A Tutorial on Dual Decomposition

Yujia Li

University of Toronto

MAP Inference for MRFs

- Energy minimization

$$\min_{\mathbf{x}} \sum_{i} \theta_i(x_i) + \sum_{f} \theta_f(x_f)$$

MAP Inference for MRFs

- Energy minimization

$$\min_{\mathbf{x}} \sum_i \theta_i(x_i) + \sum_f \theta_f(x_f)$$

- Example - unary + pairwise factors

$$\min_{\mathbf{x}} \sum_i \theta_i(x_i) + \sum_{(i,j)} \theta_{ij}(x_i, x_j)$$

MAP Inference for MRFs

- Energy minimization

$$\min_{\mathbf{x}} \sum_i \theta_i(x_i) + \sum_f \theta_f(x_f)$$

- Example - unary + pairwise + higher order factors

$$\min_{\mathbf{x}} \sum_i \theta_i(x_i) + \sum_{(i,j)} \theta_{ij}(x_i, x_j) + \sum_f \theta_f(x_f)$$

Decomposition Methods for Optimization

- Decomposable problem:

$$\min_{x,y} \; f(x) + g(y) = \min_x f(x) + \min_y g(y)$$

Two subproblems can be solved independently (in parallel).

Decomposition Methods for Optimization

- Decomposable problem:

$$\min_{x,y} \; f(x) + g(y) = \min_x f(x) + \min_y g(y)$$

  Two subproblems can be solved independently (in parallel).

- Nondecomposable problem with complicating variable:

$$\min_{x,y,z} \; f(x,z) + g(y,z)$$

Primal Decomposition

Fixing z, the problem is decomposable:

$$\min_{x,y} \; f(x,z) + g(y,z) = \min_{x} f(x,z) + \min_{y} g(y,z)$$

Primal Decomposition

Fixing z, the problem is decomposable:

$$\min_{x,y} \; f(x,z) + g(y,z) = \min_x f(x,z) + \min_y g(y,z)$$

Algorithm:

1. Solve two subproblems (in parallel) for the current z.
2. Update z, e.g. take a gradient descent step if z is continuous.

Primal Decomposition

Fixing z, the problem is decomposable:

$$\min_{x,y} \; f(x,z) + g(y,z) = \min_{x} f(x,z) + \min_{y} g(y,z)$$

Algorithm:

1. Solve two subproblems (in parallel) for the current z.
2. Update z, e.g. take a gradient descent step if z is continuous.

Let $x^* = \operatorname{argmin}_x f(x,z), \; y^* = \operatorname{argmin}_y g(y,z)$
Then gradient at z:

$$\frac{\partial f(x^*,z)}{\partial z} + \frac{\partial g(y^*,z)}{\partial z}$$

Primal Decomposition

When z is discrete and can take values from only a small set:

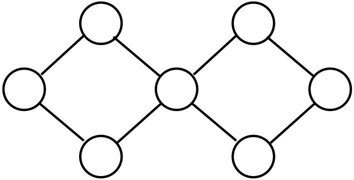
1. For each z
   - Solve the two subproblems and compute objective
2. Choose the z with the minimum objective

Primal Decomposition

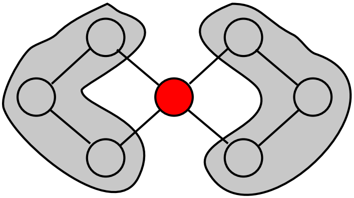
When z is discrete and can take values from only a small set:

1. For each z
   - Solve the two subproblems and compute objective
2. Choose the z with the minimum objective

Example



Loopy graph



Two chains

Primal Decomposition

Problems:

- When z is discrete, gradients not available
- When dimensionality of z is high, enumerating all z is expensive.

Dual Decomposition

Equivalent formulation of the original problem:

$$\min_{x,y,z} \; f(x,z) + g(y,z) \qquad \iff \qquad \begin{aligned} \min_{x,y,z_1,z_2} \; & f(x,z_1) + g(y,z_2) \\ \text{s.t. } & z_1 = z_2 \end{aligned}$$

Dual Decomposition

Equivalent formulation of the original problem:

$$\min_{x,y,z} \; f(x,z) + g(y,z) \qquad \Longleftrightarrow \qquad \begin{aligned} \min_{x,y,z_1,z_2} \; & f(x,z_1) + g(y,z_2) \\ \text{s.t. } & z_1 = z_2 \end{aligned}$$

Lagrangian

$$L(x,y,z_1,z_2,\lambda) = f(x,z_1) + g(y,z_2) + \lambda(z_1 - z_2)$$

Dual Decomposition

Equivalent formulation of the original problem:

$$\min_{x,y,z} f(x,z) + g(y,z) \qquad \Longleftrightarrow \qquad \begin{aligned} \min_{x,y,z_1,z_2} \; & f(x,z_1) + g(y,z_2) \\ \text{s.t. } & z_1 = z_2 \end{aligned}$$

Lagrangian

$$L(x,y,z_1,z_2,\lambda) = f(x,z_1) + g(y,z_2) + \lambda(z_1 - z_2)$$

Lagrangian dual function

$$D(\lambda) = \min_{x,y,z_1,z_2} f(x,z_1) + g(y,z_2) + \lambda(z_1 - z_2)$$

Dual Decomposition

Equivalent formulation of the original problem:

$$\min_{x,y,z} f(x,z) + g(y,z) \qquad \Longleftrightarrow \qquad \begin{aligned} \min_{x,y,z_1,z_2} \; & f(x,z_1) + g(y,z_2) \\ \text{s.t. } & z_1 = z_2 \end{aligned}$$

Lagrangian

$$L(x,y,z_1,z_2,\lambda) = f(x,z_1) + g(y,z_2) + \lambda(z_1 - z_2)$$

Lagrangian dual function

$$\begin{aligned} D(\lambda) &= \min_{x,y,z_1,z_2} f(x,z_1) + g(y,z_2) + \lambda(z_1 - z_2) \\ &= \min_{x,z_1} [f(x,z_1) + \lambda z_1] + \min_{y,z_2} [g(y,z_2) - \lambda z_2] \end{aligned}$$

Dual Decomposition

Equivalent formulation of the original problem:

$$\min_{x,y,z} f(x,z) + g(y,z) \quad \Longleftrightarrow \quad \begin{aligned} \min_{x,y,z_1,z_2} \; & f(x,z_1) + g(y,z_2) \\ \text{s.t. } & z_1 = z_2 \end{aligned}$$

Lagrangian

$$L(x,y,z_1,z_2,\lambda) = f(x,z_1) + g(y,z_2) + \lambda(z_1 - z_2)$$

Lagrangian dual function

$$\begin{aligned} D(\lambda) &= \min_{x,y,z_1,z_2} f(x,z_1) + g(y,z_2) + \lambda(z_1 - z_2) \\ &= \min_{x,z_1} [f(x,z_1) + \lambda z_1] + \min_{y,z_2} [g(y,z_2) - \lambda z_2] \\ &= \min_{x,z} [f(x,z) + \lambda z] + \min_{y,z} [g(y,z) - \lambda z] \end{aligned}$$

Lagrangian Dual Function

Denote $p^* = \min_{x,y,z} f(x,z) + g(y,z)$ as the primal optimal objective value, then

$$\begin{aligned} D(\lambda) &= \min_{x,z} \left[f(x,z) + \lambda z \right] + \min_{y,z} \left[g(y,z) - \lambda z \right] \\ &\leq \min_{x,y,z} \left[f(x,z) + \lambda z + g(y,z) - \lambda z \right] \end{aligned}$$

Lagrangian Dual Function

Denote $p^* = \min_{x,y,z} f(x,z) + g(y,z)$ as the primal optimal objective value, then

$$
\begin{aligned}
D(\lambda) &= \min_{x,z} \left[f(x,z) + \lambda z\right] + \min_{y,z} \left[g(y,z) - \lambda z\right] \\
&\leq \min_{x,y,z} \left[f(x,z) + \lambda z + g(y,z) - \lambda z\right] \\
&= \min_{x,y,z} \left[f(x,z) + g(y,z)\right] \\
&= p^*
\end{aligned}
$$

Lagrangian Dual Function

Denote $p^* = \min_{x,y,z} f(x,z) + g(y,z)$ as the primal optimal objective value, then

$$
\begin{aligned}
D(\lambda) &= \min_{x,z} [f(x,z) + \lambda z] + \min_{y,z} [g(y,z) - \lambda z] \\
&\leq \min_{x,y,z} [f(x,z) + \lambda z + g(y,z) - \lambda z] \\
&= \min_{x,y,z} [f(x,z) + g(y,z)] \\
&= p^*
\end{aligned}
$$

$D(\lambda)$ is a lower bound for p^* for any λ.

Solving the Dual Problem

Dual problem finds the tightest lower bound for p^*.

$$\max_{\lambda} D(\lambda) = \max_{\lambda} \left\{ \min_{x,z} \left[f(x,z) + \lambda z \right] + \min_{y,z} \left[g(y,z) - \lambda z \right] \right\}$$

Solving the Dual Problem

Dual problem finds the tightest lower bound for p^*.

$$\max_\lambda D(\lambda) = \max_\lambda \left\{ \min_{x,z} \left[f(x,z) + \lambda z \right] + \min_{y,z} \left[g(y,z) - \lambda z \right] \right\}$$

Algorithm

1. Solve the subproblems for the current λ
2. Update λ, e.g. take a gradient ascent step

Solving the Dual Problem

Dual problem finds the tightest lower bound for p^*.

$$\max_\lambda D(\lambda) = \max_\lambda \left\{ \min_{x,z} \left[f(x,z) + \lambda z \right] + \min_{y,z} \left[g(y,z) - \lambda z \right] \right\}$$

Algorithm

1. Solve the subproblems for the current λ
2. Update λ, e.g. take a gradient ascent step

Let $(x^*, z_1^*) = \operatorname{argmin}_{x,z} f(x,z) + \lambda z$ and $(y^*, z_2^*) = \operatorname{argmin}_{y,z} g(y,z) - \lambda z$, then

$$\frac{\partial D}{\partial \lambda} = z_1^* - z_2^*$$

The Gradient Ascent Algorithm

Remember that

$$D(\lambda) = \min_{x,z} \left[f(x,z) + \lambda z \right] + \min_{y,z} \left[g(y,z) - \lambda z \right]$$

and

$$\frac{\partial D}{\partial \lambda} = z_1^* - z_2^*$$

Taking a gradient step,

$$\begin{aligned} \Delta_f &= z_1^* z - z_2^* z \qquad && z_1^* \nearrow, \quad z_2^* \searrow \\ \Delta_g &= z_2^* z - z_1^* z \qquad && z_2^* \nearrow, \quad z_1^* \searrow \end{aligned}$$

Each gradient step makes the two subproblems agree more on z.

The Gradient Ascent Algorithm

When $\frac{\partial D}{\partial \lambda} = 0 \Leftrightarrow z_1^* = z_2^*$, we found the optimal $z^* = z_1^* = z_2^*$,

The Gradient Ascent Algorithm

When $\frac{\partial D}{\partial \lambda} = 0 \Leftrightarrow z_1^* = z_2^*$, we found the optimal $z^* = z_1^* = z_2^*$, because

$$\begin{aligned} D(\lambda) &= \min_{x,z} [f(x,z) + \lambda z] + \min_{y,z} [g(y,z) - \lambda z] \\ &= [f(x^*, z_1^*) + \lambda z_1^*] + [g(y^*, z_2^*) - \lambda z_2^*] \\ &= f(x^*, z^*) + g(y^*, z^*) \end{aligned}$$

and $D(\lambda) \leq p^* \leq f(x^*, z^*) + g(y^*, z^*)$.

Dual Decomposition Compared to Primal Decomposition

Good

- λ is continuous
- Unconstrained optimization
- $D(\lambda)$ is always concave, as

$$D(\lambda) = \min_{x,z} \left[f(x,z) + \lambda z \right] + \min_{y,z} \left[g(y,z) - \lambda z \right]$$

  is a minimum of linear functions of λ.

Bad

- Nontrivial to recover the optimal primal assignment (x^*, y^*, z^*) if $z_1^* \neq z_2^*$.
- Duality gap: sometimes (or usually) $D(\lambda^*) < p^*$ even for dual optimal λ^*

Dual Decomposition for MRF-MAP

- Primal problem:

$$\min_{\mathbf{x}} \sum_i \theta_i(x_i) + \sum_f \theta_f(x_f)$$

- $x_i \in \{0,1\}^K$, 1-of-K encoding $x_i = k \Leftrightarrow x_{ik} = 1, x_{ik'} = 0$, $k' \neq k$.

Dual Decomposition for MRF-MAP

- Primal problem:

$$\min_{\mathbf{x}} \sum_i \theta_i(x_i) + \sum_f \theta_f(x_f)$$

- $x_i \in \{0,1\}^K$, 1-of-K encoding $x_i = k \Leftrightarrow x_{ik} = 1, x_{ik'} = 0$, $k' \neq k$.
- Introduce one copy of x_f for each factor f

$$\begin{aligned} \min_{\mathbf{x},\{x^f\}_f} \quad & \sum_i \theta_i(x_i) + \sum_f \theta_f(x^f) \\ \text{s.t.} \quad & x_i^f = x_i, \quad \forall f, i \end{aligned}$$

Dual Decomposition for MRF-MAP

- Primal problem:

$$\min_{\mathbf{x}} \sum_i \theta_i(x_i) + \sum_f \theta_f(x_f)$$

- $x_i \in \{0,1\}^K$, 1-of-K encoding $x_i = k \Leftrightarrow x_{ik} = 1, x_{ik'} = 0,$ $k' \neq k$.
- Introduce one copy of x_f for each factor f

$$\begin{aligned} \min_{\mathbf{x},\{x^f\}_f} \quad & \sum_i \theta_i(x_i) + \sum_f \theta_f(x^f) \\ \text{s.t.} \quad & x_i^f = x_i, \quad \forall f, i \end{aligned}$$

- Lagrange multipliers $\lambda_i^f \in \mathbb{R}^K$ for each f, i

Dual Decomposition for MRF-MAP

Lagrangian dual:

$$L(\mathbf{x}, \{x^f\}, \lambda) = \sum_i \theta_i(x_i) + \sum_f \theta_f(x^f) + \sum_i \sum_f \lambda_i^{f\top}(x_i^f - x_i)$$

Dual Decomposition for MRF-MAP

Lagrangian dual:

$$L(\mathbf{x}, \{x^f\}, \lambda) = \sum_i \theta_i(x_i) + \sum_f \theta_f(x^f) + \sum_i \sum_f \lambda_i^{f\top}(x_i^f - x_i)$$

Denote $\lambda_i^f(x_i) = \lambda_i^{f\top} x_i$

Dual Decomposition for MRF-MAP

Lagrangian dual:

$$L(\mathbf{x}, \{x^f\}, \lambda) = \sum_i \theta_i(x_i) + \sum_f \theta_f(x^f) + \sum_i \sum_f \lambda_i^{f\top}(x_i^f - x_i)$$

Denote $\lambda_i^f(x_i) = \lambda_i^{f\top} x_i$

$$L(\mathbf{x}, \{x^f\}, \lambda) = \sum_i \left[\theta_i(x_i) - \sum_{f \in \mathcal{N}(i)} \lambda_i^f(x_i)\right] + \sum_f \left[\theta_f(x^f) + \sum_{i \in f} \lambda_i^f(x_i^f)\right]$$

Dual Decomposition for MRF-MAP

Dual function:

$$\begin{aligned} D(\lambda) &= \min_{\mathbf{x}} \sum_i \left[\theta_i(x_i) - \sum_{f \in \mathcal{N}(i)} \lambda_i^f(x_i) \right] + \sum_f \min_{x^f} \left[\theta_f(x^f) + \sum_{i \in f} \lambda_i^f(x_i^f) \right] \\ &= \min_{\mathbf{x}} \sum_i \left[\theta_i(x_i) - \sum_{f \in \mathcal{N}(i)} \lambda_i^f(x_i) \right] + \sum_f \min_{x_f} \left[\theta_f(x_f) + \sum_{i \in f} \lambda_i^f(x_i) \right] \end{aligned}$$

Dual Decomposition for MRF-MAP

Dual function:

$$\begin{aligned}
D(\lambda) &= \min_{\mathbf{x}} \sum_i \left[\theta_i(x_i) - \sum_{f \in \mathcal{N}(i)} \lambda_i^f(x_i) \right] + \sum_f \min_{x^f} \left[\theta_f(x^f) + \sum_{i \in f} \lambda_i^f(x_i^f) \right] \\
&= \min_{\mathbf{x}} \sum_i \left[\theta_i(x_i) - \sum_{f \in \mathcal{N}(i)} \lambda_i^f(x_i) \right] + \sum_f \min_{x_f} \left[\theta_f(x_f) + \sum_{i \in f} \lambda_i^f(x_i) \right] \\
&\leq \min_{\mathbf{x}} \sum_i \theta_i(x_i) + \sum_f \theta_f(x_f) \\
&= p^*
\end{aligned}$$

Dual Decomposition for MRF-MAP

Dual function:

$$
\begin{aligned}
D(\lambda) &= \min_{\mathbf{x}} \sum_i \left[\theta_i(x_i) - \sum_{f \in \mathcal{N}(i)} \lambda_i^f(x_i) \right] + \sum_f \min_{x^f} \left[\theta_f(x^f) + \sum_{i \in f} \lambda_i^f(x_i^f) \right] \\
&= \min_{\mathbf{x}} \sum_i \left[\theta_i(x_i) - \sum_{f \in \mathcal{N}(i)} \lambda_i^f(x_i) \right] + \sum_f \min_{x_f} \left[\theta_f(x_f) + \sum_{i \in f} \lambda_i^f(x_i) \right] \\
&\leq \min_{\mathbf{x}} \sum_i \theta_i(x_i) + \sum_f \theta_f(x_f) \\
&= p^*
\end{aligned}
$$

Dual problem:

$$
\max_{\lambda} D(\lambda)
$$

Dual Decomposition for MRF-MAP

Algorithm:

1. Solve subproblems
   - $\min_{x_i} \theta_i(x_i) - \sum_{f \in \mathcal{N}(i)} \lambda_i^f(x_i)$ for all i.
   - $\min_{x_f} \theta_f(x_f) + \sum_{i \in f} \lambda_i^f(x_i)$ for all f.
2. Update $\lambda_i^f(x_i)$ for all f, i, x_i, e.g. take a gradient step

Dual Decomposition for MRF-MAP

Algorithm:

1. Solve subproblems
   - $\min_{x_i} \theta_i(x_i) - \sum_{f \in \mathcal{N}(i)} \lambda_i^f(x_i)$ for all i.
   - $\min_{x_f} \theta_f(x_f) + \sum_{i \in f} \lambda_i^f(x_i)$ for all f.
2. Update $\lambda_i^f(x_i)$ for all f, i, x_i, e.g. take a gradient step

Denote $x_i^* = \operatorname{argmin}_{x_i} \theta_i(x_i) - \sum_{f \in \mathcal{N}(i)} \lambda_i^f(x_i)$
$x^{f*} = \operatorname{argmin}_{x_f} \theta_f(x_f) + \sum_{i \in f} \lambda_i^f(x_i)$, then

$$\frac{\partial D}{\partial \lambda_i^f(x_i)} = -\mathbf{I}[x_i = x_i^*] + \mathbf{I}[x_i = x_i^{f*}]$$

Decode Optimal $\mathbf{x}^*$ from Dual Solution

- The simplest method:

$$x_i^* = \operatorname*{argmin}_{x_i} \quad \theta_i(x_i) - \sum_{f \in \mathcal{N}(i)} \lambda_i^f(x_i)$$

- A better solution: take the best $\mathbf{x}^*$ over all iterations of the algorithm.
- Problem dependent, more structured methods may work better.

Example: Segmentation with Cardinality Potential

- Binary segmentation: $x_i \in \{0, 1\}$.
- Model: pairwise CRF + Cardinality Potential

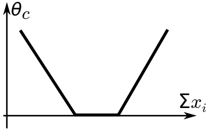
$$E(\mathbf{x}) = \sum_i \theta_i(x_i) + \sum_{(i,j)} \theta_{ij}(x_i, x_j) + \theta_c(\mathbf{x})$$

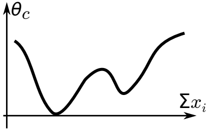
- Pairwise potential

$$\theta_{ij}(x_i, x_j) = p\,\mathbf{I}[x_i \neq x_j]$$

- Cardinality potential

$$\theta_c(\mathbf{x}) = \theta_c\left(\sum_i x_i\right)$$





Two Subproblems

- Pairwise CRF

$$\sum_i \theta_i(x_i) + \sum_{(i,j)} \theta_{ij}(x_i, x_j)$$

Efficient exact inference using graph cuts.

Two Subproblems

- Pairwise CRF

$$\sum_i \theta_i(x_i) + \sum_{(i,j)} \theta_{ij}(x_i, x_j)$$

  Efficient exact inference using graph cuts.

- Cardinality + unary

$$\begin{aligned} & \sum_i \theta_i(x_i) + \theta_c\left(\sum_i x_i\right) \\ = & \sum_i [\theta_i(1) - \theta_i(0)] x_i + \sum_i \theta_i(0) + \theta_c\left(\sum_i x_i\right) \\ = & \sum_i \theta_i' x_i + \theta_c\left(\sum_i x_i\right) + C \end{aligned}$$

  Exact inference by sorting θ_i's in $O(N \log N)$ time.

Dual problem

Break $E(\mathbf{x})$ into two parts

$$E(\mathbf{x}) = \left[\gamma \sum_i \theta_i(x_i) + \sum_{(i,j)} \theta_{ij}(x_i, x_j)\right] + \left[(1-\gamma)\sum_i \theta_i(x_i) + \theta_c\left(\sum_i x_i\right)\right]$$

Dual problem

Break $E(\mathbf{x})$ into two parts

$$E(\mathbf{x}) = \left[\gamma \sum_i \theta_i(x_i) + \sum_{(i,j)} \theta_{ij}(x_i, x_j)\right] + \left[(1-\gamma)\sum_i \theta_i(x_i) + \theta_c\left(\sum_i x_i\right)\right]$$

Dual function

$$D(\lambda) = \min_{\mathbf{x}} \left\{ \sum_i [\gamma\theta_i(x_i) - \lambda_i(x_i)] + \sum_{(i,j)} \theta_{ij}(x_i, x_j) \right\} + \min_{\mathbf{x}} \left\{ \sum_i [(1-\gamma)\theta_i(x_i) + \lambda_i(x_i)] + \theta_c\left(\sum_i x_i\right) \right\}$$

Dual problem

Break $E(\mathbf{x})$ into two parts

$$E(\mathbf{x}) = \left[\gamma \sum_i \theta_i(x_i) + \sum_{(i,j)} \theta_{ij}(x_i, x_j)\right] + \left[(1-\gamma)\sum_i \theta_i(x_i) + \theta_c\left(\sum_i x_i\right)\right]$$

Dual function

$$D(\lambda) = \min_{\mathbf{x}} \left\{ \sum_i [\gamma\theta_i(x_i) - \lambda_i(x_i)] + \sum_{(i,j)} \theta_{ij}(x_i, x_j) \right\} +$$
$$\min_{\mathbf{x}} \left\{ \sum_i [(1-\gamma)\theta_i(x_i) + \lambda_i(x_i)] + \theta_c\left(\sum_i x_i\right) \right\}$$

Gradient

$$\frac{\partial D}{\partial \lambda_i(x_i)} = -\mathbf{I}[x_i = x_i^{\text{CRF}}] + \mathbf{I}[x_i = x_i^{\text{Card}}]$$

Results

Decode primal solution $\mathbf{x}^*$ for a given λ

- I simply take $\mathbf{x}^{\mathrm{CRF}}$ as the primal solution.
- Output the $\mathbf{x}^{\mathrm{CRF}}$ with the best $E(\mathbf{x})$ over all iterations as the final result.

Show demo.

Other Methods for Solving the Dual Problem

The dual problem

$$\max_{\lambda}\left\{\sum_i \min_{x_i}\left[\theta_i(x_i)-\sum_{f\in\mathcal{N}(i)}\lambda_i^f(x_i)\right]+\sum_f \min_{x_f}\left[\theta_f(x_f)+\sum_{i\in f}\lambda_i^f(x_i)\right]\right\}$$

can also be optimized by other methods.

Other Methods for Solving the Dual Problem

The dual problem

$$\max_{\lambda} \left\{ \sum_i \min_{x_i} \left[\theta_i(x_i) - \sum_{f \in \mathcal{N}(i)} \lambda_i^f(x_i) \right] + \sum_f \min_{x_f} \left[\theta_f(x_f) + \sum_{i \in f} \lambda_i^f(x_i) \right] \right\}$$

can also be optimized by other methods.

Coordinate ascent: optimize a set of λ_i^f's and hold the others fixed

- Usually the small set of λ_i^f's can be solved to optimum
- Allows big moves - maybe faster than gradient ascent
- Parameter free - no need for learning rates

Max-Sum Diffusion

Optimize λ_i^f for a specific f and i, hold all others fixed.

Max-Sum Diffusion

Optimize λ_i^f for a specific f and i, hold all others fixed. Relevant parts in the dual

$$\min_{x_i}\left[\theta_i(x_i) - \sum_{f\in\mathcal{N}(i)} \lambda_i^f(x_i)\right] + \min_{x_f}\left[\theta_f(x_f) + \sum_{i\in f} \lambda_i^f(x_i)\right]$$

Max-Sum Diffusion

Optimize λ_i^f for a specific f and i, hold all others fixed. Relevant parts in the dual

$$\min_{x_i}\left[\theta_i(x_i) - \sum_{f\in\mathcal{N}(i)} \lambda_i^f(x_i)\right] + \min_{x_f}\left[\theta_f(x_f) + \sum_{i\in f}\lambda_i^f(x_i)\right]$$

Useful notation

$$\theta_i^{-f}(x_i) = \theta_i(x_i) - \sum_{f'\neq f}\lambda_i^{f'}(x_i), \qquad m_i^f(x_i) = \min_{x_{f\setminus i}}\left[\theta_f(x_f) + \sum_{i'\neq i}\lambda_{i'}^f(x_{i'})\right]$$

Max-Sum Diffusion

Optimize λ_i^f for a specific f and i, hold all others fixed. Relevant parts in the dual

$$\min_{x_i}\left[\theta_i(x_i) - \sum_{f\in\mathcal{N}(i)} \lambda_i^f(x_i)\right] + \min_{x_f}\left[\theta_f(x_f) + \sum_{i\in f} \lambda_i^f(x_i)\right]$$

Useful notation

$$\theta_i^{-f}(x_i) = \theta_i(x_i) - \sum_{f'\neq f} \lambda_i^{f'}(x_i), \qquad m_i^f(x_i) = \min_{x_{f\setminus i}}\left[\theta_f(x_f) + \sum_{i'\neq i} \lambda_{i'}^f(x_{i'})\right]$$

Then the objective becomes

$$\min_{x_i}\left[\theta_i^{-f}(x_i) - \lambda_i^f(x_i)\right] + \min_{x_f}\left[\theta_f(x_f) + \sum_{i\neq i'} \lambda_{i'}^f(x_{i'}) + \lambda_i^f(x_i)\right]$$

$$= \min_{x_i}\left[\theta_i^{-f}(x_i) - \lambda_i^f(x_i)\right] + \min_{x_i}\left[m_i^f(x_i) + \lambda_i^f(x_i)\right]$$

Max-Sum Diffusion

For any λ_i^f, the dual objective

$$
\begin{aligned}
&\min_{x_i}\left[\theta_i^{-f}(x_i) - \lambda_i^f(x_i)\right] + \min_{x_i}\left[m_i^f(x_i) + \lambda_i^f(x_i)\right] \\
\leq &\min_{x_i}\left[\theta_i^{-f}(x_i) - \lambda_i^f(x_i) + m_i^f(x_i) + \lambda_i^f(x_i)\right] \\
= &\min_{x_i}\left[\theta_i^{-f}(x_i) + m_i^f(x_i)\right]
\end{aligned}
$$

Max-Sum Diffusion

For any λ_i^f, the dual objective

$$\begin{aligned}
&\min_{x_i}\left[\theta_i^{-f}(x_i)-\lambda_i^f(x_i)\right]+\min_{x_i}\left[m_i^f(x_i)+\lambda_i^f(x_i)\right]\\
\leq&\min_{x_i}\left[\theta_i^{-f}(x_i)-\lambda_i^f(x_i)+m_i^f(x_i)+\lambda_i^f(x_i)\right]\\
=&\min_{x_i}\left[\theta_i^{-f}(x_i)+m_i^f(x_i)\right]
\end{aligned}$$

Choose $\lambda_i^f(x_i)=\frac{1}{2}\left[\theta_i^{-f}(x_i)-m_i^f(x_i)\right]$, then

$$\begin{aligned}
&\min_{x_i}\left[\theta_i^{-f}(x_i)-\lambda_i^f(x_i)\right]+\min_{x_i}\left[m_i^f(x_i)+\lambda_i^f(x_i)\right]\\
=&\frac{1}{2}\min_{x_i}\left[\theta_i^{-f}(x_i)+m_i^f(x_i)\right]+\frac{1}{2}\min_{x_i}\left[\theta_i^{-f}(x_i)+m_i^f(x_i)\right]\\
=&\min_{x_i}\left[\theta_i^{-f}(x_i)+m_i^f(x_i)\right]
\end{aligned}$$

Max-Sum Diffusion

$$
\begin{aligned}
\lambda_i^f(x_i) =& \frac{1}{2}\theta_i^{-f}(x_i) - \frac{1}{2}m_i^f(x_i) \\
=& \frac{1}{2}\left[\theta_i(x_i) - \sum_{f' \neq f} \lambda_i^{f'}(x_i)\right] - \frac{1}{2}\left\{\min_{x_{f \setminus i}}\left[\theta_f(x_f) + \sum_{i' \neq i} \lambda_{i'}^f(x_i')\right]\right\}
\end{aligned}
$$

is optimal.

Algorithm:

1. Loop until convergence:
   - For each f, i, update λ_i^f as above

One problem:

- Need to compute $m_i^f(x_i) = \min_{x_{f \setminus i}}\left[\theta_f(x_f) + \sum_{i' \neq i} \lambda_{i'}^f(x_i')\right]$ for all f and i, can be expensive

Max-Product Linear Programming

Optimize $\{\lambda_i^f\}_{i \in f}$ for a specific f.

Max-Product Linear Programming

Optimize $\{\lambda_i^f\}_{i\in f}$ for a specific f. Relevant parts in dual function

$$\sum_i \min_{x_i}\left[\theta_i(x_i) - \sum_{f\in\mathcal{N}(i)} \lambda_i^f(x_i)\right] + \min_{x_f}\left[\theta_f(x_f) + \sum_{i\in f}\lambda_i^f(x_i)\right]$$

Max-Product Linear Programming

Optimize $\{\lambda_i^f\}_{i \in f}$ for a specific f. Relevant parts in dual function

$$\sum_i \min_{x_i} \left[\theta_i(x_i) - \sum_{f \in \mathcal{N}(i)} \lambda_i^f(x_i) \right] + \min_{x_f} \left[\theta_f(x_f) + \sum_{i \in f} \lambda_i^f(x_i) \right]$$

Similarly we can show the following λ_i^f is optimal

$$\lambda_i^f(x_i) = \left(1 - \frac{1}{|f|}\right) \theta_i^{-f}(x_i) - \frac{1}{|f|} m_i^f(x_i)$$

MPLP is usually faster than MSD.

Connection to LP Relaxation

MRF-MAP

$$\min_{\mathbf{x}} \sum_i \theta_i(x_i) + \sum_f \theta_f(x_f)$$

LP formulation

$$\min_{\mu_i, \mu_f} \sum_i \sum_{x_i} \theta_i(x_i)\mu_i(x_i) + \sum_f \sum_{x_f} \theta_f(x_f)\mu_f(x_f)$$

such that μ_i and μ_f are marginals over x_i and x_f for some distribution $\mu(\mathbf{x})$.

- These two are equivalent: the optimal μ^* will put all the mass on $\mathbf{x}^*$, i.e. $\mu^*(\mathbf{x}^*) = 1$.
- Problem about LP: too much constraints needed for μ_i and μ_f

Connection to LP Relaxation

Relaxed LP formulation

$$
\begin{aligned}
\min_{\mu_i,\mu_f} \sum_i \sum_{x_i} \theta_i(x_i)\mu_i(x_i) + \sum_f \sum_{x_f} \theta_f(x_f)\mu_f(x_f) \\
\text{s.t.} \sum_{x_i} \mu_i(x_i) = 1, \quad \forall i \\
\sum_{x_f} \mu_f(x_f) = 1, \quad \forall f \\
\sum_{x_{f \setminus i}} \mu_f(x_f) = \mu_i(x_i), \quad \forall f, i, x_i
\end{aligned}
$$

Result: Lagrangian dual of this LP relaxation is the same as the dual function used in dual decomposition.

Connection to LP Relaxation

One interesting result about gradient ascent algorithm

- Define $\mu_i^t(x_i) = \mathbf{I}[x_i = x_i^t]$, $\mu_f^t(x_f) = \mathbf{I}[x_f = x_f^t]$, where x_i^t and x_f^t are optimal solutions for the subproblems, and

$$\begin{aligned} \bar{\mu}_i(x_i) &= \frac{1}{T}\sum_{t=1}^{T}\mu_i^t(x_i) \\ \bar{\mu}_f(x_f) &= \frac{1}{T}\sum_{t=1}^{T}\mu_f^t(x_f) \end{aligned}$$

- Then $\bar{\mu}_i(x_i)$ and $\bar{\mu}_f(x_f)$ converge to a solution of the LP relaxation as $T \to \infty$.

Conclusion

- Dual decomposition is a very general technique for optimization
- Plug-and-play if we use gradient ascent for the dual problem - can be immediately applied to a wide variety of models
- Decoding $\mathbf{x}^*$ from the dual depends on problem structure, trial and error process to figure out best method
- Coordinate ascent methods may speed things up

Boyd, S., Xiao, L., Mutapcic, A., and Mattingley, J.
Notes on decomposition methods.
Lecture Notes for EE364B, Stanford University, 2008.

Komodakis, Nikos, Paragios, Nikos, and Tziritas, Georgios.
Mrf optimization via dual decomposition: Message-passing revisited.
In *The 11th International Conference on Computer Vision*, 2007.

Komodakis, Nikos, Paragios, Nikos, and Tziritas, Georgios.
Mrf energy minimization and beyond via dual decomposition.
IEEE Transactions on Pattern Analysis and Machine Intelligence, 33(3):531–552, 2011.

Rush, Alexander M and Collins, Michael.
A tutorial on dual decomposition and lagrangian relaxation for inference in natural language processing.
Journal of Artificial Intelligence Research, 45:305–362, 2012.

Sontag, D., Globerson, A., and Jaakkola, T.

Introduction to dual decomposition for inference.
Optimization for Machine Learning, 1:219–254, 2011.

Other stuff

Strong Duality

A nice guarantee (but usually not very useful for MRF-MAP problems):

- If the primal problem $\min_{x,y,z} f(x,z) + g(y,z)$ is convex and feasible, then $D(\lambda^*) = p^*$ for dual optimal λ^*, i.e. duality gap is 0.