①

~~ALMM~~ DD-ADMM for MRFs.

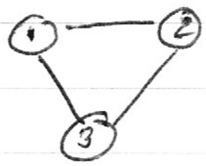
"recall" we could write

$$\text{MAP soln} = \underset{(\mu, \nu) \in M(G)}{\operatorname{argmax}} \sum_i \theta_i^T \mu_i + \sum_a \phi_a^T \nu_a$$

<u>Why</u>:

- μ_i are 0-1 variables for each value of each assignment for each <u>unary var</u> - might be vectors
- ν_a are 0-1 vectors for each n-ary factor
- The hard stuff is in $(\mu, \nu) \in M(G)$

eg.



$x_1 \in \{a, b, c)$
$x_2 \in \{d, e, f)$
$x_3 \in \{g, h, i)$

so μ_1 ~~will~~ could be $(1, 0, 0)$, $(0, 1, 0)$, $(0, 0, 1)$

etc $\theta_1 = (C_1(a), C_1(b), C_1(c))$

ν_1 for 1-2 edge, say,

then ν_1 is a 9-D vector, with a single 1 in it

$\phi_1 = (C_{12}(a, d) \; . \; . \; . \quad . \;)$ etc

Notice that only some

μ_i, ν_i are compatible

eg. we can't have

$\mu_1 = (1, 0, 0)$, $\mu_2 = (0, 1, 0)$

$\phi_1 = (0, 1, 0 \; . \; . \quad)$

Construct

$$\mathrm{conv}\left(\{\text{all compatible } \mu, \nu\text{'s}\}\right)$$

Marginal polytope

Problem: no concise rep'n available in most cases.

Relax to **Local polytope**

$$L(G) = \left\{ (\mu, \nu) \;\middle|\; \begin{array}{l} 1^T \mu_i = 1 \\ \nu_\alpha \geq 0 \quad \mu_i \geq 0 \\ H_{i\alpha} \nu_\alpha = \mu_i \end{array} \right\}$$

only one label

no negative terms in ν_α

consistency

$$H_{i\alpha}(x_i, x_\alpha) = 1 \text{ if } [x_\alpha]_i = x_i$$

The consistency term allows us to think of μ, ν as ~~form~~ a probability distribution

(Scaling)

$$H_{i\alpha} \nu_\alpha = \mu_i$$

ensures that a marginals of the factor terms are consistent w/ the μ_i

Then we could relax to

$$\max \sum_i \theta_i^T \mu_i + \sum_a \phi_a^T \nu_a$$

$$\text{st} \quad \mu, \nu \in \mathcal{L}(G)$$

In this notation, we introduce one copy of vars for each factor

ν_i^{a} ← factor (a), ← var (i)

then require they agree

$$\max \sum_a \left[\sum_i \frac{1}{d_i} \left(\theta_i^T \nu_i^a \right) + \phi_a^T \nu_a \right]$$

d_i: degree of ith node.

$$\text{s.t. } (\nu_i^{\alpha}, \nu_{\alpha}) \in M(G_a)$$

G_a: the elements of factor a

$$\nu_i^{a} = \mu_i$$

→ then do Dual Decomposition

or ADMM

Notice that

$$(\nu_i^\alpha, \nu_\alpha) \in M(G_\alpha)$$

means we are working w/ the local polytope. This is because

$$\left\{ (\mu, \nu) \,\middle|\, (\mu_{N(a)}, \nu_a) \in M(G_a) \right\}$$

The μ's that appear in factor a

is the <u>local polytope</u>

[Local constraints imposed only on each <u>factor</u>]

Augmented Lagrangian

$A_\rho(\mu, \nu, \lambda)$

$$= \sum_{a \in \text{factors}} \left[\sum_{i \in N(a)} \left[\frac{1}{d_i} \theta_i + \lambda_i^a \right]^T \nu_i^\alpha + \phi_\alpha^T \nu_\alpha \right]$$

$$- \sum_{a \in \text{factors}} \sum_{i \in N(a)} \lambda_i^{a^T} \mu_i$$

$$- \frac{\rho}{2} \sum_a \sum_i \| \nu_i^a - \mu_i \|^2$$

Now we do ADMM.

$$\nu^{k+1} = \underset{\nu}{\operatorname{argmax}}\, A_\rho(\mu^k, \nu, \lambda^k) \leftarrow \text{one per factor}$$

$$\mu^{k+1} = \underset{\mu}{\operatorname{argmax}}\, A_\rho(\mu, \nu^{k+1}, \lambda^k)$$

$$\lambda^{k+1} = \lambda^k - \{\text{usual update}\}$$

<u>Notice</u>

$\mu^{(k+1)} = \arg\max A_\rho(\mu, \nu^k, \lambda^k)$ can be

solved in closed form.

[Quadratic form ~~w/ linear constraints~~]

The slaves are interesting.

$$\min_{(\nu^a_{N(a)}, \nu_a) \in M(G_a)} \frac{\rho}{2} \sum_{i \in N(a)} \| \nu_i^a - \frac{\omega_i^a}{\rho} \|^2 - \phi_a^T \nu_a$$

cooked up out of μ^k, λ^k

<u>Variety of cases:</u>

- Factors are always 2 (MRF)

rewrite problem as:

$$\min \quad \frac{1}{2}\left[(z_1-c_1)^2 + (z_2-c_2)^2\right] - c_{12} z_{12}$$

$$\text{st.} \quad z_{12} \le z_1 \;,\; z_{12} \le z_2 \quad z_{12} \ge z_1 + z_2 - 1$$

$$(z_1, z_2, z_{12}) \in [0,1]^3$$

Straightforward to obtain closed form soln. (see Martins et al).

- Hard constraints;

- we might have factors that look like indicator fns

- eg $\phi_a(x_a) = \begin{cases} 0 & - \ x_a \text{ meets some test} \\ -\infty & \text{otherwise} \end{cases}$

eg.

$$\phi_a(x_a) = \begin{cases} 0 & \text{if } \cancel{\text{one}} \text{ exactly one of } x_a \text{ is } 1 \\ -\infty & \text{otherwise.} \end{cases}$$

one hot x-OR factor.

Slave becomes, with work,

$$\min \ \| z - c \|^2$$

$$\text{st} \qquad z \in \operatorname{conv} \delta_A$$

accept set; all x_a st factor $= 0$

in one hot case, we have

$$\min \quad \|z - c\|^2$$

$$\text{st} \quad z \in \text{Simplex}$$

$$\boxed{z \geq 0, \quad 1^T z = 1}$$

equivalently, find point in Simplex closest to c.

– via sorting, Duchi '08