Scaling MPE Inference for Constrained Continuous Markov Random Fields with Consensus Optimization: Supplementary Material

Stephen H. Bach
University of Maryland, College Park
College Park, MD 20742
bach@cs.umd.edu

Matthias Broecheler
Aurelius LLC
matthias@thinkaurelius.com

Lise Getoor
University of Maryland, College Park
College Park, MD 20742
getoor@cs.umd.edu

Dianne P. O'Leary
University of Maryland, College Park
College Park, MD 20742
oleary@cs.umd.edu

A Convergence of ADMM

After each iteration k of ADMM the sizes of the primal and dual residuals are:

$$\|r^{k+1}\|_2 \equiv \left(\sum_{i=1}^{m+r} \|\mathbf{x}_i^{k+1} - \mathbf{X}_i^{k+1}\|_2^2\right)^{1/2} \qquad \|s^{k+1}\|_2 \equiv \rho\left(\sum_{g=1}^{n} \mathcal{K}_g \left(X_g^{k+1} - X_g^{k}\right)^2\right)^{1/2}$$

where $\mathcal{K}_g$ is the number of copies of the variable X_g [1].

It is known that if the objective is closed, proper, and convex, and strong duality holds, then $\|r^k\|_2 \to 0$, $\|s^k\|_2 \to 0$, and the objective approaches $p^\star$ as $k \to \infty$, where $p^\star$ is the optimal value of the objective.

See Gabay and Mercier [2], Eckstein and Bertsekas [3], and Boyd et. al. [1] for details.

B Implementation details

B.1 Initialization

All variables in $\mathbf{X}$ were initialized to 0.5.

B.2 Stopping criteria

Boyd et. al. [1] suggest the stopping criteria

$$\|r^k\|_2 \leq \epsilon^{\text{abs}} \sqrt{\sum_{g=1}^{n} \mathcal{K}_g} + \epsilon^{\text{rel}} \max\left\{ \left(\sum_{i=1}^{m+r} \|\mathbf{x}_i^k\|_2^2\right)^{1/2}, \left(\sum_{g=1}^{n} \mathcal{K}_g \left(X_g^k\right)^2\right)^{1/2} \right\} \tag{1}$$

$$\|s^k\|_2 \leq \epsilon^{\text{abs}} \sqrt{\sum_{g=1}^{n} \mathcal{K}_g} + \epsilon^{\text{rel}} \left(\sum_{i=1}^{m+r} \|\mathbf{y}_i^k\|_2^2\right)^{1/2} \tag{2}$$

where $\epsilon^{\text{abs}}, \epsilon^{\text{rel}} > 0$ are user specified and $\mathcal{K}_g$ is the number of copies of the variable X_g. In all our experiments, consensus optimization terminated when criteria (1) and (2) were satisfied with $\epsilon^{\text{abs}} = 10^{-8}$ and $\epsilon^{\text{rel}} = 10^{-3}$.

B.3 Enconding the MPE problem as a conic program

To use an interior-point method, we encode Problem (1) in a second-order cone program (SOCP):

$$\arg\min_{\bar{x}} \quad c^T \bar{x} \quad \text{subject to} A\bar{x} = b \text{ and } \bar{x} \in \mathcal{K} \equiv \mathcal{K}_1^+ \times \cdots \times \mathcal{K}_{q1}^+ \times \mathcal{K}_1^L \times \cdots \times \mathcal{K}_{q2}^L$$

where $\bar{x} \in \mathbb{R}^{\bar{n}}$, $c \in \mathbb{R}^{\bar{n}}$, $A \in \mathbb{R}^{\bar{m} \times \bar{n}}$, $b \in \mathbb{R}^{\bar{m}}$, and $\mathcal{K}$ is a direct product of sets called cones. Each cone $\mathcal{K}^+$ is a non-negative orthant cone $x \geq 0$ and each cone $\mathcal{K}^L$ is a t-dimensional rotated second-order cone $2x_1x_2 \geq \|x_{3:t}\|_2^2$ (sometimes called a rotated Lorentz cone). Note that other definitions of SOCPs which use un-rotated second-order cones are possible, but rotated second-order cones are more convenient for our purposes.

Before continuing, there are a few shorthands we will use in describing our SOCP. The constraint $A\bar{x} = b$ restricts $\bar{x}$ to an affine subspace, and we will describe A and b as if they are a matrix and a vector, respectively, since the meaning is clear. We will mention including linear equality constraints in the SOCP. When we do, we mean that each constraint is a row in A and a component in b acting on the components of $\bar{x}$ corresponding to the components of x_i. When we mention linear inequality constraints, we mean to first convert them to equality constraints by adding a component to $\bar{x}$ for each such constraint to act as a "slack" variable. Each slack variable is constrained to lie in a non-negative orthant cone. Also, each component of c is zero unless stated otherwise.

We first include an n-dimensional component $\bar{x}_v$ in $\bar{x}$ to represent the variables $\mathbf{X}$. To enforce $\mathbf{X} \in [0,1]^n$ each dimension of $\bar{x}_v$ is constrained to lie in a non-negative orthant cone, and we constrain $(\bar{x}_v)_i \leq 1, i = 1, \ldots, n$.

We now consider encoding each hinge-loss potential function ϕ_j. For each we include a non-negative orthant component $\bar{x}_{\phi_j}$ in $\bar{x}$ and include a linear inequality constraint $\bar{x}_{\phi_j} \geq \ell_j(\bar{x}_v)$. If $p_j = 1$ then the objective coefficient of $\bar{x}_{\phi_j}$ is Λ_j. If $p_j = 2$ then we include a 3-dimensional rotated second-order cone $\bar{x}^{L,\phi_j}$ in $\bar{x}$, where the objective coefficient of $(\bar{x}^{L,\phi_j})_1$ is Λ_j, constrain $(\bar{x}^{L,\phi_j})_2 = 1/2$ and constrain $(\bar{x}^{L,\phi_j})_3 = \bar{x}_{\phi_j}$.

Finally we include each linear constraint C_k.

References

[1] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein. *Distributed Optimization and Statistical Learning Via the Alternating Direction Method of Multipliers*. Now Publishers, 2011.

[2] D. Gabay. *Applications of the method of multipliers to variational inequalities*, volume 15, chapter 9, pages 299–331. Elsevier, 1983.

[3] J. Eckstein and D. P. Bertsekas. On the Douglas-Rachford splitting method and the proximal point algorithm for maximal monotone operators. *Math. Program.*, 55(3):293–318, 1992.