Total Variation Denoising using Iterated Conditional Expectation

Cécile Louchet, Lionel Moisan

TOTAL VARIATION DENOISING USING ITERATED CONDITIONAL EXPECTATION

Cécile Louchet

Université d'Orléans
MAPMO, CNRS UMR 7349, France

Lionel Moisan

Université Paris Descartes
MAP5, CNRS UMR 8145, France

ABSTRACT

We propose a new variant of the celebrated Total Variation image denoising model of Rudin, Osher and Fatemi, which provides results very similar to the Bayesian posterior mean variant (TV-LSE) while showing a much better computational efficiency. This variant is based on an iterative procedure which is proved to converge linearly to a fixed point satisfying a marginal conditional mean property. The implementation is simple, provided numerical precision issues are correctly handled. Experiments show that the proposed variant yields results that are very close to those obtained with TV-LSE and avoids as well the so-called staircasing artifact observed with classical Total Variation denoising.

Index Terms— image denoising; total variation; posterior mean; marginal conditional mean; staircasing effect.

1. INTRODUCTION

Since the seminal paper of Rudin, Osher and Fatemi (ROF) [1], the use of Total Variation (TV) as a regularization term has become extremely popular in image-related inverse problems (see [2] and references therein). Among the reasons of this success are the ability of TV to allow for image discontinuities while penalizing unwanted oscillations and noise, and the deep theoretical knowledge and possibilities of investigation concerning functions with finite TV [3]. In the last decade, the design of fast optimization algorithms able to deal with the non-differentiability of the TV functional has made this model even more efficient and attractive [4, 5].

The major drawback of the TV model is the so-called staircasing effect, that is, the appearance of constant zones separated by artificial boundaries in TV-regularized images [6]. Several variants have been proposed to avoid this effect, and most of them propose to smooth the singularity of the TV functional, which is known to be responsible for the staircasing effect [7]. Another possibility consists in keeping the true TV term (and the nice associated mathematical properties) but changing the model framework. Exploiting this idea, the TV-LSE variant [8, 9] considers a Bayesian framework, and replaces the maximum a posteriori (MAP) estimate (which corresponds to a classical interpretation of the original ROF model) by the posterior mean estimate, which better accounts for the global shape of the posterior distribution and prevents the appearance of the staircasing structures favored by the density maximization (MAP) criterion. Unfortunately, the numerical computation of this TV-LSE variant requires the use of a Monte-Carlo Markov Chain (MCMC) algorithm that is computationally heavy, which can be a major inconvenience for some applications. Here, we propose a new variant that can be considered as a simplification of TV-LSE. Whereas in TV-LSE denoising the intensity of each pixel is a weighted average over all possible images, we propose to compute the posterior mean conditionally to the value of all other pixels. Of course, the result depends on the values chosen for the other pixels, but the procedure can be iterated simultaneously on all pixels to converge to a steady state that does not depend on the initialization.

Some instances of this iterative approach can be found in the image processing literature. The Iterated Conditional Expectation (ICE) that we consider here was proposed as a smart counterpart to the LSE estimate [10]. Recently, ICE has been successfully used to improve the convergence of Belief Propagation techniques [11] and to fill-in lacking data [12]. Note that these works do not provide a convergence proof of the iterated procedure.

In Section 2, we define the TV-ICE model and show the linear convergence of the iterated scheme. We also provide a closed-form formula for the conditional expectation operator and derive several useful properties. The numerical issues brought by the particular form of the operator are discussed and solved in Section 3. A comparison of denoising results obtained with ROF, TV-LSE and TV-ICE is then made, showing the visual absence of staircasing and a striking similarity between TV-ICE and TV-LSE results. Finally, the comparison of the convergence rates reveals that TV-ICE requires much less iterations than the best known iterative algorithms for ROF (and TV-LSE) denoising.

2. DEFINITION AND PROPERTIES OF TV-ICE

2.1. Bayesian formulation of the ROF model

Let us consider a (supposedly noisy) discrete gray-level image $v : \Omega \to \mathbb{R}$, defined on a rectangular domain $\Omega \subset \mathbb{Z}^2$. For any pixel $x \in \Omega$, x^c denotes $\Omega \setminus \{x\}$, and $\mathcal{N}_x$ represents

the neighbors of x. In all the following, we shall focus on the usual 4 neighbors (the pixels y such that $|y-x|=1$), with a mirror or periodic boundary condition to deal with pixels that lie on the boundary of Ω.

The ROF model proposes to compute a denoised version of v as the unique minimizer $u=\hat{u}_{\text{ROF}}$ of the convex energy

$$E_{\text{ROF}}(u)=\|u-v\|^2+\lambda TV(u) \tag{1}$$

where λ is a positive parameter that selects the desired level of regularization, and $TV(u)$ is the discrete total variation of u. To obtain explicit formulae, we will consider in the following the discrete anisotropic total variation

$$TV(u)=\frac{1}{2}\sum_{x\in\Omega}\sum_{y\in\mathcal{N}_x}|u(y)-u(x)|, \tag{2}$$

which is known to produce results similar to the Euclidean scheme classically used for the ROF model.

The Bayesian model associated to the (improper) prior $p(u)\propto e^{-\beta TV(u)}$ (for some $\beta>0$) and a Gaussian additive noise model $p(v|u)\propto e^{-\|u-v\|^2/(2\sigma^2)}$ (where $\sigma>0$ is the noise standard deviation) leads, thanks to Bayes rule, to the maximum a posteriori (MAP) estimate

$$\begin{aligned}\arg\max_u p(u|v)&=\arg\max_u p(v|u)p(u)\\&=\arg\max_u e^{-\frac{\|u-v\|^2}{2\sigma^2}}e^{-\beta TV(u)}\end{aligned}$$

which is precisely $\hat{u}_{\text{ROF}}$ when $\lambda=2\sigma^2\beta$. Hence, $\hat{u}_{\text{ROF}}$ is the mode of the posterior distribution

$$\pi(u)=p(u|v)\propto\exp\left(-\frac{\|u-v\|^2+\lambda TV(u)}{2\sigma^2}\right). \tag{3}$$

2.2. The TV-LSE and TV-ICE variants

In [9], the expectation of the posterior distribution π (that is, the image that reaches the Least Square Error under π) is shown to be a better denoiser than its mode (the MAP estimate $\hat{u}_{\text{ROF}}$), in particular for it avoids the staircasing effect [13]. It can be written

$$\hat{u}_{\text{LSE}}=\mathbb{E}_{u\sim\pi}(u)=\int_{\mathbb{R}^\Omega}\pi(u)\,u\,du \tag{4}$$

and depends on both λ and σ. The main limitation of this TV-LSE model is that it is computationally expensive, as heavy computations are required to estimate the solution $\hat{u}_{\text{LSE}}$ (the convergence rate of the MCMC Hastings-Metropolis algorithm is $\mathcal{O}(n^{-1/2})$, which is slow).

We can remark that for any $x\in\Omega$, $\hat{u}_{\text{LSE}}(x)$ is the expectation of the marginal distribution

$$\pi(u(x))\propto\int_{\mathbb{R}^{x^c}}\pi(u)\,du(x^c),$$

where $u(x^c)$ is the restriction of u to x^c. This marginal distribution is difficult to compute, but this is not the case for the conditional marginal distribution

$$\pi(u(x)|u(x^c))$$
$$\propto\exp\left(-\frac{(u(x)-v(x))^2+\lambda\sum_{y\in\mathcal{N}_x}|u(x)-u(y)|}{2\sigma^2}\right). \tag{5}$$

The idea of the ICE method is to replace the posterior mean by an iterated conditional marginal mean, leading us to

Theorem 1 (and definition of TV-ICE) *Given an image $v:\Omega\to\mathbb{R}$ and two positive parameters λ and σ, the sequence of images $(u^n)_{n\geq0}$ defined recursively by u^0 and*

$$\forall x\in\Omega,\quad u^{n+1}(x)=\mathbb{E}_{u\sim\pi}\left(u(x)\,\middle|\,u(x^c)=u^n(x^c)\right) \tag{6}$$

converges linearly to an image $\hat{u}_{ICE}$ that is independent of the initialization u^0 and satisfies the fixed point property

$$\forall x\in\Omega,\quad \hat{u}_{ICE}(x)=\mathbb{E}_{u\sim\pi}\left(u(x)\,\middle|\,u(x^c)=\hat{u}_{ICE}(x^c)\right).$$

2.3. The recursion operator of TV-ICE

The noisy image v and the parameters $\lambda>0$ and $\sigma>0$ being set, we first define a function $F:\mathbb{R}^\Omega\to\mathbb{R}^\Omega$ by

$$\forall x\in\Omega,\quad F(w)(x)=\mathbb{E}_{u\sim\pi}\left(u(x)\,\middle|\,u(x^c)=w(x^c)\right).$$

To simplify the notations, we shall write, for any image $w:\Omega\to\mathbb{R}$ and any $x\in\Omega$,

$$\pi_x^w(s)=\pi(u(x)=s|u(x^c)=w(x^c))$$

the density (5) of the conditional marginal distribution. We then have $F(w)(x)=\int_{\mathbb{R}}\pi_x^w(s)\,s\,ds/\int_{\mathbb{R}}\pi_x^w(s)\,ds$ for all x, and thanks to (5) this quantity depends on w only through $w(\mathcal{N}_x)$ and on v only through $v(x)$. Hence, there exists a set of functions $f_t:\mathbb{R}^4\to\mathbb{R}$ such that

$$\forall w\in\mathbb{R}^\Omega,\ \forall x\in\Omega,\quad F(w)(x)=f_{v(x)}(w(\mathcal{N}_x)) \tag{7}$$

(these functions are computed explicitly later in Theorem 2). With these notations, the recursion (6) can be rewritten $u^{n+1}=F(u^n)$, or equivalently

$$\forall x\in\Omega,\quad u^{n+1}(x)=f_{v(x)}(u^n(\mathcal{N}_x)).$$

Note that F implicitly depends on v, λ and σ. We now prove three important properties of F that will ensure the convergence of the iterated operator.

Lemma 1 *F is monotone: for all images w_1 and w_2,*

$$w_1\leq w_2\Rightarrow F(w_1)\leq F(w_2),$$

where as usual, the inequalities between functions are meant pointwise (that is, for all points of Ω).

Proof The proof essentially relies on the fact that the function f_t defined earlier is $\mathcal{C}^1$ on $\mathbb{R}^4$ and satisfies, for any x and any $y \in \mathcal{N}_x$,

$$\frac{\partial f_{v(x)}}{\partial w(y)}(w(\mathcal{N}_x)) = \frac{\lambda}{2\sigma^2}\mathrm{Cov}_{X\sim\pi_x^w}(X, \mathrm{sign}(X - w(y))) \quad (8)$$

where Cov is the usual covariance operator. The differentiability of f_t is not completely obvious because of the absolute values that appear in (5). However, one can use the fact that the function $s \mapsto \exp(-|s|)$ is Lipschitz to prove the differentiability of f_t using Lebesgue's dominated convergence Theorem and show the formula (8) above.

Now, since $\partial f_{v(x)}/\partial w(y)$ can be written as the covariance of $X \sim \pi_x^w$ with a nondecreasing function of X, it is nonnegative [14]. Let $w_1 \leq w_2$, we have

$$\begin{aligned} F(w_2)(x) - F(w_1)(x) &= f_{v(x)}(w_2(\mathcal{N}_x)) - f_{v(x)}(w_1(\mathcal{N}_x)) \\ &= \textstyle\int_0^1 \nabla f_{v(x)}(w_t(\mathcal{N}_x)) \cdot (w_2(\mathcal{N}_x) - w_1(\mathcal{N}_x))\, dt \end{aligned}$$

with $w_t = (1-t)w_1 + tw_2$. Therefore, $F(w_2)(x) - F(w_1)(x)$ is, like the term inside the integral, nonnegative. □

Lemma 2 *F is strictly nonexpansive for the ℓ^∞ norm: for any two images w_1, w_2 such that $w_1 \neq w_2$, one has*

$$\|F(w_2) - F(w_1)\|_\infty < \|w_2 - w_1\|_\infty.$$

Proof Writing $c = \|w_2 - w_1\|_\infty > 0$, we have

$$\forall x \in \Omega, \quad w_1(x) - c \leq w_2(x) \leq w_1(x) + c,$$

and as F is monotone,

$$F(w_1 - c) \leq F(w_2) \leq F(w_1 + c). \quad (9)$$

Now let us show that $F(w_1 + c) < F(w_1) + c$. As

$$f_{v(x)}(w_1(\mathcal{N}_x) + c) = f_{v(x)-c}(w_1(\mathcal{N}_x)) + c$$

(this derives from the change of variable $s' = s - c$ in (7) using (5)), this amounts to show that $t \mapsto f_t(w_1(\mathcal{N}_x))$ is increasing. But from (7) we get

$$\frac{\partial f_t(w_1(\mathcal{N}_x))}{\partial t} = \frac{1}{\sigma^2}\mathrm{Var}_{X\sim\pi_x^{w_1}}(X)$$

which is positive as $\sigma > 0$. Symmetrically, we also have $F(w_1) - c < F(w_1 - c)$, so that (9) entails

$$F(w_1) - c < F(w_2) < F(w_1) + c,$$

that is, $\|F(w_2) - F(w_1)\|_\infty < c$ as required. □

Lemma 3 *Define, for any image $w : \Omega \to \mathbb{R}$, the set*

$$K_w = \left[\min(\min_\Omega v, \min_\Omega w), \max(\max_\Omega v, \max_\Omega w)\right]^\Omega. \quad (10)$$

Then for any w, $F(K_w) \subset K_w$.

Proof In this proof, we write F_v instead of F to recall that F depends on v. Since $F_{v-a}(w - a) = F_v(w) - a$ and $F_{-v}(-w) = -F_v(w)$, it is enough to prove that if $v \geq 0$ and $w \geq 0$, then $F_v(w) \geq 0$. But since

$$\pi_x^0(s) \propto \exp\left(-\frac{(s - v(x))^2 + 4\lambda|s|}{2\sigma^2}\right),$$

the assumption $v(x) \geq 0$ implies $\pi_x^0(s) \geq \pi_x^0(-s)$ for any $s \geq 0$, so that $F_v(0)(x) := \int_{\mathbb{R}} \pi_x^0(s)\, s\, ds / \int_{\mathbb{R}} \pi_x^0(s)\, ds \geq 0$ and thanks to Lemma 1 we get, since $w \geq 0$, the desired inequality $F_v(w) \geq F_v(0) \geq 0$. □

2.4. Proof of Theorem 1

The compact set K_{u^0} is stable by F (Lemma 3) and thus F is a contraction in K_{u^0} (Lemma 2), that is, there exists $\alpha \in (0, 1)$ such that $\|F(w_1) - F(w_2)\|_\infty \leq \alpha\|w_1 - w_2\|_\infty$ for all images w_1 and w_2. Thus, the Banach fixed-point theorem applies and the sequence (u^n) defined in Theorem 1 converges to a fixed point of F, which is unique thanks to Lemma 2. The convergence is linear as $\|u^{n+1} - \hat{u}_{\text{ICE}}\|_\infty \leq \alpha\|u^n - \hat{u}_{\text{ICE}}\|_\infty$, or in other terms, $\|u^n - \hat{u}_{\text{ICE}}\|_\infty = \mathcal{O}(\alpha^n)$ as $n \to \infty$.

2.5. A closed-form formula for the recursion operator

Theorem 2 *If $u(\mathcal{N}_x) = \{a, b, c, d\}$ with $a \leq b \leq c \leq d$, then the local iteration function f_t that defines $\hat{u}_{ICE}$ is*

$$f_t(u(\mathcal{N}_x)) = t + \lambda\frac{2X_{-2} + X_{-1} - X_1 - 2X_2}{X_{-2} + X_{-2} + X_0 + X_1 + X_2}, \quad (11)$$

where

$$\begin{cases} X_{-2} = \mathrm{erfc}(\frac{t-a+2\lambda}{\sigma\sqrt{2}})\exp\left(\frac{2\lambda(2(t+\lambda)-a-b)}{2\sigma^2}\right), \\ X_{-1} = \left(\mathrm{erf}(\frac{b-t-\lambda}{\sigma\sqrt{2}}) - \mathrm{erf}(\frac{a-t-\lambda}{\sigma\sqrt{2}})\right)\exp\left(\frac{\lambda(2(t-b)+\lambda)}{2\sigma^2}\right), \\ X_0 = \mathrm{erf}(\frac{c-t}{\sigma\sqrt{2}}) - \mathrm{erf}(\frac{b-t}{\sigma\sqrt{2}}), \\ X_1 = \left(\mathrm{erf}(\frac{d-t+\lambda}{\sigma\sqrt{2}}) - \mathrm{erf}(\frac{c-t+\lambda}{\sigma\sqrt{2}})\right)\exp\left(\frac{\lambda(2(c-t)+\lambda)}{2\sigma^2}\right), \\ X_2 = \mathrm{erfc}(\frac{d-t+2\lambda}{\sigma\sqrt{2}})\exp\left(\frac{2\lambda(c+d-2(t-\lambda))}{2\sigma^2}\right). \end{cases} \quad (12)$$

where erf *and* erfc *are the standard error and complementary error functions.*

Proof The computation of $f_t(\{a, b, c, d\})$ requires the integration of

$$g(s)e^{-\frac{(s-t)^2 + \lambda(|s-a| + |s-b| + |s-c| + |s-d|)}{2\sigma^2}}$$

over $s \in \mathbb{R}$, for $g(s) = 1$ (denominator) and $g(s) = s$ (numerator). This can be done explicitly by sorting a, b, c, d in increasing order, and then integrating on the intervals $(-\infty, a)$, (a, b), (b, c), (c, d) and (d, ∞) to get rid of the absolute values. The complete computation was checked using the formal calculus software Maple™. □

3. ALGORITHM AND EXPERIMENTS

The proposed TV-ICE denoising model being defined by the iteration of an explicit operator (given in Theorem 2), the algorithm is very simple. However, a direct implementation in a software development framework providing (in double precision) the mathematical functions erf and $\mathrm{erfc} = 1 - \mathrm{erf}$ will not work. In fact, (11) and (12) may involve numbers that have extremely different orders of magnitude, which causes numerical issues that must be properly handled.

We first explain how to accurately compute the values $(\log(X_k))_{-2\leq k\leq 2}$ (see (12)). These quantities involve the functions $t \mapsto \log(\mathrm{erfc}(t))$ and $(a,b) \mapsto \log(\mathrm{erf}(b) - \mathrm{erf}(a))$ (for $a < b$), which cannot be systematically computed using the direct composition of built-in functions. A simple example is given by $\log \mathrm{erfc}(1000) \simeq -10^6$, which is difficult to compute since e^{-10^6} is represented by 0 (underflow) in most programming languages. Once the values of $\log(X_k)$ are computed accurately thanks to Algorithms 1 and 2, it is not difficult to implement (11): we only have to subtract a convenient value to these quantities to avoid possible overflows when calling the `exp` function, as detailed in Algorithm 3.

Algorithm 1: `logerfc` **function**

uses: `log` and `erfc` (built-in functions)
input: a real number t
output: accurate[a] estimate of $\log(\mathrm{erfc}(t))$

if $t < 20$ **then return** $\mathtt{log}(\mathtt{erfc}(t))$
else
$$S \leftarrow 1 + \sum_{n=1}^{8} (-1)^n \frac{1}{2}\cdot\frac{3}{2}\cdot\frac{5}{2}\cdot\ldots\cdot\frac{2n-1}{2}\cdot t^{-2n}$$
return $-t^2 + \mathtt{log}\left(\frac{S}{t\sqrt{\pi}}\right)$

[a] The formula for S (evaluated with Horner method) comes from a 9-terms asymptotic expansion of erfc in $+\infty$.

Algorithm 2: `logerf2` **function**

uses: `log`, `erf`, `expm1` (built-in functions[a])
`logerfc` (function defined in Algorithm 1)
input: two real numbers a and b such that $a < b$
output: accurate estimate of $\log(\mathrm{erf}(b) - \mathrm{erf}(a))$

if $b < 0$ **then** $(a,b) \leftarrow (-b,-a)$
if $\frac{b-a}{|a|+|b|} < 10^{-14}$ **then return** $\mathtt{log}\left(\frac{2(b-a)}{\sqrt{\pi}}\right) - b^2$
else
  if $a < 1$ **then return** $\mathtt{log}(\mathtt{erf}(b) - \mathtt{erf}(a))$
  else
    $m \leftarrow \mathtt{logerfc}(b)$
    return $m + \mathtt{log}(\mathtt{expm1}(\mathtt{logerfc}(a) - m))$

[a] `expm1` is assumed to be a built-in implementation of $t \mapsto e^t - 1$ that is accurate even when $|t|$ is very small. It is a standard function in many programming languages (C, Java, Matlab™, etc.).

Algorithm 3: TV-ICE denoising algorithm

inputs: the noisy image v, the model parameters σ, λ, and the number of iterations N
initialization: $u \leftarrow v$
repeat N times
  for each $x \in \Omega$,
    collect the intensity values of the 4 neighbors of x and sort them into $a \leq b \leq c \leq d$
    compute $(\log X_k)_{-2\leq k\leq 2}$ using Equation (12) and Algorithms 1 and 2
    $M \leftarrow \max_{-2\leq k\leq 2} \log X_k$
    $X'_k \leftarrow \mathtt{exp}(\log X_k - M)$ for $k = -2,-1,0,1,2$
    $\tilde{u}(x) \leftarrow v(x) + \lambda \frac{2X'_{-2}+X'_{-1}-X'_1-2X'_2}{X'_{-2}+X'_{-2}+X'_0+X'_1+X'_2}$,
  $u \leftarrow \tilde{u}$
return u, the estimate of $\hat{u}_{\mathrm{ICE}}$

We tried Algorithm 3 on several images, and observed a striking similarity between the results produced by the TV-LSE and TV-ICE models (see Fig 1). In particular, TV-ICE reproduces a nice property of TV-LSE, that is, the preservation of edges (like ROF) without the undesirable staircasing artifact. The comparison of the convergence rates (see Fig. 2) is in favor of TV-ICE, even if an iteration of TV-ICE is computationally more expensive than an iteration of ROF (the ratio is about 13 for our non-optimized implementation of TV-ICE).

Thus, the TV-ICE model provides a very good approximation of the staircasing-free TV-LSE model for a computation time that is far below. This opens several perspectives, in particular the possible generalizations to more complex inverse problems or to color images for example. The generalization to a different neighborhood system, and in particular to three-dimensional images, is rather straightforward. From a theoretical viewpoint, it would be interesting to study the properties of the limit and to explain its similarity with the posterior mean solution.

REFERENCES

[1] L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D*, vol. 60, no. 1, pp. 259–268, 1992.

[2] A. Chambolle, V. Caselles, D. Cremers, M. Novaga, and T. Pock, "An introduction to total variation for image analysis," *Theoretical foundations and numerical methods for sparse recovery*, vol. 9, pp. 263–340, 2010.

[3] L. Ambrosio, N. Fusco, and D. Pallara, *Functions of bounded variation and free discontinuity problems*, vol. 254, Clarendon Press Oxford, 2000.

[4] A. Chambolle, "An algorithm for total variation minimization and applications," *J. Math. Imag. Vis.*, vol. 20, no. 1-2, pp. 89–97, 2004.

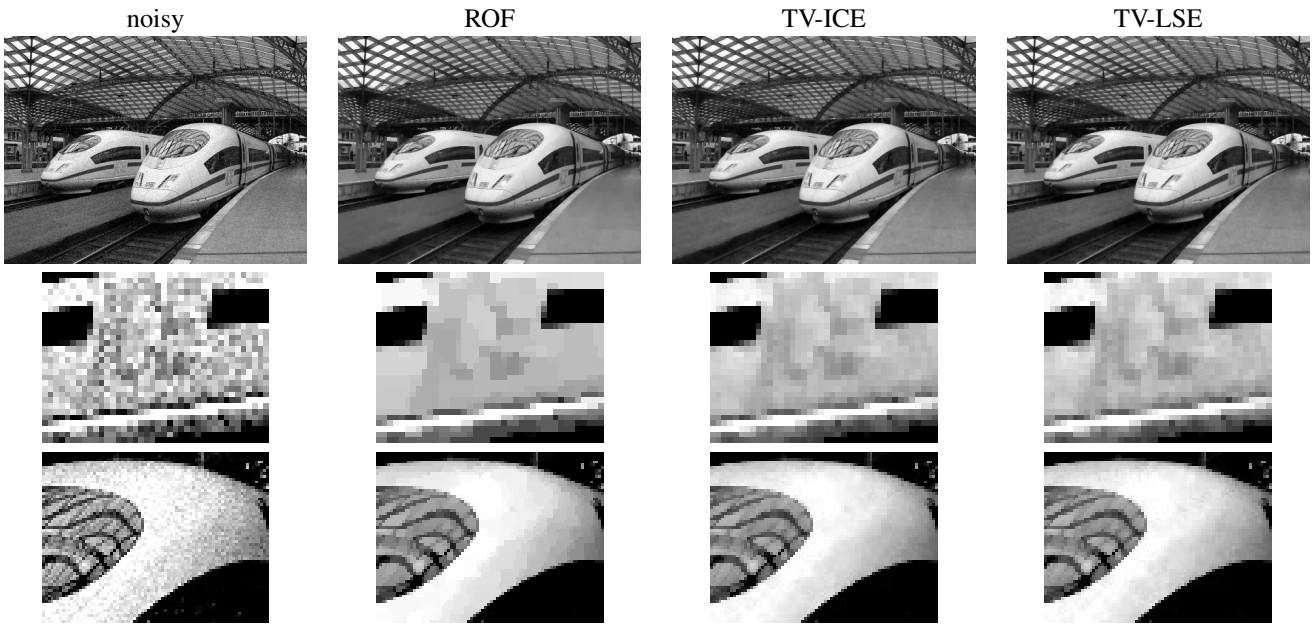

Fig. 1. **Comparison of ROF, TV-LSE and TV-ICE methods.** The top row shows the full image, whereas close-ups are displayed on the other rows. A clean 465×348 image (original author: Heinz Albers, source: wikipedia) was corrupted by a Gaussian additive white noise with standard deviation $\sigma = 10$, then denoised with the ROF, TV-LSE and TV-ICE algorithms. The parameters were chosen so that the norm of the estimated noise $\|\hat{u} - u\|$ was the same for each algorithm, leading to: $\lambda = 20$, $\sigma = 10$ for TV-LSE (initial choice), $\lambda = 18.6$, $\sigma = 10$ for TV-ICE, and $\lambda = 15.6$ for ROF. Note the similarity between LSE and ICE, and the staircasing effect clearly visible on the ROF close-ups.

[5] P. Weiss, L. Blanc-Féraud, and G. Aubert, "Efficient schemes for total variation minimization under constraints in image processing," *SIAM J. Sci. Comput.*, vol. 31, no. 3, pp. 2047–2080, 2009.

[6] D. Dobson and F. Santosa, "Recovery of blocky images from noisy and blurred data," *SIAM J. Appl. Math.*, vol. 56, no. 4, pp. 1181–1198, 1996.

[7] C. Bouman and K. Sauer, "A generalized Gaussian image model for edge-preserving MAP estimation," *IEEE Trans. Image Proc.*, vol. 2, no. 3, pp. 296–310, 1993.

[8] C. Louchet and L. Moisan, "Total variation denoising using posterior expectation," *Proc. Eur. Signal Process. Conf., Lausanne, Switzerland*, 2008.

[9] C. Louchet and L. Moisan, "Posterior expectation of the total variation model: Properties and experiments," *SIAM J. Imaging Sciences*, vol. 6, pp. 2640–2684, 2013.

[10] S. Geman, K. Manbeck, and D. McClure, "A comprehensive statistical model for single-photon emission tomography," *Markov Random Fields: Theory and Applications*, pp. 93–130, 1993.

[11] T. Ružić, A. Pižurica, and W. Philips, "Neighborhood-consensus message passing as a framework for generalized iterated conditional expectations," *Pattern Recognition Letters*, vol. 33, no. 3, pp. 309–318, 2012.

[12] H. Quick, S. Banerjee, and B. Carlin, "Modeling temporal gradients in regionally aggregated california asthma hospitalization data," *Ann. Appl. Stat.*, vol. 7, no. 1, pp. 154–176, 2013.

[13] M. Nikolova, "Weakly constrained minimization: application to the estimation of images and signals involving constant regions," *J. Math. Imag. Vis.*, vol. 21, no. 2, pp. 155–175, 2004.

[14] K. D. Schmidt, "On the covariance of monotone functions of a random variable," Unpublished note, University of Dresden, 2003.

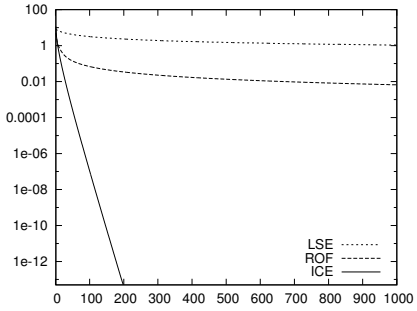

Fig. 2. Convergence rates (L^2 distance to solution, in log scale, with respect to the number of iterations) for the ROF (Nesterov dual algorithm [5]), TV-LSE (MCMC algorithm [8]), and TV-ICE (Algorithm 3). Contrary to ROF and TV-LSE methods, TV-ICE clearly exhibits a linear convergence rate.