

Bayesian Tomographic Reconstruction Using Riemannian MCMC

Stefano Pedemonte^{1,2}, Ciprian Catana¹, Koen Van Leemput^{1,2,3}

¹ Athinoula A. Martinos Center for Biomedical Imaging, MGH/Harvard, MA, USA

² Department of Computer Science and Department of Neuroscience and Biomedical Engineering, Aalto University, FI

³ Department of Applied Mathematics and Computer Science, Technical University of Denmark, DK

Abstract. This paper describes the use of Monte Carlo sampling for tomographic image reconstruction. We describe an efficient sampling strategy, based on the Riemannian Manifold Markov Chain Monte Carlo algorithm, that exploits the peculiar structure of tomographic data, enabling efficient sampling of the high-dimensional probability densities that arise in tomographic imaging. Experiments with positron emission tomography (PET) show that the method enables the quantification of the uncertainty associated with tomographic acquisitions and allows the use of arbitrary risk functions in the reconstruction process.

1 Introduction

A tomographic imaging device produces an indirect measurement of a spatially-dependent quantity. In the case of positron emission tomography, the interaction of photons with the array of detection crystals provides information about the rate of nuclear decay. In the early days, the tomographic reconstruction problem was addressed with a mathematical formulation arising in the context of integral geometry: the Radon transform. Such idealized formulation provides an exact inversion formula under the assumptions of noiseless measurements and infinitesimally small detection elements. Starting with Shepp and Vardi [1], the formulation of the reconstruction problem in the probabilistic framework has enabled the development of algorithms that account for the uncertainty associated to the measurements and employ accurate models of the characteristics of the imaging devices. While in the context of tomographic imaging the research community has focused on the development of algorithms to compute a *best guess* of the unknown parameters, under the *maximum likelihood* (ML) and *maximum a posteriori* (MAP) point estimation criteria, in the context of low-level vision, starting with Geman and Geman [2], there has been a growing interest in the fully Bayesian approach to probabilistic reasoning, based on Markov Chain Monte Carlo (MCMC) sampling. In contrast to point-estimation criteria, which aim to maximize a density function, the fully Bayesian approach aims to characterize the entire posterior probability density induced by the imaging experiment. This enables the quantification of the uncertainty, as reported

in recent applications in low-level vision, including image restoration [3], super-resolution [4], optical-flow [5], *etc.* and in numerous applications in the field of spatial statistics.

The computational complexity of the fully Bayesian approach via sampling techniques has hindered its application in high-dimensional problems that present strong non-local interdependence of the parameters, such as occurs in tomographic imaging due to the integral measurements along lines-of-response (LORs). Recently, Girolami and Calderhead [7] have presented a new class of MCMC algorithms, based on concepts of differential geometry, that exploit the natural structure of probability density functions, achieving significantly improved computational efficiency in high-dimensional problems with strongly correlated parameters. Such formulation is an extension of Hamiltonian Monte Carlo (HMC), a sampling technique based on data-augmentation that was first introduced by Duane *et al.* [8].

In this paper we describe a sampling algorithm for tomographic imaging based on the formulation of Girolami and Calderhead [7]. We describe a strategy that exploits the natural geometry of the probability density function and the spatial structure of tomographic data to enable sampling from the high-dimensional probability densities that arise in tomographic imaging. In the experiments we highlight how the algorithm enables the quantification of uncertainty and enables the use of risk functions in the reconstruction other than the one that leads to MAP estimation.

2 Efficient sampling of tomographic imaging data

This section first introduces the imaging model for PET and gives an overview of state-of-the-art Hamiltonian Monte Carlo; then it introduces the details of the computational strategy that enables efficient sampling of tomographic data.

2.1 Imaging model and overall sampling procedure

PET imaging operates in the photon-limited regime. Discretizing the problem on an image lattice, with discrete locations indexed by v , and letting d index the LORs; the probability density associated to the number of counts $\mathbf{y} = [y_1, \dots, y_d, \dots, y_D]^T$, conditional to the rates of emission $\mathbf{x} = [x_1, \dots, x_v, \dots, x_N]^T$, is a product of Poisson functions (see e.g. [1]):

$$p(\mathbf{y}|\mathbf{x}) = \prod_{d=1}^D \frac{e^{-[\mathbf{A}\mathbf{x}]_d} [\mathbf{A}\mathbf{x}]_d^{y_d}}{y_d!}, \quad (1)$$

where each element of the system matrix $\mathbf{A} = \{a_{dv}\}$ expresses the probability that a pair of photons emitted from location v is detected in LOR d . Given a prior probability density $p(\mathbf{x})$, MAP estimation consists in finding the value of \mathbf{x} that maximizes the posterior probability density $p(\mathbf{x}|\mathbf{y}) \propto p(\mathbf{y}|\mathbf{x})p(\mathbf{x})$. Numerous algorithms have been reported for the maximization of the posterior, the

most notables being the MLEM algorithm for non-informative uniform priors, first described by Shepp and Vardi [1], and the one-step-late MAP-EM algorithm for arbitrary log-differentiable priors, introduced by Green [6]. Rather than computing the maximizer, the aim of the algorithm described in the following is to characterize the full posterior density by MCMC sampling.

The samples are generated using a Markov Chain with equilibrium $p(\mathbf{x}|\mathbf{y})$, obtained with the Metropolis-Hastings (MH) method. A new candidate is generated according to a proposal distribution $q(\mathbf{x}^*|\mathbf{x}^s)$ and accepted/rejected according to the MH criterion. Letting s index the samples:

$$\mathbf{x}^* \sim q(\mathbf{x}^*|\mathbf{x}^s) \quad (2)$$

$$u \sim \text{unif}(0, 1) \quad (3)$$

$$\alpha = \min \left[1, \frac{p(\mathbf{x}^*|\mathbf{y})q(\mathbf{x}^s|\mathbf{x}^*)}{p(\mathbf{x}^s|\mathbf{y})q(\mathbf{x}^*|\mathbf{x}^s)} \right] \quad (4)$$

$$\mathbf{x}^{s+1} := \begin{cases} \mathbf{x}^* & \text{if } u < \alpha \\ \mathbf{x}^s & \text{if } u \geq \alpha \end{cases} \quad (5)$$

Under the MH criterion, if the proposal $q(\mathbf{x}^*|\mathbf{x}^s)$ is *reversible* (see e.g. [9]), the sequence is guaranteed to be distributed, at equilibrium, according to the probability density $p(\mathbf{x}|\mathbf{y})$. A usual choice is to let the proposal $q(\mathbf{x}^*|\mathbf{x}^s)$ be a Gaussian distribution centered in \mathbf{x}^s , so that points closer to \mathbf{x}^s are more likely to be visited next. Such choice makes the sequence of samples a random-walk, producing high acceptance rate, but slow exploration of the space of the solutions. In order to scale to high-dimensions, it is necessary to make large steps in the proposal of the MH algorithm, while maintaining the probability of acceptance α high. Duane *et al.* [8] introduced a proposal based on data-augmentation that, unlike random walk, uses the derivatives of the posterior density to allow for large steps with high probability of acceptance. The technique consists in introducing a variable \mathbf{z} of the same dimensionality of \mathbf{x} , independent from \mathbf{x} and normally distributed with covariance \mathbf{M} (denominated *mass-matrix*): $\log p(\mathbf{x}, \mathbf{z}|\mathbf{y}) = \log p(\mathbf{x}|\mathbf{y}) - \frac{1}{2} \log\{(2\pi)^N |\mathbf{M}|\} - \frac{1}{2} \mathbf{z}^T \mathbf{M}^{-1} \mathbf{z}$; the new candidate is computed by drawing a random sample of \mathbf{z} and moving along the iso-curve of the joint probability density of \mathbf{x} and \mathbf{z} , described by the Hamiltonian equations with fictitious time t :

$$\mathbf{z}^* \sim \mathcal{N}(\mathbf{z}; 0, \mathbf{M}) \quad (6)$$

$$\begin{cases} \frac{d\mathbf{x}_i}{dt} = -\frac{\partial \log p(\mathbf{x}, \mathbf{z}|\mathbf{y})}{\partial \mathbf{z}_i} = [\mathbf{M}^{-1} \mathbf{z}]_i \\ \frac{d\mathbf{z}_i}{dt} = \frac{\partial \log p(\mathbf{x}, \mathbf{z}|\mathbf{y})}{\partial \mathbf{x}_i} = [\nabla_{\mathbf{x}} \log p(\mathbf{x}|\mathbf{y})]_i \end{cases} \quad (7)$$

This procedure generates samples from $p(\mathbf{x}, \mathbf{z}|\mathbf{y})$. The samples from the marginal density $p(\mathbf{x}|\mathbf{y})$ are obtained simply by discarding the samples of \mathbf{z} . To discretize the Hamiltonian dynamics, we employ the Stormer-Verlet leapfrog integrator, employed by Duane and others. With N_K integration steps ($k = 1, \dots, N_K$):

$$\begin{cases} \mathbf{x}^{*k+1} = \mathbf{x}^{*k} + \epsilon \mathbf{M}^{-1} \left[\mathbf{z}^{*k} + \frac{\epsilon}{2} \nabla_{\mathbf{x}} \log p(\mathbf{x}^{*k}|\mathbf{y}) \right] \\ \mathbf{z}^{*k+1} = \mathbf{z}^{*k} + \frac{\epsilon}{2} \nabla_{\mathbf{x}} \log p(\mathbf{x}^{*k}|\mathbf{y}) + \frac{\epsilon}{2} \nabla_{\mathbf{x}} \log p(\mathbf{x}^{*k+1}|\mathbf{y}) \end{cases} \quad (8)$$

The Stormer-Verlet integrator is perfectly time-reversible (see Neal [9] and Girolami [7]), therefore guaranteeing reversibility of the overall proposal function for any positive definite mass matrix \mathbf{M} . The choice of the mass-matrix and of the integration step, addressed next, is critical for the performance of the sampler, especially when the problem exhibits strong correlations [7].

2.2 Efficient sampling of tomographic data

While HMC is well established across multiple domains of computational science, its application to tomographic imaging is challenging. First, the rate of emission \mathbf{x} is non-negative. We propose to impose non-negativity by simulating an elastic boundary [9]:

$$\left. \begin{aligned} x_i^{*k+1} &:= -x_i^{*k+1} \\ z_i^{*k+1} &:= -z_i^{*k+1} \end{aligned} \right\} \text{if } x_i^{*k+1} < 0 \quad (9)$$

Second, the variables are strongly correlated, leading to high rates of rejection for naive implementation of the HMC algorithm. As pointed out in the recent formulation of Girolami and Calderhead [7], the Fisher Information Matrix (FIM) of the likelihood term, defined, for the PET imaging model, as:

$$\mathbf{H}_{\mathbf{x}} : h_{ij} = -\frac{\partial^2}{\partial x_i \partial x_j} \log p(\mathbf{x}|\mathbf{y}) = \sum_{d=1}^D \frac{a_{di} a_{dj} y_d}{\left(\sum_{v=1}^N a_{dv} \mathbf{x}_v \right)^2}, \quad (10)$$

constitutes a piece-wise constant approximation of the metric tensor of the Riemann structure of the parameter space. The choice $\mathbf{M} = \mathbf{H}_{\mathbf{x}}^{-1}$ yields effective transitions that respect and exploit the geometry of the manifold, with the effect of increasing the performance of the sampler and making the choice of the integration step ϵ non-critical.

In tomographic imaging, the FIM is too large to be evaluated numerically and stored. However the spatial organization of the parameters allows for approximations. Assuming periodic boundary conditions, the FIM (10) is well approximated by a block-circulant matrix with circulant blocks (BCCB), an approximation commonly utilized, in various ways, in imaging applications [10]. By the Fourier diagonalisation properties of BCCB matrices:

$$\mathbf{H}_{\mathbf{x}}^{\text{BCCB}} = \mathbf{F}^\dagger \mathbf{A} \mathbf{F} \quad (11)$$

$$\mathbf{A} = \text{diag}(\sqrt{N} \mathbf{F} \mathbf{h}_1), \quad (12)$$

where \mathbf{F} is the Discrete Fourier Transform matrix, \mathbf{F}^\dagger its complex conjugate, and \mathbf{h}_1 the first column of the FIM. Following (12), the BCCB approximation reduces the matrix-vector multiplication in (8) to a convolution. Furthermore, sampling the hidden variable \mathbf{z}^* becomes trivial using the method of the affine transformation:

$$\mathbf{v} \sim \mathcal{N}(0, I) \quad (13)$$

$$\mathbf{z} = \mathbf{F}^\dagger \mathbf{A}^{-\frac{1}{2}} \mathbf{v} \quad (14)$$

The sample of \mathbf{z}^* is obtained by generating a vector of independent normal variates and computing an inverse Fourier transformation (14).

A further approximation is introduced by computing the BCCB FIM just once, for \mathbf{x} equal to the MAP estimate (the ML estimate in case of uninformative prior).

2.3 Summary of the algorithm

In summary, the algorithm consists in first constructing an estimate of the BCCB FIM, obtained by performing an initial reconstruction with the MAP-EM algorithm [6] and by computing the first column of the Hessian matrix and its Fourier transform; then, at each iteration, the new candidate is obtained by sampling the hidden variable \mathbf{z}^{*0} and integrating the Hamiltonian dynamics with elastic boundary; the resulting candidate is accepted or rejected according to the MH criterion for the joint density. Samples of \mathbf{z} are discarded:

$$\text{Fisher Information} \left[\begin{array}{l} \mathbf{x}^{MAP} = \arg \min_{\mathbf{x}} -\log p(\mathbf{x}|\mathbf{y}) \\ \mathbf{h}_1^{\text{BCCB}} : h_{1j}^{\text{BCCB}} = \sum_{d=1}^D \frac{a_{d1} a_{dj} y_d}{(\sum_{i=1}^N a_{di} x_i^{MAP})^2} \\ \mathbf{A} = \text{diag} \left(\sqrt{N} \mathbf{F} \mathbf{h}_1^{\text{BCCB}} \right) \end{array} \right. \quad (15)$$

$$\text{New candidate} \left[\begin{array}{l} \text{for } s = 1 : N_S \\ \mathbf{v} \sim \mathcal{N}(0, I) \\ \mathbf{z}^{*0} = \mathbf{F}^\dagger \mathbf{A}^{-\frac{1}{2}}(\mathbf{v}) \\ \mathbf{x}^{*0} = \mathbf{x}^{s-1} \end{array} \right. \quad (16)$$

$$\text{Hamiltonian dynamics} \left[\begin{array}{l} \text{for } k = 0 : N_K \\ \mathbf{w} = \mathbf{z}^{*k} + \frac{\epsilon}{2} \nabla_{\mathbf{x}} \log p(\mathbf{x}^{*k}|\mathbf{y}) \\ \mathbf{x}^{*k+1} = \mathbf{x}^{*k} + \epsilon \mathbf{F}^\dagger \mathbf{A} \mathbf{F} \mathbf{w} \\ \mathbf{z}^{*k+1} = \mathbf{w} + \frac{\epsilon}{2} \nabla_{\mathbf{x}} \log p(\mathbf{x}^{*k+1}|\mathbf{y}) \\ \left. \begin{array}{l} x_i^{*k+1} := -x_i^{*k+1} \\ z_i^{*k+1} := -z_i^{*k+1} \end{array} \right\} \text{if } x_i^{*k+1} < 0 \end{array} \right. \quad (17)$$

$$\text{Accept / reject} \left[\begin{array}{l} u \sim \text{unif}(0, 1) \\ \alpha = \min \left[1, \frac{p(\mathbf{x}^*|\mathbf{y})}{p(\mathbf{x}^s|\mathbf{y})} e^{-\frac{1}{2}(\mathbf{z}^*)^T \mathbf{F}^\dagger \mathbf{A} \mathbf{F} \mathbf{z}^* + \frac{1}{2}(\mathbf{z}^{*0})^T \mathbf{F}^\dagger \mathbf{A} \mathbf{F} \mathbf{z}^{*0}} \right] \\ \mathbf{x}^{s+1} := \begin{cases} \mathbf{x}^* & \text{if } u < \alpha \\ \mathbf{x}^s & \text{if } u \geq \alpha \end{cases} \end{array} \right. \quad (18)$$

With the BCCB approximation, the calculation of the gradient $\nabla_{\mathbf{x}} \log p(\mathbf{x}|\mathbf{y}) = \nabla_{\mathbf{x}} \log p(\mathbf{x}) + \mathbf{A}^T \left(\frac{\mathbf{y}}{\mathbf{A}\mathbf{x}} - \bar{\mathbf{1}} \right)$ constitutes the most computationally expensive step, entailing a projection (i.e. the simulation of the light propagation process) and a back-projection. In the experiments that follow, in order to speed up the computations, we utilize a GPU-accelerated ray-tracer, achieving a rate of 50 gradient evaluations per sec with a 64^3 image lattice, using a GPU Nvidia Tesla K20.

3 Experiments

The following experiments were performed to investigate the use of MCMC for tomographic image formation. A healthy volunteer was administered 170 *MBq* of FDG and scanned for 60 *min* with a Siemens Biograph mMR scanner. The list-mode data was binned (discarding intra-ring coincidences, to reduce the complexity) into three sinograms of duration 20 *min*, 40 *min* and 60 *min* since the start of the acquisition. For each sinogram, an initial reconstruction on a image lattice of 64^3 4 *mm* isotropic voxels was obtained by running the standard MLEM algorithm until approximate convergence. The first column of the Hessian matrix was computed, for each of the three measurements, using the initial reconstruction. A total of $N_S = 5000$ samples was drawn for each of the three measurements using the uninformative prior (the number of samples was selected empirically to obtain smooth marginals - Fig. 3). The number of integration steps was set to $N_K = 10$ and the integration step size was set empirically to $\epsilon = 0.08$ to obtain acceptance rate near 0.5 – see in Fig. 1-(I). Fig. 1-(II-IV) provides a visualization of the BCCB approximation for the 40 *min* experiment.

3.1 Uncertainty quantification

The samples of the posterior density can be used to quantify and visualize the uncertainty associated to the tomographic measurement. Fig. 2(III) displays the sample variance for the 40 *min* acquisition and Fig. 3 reports the histograms of the samples relative to the three voxels A,B,C, indicated in Fig. 2(I), for the three acquisition times. Providing uncertainty images such as 2(III) alongside the reconstructions themselves is of direct potential clinical use in e.g., tumor imaging to improve diagnosis/detection and interpretation of clinical images. The plots in Fig. 3 demonstrate that the method allows one to capture the increase of information obtained by increasing the scanning time – something that conventional reconstruction algorithms based on point estimation cannot provide. With (much) more powerful computers, this may ultimately provide application-specific real-time feedback on when to stop scanning.

3.2 Decision theoretic criteria for image reconstruction

Decision theory allows us to obtain estimates that are optimal according to certain risk criteria. In the decision theoretic framework, estimates are obtained

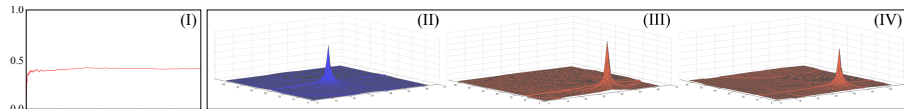


Fig. 1. 20 *min* scan: (I) Acceptance rate; (II) Column of the FIM matrix, central voxel; (III) Column of the FIM matrix, peripheral voxel; (IV) Column of the BCCB FIM matrix, peripheral voxel.

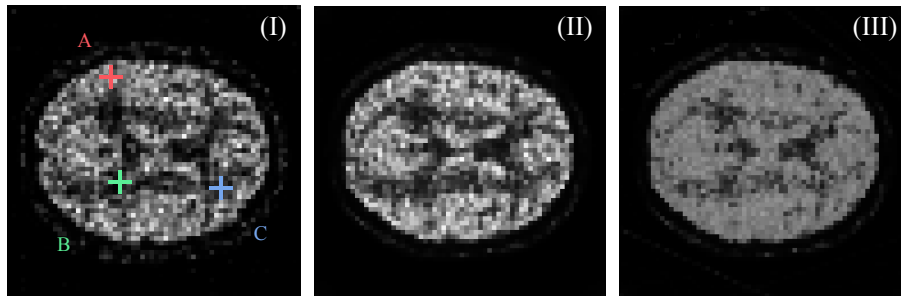


Fig. 2. 40 *min* scan: (I) MAP estimate - MLEM; (II) MMSE estimate (sample mean); (III) Sample variance of each voxel's marginal posterior distribution. A,B,C see Fig.3.

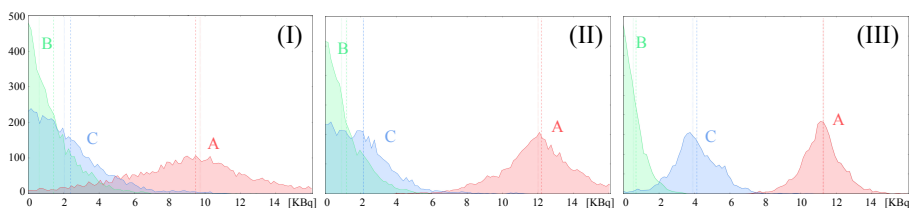


Fig. 3. Histograms of the posterior samples for voxels A,B,C in Fig. 2. (I) 20 *min* scan; (II) 40 *min* scan; (III) 60 *min* scan. Dotted lines indicate the MLE estimates and dashed lines the MMSE estimates (i.e., the sample means). Note the reduction in reconstruction uncertainty with longer acquisition time.

by minimizing the expected risk:

$$\hat{\mathbf{x}} = \arg \min_{\tilde{\mathbf{x}}} \int L(\tilde{\mathbf{x}} - \mathbf{x}) p(\mathbf{x}|\mathbf{y}) d\mathbf{x}, \quad (19)$$

where $L(\tilde{\mathbf{x}} - \mathbf{x})$ represents the loss associated to choosing $\tilde{\mathbf{x}}$ when the true underlying value is \mathbf{x} . The MAP estimation criterion corresponds to the particular choice of the loss function that assigns the same penalty whenever the estimate is not the true underlying variable and none otherwise: $L(\tilde{\mathbf{x}} - \mathbf{x}) = 1 - \delta_{\tilde{\mathbf{x}}\mathbf{x}}$, where δ is the Kronecker Delta function. In this case, the minimum risk solution simplifies to $\hat{\mathbf{x}} = \arg \max_{\tilde{\mathbf{x}}} p(\tilde{\mathbf{x}}|\mathbf{y})$. In general, however, the integral (19) is intractable. The posterior sampling approach via MCMC enables us to approximate the integral in equation (19) as a finite sum over the samples, allowing the definition of arbitrary risk criteria.

As an example we can perform image reconstruction by computing the Bayesian minimum mean squared error estimate (MMSE), which assigns cost in proportion to the distance of an estimate from the true value, in the sense of the $L2$ -norm: $\hat{\mathbf{x}} = \arg \min_{\tilde{\mathbf{x}}} \int \|\tilde{\mathbf{x}} - \mathbf{x}\|^2 p(\mathbf{x}|\mathbf{y}) d\mathbf{x} = \int \mathbf{x} p(\mathbf{x}|\mathbf{y}) d\mathbf{x} \approx \frac{1}{N_S} \sum_{s=1}^{N_S} \mathbf{x}^s$. MMSE is equal to the mean of the posterior distribution and approximated by the sample mean. Fig. 2 displays the central transverse slice of (I) the MAP estimate (with non-informative prior) obtained with the MLEM algorithm it-

erated until approximate convergence and of (II) the MMSE estimate obtained by averaging the samples. The vertical lines in Fig.3 indicate the MMSE and MAP estimates for voxels A,B and C, indicating how the two differ quantitatively. Note that sampling allows for more general reconstruction loss functions than the two reported in Fig. 2, e.g., functions that in tumor detection penalize false negatives more heavily than false positives. Exploring this further is left for future work.

4 Conclusion

In this paper, we have introduced an algorithm for the posterior sampling of the unknown parameters in tomographic imaging. The algorithm enables a number of new possibilities. These include the quantification of the uncertainty of tomographic measurements and the possibility to define arbitrary risk criteria at a very low level of the image formation process. We have described the algorithm in detail and explored its use with simple experiments in PET. Future work will evaluate the comparison of MAP estimation and reconstruction under the MMSE criterion, will explore other risk criteria, and will consider the Bayesian variational framework in search of solutions with increased computational efficiency.

5 Acknowledgements

This research was supported by the NIH NCRR (P41-RR14075), the NIH NIBIB (R01EB013565), TEKES (ComBrain), the Lundbeck Foundation (R141-2013-13117), NVIDIA Corporation.

References

1. L.A. Shepp, Y. Vardi. Maximum Likelihood Reconstruction for Emission Tomography. *IEEE TMI*, 1982, 1(2):113–22.
2. S. Geman, D. Geman. Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE TPAMI*, 1984, 6(6):721–41.
3. S. Roth, M. J. Black. Fields of experts. *IJCV*, 2009, 82(2):205–229.
4. M. F. Tappen, B. C. Russell, W. T. Freeman. Exploiting the sparse derivative prior for super-resolution and image demosaicing. *Int. Workshop SCTV*, 2003.
5. Y. Liand, D.P.Huttenlocher. Learning for optical flow using stochastic optimization. *ECCV* 2008.
6. P.J. Green. Bayesian reconstructions from emission tomography data using a modified EM algorithm. *IEEE TMI*, 1990, 9(1):84–93.
7. M. Girolami, B. Calderhead. Riemann Manifold Langevin and Hamiltonian Monte Carlo Methods. *J. of the Royal Stat. Soc.*, 2014, 73(2), 123–214.
8. S. Duane *et al.* Hybrid Monte Carlo. *Physics Letters B*, 1987, 195(2):216–222.
9. R.M. Neal. Probabilistic Inference Using Markov Chain Monte Carlo Methods. Technical Report CRG-TR-93-1, 1993, University of Toronto.
10. B. Fischer, J. Modersitzki. Fast inversion of matrices arising in image processing. *Numerical Algorithms*. 2000, 22(1):1–11.