Diverse Particle Selection for Inference in Continuous Graphical Models

Jason Pacheco

Collaborators:
Erik Sudderth, Brown University
Silvia Zuffi, Michael Black, MPI Tubingen
Reasoning Under Uncertainty

Data $y$ => Unknowns $x$ => $p(x, y)$ Probability Model => Estimate $x^*$

Discrete Unknowns

Continuous Unknowns

Efficient inference based on *dynamic programming*

Unrealistic approximations of model yield analytic updates
Reasoning Under Uncertainty

Data $y$  

Unknowns $x$ 

Efficient inference based on *dynamic programming*

Unrealistic approximations of model yield analytic updates.
Continuous Problems

1. Human pose estimation & tracking

2. Protein structure & side chain prediction

Continuous Problems

Human pose estimation & tracking

Protein structure & side chain prediction

Nonlinear time series & target tracking.
Maximum a Posteriori (MAP)

Maximizer of the posterior probability:

\[ x^* = \arg \max_x p(x \mid y) \]

Issues with continuous models:

- Analytically intractable posterior density
Maximizer of the posterior probability:

\[ x^* = \arg\max_x p(x \mid y) \]

Issues with continuous models:

- Analytically intractable posterior density
- Many local optima. These can be useful too…
Goal

Develop maximum a posteriori (MAP) inference algorithms for continuous probability models that:

- Does not assume tractable models
- Is *black-box* (e.g. no gradient calculations)
- Robust to multiple local optima
Pairwise Markov Random Field

\[ p(x) \propto \prod_{s \in \mathcal{V}} \psi_s(x_s) \prod_{(s,t) \in \mathcal{E}} \psi_{st}(x_s, x_t) \]

- Graphical encoding of probability density
- Product of potential functions \( \psi(x) \)
- Facilitates efficient inference algorithms

\( \mathcal{G}(\mathcal{V}, \mathcal{E}) : \) Vertices \( s \in \mathcal{V} \), edges \( (s, t) \in \mathcal{E} \)
Graphical Models

Human pose estimation / tracking

Protein structure / side chain prediction

Nonlinear time series
Human Pose Estimation

\[ p(x, y) \propto \prod_{s \in \mathcal{V}} \psi_s(x_s, y) \prod_{(s,t) \in \mathcal{E}} \psi_{st}(x_s, x_t) \]

- Complicated Likelihood
- Non-Gaussian Compatibility

Deformable Structures (DS):
Continuous state \( x_s \in \mathcal{X}_s \) for part shape, location, orientation and scale.

PCA Shape: \( z_{st}(1) \), \( z_{st}(2) \)

Orientation

“Spring”
Message Passing

Global MAP inference decomposes into local computations via graph structure...

\[
\max_x p(x) \propto \max_{x_1, x_2} \psi_{12}(x_1, x_2) \psi_{23}(x_2, x_3) \psi_{24}(x_2, x_4)
\]

\[
= \max_{x_1} \max_{x_2} \psi_{12}(x_1, x_2) \left[ \max_{x_3} \psi_{23}(x_2, x_3) \right] \left[ \max_{x_4} \psi_{24}(x_2, x_4) \right]
\]

\[
m_{32}(x_2) \quad m_{42}(x_2)
\]

\[
m_{21}(x_1)
\]
Message Passing

**Max-marginal** distribution quantifies uncertainty over maxima.

\[ q_1(x_1) \propto \max_{x_2, x_3, x_4} \psi_{12}(x_1, x_2) \psi_{23}(x_2, x_3) \psi_{24}(x_2, x_4) \]

\[ = \max_{x_2} \psi_{12}(x_1, x_2) \left[ \max_{x_3} \psi_{23}(x_2, x_3) \right] \left[ \max_{x_4} \psi_{24}(x_2, x_4) \right] \]
Max-Product (MP) Belief Propagation

Passing messages in a graphical model…

Message \( m_{ts}(x_s) \propto \max_{x_t} \psi_{st}(x_s, x_t) \psi_t(x_t) \prod_{k \in \Gamma(t) \setminus s} m_{kt}(x_t) \)

Max-Marginal \( q_t(x_t) \propto \psi_t(x_t) \prod_{k \in \Gamma(t)} m_{kt}(x_t) \)
Max-Product Belief Propagation

Discrete

\[ x \in \{1, \ldots, N\}^D \]

Message Update:

\[ m_{ts} = \max_{x_t} \left[ \psi_t \prod_{k} m_{kt} \right] \]

Matrix-vector multiplication & discrete maximization

Continuous

\[ x \in \mathcal{R}^D \]

Message Update:

\[ m_{ts}(x_s) = \ldots \]

\[ \max_{x_t} \psi_{st}(x_s, x_t) \psi_t(x_t) \prod_{k} m_{kt}(x_t) \]

Nonlinear optimization
Regular Discretization

Approximate continuous max-product messages over regular grid of points

- ~10 dimensions.
- 10 grid points per dimension
- 10 Million points!

Example: Torso

Infeasible for high dimensional models.
CONDENSATION algorithm [Isard & Blake, 1998]

- Particles degenerate over time
- Resampling step reduces effective number of particles
- Particle representation is too *compact* (e.g. entire trajectory)
Particle Representations

**Particle filter:**
Each particle is a full joint instantiation

**Max-Product:**
- Each particle is a single random variable
- Efficiently enumerates all combinations
Particle Max-Product (PMP)

Combine particle filter ideas with max-product more effectively.

Particle approximation of continuous max-product (MP) messages.
Sample new hypotheses at every node to grow particle set.
Particle Max-Product (PMP)

1. **Augment Particles**

2. **Max-Product Update**

\[
m_{ts}(x_s) \propto \max_{x_t} \psi_{st}(x_s, x_t) \psi_t(x_t) \prod_{k \in \Gamma(t) \setminus s} m_{kt}(x_t)
\]

Update MP messages on augmented particles.
Particle Max-Product (PMP)

1. **Augment Particles**
   - Given $N$ particles

2. **Max-Product Update**
   - Grow to $\alpha N$ particles; $\alpha > 1$

3. **Select Particles**
   - Reduce to $N$ good particles

Select subset of *good* particles and repeat

Need a particle selection method…
Synthetic Pose Estimation

Binary image of 4 silhouettes.

**Model** Truncated Gaussian pairwise potentials $\psi_{st}(x_s, x_t)$:

Likelihood $\psi_s(x_s)$ distance-map from silhouette contours.
Greedy PMP (G-PMP)

- Keep best particle
- Sample random walk
- Naïve proposal distribution
- Too greedy; Sensitive to initialization

[ Peng et al., ICML 2011 ]
Keep N-best particles
Sensitive to initialization
Still too greedy; Selection reduces effective number of particles

Maintain diversity in particles.

[ Pacheco et al., ICML 2014 ]
Diverse Particle Selection

Initial Particles

Integer Program

Diverse Selection

Integer Program (IP) solved with efficient greedy approximation:

LP : Linear Program relaxation
IP: Optimal solution by brute force
Greedy: Efficient approximation
Continuous Message

\[ m_{ts}(x_s) = \max_{x_t} \psi_{st}(x_s, x_t) \psi_t(x_t) \]

**Joint Distribution**

\[ p(x_s, x_t) \]

**Bivariate Gaussian mixture:**

\[ \psi_{st}(x_s, x_t) = N(x | \mu_0, \Sigma_0) + N(x | \mu_1, \Sigma_1) \]
Discrete Message

Joint Distribution

Regular grid of 50 states gives discretization:

\[ \mathbf{X} = \{ x^{(1)}, \ldots, x^{(50)} \} \]
Particle Selection

Indicator vector controls state selection: $z \in \{0, 1\}^{50}$

$z(i) = 1$ indicates selected states (red line)
Particle Selection

Selecting more states reduces distortion between discrete message vectors.
Diverse Particle Selection

Minimize total message distortion:

\[
\minimize \quad \sum_{a=1}^{\alpha N} \sum_{s \in \Gamma(t)} (m_{ts}(a) - \hat{m}_{ts}(a, z))
\]

subject to \( \|z\|_1 \leq N, \quad z \in \{0, 1\}^{\alpha N} \)

\[\times\] NP-hard

\[\check{\times}\] Submodular

Good approximation qualities.
Submodularity

Set function $f : 2^Z \rightarrow \mathbb{R}$ is submodular iff diminishing marginal gains.

$$f(Y \cup \{e\}) - f(Y) \geq f(X \cup \{e\}) - f(X)$$

Efficient greedy approximation

Within $(1 - 1/e) \approx 63\%$ times optimal
Greedy Particle Selection

minimize \( \sum_{z} \sum_{s \in \Gamma(t)} \sum_{a=1}^{\alpha N} (m_{ts}(a) - \hat{m}_{ts}(a, z)) \)
Greedy Particle Selection

\[ \minimize_z \sum_{s \in \Gamma(t)} \sum_{a=1}^{\alpha N} (m_{ts}(a) - \hat{m}_{ts}(a, z)) \]
Greedy Particle Selection

\[
\text{minimize } \sum_{z} \sum_{s \in \Gamma(t)} \sum_{a=1}^{\alpha N} (m_{ts}(a) - \hat{m}_{ts}(a, z))
\]
Greedy Particle Selection

$$\text{minimize} \sum_{z} \sum_{s \in \Gamma(t)}^{\alpha N} (m_{ts}(a) - \hat{m}_{ts}(a, z))$$
Diverse Particle Selection

Diversity:

- IP objective encourages diversity
- No explicit *distance* constraint
- Distance difficult to specify in general

Multiple neighbors:

- Two-node case generalizes:

\[
\min_z \sum_a (m_t(a) - \hat{m}_t(a, z))
\]

- Concatenate message vectors:

\[
m_t = [m_{ta}, m_{tb}, m_{tc}]
\]

- Still submodular, etc…
Avoids particle degeneracies by maintaining *ensemble* of *diverse solutions* near local modes.

- No explicit diversity constraint
- Objective encourages diversity
- Efficient *Lazy greedy algorithm*
- Bounds on optimality

**Example Runs**

Colors

[Diverse Particle Max-Product (D-PMP)]

Pacheco et al., ICML 2014
Synthetic Puppets

Box plots summarize results from 11 random initializations.
Top 3 arm hypotheses: MAP estimate, 2\textsuperscript{nd} and 3\textsuperscript{rd} modes for upper arm (magenta, cyan), lower arm (green, white).

- “Buffy” dataset [Ferrari et al. 2008].
- Detections versus number of ranked hypotheses.
- Baseline: Flexible Mixture of Parts (FMP) [Yang & Ramanan 2013; Park & Ramanan 2011]

[ Pacheco, Zuffi, Black & Sudderth, ICML 2014 ]
Real Images (Multiple People)

D-PMP Particles

Mode Estimates

Precision-Recall for multi-person frames:

**T-PMP** : High precision, low recall, particles on one figure

**D-PMP** : Outperforms **FMP** and other particle methods

Note: G-PMP not reported due to poor performance.

[ Pacheco, Zuffi, Black & Sudderth, ICML 2014 ]
Human Pose Tracking

- Prior work fails to show improvement with inference on temporal model
- We believe this is due to a failure of inference

Joint work with S. Zuffi and M. Black
Human Pose Tracking

Flowing Puppets
[Zuffi et al., ICCV 2013]

Input Sequence

Hand Flow Likelihood

Skin Color and contour likelihood terms as in Deformable Structures

Input Dense Flow Hand Flow Map

Sapp et al., “Parsing Human Motion w/ Stretchable Models”, CVPR11

Temporal Prior

Motion:

Appearance Constancy:
Loopy Max-Product BP

Many interesting models exhibit *cyclic* dependency structure...

Loopy Max-Product BP: Iteratively update until converged.

Minimal guarantees.
MAP Probability Bound

Spanning Tree Distribution

\[ \rho(T_1) \quad \rho(T_2) \quad \rho(T_3) \]
\[ p(x; \psi(T_1)) \quad p(x; \psi(T_2)) \quad p(x; \psi(T_3)) \]

Bound MAP via Jensen’s Inequality:

\[ \max_x \log p(x) \leq \sum_T \rho(T) \max_x \log p(x; \psi(T)) \]

Dual Problem:

\[ \min_{\psi} \sum_T \rho(T) \max_x \log p(x; \psi(T)) \]

[Wainwright et al., 2005]
Reweighted Max-Product (RMP)

Edge Appearance

\[ m_{ts}(x_s) = \max_{x_t} \psi_t(x_t) \psi_{st}(x_s, x_t) \frac{1}{\rho_{st}} \prod_{u \in \Gamma(t) \setminus s} \frac{m_{ut}(x_t)^{\rho_{ut}}}{m_{st}(x_t)^{1-\rho_{st}}} \]

Solve dual problem via reweighted message passing

[Wainwright et al., 2005]
RMP Bound Tightness

Pseudo-Max-Marginal distribution:

\[ \nu_s(x_s) \propto \psi_s(x_s) \prod_{u \in \Gamma(s)} m_{us}(x_s)^{\rho_{us}} \approx q_s(x_s) \]

Consistent maximizer:

\[ x^*_s = \arg\max_{x_s} \nu_s(x_s) \]

\[ (x^*_s, x^*_t) = \arg\max_{x_s, x_t} \nu_{st}(x_s, x_t) \]

RMP bound tight and \( x^* \) global MAP:

\[
\max_x \log p(x) = \sum_T \rho(T) \max_x \log p(x; \psi(T))
\]
Loopy Particle Max-Product

1. Augment Particles
2. RMP Update
3. Select Diverse

Select diverse subset and repeat…

Log-Probability
"RMP Bound"
Diverse Particle Selection

Minimize reweighted message distortion:

$$\min_z \sum_{s \in \Gamma(t)} \sum_{a=1}^{\alpha N} (m_{ts}(a)^{\rho_{st}} - \hat{m}_{ts}(a, z)^{\rho_{st}})$$

subject to \( \|z\|_1 \leq N, z \in \{0, 1\}^{\alpha N} \)

- Accounts for spanning tree distribution
- Remains submodular
- Same greedy approximation
Pseudo-Max-Marginal Error

Pseudo-max-marginal error bound:

$$\|\nu_s - \hat{\nu}_s\|_1 \leq \sum_{t \in \Gamma(s)} \sum_{a=1}^{\alpha N} \left( m_{ts}(a)^{\rho_{st}} - \hat{m}_{ts}(a)^{\rho_{st}} \right)$$

Pseudo-max-marginal definitions:

$$\nu_s(a) \propto \psi_s(a) \prod_{t \in \Gamma(s)} m_{ts}(a) \quad \hat{\nu}_s(a) \propto \psi_s(a) \prod_{t \in \Gamma(s)} \hat{m}_{ts}(a)$$

**Intuition:** Selection IP objective upper bounds pseudo-max-marginal distortion.
VideoPose2 Experiments [Sapp et al. 2011]

D-PMP

Elbow Accuracy

T-PMP

Wrist Accuracy

Distance Threshold

G-PMP
T-PMP
D-PMP
Pose Tracking Particles

T-PMP

D-PMP

Greater diversity in particles allows D-PMP to reason more globally.
Oracle solution quantifies diversity
(30-pixel detection radius)
Protein Structure Prediction

All information for predicting 3D structure encoded in amino acid sequence and physics

Protein Side Chains

Side chain prediction: Estimate side chains given fixed backbone.

20 Amino Acid Types

- Leu (L)
- Val (V)
- Phe (F)
- Trp (W)
Dihedrals and Rotamers

Dihedral Angles:
- Compact angular encoding
- 1D-4D continuous state

Rotamer discretization based on marginal statistics fails to capture fine details...

[Shapovalov & Dunbrack 2007]
Side Chain Prediction

\[ p(x) \propto \prod_{s \in V} \psi_s(x_s) \prod_{(s,t) \in E} \psi_{st}(x_s, x_t) \]

Edges between amino acids within distance threshold.

[ Image: Harder et al., BMC Informatics 2010 ]
Side Chain Prediction

\[ p(x) \propto \prod_{s \in V} \psi_s(x_s) \prod_{(s,t) \in E} \psi_{st}(x_s, x_t) \]

Rotamer Likelihood

Atomic Interaction

Statistical and physical potential functions.

[ Image: Harder et al., BMC Informatics 2010 ]
D-PMP for Side Chains

Continuous optimization of side chains:
- Capture non-rotameric side chains
- Conformational diversity
- Likelihood-based proposals

1. Augment Particles
2. RMP Update
3. Select Diverse
Resolving Ties

Particle diversity leads to more conflicts:

**Side Chain Particles**

T-PMP

D-PMP

![Graphical representations of T-PMP and D-PMP side chain particles with max-marginal and edge max-marginal images.](Image)

![Graphs showing log-probability over iterations for T-PMP and D-PMP.](Image)
Resolving Ties

Find consistent maximizer:

\[
x_s^* = \arg\max_{x_s} \nu_s(x_s)
\]

\[
(x_s^*, x_t^*) = \arg\max_{x_s, x_t} \nu_{st}(x_s, x_t)
\]

1) Assign non-tied nodes
Resolving Ties

Find consistent maximizer:

\[ x_s^* = \arg\max_{x_s} \nu_s(x_s) \]
\[ (x_s^*, x_t^*) = \arg\max_{x_s, x_t} \nu_{st}(x_s, x_t) \]

1) Assign non-tied nodes
2) Create subgraph of tied nodes
3) Exact inference (e.g. via junction tree)
Rosetta

- Energy model used in FoldIt game
- Simulated annealing (SA) Monte Carlo
- Independent chains for multiple optima

Replace SA with D-PMP. Use Rosetta as black-box energy method.
Protein Side Chain Prediction

Log-probability of MAP estimate for:

20 Proteins (11 Runs)

G-PMP, T-PMP, D-PMP, Rosetta simulated annealing [Rohl et al., 2004]

370 Proteins

[ Pacheco et al., ICML 2015 ]
Protein Side Chain Prediction

Root mean square deviation (RMSD) from x-ray structure.

Oracle selects best configuration in current particle set.
Non-Rotameric Side Chains

Not all side chains obey standard rotamer discretization.

Penicillin Acylase Complex, Trp154 [Shapovalov & Dunbrack 2007]
Protein Side Chain Prediction

T-PMP

Ground Truth
Protein Side Chain Prediction

D-PMP

Ground Truth
Summary

Particle-based approximation for **MAP inference** in **continuous graphical models**.

Diverse selection procedure avoids degenerate particles.

Many applications possible...

[Images of human pose estimation, human pose tracking, and structure prediction]

Code Available: cs.brown.edu/~pacheco