Supplementary material for "Newton-type Methods for Inference in Higher-Order Markov Random Fields"

Hariprasad Kannan
CentraleSupélec-INRIA Saclay
Université Paris-Saclay
hkannan@gmail.com

Nikos Komodakis
Ecole des Ponts ParisTech
Université Paris Est
nikos.komodakis@enpc.fr

Nikos Paragios
CentraleSupélec-INRIA Saclay
Université Paris-Saclay
nikos.paragios@ecp.fr

Abstract

This document contains description of some points related to implementation and some experimental results.

1. Numerically stable sum-product computation

When the damping parameter is increased to high values (low smoothness), the terms present in equations 13 and 14 will all take zero value and the summations in these equations will be of value zero, which is erroneous. This is because these terms are in fact exponentials of terms corresponding to the nodes and cliques. So, $\psi_c(\boldsymbol{x}_c)$ and $\psi_i(x_i)$ are of the form $\exp(\theta_c(\boldsymbol{x}_c))$ and $\exp(\theta_i(x_i))$, respectively, in a suitably defined graphical model. This numerical issue is addressed by storing offsets ($M_{bc}(\boldsymbol{x}_n)$) for each term of the message $m_{bc}(\boldsymbol{x}_n)$. Hence, $m_{bc}(\boldsymbol{x}_n)$ in equation 13 is computed as follows,

$$\begin{aligned} m_{bc}(\boldsymbol{x}_n) = \sum_{\boldsymbol{x}_{b\setminus n}} \exp\big[\theta_b(\boldsymbol{x}_b) + \sum_{i\in b\setminus n} \theta_i(x_i) \\ + \quad \log(m_{ab}(\boldsymbol{x}_{b\setminus n})) \quad - \quad M_{bc}(\boldsymbol{x}_n)\big] \end{aligned} \tag{1}$$

$$\begin{aligned} \text{where} \quad M_{bc}(\boldsymbol{x}_n) = \max_{\boldsymbol{x}_{b\setminus n}} \big[\theta_b(\boldsymbol{x}_b) + \sum_{i\in b\setminus n} \theta_i(x_i) \\ + \quad \log(m_{ab}(\boldsymbol{x}_{b\setminus n}))\big] \end{aligned} \tag{2}$$

This ensures that at least one term in the summation is of value 1, hence, the summation is assured to be non-zero. Later, the required marginals are computed as $p_i^c(x_i) = \frac{\psi_i(x_i)}{Z} \sum_{\boldsymbol{x}_n} \big[\big(\prod_{j\in n\setminus i} \psi_j(x_j)\big) m_{bc}(\boldsymbol{x}_n)\big]$, where Z ensures that the marginals sum to one. In this expression each summand is computed as $\big[\sum_{j\in n\setminus i} \theta_j(x_j) + \log(m_{bc}(\boldsymbol{x}_n)) + M_{bc}(\boldsymbol{x}_n)\big]$. This approach can be carefully extended to equation 14, also. We will need to maintain two offsets, $M_{bc}(\boldsymbol{x}_n)$ as before and also, $\bar{M}$ corresponding to the second sum in equation 14.

2. Exit condition for large problems

A simultaneous criterion on function value difference, variable difference and gradient has been used (adapted from §8.2.3.2 [1]). Given adjacent iterations k and $k+1$, for the dual function $g(\boldsymbol{\delta})$, function value difference is $\Delta g = g_{k+1}(\boldsymbol{\delta}) - g_k(\boldsymbol{\delta})$ and l_∞ norm of the step taken is $\Delta\boldsymbol{\delta} = |\boldsymbol{\delta}_{k+1} - \boldsymbol{\delta}_k|_\infty$. Thus, the conditions enforced are $\Delta g < 10^{-3}, \Delta\boldsymbol{\delta} < 10^{-3}, \|\nabla g(\boldsymbol{\delta}_{k+1})\|_\infty < 0.5$ and we exit if all these conditions are satisfied.

3. Additional result

Here, we present another result for stereo with curvature prior, which we didn't present in the main paper due to lack of space. As before, the cliques are of shape 1×3 and 3×1 and the clique energies are truncated and the unaries are absolute differences. The results are for Venus, image size 191×217, 20 depth levels. As before, pattern-based sum-product was computed on clique chains. AD3 showed poor convergence behavior for this large problem, also. FISTA ran for more than 180 iterations without converging.

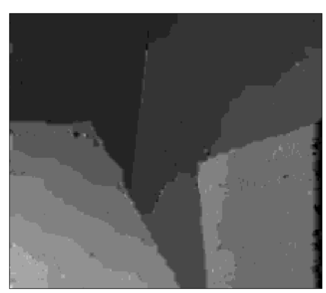

Figure 1. Venus result for quasi-Newton. Iterations: 102. Non-smooth dual: 39153. Integer primal: 39230.5.

References

[1] P. E. Gill, W. Murray, and M. H. Wright. Practical optimization. 1981. 1