Numer Algor (2006) 43:235–249
DOI 10.1007/s11075-006-9056-0

ORIGINAL PAPER

Fast Moreau envelope computation I: numerical algorithms

Yves Lucet

Received: 29 September 2005 / Revised: 29 November 2006 /
Accepted: 29 November 2006 / Published online: 5 January 2007

Abstract The present article summarizes the state of the art algorithms to compute the discrete Moreau envelope, and presents a new linear-time algorithm, named NEP for NonExpansive Proximal mapping. Numerical comparisons between the NEP and two existing algorithms: The Linear-time Legendre Transform (LLT) and the Parabolic Envelope (PE) algorithms are performed. Worst-case time complexity, convergence results, and examples are included. The fast Moreau envelope algorithms first factor the Moreau envelope as several one-dimensional transforms and then reduce the brute force quadratic worst-case time complexity to linear time by using either the equivalence with Fast Legendre Transform algorithms, the computation of a lower envelope of parabolas, or, in the convex case, the non expansiveness of the proximal mapping.

Keywords Moreau–Yosida approximate · Legendre–Fenchel transform · conjugate · discrete Legendre transform · computational convex analysis

Mathematics Subject Classifications (2000) 52B55 · 65D99

1 Introduction

The Moreau envelope [71] of an extended real-valued function $f : \mathbb{R}^d \to \mathbb{R} \cup \{\infty\}$, also called the Moreau–Yosida approximate, Yosida Approximate [1] or Moreau–Yosida regularization

$$M_\lambda(s) := \inf_{x \in \mathbb{R}^d} \left[\frac{\|s-x\|^2}{2\lambda} + f(x) \right] \tag{1}$$

Y. Lucet (✉)
Computer Science, I. K. Barber School of Arts and Sciences-Unit 4, The University of British Columbia Okanagan, 3333 University Way, Kelowna, BC V1V 1V7, Canada
e-mail: yves.lucet@ubc.ca

has been studied extensively both theoretically and algorithmically for its regularization properties. Its origin goes back to the work of Yosida [77] on maximal monotone operators, and its behaviour is well known in the field of convex analysis [58–60, 69] and variational analysis [71, Chapter 12]. Under general conditions, M_λ is C^1 with Lipschitz continuous gradient, and critical points of f are fixed points of the proximal mapping

$$P_\lambda(s) := \operatorname*{Argmin}_{x\in\mathbb{R}^d}\left[\frac{\|s-x\|^2}{2\lambda}+f(x)\right]. \tag{2}$$

When f is convex lower semi-continuous and proper, the proximal mapping is a maximal monotone operator and its fixed points are the minima of f. More precise smoothness results are known under various hypothesis on f [12, 24, 45, 46, 48, 67] and recent developments focus on the notion of prox-regularity [4–6, 44, 65, 66].

The Moreau envelope is an attractive regularization transform considering that M_λ converges pointwise to $f(x)$ when $\lambda \searrow 0$, and that it shares the same critical points of f. On the practical side, the proximal point algorithm exploits the fixed point property of the proximal mapping to converge to a minimum of f [70]. Its convergence properties are well known [21, 34], and variants have been introduced to speed it up [7, 9–11, 13, 33, 47, 57, 61, 62, 68, 73, 76]. Extensions to non-quadratic kernels like entropy methods and Bregman distances have also been studied [17, 30, 74].

In addition to proximal point algorithm variants and extensions, bundle methods are another family of numerical optimization algorithms that is intrinsically linked to the Moreau envelope (see [47], and [28, Chapter XV]). Recent developments in that direction focus on $\mathcal{VU}$-decomposition [35–39, 49–56].

While the present article is concerned with the numerical computation of the Moreau envelope, contrary to [25] we do not consider computing its value at one point but instead we tackle the problem of computing the Moreau envelope on a grid.

Similar algorithms have been developed to compute the Legendre conjugate (also named Legendre–Fenchel conjugate, and Legendre–Fenchel Transform)

$$f^*(s) = \sup_{x\in\mathbb{R}^n}[\langle s,x\rangle - f(x)] \tag{3}$$

motivated by the study of some Hamilton–Jacobi differential equations. A log-linear algorithm named the Fast Legendre Transform (FLT for short, by analogy with the Fast Fourier Transform) was first introduced [8, 14, 40, 64, 72] to be subsequently improved by a linear-time algorithm: The Linear-time Legendre transform (LLT) [41]. The FLT and the faster LLT algorithms have been used in efficient numerical simulations of the Burger's equation [2, 3, 19, 20, 22, 23, 63, 75]. The LLT has also found applications in robotics [31], network communication [29], pattern recognition [43], numerical simulation of multiphasic flows [26], and analysis of the distribution of chemical compounds in the atmosphere [32].

Applications spanning a wide range of fields (Image processing, robot navigation, partial differential equations, etc.) are forthcoming in a companion paper [42].

The paper is organized as follow. Section 2 introduces our framework and convergence results, Section 3 presents the LLT, and Section 4 the PE algorithms. We introduce our new algorithm in Section 5, validate our complexity results in Section 6, and conclude the paper in Section 7.

2 The discrete Moreau envelope

Let us fix our notations. Our objective is the numerical computation of the discrete Moreau envelope

$$M_{\lambda,X}(s) := \min_{x\in X}\left[\frac{\|s-x\|^2}{2\lambda} + f(x)\right], \tag{4}$$

at many points $s \in S \subset \mathbb{R}^d$, knowing the values of f on the set X. Both sets S and X are discrete sets in $\mathbb{R}^d$ containing m and n points respectively. A brute force computation of the Moreau envelope involves computing a minimum over n points x for m values of the parameters s. Hence, it has a quadratic worst-case time complexity. Our objective is to investigate algorithms that produce the same results with a linear worst-case time complexity.

We name the set of minima in (4) the discrete proximal mapping

$$P_{\lambda,X}(s) := \operatorname*{Argmin}_{x\in X}\left[\frac{\|s-x\|^2}{2\lambda} + f(x)\right]. \tag{5}$$

When there is no ambiguity, we will drop the index λ and X.

To obtain more efficient algorithms, we are going to restrict the set X we consider. The Moreau envelope (1) can always be factored as d one-dimensional envelopes. However, we need to restrict X to be a grid $X = X_1 \times \cdots \times X_d \subset \mathbb{R}^d$ to take advantage of that property. In that case, we have for $x = (x_1, \ldots, x_d)$ and $s = (s_1, \ldots, s_d)$

$$M_{\lambda,X}(s) = \inf_{x_1\in X_1}\left[\frac{|s_1-x_1|^2}{2\lambda} + \cdots + \inf_{x_d\in X_d}\left[\frac{|s_d-x_d|^2}{2\lambda} + f(x)\right]\ldots\right], \tag{6}$$

and a fast one-dimensional algorithm will give a fast d-dimensional algorithm by applying it repeatedly. So for computation on a grid, we can restrict ourselves to computing the Moreau envelope for functions of one variable. Consequently, in all the remainder of the paper and unless otherwise specified, X will be a discrete subset in $\mathbb{R}$.

For numerical computation purposes, we can ignore the parameter λ since

$$M_{\lambda,X}(s) = \frac{1}{2\lambda}\min_{x\in X}\left[\|s-x\|^2 + 2\lambda f(x)\right].$$

So an algorithm computing $M_{1/2}$ for a function f also computes M_λ when applied to the function $2\lambda f$ *i.e.* $M_\lambda f = \frac{1}{2\lambda}M_{1/2}2\lambda f$.

Remark 1 If we restrict ourselves to computing the discrete Moreau envelope on the same grid the function f is sampled (*i.e.* take $S = X$) and of using a regularly sampled grid, then it is not restrictive to assume the grid X is $\{1, \ldots, n\}$. Under these assumptions we have

$$M_j := M_\lambda(x_j) = \frac{h^2}{2\lambda}\min_i\left[\|j-i\|^2 + \frac{2\lambda}{h^2}f(x_0+ih)\right]$$

where $x_i = x_0 + ih$ for $1 \le i \le n$ and $h > 0$ is the stepsize of the grid. While we will not restrict ourselves to regular grids, the above setting is similar to the distance transform framework in image processing and allows for a simpler implementation (if not faster due to the restriction of some operations to integer arithmetic), see [43].

The discrete Moreau envelope converges to the Moreau envelope, and the smoother the function f, the faster the convergence. To state convergence results, we consider an interval $[a, b] \subset \mathbb{R}$ with $a < b$, and name $M_{[a,b]}$ the Moreau envelope of the function $f + I_{[a,b]}$ *i.e.*

$$M_{[a,b]} = M_\lambda(f + I_{[a,b]}),$$

where $I_{[a,b]}$ is the indicator function of the interval $[a, b]$: $I_{[a,b]}(x) = 0$ when $x \in [a, b]$ and $+\infty$ otherwise. We consider a grid $X \subset [a, b]$ with n regularly spaced points and note M_n the discrete Moreau envelope: $M_n = M_{\lambda,X}$.

We will say a function $f : \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$ is proper if there is $x \in \mathbb{R}^d$ for which $f(x) < +\infty$.

Proposition 1 *Assume* $f : \mathbb{R} \to \mathbb{R} \cup \{+\infty\}$ *is proper.*

(1) *If f is upper semi-continuous on* $[a, b]$*, M_n converges pointwise to* $M_{[a,b]}$.
(2) *If f is continuous on* $[a.b]$*, M_n converges uniformly on* $[a.b]$ *to* $M_{[a,b]}$.
(3) *If f is Lipschitz continuous on* $[a, b]$*,*

$$\| M_{[a,b]} - M_n \|_{L^\infty([a,b])} \le \frac{c_1}{n}$$

where c_1 is a constant depending only on the Lipschitz constant of f, on a, and on b.

(4) *If f is C^2 on a neighborhood of* $[a, b]$*,*

$$\| M_{[a,b]} - M_n \|_{L^\infty([a,b])} \le \frac{c_2}{n^2}$$

where c_2 is a constant depending only on $\max_{[a,b]} f''$*, and on $b - a$.*

(5) *The convergence results above also hold when f is defined in* $\mathbb{R}^d$.

Proof Use Formula (7) below with the convergence results for the discrete Legendre transform contained in [14, 40, 41]. □

Since the previous proposition only guarantees the convergence of $M_\lambda f$ to $M_\lambda(f + I_{[a,b]})$, to obtain the convergence on an unbounded domain involves taking a set $[a, b]$ large enough to contain $P_\lambda(s)$ for all s we want to approximate the Moreau envelope at. The next result states that for a grid large enough, the Moreau envelope of the function f is equal to the Moreau envelope of the function $f + I_{[a,b]}$ for $b - a$ large enough.

Proposition 2 *The following equivalence holds for any $s \in \mathbb{R}$, and any $a > 0$*

$$\partial f^*(s) \cap [-a, a] \neq \emptyset \Leftrightarrow M_{[-a,a]} = M.$$

Proof The full proof is contained in [27]. See also [41, Proposition 2] for the same statement in the context of discrete Legendre transforms. □

3 The Linear-time Legendre Transform algorithm (LLT)

The computation of the Moreau envelope is equivalent to the computation of the Legendre–Fenchel conjugate as the following proposition shows.

Proposition 3 (See [71, Example 11.26]) *Let f be a function defined on $\mathbb{R}^n$, and assume $\lambda > 0$. Then*

$$M_\lambda(s) = \frac{\|s\|^2}{2\lambda} - \frac{1}{\lambda} g_\lambda^*(s). \tag{7}$$

where $g_\lambda(x) := \|x\|^2/2 + \lambda f(x)$.

Proof Expand the quadratic form to obtain

$$\begin{aligned} M_\lambda(s) &:= \inf_{x\in\mathbb{R}} \left[\frac{\|s-x\|^2}{2\lambda} + f(x) \right], \\ &= \frac{\|s\|^2}{2\lambda} + \inf_{x\in\mathbb{R}} \left[-\frac{1}{\lambda} \left(\langle s, x\rangle - \left(\frac{\|x\|^2}{2} + \lambda f(x) \right) \right) \right], \\ &= \frac{\|s\|^2}{2\lambda} - \frac{1}{\lambda} \sup_{x\in\mathbb{R}} [\langle s, x\rangle - g_\lambda(x)], \end{aligned}$$

and use the definition of g_λ to obtain (7). □

Remark 2 The Moreau–Yosida approximate of a nonconvex function may not be convex. In fact, Formula (7) shows it is a difference of convex functions. Nevertheless we can use (7) with discrete Legendre–Fenchel transform algorithms such as [41] to compute the Moreau envelope of even nonconvex functions f in linear time.

Remark 3 The equivalence between the computation of the Legendre–Fenchel conjugate and the Moreau envelope has been noted by several authors [22, 64, 71, 72].

Remark 4 As Formula (7) only involves computing one Legendre–Fenchel conjugate, it is more efficient than the suggestions in [40, 41] to use the fact that for lower semi-continuous convex functions $F_\lambda = (f^* + \frac{\lambda}{2}\|.\|^2)^*$.

For the sake of completion we recall the Linear-time Legendre transform algorithm (full details are presented in [41]). A similar factorization as Formula (6) holds for the Legendre–Fenchel conjugate, so without loss of generality we can assume the sets X_n, and S_m are subsets of $\mathbb{R}$. (Figure 1) shows the algorithm.

The key step is the explicit computation of the convex hull. Using the fact that the set X_n is already sorted, the convex hull is computed in linear time. The resulting vertices give rise to an increasing sequence C, since the derivative of a convex function is always non-decreasing. Then computing the argmax amounts to merging the two increasing sequences C and S.

In higher dimension, the algorithm computes several one-dimensional conjugate. For example, for a function f of two variables, the LLT algorithm first computes the conjugate for each row, then for each column. The resulting complexity in dimension d is $O(dN)$ where N is the number of points on the grid.

Fig. 1 LLT algorithm

Input: The sets X_n, Y_n, and S_m where $Y_n[i]$ is an approximation of $f(X_n[i])$.
Output: The set Z_m where $Z_m[j]$ is an approximation of $f^*(S_m[j])$.
Step 1: Compute the (lower) convex hull of the points $(X[i], Y[i])$ using the Beneath-Beyond convex hull algorithm. Rename the sequence X and Y to be the resulting vertices.
Step 2: Compute the set of slopes $C[i] := (Y[i+1] - Y[i])/(X[i+1] - X[i])$ and for each $S[j]$ find the index i such that $C[i] \leq S[j] < C[i+1]$. Then $X[i]$ is a point where the maximum is attained and $Z[j] = S[j] * X[i] - Y[i]$.

4 The Parabolic Envelope algorithm (PE)

Felzenszwalb and Huttenlocher [18] proposed a linear-time algorithm to compute the Moreau envelope. Their algorithm uses Formula (6) to reduce the computation to one dimension. Then they compute the lower envelope of parabolas by noting that the intersection of two parabolas can be computed in constant time. In this paper, their algorithm will be named PE for parabolic envelope. It is summarized in (Fig. 2).

In Step 1, each parabola is added once and may be deleted at most once. Since adding or deleting a parabola is done in constant time and we consider n parabolas, Step 1 executes in $O(n)$. Step 2 takes $O(m)$ to give a $O(n+m)$ worst-case time complexity for a function of one variable.

Felzenszwalb and Huttenlocher further note that in dimension d the complexity becomes $O(dN)$ where N is the number of points in the grid ($N = n+m$ in the algorithm description).

Remark 5 The idea to compute a parabolic envelope is also contained in [15, 16].

5 The NonExpansive Proximal mapping algorithm (NEP)

The regularity of the proximal mapping allows the development of faster algorithms. Let us first recall the following properties of the proximal mapping.

Proposition 4 *Assume f is a proper lower semi-continuous function and $\lambda > 0$. Then the proximal mapping P_λ is monotone.*

If f is also convex, then P_λ is maximal monotone and non-expansive (hence single-valued).

Proof See [71, Proposition 12.19].

We give here a simple proof for the convex case.

Name ∂f the subdifferential in the sense of convex analysis

$$\partial f(x) := \{s \in \mathbb{R}^d : f(y) \geq f(x) + \langle s, y - x \rangle \text{ for all } y\}$$

Fig. 2 PE algorithm

Input: The sets X_n, Y_n, and S_m where $Y_n[i]$ is an approximation of $f(X_n[i])$.
Output: The set Z_m where $Z_m[j]$ is an approximation of $f^*(S_m[j])$.
Step 1: Compute the (lower) envelope of the n parabolas.
Step 2: Fill in Z_m.

and apply [71, Example 10.2] to obtain

$$\frac{x - p(x)}{\lambda} \in \partial f(p(x))$$

where $p(x) \in P_\lambda(x)$ is a selection in P_λ.

Note that although [71, Example 10.2] is stated with the general subgradient (which is the limit of regular subgradients), it still holds with the subdifferential in the sense of convex analysis since both subgradients are equal in the convex case.

For any two points x and x', we have

$$\frac{x - p(x)}{\lambda} \in \partial f(p(x)), \text{ and } \frac{x' - p(x')}{\lambda} \in \partial f(p(x')).$$

Now use the monotonicity of the subdifferential to obtain

$$\left\langle \frac{x - p(x)}{\lambda} - \frac{x' - p(x')}{\lambda}, p(x) - p(x') \right\rangle \geq 0,$$

in other words

$$\langle x - x', p(x) - p(x') \rangle \geq \| p(x) - p(x') \|^2. \tag{8}$$

So p is *strongly monotone* for any proper function f. In particular, for univariate functions $(x - x')(p(x) - p(x') \geq 0$ *i.e.* p is increasing. □

Now, using the same scheme as in [14, 40, 64] we can build a log-linear $O((n + m)\ln(n + m))$ worst-case time algorithm to compute the Moreau envelope at m points, where n is the number of points at which we sample the function f. Of course, it is outperformed by the PE and the LLT linear-time algorithms.

However, in the convex case, we can build a very simple linear-time algorithm using the smoothness of the proximal mapping by carefully selecting grids as follow. Apply the Cauchy-Schwarz inequality to (8) to obtain

$$\| p(x) - p(x') \| \leq \| x - x' \|,$$

in other words, P_λ is non-expansive. So any selection p of the proximal mapping P_λ is 1-Lipschitz. Take two partitions, $x_1 < \cdots < x_n$ and $s_1 < \cdots < s_m$, and assume

$$x_{i+1} - x_i = s_{j+1} - s_j =: h$$

for any integer i, j with $1 \leq i \leq n - 1$ and $1 \leq j \leq m - 1$. The algorithm NEP described in (Fig. 3) computes the Moreau-Yosida approximate at all the point on the grid $(s_j)_j$ by approximating the infimum with the computation of the minimum on the grid $(x_i)_i$.

The complexity of the NEP algorithm is linear since Step 1 costs $O(n)$ and Step 2 runs in $O(m)$ (each $p(s_j)$ is computed in constant time for $j > 1$).

Fig. 3 NEP algorithm

Step 1: Compute $p(s_1)$ by a linear search.
Step 2: Compute a selection of the proximal mapping using

$$0 \leq p(s_{j+1}) - p(s_j) \leq h$$

so for each j, either $p(s_{j+1}) = p(s_j)$ or $p(s_{j+1}) = p(s_j) + h$.

Table 1 LLT and brute force (direct) algorithms for computing the discrete Legendre transform as included in lft.sci

Algorithms	Discrete Legendre transform functions
lft_llt	Compute the discrete Legendre transform using the LLT algorithm (*main function*).
lft_llt_d	Same as lft_llt but computation is performed on $\{1, \ldots, n\}$ for a function defined on $\{1, \ldots, n\}$.
bb	Compute the lower convex envelope of a set of points in the plane using the Beneath-Beyond algorithm. Assume the points are sorted along the x-axis.
lft_direct	Compute the discrete Legendre transform with a quadratic (brute force) algorithm (for comparison only).
lft_direct_d	Same as lft_direct on the grid $\{1, \ldots, n\}$ for a function defined on $\{1, \ldots, n\}$.
fusion	Merge two increasing sequences using a linear algorithm (internal function).
fusionsci	Merge two increasing sequences using Scilab syntax resulting in a fast but nonlinear algorithm (internal function).

Lemma 1 *The NEP algorithm computes the Moreau envelope in linear-time when the function f is convex.*

Proof We only need to prove that the algorithm computes the Moreau envelope. Since

$$0 \leq p(s_j) - p(s_{j-1}) \leq s_j - s_{j-1} \leq h$$

the only possibilities for $p(s_j) - p(s_{j-1})$ are either 0 or h. In the first case, $p(s_j) = p(s_{j-1})$ and in the second $p(s_j)$ is the successor of $p(s_{j-1})$ in the grid $x_1 < \cdots < x_n$. So the result is indeed a selection of the proximal mapping. □

Table 2 LLT, PE, NEP, and brute force algorithms for computing the discrete Moreau envelope as included in me.sci

Algorithms	Discrete Moreau envelope functions
	Main functions (one dimension)
me_llt	Compute the discrete Moreau envelope (LLT algorithm).
me_pe	Compute the discrete Moreau envelope (PE algorithm).
me_nep	Compute the discrete Moreau envelope (NEP algorithm).
	Main functions (two dimensions)
me_llt2d	Same as me_llt for a function of two variables.
me_pe2d	Same as me_pe for a function of two variables.
me_nep2d	Same as me_nep for a function of two variables.
	Functions provided for comparison only
me_direct	Compute the discrete Moreau envelope with a quadratic (brute force) algorithm.
me_direct2d	Same as me_direct for a function of two variables but uses separability to achieve a $O(N^{3/2})$ complexity.
me_brute2d	Compute the discrete Moreau envelope with a quadratic (brute force) algorithm for a function of two variables.

Table 3 Examples and unit test files

Script name	Function demonstrated and tested
test_bb.sci	bb
test_llt.sci	lft_llt, lft_llt_d, lft_direct, lft_direct_d
test_me_llt.sci	me_llt, me_llt2d, me_direct, me_direct2d
test_me_nep.sci	me_nep, me_nep2d, me_direct
test_me_pe.sci	me_pe, me_pe2d
test_fusion.sci	Fusion and fusionsci (internal functions, unit tests only).

6 Numerical validation

The CCA (Computational Convex Analysis) toolbox for Scilab 4.0 associated with the paper runs under both Ms Windows and Linux, and is freely available from www.netlib.org/numeralgo/na24. (Table 1) lists the functions available in the package for computing the discrete Legendre transform, (Table 2) lists the functions for computing the discrete Moreau envelope, and (Table 3) shows the unit tests and examples available. The package consists of two main files, lft.sci and me.sci, with several unit test files named test_$*$.sci called from test.sci. The unit test files serve a double purpose: To check the functions and to provide examples. Finally (Table 4) lists the demonstration files that also illustrate the use of the functions. (Due to space constraints, the result of some of the demonstrations were not included in the paper but are available in the package).

We now present numerical results that validate our complexity results. Computations were performed on a 1.8GHz Pentium 4 using Scilab v4.0.

(Figure 4) compares the three Moreau envelope algorithms with the direct algorithm to compute the function $f(x) = x^2$ on the grid $x_i = x_0 + ih$, with $x_0 = 1$,

Table 4 Demonstration files

Demo	Script name	Description
1	time_me.sci	Comparison of Moreau Envelope algorithms (LLT, PE, NEP, direct). Produces (Fig. 4).
2	abs.sci	Moreau Envelope and proximal mapping of the absolute value function. Produces (Figs. 6a–6b).
3	sqr.sci	Moreau Envelope and proximal mapping of the square function. Produces (Figs. 5a–5b).
4	nonconvexabs.sci	Moreau Envelope and proximal mapping of $f(x) = \|\|x\| - 1\|$. Produces (Figs. 7a–7b).
5	nonconvexsqr.sci	Moreau Envelope and proximal mapping of $f(x) = (x^2 - 1)^2$. Produces (Figs. 8a–8b).
6	bb.sci	Computes the lower convex envelope of the smooth nonconvex function $f(x) = (x^2 - 1)^2$ on the grid $X = \{-2, -1.5, \ldots, 2\}$.
7	lft_direct.sci	Illustrates the convergence of the discrete Legendre-Fenchel transform when the domain is enlarged.
8	time_lft.sci	Comparison of Legendre-Fenchel transform algorithms (lft_direct, lft_llt using fusion, and lft_llt using fusionsci).

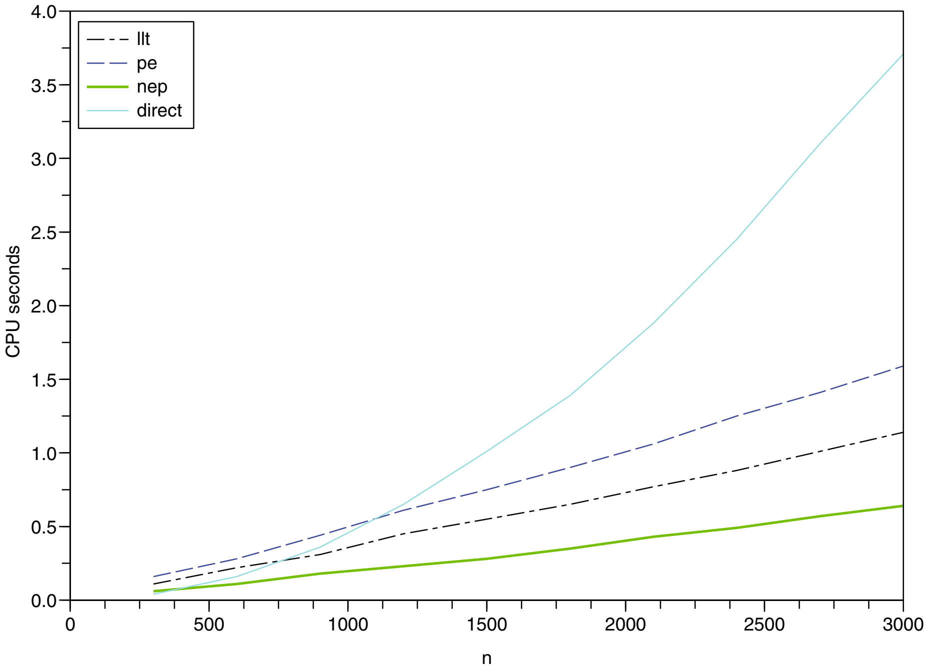
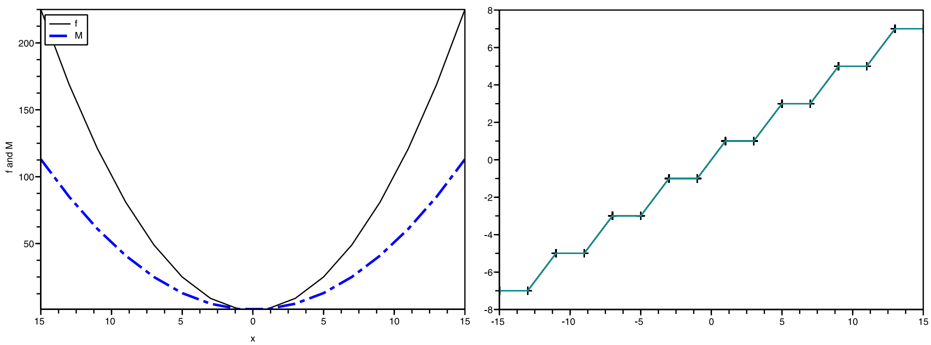


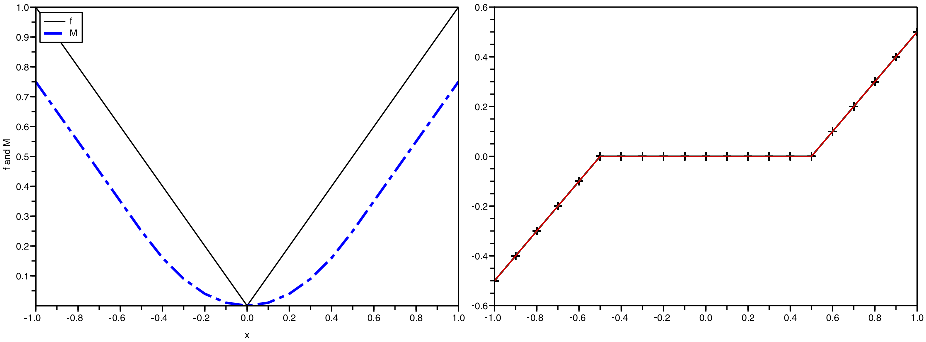
Fig. 4 Comparison of Fast Moreau Envelope Algorithms to compute the Moreau envelope of the function $f(x) = x^2$ on the grids $X = S = \{1, 2, \ldots, n\}$ with $n \in \{300, 600, \ldots, 3000\}$. Direct computation has a quadratic cost while all other algorithms run in linear time

$h = 1$, and $1 \leq i \leq n$. Clearly the fast algorithms perform better than the direct computation even taking into account Scilab optimization for matrix computation. In our implementation, NEP comes best, followed by LLT, PE, and then direct computation. Since NEP requires additional properties (it only works for convex



(a) The function f and its Moreau envelope (b) Proximal mapping

Fig. 5 Moreau envelope and Proximal mapping of the smooth convex function $f(x) = x^2$ on the grids $X = S = \{-15, -13, \ldots, 15\}$



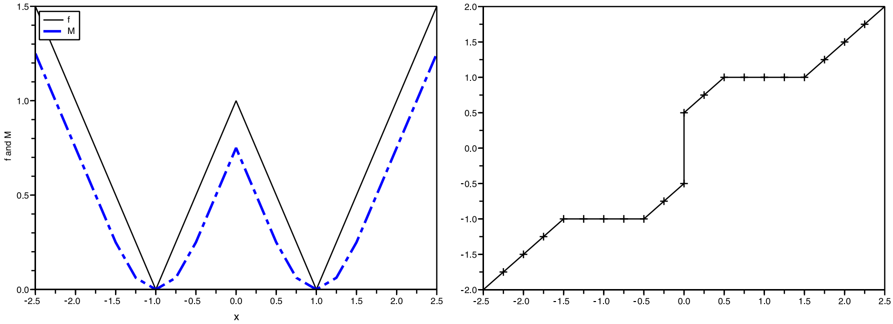
(a) The function f and its Moreau envelope (b) Proximal mapping

Fig. 6 Moreau envelope and Proximal mapping of the nonsmooth convex function $f(x) = |x|$ on the grids $X = S = \{-1, -0.9, \ldots, 1\}$

data), it is not surprising it performs better. (Note that we could have further tuned the LLT algorithm by skipping the computation of the convex hull since the function is convex.) The demo script time_me.sci (Option *1. Comparison of ME algorithms* in the CCA demos menu) generates (Fig. 4); see demo 1 of (Table 4).

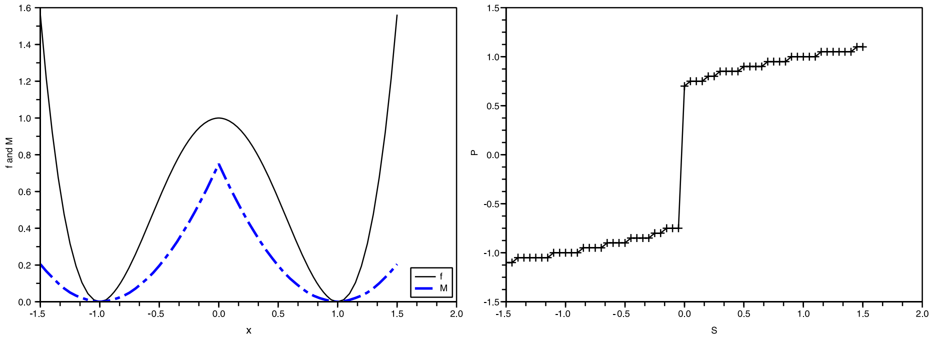
(Figure 5a) shows the smooth convex function $f(x) = x^2$ and its Moreau envelope while (Fig. 5b) shows the corresponding proximal mapping. The nonexpansiveness of the proximal mapping translates graphically as a step function where the height of the step cannot be greater than its width. Both figures were generated using the demo script sqr.sci (Option *3. ME of the square function* in the CCA demos menu; see demo 3 of (Table 4)).

(Figure 6) illustrates the regularization property of the Moreau envelope for the nonsmooth convex function $f(x) = |x|$. While the kink at the minimum is smoothed, the Moreau envelope and the function still share the same minimum, which corresponds to the fixed point of the proximal mapping shown in (Fig. 6b). The demo script abs.sci (Option *2. ME of the abs function* in the CCA demos menu; see demo 2 of (Table 4)) generates both figures.



(a) The function f and its Moreau envelope (b) Proximal mapping

Fig. 7 Moreau envelope and Proximal mapping of the nonsmooth nonconvex function $f(x) = ||x| - 1|$ on the grids $X = S = \{-2.5, -2.25, \ldots, 2.5\}$



(a) The function f and its Moreau envelope (b) Proximal mapping

Fig. 8 Moreau envelope and Proximal mapping of the smooth convex function $f(x) = (x^2 - 1)^2$ on the grids $X = S = \{-1.5, -1.45, \ldots, 1.5\}$

The last two examples illustrate nonconvex functions and the failure of the NEP algorithm. (Figure 7a) shows the graph of the function $f(x) = ||x| - 1|$ and its Moreau envelope. (Figure 7b) shows the associated proximal mapping. (both are generated by the demo script nonconvexabs.sci, Option *4. ME of a nonsmooth nonconvex function* in the CCA demos menu; see demo 4 of (Table 4). The jump at the origin indicates that the proximal mapping is not nonexpansive, so the NEP algorithm cannot be applied. Note that the jump is not related to the lack of regularity: (Fig. 8b) illustrates a similar jump for the smooth function $f(x) = (x^2 - 1)^2$ and (Fig. 8a) shows the corresponding Moreau envelope. Both figures are created by the demo script nonconvexsqr.sci (Option *5. ME of a smooth nonconvex function* in the CCA demos menu; see demo 5 of (Table 4)).

7 Conclusion

The principles used to build the fast algorithms presented can be applied in different settings.

The PE algorithm idea is to build the lower envelope of parabolas. It achieves a linear-time complexity because computing the intersection of two parabolas can be done in constant time $\Theta(1)$. The same principle applies to building the envelope of any family of functions provided the intersection of two functions can be computed in linear time. For example, computing the discrete Legendre transform using that principle amounts to computing the lower envelope of affine functions, which is essentially the same as applying the Beneath-Beyond algorithm.

The LLT algorithm relies on convexity. Namely it uses the Beneath-Beyond algorithm to compute the vertices of the lower convex envelope, and achieves linear time since the vertices are naturally sorted along one axis. Any algorithm benefiting from convexity can use the Beneath-Beyond algorithm as a pre-processing step to improve its speed.

Finally, the NEP algorithm principle is to select a regular grid and use the nonexpansiveness property of the proximal mapping. Any transform with a nonexpansive argmax can use the same strategy to achieve linear time.

Acknowledgements The author would like to thank Jeff Dicker and Andrew Tonner for implementing the initial version of the algorithms. The associated software package greatly benefited from several suggestions from the software editor. This work was partly supported by the Okanagan University College Grant in Aid program, and by the author NSERC Discovery grant. Part of this work begun while the author was a Postdoctoral Fellow at the Pacific Institute for the Mathematical Sciences (PIMS).

References

1. Attouch, H., Wets, R.J.-B.: A convergence theory for saddle functions. Trans. Am. Math. Soc. **280**, 1–41 (1983)
2. Aurell, E., Gurbatov, S.-G. N., Simdyankin, S.-G.I.: Numerical proof of self-similarity in Burgers' turbulence. Tech. Rep. http://arxiv.org/abs/patt-sol/9602005,arXiv.org (1996)
3. Bec, J., Frisch, U., Khanin, K.: Kicked Burger turbulence. J. Fluid Mech. **416**, 239–267 (2000)
4. Bernard, F., Thibault, L.: Prox-regularity of functions and sets in Banach spaces. Set-Valued Anal. **12**, 25–47 (2004)
5. Bernard, F., Thibault, L.: Prox-regular functions in Hilbert spaces. J. Math. Anal. Appl. **303**, 1–14 (2005)
6. Bernard, F., Thibault, L.: Uniform prox-regularity of functions and epigraphs in Hilbert spaces. Nonlinear Anal. **60**, 187–207 (2005)
7. Bonnans, J.F., Gilbert, J.C., Lemaréchal, C., Sagastizábal, C.A.: A family of variable metric proximal methods. Math. Program. **68**, 15–47 (1995)
8. Brenier, Y.: Un algorithme rapide pour le calcul de transformées de Legendre–Fenchel discrètes. C. R. Acad. Sci. Paris Sér. I Math. **308**, 587–589 (1989)
9. Burke, J.V., Qian, M.: A variable metric proximal point algorithm for monotone operators. Tech. Rep., Department of Mathematics, University of Washington, Seattle, Washington; and Department of Mathematics, California State University, Fullerton, CA (1992)
10. Burke, J.V., Qian, M.: On the superlinear convergence of the variable metric proximal point algorithm using Broyden and BFGS matrix secant updating. Tech. Rep., Department of Mathematics, University of Washington, Seattle, Washington; and Department of Mathematics, California State University, Fullerton, CA (1996)
11. Burke, J.V., Qian, M.: On the local super-linear convergence of a matrix secant implementation of the variable metric proximal point algorithm for monotone operators. Tech. Rep., Department of Mathematics, University of Washington, Seattle, Washington; and Department of Mathematics, California State University, Fullerton, CA (1998)
12. Clarke, F.H., Stern, R.J., Wolenski, P.R.: Proximal smoothness and the lower-C^2 property. J. Convex Anal. **2**, 117–144 (1995)
13. Cominetti, R.: Coupling the proximal point algorithm with approximation methods. J. Optim. Theory Appl. **95**, 581–600 (1997)
14. Corrias, L.: Fast Legendre–Fenchel transform and applications to Hamilton–Jacobi equations and conservation laws. SIAM J. Numer. Anal. **33**, 1534–1558 (1996)
15. Deniau, L.: Proposition d'un opérateur géométrique pour l'analyse et l'identification de signaux et images. PhD thesis, Université de Paris-Sud, Centre d'Orsay (Dec. 1997)
16. Deniau, L., Blanc-Talon, J.: Fractal analysis with Hausdorff distance under affine transformations, Tech. Rep., ETCA-CREA-SP (1995)
17. Eckstein, J.: Nonlinear proximal point algorithms using Bregman functions, with applications to convex programming. Math. Oper. Res. **18**, 202–226 (1993)
18. Felzenszwalb, P.F., Huttenlocher, D.P.: Distance transforms of sampled functions, Tech. Rep. TR2004-1963, Cornell Computing and Information Science (Sept. 2004)
19. Frisch, U., Bec, J.: Burgulence. In: Lesieur, A.M., David, F. (eds.) Les Houches 2000: New Trends in Turbulence, pp. 341–383. Springer EDP-Sciences, France (2001)
20. Frisch, U., Bec, J., Villone, B.: Singularities and the distribution of density in the Burgers/adhesion model. Physica D, **152–153**, 620–635 (2001)
21. Güler, O.: On the convergence of the proximal point algorithm for convex minimization. SIAM J. Control Optim. **29**, 403–419 (1991)
22. Gurbatov, S.N., Simdyankin, S.I., Aurell, E., Frisch, U., Tóth, G.: On the decay of Burgers turbulence. J. Fluid Mech. **344**, 339–374 (1997)

23. Gurbatov, S.N., Troussov, A.V.: The decay of multiscale signals – deterministic model of the Burgers turbulence, Tech. Rep., arXiv.org (2000)
24. Hamdi, A.: A Moreau-Yosida regularization of a difference of two convex functions. Appl. Math. E-Notes **5**, 164–170 (2005) (electronic)
25. Hare, W.L., Sagastizabal, C.: Computing proximal points on nonconvex functions, Tech. Rep., Optimization Online Digest – June 2005 (2005)
26. Helluy, P.: Simulation numérique des écoulements multiphasiques : De la théorie aux applications, PhD thesis, Institut des Sciences de l'Ingenieur de Toulon et du Var, Laboratoire Modélisation Numérique et Couplages, BP 56, 83162 La Valette CEDEX, France Habilitation à Diriger des Recherches (Jan. 2005)
27. Hiriart-Urruty, J.B.: Lipschitz r-continuity of the approximate subdifferential of a convex function. Math. Scand. **47**, 123–134 (1980)
28. Hiriart-Urruty, J.B., Lemaréchal, C.: Convex analysis and minimization algorithms. In: Grundlehren der Mathematischen Wissenschaften, vol. 305–306. [Fundamental Principles of Mathematical Sciences] Springer, Berlin Heidelberg New York (1993) (Vol. I: Fundamentals, Vol. II: Advanced theory and bundle methods)
29. Hisakado, T., Okumura, K., Vukadinovic, V., Trajkovic, L.: Characterization of a simple communication network using Legendre transform. In: Proc. IEEE Int. Symp. Circuits and Systems, vol. 3, pp. 738–741 (May 2003)
30. Iusem, A.N., Svaiter, B.F., Teboulle, M.: Entropy-like proximal methods in convex programming. Math. Oper. Res. **19**, 790–814 (1994)
31. Koopen, B.: Contact of bodies in 2D-space: Implementing the Discrete Legendre Transform. AI Master's thesis, Intelligent Autonomous Systems Group, University of Amsterdam (Feb. 2002)
32. Legras, B., Pisso, I., Berthet, G., Lefèvre, F.: Variability of the Lagrangian turbulent diffusion in the lower stratosphere. Atmos. Chem. Phys. **5**, 1605–1622 (2005)
33. Lehdili, N., Moudafi, A.: Combining the proximal algorithm and Tikhonov regularization. Optimization **37**, 239–252 (1996)
34. Lemaire, B.: The proximal algorithm. In: New Methods in Optimization and Their Industrial Uses (Pau/Paris, 1987) Internat, vol. 87, pp. 73–87. Schriftenreihe Numer. Math., Birkhäuser, Basel (1989)
35. Lemaréchal, C., Oustry, F.: Growth conditions and $\mathcal{U}$-Lagrangians. Set-Valued Anal. **9**, 123–129 (2001) (Wellposedness in optimization and related topics (Gargnano, 1999))
36. Lemaréchal, C., Oustry, F., Sagastizábal, C.: The $\mathcal{U}$-Lagrangian of a convex function. Trans. Am. Math. Soc. **352**, 711–729 (2000)
37. Lemaréchal, C., Sagastizábal, C.: More than first-order developments of convex functions: primal-dual relations. J. Convex Anal. **3**, 255–268 (1996)
38. Lemaréchal, C., Sagastizábal, C.: Practical aspects of the Moreau–Yosida regularization: theoretical preliminaries. SIAM J. Optim. **7**, 367–385 (1997)
39. Lemaréchal, C., Sagastizábal, C.: Variable metric bundle methods: from conceptual to implementable forms. Math. Program. **76**, 393–410 (1997)
40. Lucet, Y.: A fast computational algorithm for the Legendre–Fenchel transform. Comput. Optim. Appl. **6**, 27–57 (1996)
41. Lucet, Y.: Faster than the Fast Legendre Transform, the Linear-time Legendre Transform. Numer. Algor. **16**, 171–185 (1997)
42. Lucet, Y.: Fast Moreau envelope computation II: Applications. Tech. Rep., University of British Columbia Okanagan (2005)
43. Lucet, Y.: A linear Euclidean distance transform algorithm based on the Linear-time Legendre Transform. In: Proceedings of the Second Canadian Conference on Computer and Robot Vision (CRV 2005), Victoria BC, May 2005. IEEE Computer Society Press, Los Alamitos, CA
44. Meng, F., Sun, D., Zhao, G.: Semismoothness of solutions to generalized equations and the Moreau–Yosida regularization. Math. Program. (2005) (online first)
45. Meng, F., Zhao, G.: On second-order properties of the Moreau–Yosida regularization for constrained nonsmooth convex programs. Numer. Funct. Anal. Optim. **25**, 515–529 (2004)
46. Meng, F.W., Hao, Y.: Piecewise smoothness for Moreau–Yosida approximation to a piecewise C^2 convex function. Adv. Math. (China), **30**, 354–358 (2001)
47. Mifflin, R.: A quasi-second-order proximal bundle algorithm. Math. Program. **73**, 51–72 (1996)
48. Mifflin, R., Qi, L., Sun, D.: Properties of the Moreau-Yosida regularization of a piecewise C^2 convex function. Math. Program. **84**, 269–281 (1999)
49. Mifflin, R., Sagastizábal, C.: $\mathcal{VU}$-decomposition derivatives for convex max-functions. In: Ill-posed Variational Problems and Regularization Techniques (Trier, 1998). Lecture Notes in

Economics and Mathematical Systems, vol. 477, pp. 167–186. Springer, Berlin Heidelberg New York (1999)
50. Mifflin, R., Sagastizábal, C.: Functions with primal-dual gradient structure and $\mathcal{U}$-Hessians. In: Nonlinear Optimization and Related Topics (Erice, 1998). Applied Optimization, vol. 36, pp. 219–233. Kluwer, Dordrecht (2000)
51. Mifflin, R., Sagastizábal, C.: On $\mathcal{VU}$-theory for functions with primal-dual gradient structure, SIAM J. Optim. **11**, 547–571 (2000) (electronic)
52. Mifflin, R., Sagastizábal, C.: Primal-dual gradient structured functions: second-order results; links to epi-derivatives and partly smooth functions. SIAM J. Optim. **13**, 1174–1194 (2003) (electronic)
53. Mifflin, R., Sagastizábal, C.: On the relation between $\mathcal{U}$-Hessians and second-order epi-derivatives. Eur. J. Oper. Res. **157**, 28–38 (2004)
54. Mifflin, R., Sagastizábal, C.: $\mathcal{VU}$-smoothness and proximal point results for some nonconvex functions. Optim. Methods Softw. **19**, 463–478 (2004)
55. Mifflin, R., Sagastizábal, C.: Relating $\mathcal{U}$-Lagrangians to second-order epi-derivatives and proximal-tracks. J. Convex Anal. **12**, 81–93 (2005)
56. Mifflin, R., Sagastizábal, C.: A $\mathcal{VU}$-algorithm for convex minimization. Math. Program. (2005)
57. Mifflin, R., Sun, D., Qi, L.: Quasi-Newton bundle-type methods for nondifferentiable convex optimization. SIAM J. Optim. **8**, 583–603 (1998) (electronic)
58. Moreau, J.-J.: Propriétés des applications “prox”. C. R. Acad. Sci. Paris **256**, 1069–1071 (1963)
59. Moreau, J.-J.: Proximité et dualité dans un espace Hilbertien. Bull. Soc. Math. France **93**, 273–299 (1965)
60. Moreau, J.-J.: Convexity and duality. In: Functional Analysis and Optimization, pp. 145–169. Academic, New York (1966)
61. Moudafi, A.: Coupling proximal methods and variational convergence. Z. Oper. Res. **38**, 269–280 (1993)
62. Moudafi, A.: Coupling proximal algorithm and Tikhonov method. Nonlinear Times Digest **1**, 203–209 (1994)
63. Noullez, A., Gurbatov, S.N., Aurell, E., Simdyankin, S.I.: The global picture of self-similar and not self-similar decay in Burgers turbulence. Tech. Rep. nlin.CD/0409022, arXiv.org eprint archive, (Sept. 2004)
64. Noullez, A., Vergassola, M.: A fast Legendre transform algorithm and applications to the adhesion model. J. Sci. Comput. **9**, 259–281 (1994)
65. Poliquin, R., Rockafellar, R.T.: Generalized Hessian properties of regularized nonsmooth functions. SIAM J. Optim. **6**, 1121–1137 (1996)
66. Poliquin, R., Rockafellar, R.T.: Prox-regular functions in variational analysis. Trans. Am. Math. Soc. **348**, 1805–1838 (1996)
67. Qi, L.: Second-order analysis of the Moreau-Yosida regularization. In: Nonlinear analysis and convex analysis (Niigata, 1998), pp. 16–25. World Sci., River Edge, NJ (1999)
68. Qi, L., Chen, X.: A preconditioning proximal Newton method for nondifferentiable convex optimization. Math. Program. **76**, 411–429 (1997)
69. Rockafellar, R.T.: Convex Analysis. Princeton University Press, Princeton, NY (1970)
70. Rockafellar, R.T.: Monotone operators and the proximal point algorithm. SIAM J. Control Optim. **14**, 877–898 (1976)
71. Rockafellar, R.T., Wets, R.J.-B.: Variational Analysis. Springer, Berlin Heidelberg New York (1998)
72. She, Z.-S., Aurell, E., Frisch, U.: The inviscid Burgers equation with initial data of Brownian type. Comm. Math. Phys. **148**, 623–641 (1992)
73. Spingarn, J.E.: Submonotone mappings and the proximal point algorithm. Numer. Funct. Anal. Optim. **4**, 123–150 (1981–1982)
74. Teboulle, M.: Entropic proximal mappings with applications to nonlinear programming. Math. Oper. Res. **17**, 670–690 (1992)
75. Vergassola, M., Dubrulle, B., Frisch, U., Noullez, A.: Burger’s equation, devil’s staircases and the mass distribution function for large-scale structures. Astron. Astrophys. **289**, 325–356 (1994)
76. Wei, Z., Qi, L.: Convergence analysis of a proximal Newton method. Numer. Funct. Anal. Optim. **17**, 463–472 (1996)
77. Yosida, K.: Functional Analysis, Classics in Mathematics. Springer, Berlin Heidelberg New York (1995) (Reprint of the sixth (1980) edition)