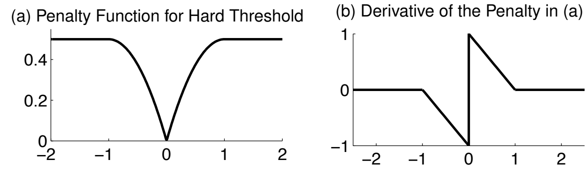

Fig. 2. (a) The penalty function associated with the hard threshold. (b) The generalized gradient of the penalty in (a).

Thus for Q with $\partial Q = \widetilde{T}^{-1}$, obtained via (10), we can set $P(x) = Q(x) - x^2/2$, proving Prop. 1.

While (10) applies for a general convex function, here, our maximal monotone operators are defined on $\mathbb{R}$ and an integration procedure that is easier to realize can be given (see also [3] for a similar procedure).

C. Integration of a Maximal Monotone Mapping on $\mathbb{R}$

By Prop. 6, we know that there is an interval I on which $\widetilde{T}^{-1}(x)$ is non-empty. Let $J \subset I$ be the set of x where $\widetilde{T}^{-1}(x)$ is not a singleton. For threshold functions useful in practice, J is at most a finite set and we will work under this assumption. We start by setting $Q(x) = \infty$ if x is not in the closure of I. Now pick x_-, x_+ from J such that $(x_-, x_+) \cap J = \emptyset$. On the interval (x_-, x_+), $\widetilde{T}^{-1}$ is single-valued and its antiderivative determines Q up to a constant, since $\partial Q = \widetilde{T}^{-1}$. Therefore, Q is determined up to a constant for a finite collection of intervals which partition I. We choose the constant in each interval so that the resulting Q is continuous in the interior of I. We extend Q on the boundaries of I by continuity. Once we obtain Q, we subtract $x^2/2$ to obtain the penalty function, P.

If we apply this procedure to the inverse of the maximal monotone hard threshold (Fig. 1d), then the resulting penalty function is (as is noted in [9], [10]),

$$P(x) = \begin{cases} \tau|x| - x^2/2, & \text{if } |x| \leq \tau, \\ \tau^2/2, & \text{if } \tau < |x|. \end{cases} \tag{12}$$

This penalty is sketched in Fig. 2a. This P makes h in (3) convex, but not strictly convex. When $y = \pm\tau$, the minimum of h occurs at $\pm[0, \tau]$.

D. Derivative of the Penalty Function

We noted in the introduction that it is also of interest to find the derivative of the penalty. For the penalty function of interest, since $Q(x) = P(x) + x^2/2$, we can formally write $\partial P(x) = \partial Q(x) - x$. For the hard threshold, ∂P is sketched in Fig. 2b. When Q is smooth, ∂P is the derivative of the penalty function. However, if Q is not smooth (as in the hard threshold), ∂P is not necessarily a subdifferential in the sense of Defn. 1 above. In particular, we observe that ∂P is not monotone for the hard threshold, signalling that it is not associated with a convex function. When ∂P is not smooth, under certain Lipschitz conditions, it coincides with the generalized derivative in [14] or the generalizations of the subdifferential that can be found in [8] (see e.g. Chp 8).

III. Examples

We demonstrate the procedure on two more examples. But before that, let us note a point useful in computations.

Suppose T is a monotone function. In order to obtain the associated penalty, we need to integrate T^{-1}. If it is easier to integrate $T(x)$, a useful equality is,

$$\int_a^b T^{-1}(x)\,dx = x\,T^{-1}(x)\Big|_a^b - \int_{T^{-1}(a)}^{T^{-1}(b)} T(x)\,dx. \quad (13)$$

Example 1 (Hyperbolic Tangent)**.** We construct the penalty function associated with the monotone mapping

$$T_\alpha(x) = \tanh(\alpha x) = \frac{e^{\alpha\,x} - e^{-\alpha\,x}}{e^{\alpha\,x} + e^{-\alpha\,x}}. \quad (14)$$

$T_\alpha(x)$ is shown in Fig. 3a, for $\alpha = 5,\ 15$. With increasing α, $T_\alpha(x)$ converges to the sgn function, which can be interpreted as a binary classifier of the input. Note that, $T_\alpha(x)$ is maximal, by the criterion in Prop. 3. We find,

$$T_\alpha^{-1}(x) = \ln\left(s_\alpha(x)\right), \text{ for } -1 < x < 1, \quad (15)$$

where,

$$s_\alpha(x) = \left(\frac{1+x}{1-x}\right)^{1/2\alpha}. \quad (16)$$

Because $|T_\alpha(x)| < 1$ for all x, $T_\alpha^{-1}(x)$ is constrained to the interval $I = (-1, 1)$. Outside the closure of this interval, we set $Q_\alpha(x) = \infty$, as required by the construction. On I, we first set $Q_\alpha(0) = 0$. Then, using (13), we find,

$$Q_\alpha(x) = \int_0^x T_\alpha^{-1}(u)\,du \quad (17)$$

$$= x\ln\Big(s_\alpha(x)\Big) - \ln\Big(s_\alpha(x) - 1/s_\alpha(x)\Big), \quad (18)$$

for $0 < x < 1$. We extend $Q_\alpha(x)$ to $(-1, 0)$ by symmetry. Note that $\lim_{x\to 1} Q_\alpha(x) = \infty$. Finally, we obtain the penalty function as $P_\alpha(x) = Q_\alpha(x) - x^2/2$. $P_\alpha(x)$ is shown in Fig. 3b for $\alpha = 5,\ 15$. With increasing α, $P_\alpha(x)$ gets closer to $P(x) = -x^2$ on $|x| \le 1$. We note that $P(x)$ is the penalty associated with the sgn function (which can be derived independently following the procedure in this letter). Notice also that for fixed α, since $\lim_{x\to 1} P_\alpha(x) = \infty$, convergence of $P_\alpha(x)$ to $P(x)$ is not uniform but pointwise for $|x| < 1$.

Example 2 (Log Threshold)**.** In a final example, we consider the threshold function from [4], given by,

$$T(x) = \begin{cases} 0, & \text{if } |x| \le \tau, \\ \frac{1}{2}\operatorname{sgn}(x)\Big((|x| - \delta) \\ \quad + \sqrt{(|x|+\delta)^2 - 2\lambda}\Big), & \text{if } \tau < |x|, \end{cases} \quad (19)$$

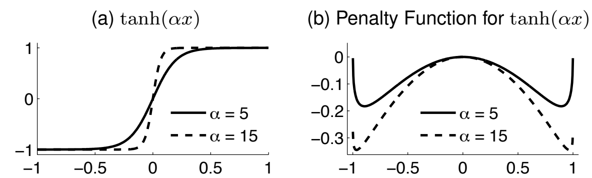

Fig. 3. (a) Scaled hyperbolic tangent functions are monotone and concentrate the input around -1 and 1. (b) The associated penalty favors points near -1 and 1. Outside the unit interval, the penalty assumes the value ∞.

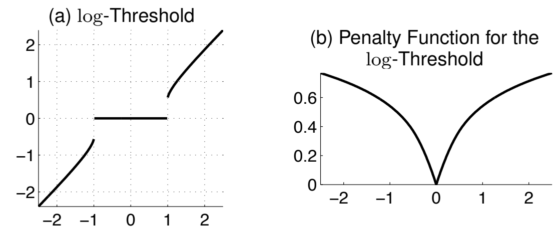

Fig. 4. (a) The log-threshold in (19) with $\lambda = 1/2$, $\delta = 1/100$. (b) The associated penalty given in (20) for the threshold function in (a).

where $\tau = \sqrt{2\lambda} - \delta$. This threshold function (see Fig. 4a) is derived in [4] by considering a logarithmic penalty of the form $\widetilde{P}(x) = (\lambda/2)\,\ln(|x| + \delta)$. Note that the function $h(x) = (x-y)^2/2 + \widetilde{P}(x)$ is not convex for small δ. The authors in [4] construct the threshold function above by considering the local minima of the cost. However, since $T(x)$ is monotone, Prop. 1 suggests that there is a weakly convex function $P(x)$ associated with $T(x)$. Following our procedure, we find that the weakly convex function

$$P(x) = \begin{cases} \Big(\sqrt{2\lambda} - \delta\Big)\ |x| - x^2/2, & \text{if } |x| < \sqrt{\frac{\lambda}{2}} - \delta, \\ \frac{\lambda}{2}\ \ln\big(|x| + \delta\big) + c, & \text{if } |x| \ge \sqrt{\frac{\lambda}{2}} - \delta, \end{cases} \quad (20)$$

leads to the threshold above. In (20), the constant c is chosen so that $P(x)$ is continuous (see Fig. 4b).

IV. Discussion

Prop. 1 is also of interest for more general linear inverse problems. In particular, if we have noisy data acquired under linear distortion represented by an operator H, a widely used reconstruction algorithm is of the form

$$x^{n+1} = T\Big(x^n + \alpha H^T\,(y - H\,x^n)\Big), \quad (21)$$

where T is a denoising operator [15]–[17]. If T is the proximity operator of a convex function [7], this algorithm becomes an instance of a forward-backward splitting algorithm, which is known to converge [13], [18]. However, many interesting denoisers, like the hard threshold and the log-threshold, do not satisfy such a requirement (see e.g. Prop. 1 in [2] or Thm. 1 in [3]). In that case, Prop. 1 can be used to construct a Lyapunov function [19] for the algorithm in (21), whose value decreases monotonically with iterations. This may help prove convergence of the iterations in (21).

Acknowledgements

The author thanks the anonymous reviewers for their comments that helped improve the letter.

REFERENCES

[1] T. Blumensath and M. E. Davies, "Iterative thresholding for sparse approximations," *Journal of Fourier Analysis and Applications*, vol. 14, no. 5, pp. 629–654, Dec. 2008.

[2] I. W. Selesnick and İ. Bayram, "Sparse signal estimation by maximally sparse convex optimization," *IEEE Trans. Signal Processing*, vol. 62, no. 5, pp. 1078–1092, Mar. 2014.

[3] R. Chartrand, "Shrinkage mappings and their induced penalty functions," in *Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Proc. (ICASSP)*, 2014.

[4] D. Malioutov and A. Aravkin, "Iterative log thresholding," in *Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Proc. (ICASSP)*, 2014.

[5] E. J. Candès, M. B. Wakin, and S. Boyd, "Enhancing sparsity by reweighted l1 minimization," *Journal of Fourier Analysis and Applications*, vol. 14, no. 5, pp. 877–905, Dec. 2008.

[6] A. Antoniadis, "Wavelet methods in statistics : Some recent developments and their applications," *Statistics Surveys*, vol. 1, pp. 16–55, 2007.

[7] P. L. Combettes and J.-C. Pesquet, "Proximal splitting methods in signal processing," in *Fixed-Point Algorithms for Inverse Problems in Science and Engineering*, H. H. Bauschke, R. S. Burachik, P. L. Combettes, V. Elser, D. R. Luke, and H. Wolkowicz, Eds. Springer, New York, 2011.

[8] R. T. Rockafellar and R. J.-B. Wets, *Variational Analysis*. Springer, 1998.

[9] A. Antoniadis and J. Fan, "Regularization of wavelet approximations," *Journal of the American Statistical Association*, vol. 96, no. 454, pp. 939–955, Sep. 2001.

[10] J. Fan and R. Li, "Variable selection via nonconcave penalized likelihood and its oracle properties," *Journal of the American Statistical Association*, vol. 96, no. 456, pp. 1348–1360, Dec. 2001.

[11] J.-P. Vial, "Strong and weak convexity of sets and functions," *Mathematics of Operations Research*, vol. 8, pp. 231–259, May 1983.

[12] J.-B. Hiriart-Urruty and C. Lemaréchal, *Fundamentals of Convex Analysis*. Springer, 2004.

[13] H. H. Bauschke and P. L. Combettes, *Convex analysis and monotone operator theory in Hilbert spaces*. Springer, 2011.

[14] F. H. Clarke, *Optimization and Nonsmooth Analysis*. SIAM, 1990.

[15] M. A. T. Figueiredo and R. D. Nowak, "An EM algorithm for wavelet-based image restoration," *IEEE Trans. Image Proc.*, vol. 12, no. 8, pp. 906–916, Aug. 2003.

[16] M. Elad, B. Matalon, J. Shtok, and M. Zibulevsky, "A wide-angle view at iterated shrinkage algorithms," in *Proc. SPIE Wavelets XII*, 2007.

[17] I. Daubechies, M. Defrise, and C. D. Mol, "An iterative thresholding algorithm for linear inverse problems with a sparsity constraint," *Commun. Pure Appl. Math.*, vol. 57, no. 11, pp. 1413–1457, Aug. 2004.

[18] P. L. Combettes and V. R. Wajs, "Signal recovery by proximal forward-backward splitting," *SIAM J. Multiscale Model. Simul.*, vol. 4, no. 4, pp. 1168–1200, Nov. 2005.

[19] D. G. Luenberger, *Introduction to Dynamic Systems*. Wiley, 1979.

APPENDIX

In this supplementary material, we provide proofs of some statements from Section II.

Proofs of Prop. 2 and Prop. 5 can be found in [8] (Chp. 12) or [13] (Chp. 20). These propositions hold for general monotone mappings. Below we provide the proofs of the propositions that are valid for the monotone mappings we are interested in.

Proof of Prop. 3. ($\Rightarrow$) We start by showing that (i) follows. Assume that M is maximal monotone and $M(x)$ is non-empty for all x. Given $x_0 < x^* < x_1$, let $y_0 \in M(x_0)$, $y_1 \in M(x_1)$ (these exist because $M(x)$ is non empty for all x). But by monotonicity, $y_0 \leq \inf M(x^*)$, and $\sup M(x^*) \leq y_1$, so that $M(x^*)$ is contained in a finite interval. Let us now show that $M(x^*)$ has to be also closed. Suppose on the contrary that $M(x^*)$ is not a closed interval. Let M' be defined by setting $M'(x^*)$ to be the closed interval $[\inf M(x^*), \sup M(x^*)]$ and $M'(x) = M(x)$ for $x \neq x^*$. Note that if $x^* < x$, then $\sup M(x^*) \leq \inf M(x)$ and if $x < x^*$, then $\sup M(x) \leq \inf M(x^*)$. Combined with the monotonicity of M, this implies that $M'(x)$ is monotone. But $M'(x^*)$ is a proper superset of $M(x^*)$ so that $\operatorname{gra} M'$ is a proper superset of $\operatorname{gra} M$. Thus, M cannot be maximal monotone, which contradicts the assumption. Therefore $M(x^*)$ has to be a closed interval (possibly a singleton).

We now consider (ii). Assume that M is maximal monotone, non-constant and $M(x)$ is non-empty for all x. Suppose however that (ii) does not hold. Then, for some y with $M^- < y < M^+$, we have $y \notin M(x)$ for any x. Now let,

$$\begin{aligned} x_- &= \sup x, \text{ subject to } \sup M(x) < y, \\ x_+ &= \inf x, \text{ subject to } \inf M(x) > y. \end{aligned}$$

Since $M^- < y < M^+$, we have $-\infty < x_-$ and $x_+ < \infty$. Further, by the monotonicity of M, we have $x_- \leq x_+$. Now let M' be defined such that

$$\operatorname{gra} M' = \operatorname{gra} M \cup \Big\{ \big((x_- + x_+)/2, y\big) \Big\}.$$

Then, M' is also monotone and $\operatorname{gra} M'$ is a proper superset of $\operatorname{gra} M$. But this contradicts the maximality of M. Therefore we conclude that (ii) must hold.

($\Leftarrow$) Assume now that (i) and (ii) hold, M is monotone, non-constant and $M(x)$ is non-empty for all x. Suppose, however, that M is not maximal monotone. Since M is not maximal monotone, we can find a monotone M' such that $\operatorname{gra} M'$ is a proper superset of $\operatorname{gra} M$. Therefore, we can find a pair (x^*, y^*), such that $y^* \in M'(x^*)$ but $y^* \notin M(x^*)$. Since by the assumptions, $M(x^*)$ is a non-empty closed finite interval, we will have either $y^* < \min M(x^*)$ or $y^* > \max M(x^*)$. Suppose $y^* < \min M(x^*)$. Also, let $\tilde{y} = (y^* + \min M(x^*))/2$. Note that, since M' is monotone, $\operatorname{gra} M' \supset \operatorname{gra} M$, and $y^* < \tilde{y} < \min M(x^*)$, it follows that $\tilde{y} \notin M(x)$ for any x. Then, by (ii), we must have $\tilde{y} \leq M^-$. Now pick $x_0 < x^*$. Also, let $y_0 \in M(x_0)$. We should also have $y_0 \in M'(x_0)$. But since M' is monotone, $y_0 \leq y^*$. Thus we obtain the chain of inequalities $M^- \leq y_0 \leq y^* < \tilde{y} \leq M^-$ which cannot be true. By a similar argument, we can obtain a contradiction for the case $y^* > \max M(x^*)$. Therefore M has to be maximal. □

Proof of Prop.4. Maximality of $\widetilde{M}$ follows by Prop. 3.

We need to show that $\widetilde{M}$ is unique. Suppose now that M' is another maximal extension of M, different than $\widetilde{M}$. Then, we can find a pair (x^*, y^*) such that $y^* \in M'(x^*)$ but $y^* \notin \widetilde{M}(x^*)$. Since $\widetilde{M}(x^*)$ is a closed interval (by Prop. 3), this implies that either $y^* < \min \widetilde{M}(x^*)$ or $y^* > \max \widetilde{M}(x^*)$. Suppose $y^* < \min \widetilde{M}(x^*)$. By the definition of $\widetilde{M}$ and the intervals in (8), we can find a point $x_0 < x^*$ with $y_0 = M(x_0)$ such that $y^* < y_0 \leq \min \widetilde{M}(x^*)$. But since M' is an extension of M, this implies that $y_0 \in M'(x_0)$, violating the monotonicity of M'. It can be shown by a similar argument that if $y^* > \max \widetilde{M}(x^*)$, M' cannot be monotone. Thus there exists no maximal monotone extension of M other than $\widetilde{M}$.

□

Proof of Prop. 6. Let M^- and M^+ be defined as,

$$M^- = \lim_{x \to -\infty} \big(\inf M(x)\big), \quad M^+ = \lim_{x \to \infty} \big(\sup M(x)\big).$$

Consider an interval $I \subset \mathbb{R}$, whose interior is (M^-, M^+). If $M^- \in M(x)$ for any x, then include M^- in I, otherwise exclude it. Similarly, if $M^+ \in M(x)$ for any x, then include M^+ in I, otherwise exclude it. This construction, along with condition (ii) in Prop. 3, implies that,

(a) if $y \in I$, then there exists x such that $y \in M(x)$,
(b) if $y \notin I$, then there is no x such that $y \in M(x)$.

But these are equivalent to,

(a) if $y \in I$, then there exists x such that $x \in M^{-1}(y)$, or that $M^{-1}(y)$ is non-empty,
(b) if $y \notin I$, then $M^{-1}(y) = \emptyset$.

□