

OPTIMIZED SINGLE SITE UPDATE ALGORITHMS FOR IMAGE DEBLURRING

Stéphane Brette and Jérôme Idier

Laboratoire des Signaux et Systèmes (CNRS/ESE/UPS)
Plateau de Moulon, 91192 Gif-sur-Yvette Cedex, France
e-mail brette@lss.supelec.fr

1. ABSTRACT

In this paper we present optimized algorithms for image deblurring in the case of a separable Point Spread Function (PSF). Our work is in the usual context of Bayesian estimation with Gibbs Random Fields (GRF). The derived algorithms fall into the class of Single Site Update Algorithms (SSUAs), which exhibit a high convergence rate per iteration [1] and small memory requirements, while hard domain constraints such as positivity are easily introduced. On the other hand, standard forms of SSUAs rapidly become intractable when the size of the PSF is large. In the present study, we show how PSF separability can benefit SSUAs, in order to reduce the cost of each pixel update from $O(2pq)$ to $O(p+q)$ ($p \times q$ is the size of the PSF). We show that the resulting deterministic SSUA compares very favorably with Global Update Algorithms (GUAs). The new separable form can also benefit other SSUAs, especially stochastic versions such as Simulated Annealing (SA) and Monte Carlo Markov Chain (MCMC) algorithms.

2. MOTIVATION

In a Bayesian approach to deblurring, data \mathbf{y} are observed through an imaging system with a two dimensional (2D) impulse response H . With white Gaussian additive noise \mathbf{n} the degradation model is described by the linear relation $\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n}$ where \mathbf{x} is the column ordered vector corresponding to the $M \times N$ image to be restored and \mathbf{H} is the convolutional matrix (Toeplitz block Toeplitz) corresponding to the PSF H . Standard application of Bayes rule with a GRF as prior for \mathbf{x} yields the posterior probability density function (PDF) $\pi(\mathbf{x}|\mathbf{y})$ and the negloglikelihood function $J(\mathbf{x}) = -\log \pi(\mathbf{x}|\mathbf{y})$, up to an additional constant C :

$$J(\mathbf{x}) = \frac{\|\mathbf{y} - \mathbf{H}\mathbf{x}\|^2}{2\sigma^2} + \sum_{s,r \in C} \lambda_{rs} \phi(x_r, x_s) + C \quad (1)$$

Handling $J(\mathbf{x})$ is the corner stone of Bayesian estimators, including MAP, Posterior Mean and Marginal MAP. For instance, the MAP estimator $\hat{\mathbf{x}}_{map}$ minimizes $J(\mathbf{x})$. The problem of computing $\hat{\mathbf{x}}_{map}$ is formally simplified when the Gibbs potential ϕ is convex because $J(\mathbf{x})$ has no local minima, but it may represent a numerical challenge if a large number of pixels are processed.

When GUAs such as gradient or pseudo-conjugate gradient algorithms [2] are used for minimizing (1), each iteration requires the computation of the gradient $\frac{\partial J(\mathbf{x})}{\partial \mathbf{x}}$ and at least one time the computation of the criterion $J(\mathbf{x})$, depending on the line search minimization method. Fast convolution techniques (using FFT) and separable PSFs can considerably reduce the computational cost, but memory requirements of GUAs remain high (roughly speaking, from three to six times the image size) and domain constraints are difficult to introduce.

On the contrary, SSUAs are memory cheaper (two times the image size) and hard domain constraints are easily introduced, but no existing version takes advantage of separability, so SSUAs are presently restricted to PSFs with reduced support. This is also true for stochastic versions [3, 4] where SSUAs are widely used for sampling the PDF $\pi(\mathbf{x}|\mathbf{y})$ using Markov chain methods. In the next section, we derive an efficient "separable" SSUA for deblurring, which also extends to stochastic simulation.

3. SINGLE SITE UPDATE ALGORITHMS

According to a SSUA, image pixels are visited and modified in turn. In order to update the currently visited pixel x_{ij} , efficient computation of the terms of $J(\mathbf{x})$ involving x_{ij} must be considered. In the following, we successively address three cases where J is respectively quadratic (*ie.* \mathbf{x} is a Gauss-Markov field), convex or non-convex. We first focus on the quadratic case (§ 3.1). Both other cases are addressed by admix-

ture of auxiliary variables (§ 3.2 and § 3.3).

3.1. Gauss-Markov prior

When a Gauss-Markov model (i.e., a GRF with Gibbs potential $\phi(u) = u^2$) is used as prior, J is quadratic as a function of x_{ij} . Its minimum value is reached at m_{ij} :

$$m_{ij} = x_{ij} + \frac{[\mathbf{H}^t \mathbf{y}]_{ij} - [\mathbf{H}^t \mathbf{H} \mathbf{x}]_{ij} + 2\sigma^2 \sum \lambda_{r,ij}(x_r - x_{ij})}{[\mathbf{H}^t \mathbf{H}]_{ij,ij} + 2\sigma^2 \sum \lambda_{r,ij}}, \quad (2)$$

where the sums extend to the neighborhood of the currently visited pixel x_{ij} . Note that m_{ij} is also the mean of the conditional density $\pi(x_{ij} | \mathbf{x}_{\mathcal{S} \setminus ij}, \mathbf{y})$ whose variance σ_{ij}^2 is inversely proportional to the above denominator.

Since $[\mathbf{H}^t \mathbf{y}]_{ij}$ does not depend on the current \mathbf{x} we pre-calculate and store it. Only the Gibbs part (which often includes only few pixels) and $[\mathbf{H}^t \mathbf{H} \mathbf{x}]_{ij}$ have to be calculated. The latter requires $4pq$ multiplications for each visited pixel or $2pq$ thanks to central symmetry of $\mathbf{H}^t \mathbf{H}$, which is the 2D autocorrelation function of the PSF. In the following we show that separability $\mathbf{H} = \mathbf{v}^t \mathbf{h}$ yields an efficient way of computing $[\mathbf{H}^t \mathbf{H} \mathbf{x}]_{ij}$ from the autocorrelation vectors (\mathbf{g}, \mathbf{w}) of the horizontal and vertical components (\mathbf{h}, \mathbf{v}) respectively. Let us introduce an auxiliary vector \mathbf{R}^{ij} of length $2p - 1$ defined by:

$$R_k^{ij} = \sum_{l=-(q-1)}^{q-1} g_l x_{i+k,j+l} \quad (3)$$

for $k = -p + 1, \dots, p - 1$. It is easy to check that

$$[\mathbf{H}^t \mathbf{H} \mathbf{x}]_{ij} = \sum_{k=-(p-1)}^{p-1} w_k R_k^{ij}. \quad (4)$$

Since \mathbf{g} and \mathbf{w} are symmetric, (3) and (4) only involve q and p multiplications. Now, the keypoint of the derivation is the shift invariance $R_k^{i+1,j} = R_{k+1}^{ij}$ for each $k < p - 1$ (c5), which is kept up-to-date during the scan of a whole column (c4). From (3) hence, after each update of a pixel x_{ij} , one has only to keep R_0^{ij} up-to-date (c4) and to calculate R_{p-1}^{ij} (c6). This provides a very efficient way of updating the pixels by recursive scanning of the columns (a transposed form can also be considered), according to a Gauss-Seidel procedure. The following table presents a Successive Over-relaxation (SOR) version, which introduces a relaxation coefficient $0 < \omega < 2$ (c3) to allow faster convergence [2].

- | |
|---|
| <p>(a) Initialize : calculate and store $\mathbf{H}^t \mathbf{y}$, \mathbf{g} and \mathbf{w}.</p> <p>(b) For each column $j = 1, \dots, N$: initialize \mathbf{R}^{0j} (2b)</p> <p>(c) For each pixel $i = 1, \dots, M$ in the jth column :</p> <p>(c1) Apply the mask $[\mathbf{H}^t \mathbf{H} \mathbf{x}]_{ij} = \mathbf{w}^t \mathbf{R}^{ij}$ (2a)</p> <p>(c2) Add Gibbs and retrofiltered contributions to compute m_{ij}.</p> <p>(c3) Update x_{ij} subject to domain constraints $x_{ij}^{new} \leftarrow x_{ij} + \omega(m_{ij} - x_{ij})$</p> <p>(c4) Update $R_0^{ij} \leftarrow R_0^{ij} + w_0(x_{ij}^{new} - x_{ij})$ to take x_{ij}^{new} into account.</p> <p>(c5) $R_k^{i+1,j} = R_{k+1}^{ij}$ for $k = -p + 1, \dots, p - 2$.</p> <p>(c6) Compute $R_{p-1}^{i+1,j}$ according to (2b).</p> <p>Iterate step (c) until end of the current column.
Iterate step (b) until convergence.</p> |
|---|

3.2. Convex cases

Though the above form only applies to quadratic criteria, admixture of auxiliary variables \mathbf{l} [5, 6] provides a simple way to extend the same optimized structure to maximization (and sampling) of a much larger family of posterior densities.

A recently acknowledged role of auxiliary variables is to introduce an augmented criterion $J(\mathbf{x}, \mathbf{l})$ [5] arising from the following duality principle [7, chap. 7]:

Theorem 1 *Let $f(x)$ be concave on a convex set D . The concave conjugate $g(l)$ on the convex set D^* is defined as*

$$g(l) = \inf_{x \in D} (lx - f(x)) \quad (5)$$

where $D^* = \{l | g(l) > -\infty\}$ is a convex set. As a consequence, $f(x) = \inf_{l \in D^*} (lx - g(l))$.

In our case, let $\phi(u) = \inf_{l \in D^*} (lu^2 + \psi(l))$ an even function. It follows that $\phi(\sqrt{|u|})$ and $-\psi(l)$ are convex conjugate on $D^* \subset \mathbb{R}^+$ and $J(\mathbf{x}) = \inf_{l \in D^*} J(\mathbf{x}, \mathbf{l})$, with

$$J(\mathbf{x}, \mathbf{l}) = \frac{|\mathbf{y} - \mathbf{H} \mathbf{x}|^2}{2\sigma^2} + \lambda \sum_{\{r,s\} \in \mathcal{C}} l_{rs} (x_r - x_s)^2 + \psi(l_{rs}). \quad (6)$$

By construction $J(\mathbf{x})$ and $J(\mathbf{x}, \mathbf{l})$ share the same minimum. The latter can be obtained by alternate minimization of (6) with respect to \mathbf{x} and \mathbf{l} . As a function of \mathbf{x} , $J(\mathbf{x}, \mathbf{l})$ is an inhomogeneous quadratic function to which § 3.1 applies for $\lambda_{rs} \equiv \lambda l_{rs}$ in (1). As a function of the auxiliary variables \mathbf{l} , $J(\mathbf{x}, \mathbf{l})$ reads

$\sum J_{rs}(l_{rs}, x_r, x_s)$, which is independently minimized by

$$\hat{l}_{rs} = \frac{\phi'(x_r - x_s)}{2(x_r - x_s)} \quad (7)$$

according to a simple duality result [7, chap. 7].

The whole algorithm alternates

updates of \mathbf{x} and \mathbf{l} , the latter being a local and inexpensive operation. Moreover, the auxiliary variables require no additional memory if the updates are carefully intertwined. When first order cliques are used, it is never necessary to store more than one column of auxiliary variables at the same time, and not more than three columns for second order cliques.

3.3. Stochastic version for SA and other MCMC

When the Gibbs potential ϕ is not convex, SA [3, 8, 5, 4] still provides a global optimization algorithm. The idea is to build a stationary Markov Chain that converges to $\pi^{1/T}$ and to slowly decrease the temperature T . In the general case of continuous random variables, samples of π are obtained by a Metropolis sampler [9], involving a proposal PDF. The latter must be carefully chosen and tuned, otherwise convergence can be very slow.

In the same context, a major advantage of introducing auxiliary variables is to allow Gibbs sampling [3] from $\pi(\mathbf{x}, \mathbf{l}|\mathbf{y}) \propto \exp -J(\mathbf{x}, \mathbf{l})$ rather than Metropolis sampling from $\pi(\mathbf{x}|\mathbf{y})$. If \mathbf{X} and \mathbf{L} are alternatively sampled, a stationary Markov chain is generated that converges to $\pi(\mathbf{x}, \mathbf{l}|\mathbf{y})$. On one hand $(\mathbf{X}|\mathbf{L}, \mathbf{Y})$ is Gauss-Markov, and over-relaxation can still be used to sample \mathbf{X} by replacing step (c3) by :

$$x_{ij}^{new} \leftarrow x_{ij} + \omega(m_{ij} - x_{ij}) + \sigma_{ij} \sqrt{\omega(2-\omega)} n \quad (8)$$

with $n \equiv \mathcal{N}(0, 1)$. Such an over-relaxed Gaussian sampler allows faster convergence and boils down to the usual Gibbs sampler [3, 8] when $\omega = 1$ [10]. On the other hand $(\mathbf{L}|\mathbf{X})$ are independent random variables sampled from

$$\pi(l_{rs}|x_r, x_s) \propto \exp -\lambda (l_{rs}(x_r - x_s)^2 + \psi(l_{rs})) .$$

4. SIMULATED EXAMPLES

Efficiency of the presented deterministic algorithm is demonstrated in a deblurring problem. The simulated image is a chessboard of 128×128 pixels (Figure 1(a)), blurred by a 19×19 Gaussian blur.

To a certain extent, prescribing a Gibbs potential is still an open issue. Many *edge preserving* convex functions already provide satisfactory results for image processing [11, 6, 1]. Here we propose a new convex

function $\phi(u) = |u/\delta| - \log(1 + |u/\delta|)$. It is not designed to perform better than others, since it has the usual L1-L2 shape, but it is numerically attractive since its derivative is simply $\phi'(u) = u/\delta(|u| + \delta)$. Hence, updating the auxiliary variables using (7) is an inexpensive operation involving no transcendental function. This feature would be negligible if the heaviest term to compute was still the backprojection $\mathbf{H}^t \mathbf{H} \mathbf{x}$. Actually when separability is taken into account, it is not so any more and we have practically found that transcendental evaluations would represent a rather important part of the computation load, especially when second (or greater) order of cliques are used.

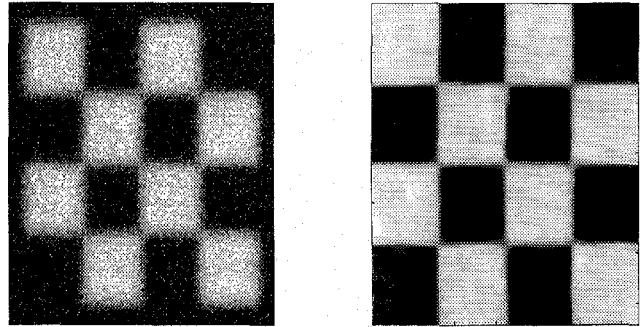


Figure 1: (a) Observed and (b) deblurred images

The convergence rate of our deterministic algorithm (with $\omega = 1.7$) has been compared to the well-known *Broyden-Fletcher-Golfrab-Shanno* Pseudo-conjugate gradient [2] with mixed quadratic-cubic line minimization algorithm for the chessboard problem. Signal to noise ratio was 20 dB and hyper-parameters δ, λ were empirically chosen to obtain visually good results (Figure 2(b)). Both algorithms were initialized with a black image.

The following table reports Mflops, final criteria and final gradient norms. Each PCG iteration (which also benefits from separability) costs about four times a SSUA iteration. For similar values of final criteria, SSUA costs six times fewer Mflops than PCG, and for similar numbers of Mflops, the norm of the gradient is drastically reduced by the SSUA. On the other hand, our algorithm is roughly 13 times faster than the equivalent SSUA without separability.

	$J(\hat{\mathbf{x}})$	$ \text{grad}J(\hat{\mathbf{x}}) _2$	Mflops
\mathbf{x}_0	3.47910^5	6129	
PCG	11812	3.6907	438 (50 it)
SSUA	11811	1.8219	77 (34 it)
	11809	0.0149	437 (195 it)

Figure 2(a) is a view of Saturn (200×460) from

the 2.2m Haway telescope taken during the first ring plane crossing in May 1995. Blur results from a 8s pause time and a 29×37 separable PSF was estimated from the small satellite Thetys (far left of saturn's ring) considered as a point source. Our prior was used together with positivity constraint and parameters were chosen to obtain good visual reconstruction. Saturn's satellites Thetys and Enceladus are clearly visible on the restored image (Figure 2(b)), as well as a white spot on the north of equator which appeared to be a storm. The proposed algorithm has spent 110 sec. for 200 iterations on a HP/J210 workstation and is more than 30 times faster than the equivalent SSUA without separability.

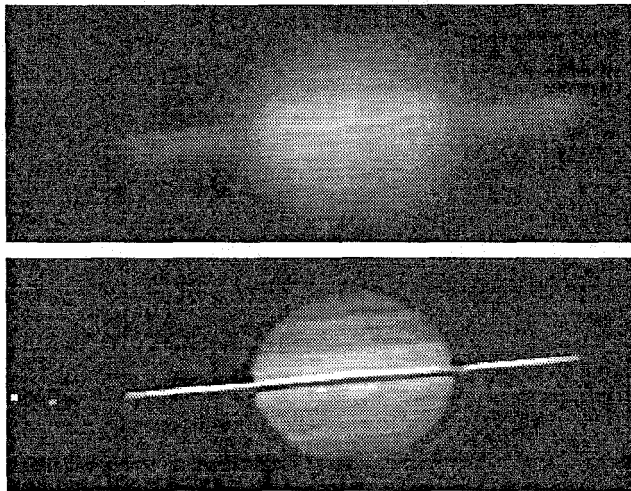


Figure 2: (a) Observed and (b) restored images of Saturn

5. CONCLUSIONS

In this paper we have presented a new algorithmic structure available for deblurring images when the PSF is separable. We have shown that the deterministic version is very efficient to minimize a convex criterion compared to usual and global update algorithms. Moreover, SSUA are memory efficient, can manage hard domain constraint and are easily parallelisable. Finally, stochastic simulation can also benefit from the new separable form. One possible extension would be to generalize the algorithmic structure to PSF of rank greater than one. This can be achieved by recursively updating several generator vectors rather than only \mathbf{R} .

6. REFERENCES

[1] K. Sauer and C. Bouman, "A local update strategy for iterative reconstruction from projections",

IEEE Transactions on Signal Processing, vol. SP-41, no. 2, pp. 534-548, February 1993.

- [2] W. Press, S. Teukolsky, W. Vetterling, and B. Flannery, *Numerical recipes in C, the art of scientific computing, second edition*, Cambridge Univ. Press, New York, 1992.
- [3] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-6, no. 6, pp. 721-741, November 1984.
- [4] C. Yang, "Efficient stochastic algorithms on locally bounded image space", *CVGIP: Graphical Models and Image Processing*, vol. 55, no. 6, pp. 494-506, November 1993.
- [5] S. Geman and G. Reynolds, "Constrained restoration and recovery of discontinuities", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-14, no. 3, pp. 367-383, March 1992.
- [6] P. Charbonnier, L. Blanc-Féraud, G. Aubert, and M. Barlaud, "Two deterministic half-quadratic regularization algorithms for computed imaging", in *Proceedings of IEEE ICIP*, 1994, vol. 2, pp. 168-172.
- [7] D. Luenberger, *Optimization by Vector Space Methods*, Wiley, J., New York, 1 edition, 1969.
- [8] F. Jeng and J. Woods, "Simulated annealing in compound Gaussian random fields", *IEEE Transactions on Information Theory*, vol. IT-36, no. 1, pp. 94-107, January 1990.
- [9] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, "Equations of state calculations by fast computing machines", *Journal of chemical physics*, vol. 21, pp. 1087-1092, June 1953.
- [10] S. Adler, "Over-relaxation method for Monte-Carlo evaluation of the partition function for multi-quadratic actions", *Physical Review D*, vol. 23, pp. 2901-2904, June 1981.
- [11] P. J. Green, "Bayesian reconstructions from emission tomography data using a modified EM algorithm", *IEEE Transactions on Medical Imaging*, vol. MI-9, no. 1, pp. 84-93, March 1990.