Closed-Form Relaxation for MRF-MAP Tissue Classification Using Discrete Laplace Equations

Alexis Roche

Siemens Research, CIBM, Lausanne, Switzerland
alexis.roche@epfl.ch

Abstract. While Markov random fields are very popular segmentation models in medical image processing, the associated maximum a posteriori (MAP) estimation problem is usually solved using iterative methods that are prone to local maxima. We show that a variant of the random walker algorithm can be seen as a relaxation method for the MAP problem under the Potts model. The key advantage of this technique is that it boils down to a sparse linear system with a uniquely defined explicit solution. Our experiments further demonstrate that the resulting MAP approximation can be used to improve the classical mean-field algorithm in terms of MAP estimation quality.

1 Introduction

Many image segmentation problems can be conveniently formulated using Markov random fields (MRF), however the associated task of computing the maximum a posteriori (MAP) segmentation is combinatorial NP-hard. Early MRF-MAP tracking methods include the ICM algorithm [1] known to be fast but highly prone to local maxima, and simulated annealing [2] which may be hopelessly slow in practice. Over the past two decades, several approaches have been proposed to work around these limitations.

One such approach, which stems from classical optimization theory, is relaxation. The basic idea is to substitute the combinatorial optimization problem with a continuous one which, in image segmentation context, involves extending the MAP search to the space of probabilistic assignments from voxels to classes. An approximation to the MAP is found by binarizing the optimal such assignment. Relaxation for MRF-MAP has been implemented using convex programing [3,4,5], which guarantees a unique solution but tends to be computationally expensive as it relies on constrained optimization.

Meanwhile, message-passing algorithms have emerged from the machine learning community for inference on probabilistic graphical models [6]. In particular, the variational expectation-maximization (VEM) algorithm, also known as mean-field algorithm, has long been used in brain imaging, though sometimes through ad-hoc variants [7,8,9,10]. Other message-passing schemes used in computer vision include belief propagation and tree-reweighted message-passing [11]. Message-passing can be viewed as a special kind of relaxation for MRF-MAP,

which is rather fast as it does do not require handling explicit inequality constraints, but is initialization-dependent since the underlying objective function is typically non-convex.

Another research trend has been to apply methods from deterministic graph theory to image segmentation. Nowadays, graph cuts [12] are widely regarded as the most robust methods for MRF-MAP [11]. In the binary segmentation case where MRF-MAP amounts to a max-flow min-cut problem, they are guaranteed to find the global maximum if unique. However, conventional graph cut methods such as the expansion and swap algorithms iterate over labels or pairs of labels when more than two labels are involved, and thus become prone to local convergence in addition to being slower.

A method that closely relates with graph cuts is the random walker (RW) algorithm of Grady [13,14], which was previously exposed in a different and less general form by Marroquin *et al* [15]. In this work, we show how to re-tune RW to yield a powerful MRF-MAP relaxation method that boils down to solving a sparse linear system. Although RW stems from a variational problem similar to the min-cut in the two-label case [16], its deep connections with MRF-based segmentation have been somewhat overlooked so far. We further advocate a method that combines the proposed relaxation with the traditional VEM algorithm.

2   MRF-MAP Segmentation

The MRF-MAP problem under the Potts model for labeling an image Y in K classes amounts to minimizing the following energy [17]:

$$L(\delta, \theta) = -\sum_i \delta_i^\top \log \ell_i(\theta) + \beta \sum_{i,j} w_{ij}(1 - \delta_i^\top \delta_j), \qquad (1)$$

where $\delta = (\delta_1, \delta_2, \ldots)$ is a collection of "delta-distributions", that is, for each voxel i, δ_i is a K-dimensional vector with a single non-zero component $\delta_{ik} = 1$ corresponding to the voxel label. The weights w_{ij} encode spatial interactions between voxels and are usually symmetric, equal to one if voxels i and j are neighbors according to a given discrete topology, and zero otherwise. The first term in the right hand side involves the likelihood $\ell_{ik}(\theta) = p(y_i|k, \theta)$ of the labels at voxel i, where θ is a nuisance parameter vector to be estimated. Under the usual Gaussian noise model, $p(y_i|k, \theta) = N(y_i; \mu_k, \sigma_k)$ and $\theta = (\mu_1, \sigma_1, \ldots, \mu_K, \sigma_K)$ is the concatenation of mean intensities and standard deviations over classes. Note, however, that the likelihood can be substituted with any external field without changing the analysis that follows.

2.1   Free Energy Relaxation

A known relaxation method to approximate the minimization of (1) is to minimize the so-called free energy function over arbitrary probability masses q_i,

$$\tilde{L}(q, \theta) = -\sum_i q_i^\top \log \ell_i(\theta) + \beta \sum_{i,j} w_{ij}(1 - q_i^\top q_j) + \sum_i q_i^\top \log q_i, \qquad (2)$$

which defines a continuous extension of (1) in the sense that $\tilde{L}(\delta, \theta)$ and $L(\delta, \theta)$ coincide on the space of delta-distributions. While the global minimization of (2) is intractable, one may resort to a greedy approach based on an alternate minimization along the q_i's, yielding explicit updates [10]:

$$q_i \propto \ell_i(\theta)e^{2\beta \sum_j w_{ij} q_j}$$

Applying this equation iteratively by cycling through the voxels corresponds to the E-step of a VEM algorithm and is guaranteed, under broad conditions, to converge to a local minimum of free energy. In the VEM algorithm, probability updates are interleaved with minimizations along θ to concurrently refine intensity parameters. The VEM algorithm is nevertheless dependent on starting values for both q and θ, and there is no theoretical warranty as to the accuracy of the resulting MAP approximation even at fixed θ.

2.2   Laplace Relaxation

We now describe another MAP relaxation approach that yields a convex problem unlike (2). Let us start with defining a surrogate MAP energy function:

$$L_s(\delta, \theta) = \sum_i (1 - \delta_i^\top \pi_i(\theta)) + \beta \sum_{i,j} w_{ij}(1 - \delta_i^\top \delta_j) - \sum_i \log z_i(\theta),$$

where $\pi_i(\theta) = \ell_i(\theta)/z_i(\theta)$ is the likelihood at voxel i normalized to unit sum and $z_i(\theta)$ is the associated partition function. Using the inequality $\log(x) \leq x - 1$, we see that $L_s(\delta, \theta) \leq L(\delta, \theta)$ for any delta-distribution, with equality iff π_i is a delta-distribution and $\delta_i = \pi_i$ at each voxel. Moreover, we have:

$$\|\delta\|^2 = 1, \qquad 1 - \delta^\top v = \frac{1}{2}\|\delta - v\|^2 + \frac{1}{2} - \frac{1}{2}\|v\|^2,$$

for any delta-distribution δ and vector v. Therefore, the following function defines a continuous extension of L_s over arbitrary distributions:

$$\tilde{L}_s(q, \theta) = \frac{1}{2} \sum_i \|q_i - \pi_i(\theta)\|^2 + \frac{\beta}{2} \sum_{i,j} w_{ij}\|q_i - q_j\|^2 + C(\theta), \qquad (3)$$

where $C(\theta) = \sum_i(-\log z_i(\theta) + \frac{1}{2} - \frac{1}{2}\|\pi_i(\theta)\|^2)$. Clearly, $\tilde{L}_s$ is quadratic and strictly convex in q. Minimizing it at fixed θ yields the first-order condition:

$$\forall k, \quad (\mathbb{I} + \lambda L)Q_k = \Pi_k, \qquad \text{with} \quad \lambda = 2\beta, \qquad (4)$$

which is a set of sparse linear systems, where L is the Laplacian matrix of the image grid considered as a graph with weights w_{ij} and $\mathbb{I}$ is the identity matrix with size equal to the number of voxels. Q_k stands for the probability image associated with class k, i.e. $Q_{ki} = q_{ik}$, and Π_k similarly represents the normalized likelihood image for class k.

Equation (4) turns out to be a vector-valued discrete Laplace equation and is equivalent to the method proposed by Marroquin *et al* [15] as an approximation to the same MAP problem. The key property is that the unique solution is a probability map without the need to incorporate explicit equality or inequality constraints, which provides a massive computational advantage over other relaxation approaches [3,4,5]. In [15], the smoothing parameter λ was not related to the MRF parameter β and was tuned empirically. We showed here that setting $\lambda = 2\beta$ ensures that the surrogate energy L_s is uniformly upper bounded by the MAP objective (1), therefore (4) qualifies as a relaxation method for the MAP problem.

A generalization of (4) is the multilabel RW algorithm [13,14]. Our strategy should however be expected to differ significantly from the original RW in practice, since both the weights w_{ij} and the "diagonal matrix" (here, the identity) are chosen in different ways, independently from the data in our case.

Owing to the inequality $L_s(\delta, \theta) \leq L(\delta, \theta)$, the MAP may be bracketed using the Laplace relaxation solution $q_\star$ and its binarization $\delta_\star$:

$$\tilde{L}_s(q_\star) \leq \min_\delta L(\delta) \leq L(\delta_\star),$$

hence providing some confidence bounds on the MAP approximation. Note that such a lower bound is not available for the VEM output as it may not be a global minimizer of free energy (2).

Also, we shall note that there is no explicit solution to minimizing (3) with respect to θ, unlike the case of free energy. Therefore, Laplace relaxation does not come with a simple built-in method for intensity parameter estimation.

3 Experiments

This section compares both relaxation methods presented above in brain tissue classification. We used a subset of 248 brain MR T1-weighted images from the Alzheimer's Disease Neuroimaging Initiative database (ADNI, `adni.loni.ucla.edu`) acquired on both 1.5 Tesla and 3 Tesla scanners from different manufacturers, with voxel volume ranging from 1 to 1.9 mm^3. The dataset includes 163 healthy controls and 85 diagnosed AD patients (55% males, 45% females) with mean age 77 ± 7 years. As a pre-processing, the images were corrected for bias field using the N3 method [18] and skull-stripped by non-rigid registration with a template [19].

We here focus on further classifying the skull stripped data into cerebrospinal fluid (CSF), gray matter (GM), and white matter (WM). To this end, we used a 4-class Potts prior model using a 6-neighborhood system with two classes representing GM to account for the usually rather large intensity variations between cortical GM and deep GM in T1-weighted images. The spatial regularization parameter was set to $\beta = 0.5$ based on previous tests. No external field was incorporated to the model at this stage to avoid biasing tissue classification towards an atlas [20]. This model was found to yield high overlap with ground

truth segmentation on Brainweb data [21] using the conventional VEM algorithm (Jaccard indices larger than 0.88).

The VEM algorithm was mainly implemented in Python based on the Scientific Python package (`www.scipy.org`) with a subroutine in C for higher performance. Laplace relaxation was implemented in pure Python inspired by the random walker implementation by E. Gouillart (`github.com/emmanuelle`) using a smoothed aggregation solver [22]. The θ parameter supplied to each method was computed using a simple moment matching technique [10].

The computation time on a single processor Intel Core i7-975 CPU 3.33GHz was about 1.5 seconds per iteration for VEM (including update of θ), and about 15 seconds for Laplace relaxation. The VEM algorithm was run for 50 iterations, which achieved satisfactory convergence in all cases (relative variations of free energy lower than 2.5×10^{-4}), resulting in a total computation time of 75 seconds per image.
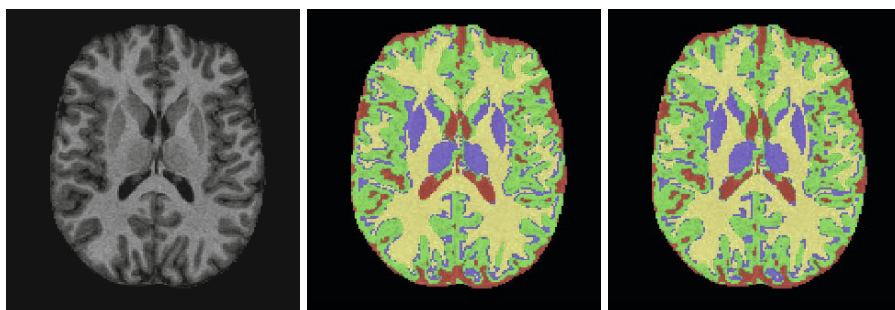

Fig. 1. Comparison of MAP estimates found by Laplace relaxation (middle) and the VEM algorithm (right) for a skull-stripped MR T1-weighted image (left). Label colors are red for CSF, blue and green for GM, and yellow for WM.

Figure 1 illustrates that MAP estimates found by Laplace relaxation generally look very similar from visual inspection to those provided by the VEM algorithm. To quantify this, we computed minimum Jaccard overlap coefficients,

$$J = \min_k \frac{|A_k \cap B_k|}{|A_k \cup B_k|},$$

where A_k and B_k denote the sets of voxel labeled as k in the respective classifications. Overlap coefficients ranged from 0.33 to 0.87 on the whole dataset with mean 0.64 and standard deviation 0.13. They were found from ANOVA to correlate negatively with voxel volume (p-value $< 10^{-10}$) and, to a lesser extent, with pathology (p-value $< 10^{-2}$), the agreement between both segmentation methods being higher for AD patients. Correlations with age and gender were not significant.

While the VEM algorithm was slower than Laplace relaxation, it converged to a solution of lower energy (1) in all of the 248 cases and was therefore more accurate at tracking the MAP despite being theoretically prone to local minima. To further investigate the benefit of Laplace relaxation, we tested a VEM variant, hereafter referred to as LR-VEM, where the class probability map q is initialized as the binarized solution of Laplace relaxation, as opposed to the standard initialization with a uniform distribution over labels.
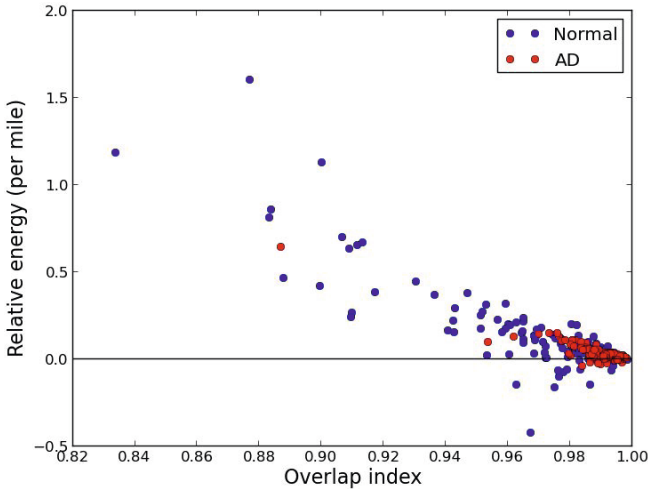

Fig. 2. Comparison between standard VEM and LR-VEM algorithms: plot of final relative MAP energy values against Jaccard indices for 248 ADNI subjects

Figure 2 plots a relative measure of MAP estimation quality of LR-VEM versus VEM, defined as $L_{final}^{\text{VEM}}/L_{final}^{\text{LR-VEM}} - 1$, where L_{final} denotes the energy level reached after 50 iterations, against the overlap indices computed between the respective corresponding MAP estimates. In 83.5% cases, LR-VEM achieved lower energy than VEM, while the converse happened in 16.5% cases. Segmentation results showed non-negligible differences in 10% cases as shown by overlap indices lower than 0.95. In all such cases, the MAP estimate from LR-VEM had the lower energy. Conversely, when VEM achieved lower energy than LR-VEM, the respective MAP estimates were almost identical. This provides some evidence that initialization with Laplace relaxation makes the VEM algorithm more robust in tracking the MAP.

ANOVA revealed that overlap indices correlate strongly with pathology (p-value $< 10^{-6}$) and age (p-value $< 10^{-5}$) in the sense that differences between VEM and LR-VEM are reduced for diseased or aged subjects. A slight negative correlation with voxel volume (p-value $< 10^{-2}$) was found in this case, and again no significant correlation with gender.

Moreover, the LR-VEM algorithm required an average of 7 ± 10 iterations less than VEM to achieve the same tolerance on free energy variations as in the last VEM iteration, meaning that the computational overhead of Laplace relaxation is compensated for by faster convergence.

4 Discussion

Laplace relaxation offers a fast alternative to the VEM algorithm for MRF-MAP classification that is independent from an initial label probability assignment. In our brain tissue classification experiments, Laplace relaxation produced results quite similar to the VEM algorithm (as shown by minimum Jaccard indices of 0.64 ± 0.13). The MAP classifications output by VEM were, however, more accurate. In a different scenario, Laplace relaxation can be used as an initialization step for the VEM algorithm, leading to noticeable improvements in MAP estimation in about 10% cases without significant overhead in computation time due to faster convergence.

We did not expect massive improvements in the whole-brain classification setting where the VEM algorithm has previously been reported to be robust. We anticipate that the effect of Laplace relaxation may be more substantial in segmentation applications that target specific anatomical structures since local volume or shape assessments are likely to be sensitive to small variations in tissue probability maps. The benefit of Laplace relaxation in brain morphometry is thus to be further investigated.

Laplace relaxation is currently applicable to a subclass of MRF models that includes extensions of the Potts model that involve non-stationary scalar-weighted interactions and addition of any external field. Future work will aim to extend the methodology to other MRF models for which iterative methods such as the VEM algorithm may have serious local convergence issues, in particular models that incorporate strong topological constraints via tissue-dependent interaction potentials.

References

1. Besag, J.: Spatial interaction and the statistical analysis of lattice systems. Journal of the Royal Statistical Society: Series B 36(2), 192–236 (1974)
2. Geman, S., Geman, D.: Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images. IEEE Transactions on Pattern Analysis and Machine Intelligence 6(6), 721–741 (1984)
3. Ravikumar, P., Lafferty, J.: Quadratic Programming Relaxations for Metric Labeling and Markov Random Field MAP Estimation. In: International Conference on Machine Learning, pp. 737–744 (2006)

4. Cour, T., Shi, J.: Solving Markov Random Fields with Spectral Relaxation. Journal of Machine Learning Research 2, 75–82 (2007)
5. Kumar, M.P., Kolmogorov, V., Torr, P.H.: An Analysis of Convex Relaxations for MAP Estimation of Discrete MRFs. Journal of Machine Learning Research 10, 71–106 (2008)
6. Minka, T.P.: Divergence measures and message passing. Technical Report MSR-TR-2005-173, Microsoft Research Ltd., Cambridge, UK (December 2005)
7. Van Leemput, K., Maes, F., Vandermeulen, D., Suetens, P.: Automated model-based tissue classification of MR images of the brain. IEEE Transactions on Medical Imaging 18(10), 897–908 (1999)
8. Zhang, Y., Brady, J., Smith, S.: Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm. IEEE Transactions on Medical Imaging 20(1), 45–57 (2001)
9. Forbes, F., Fort, G.: Combining Monte Carlo and Mean-Field-Like Methods for Inference in Hidden Markov Random Fields. IEEE Transactions on Image Processing 16(3), 824–837 (2007)
10. Roche, A., Ribes, D., Bach-Cuadra, M., Krueger, G.: On the Convergence of EM-Like Algorithms for Image Segmentation using Markov Random Fields. Medical Image Analysis 15(6), 830–839 (2011)
11. Szeliski, R., Zabih, R., Scharstein, D., Veksler, O., Kolmogorov, V., Agarwala, A., Tappen, M., Rother, C.: A comparative study of energy minimization methods for Markov random fields with smoothness-based priors. IEEE Transactions on Pattern Analysis and Machine Intelligence 30(6), 1068–1080 (2008)
12. Boykov, Y., Veksler, O., Zabih, R.: Fast approximate energy minimization via graph cuts. IEEE Transactions on Pattern Analysis and Machine Intelligence 23(11), 1222–1239 (2001)
13. Grady, L.: Multilabel Random Walker Image Segmentation Using Prior Models. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), vol. 1, pp. 763–770 (2005)
14. Grady, L.: Random Walks for Image Segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence 28(11), 1768–1783 (2006)
15. Marroquin, J., Vemuri, B., Botello, S., Calderon, F., Fernandez-Bouzas, A.: An Accurate and Efficient Bayesian Method for Automatic Segmentation of Brain MRI. IEEE Transactions on Medical Imaging 21(8), 934–945 (2002)
16. Couprie, C., Grady, L., Najman, L., Talbot, H.: Power Watersheds: A Unifying Graph-Based Optimization Framework. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1–17 (2010)
17. Li, S.Z.: Markov random field modeling in computer vision. Springer, Berlin (1995)
18. Sled, J., Zijdenbos, A., Evans, A.: A nonparametric method for automatic correction of intensity nonuniformity in MRI data. IEEE Transactions on Medical Imaging 17, 87–97 (1998)
19. Hermosillo, G., Chefd'Hotel, C., Faugeras, O.: Variational Methods for Multimodal Image Matching. International Journal of Computer Vision 50(3), 329–343 (2002)
20. Ribes, D., et al.: Comparison of tissue classification models for automatic brain MR segmentation. In: Int. Society for Magnetic Resonance in Medicine (2011)
21. Kwan, R.S., Evans, A., Pike, G.: MRI simulation-based evaluation of image-processing and classification methods. IEEE Transactions on Medical Imaging 18(11), 1085–1097 (1999)
22. Bell, W.N., Olson, L.N., Schroder, J.B.: PyAMG: Algebraic multigrid solvers in Python v2.0, Release 2.0 (2011)