Noname manuscript No.
(will be inserted by the editor)

Supplementary text to: "Coupling image restoration and segmentation: a generalized linear model/Bregman perspective"

Grégory Paul, Janick Cardinale, and Ivo F. Sbalzarini

Abstract This supplementary text provides details about optimization trajectories of the alternating minimization (AM) solver based on level-set (LS) and and on the alternating split Bregman (ASB) (section 4), collects the main results from Banerjee *et al.* (2005) about the duality between the natural and the mean parametrization in a one-dimensional regular exponential family (section 1), adapts the proofs from the generalized linear model literature of the photometric estimation results (section 2), and provides explicit formulae for the w_1 sub-problem of the ASB for Gaussian, Poisson and, Bernoulli noise models (section 3).

1 Duality in the Regular Exponential Family

We collect results about the duality relation in the Regular Exponential Family (REF) as stated by Banerjee *et al.* (2005) (*cf.* their Definition 4, Lemma 1 and, Theorem 2), but with our notation and for a one dimensional REF. We further gather required results about the duality relation induced by the Legendre-Fenchel transform of the cumulant-generating function b of a REF (Banerjee *et al.* 2005; Rockafellar 1997). We use the notation $\mathrm{int}(\cdot)$ for the interior of a set and $\mathrm{dom}(\psi)$ for the effective domain of the function ψ, namely the set of points in the domain of definition of ψ for which the function is finite.

Definition 1 Let b be a real-valued function on $\mathbb{R}$. Its conjugate function $b^\star$ is defined as:

$$b^\star(\mu) = \sup_{\theta \in \mathrm{dom}(b)} \left(\theta\mu - b(\theta)\right) .$$

We can now state the relevant duality results induced by the cumulant-generating function b between the natural parameter θ and the mean parameter μ of a one-dimensional REF.

MOSAIC Group
ETH Zurich
Universitätstr. 6, CH–8092 Zürich, Switzerland
E-mail: grpaul@ethz.ch, janickc@inf.ethz.ch, ivos@mpi-cbg.de

Lemma 1 *Let b be the cumulant-generating function of a REF with natural parameter space $\Theta = \mathrm{int}(\mathrm{dom}(b))$. Then b is a proper, closed, and convex function with $\mathrm{int}(\Theta) = \Theta$ and (Θ, b) is a convex function of Legendre type*[1] *(Rockafellar 1997). Its conjugate function $(\Theta^\star, b^\star)$ satisfies:*

1. *$(\Theta^\star, b^\star)$ is a convex function of Legendre type with $\Theta^\star = \mathrm{int}(\mathrm{dom}(b^\star))$.*
2. *(Θ, b) and $(\Theta^\star, b^\star)$ are Legendre duals of each other.*
3. *The gradient mapping $b' \colon \Theta \to \Theta^\star$ is a one-to-one mapping from the open convex set Θ onto the open convex set $\Theta^\star$.*
4. *The gradient mappings b' and $(b^\star)'$ are continuous and $(b^\star)' = (b')^{-1}$.*

Therefore, two points $(\theta, \mu) \in \Theta \times \Theta^\star$ in (Legendre) duality are uniquely related by the Legendre transformations induced by the diffeomorphism b':

$$b'(\theta) = \mu(\theta) \text{ and } (b^\star)'(\mu(\theta)) = \theta(\mu) \ .$$

The last result concerns the dual relationship between the Bregman divergences B_b and $\mathrm{B}_{b^\star}$ induced by b and its conjugate $b^\star$. For any two pairs of points $((p, q), (p^\star, q^\star)) \in \Theta^2 \times (\Theta^\star)^2$ in duality:

$$\mathrm{B}_b(p \parallel q) = \mathrm{B}_{b^\star}(q^\star \parallel p^\star) \ .$$

2 Proofs

We adapt classical proofs from the GLM literature (Nelder and Wedderburn 1972; McCullagh and Nelder 1989) to our image-processing problem. The results are not new and only provided here for the convenience of the reader. The only differences with the statistics literature are the continuous formulation of the results, requiring basics results about interchanging derivation and integration, and the sign convention. We recall that ℓ denotes the anti-log-likelihood function.

*Proof (Proof of **Result 2**)*
The proof is classical (McCullagh and Nelder 1989) and amounts to writing the appropriate chain rule in order to ease the substitutions of the mean function $\mu(\boldsymbol{x})$ and the variance function V. The only difference is the sign convention.

$$\begin{aligned}
\boldsymbol{s}(\boldsymbol{\beta}, \boldsymbol{x}) &= \frac{\partial \ell}{\partial \theta} \frac{\mathrm{d}\theta}{\mathrm{d}\mu} \frac{\mathrm{d}\mu}{\mathrm{d}\eta} \frac{\partial \eta}{\partial \boldsymbol{\beta}} && (1)\\
&= \frac{\partial \ell}{\partial \theta} \left(\frac{\mathrm{d}\mu}{\mathrm{d}\theta} \right)^{-1} \left(\frac{\mathrm{d}\eta}{\mathrm{d}\mu} \right)^{-1} \frac{\partial \eta}{\partial \boldsymbol{\beta}} && (2)\\
&= \frac{b'(\theta) - u_0(\boldsymbol{x})}{a(\boldsymbol{x}, \phi)} \frac{1}{b''(\theta)} \left(g'(\mu(\boldsymbol{x}, \boldsymbol{\beta})) \right)^{-1} \boldsymbol{X}(\boldsymbol{x}) && (3)\\
&= \frac{\mu(\boldsymbol{x}, \boldsymbol{\beta}) - u_0(\boldsymbol{x})}{\sigma^2(\boldsymbol{x}, \boldsymbol{\beta}) g'(\mu(\boldsymbol{x}, \boldsymbol{\beta}))^2} g'(\mu(\boldsymbol{x}, \boldsymbol{\beta})) \boldsymbol{X}(\boldsymbol{x}) \ . && (4)
\end{aligned}$$

[1] A convex function (Θ, f) of Legendre type has a non-empty domain, is strictly convex, differentiable, and the norm of its gradient is diverging to infinity for sequences in Θ converging to boundary points.

We emphasized here the chain rule (1) and the derivatives of interest (2). The following steps use the properties of the GLM: equations (3) and (4) use the definition of the link function and the properties of the mean and variance functions.

Concerning the whole-image score function $\boldsymbol{s}(\boldsymbol{\beta})$, the regularity condition needed is that $\ell(\boldsymbol{x}, \boldsymbol{\beta})$ is regular enough such that differentiation and integration can be interchanged:

$$\boldsymbol{s}(\boldsymbol{\beta}) = \frac{\mathrm{d}}{\mathrm{d}\boldsymbol{\beta}} \int_{\Omega_I} \ell(\boldsymbol{x}, \boldsymbol{\beta}) \, \mathrm{d}\boldsymbol{x} = \int_{\Omega_I} \frac{\partial}{\partial \boldsymbol{\beta}} \ell(\boldsymbol{x}, \boldsymbol{\beta}) \, \mathrm{d}\boldsymbol{x} \ .$$

Examples of such conditions are: $\ell(\boldsymbol{x}, \boldsymbol{\beta})$ is Ω_I-almost everywhere $\boldsymbol{\beta}$-differentiable and $|\boldsymbol{s}(\boldsymbol{x}, \boldsymbol{\beta})|$ is Ω_I-almost everywhere bounded by an integrable function of $\boldsymbol{x}$ only. □

*Proof (Proof of **Result 3**)*

We first factor out the deterministic term $\boldsymbol{X}(\boldsymbol{x})$ from the variance:

$$\mathbb{V}\big[W(\boldsymbol{x}, \boldsymbol{\beta})(\mu(\boldsymbol{x}, \boldsymbol{\beta}) - u_0(\boldsymbol{x}))g'(\mu(\boldsymbol{x}, \boldsymbol{\beta}))\boldsymbol{X}(\boldsymbol{x})\big] = \\ \boldsymbol{X}(\boldsymbol{x}) \ \mathbb{V}\big[W(\boldsymbol{x}, \boldsymbol{\beta})(\mu(\boldsymbol{x}, \boldsymbol{\beta}) - u_0(\boldsymbol{x}))g'(\mu(\boldsymbol{x}, \boldsymbol{\beta}))\big] \ \boldsymbol{X}^T(\boldsymbol{x}) \ .$$

The central term is straightforwardly evaluated as

$$\mathbb{V}\big[W(\boldsymbol{x}, \boldsymbol{\beta})(\mu(\boldsymbol{x}, \boldsymbol{\beta}) - u_0(\boldsymbol{x}))g'(\mu(\boldsymbol{x}, \boldsymbol{\beta}))\big] = \\ W(\boldsymbol{x}, \boldsymbol{\beta})^2 \sigma^2(\boldsymbol{x}, \boldsymbol{\beta}) g'(\mu(\boldsymbol{x}, \boldsymbol{\beta}))^2 = W(\boldsymbol{x}, \boldsymbol{\beta}) \ ,$$

where the last equation comes from the definition of $W(\boldsymbol{x}, \boldsymbol{\beta})$. This ends the proof for the Fisher-information matrix of a single pixel.

Concerning the whole-image Fisher information matrix, we use the property that the Fisher information matrix $\mathcal{I}(\boldsymbol{\beta})$ is also the expectation of the observed information matrix:

$$\mathcal{I}(\boldsymbol{\beta}) = \mathbb{E}\Big[\nabla\nabla^T \ell(\boldsymbol{\beta})\Big] \ ,$$

where $\nabla\nabla^T$ is the Hessian operator (matrix of second order partial derivatives). We thus need two conditions: one to interchange integration and differentiation (6), and one to interchange the order of integration (7):

$$\begin{aligned} \mathcal{I}(\boldsymbol{\beta}) &= \mathbb{E}\left[\nabla\nabla^T \int_{\Omega_I} \ell(\boldsymbol{x}, \boldsymbol{\beta}) \, \mathrm{d}\boldsymbol{x}\right] && (5) \\ &= \mathbb{E}\left[\int_{\Omega_I} \nabla\nabla^T \ell(\boldsymbol{x}, \boldsymbol{\beta}) \, \mathrm{d}\boldsymbol{x}\right] && (6) \\ &= \int_{\Omega_I} \mathbb{E}\Big[\nabla\nabla^T \ell(\boldsymbol{x}, \boldsymbol{\beta})\Big] \, \mathrm{d}\boldsymbol{x} && (7) \\ &= \int_{\Omega_I} \mathcal{I}(\boldsymbol{x}, \boldsymbol{\beta}) \, \mathrm{d}\boldsymbol{x} \ , && (8) \end{aligned}$$

where equations (5) and (8) come from the property of the Fisher information matrix applied to the whole image or to a single pixel. □

*Proof (Proof of the Fisher scoring iteration, **Result 4**)*

We multiply both sides of the Fisher scoring iteration (equation (14) in the article) by the Fisher information matrix in order to obtain:

$$\mathcal{I}(\boldsymbol{\beta}_r)\boldsymbol{\beta}_{r+1} = \mathcal{I}(\boldsymbol{\beta}_r)\boldsymbol{\beta}_r - \boldsymbol{s}(\boldsymbol{\beta}_r) \ . \tag{9}$$

The left-hand side of (9) is already the left-hand side of equation (15) of **Result 5** in the article. The right-hand side of (9) requires only few groupings of integrals. First, note that

$$\begin{aligned}
\mathcal{I}(\boldsymbol{\beta}_r)\boldsymbol{\beta}_r &= \int_{\Omega_I} W(\boldsymbol{x},\boldsymbol{\beta}_r)\boldsymbol{X}(\boldsymbol{x})\boldsymbol{X}^T(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}\ \boldsymbol{\beta}_r \\
&= \int_{\Omega_I} W(\boldsymbol{x},\boldsymbol{\beta}_r)\boldsymbol{X}(\boldsymbol{x})\cdot\boldsymbol{X}^T(\boldsymbol{x})\boldsymbol{\beta}_r\,\mathrm{d}\boldsymbol{x} \\
&= \int_{\Omega_I} W(\boldsymbol{x},\boldsymbol{\beta}_r)\boldsymbol{X}(\boldsymbol{x})\eta(\boldsymbol{x},\boldsymbol{\beta}_r)\,\mathrm{d}\boldsymbol{x} \\
&= \int_{\Omega_I} W(\boldsymbol{x},\boldsymbol{\beta}_r)\boldsymbol{X}(\boldsymbol{x})g(\boldsymbol{x},\boldsymbol{\beta}_r)\,\mathrm{d}\boldsymbol{x} \ .
\end{aligned}$$

Then,

$$\mathcal{I}(\boldsymbol{\beta}_r)\boldsymbol{\beta}_r - \boldsymbol{s}(\boldsymbol{\beta}_r) = \int_{\Omega_I} W(\boldsymbol{x},\boldsymbol{\beta}_r)\boldsymbol{X}(\boldsymbol{x})\Big\{g(\boldsymbol{x},\boldsymbol{\beta}_r) + (u_0(\boldsymbol{x}) - \mu(\boldsymbol{x},\boldsymbol{\beta}_r))\,g'(\boldsymbol{x},\boldsymbol{\beta}_r))\Big\}\,\mathrm{d}\boldsymbol{x}\ ,$$

where we recognized the definition of the adjusted dependent variable in the center of the integral. This ends the proof for the continuous IRWLS iteration.

In order to recover the classical discrete IRWLS from the statistics literature, we introduce the set of discrete pixel positions $\{\boldsymbol{x}_i\}_{i=1}^n$ and approximate the integral by a sum over the pixels:

$$\sum_{i=1}^{n} \boldsymbol{X}_i W_{i,r} \boldsymbol{X}_i^T\ \boldsymbol{\beta}_{r+1} = \sum_{i=1}^{n} \boldsymbol{X}_i W_{i,r} Z_{i,r}\ ,$$

with $\boldsymbol{X}_i \coloneqq \boldsymbol{X}(\boldsymbol{x}_i)$, $W_{i,r} \coloneqq W(\boldsymbol{x}_i,\boldsymbol{\beta}_r)$, and $Z_{i,r} \coloneqq Z(\boldsymbol{x}_i,\boldsymbol{\beta}_r)$. In matrix notation, this reads:

$$\begin{bmatrix}\boldsymbol{X}_1 \ \dots\ \boldsymbol{X}_n\end{bmatrix}\begin{bmatrix}W_{1,r} & & \\ & \ddots & \\ & & W_{n,r}\end{bmatrix}\begin{bmatrix}\boldsymbol{X}_1^T \\ \dots \\ \boldsymbol{X}_n^T\end{bmatrix}\ \boldsymbol{\beta}_{r+1} = \begin{bmatrix}\boldsymbol{X}_1 \ \dots\ \boldsymbol{X}_n\end{bmatrix}\begin{bmatrix}W_{1,r} & & \\ & \ddots & \\ & & W_{n,r}\end{bmatrix}\begin{bmatrix}Z_1 \\ \dots \\ Z_n\end{bmatrix}\ ,$$

which is the matrix equation stated in the result.

For a two-region segmentation problem we remark that

$$\boldsymbol{X}(\boldsymbol{x}) = \begin{bmatrix}\mathrm{K}\left[M_1\right](\boldsymbol{x}) \\ \mathrm{K}\left[M_2\right](\boldsymbol{x})\end{bmatrix}\ ,$$

from which it is easily seen that the IRWLS equation becomes:

$$\begin{bmatrix} \mathrm{K}_{11}^r & \mathrm{K}_{12}^r \\ \mathrm{K}_{12}^r & \mathrm{K}_{22}^r \end{bmatrix} \boldsymbol{\beta}^{r+1} = \begin{bmatrix} U_1^r \\ U_2^r \end{bmatrix} \ .$$

If the determinant of the square matrix $\mathrm{K}_{11}^r \, \mathrm{K}_{22}^r - (\mathrm{K}_{12}^r)^2$ on the left-hand side is non-zero, we can invert the matrix system. The Cauchy-Schwarz inequality for the quantities involved in the determinant is:

$$(\mathrm{K}_{12}^r)^2 \leq \mathrm{K}_{11}^r \, \mathrm{K}_{22}^r \ .$$

Hence, the determinant is non-zero if and only if the previous inequality is strict. The condition for equality in the Cauchy-Schwarz inequality is that there exist two positive scalars $a, b > 0$ such that for almost all $\boldsymbol{x} \in \Omega_I$,

$$a \, \mathrm{K}[M_1](\boldsymbol{x}) = b \, \mathrm{K}[M_2](\boldsymbol{x}) \ . \tag{10}$$

Suppose that two such numbers $a, b > 0$ exist and assume for simplicity that the operator K is normalized as $\mathrm{K}[1] = 1$. From the decomposition $M_1(\boldsymbol{x}) + M_2(\boldsymbol{x}) = 1$ and the linearity of K, this entails that

$$\mathrm{K}[M_1](\boldsymbol{x}) + \mathrm{K}[M_2](\boldsymbol{x}) = 1 \ .$$

Combining this equality with condition (10) implies that $\mathrm{K}[M_2](\boldsymbol{x})$ is constant $\boldsymbol{x}$-almost everywhere, hence:

$$\mathrm{K}[M_2](\boldsymbol{x}) = \frac{b}{a+b} \ ,$$

which is a contradiction. Therefore, the inequality in the Cauchy-Schwarz inequality is strict and the determinant is not zero.

The variance-covariance matrix of $\widehat{\boldsymbol{\beta}}_{\mathrm{MLE}}$ is the inverse Fisher information matrix evaluated at the MLE. This is a standard result in maximum-likelihood theory and is hence not reproduced here. □

The following proof shows that our general result for GLM data-fitting energies includes the classical result of Chan and Vese (2001) about photometric estimation. We also recover the result for a Chan-Vese model with deconvolution, as derived by Jung *et al.* (2009).

Proof (Proof of the identification of the photometric estimation for the Chan-Vese model.) Write the continuous IRWLS equation for the extended Chan-Vese model. In this case, the variance function is $\sigma^2(\boldsymbol{x}, \boldsymbol{\beta}) = 1$ and the link function g is the identity. Therefore, $W(\boldsymbol{x}, \boldsymbol{\beta}) = 1$ and the adjusted dependent variable $Z = u_0$. As a consequence, the continuous IRWLS becomes

$$\int_{\Omega_I} \boldsymbol{X}(\boldsymbol{x}) \boldsymbol{X}^T(\boldsymbol{x}) \, \mathrm{d}\boldsymbol{x} \ \boldsymbol{\beta}_{r+1} = \int_{\Omega_I} \boldsymbol{X}(\boldsymbol{x}) u_0(\boldsymbol{x}) \, \mathrm{d}\boldsymbol{x} \ , \tag{11}$$

which is linear in $\boldsymbol{\beta}$ and independent of r. Note that

$$\boldsymbol{X}\boldsymbol{X}^T = \begin{bmatrix} X_1(\boldsymbol{x})^2 & X_1(\boldsymbol{x}) X_2(\boldsymbol{x}) \\ X_1(\boldsymbol{x}) X_2(\boldsymbol{x}) & X_2(\boldsymbol{x})^2 \end{bmatrix} \ ,$$

with $X_i(\boldsymbol{x}) = (\mathrm{K} * M_i)(\boldsymbol{x})$. Equation (11) then becomes

$$\begin{bmatrix} \mathrm{K}_{11} & \mathrm{K}_{12} \\ \mathrm{K}_{12} & \mathrm{K}_{22} \end{bmatrix} \widehat{\boldsymbol{\beta}}_{\mathrm{MLE}} = \begin{bmatrix} U_1 \\ U_2 \end{bmatrix} \ . \tag{12}$$

In the proof of the general Fisher scoring iteration we have shown that the matrix left-multiplying the vector $\widehat{\boldsymbol{\beta}}_{\mathrm{MLE}}$ can be inverted. We obtain the MLE estimate $\widehat{\boldsymbol{\beta}}_{\mathrm{MLE}}$ by inverting equation (12), recovering the result of Jung *et al.* (2009).

For an identity kernel $\mathrm{K} = \mathrm{Id}$, we have for $i \in 1, 2$: $\mathrm{K} * M_i = M_i$. As a consequence,

$$\mathrm{K}_{ij} = \int_{\Omega_I} M_i(\boldsymbol{x}) M_j(\boldsymbol{x}) \, \mathrm{d}\boldsymbol{x} = \begin{cases} 0 & \text{if } i \neq j \\ |\Omega_1| & \text{if } i = j = 1 \\ |\Omega_2| & \text{if } i = j = 2 \ , \end{cases}$$

where $|\Omega|$ denotes the area of the domain Ω and for $i \in \{1, 2\}$:

$$U_i = \int_{\Omega_I} u_0(\boldsymbol{x}) M_1(\boldsymbol{x}) \, \mathrm{d}\boldsymbol{x} = \int_{\Omega_i} u_0(\boldsymbol{x}) \, \mathrm{d}\boldsymbol{x} \ .$$

The MLE estimate then is:

$$\frac{1}{\mathrm{K}_{11}\,\mathrm{K}_{22} - \mathrm{K}_{12}^2} \begin{bmatrix} \mathrm{K}_{22} & -\mathrm{K}_{12} \\ -\mathrm{K}_{12} & \mathrm{K}_{11} \end{bmatrix} \begin{bmatrix} U_1 \\ U_2 \end{bmatrix} = \frac{1}{|\Omega_1|\,|\Omega_2|} \begin{bmatrix} |\Omega_2| & 0 \\ 0 & |\Omega_1| \end{bmatrix} \begin{bmatrix} \int_{\Omega_1} u_0(\boldsymbol{x}) \, \mathrm{d}\boldsymbol{x} \\ \int_{\Omega_2} u_0(\boldsymbol{x}) \, \mathrm{d}\boldsymbol{x} \end{bmatrix} \ ,$$

which recovers the result of Chan and Vese (2001).

3 Solutions to the w_1-subproblem

We provide analytic solutions to the w_1-subproblem of our ASB algorithm for three convex noise models: Gaussian, Poisson, and Bernoulli, coupled using the identity link function. The Euler-Lagrange equation, re-rewritten is terms of $\mu = \boldsymbol{\Delta\beta}\, w_1 + \beta_1$, reads:

$$a(\mu - u_0) + (\mu - b)V(\mu) = 0 \ ,$$

with $a \coloneqq \gamma(\boldsymbol{\Delta\beta})^2 w_d$ and $b \coloneqq \beta_1 + \boldsymbol{\Delta\beta}\left(b_1^k + \mathrm{K}[M^{k+1}]\right)$. The variance function $V(\mu)$ is polynomial for all three cases: $V = 1$ for Gaussian noise, $V = \mu$ for Poisson noise, and $V = \mu(1 - \mu)$ for Bernoulli noise, leading to polynomial equations of degree 1, 2, and 3, respectively. Nevertheless, the range of admissible values for μ in the Poisson case ($\mu > 0$) and the Bernoulli case ($\mu \in \,]0\,, 1[$) allows us to determine a unique solution. We do not show the derivation, but provide the solutions in terms of μ. The solution of interest in w_1 is obtained by the conversion $w_1 = (\mu - \beta_1)/\boldsymbol{\Delta\beta}$.

Gaussian noise:

$$\mu = \frac{a u_0 + b}{a + 1} \ .$$

Poisson noise:

$$\mu = \frac{1}{2}\left(b - a + \sqrt{(b-a)^2 + 4 a u_0}\right) \ .$$

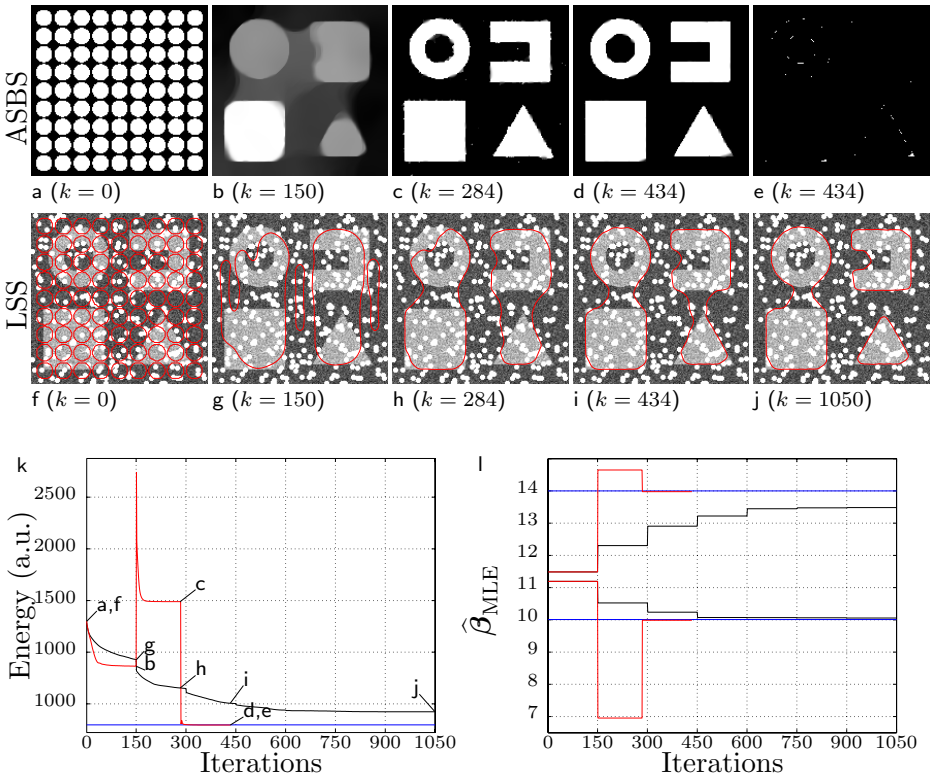

Fig. 1 Optimization trajectories for segmentation coupled with Gaussian denoising and TV-inpainting. Trajectories of the ASBS (a–e) and the LSS (f–j) AM corresponding to $(\mathsf{a}, 6)$ and $(\mathsf{a}, 5)$ in Fig. 1 of the main text are shown. (a,f) The initialization is the same for both algorithms, an array of circles represented as a labeling function for ASBS (a) or as the zero level-set (in red) for LSS (f). (b–d) Soft membership functions just after photometric re-estimation. In (c,d) we observe that the photometric re-estimation took place before the imposed re-estimation period of 150 iterations because the relative energy dropped below 10^{-6} (g–i) Zero level-set after each re-estimation of the ASBS. (e) Misclassified pixels in the final segmentation. (j) Final zero level-set at convergence. (k) Energy trace of the ASBS (red solid curve) and the LSS (black solid curve). Letter labels refer to the figures above. The horizontal blue line is the ground-truth energy. (l) Trace of $\widehat{\boldsymbol{\beta}}_{\mathrm{MLE}}$ for the ASBS (red solid curve) and the LSS (black solid curve). The horizontal blue lines are the ground-truth $\boldsymbol{\beta}$ values.

Bernoulli noise: For Bernoulli noise, the solution depends on the value of the data $u_0 \in \{0, 1\}$:

$$u_0 = 0 : \mu = \frac{1}{2}\left(1 + b - \sqrt{(1-b)^2 + 4a}\right)$$
$$u_0 = 1 : \mu = \frac{1}{2}\left(b + \sqrt{b^2 + 4a}\right) .$$

4 Optimization trajectories

We detail optimization trajectories for cases (a,5–6) and (a,11–12) of figure 1 in the main text. Each figure is organized in a similar manner. The first two rows illustrate

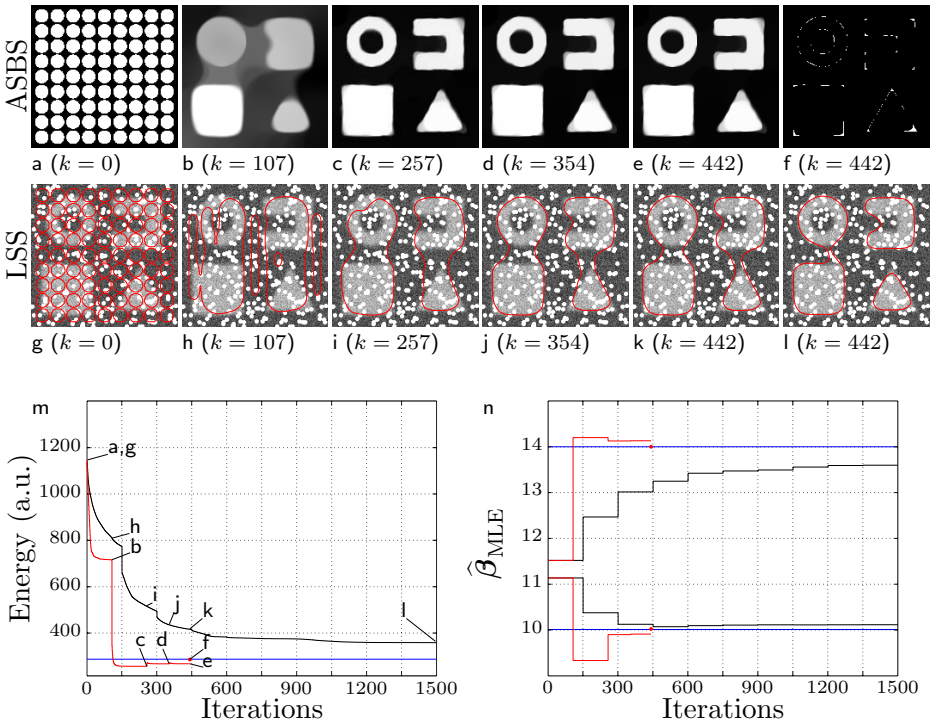

Fig. 2 Optimization trajectories for segmentation coupled with Gaussian denoising, deconvolution, and TV-inpainting. The legend is similar to the one of Fig. 1

the nature of the object used to encode the geometric optimization problem. The ASB solver (ASBS, first row) evolves a soft membership function from a binary mask (subfigure a) until convergence (Fig. 1d and Fig. 2e). The LS solver (LSS, second row) evolves an active contour (in red) from an initial contour (Fig. 1f and Fig. 2g) until convergence (Fig. 1j and Fig. 2l). The intermediate images show the state of the geometric solver at intermediate steps (Fig. 1b–c, and g–i and Fig. 2b–e and h–k) when the ASBS-based AM re-estimates the photometric constants. We observe that the ASBS reaches the relative convergence threshold of 10^{-6} prematurely (iteration numbers at photometric re-estimation are not all a multiple of 150). After the first photometric re-estimation the ASBS has already detected the four shapes (Fig. 1b and Fig. 2b), whereas the LSS has not (Fig. 1g and Fig. 2h). After an additional re-estimation, the ASBS is already very close in geometry to the ground truth (Fig. 1c and Fig. 2c), whereas the LSS has still not converged (Fig. 1h and Fig. 2i). Figures 1e and 2f show the misclassified pixels of the resulting segmentation after thresholding the final mask (which is shown in Fig. 1d and Fig. 2e, respectively), using the *a posteriori* upper bound.

The bottom-left plot shows the energy trajectories (in arbitrary units) for both the LSS (black line) and the ASBS (red line). The blue line indicates the ground truth energy. The letters in this plot correspond to the images above (see Fig. 1k and Fig. 2m). The bottom-right plot shows the evolution of the photometric parameters in the LSS (black solid line) and the ASBS (red solid line). The blue

lines correspond to the ground-truth values of the fore- and background, respectively. For both examples, we observe that the LSS converges to a sub-optimal state. We can easily identify reasons for such a sub-optimal result. For example, the cluster of missing data atop the upper-right corner of the square shape seems to prevent the active contour from moving further and hence from separating the square from the ring. The missing data inside the concavity of the U-shape seem to have a similar effect.

The ASBS converges closer to ground truth both in energy and in photometry. Without deconvolution, the solutions before and after thresholding are indistinguishable, as the soft mask is already almost binary (see Fig. 1d) and close to the ground-truth binary solution. This is because the convex relaxation is exact in this case. With deconvolution, the final mask $M^\star$ is clearly not binary (see Fig. 2e). The energy and the photometric constants of $M^\star$ at convergence hence differ from the ground-truth values (see Fig. 2m, label e and Fig. 2n). Nevertheless, after thresholding the final soft mask with the threshold $t^\star$ minimizing the *a posteriori* error upper-bound, the final result is very close to ground truth (see the red dots in Fig. 2m, label f and Fig. 2n). The convex relaxation is not exact when K is not the identity. In figure 2m, the difference in energy between points e and f illustrates the *a posteriori* upper bound on the difference in energy between the global solution of the original non-convex problem and the proposed solution.

References

Banerjee A, Merugu S, Dhillon I, Ghosh J (2005) Clustering with Bregman divergences. The Journal of Machine Learning Research 6:1705–1749

Chan TF, Vese L (2001) Active contours without edges. Image Processing, IEEE Transactions on 10(2):266 –277

Jung M, Chung G, Sundaramoorthi G, Vese L, Yuille A (2009) Sobolev gradients and joint variational image segmentation, denoising and deblurring. In: SPIE Electronic Imaging Conference Proceedings, Computational Imaging VII, SPIE, vol 7246

McCullagh P, Nelder J (1989) Generalized Linear Models, 2nd edn. Chapman and Hall/CRC

Nelder J, Wedderburn R (1972) Generalized linear models. Journal of the Royal Statistical Society Series A (General) 135(3):370–384

Rockafellar R (1997) Convex Analysis, vol 28. Princeton University Press