Infimal Convolution Regularizations with Discrete ℓ_1-type Functionals

S. Setzer, G. Steidl and T. Teuber *

Dedicated to Prof. Dr. Lothar Berg on the occasion of his 80th birthday

August 17, 2010

Abstract

As first demonstrated by Chambolle and Lions the staircasing effect of the Rudin-Osher-Fatemi model can be reduced by using infimal convolutions of functionals containing higher order derivatives. In this paper, we examine a modification of such infimal convolutions in a general discrete setting. For the special case of finite difference matrices, we show the relation of our approach to the continuous total generalized variation approach recently developed by Bredies, Kunisch and Pock. We present splitting methods to compute the minimizers of the ℓ_2^2 - (modified) infimal convolution functionals which are superior to previously applied second order cone programming methods. Moreover, we illustrate the differences between the ordinary and the modified infimal convolution approach by numerical examples.

1 Introduction

It is well-known that the staircasing effect visible in the minimizer of the Rudin-Osher-Fatemi (ROF) model [29]

$$\operatorname*{argmin}_{u \in L_2} \Big\{ \frac{1}{2} \|f-u\|_{L_2}^2 + \alpha |u|_{BV} \Big\}, \qquad \alpha > 0$$

with the semi-norm

$$\begin{aligned} |u|_{BV} &:= \sup_{V \in C_0^1, \|V\|_\infty \le 1} \int_\Omega u \operatorname{div} V \, dx \\ &= \int_\Omega |\nabla u| \, dx \qquad \text{if } u \text{ and its weak first derivatives are in } L_1(\Omega) \end{aligned}$$

for denoising images $f : \Omega \to \mathbb{R}$ corrupted by white Gaussian noise can be reduced by incorporating higher order derivatives into the functional. One successful approach in this direction was given by Chambolle and Lions in [8] who suggested to use the infimal convolution of functionals with first and second order derivatives as regularizer, i.e.,

$$\inf_{u_1 + u_2 = u} \int_\Omega \alpha_1 |\nabla u_1| + \alpha_2 |\nabla(\nabla u_2)| \, dx.$$

*University of Mannheim, Dept. of Mathematics and Computer Science

An alternative approach with $|\triangle u_2|$ instead of $|\nabla(\nabla u_2)|$ was given in [10]. For various other variational and PDE approaches involving higher order derivatives see [11, 15, 22, 24, 30, 41, 43]. Among these approaches we only mention that instead of infimal convolutions also functionals of the form $\Phi(u) = \sum_{i=1}^m \Phi_i(u)$ were proposed, see, e.g., [15, 25]. In one dimension, the difference between the minimizers of the functionals

$$\frac{1}{2}\|f-u\|_{L_2}^2 + (\Phi_1 \Box \Phi_2)(u) \quad \text{and} \quad \frac{1}{2}\|f-u\|_{L_2}^2 + (\Phi_1 + \Phi_2)(u) \tag{1}$$

with $\Phi_1(u) := \alpha_1 \int_\Omega |u'(x)|\,dx$ and $\Phi_2(u) := \alpha_2 \int_\Omega |u''(x)|\,dx$ is shown in Fig. 1. The advantages of the infimal convolution regularization are clearly visible. Finally, note that infimal convolutions with other operators than derivatives were applied, e.g., for image decomposition in [1, 2, 36].

In [34], we have applied a *modified infimal convolution* (MIC) regularization with first and second order derivatives just for some computational reasons related to second order cone programming. In general this modification leads to better numerical results than the original one by Chambolle and Lions. We have also generalized our model to tensor-valued images in [35]. Recently, this MIC approach was given a theoretical fundament (in the continuous setting for derivatives of arbitrary order) by Bredies, Kunisch and Pock [5] based on tensor algebra. The corresponding regularizer was called *total generalized variation* (TGV). For other generalizations of TV we refer to [31].

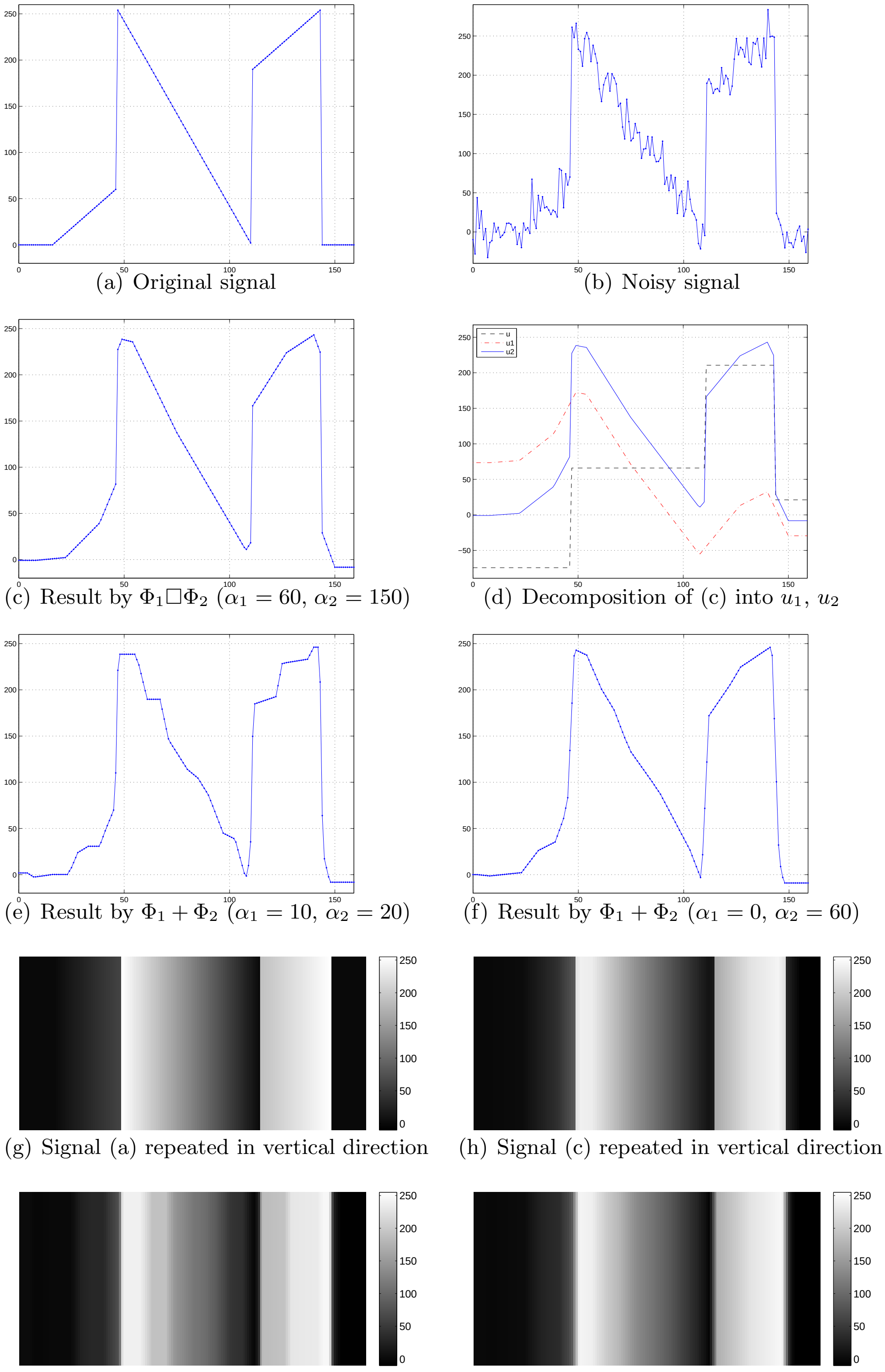
In this paper, we examine more general MIC functionals than in [34] in a discrete setting. These functionals combine ℓ_1-type norms with linear operators fulfilling some general factorization properties. The modifications of the ordinary infimal convolution appear by tightening the constraints on the dual variable. The corresponding primal problem contains a modified infimal convolution regularizer with some additional variables related to the linear operators. We propose an alternating direction method of multipliers and a primal-dual hybrid gradient algorithm to compute the minimizers of the functionals as well as some important intermediate values which are helpful to interpret the overall results. We show that this method can beat second order cone programming used in [34] significantly in terms of computational time.

This paper is organized as follows: In Section 2, we recall properties of infimal convolutions and consider minimization problems with ℓ_2^2 data fitting term and special ℓ_1-type infimal convolutions as regularization terms. Based on the dual formulation of these problems, we introduce modified dual problems by tightening the constraints on the dual variable in Section 3. We give a useful formulation of the modified primal problem which clearly shows its difference to the original problem.

In Section 4, we consider modified ℓ_1-type infimal convolutions with finite difference matrices. We start with the practically most important case of ordinary difference matrices in Subsection 4.1 and show the relation to TGV regularizers introduced in [5]. This subsection is related to our previous work [34], where we have introduced a modified infimal convolution just for computational reasons within second order cone programming. In Subsection 4.2 we enlarge our considerations to more general difference matrices.

To compute the minimizers for the infimal convolution regularization term we apply an alternating direction method of multipliers in Section 5. Moreover, we use a primal-dual hybrid gradient algorithm for the corresponding MIC-regularized problem.

In Section 6, we explain the differences between the ordinary and the modified infimal convolution approaches by numerical examples. The paper finishes with conclusions.



(a) Original signal

(b) Noisy signal

(c) Result by $\Phi_1 \square \Phi_2$ ($\alpha_1 = 60$, $\alpha_2 = 150$)

(d) Decomposition of (c) into u_1, u_2

(e) Result by $\Phi_1 + \Phi_2$ ($\alpha_1 = 10$, $\alpha_2 = 20$)

(f) Result by $\Phi_1 + \Phi_2$ ($\alpha_1 = 0$, $\alpha_2 = 60$)

(g) Signal (a) repeated in vertical direction

(h) Signal (c) repeated in vertical direction

(i) Signal (e) repeated in vertical direction

(j) Signal (f) repeated in vertical direction

Figure 1: Results of minimizing the functionals in (1) applied to the noisy 1D signal (b) corrupted by additive Gaussian noise of standard deviation 20. By the infimal convolution approach both the jump discontinuities and the linear parts in the signal are nicely restored, see (c) and (h). The corresponding decomposition into the sum of two signals is shown in (d).

2 ℓ_1-type infimal convolutions

We start by considering some general properties of infimal convolutions. The *infimal convolution* of the convex functionals $\Phi_i : \mathbb{R}^N \to (-\infty, +\infty]$, $i = 1, \ldots, m$, $m \geq 2$ is the functional Φ defined by

$$\Phi(u) = (\Phi_1 \Box \ldots \Box \Phi_m)(u) = \inf_{u = u_1 + \ldots + u_m} \sum_{i=1}^{m} \Phi_i(u_i). \tag{2}$$

It can be considered as the convex analysis counterpart of the usual convolution. In the following, let $\Psi^*(v) := \sup_{w \in \mathbb{R}^M} \{\langle v, w \rangle - \Psi(w)\}$ denote the *Fenchel conjugate* of Ψ. For a proper, convex, lower semi-continuous (l.s.c.) function Ψ we have that $\Psi^{**} = \Psi$. Moreover, we stress the fact that the *support function* $\sup_{v \in C} \langle \cdot, v \rangle$ of a nonempty, closed, convex set $C \subset \mathbb{R}^N$ is the Fenchel conjugate of the *indicator function* ι_C of C and vice versa. If Ψ is proper, convex, l.s.c. and positively homogeneous, then it is the support function of a nonempty, closed, convex set. The converse is also true.
Let

$$(\Psi 0^+)(v) := \lim_{\lambda \to \infty} \frac{\Psi(u + \lambda v) - \Psi(u)}{\lambda}, \qquad u \in \operatorname{dom} \Psi$$

be the *recession function* of Ψ.
By the following proposition the convexity of the Φ_i implies the convexity of Φ. Properness of convex functions is not always preserved by infimal convolution since the infimum may be $-\infty$. Lower semi-continuity (l.s.c.) is only preserved under additional conditions. For more information on infimal convolutions we refer to [37].

Theorem 2.1. *Let Φ be the infimal convolution of proper, convex functions Φ_i, $i = 1, \ldots, m$. Then Φ has the following properties:*

i) Φ *is convex.*

ii) *If the Φ_i, $i = 1, \ldots, m$ are also l.s.c. and*

$$\begin{aligned} (\Phi_1 0^+)(u_1) + \cdots + (\Phi_m 0^+)(u_m) &\leq 0, \\ (\Phi_1 0^+)(-u_1) + \cdots + (\Phi_m 0^+)(-u_m) &> 0 \end{aligned}$$

imply that $u_1 + \cdots + u_m \neq 0$, then Φ is proper, convex and l.s.c. and the infimum in the definition of $\Phi(u)$ is attained for any $u \in \mathbb{R}^N$. In particular, the above implication holds true if $\Phi_i(u) = \Phi_i(-u)$ for all $u \in \mathbb{R}^N$.

iii) *If $\Phi_i(u) := \|R_i u\|$ with $R_i \in \mathbb{R}^{N_i, N}$, $i = 1, \ldots, m$ and some norm $\|\cdot\|$ in $\mathbb{R}^{N_i}$, then Φ is continuous.*

iv) $(\Phi_1 \Box \ldots \Box \Phi_m)^* = \Phi_1^* + \ldots + \Phi_m^*$.

Proof: For i) we refer to [28, p. 33] and the proof of the first part of ii) can be found in [28, p. 76]. The last part of ii) is clear since it follows from $\Psi(u) = \Psi(-u)$ that

$$\begin{aligned} (\Psi 0^+)(-v) &= \lim_{\lambda \to \infty} \frac{\Psi(u - \lambda v) - \Psi(u)}{\lambda} = \lim_{\lambda \to \infty} \frac{\Psi(-u - \lambda v) - \Psi(-u)}{\lambda} \\ &= \lim_{\lambda \to \infty} \frac{\Psi(u + \lambda v) - \Psi(u)}{\lambda} = (\Psi 0^+)(v). \end{aligned}$$

To prove iii) we consider

$$\Phi(u+h) = \inf_{u+h=u_1+\ldots+u_m} \sum_{i=1}^m \|R_i u_i\| = \inf_{u_1,\ldots,u_{m-1}} \left\{ \sum_{i=1}^{m-1} \|R_i u_i\| + \|R_m(u+h-\sum_{i=1}^{m-1} u_i)\| \right\}.$$

Since

$$\|R_m(u-\sum_{i=1}^{m-1} u_i)\| - \|R_m h\| \leq \|R_m(u+h-\sum_{i=1}^{m-1} u_i)\| \leq \|R_m(u-\sum_{i=1}^{m-1} u_i)\| + \|R_m h\|$$

we conclude that

$$\Phi(u) - \|R_m h\| \leq \Phi(u+h) \leq \Phi(u) + \|R_m h\|.$$

This implies that $|\Phi(u+h)-\Phi(u)| \to 0$ if $\|h\| \to 0$ and we are done.
The proof of iv) is given in [28, p. 145]. □

The infimal convolution functionals applied in this paper will fulfill both ii) and iii).

Let $\|\cdot\|_p$, $1 \leq p \leq \infty$ denote the usual ℓ_p vector norms on $\mathbb{R}^N$. For $V = (V_1^{\mathrm{T}}, \ldots, V_n^{\mathrm{T}})^{\mathrm{T}} \in \mathbb{R}^{nN}$, $V_i \in \mathbb{R}^N$ and positive weight vectors $\omega = (\omega_k)_{k=1}^n$, we define norms on $\mathbb{R}^{nN}$ as follows:

$$\|V\|_{p,\omega} := \| \left(\omega_1 V_1^2 + \ldots + \omega_n V_n^2\right)^{\frac{1}{2}} \|_p,$$

where the vector multiplication and the square root are meant componentwise. For given $f \in \mathbb{R}^N$, we are interested in minimizers of the functional

$$(\ell_2^2\text{-IC/P}) \qquad \operatorname*{argmin}_{u\in\mathbb{R}^N} \left\{ \frac{1}{2}\|f-u\|_2^2 \ + \Phi_{IC}(u) \right\}$$

with the infimal convolution $\Phi_{IC} := \Phi_1 \Box \ldots \Box \Phi_m$ of the special ℓ_1-type functionals

$$\Phi_i(u) := \alpha_i \|R_i u_i\|_{1,\omega_i}. \tag{3}$$

Note that $\|V\|_{p,\omega_i} = \| \left(\omega_{i,1} V_1^2 + \ldots + \omega_{i,n_i} V_{n_i}^2\right)^{\frac{1}{2}} \|_p$ for $V := R_i u$. Since the functional in (ℓ_2^2-IC/P) is coercive, strictly convex and by Theorem 2.1 iii) continuous, it has a unique minimizer which we denote by $\hat{u}_{IC}$.
In this paper, we will propose a modification of the (ℓ_2^2-IC/P) functional. Since our modification is motivated from the dual functional of (ℓ_2^2-IC/P) we have to establish the dual problem first. In general, for proper, convex, l.s.c. functions $g : \mathbb{R}^N \to (-\infty, +\infty]$ and $f : \mathbb{R}^M \to (-\infty, +\infty]$ and a linear operator $A \in \mathbb{R}^{M,N}$, the primal and its dual optimization problems read

$$\text{(P)} \quad \min_{u\in\mathbb{R}^N} \{g(u) + f(Au)\}, \qquad\qquad \text{(D)} \quad -\min_{v\in\mathbb{R}^M} \{g^*(-A^{\mathrm{T}}v) + f^*(v)\}. \tag{4}$$

Thus, using Theorem 2.1 iv) and the fact that $(\frac{1}{2}\|f-\cdot\|_2^2)^*(v) = \frac{1}{2}\|f+v\|_2^2 - \frac{1}{2}\|f\|_2^2$, the dual problem of ($\ell_2^2$-IC/P) reads

$$\operatorname*{argmin}_{v\in\mathbb{R}^N} \left\{ \frac{1}{2}\|f-v\|_2^2 + \sum_{i=1}^m \Phi_i^*(v) \right\}.$$

The functionals Φ_i are positively homogeneous so that their Fenchel conjugates are indicator functions ι_C of some sets C, more precisely,

$$\Phi_i^* = \iota_{C_{\alpha_i}} \quad \text{with} \quad C_{\alpha_i} := \{v = R_i^{\mathrm{T}} V : \|V\|_{\infty,1/\omega_i} \leq \alpha_i\}.$$

Conversely, we can rewrite Φ_i as

$$\Phi_i = \iota_{C_{\alpha_i}}^* = \sup_{\|V\|_{\infty,1/\omega_i} \leq \alpha_i} \langle \cdot, R_i^{\mathrm{T}} V \rangle.$$

Hence our dual problem becomes

$$\operatorname*{argmin}_{v \in \mathbb{R}^N} \Big\{ \frac{1}{2} \|f - v\|_2^2 + \sum_{i=1}^m \iota_{C_{\alpha_i}}(v) \Big\},$$

or as a constrained problem

$$(\ell_2^2\text{-IC/D}) \qquad \frac{1}{2}\|f - v\|_2^2 \;\to\; \min \qquad \text{subject to} \qquad \begin{aligned} & v = R_1^{\mathrm{T}} V_1 = \ldots = R_m^{\mathrm{T}} V_m, \\ & \|V_i\|_{\infty,1/\omega_i} \leq \alpha_i, \; i = 1, \ldots, m. \end{aligned}$$

The relation between the minimizers $\hat{u}_{IC}$ of (ℓ_2^2-IC/P) and $\hat{v}_{IC}$ of (ℓ_2^2-IC/D) is given by $\hat{u}_{IC} = f - \hat{v}_{IC}$.

In applications matrices R_i arising from differential operators as those in the following example are frequently applied.

Example 2.2. *Let $m = 2$. Take the* forward difference matrix *(with Neumann/mirror boundary conditions)*

$$D := \begin{pmatrix} -1 & 1 & 0 & 0 & \cdots & 0 \\ 0 & -1 & 1 & 0 & \cdots & 0 \\ & & & & & \\ \vdots & & \ddots & \ddots & & \vdots \\ 0 & \cdots & 0 & -1 & 1 & 0 \\ 0 & \cdots & 0 & 0 & -1 & 1 \\ 0 & \cdots & 0 & 0 & 0 & 0 \end{pmatrix} \in \mathbb{R}^{n,n} \tag{5}$$

as a discretization of the first derivative with spatial step size $h = 1$. Then $(-D)^T D$ is the central difference matrix for second order derivatives. Let $A \otimes B$ denote the Kronecker product of A and B. If we reshape square images F of size $n \times n$ columnwise into vectors f of size $N = n^2$ we can use

$$\begin{aligned} D_x &:= I_n \otimes D, & D_{xx} &:= I_n \otimes (-D^T)D, & D_{xy} &:= (-D^T) \otimes D, \\ D_y &:= D \otimes I_n, & D_{yy} &:= (-D^T)D \otimes I_n, & D_{yx} &:= D \otimes (-D^T), \end{aligned}$$

as discrete partial first and second order derivative operators. For simplicity of notation we use square images although the approach works for rectangular images, too. Set

$$\mathcal{D}_1 := \begin{pmatrix} D_x \\ D_y \end{pmatrix}, \; \mathcal{D}_{2,a} := \begin{pmatrix} D_{xx} \\ D_{yy} \end{pmatrix}, \; \mathcal{D}_{2,b} := \begin{pmatrix} D_{xx} \\ D_{xy} + D_{yx} \\ D_{yy} \end{pmatrix}, \; \mathcal{D}_{2,c} := \begin{pmatrix} D_{xx} \\ D_{yx} \\ D_{xy} \\ D_{yy} \end{pmatrix}.$$

In particular, $\mathcal{D}_1 \sim \nabla$ serves as a frequently used discrete gradient operator and its negative adjoint as the corresponding discrete divergence $-\mathcal{D}_1^T \sim \operatorname{div}$, see, e.g., [7]*. In applications*

$R_1 := \mathcal{D}_1$ and $R_2 := \mathcal{D}_{2,\bullet}$ with weights $(1,1)$ on $\mathbb{R}^{2N}$, $(1,\frac{1}{2},1)$ on $\mathbb{R}^{3N}$ and $(1,1,1,1)$ on $\mathbb{R}^{4N}$ were used. Note that except for $\|\mathcal{D}_{2,a}u\|_{1,(1,1)}$ the corresponding continuous functionals of $\|R_iu\|_{1,w_i}$, $i=1,2$ are rotationally invariant. The continuous equivalent of (ℓ_2^2-IC/P) with $m=2$ and first and second order derivative operators was for example used in the Chambolle-Lions approach [8].

3 Modified ℓ_1-type infimal convolutions

In this section, we propose a modification of the ℓ_2^2-IC functional which is superior in certain image processing tasks as demonstrated in Fig. 2.

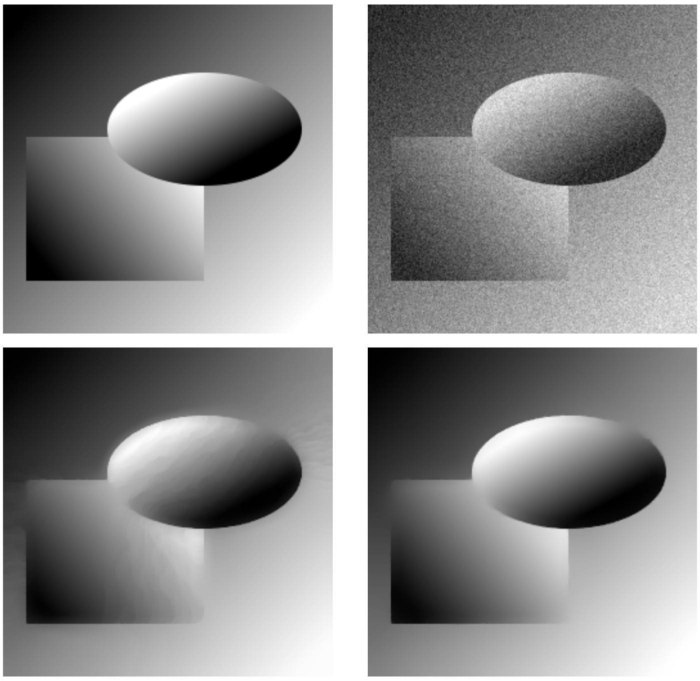


Figure 2: Top: Original image u (left), courtesy of S. Didas [14], and noisy image f (right) corrupted by additive Gaussian noise of standard deviation 20. Bottom: Denoised images by ℓ_2^2-IC (left) and ℓ_2^2-MIC (right) with $R_1 = \mathcal{D}_1$, $R_2 = \mathcal{D}_{2,a}$ and $\alpha_1 = 60$, $\alpha_2 = 300$, see [34].

To this end, we assume that the matrices R_i are related to R_m via matrices $L_i \in \mathbb{R}^{N_m,N_{m-i}}$ such that

$$R_m = L_i^m R_{m-i} = L_i R_{m-i}, \quad i = 0,\ldots,m-1, \tag{6}$$

where we skip the superscript m in the notation of L if its relation to the index i is clear. Furthermore, we agree that $L_0 := I_N$. Note that such a matrix L_i^m exists if $\operatorname{rg} R_{m-i}^{\mathrm{T}} = \operatorname{rg}(R_{m-i}^{\mathrm{T}}\, R_m^{\mathrm{T}})$.

In particular, we obtain for our discrete differential operators in Example 2.2 the following factorizations.

Example 3.1. *For the matrices $R_1 = \mathcal{D}_1$ and $R_2 = \mathcal{D}_{2,\bullet}$ in Example 2.2 it holds that*

$\mathcal{D}_{2,\bullet} = L_{1,\bullet}\mathcal{D}_1$ *with*

$$L_{1,a} = \begin{pmatrix} -D_x^T & 0 \\ 0 & -D_y^T \end{pmatrix}, \; L_{1,b} = \begin{pmatrix} -D_x^T & 0 \\ -D_y^T & -D_x^T \\ 0 & -D_y^T \end{pmatrix}, \; L_{1,c} = \begin{pmatrix} -D_x^T & 0 \\ 0 & -D_x^T \\ -D_y^T & 0 \\ 0 & -D_y^T \end{pmatrix}.$$

We consider the dual problem (ℓ_2^2-IC/D) in its constrained form. Having the relation $v = R_m^{\mathrm{T}} V_m = R_i^{\mathrm{T}} L_{m-i}^{\mathrm{T}} V_m$ in mind, it is self-evident to deal also with the slightly modified functional

$$(\ell_2^2\text{-MIC/D}) \qquad \frac{1}{2}\|f-v\|_2^2 \;\to\; \min \qquad \text{subject to} \qquad \begin{aligned} & v = R_m^{\mathrm{T}} V, \\ & \|L_{m-i}^{\mathrm{T}} V\|_{\infty,1/\omega_i} \le \alpha_i, \; i = 1,\ldots,m. \end{aligned}$$

In other words, in contrast to (ℓ_2^2-IC/D) we have the additional restrictions $V_i = L_{m-i}^{\mathrm{T}} V_m$, $i = 1,\ldots,m-1$. Note that $R_i^{\mathrm{T}} V_i = R_i^{\mathrm{T}} W_i$ implies $V_i = W_i$ if and only if $\mathcal{N}(R_i^{\mathrm{T}}) = \{0\}$, respectively, if and only if $\mathcal{R}(R_i) = \mathbb{R}^{N_i}$. Hence, if the above conditions hold true for $i = 1,\ldots,m-1$, then the two problems (ℓ_2^2-IC/D) and (ℓ_2^2-MIC/D) coincide.
As an unconstrained problem (ℓ_2^2-MIC/D) reads

$$\operatorname*{argmin}_{V\in\mathbb{R}^{N_m}} \Big\{\frac{1}{2}\|f-R_m^{\mathrm{T}} V\|_2^2 + \iota_K(V)\Big\} \quad \text{with} \quad K := \{V : \|L_{m-i}^{\mathrm{T}} V\|_{\infty,1/\omega_i} \le \alpha_i, \; i = 1,\ldots,m\} \tag{7}$$

respectively,

$$\operatorname*{argmin}_{V\in\mathbb{R}^{N_m}} \Big\{\frac{1}{2}\|f - R_m^{\mathrm{T}} V\|_2^2 + \sum_{i=1}^m \iota_{K_{\alpha_i}}(V)\Big\} \quad \text{with} \quad K_{\alpha_i} := \{V : \|L_{m-i}^{\mathrm{T}} V\|_{\infty,1/\omega_i} \le \alpha_i\}. \tag{8}$$

For $m = 2$ and the special matrices $R_1 = \mathcal{D}_1$, $R_2 = \mathcal{D}_{2,a}$, respectively, $R_1 = \mathcal{D}_1$, $R_2 = \mathcal{D}_{2,c}$ of Example 2.2 the modified dual functional (ℓ_2^2-MIC/D) was suggested in [34] for the denoising of images.
Since it is hard to see why (ℓ_2^2-MIC/D) could lead to better denoising results than (ℓ_2^2-IC/D) we give a formulation of the primal problem (ℓ_2^2-MIC/P) which in our opinion better clarifies the differences between the approaches.

Proposition 3.2. *The primal problem of* (ℓ_2^2-MIC/D) *is given by*

$$(\ell_2^2\text{−MIC/P}) \quad \operatorname*{argmin}_{u\in\mathbb{R}^N} \Big\{\frac{1}{2}\|f-u\|_2^2 + \iota_K^*(R_m u)\Big\} \quad \text{with} \quad \iota_K^*(R_m u) := \sup_{\substack{\|L_{m-i}^T V\|_{\infty,1/\omega_i} \le \alpha_i \\ i=1,\ldots,m}} \langle R_m u, V\rangle$$

and can be rewritten as

$$\operatorname*{argmin}_{u\in\mathbb{R}^N} \Big\{\frac{1}{2}\|f-u\|_2^2 + \Phi_{MIC}(u)\Big\}$$

with

$$\Phi_{MIC}(u) := \inf_{\substack{u=u_1+\ldots+u_m \\ s_i\in\mathcal{N}(R_i^T)}} \Big\{\sum_{i=1}^{m-1} \alpha_i \|R_i u_i - s_i\|_{1,\omega_i} + \alpha_m \|R_m u_m + \sum_{i=1}^{m-1} L_{m-i} s_i\|_{1,\omega_m}\Big\}, \tag{9}$$

where $\mathcal{N}(R_i^T)$ *denotes the null space (kernel) of the operator* R_i^T.

The difference between Φ_{IC} and Φ_{MIC} consists in the additional degree of freedom obtained by the vectors $s_i \in \mathcal{N}(R_i^{\mathrm{T}})$, $i = 1, \ldots, m-1$. We see that (ℓ_2^2-MIC/P) is coercive, strictly convex and l.s.c. Thus its minimizer which we call $\hat{u}_{MIC}$ is unique. It is related to the minimizer $\hat{v}_{MIC}$ of the dual problem by $\hat{u}_{MIC} = f - \hat{v}_{MIC}$.

Proof: By (4) the primal problem of (7) reads as (ℓ_2^2-MIC/P). By (8) the primal problem is also given by

$$\frac{1}{2}\|f-u\|_2^2 + \Phi(R_m u),$$

where by Theorem 2.1 iv)

$$\Phi(R_m u) = \Big(\sum_{i=1}^m \iota_{K_{\alpha_i}}\Big)^*(R_m u) = \inf_{R_m u = U_1 + \ldots + U_m} \sum_{i=1}^m \iota_{K_{\alpha_i}}^*(U_i). \tag{10}$$

Using $\mathbb{R}^{N_m} = \mathcal{R}(L_{m-i}) \oplus \mathcal{N}(L_{m-i}^{\mathrm{T}})$, we obtain that

$$\iota_{K_{\alpha_i}}^*(U) \;=\; \sup_{\|L_{m-i}^{\mathrm{T}} V\|_{\infty,1/\omega_i} \le \alpha_i} \langle U, V\rangle \;=\; +\infty \quad \text{if } U \notin \mathcal{R}(L_{m-i}).$$

Since we are looking for the infimum in (10) this implies that $U = L_{m-i}x$ and consequently

$$\begin{aligned}
\iota_{K_{\alpha_i}}^*(U) &= \sup_{\|L_{m-i}^{\mathrm{T}} V\|_{\infty,1/\omega_i} \le \alpha_i} \langle L_{m-i}x, V\rangle = \sup_{\substack{\|z\|_{\infty,1/\omega_i} \le \alpha_i \\ z \in \mathcal{R}(L_{m-i}^{\mathrm{T}})}} \langle x, z\rangle \\
&= \sup_z \{\langle x, z\rangle - \iota_{\{z:\|z\|_{\infty,1/\omega_i} \le \alpha_i\}}(z) - \iota_{\{z: z \in \mathcal{R}(L_{m-i}^{\mathrm{T}})\}}(z)\} \\
&= \left(\iota_{\{z:\|z\|_{\infty,1/\omega_i} \le \alpha_i\}} + \iota_{\{z: z \in \mathcal{R}(L_{m-i}^{\mathrm{T}})\}}\right)^*(x) \\
&= \inf_{x = v + w} \{\iota^*_{\{z:\|z\|_{\infty,1/\omega_i} \le \alpha_i\}}(v) + \iota^*_{\{z: z \in \mathcal{R}(L_{m-i}^{\mathrm{T}})\}}(w)\} \\
&= \inf_{x = v + w} \{\alpha_i \|v\|_{1,\omega_i} + \iota^*_{\{z: z \in \mathcal{R}(L_{m-i}^{\mathrm{T}})\}}(w)\}.
\end{aligned} \tag{11}$$

Since

$$\iota^*_{\{z: z \in \mathcal{R}(L_{m-i}^{\mathrm{T}})\}}(w) = \sup_{v \in \mathcal{R}(L_{m-i}^{\mathrm{T}})} \langle v, w\rangle = \sup_{y \in \mathbb{R}^{N_i}} \langle L_{m-i}^{\mathrm{T}} y, w\rangle,$$

we conclude that $w \in \mathcal{N}(L_{m-i})$ since otherwise this functional becomes $+\infty$ and cannot lead to the infimum in (11). Hence it follows that

$$\iota_{K_{\alpha_i}}^*(U) = \inf_{\substack{U = L_{m-i}x \\ w \in \mathcal{N}(L_{m-i})}} \alpha_i \|x - w\|_{1,\omega_i}$$

and the functional in (10) reads

$$\begin{aligned}
\Phi(R_m u) &= \inf_{\substack{R_m u = \sum_{i=1}^m L_{m-i} x_i \\ w_i \in \mathcal{N}(L_{m-i})}} \sum_{i=1}^m \alpha_i \|x_i - w_i\|_{1,\omega_i} = \inf_{R_m u = \sum_{i=1}^m L_{m-i} x_i} \sum_{i=1}^m \alpha_i \|x_i\|_{1,\omega_i} \\
&= \inf_{x_i \in \mathbb{R}^{N_i}} \Big\{ \sum_{i=1}^{m-1} \alpha_i \|x_i\|_{1,\omega_i} + \alpha_m \|R_m u - \sum_{i=1}^{m-1} L_{m-i} x_i\|_{1,\omega_m} \Big\}.
\end{aligned}$$

The structure of Φ_{MIC} follows by setting $x_i := R_i u_i - s_i$ with $s_i \in \mathcal{N}(R_i^{\mathrm{T}})$, $u_m := u - \sum_{i=1}^{m-1} u_i$ and by using (6). $\square$

In the context of infimal convolutions we mention that

$$\Phi_{MIC}(u) = (\Psi_1 \Box \Psi_2)(S_m u) \quad \text{with} \quad S_m := \frac{1}{m-1} \begin{pmatrix} R_1 \\ \vdots \\ R_{m-1} \end{pmatrix}, \; m \geq 2$$

and

$$\Psi_1(x_1, \ldots, x_{m-1}) := \sum_{i=1}^{m-1} \alpha_i \|x_i\|_{1,\omega_i}, \quad \Psi_2(x_1, \ldots, x_{m-1}) := \alpha_m \| \sum_{i=1}^{m-1} L_{m-i} x_i \|_{1,\omega_m}.$$

There exists an intermediate problem between ℓ_2^2-IC and ℓ_2^2-MIC. This is discussed in the following remark.

Remark 3.3. *Having the relation $v = R_m^T V_m = R_i^T L_{m-i}^T V_m = R_i^T V_i$ in mind and setting $V_i = L_{m-i}^T W_i$ in (ℓ_2^2-IC/D), we obtain the following modification of the ℓ_2^2-IC functional*

$$(\ell_2^2\text{-}\tilde{\text{IC}}/\text{D}) \qquad \frac{1}{2}\|f - v\|_2^2 \;\to\; \min \qquad \textit{subject to} \qquad \begin{aligned} &v = R_m^T W_1 = \ldots = R_m^T W_m, \\ &\|L_{m-i}^T W_i\|_{\infty,1/\omega_i} \leq \alpha_i, \; i = 1, \ldots, m \end{aligned}$$

or in unconstrained form

$$\operatorname*{argmin}_{v \in \mathbb{R}^N} \left\{ \frac{1}{2}\|f - v\|_2^2 + \sum_{i=1}^{m} \iota_{\mathcal{C}_{\alpha_i}}(v) \right\}, \qquad \mathcal{C}_{\alpha_i} := \{v = R_m^T V : \|L_{m-i}^T V\|_{\infty,1/\omega_i} \leq \alpha_i\}. \tag{12}$$

Following similar lines as in the proof of Proposition 3.2 the corresponding primal problem reads

$$(\ell_2^2\text{-}\tilde{\text{IC}}/\text{P}) \qquad \operatorname*{argmin}_{u \in \mathbb{R}^N} \{\frac{1}{2}\|f - u\|_2 + \Phi_{\tilde{IC}}(u)\}, \qquad \Phi_{\tilde{IC}}(u) := \inf_{\substack{u = u_1 + \ldots + u_m \\ w_i \in \mathcal{N}(L_{m-i})}} \sum_{i=1}^{m} \alpha_i \|R_i u_i - w_i\|_{1,\omega_i}. \tag{13}$$

Having a look at the dual problems we conclude that $\|\hat{u}_{IC}\|_2 \leq \|\hat{u}_{\tilde{IC}}\|_2 \leq \|\hat{u}_{MIC}\|_2$.

In image restoration applications we are mainly interested in the case $m = 2$. Let us summarize how the penalizers of the primal problems look like for $m = 2$:

$$\begin{aligned} \Phi_{IC}(u) &= \inf_{u = u_1 + u_2} \{\alpha_1 \|R_1 u_1\|_{1,\omega_1} + \alpha_2 \, \|R_2 u_2\|_{1,\omega_2}\}, \\ &= \inf_{\substack{R_1 u = x_1 + x_2 \\ x_i \in \mathcal{R}(R_1)}} \{\alpha_1 \|x_1\|_{1,\omega_1} + \alpha_2 \|L_1 x_2\|_{1,\omega_2}\}, \end{aligned} \tag{14}$$

$$\begin{aligned} \Phi_{\tilde{IC}}(u) &= \inf_{\substack{u = u_1 + u_2 \\ w_1 \in \mathcal{N}(L_1)}} \{\alpha_1 \|R_1 u_1 - w_1\|_{1,\omega_1} + \alpha_2 \|R_2 u_2\|_{1,\omega_2}\}, \\ &= \inf_{\substack{R_1 u = x_1 + x_2 \\ x_i \in \mathcal{R}(R_1) \\ w_1 \in \mathcal{N}(L_1)}} \{\alpha_1 \|x_1 - w_1\|_{1,\omega_1} + \alpha_2 \|L_1 x_2\|_{1,\omega_2}\}, \\ \Phi_{MIC}(u) &= \inf_{\substack{u = u_1 + u_2 \\ s_1 \in \mathcal{N}(R_1^{\mathrm{T}})}} \{\alpha_1 \|R_1 u_1 - s_1\|_{1,\omega_1} + \alpha_2 \|R_2 u_2 + L_1 s_1\|_{1,\omega_2}\}. \\ &= \inf_{R_1 u = x_1 + x_2} \{\alpha_1 \|x_1\|_{1,\omega_1} + \alpha_2 \|L_1 x_2\|_{1,\omega_2}\}. \end{aligned} \tag{15}$$

Recall that we originally obtained ℓ_2^2-MIC from ℓ_2^2-IC via ℓ_2^2-$\widetilde{\text{IC}}$ by adding further constraints on the dual variables. Here, we see that this led to relaxed conditions on new variables x_1, x_2, compare e.g. (14) and (15). For $\Phi_{MIC}(u)$ we no longer have the restriction that $x_i \in \mathcal{R}(R_1)$ for $i = 1, 2$ and thus, $R_1 u$ can be decomposed into any x_1 and x_2. In general, this results of course in different minimizers and minima. In Section 6 we will see that these modifications improve the restoration results for the discrete difference operators studied in the next section.

4 Discrete difference matrices

In this section, we are interested in matrices R_i related to differential operators since this is the most relevant case in practice. We restrict our attention to finite difference matrices arising from differential operators at rectangular domains with Neumann/mirror boundary conditions. Similar results can be obtained for matrices related to zero or periodic boundary conditions. We start with simple i-th order difference matrices R_i in Subsection 4.1. Then, in Subsection 4.2, we turn to more general difference matrices. The corresponding general differential operators appear for example in the definition of $\mathcal{L}$-splines [32] which can be represented in terms of the Green function of such operators [32]. Applications of such operators and their discrete counterparts can be found in [38, 39].

4.1 Simple difference matrices

Let D be the first order forward difference matrix (5) from Example 2.2. For $j \in \mathbb{N}$, we consider the following i-th order finite difference matrices

$$D_1 := D,\ D_{2j} := (-D^{\mathrm{T}} D)^j,\ D_{2j+1} := D D_{2j}. \tag{16}$$

Moreover, we use the notation $D_0 := I_n$.

Remark 4.1. *Replacing the first and last j rows of $D_{2j} \in \mathbb{R}^{n,n}$, $n > 2j$ and the first j and last $j+1$ rows of $D_{2j+1} \in \mathbb{R}^{n,n}$, $n > 2j+1$ by zero rows, we obtain that the kernel of the i-th modified matrix is given by the span of the discrete polynomials of degree $\leq i-1$, i.e., by*

$$\operatorname{span}\{\left(k^r\right)_{k=1}^n : r = 0, \ldots, i-1\}.$$

One-dimensional setting. Let us first have a look at matrices related to differential operators on the interval, more precisely we deal with $R_i := D_i$. Since the matrices D_i are singular of rank $n-1$, there are many ways to choose L_{m-i} such that (6) is fulfilled. Indeed different choices of L_{m-i} may lead to different functionals Φ_{MIC}. Related to the factorization in (16) a self-evident choice is

$$L_{m-i} := \begin{cases} D_{m-i} & \text{for } i \text{ even}, \\ D_{m-i-1}(-D^{\mathrm{T}}) & \text{for } i \text{ odd}\ . \end{cases} \tag{17}$$

Indeed this choice can also be explained in another way: Based on the singular value decomposition $D = U\Sigma V^{\mathrm{T}}$ we obtain that

$$D_i = \begin{cases} (-1)^{i/2} V \Sigma^i V^{\mathrm{T}} & i \text{ even}, \\ (-1)^{(i-1)/2} U \Sigma^i V^{\mathrm{T}} & i \text{ odd}. \end{cases}$$

Using this relation it is easy to check that L_{m-i} in (17) can also be written as $L_{m-i} = D_m D_i^\dagger$, where $D_i^\dagger$ denotes the *Moore-Penrose inverse* of D_i.

Proposition 4.2. *Let the matrices $R_i = D_i$ be given by* (16) *and the matrices L_i by* (17)*. Then*

$$\Phi_{MIC}(u) = \Phi_{\tilde{IC}}(u) = \inf_{\substack{u=u_1+\ldots+u_m \\ s_i \in \mathcal{N}(R_i^T)}} \Big\{ \sum_{i=1}^{m-1} \alpha_i \|R_i u_i - \epsilon_i s_i\|_1 + \alpha_m \|R_m u_m\|_1 \Big\},\ \epsilon_i = \begin{cases} 1 & \text{for } i \text{ even,} \\ 0 & \text{for } i \text{ odd} \end{cases}$$

holds true for all $m \in \mathbb{N}$ and in particular $\Phi_{IC} = \Phi_{\tilde{IC}} = \Phi_{MIC}$ for $m = 2$.

Proof: Since

$$\begin{aligned} \mathcal{N}(D_{2j}) &= \mathcal{N}(D_{2j}^{\mathrm{T}}) = \mathcal{N}(D_{2j+1}) = \mathcal{N}(L_{m-2j}) = \{c\,1_n : c \in \mathbb{R}\}, \\ \mathcal{N}(D_{2j+1}^{\mathrm{T}}) &= \mathcal{N}(L_{m-2j-1}) = \{(0, \ldots, 0, c)^{\mathrm{T}} : c \in \mathbb{R}\}, \end{aligned} \tag{18}$$

where 1_n denotes the vector of length n consisting only of entries 1, we see that $\mathcal{N}(R_i^{\mathrm{T}}) = \mathcal{N}(L_{m-i})$. Hence $s_i = w_i$ in the definitions (9) and (13). Moreover, the last sum in (9) vanishes and $\mathcal{N}(L_0) = \{0\}$ so that Φ_{MIC} and $\Phi_{\tilde{IC}}$ coincide. Further, considering $\|R_{2j+1}u_{2j+1} - s_{2j+1}\|_1$, $s_{2j+1} \in \mathcal{N}(D_{2j+1}^{\mathrm{T}})$, we conclude by

$$\mathcal{R}(D_{2j+1}) = \{(x_1, \ldots, x_{n-1}, 0)^{\mathrm{T}} : x_i \in \mathbb{R}\} \tag{19}$$

and the definition of $\Phi_{\tilde{IC}}$ that $s_{2j+1} = 0$. This finishes the proof. □

Two-dimensional setting. We consider matrices related to partial differential operators on rectangles, more precisely we restrict our attention to the following two cases a and b. For more sophisticated discretizations of partial derivative operators via finite mimetic differences we refer to [23, 42].

Case a: We use

$$R_{i,a} := \begin{pmatrix} D_{i,x} \\ D_{i,y} \end{pmatrix} \quad \text{with} \quad \begin{aligned} &D_{i,x} := I_n \otimes D_i,\ D_{i,y} := D_i \otimes I_n, \\ &\tilde{D}_{m-i,x} := I_n \otimes D_{m-i-1}(-D^{\mathrm{T}}),\ \tilde{D}_{m-i,y} := D_{m-i-1}(-D^{\mathrm{T}}) \otimes I_n \end{aligned}$$

and

$$L_{m-i,a} := \begin{cases} \operatorname{diag}(D_{m-i,x}, D_{m-i,y}) & \text{for } i \text{ even,} \\ \operatorname{diag}(\tilde{D}_{m-i,x}, \tilde{D}_{m-i,y}) & \text{for } i \text{ odd.} \end{cases}$$

In particular, this involves the setting in Example 2.2, namely $R_{1,a} = \mathcal{D}_1$, $R_{2,a} = \mathcal{D}_{2,a}$ and $L_{1,a}^2$ coincides with the corresponding matrix in Example 3.1.
By definition we see that the elements of $\mathcal{N}(L_{m-(2j+1),a})$ and $\mathcal{R}(R_{2j+1,a})$ are special compositions of the vectors in (18) and (19), respectively. Then we can conclude similarly as in the one-dimensional setting that $w_{2j+1} = 0$ in the definition of $\Phi_{\tilde{IC}}$. Therefore, for $m = 2$, the functionals Φ_{IC} and $\Phi_{\tilde{IC}}$ coincide again. The functional Φ_{MIC} is indeed different. This is discussed in more detail in the Examples 6.1 and 6.2.

Case b: Here, we use the matrices $R_{1,b} := \mathcal{D}_1$, $R_{2,b} := \mathcal{D}_{2,b}$ from Example 2.2 and $L_{1,b}^2$ from Example 3.1. Appropriate matrices for $m = 3$ fulfilling $R_{3,b} = L_1^3 R_{2,b} = L_2^3 R_{1,b}$ can be chosen as follows:

$$R_{3,b} = \begin{pmatrix} D_{xxx} \\ D_{xxy} + D_{xyx} + D_{yxx} \\ D_{xyy} + D_{yxy} + D_{yyx} \\ D_{yyy} \end{pmatrix} := \begin{pmatrix} I_n \otimes D(-D^{\mathrm{T}})D \\ D \otimes (-D^{\mathrm{T}})D + (-D^{\mathrm{T}}) \otimes DD + D \otimes D(-D^{\mathrm{T}}) \\ D(-D^{\mathrm{T}}) \otimes D + DD \otimes (-D^{\mathrm{T}}) + (-D^{\mathrm{T}})D \otimes D \\ D(-D^{\mathrm{T}})D \otimes I_n \end{pmatrix},$$

$$L^3_{2,b} = \begin{pmatrix} I_n \otimes D(-D^{\mathrm{T}}) & 0 \\ D \otimes (-D^{\mathrm{T}}) + (-D^{\mathrm{T}}) \otimes D & I_n \otimes D(-D^{\mathrm{T}}) \\ D(-D^{\mathrm{T}}) \otimes I_n & D \otimes (-D^{\mathrm{T}}) + (-D^{\mathrm{T}}) \otimes D \\ 0 & D(-D^{\mathrm{T}}) \otimes I_n \end{pmatrix}, \quad L^3_{1,b} = \begin{pmatrix} I_n \otimes D & 0 & 0 \\ D \otimes I_n & I_n \otimes D & 0 \\ 0 & D \otimes I_n & I_n \otimes D \\ 0 & 0 & D \otimes I_n \end{pmatrix}.$$

For case b the functionals Φ_{MIC} can be considered as discrete variants of the continuous ℓ_2^2-TGV_α^m, $m = 2, 3$ functionals introduced in [5]. To verify this relation let us recall the definition of TGV_α^m from [5].

Definition of TGV_α^m: Let

$$\mathrm{Sym}^m(\mathbb{R}^d) := \{v : \underbrace{\mathbb{R}^d \times \cdots \times \mathbb{R}^d}_{m} \to \mathbb{R} : v \; m - \text{linear}, \text{symmetric}\}$$

be the space of m-linear, symmetric mappings over $\mathbb{R}^d$ to $\mathbb{R}$, i.e., the space of symmetric, covariant m-tensors. These symmetric m-tensors are completely determined by the values $v(e_{j_1}, \ldots, e_{j_m}) = v_{j_1,\ldots,j_m}$, where e_j denotes the j-th unit vector in $\mathbb{R}^d$ and $j_i \in \{1, \ldots, d\}$, $j_1 \leq \ldots \leq j_m$. We consider symmetric m-tensor fields $V : \Omega \to \mathrm{Sym}^m(\mathbb{R}^d)$ with $\Omega \subset \mathbb{R}^d$. The *total generalized variation of order m with weighting vector* $\alpha > 0$ is defined by

$$\begin{aligned} TGV_\alpha^m(u) \quad := \quad \sup \Big\{ \int_\Omega u \, \mathrm{div}^m V \, dx : \, & V \in C_c^m(\Omega, \mathrm{Sym}^m(\mathbb{R}^d)), \\ & \|\mathrm{div}^{m-i} V\|_\infty \leq \alpha_i, \; i = 1, \ldots, m \Big\}, \end{aligned}$$

where $C_c^k(\Omega, \mathrm{Sym}^m(\mathbb{R}^d))$ denotes the space of k times continuously differentiable symmetric m-tensor fields with compact support in Ω and

$$\begin{aligned} \mathrm{div}^i V(x) &:= \mathrm{tr}^i((\nabla^i \otimes V)(x)), \\ \mathrm{tr}(v)(a) &:= \sum_{j=1}^d v(e_j, a, e_j), \quad a \in \underbrace{\mathbb{R}^d \times \ldots \times \mathbb{R}^d}_{m-2}, \\ (\nabla^i \otimes V)(x)(a_1, \ldots, a_{m+i}) &:= D^i(V)(x)(a_1, \ldots, a_i)(a_{i+1}, \ldots, a_{m+i}). \end{aligned}$$

Here $D^i(V) : \Omega \to \mathcal{L}\left((\mathbb{R}^d)^i, \mathrm{Sym}^m(\mathbb{R}^d)\right)$ is the i-th Fréchet derivative of V (componentwise Fréchet derivative) and $\mathcal{L}\left((\mathbb{R}^d)^i, \mathrm{Sym}^m(\mathbb{R}^d)\right)$ is the space of i-linear, continuous mappings from $(\mathbb{R}^d)^i$ into $\mathrm{Sym}^m(\mathbb{R}^d)$.
Note that for a symmetric m-tensor field V we have that $\mathrm{div}^i V$ is an $m - i$ tensor field.

In our applications, we are only interested in rectangular domains $\Omega \subset \mathbb{R}^2$, i.e., $d = 2$. To see the relation to our setting, consider Φ_{MIC} from (ℓ_2^2-MIC/P) in Proposition 3.2:

$$\Phi_{MIC}(u) = \sup_{V \in \mathbb{R}^{N_m}} \left\{ \langle u, R_m^{\mathrm{T}} V \rangle : \|L_{m-i}^{\mathrm{T}} V\|_{\infty, 1/\omega_i} \leq \alpha_i, \; i = 1, \ldots, m \right\}.$$

We are looking for appropriate matrices R_m^{T} playing the discrete role of div^m and matrices L_{m-i}^{T} which can be considered as discrete versions of div^{m-i}. Such a discrete setting is given in the above case b. We explain the relation for $m = 2$. For symmetric 2-tensor fields

$V = (V_{1,1}^{\mathrm{T}}, V_{1,2}^{\mathrm{T}}, V_{2,2}^{\mathrm{T}})^{\mathrm{T}}$ we obtain that the 0-tensor field (scalar function) $\mathrm{div}^2 V$ is given by

$$\begin{aligned}
\mathrm{div}^2 V &= \mathrm{tr}^2((\nabla^2 \otimes V)(\cdot)) \;=\; \mathrm{tr}\left(\mathrm{tr}((\nabla^2 \otimes V)(\cdot))\right) \\
&= \mathrm{tr}((\nabla^2 \otimes V)(\cdot))(e_1, e_1) \;+\; \mathrm{tr}((\nabla^2 \otimes V)(\cdot))(e_2, e_2) \\
&= (\nabla^2 \otimes V)(\cdot)(e_1, e_1, e_1, e_1) \;+\; (\nabla^2 \otimes V)(\cdot)(e_2, e_1, e_1, e_2) \\
&\quad + (\nabla^2 \otimes V)(\cdot)(e_1, e_2, e_2, e_1) \;+\; (\nabla^2 \otimes V)(\cdot)(e_2, e_2, e_2, e_2) \\
&= D^2 V(\cdot)(e_1, e_1)(e_1, e_1) + D^2 V(\cdot)(e_2, e_1)(e_1, e_2) \\
&\quad + D^2 V(\cdot)(e_1, e_2)(e_2, e_1) + D^2 V(\cdot)(e_2, e_2)(e_2, e_2) \\
&= \frac{\partial^2}{\partial x^2} V_{1,1} \;+\; \left(\frac{\partial^2}{\partial y \partial x} + \frac{\partial^2}{\partial x \partial y}\right) V_{1,2} \;+\; \frac{\partial^2}{\partial y^2} V_{2,2},
\end{aligned}$$

and the 1-tensor field (vector function with two components) $\mathrm{div}^1 V$ reads

$$\begin{aligned}
\mathrm{div}^1 V(\cdot)(a) &= \mathrm{tr}((\nabla \otimes V)(\cdot))(a) \;=\; (\nabla \otimes V)(\cdot)(e_1, a, e_1) + (\nabla \otimes V)(\cdot)(e_2, a, e_2) \\
&= DV(\cdot)(e_1)(a, e_1) + DV(\cdot)(e_2)(a, e_2), \\
\mathrm{div}^1 V(\cdot)(e_1) &= \frac{\partial}{\partial x} V_{1,1} + \frac{\partial}{\partial y} V_{1,2}, \\
\mathrm{div}^1 V(\cdot)(e_2) &= \frac{\partial}{\partial x} V_{1,2} + \frac{\partial}{\partial y} V_{2,2}.
\end{aligned}$$

On the other hand, for a vector $V = (V_{1,1}^{\mathrm{T}}, V_{1,2}^{\mathrm{T}}, V_{2,2}^{\mathrm{T}})^{\mathrm{T}}$ with $V_{1,1}, V_{1,2}, V_{2,2} \in \mathbb{R}^N$, $N = n^2$, which acts as a discrete version of the above tensor field, we have indeed that

$$\begin{aligned}
R_{2,b}^{\mathrm{T}} V &= (D_{xx}, D_{xy} + D_{yx}, D_{yy}) V = D_{xx} V_{1,1} + (D_{xy} + D_{yx}) V_{1,2} + D_{yy} V_{2,2}, \\
(L_{1,b}^2)^{\mathrm{T}} V &= \begin{pmatrix} -D_x & -D_y & 0 \\ 0 & -D_x & -D_y \end{pmatrix} V \;=\; -\begin{pmatrix} D_x V_{1,1} + D_y V_{1,2} \\ D_x V_{1,2} + D_y V_{2,2} \end{pmatrix}.
\end{aligned}$$

Furthermore, we have in case b that $\omega_1 := (1,1)$ and $\omega_2 := (1, \frac{1}{2}, 1)$, which correspond to the weights used in [5, Section 4.1].

4.2 General difference matrices

Although the finite difference matrices of the previous subsection are mainly applied in practical image processing tasks, other difference operators may be useful for special applications as well, see [39]. We consider the polynomial

$$\begin{aligned}
P_{\mathcal{L}}(x) &:= x^m + a_{m-1} x^{m-1} + \ldots + a_1 x + a_0, & a_i \in \mathbb{R}, \\
&= \prod_{k=1}^{m} (x - \xi_k), & \xi_k \in \mathbb{C} \qquad (20)
\end{aligned}$$

and the corresponding differential operator

$$\mathcal{L}(u) = u^{(m)} + a_{m-1} u^{(m-1)} + \ldots + a_1 u' + a_0 u.$$

The motivation for the following consideration of a discrete version of $\mathcal{L}$ comes from [38]. We also refer to [38] for more material on the role of $\mathcal{L}$ in signal processing including references, in particular, in connection with $\mathcal{L}$-splines. Let ξ_{m_j}, $j = 1, \ldots, \tilde{m}$ denote the pairwise different values of ξ_j in (20) and assume that ξ_{m_j} appears with multiplicity d_j. Then, the kernel of $\mathcal{L}$ is given by

$$\mathcal{N}(\mathcal{L}) := \operatorname{span}\{x^r e^{\xi_{m_i} x} : m_i = 1, \ldots, \tilde{m};\ r = 0, \ldots, d_i - 1\}. \tag{21}$$

In the following, we restrict our attention to operators with $\xi_k \in \mathbb{R}$, $k = 1, \ldots, m$. As the discrete counterpart of $\mathcal{L}$ we use

$$\begin{aligned}
D(\xi_1) &:= D - \xi_1 I_n, \\
D(\xi_1, \xi_2) &:= (-D^{\mathrm{T}} - \xi_2 I_n)(D - \xi_1 I_n), \\
D(\xi_1, \ldots, \xi_{2j}) &:= \prod_{l=1}^{j} \left((-D^{\mathrm{T}} - \xi_{2l} I_n)(D - \xi_{2l-1} I_n)\right), \\
D(\xi_1, \ldots, \xi_{2j+1}) &:= (D - \xi_{2j+1} I_n) \prod_{l=1}^{j} \left((-D^{\mathrm{T}} - \xi_{2l} I_n)(D - \xi_{2l-1} I_n)\right),
\end{aligned}$$

where we use the agreement that $\prod_{l=1}^{j} A_l := A_j \ \ldots \ A_1$. Note that the ordering of the matrix multiplication plays only a role for the first and last j rows of $D(\xi_1, \ldots, \xi_{2j})$ and for the first j and last $j+1$ rows of $D(\xi_1, \ldots, \xi_{2j+1})$.

Remark 4.3. *Let us briefly discuss the relation to* (21)*. Replacing the first and last j rows in $D(\xi_1, \ldots, \xi_{2j})$ and the first j and last $j+1$ rows in $D(\xi_1, \ldots, \xi_{2j+1})$ by zero rows, we obtain for $\xi_{2j+1} \neq -1$ and $\xi_{2j} \neq 1$ that the kernels of the corresponding modified matrices $\widetilde{D}(\xi_1, \ldots, \xi_n)$ are given by*

$$\begin{aligned}
\widetilde{D}(\xi_1): &\quad \operatorname{span}\{\left((1+\xi_1)^k\right)_{k=0}^{n-1}\}, \\
\widetilde{D}(\xi_1, \xi_2): &\quad \operatorname{span}\{\left((1+\xi_1)^k\right)_{k=0}^{n-1}, \left((1-\xi_2)^{n-k-1}\right)_{k=0}^{n-1}\} \quad \text{if } \xi_2 \neq \frac{\xi_1}{1+\xi_1} \\
&\quad \operatorname{span}\{\left((1+\xi_1)^k\right)_{k=0}^{n-1}, \left(k(1+\xi_1)^k\right)_{k=0}^{n-1}\} \quad \text{if } \xi_2 = \frac{\xi_1}{1+\xi_1} \\
&\quad \vdots
\end{aligned}$$

Note that $e^{\xi_i} = 1 + \xi_i + \mathcal{O}(\xi_i^2)$ as $\xi_i \to 0$.

We have that $D(\xi) = D_i$ for $\xi = (\xi_1, \ldots, \xi_i) = (0, \ldots, 0)$. Therefore, we assume in the following that $\xi_i \neq 0$, $i = 1, \ldots, m$. Moreover, we choose $\xi_{2j+1} \neq -1$ and $\xi_{2j} \neq 1$ so that the matrices $D(\xi)$ are invertible. Then, in the one-dimensional setting with $R_i := D(\xi)$ the matrices L_{m-i}^m in (6) are uniquely determined. Moreover, $\mathcal{N}(L_{m-i}^m) = \mathcal{N}(R_i^{\mathrm{T}}) = \{0\}$ so that the problems ℓ_2^2-IC, ℓ_2^2-$\widetilde{\mathrm{IC}}$ and ℓ_2^2-MIC are again equivalent. Fig. 3 shows the behavior of the functional ℓ_2^2-IC with general difference operators for denoising a signal in the kernel of $D(\xi_1, \xi_2)$. If we choose the parameters α_i, $i = 1, 2$ large enough, we obtain a very good result for such signals in contrast to ℓ_2^2-IC with ordinary first and second order difference operators.

A two-dimensional approach involving the operators $D(\xi)$ instead of D_i can be obtained in the same way as in the previous subsection. Example 6.4 shows the differences between the minimizers of the ℓ_2^2-IC and the ℓ_2^2-MIC functional for the setting in case a with $D(\xi)$.

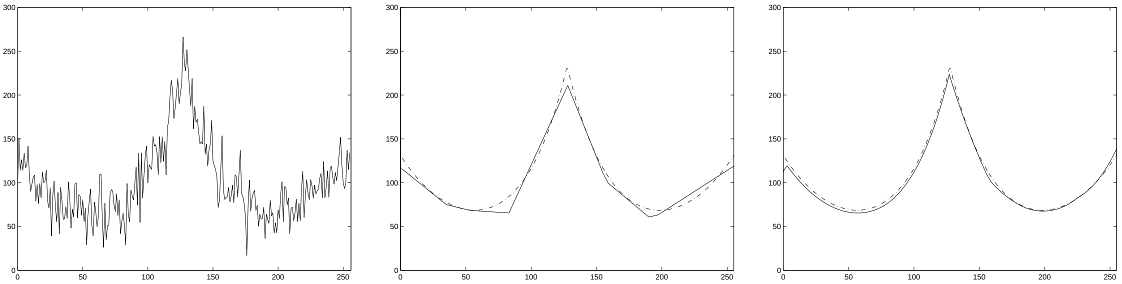


Figure 3: Left: Noisy signal f of the mirrored original signal $u = 5\left((1+0.03)^k\right)_{k=1}^{128} + 5\left((1+0.02)^{129-k}\right)_{k=1}^{128}$ corrupted by additive Gaussian noise of standard deviation 20. Middle/Right: Denoised image by ℓ_2^2-IC with $\alpha_1 = \alpha_2 = 1000$ and difference operators D_i, $i = 1, 2$ (middle) as well as $D(0.03)$, $D(0.03, -0.02)$ (right). The dash-dotted signal is the original one.

5 Numerical algorithms

There exist several algorithms to compute the minimizer of the above problems. *Second order cone programming* (SOCP) was used, e.g., by some of the authors in [34].
The *fast iterative shrinkage threshold algorithm* (FISTA) of Beck and Teboulle [3, 4] was applied with outer and inner FISTA loops, e.g., in [5]. Note that FISTA is based on a multistep algorithm proposed by Nesterov [26].
Sparked by [12, 13, 21, 40, 44], splitting methods which make use of the additive structure of the objective function have become popular recently in image processing. The idea is to solve in each iteration several subproblems which deal with the different components of the objective function individually. For our minimization problems ℓ_2^2-IC and ℓ_2^2-MIC it turns out that the alternating direction method of multipliers (ADMM) and the primal-dual hybrid gradient method (PDHG) are very useful. ADMM and the PDHG method can be derived by considering the Lagrangian function and the augmented Lagrangian function, respectively, and minimizing alternatingly with respect to the primal and the dual variable. Furthermore, ADMM can also be deduced via Douglas-Rachford splitting applied to the dual problem or via Bregman proximal point methods. The PDHG algorithm turns out to be equivalent to Arrow-Hurwicz method. More on theses algorithms can be found in [6, 9, 16, 17, 18, 19, 20, 21, 33, 44] and the references therein.
The starting point to apply ADMM and PDHG is to rewrite a general problem of the form

$$\operatorname*{argmin}_{v\in\mathbb{R}^D} \sum_{i=1}^{r} F_i(C_i v) + F_{r+1}(v), \qquad C_i \in \mathbb{R}^{M_i,D},\ F_i : \mathbb{R}^{M_i} \to (-\infty, +\infty] \tag{22}$$

as a constrained problem

$$\operatorname*{argmin}_{v\in\mathbb{R}^D, z_i\in\mathbb{R}^{M_i}} \sum_{i=1}^{r} F_i(z_i) + F_{r+1}(v) \quad \text{subject to} \quad z_i = C_i v,\ i = 1, \ldots, r \tag{23}$$

with $C := (C_1^{\mathrm{T}} \ldots C_r^{\mathrm{T}})^{\mathrm{T}}$ for $C_i \in \mathbb{R}^{M_i,D}$ and $z^{(k)} := (z_1^{(k)}, \ldots, z_r^{(k)})^{\mathrm{T}}$.

Using this notation, ADMM reads:

Algorithm (ADMM for (23) **)**
Initialization: $z^{(0)}$, $b^{(0)}$
For $k = 0, \dots$ repeat until a stopping criterion is reached:

$$\begin{aligned} v^{(k+1)} &= \operatorname*{argmin}_{v \in \mathbb{R}^D} \{ F_{r+1}(v) + \frac{\gamma}{2} \| b^{(k)} + Cv - z^{(k)} \|_2^2 \} \\ z^{(k+1)} &= \operatorname*{argmin}_{z_i \in \mathbb{R}^{M_i}} \{ \sum_{i=1}^{r} F_i(z_i) + \frac{\gamma}{2} \| b^{(k)} + Cv^{(k+1)} - z \|_2^2 \} \\ b^{(k+1)} &= b^{(k)} + Cv^{(k+1)} - z^{(k+1)} \end{aligned} \tag{24}$$

Let us assume that the functions F_i are proper, convex and closed. Furthermore, suppose that (22) and its dual problem have a solution and that the duality gap is zero. Then, the sequence $(b^{(k)})_{k \in \mathbb{N}}$ converges for every step length parameter $\gamma > 0$ to a point $\hat{b}$ whose scaled version $\frac{1}{\gamma}\hat{b}$ is a solution of the dual problem of (22). Moreover, every cluster point of $(v^{(k)})_{k \in \mathbb{N}}$ is a minimizer of (22).

Algorithm (PDHG for (23) **)**
Initialization: $v^{(0)}$, $z^{(0)}$, $b^{(0)}$
For $k = 0, \dots$ repeat until a stopping criterion is reached:

$$\begin{aligned} v^{(k+1)} &= \operatorname*{argmin}_{v \in \mathbb{R}^D} \{ F_{r+1}(v) + \frac{1}{2\tau} \| v - v^{(k)} + \tau\gamma C^{\mathrm{T}} b^{(k)} \|_2^2 \} \\ z^{(k+1)} &= \operatorname*{argmin}_{z_i \in \mathbb{R}^{M_i}} \{ \sum_{i=1}^{r} F_i(z_i) + \frac{\gamma}{2} \| b^{(k)} + Cv^{(k+1)} - z \|_2^2 \} \\ b^{(k+1)} &= b^{(k)} + Cv^{(k+1)} - z^{(k+1)} \end{aligned}$$

Note that using the notation $p^{(k)} := \frac{1}{\gamma} b^{(k)}$ the above PDHG algorithm is often written as follows in the literature:

$$\begin{aligned} v^{(k+1)} &= \operatorname*{argmin}_{v \in \mathbb{R}^D} \{ F_{r+1}(v) + \frac{1}{2\tau} \| v - v^{(k)} + \tau C^{\mathrm{T}} p^{(k)} \|_2^2 \}, \\ p^{(k+1)} &= \operatorname*{argmin}_{p_i \in \mathbb{R}^{M_i}} \{ (\sum_{i=1}^{r} F_i)^*(p_1, \dots, p_r) + \frac{1}{2\gamma} \| p - p^{(k)} - \gamma C v^{(k+1)} \|_2^2 \}. \end{aligned}$$

The following convergence result for PDHG was shown in [9]: Assume again that the functions F_i are proper, convex and closed, that the primal problem (22) and its dual problem have a solution and that the duality gap is zero. Moreover, we suppose that the domain of $(\sum_{i=1}^{r} F_i)^*$ is bounded and that $\tau\gamma < \frac{1}{\|C\|^2}$. Then, the sequences $(v^{(k)}))_{k \in \mathbb{N}}$ and $(\frac{1}{\gamma} b^{(k)})_{k \in \mathbb{N}}$ generated by the above PDHG algorithm converge to a solution of the primal and the dual problem, respectively. Note that a similar convergence result for a slightly different algorithm was given earlier in [18]. It was observed in [9, 18, 44] that a dynamic choice of the step length parameters $(\tau^{(k)}, \gamma^{(k)})$ is advantageous. This is not implemented in the experiments reported in this paper. We also do not use the acceleration techniques, e.g. based on FISTA, which are proposed in [9].

We now want to apply ADMM and PDHG to the primal ℓ_2^2-IC and ℓ_2^2-MIC problems. There are different ways to implement the above algorithms depending on the formulation of the constrained problem. In our experiments, it turned out that for the ℓ_2^2-IC problem an ADMM performs best, cf. Subsection 5.1. For the ℓ_2^2-MIC problem we found that a PDHG method which we describe in Subsection 5.2 is very fast.

5.1 ADMM for ℓ_2^2-IC/P

For $m = 2$ it holds that ℓ_2^2-IC/P can be written as a constrained problem of the form (23) with $r = 2$ which reads

$$\begin{aligned}
&\operatorname*{argmin}_{u\in\mathbb{R}^N} \Big\{\frac{1}{2}\|f-u\|_2^2 + \inf_{u=u_1+u_2}\{\alpha_1\|R_1u_1\|_{1,\omega_1} + \alpha_2\|R_2u_2\|_{1,\omega_2}\}\Big\} \qquad (25)\\
&= \operatorname*{argmin}_{u_1,u_2,z_1,z_2} \Big\{\underbrace{\frac{1}{2}\|f-u_1-u_2\|_2^2}_{F_3(v)} + \underbrace{\alpha_1\|\,|z_1|\,\|_1}_{F_1(C_1v)} + \underbrace{\alpha_2\|\,|z_2|\,\|_1}_{F_2(C_2v)}\Big\}\\
&\quad\text{subject to}\quad \underbrace{\begin{pmatrix}\widetilde{R}_1 & 0\\ 0 & \widetilde{R}_2\end{pmatrix}}_{C}\underbrace{\begin{pmatrix}u_1\\u_2\end{pmatrix}}_{v} = \underbrace{\begin{pmatrix}z_1\\z_2\end{pmatrix}}_{z},
\end{aligned}$$

where we use the notation $\widetilde{R}_i = \left(\sqrt{\omega}_{i,1}R_{i,1}^{\mathrm{T}}, \dots, \sqrt{\omega}_{i,n_i}R_{i,n_i}^{\mathrm{T}}\right)^{\mathrm{T}}$ assuming that $R_i = \left(R_{i,1}^{\mathrm{T}}, \dots, R_{i,n_i}^{\mathrm{T}}\right)^{\mathrm{T}} \in \mathbb{R}^{n_iN,N}$ with $R_{i,j} \in \mathbb{R}^{N,N}$. Here, $\|\,|\cdot|\,\|_1 := \|\cdot\|_{1,\omega}$ for $\omega = (1,\dots,1)$. Hence, the corresponding ADMM reads

$$\begin{aligned}
\begin{pmatrix}u_1^{(k+1)}\\u_2^{(k+1)}\end{pmatrix} &= \operatorname*{argmin}_{u_1,u_2}\Big\{\frac{1}{2}\|f-u_1-u_2\|_2^2 + \frac{\gamma}{2}\|b^{(k)} + C\begin{pmatrix}u_1\\u_2\end{pmatrix} - \begin{pmatrix}z_1^{(k)}\\z_2^{(k)}\end{pmatrix}\|_2^2\Big\}, \qquad (26)\\
\begin{pmatrix}z_1^{(k+1)}\\z_2^{(k+1)}\end{pmatrix} &= \operatorname*{argmin}_{z_1,z_2}\Big\{\alpha_1\|\,|z_1|\,\|_1 + \alpha_2\|\,|z_2|\,\|_1 + \frac{\gamma}{2}\|b^{(k)} + C\begin{pmatrix}u_1^{(k+1)}\\u_2^{(k+1)}\end{pmatrix} - \begin{pmatrix}z_1\\z_2\end{pmatrix}\|_2^2\Big\}, \qquad (27)\\
b^{(k+1)} &= b^{(k)} + C\begin{pmatrix}u_1^{(k+1)}\\u_2^{(k+1)}\end{pmatrix} - \begin{pmatrix}z_1^{(k+1)}\\z_2^{(k+1)}\end{pmatrix}.
\end{aligned}$$

In the first step (26) we have to solve following system of linear equations:

$$\begin{aligned}
0 &= u_1^{(k+1)} + u_2^{(k+1)} + \gamma\widetilde{R}_1^{\mathrm{T}}\widetilde{R}_1u_1^{(k+1)} - (\underbrace{f - \gamma\widetilde{R}_1^{\mathrm{T}}(b_1^{(k)} - z_1^{(k)})}_{=:t_1}),\\
0 &= u_1^{(k+1)} + u_2^{(k+1)} + \gamma\widetilde{R}_2^{\mathrm{T}}\widetilde{R}_2u_2^{(k+1)} - (\underbrace{f - \gamma\widetilde{R}_2^{\mathrm{T}}(b_2^{(k)} - z_2^{(k)})}_{=:t_2}).
\end{aligned}$$

This can be rewritten as

$$u_1^{(k+1)} = t_2 - (I + \gamma\widetilde{R}_2^{\mathrm{T}}\widetilde{R}_2)u_2^{(k+1)}, \qquad (28)$$

$$u_2^{(k+1)} = t_1 - (I + \gamma\widetilde{R}_1^{\mathrm{T}}\widetilde{R}_1)u_1^{(k+1)}. \qquad (29)$$

If we substitute (29) into (28) and solve for $u_1^{(k+1)}$ we obtain

$$\begin{aligned} u_1^{(k+1)} &= \tfrac{1}{\gamma}(\widetilde{R}_2^{\mathrm{T}}\widetilde{R}_2 + (I+\gamma\widetilde{R}_2^{\mathrm{T}}\widetilde{R}_2)\widetilde{R}_1^{\mathrm{T}}\widetilde{R}_1)^{\dagger}((I+\gamma\widetilde{R}_2^{\mathrm{T}}\widetilde{R}_2)t_1 - t_2), \\ u_2^{(k+1)} &= t_1 - (I+\gamma\widetilde{R}_1^{\mathrm{T}}\widetilde{R}_1)u_1^{(k+1)}. \end{aligned} \tag{30}$$

Observe that (25) has a unique solution with respect to u. This implies that $(u^{(k)})_{k\in\mathbb{N}}$, defined by the sequences $(u_i^{(k)})_{k\in\mathbb{N}}$, $i=1,2$ of the above ADMM via $u^{(k)} = u_1^{(k)} + u_2^{(k)}$, converges to the solution of ℓ_2^2-IC/P, cf., e.g., [33]. On the other hand, the matrix $\widetilde{R}_2^{\mathrm{T}}\widetilde{R}_2 + (I+\gamma\widetilde{R}_2^{\mathrm{T}}\widetilde{R}_2)\widetilde{R}_1^{\mathrm{T}}\widetilde{R}_1$ is not invertible in general, in other words, $u_1^{(k)}$ and $u_2^{(k)}$ are not unique in contrast to their sum $u^{(k)}$. Nevertheless, a pair of solutions $(u_1^{(k)}, u_2^{(k)})$ of (26) always exists and in our implementation we compute $u_1^{(k)}$ to be the one with minimal ℓ_2-norm, i.e., we apply the Moore-Penrose inverse. It is easy to see that this also implies that the sequences $(u_i^{(k)})_{k\in\mathbb{N}}$, $i=1,2$, converge. In the example described below, see also Table 1, we use the difference operators of case a in Example 2.2. In this case, we can solve (30) explicitely via the discrete cosine transform since both $\widetilde{R}_1^{\mathrm{T}}\widetilde{R}_1$ and $\widetilde{R}_2^{\mathrm{T}}\widetilde{R}_2$ can be diagonalized by this transformation, see, e.g., [27].

Interestingly, the second step (27) in the ADMM algorithm is very easy to compute since it can be solved separately with respect to $z_1^{(k+1)}$ and $z_2^{(k+1)}$, i.e., we have

$$z_i^{(k+1)} = \underset{z_i}{\operatorname{argmin}} \left\{\alpha_i \| \, |z_i| \, \|_1 + \frac{\gamma}{2}\|b_i^{(k)} + \widetilde{R}_i u_i^{(k+1)} - z_i\|_2^2\right\}, \quad i=1,2.$$

This problem is well-known to have the analytic solution $z_i^{(k+1)} = \mathit{shrink}_{\frac{\alpha_i}{\gamma}}(b_i^{(k)} + \widetilde{R}_i u_i^{(k+1)})$. The operator $\mathit{shrink}_\lambda : \mathbb{R}^{dM} \to \mathbb{R}^{dM}$ is called *coupled shrinkage* and given componentwise for $p^{\mathrm{T}} := \begin{pmatrix} p_1^{\mathrm{T}}, & \dots, & p_d^{\mathrm{T}} \end{pmatrix}$, $p_i := (p_{ij})_{j=1}^M$, $i=1,...,d$, by

$$\mathit{shrink}_\lambda(p_{ij}) := \begin{cases} p_{ij} - \lambda p_{ij}/\sqrt{p_{1j}^2 + \cdots + p_{dj}^2} & \text{if } \sqrt{p_{1j}^2 + \cdots + p_{dj}^2} \ge \lambda, \\ 0 & \text{otherwise.} \end{cases}$$

In summary, we obtain the following algorithm:

```
Algorithm (ADMM for $\ell_2^2$-IC/P)
Initialization: $M = \widetilde{R}_2^{\mathrm{T}}\widetilde{R}_2 + (I+\gamma\widetilde{R}_2^{\mathrm{T}}\widetilde{R}_2)\widetilde{R}_1^{\mathrm{T}}\widetilde{R}_1$, $u_i^{(0)} = \frac{1}{2}f$, $z_i^{(0)} = \frac{1}{2}\widetilde{R}_i f$, $b_i^{(0)} = 0$, $i=1,2$
For $k = 0, \dots$ repeat until a stopping criterion is reached:
    $u_1^{(k+1)} = \frac{1}{\gamma}M^{\dagger}\Big((I+\gamma\widetilde{R}_2^{\mathrm{T}}\widetilde{R}_2)\big(f - \gamma\widetilde{R}_1^{\mathrm{T}}(b_1^{(k)} - z_1^{(k)})\big) - \big(f - \gamma\widetilde{R}_2^{\mathrm{T}}(b_2^{(k)} - z_2^{(k)})\big)\Big)$
    $u_2^{(k+1)} = f - \gamma\widetilde{R}_1^{\mathrm{T}}(b_1^{(k)} - z_1^{(k)}) - (I+\gamma\widetilde{R}_1^{\mathrm{T}}\widetilde{R}_1)u_1^{(k+1)}$
    $z_i^{(k+1)} = \mathit{shrink}_{\frac{\alpha_i}{\gamma}}(b_i^{(k)} + \widetilde{R}_i u_i^{(k+1)}), \quad i=1,2$
    $b_i^{(k+1)} = b_i^{(k)} + \widetilde{R}_i u_i^{(k+1)} - z_i^{(k+1)}, \quad i=1,2$
Output: $u_1^{(k+1)}$, $u_2^{(k+1)}$, $u^{(k+1)} := u_1^{(k+1)} + u_2^{(k+1)}$
```

In the first two rows of Table 1 we compare the running times of the above ADMM algorithm with SOCP as implemented in the commercial software MOSEK 6.0 for the denoising experiment of Fig. 11. Our computations were performed with MATLAB 7.7 on an Intel Core Duo CPU with 2.66 GHz and 4GB RAM. Note that we use the difference operators of case a in Example 2.2. We report the computation times for SOCP and ADMM to reach a maximal difference in the gray value in every pixel of smaller than 1.0 and 0.1 with respect to a reference solution. Clearly, we see that ADMM is much faster for both cases. Note that the gray values of the original image in Fig. 11 range from 0 to 255 and therefore even a maximal error of 1.0 yields a solution which is visually the same as the reference solution.

5.2 PDHG for ℓ_2^2-MIC/P

For $m = 2$ problem ℓ_2^2-MIC/P can be written as

$$\begin{aligned}
&\operatorname*{argmin}_u \Big\{\frac{1}{2}\|f-u\|_2^2 + \inf_{R_1 u = x_1 + x_2} \{\alpha_1 \|x_1\|_{1,\omega_1} + \alpha_2 \|L_1 x_2\|_{1,\omega_2}\}\Big\} \\
&= \operatorname*{argmin}_{u,x_1,x_2,y} \{\underbrace{\frac{1}{2}\|f-u\|_2^2}_{F_3(v)} + \underbrace{\alpha_1 \|\,|x_1|\,\|_1}_{F_1(C_1 v)} + \underbrace{\alpha_2 \|\,|y|\,\|_1}_{F_2(C_2 v)}\} \\
&\text{subject to} \quad \underbrace{\begin{pmatrix} \widetilde{R}_1 & -\widetilde{I} \\ 0 & \widetilde{L}_1 \end{pmatrix}}_{C} \underbrace{\begin{pmatrix} u \\ x_2 \end{pmatrix}}_{v} = \underbrace{\begin{pmatrix} x_1 \\ y \end{pmatrix}}_{z},
\end{aligned}$$

where we use the notation $\widetilde{R}_1 = \big(\sqrt{\omega}_{1,1} R_{1,1}^{\mathrm{T}}, \dots, \sqrt{\omega}_{1,n_1} R_{1,n_1}^{\mathrm{T}}\big)^{\mathrm{T}}$, $\widetilde{I} = \big(\sqrt{\omega}_{1,1} I_N, \dots, \sqrt{\omega}_{1,n_1} I_N\big)^{\mathrm{T}}$ and $\widetilde{L}_1 = \big(\sqrt{\omega}_{2,1} L_{1,1}^{\mathrm{T}}, \dots, \sqrt{\omega}_{2,n_2} L_{1,n_2}^{\mathrm{T}}\big)^{\mathrm{T}}$ for $L_{1,i} \in \mathbb{R}^{N,n_1 N}$. For the above splitting, the matrix inversion which appears when applying ADMM is much harder to compute than for the problem ℓ_2^2-IC/P in Subsection 5.1. We therefore use the PDHG method for this problem. It reads

$$\begin{aligned}
\begin{pmatrix} u^{(k+1)} \\ x_2^{(k+1)} \end{pmatrix} &= \operatorname*{argmin}_{u,x_2} \{\frac{1}{2}\|f-u\|_2^2 + \frac{1}{2\tau}\| \begin{pmatrix} u \\ x_2 \end{pmatrix} - \begin{pmatrix} u^{(k)} \\ x_2^{(k)} \end{pmatrix} + \tau\gamma C^{\mathrm{T}} b^{(k)} \|_2^2\}, \\
\begin{pmatrix} x_1^{(k+1)} \\ y^{(k+1)} \end{pmatrix} &= \operatorname*{argmin}_{x_1,y} \{\alpha_1 \|\,|x_1|\,\|_1 + \alpha_2 \|\,|y|\,\|_1 + \frac{\gamma}{2}\|b^{(k)} + C \begin{pmatrix} u^{(k+1)} \\ x_2^{(k+1)} \end{pmatrix} - \begin{pmatrix} x_1 \\ y \end{pmatrix} \|_2^2\}, \\
b^{(k+1)} &= b^{(k)} + C \begin{pmatrix} u^{(k+1)} \\ x_2^{(k+1)} \end{pmatrix} - \begin{pmatrix} x_1^{(k+1)} \\ y^{(k+1)} \end{pmatrix}.
\end{aligned}$$

Now the first step is very easy to solve. We have

$$\begin{aligned}
u^{(k+1)} &= \frac{1}{1+\tau}(\tau f + u^{(k)} - \tau\gamma \widetilde{R}_1^{\mathrm{T}} b_1^{(k)}), \\
x_2^{(k+1)} &= x_2^{(k)} + \tau\gamma(\tilde{I} b_1^{(k)} - \widetilde{L}_1^{\mathrm{T}} b_2^{(k)}).
\end{aligned}$$

In the second step, the minimization with respect to x_1 and y decouples again and we can solve the corresponding problems in the same way as in Subsection 5.1 using the coupled shrinkage operator.

In summary, we obtain the following algorithm:

Algorithm (PDHG for (ℓ_2^2-MIC/P))

Initialization: $u^{(0)} = f$, $x_1^{(0)} = \frac{1}{2}\widetilde{R}_1 f$, $x_2^{(0)} = \frac{1}{2}\widetilde{L}_1\widetilde{R}_1 f$, $b_1^{(0)} = b_2^{(0)} = 0$

For $k = 0, \ldots$ repeat until a stopping criterion is reached:

$$\begin{aligned}
u^{(k+1)} &= \frac{1}{1+\tau}(\tau f + u^{(k)} - \tau\gamma \widetilde{R}_1^{\mathrm{T}} b_1^{(k)}) \\
x_2^{(k+1)} &= x_2^{(k)} + \tau\gamma(\tilde{I} b_1^{(k)} - \widetilde{L}_1^{\mathrm{T}} b_2^{(k)}) \\
x_1^{(k+1)} &= shrink_{\frac{\alpha_1}{\gamma}}(b_1^{(k)} + \widetilde{R}_1 u^{(k+1)} - x_2^{(k+1)}) \\
y^{(k+1)} &= shrink_{\frac{\alpha_2}{\gamma}}(b_2^{(k)} + \widetilde{L}_1 x_2^{(k+1)}) \\
b_1^{(k+1)} &= b_1^{(k)} + \widetilde{R}_1 u^{(k+1)} - x_2^{(k+1)} - x_1^{(k+1)} \\
b_2^{(k+1)} &= b_2^{(k)} + \widetilde{L}_1 x_2^{(k+1)} - y^{(k+1)}
\end{aligned}$$

Output: $u^{(k+1)}$, $x_1^{(k+1)}$, $x_2^{(k+1)}$

The last two rows of Table 1 illustrate that this algorithm is much faster than SOCP via MOSEK. Note that the pairs (τ, γ) used to obtain the results in Table 1 do not satisfy the assumption $\tau\gamma < \frac{1}{\|C\|^2}$ of the convergence proof in [9]. However, we use them since the resulting algorithms still seem to converge and are much faster. Similar observations were reported in [18].

		max. error			
		< 1.0		< 0.1	
ℓ_2^2-IC/P	SOCP	73.5 sec		81.4 sec	
	ADMM	3.6 sec	145 iter.	19.1 sec	806 iter.
ℓ_2^2-MIC/P	SOCP	18.0 sec		42.8 sec	
	PDHG	1.7 sec	162 iter.	19.6 sec	1929 iter.

Table 1: Computation time to achieve a maximal difference smaller than 1.0 and 0.1 in each pixel with respect to a reference solution for the experiment shown in Fig. 11. We use the difference operators of case a in Example 2.2. For ADMM the best step length parameters γ were found to be 7.6 and 25.1 for a maximal error of 1.0 and 0.1, respectively. The best combinations of τ and γ in the PDHG algorithm are $(0.07, 5.9)$ and $(0.04, 9.1)$, respectively.

6 Numerical examples

Finally, we want to illustrate the differences between the ℓ_2^2-IC and ℓ_2^2-MIC models in two dimensions by numerical examples.

Note that in our numerical examples the ℓ_2^2-IC and ℓ_2^2-MIC models corresponding to the different cases a and b described in the Examples 2.2 and 3.1 show only marginal differences as depicted in Fig. 4. For this reason, we restrict our attention to the difference operators of case a in the numerical experiments. Furthermore, we concentrate on the practically important case $m = 2$. For experiments with $m = 3$ we refer to [5].

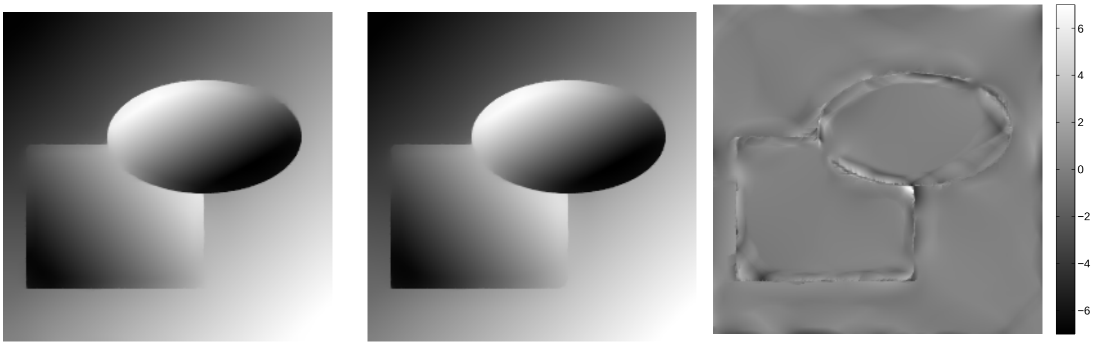


Figure 4: Left: Result $\hat{u}_{\mathrm{MIC,a}}$ from Fig. 2 (bottom right) obtained by ℓ_2^2-MIC with $R_1 = \mathcal{D}_1$, $\alpha_1 = 60$ and $R_2 = \mathcal{D}_{2,a}$, $\alpha_2 = 300$. Middle: Result $\hat{u}_{\mathrm{MIC,b}}$ of the same experiment with $R_2 = \mathcal{D}_{2,b}$, $\alpha_2 = 260$. Since $\|R_{2,a}u\|_{1,\omega_{2a}} \leq \|R_{2,b}u\|_{1,\omega_{2b}}$ with $\omega_{2a} = (1,1)$ and $\omega_{2b} = (1, \frac{1}{2}, 1)$, the value of α_2 has been adjusted. The difference image $\hat{u}_{\mathrm{MIC,a}} - \hat{u}_{\mathrm{MIC,b}}$ on the right shows that there are small differences between these two images which are in the images themselves hard to recognize at all.

Example 6.1.

We start with the original and noisy images in Fig. 5.

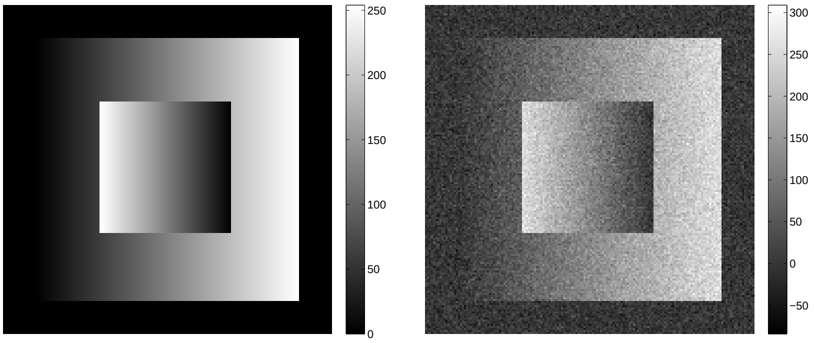


Figure 5: Original image u (left) and noisy image f (right) corrupted by additive Gaussian noise of standard deviation 20.

In this example we study the difference between the penalizers

$$\begin{aligned}\Phi_{IC}(u) &= \inf_{u=u_1+u_2}\{\alpha_1\|R_1u_1\|_{1,\omega_1}+\alpha_2\|R_2u_2\|_{1,\omega_2}\}\\ &= \inf_{\substack{R_1u=x_1+x_2\\ x_i\in\mathcal{R}(R_1)}}\{\alpha_1\|x_1\|_{1,\omega_1}+\alpha_2\|L_1x_2\|_{1,\omega_2}\},\\ \Phi_{MIC}(u) &= \inf_{\substack{u=u_1+u_2\\ s_1\in\mathcal{N}(R_1^{\mathrm{T}})}}\{\alpha_1\|R_1u_1-s_1\|_{1,\omega_1}+\alpha_2\|R_2u_2+L_1s_1\|_{1,\omega_2}\}\\ &= \inf_{R_1u=x_1+x_2}\{\alpha_1\|x_1\|_{1,\omega_1}+\alpha_2\|L_1x_2\|_{1,\omega_2}\}.\end{aligned}$$

Figs. 6 and 7 show decompositions by Φ_{IC} and Φ_{MIC} of the image u given in Fig. 5 (left). Note that for a better visual impression the gray values of the images depicting R_1u, x_1 and x_2 are restricted to the interval $[-10,10]$ and all values outside of this interval are represented by the gray values -10 and 10.

In image restoration the aim of a regularization term is usually to penalize the noise contained in u without penalizing structures of the original noisefree image. This example will show that for appropriate α_1, α_2, the functional

Φ_{MIC} penalizes linear regions of our noisefree test image much less than Φ_{IC}.

In the first row of Figs. 6 and 7, we can see that the images u_1 and u_2 look quite similar for both functionals. However, the decompositions of R_1u into the vectors $x_i=\left(x_{i,1}^{\mathrm{T}},\ x_{i,2}^{\mathrm{T}}\right)^{\mathrm{T}}$ for $i=1,2$ depicted in the second and third row of Figs. 6 and 7 are fundamentally different.

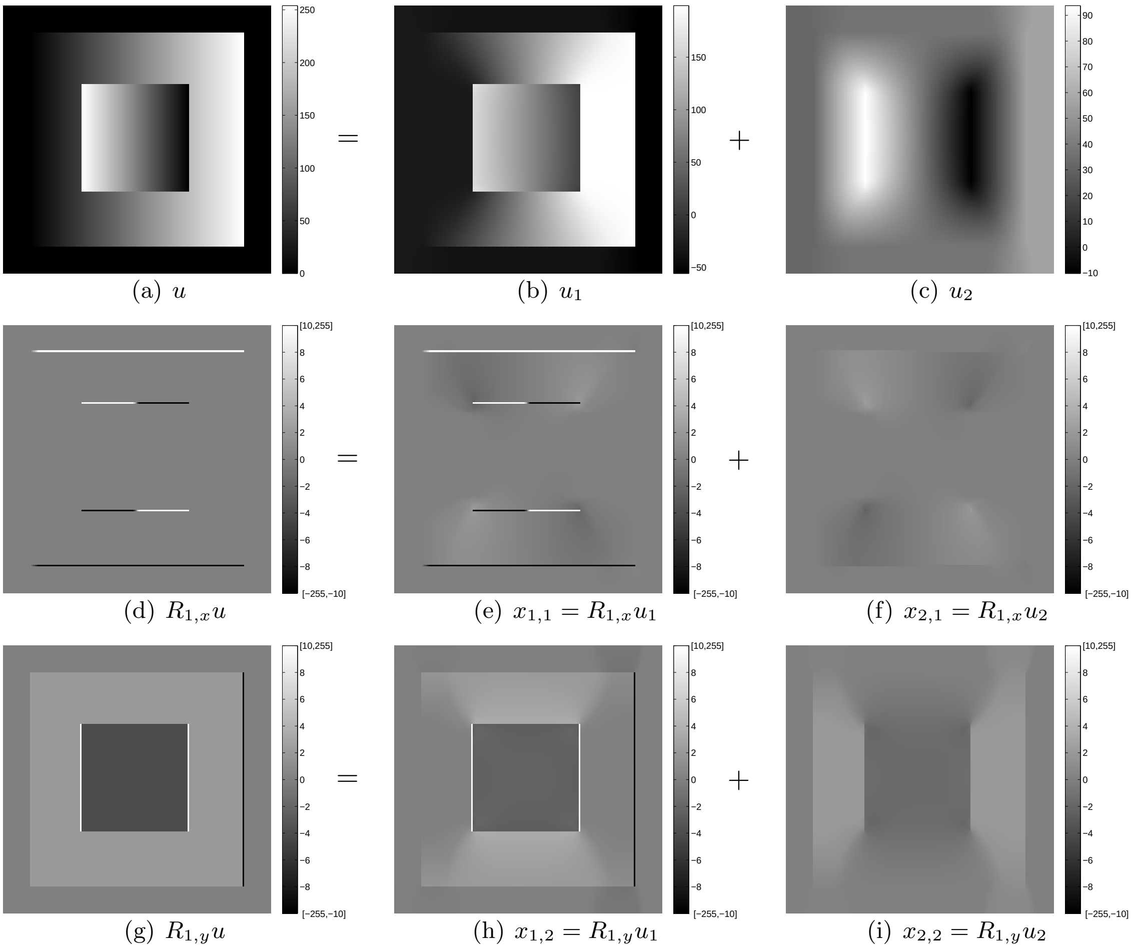
In $\Phi_{MIC}(u)$ the additional variable s_1 allows for a decomposition such that $x_1=R_1u_1-s_1$ contains only the gradients of the edges whereas $x_2=R_1u_2+s_1$ comprises the gradients of the linear parts, see also Fig. 8. Hence, by $\alpha_1\|x_1\|_{1,\omega_1}$ the functional $\Phi_{MIC}(u)$ penalizes only the gradients at the edges and since within linear regions the second derivatives are zero, $\alpha_2\|L_1x_2\|_{1,\omega_2}$ penalizes only the boundaries of the linear regions of u.

In contrast, for $\Phi_{IC}(u)$ it is not possible to choose the same x_i, $i=1,2$, due to the restriction that x_i must be in $\mathcal{R}(R_1)$, or, equivalently, the absence of the variable s_1. Thus, we see in the second and third row of Fig. 6 that x_1 and x_2 do not separate R_1u into a part which contains the gradients of the edges and the linear components, respectively. Especially $x_{1,2}$ comprises a significant part of the gradients of the linear regions which is then penalized by $\alpha_1\|x_1\|_{1,\omega_1}$. This leads to a higher value of the penalizer $\Phi_{IC}(u)$ compared to the value of $\Phi_{MIC}(u)$, i.e. for appropriate α_i the functional Φ_{IC} wrongly penalizes linear regions a lot more than $\Phi_{MIC}(u)$ does.

Example 6.2.

Our next Figs. 9 and 10 illustrate what happens if we apply (ℓ_2^2-IC/P) and (ℓ_2^2-MIC/P) to the noisy image depicted in Fig. 5 (right). First of all, a slight smoothing of the edges of the restored image $\hat{u}$ is visible for both problems, in particular if we look at the images of $R_{1,x}u$ and $R_{1,y}u$. However, due to our choice of $\alpha_1<\alpha_2$ this smoothing is of minor extent so that it is hardly visible by looking at the restored image $\hat{u}$ compared to the original image u. Visually more eye-catching are the staircasing artifacts of the restoration result of (ℓ_2^2-IC/P). These artifacts can be explained as follows: If we assume that u_2 is given as in Fig. 9, then u_1 is the solution of the functional

$$\inf_{u_1}\{\frac{1}{2}\|(f-u_2)-u_1\|_2^2+\alpha_1\|R_1u_1\|_{1,\omega_1}\}.$$



(a) u (b) u_1 (c) u_2

(d) $R_{1,x}u$ (e) $x_{1,1} = R_{1,x}u_1$ (f) $x_{2,1} = R_{1,x}u_2$

(g) $R_{1,y}u$ (h) $x_{1,2} = R_{1,y}u_1$ (i) $x_{2,2} = R_{1,y}u_2$

Figure 6: Decomposition of the original images u and $R_1 u$ in Fig. 5 by $\Phi_{IC}(u)$ with $\alpha_1 = 60$ and $\alpha_2 = 150$.

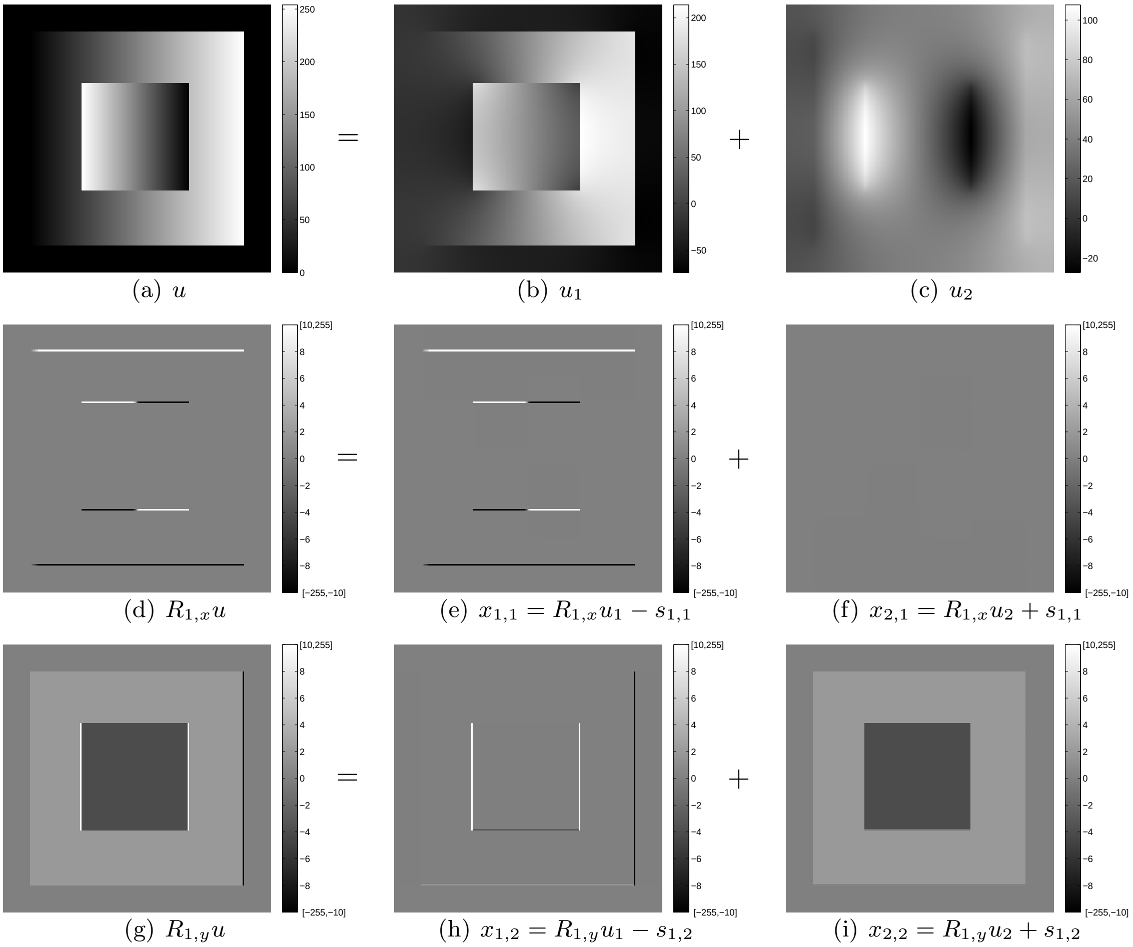
This functional is nothing else than the ROF functional applied to $f - u_2$, which is known to produce staircasing at linear regions of $\hat{u} - u_2$. By choosing a larger α_1 and thus, bringing u_2 closer to u, these artifacts can be reduced but visible blurring artifacts at the edges are introduced. In contrast to (ℓ_2^2-IC/P) the result of (ℓ_2^2-MIC/P) is nearly perfect without any staircasing. The reason for this is that all gradients at the linear regions of the original image u are contained in x_2 rather than x_1.

Example 6.3.

For natural images, the ℓ_2^2-IC and the ℓ_2^2-MIC approach with ordinary difference matrices work quite similar and for most images there will be no visual differences. The image of a car shown in Fig. 11 contains affine sets and sharp edges so that the ℓ_2^2-MIC approach is again superior to ℓ_2^2-IC.

Example 6.4.

Finally, we give a denoising example for ℓ_2^2-IC/MIC with the difference matrices $D(\xi_1)$ and $D(\xi_1, \xi_2)$ in Fig. 12, where $\xi_1 = 0.03$ and $\xi_2 = -0.03$. We mention that the denoised image

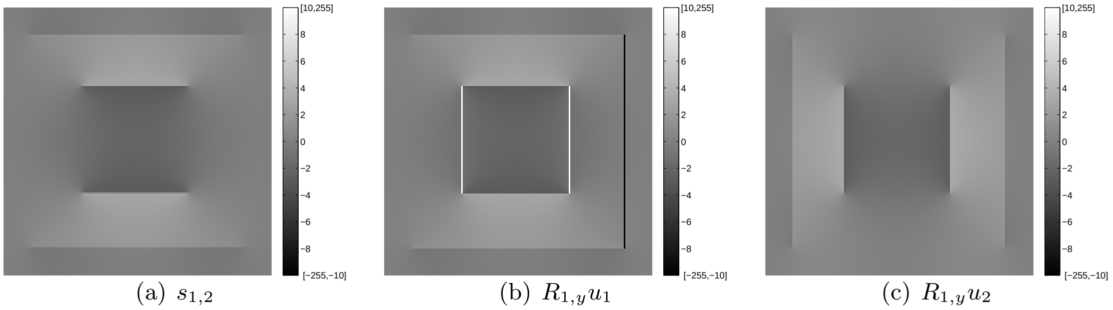


(a) u (b) u_1 (c) u_2

(d) $R_{1,x}u$ (e) $x_{1,1} = R_{1,x}u_1 - s_{1,1}$ (f) $x_{2,1} = R_{1,x}u_2 + s_{1,1}$

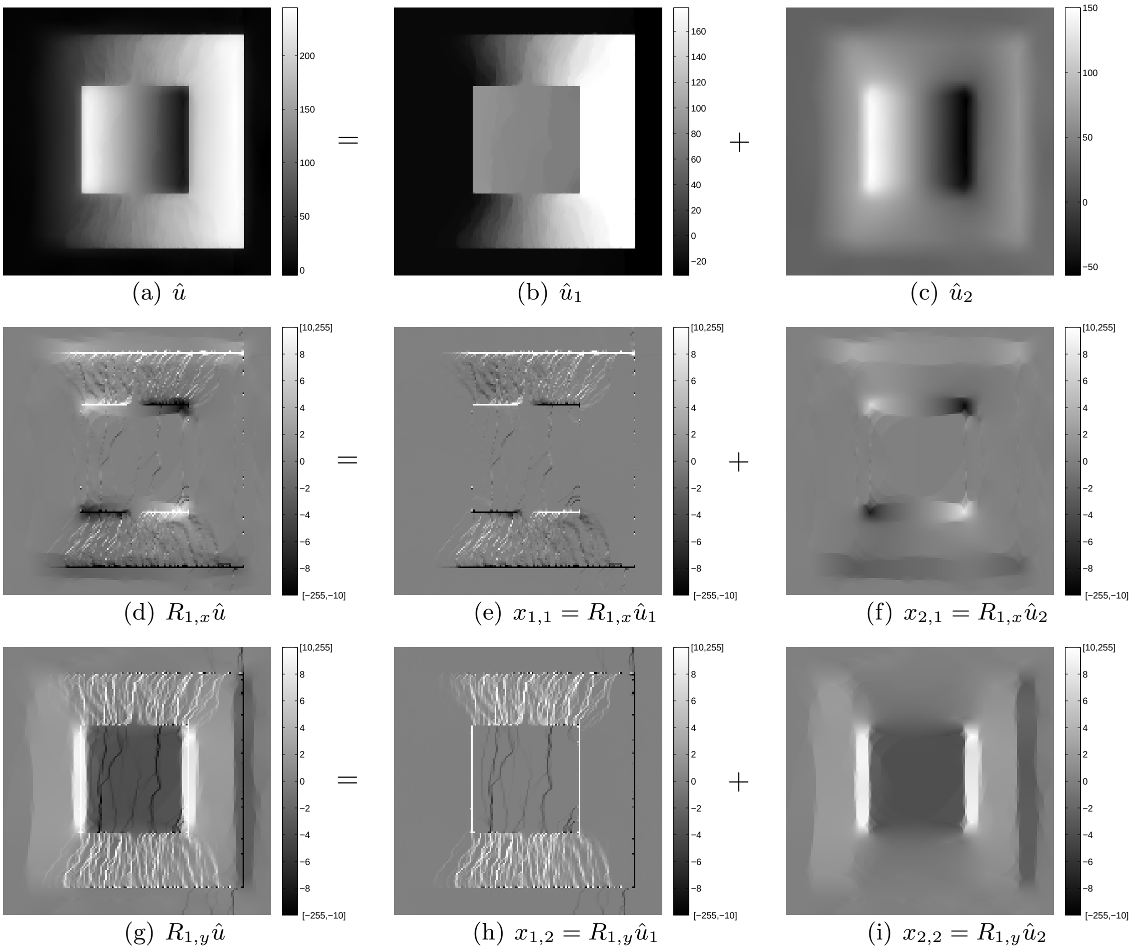
(g) $R_{1,y}u$ (h) $x_{1,2} = R_{1,y}u_1 - s_{1,2}$ (i) $x_{2,2} = R_{1,y}u_2 + s_{1,2}$

Figure 7: Decomposition of the original images u and $R_1 u$ in Fig. 5 by $\Phi_{MIC}(u)$ with $\alpha_1 = 60$ and $\alpha_2 = 150$.



(a) $s_{1,2}$ (b) $R_{1,y}u_1$ (c) $R_{1,y}u_2$

Figure 8: The components $s_{1,2}$, $R_{1,y}u_1$ and $R_{1,y}u_2$ which allow the favorable decomposition $R_{1,y}u = x_{1,2} + x_{2,2}$ depicted in the third row of Figure 7.



(a) $\hat{u}$ (b) $\hat{u}_1$ (c) $\hat{u}_2$

(d) $R_{1,x}\hat{u}$ (e) $x_{1,1} = R_{1,x}\hat{u}_1$ (f) $x_{2,1} = R_{1,x}\hat{u}_2$

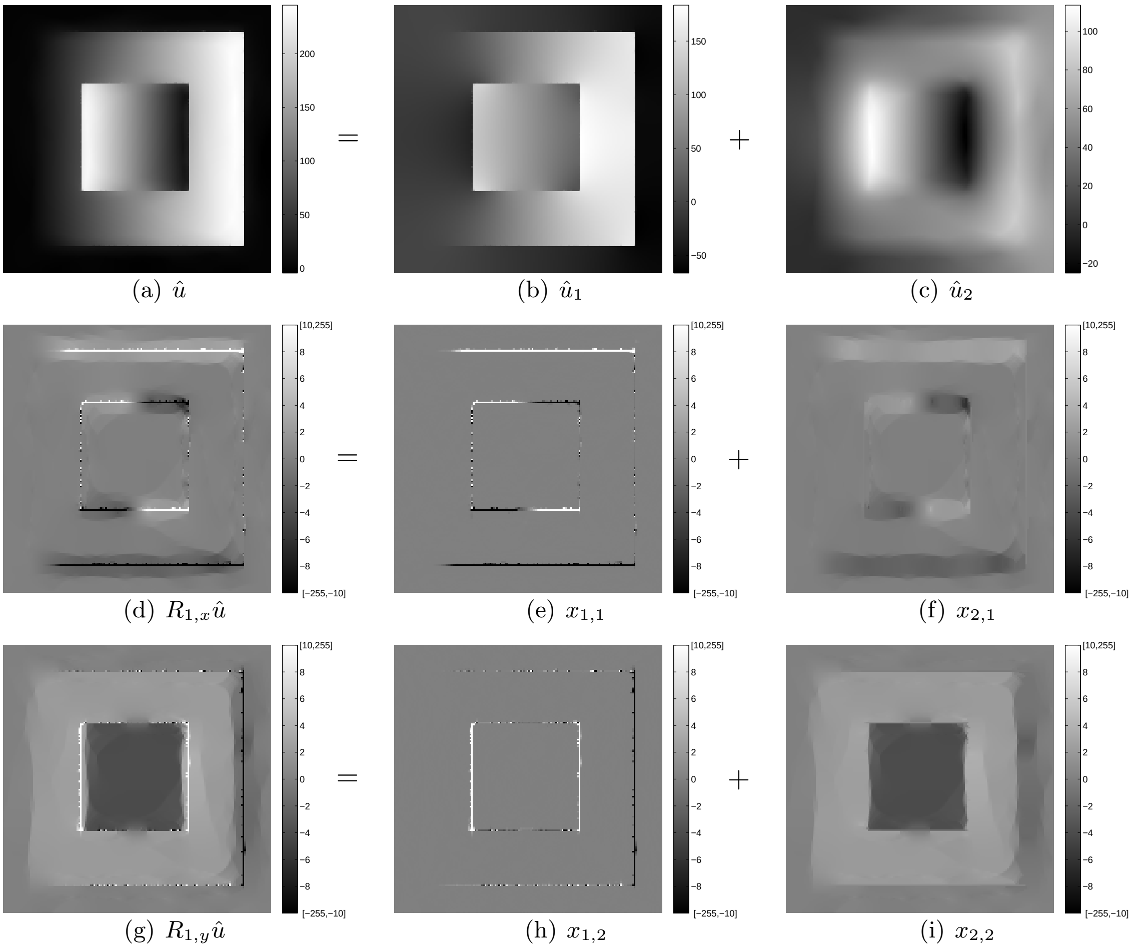
(g) $R_{1,y}\hat{u}$ (h) $x_{1,2} = R_{1,y}\hat{u}_1$ (i) $x_{2,2} = R_{1,y}\hat{u}_2$

Figure 9: Results of (ℓ_2^2-IC/P) applied to the noisy image f in Fig. 5 (right) for $\alpha_1 = 60$ and $\alpha_2 = 150$.

for ℓ_2^2-MIC with D_i, $i = 1, 2$ (i.e. $\xi_1 = \xi_2 = 0$) looks quite similar while ℓ_2^2-IC with these matrices shows fewer staircasing effects.

7 Conclusions

We have presented a general discrete approach to modify infimal convolutions containing ℓ_1-type functionals with linear operators. For the special case of finite difference matrices we obtain the results from our previous paper [34] and a discrete version of [5]. However, in contrast to [34], we also considered the primal problem which in our opinion shows better the differences between the original and the modified version. We illustrated these differences by numerical examples showing also decompositions of the primal variables appearing in the functional. An open question is role of different factorizations in (6). Besides, it remains to examine other useful operators for image processing tasks as, e.g., frame analysis operators. A first step in this direction was done by considering more general difference matrices known from $\mathcal{L}$ splines. This paper also contributes to finding fast algorithms to solve problems with



(a) $\hat{u}$ (b) $\hat{u}_1$ (c) $\hat{u}_2$

(d) $R_{1,x}\hat{u}$ (e) $x_{1,1}$ (f) $x_{2,1}$

(g) $R_{1,y}\hat{u}$ (h) $x_{1,2}$ (i) $x_{2,2}$

Figure 10: Results of (ℓ_2^2-MIC/P) applied to the noisy image f in Fig. 5 (right) $\alpha_1 = 60$ and $\alpha_2 = 150$.

infimal convolutions containing ℓ_1-type functionals. In particular, we apply two splitting methods, the alternating direction method of multipliers and the primal-dual hybrid gradient algorithm. Both of them use the additive structure of our objective functions and solve in each iteration subproblems corresponding to these terms. We show numerically that the resulting algorithms are much faster than the commercial software MOSEK which implements second order cone programming.

References

[1] J.-F. Aujol, G. Aubert, L. Blanc-Féraud, and A. Chambolle. Image decomposition into a bounded variation component and an oscillating component. *Journal of Mathematical Imaging and Vision*, 22(1):71–88, 2005.

[2] J.-F. Aujol and A. Chambolle. Dual norms and image decomposition models. *International Journal of Computer Vision*, 63(1):85–104, 2005.

[3] A. Beck and M. Teboulle. Fast gradient-based algorithms for constrained total variation image denoising and deblurring problems. *IEEE Transactions on Image Processing*, 18(11):2419–2434, 2009.

[4] A. Beck and M. Teboulle. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM Journal on Imaging Sciences*, 2(1):183–202, 2009.

[5] K. Bredies, K. Kunisch, and T. Pock. Total generalized variation. *SIAM Journal on Imaging Sciences*, 2009. Accepted.

[6] Y. Censor and A. Lent. An iterative row-action method for interval convex programming. *Journal of Optimization Theory and Applications*, 34(3):321–353, 1981.

[7] A. Chambolle. An algorithm for total variation minimization and applications. *Journal of Mathematical Imaging and Vision*, 20(1-2):89–97, 2004.

[8] A. Chambolle and P.-L. Lions. Image recovery via total variation minimization and related problems. *Numerische Mathematik*, 76:167–188, 1997.

[9] A. Chambolle and T. Pock. A first-order primal-dual algorithm for convex problems with applications to imaging. Preprint, University of Graz, May 2010.

[10] T. F. Chan, S. Esedoglu, and F. Park. A fourth order dual method for staircase reduction in texture extraction and image restoration problems. CAM Report 05-28, UCLA, 2005.

[11] T. F. Chan, A. Marquina, and P. Mulet. High-order total variation-based image restoration. *SIAM Journal on Scientific Computing*, 22(2):503–516, 2000.

[12] P. L. Combettes. Solving monotone inclusions via compositions of nonexpansive averaged operators. *Optimization*, 53(5–6):475–504, 2004.

[13] P. L. Combettes and J.-C. Pesquet. A Douglas-Rachford splitting approach to nonsmooth convex variational signal recovery. *IEEE Journal of Selected Topics in Signal Processing*, 1(4):564–574, 2007.

[14] S. Didas, G. Steidl, and S. Setzer. Combined ℓ_2 data and gradient fitting in conjunction with ℓ_1 regularization. *Advances in Computational Mathematics*, 30(1):79–99, 2009.

[15] S. Didas, J. Weickert, and B. Burgeth. Properties of higher order nonlinear diffusion filtering. *Journal of Mathematical Imaging and Vision*, 35:208–226, 2009.

[16] J. Eckstein and D. P. Bertsekas. On the Douglas–Rachford splitting method and the proximal point algorithm for maximal monotone operators. *Mathematical Programming*, 55:293–318, 1992.

[17] E. Esser. Applications of Lagrangian-based alternating direction methods and connections to split Bregman. CAM Report 09-31, UCLA, 2009.

[18] E. Esser, X. Zhang, and T. F. Chan. A general framework for a class of first order primal-dual algorithms for TV minimization. CAM Report 09-67, UCLA, 2009.

[19] D. Gabay. Applications of the method of multipliers to variational inequalities. In M. Fortin and R. Glowinski, editors, *Augmented Lagrangian Methods: Applications to the Numerical Solution of Boundary–Value Problems*, volume 15 of *Studies in Mathematics and its Applications*, chapter 9, pages 299–331. North–Holland, Amsterdam, 1983.

[20] D. Gabay and B. Mercier. A dual algorithm for the solution of nonlinear variational problems via finite element approximation. *Computers & Mathematics with Applications*, 2:17–40, 1976.

[21] T. Goldstein and S. Osher. The Split Bregman method for L1-regularized problems. *SIAM Journal on Imaging Sciences*, 2(2):323–343, 2009.

[22] W. Hintermüller and K. Kunisch. Total bounded variation regularization as a bilaterally constrained optimization problem. *SIAM Journal on Applied Mathematics*, 64(4):1311–1333, 2004.

[23] J. M. Hyman and M. J. Shashkov. Natural discretizations for the divergence, gradient, and curl on logically rectangular grids. *Computers & Mathematics with Applications*, 33(4):81–104, 1997.

[24] M. Lysaker, A. Lundervold, and X. Tai. Noise removal using fourth-order partial differential equations with applications to medical magnetic resonance images in space and time. *IEEE Transactions on Image Processing*, 12(12):1579 – 1590, 2003.

[25] J. Ma and G. Plonka-Hoch. Curvelet-wavelet regularized split Bregman iteration for compressed sensing. *International Journal of Wavelets, Multiresolution and Information Processing*, 2010. Accepted.

[26] Y. E. Nesterov. A method of solving a convex programming problem with convergence rate $O(1/k^2)$. *Soviet Mathematics Doklady*, 27(2):372–376, 1983.

[27] D. Potts and G. Steidl. Optimal trigonometric preconditioners for nonsymmetric Toeplitz systems. *Linear Algebra and its Applications*, 281:265–292, 1998.

[28] R. T. Rockafellar. *Convex Analysis.* Princeton University Press, Princeton, 1970.

[29] L. I. Rudin, S. Osher, and E. Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D*, 60:259–268, 1992.

[30] O. Scherzer. Denoising with higher order derivatives of bounded variation and an application to parameter estimation. *Computing*, 60:1–27, 1998.

[31] O. Scherzer, M. Grasmair, H. Grossauer, M. Haltmeier, and F. Lenzen. *Variational Methods in Imaging*, volume 167 of *Applied Mathematical Sciences.* Springer, 2009.

[32] L. L. Schumaker. *Spline Functions: Basic Theory.* Wiley and Sons, New York, 1981.

[33] S. Setzer. Operator splittings, Bregman methods and frame shrinkage in image processing. *International Journal of Computer Vision*, 2009. Accepted.

[34] S. Setzer and G. Steidl. Variational methods with higher order derivatives in image processing. In M. Neamtu and L. L. Schumaker, editors, *Approximation XII, San Antonio, USA*, pages 360–386. Nashboro Press, Brentwood, 2008.

[35] S. Setzer, G. Steidl, B. Popilka, and B. Burgeth. Variational methods for denoising matrix fields. In D. H. Laidlaw and J. Weickert, editors, *Visualization and Processing of Tensor Fields, Advances and Perspectives*, Mathematics and Visualization, pages 341–360. Springer, Berlin, 2009.

[36] J. L. Starck, M. Elad, and D. L. Donoho. Image decomposition: separation from texture from piecewise smooth content. *IEEE Transactions on Image Processing*, 14(10):1570–1582, 2005.

[37] T. Strömberg. *The operation of infimal convolution.* Ph.D. Thesis. Polska Akademia Nauk, Warszawa, Poland, 1996.

[38] M. Unser and T. Blu. Cardinal exponential splines: part i - theory and filtering algorithms. *IEEE Transactions on Signal Processing*, 53(4):1425–1438, 2005.

[39] M. Unser and T. Blu. Generalized smoothing splines and the optimal discretization of the Wiener filter. *IEEE Transactions on Signal Processing*, 53(6):2146–2159, 2005.

[40] Y. Wang, J. Yang, W. Yin, and Y. Zhang. A new alternating minimization algorithm for total variation image reconstruction. *SIAM Journal on Imaging Sciences*, 1(3):248–272, 2008.

[41] Y.-L. You and M. Kaveh. Fourth-order partial differential equations for noise removal. *IEEE Transactions on Image Processing*, 9(10):1723–1730, 2000.

[42] J. Yuan, C. Schnörr, and G. Steidl. Simultaneous higher order optical flow estimation and decomposition. *SIAM Journal on Scientific Computing*, 29(6):2283 – 2304, 2007.

[43] J. Yuan, C. Schnörr, and G. Steidl. Total-variation based piecewise affine regularization. In A. Lie, M. Lysaker, K. Morken, and X.-C. Tai, editors, *Second International Conference on Scale Space Methods and Variational Methods in Computer Vision, SSVM 2009, Voss, Norway, June 1-5, 2009. Proceedings*, volume 5567 of *Lecture Notes in Computer Science*, pages 552–564. Springer, 2009.

[44] M. Zhu and T. F. Chan. An efficient primal-dual hybrid gradient algorithm for total variation image restauration. CAM Report 08-34, UCLA, 2008.

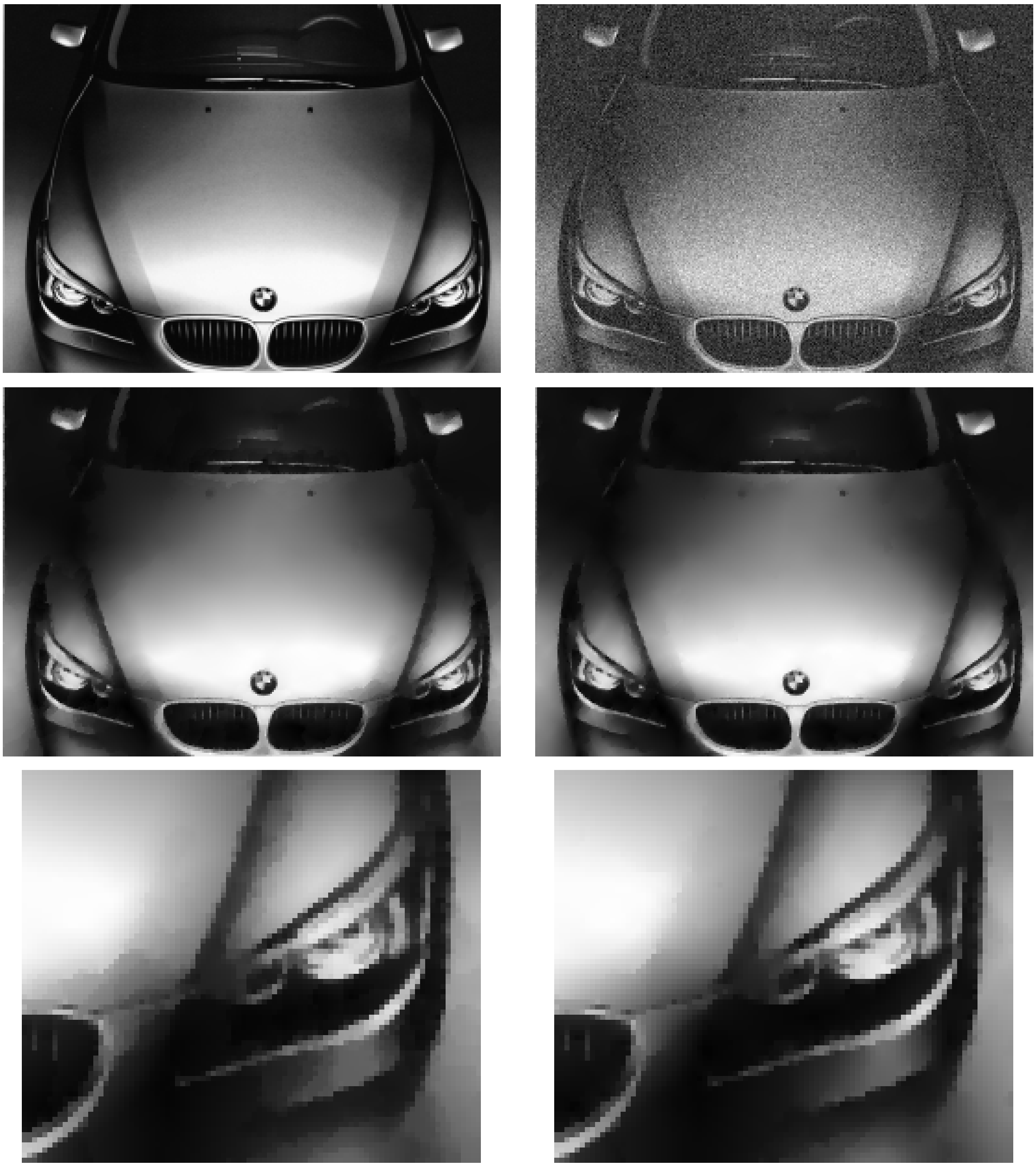


Figure 11: Top: Original image u (left), image size: 200×270, copyright P. Allert, Allert and Hoess Photography GbR, München, and noisy image f (right) corrupted by additive Gaussian noise of standard deviation 20. Middle: Denoised images by ℓ_2^2-IC (left) and ℓ_2^2-MIC (right) with ordinary difference matrices D_i, $i = 1, 2$ with $\alpha_1 = 23$ and $\alpha_2 = 60$. Bottom: Part of the denoised images by ℓ_2^2-IC (left) and ℓ_2^2-MIC (right).

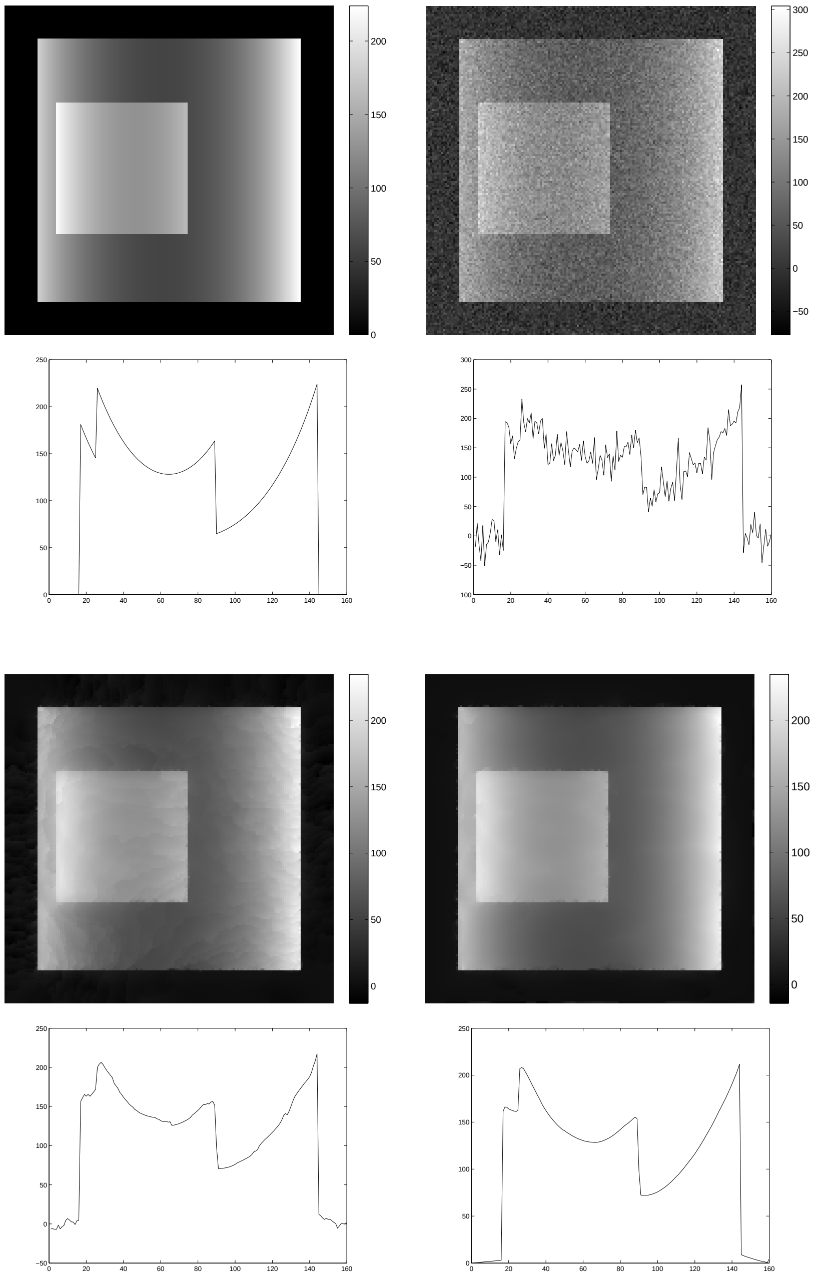


Figure 12: First row: Original image u (left) and noisy image f (right) corrupted by additive Gaussian noise of standard deviation 20. Second row: Plots of the 99th row of the images in the first row. Third row: Denoised images by ℓ_2^2-IC (left) and ℓ_2^2-MIC (right) for $\alpha_1 = 27$ and $\alpha_2 = 100$. Fourth row: Plots of the 99th row of the images in the third row.