Perturbative Corrections for Expectation Propagation

Manfred Opper, Computer Science

joint work with:
Ulrich Paquet (U Cambridge)
Ole Winther (TU Denmark)
Outline

• EP in a nutshell

• EP fixed points and relation to exact inference

• Correction to partition function for models with pairwise couplings

• Perturbation expansion for correction

• Illustration on simple models (GP classification & Ising)

• Outlook
Expectation Propagation (EP) in a nutshell

\( x = (x_1, \ldots, x_N) \)

Assume factorisation of intractable density

\[
p(x) = \frac{1}{Z} f_0(x) \prod_n f_n(x)
\]

by a tractable density (in exponential family)

\[
q(x) \propto f_0(x) \prod_n g_n(x)
\]

Parameters of the \( g_n \)'s are optimised iteratively by:

- Define \textit{tilted distribution} \( q_j(x) \propto f_j(x) \frac{q(x)}{g_j(x)} \)

- Optimise \( q \) s.t. \( q \approx q_j \) by moment matching.
Fixed point equations

Assume exponential family $g_n(x) \propto e^{\Lambda_n^\top \phi(x)} \rightarrow$

$$q(x) = \frac{1}{Z_0} f_0(x) e^{\Lambda^\top \phi(x)}$$

with $\Lambda = \sum_n \Lambda_n$. Thus

$$q_j(x) = \frac{1}{Z_j} f_j(x) e^{(\Lambda - \Lambda_j)^\top \phi(x)} f_0(x)$$

Moment matching conditions

$$\langle \phi(x) \rangle_q = \langle \phi(x) \rangle_{q_j} \quad \text{for } j = 1, \ldots, N.$$. 
The partition function

Express intractable terms

\[ f_n(x) = \frac{Z_n q_n(x)}{Z_0 q(x)} \exp \left( \Lambda^T_n \phi(x) \right) . \]

and insert into the partition function

\[ Z = \int dx \ f_0(x) \prod_n f_n(x) \]

\[ = \prod_n \left[ \frac{Z_n}{Z_0} \right] \times \int dx \ f_0(x) \prod_n (1 + \varepsilon_n(x)) \exp \left( \Lambda^T \phi(x) \right) . \]

where \( \varepsilon_n(x) = \frac{q_n(x)-q(x)}{q(x)} \).
**EP is optimal to linear order**

Expand log partition function into multivariate Taylor expansion with respect to $\varepsilon_n$, i.e.

$$\log Z = \log Z_{EC} + C_1 + C_2 + \ldots$$

$C_1$ contains corrections linear in $\varepsilon_n$ etc.

$$C_1 = \sum_n \frac{\int dx \ f_0(x) \varepsilon_n(x) \exp(\Lambda^T \phi(x))}{\int dx \ f_0(x) \exp(\Lambda^T \phi(x))}$$

$$= \sum_n \int dx \ q(x) \varepsilon_n(x) = \sum_n \int dx \ q(x) \frac{q_n(x) - q(x)}{q(x)} = 0$$

by normalization of $q_n$ and $q$. 


Express joint density via $q$ & $q_n$

$$Z = Z_{EP} \times R$$

$$p(x) = \frac{q(x) \prod_n \left( \frac{q_n(x)}{q(x)} \right)}{R}$$

where

$$R = \int dx \ q(x) \prod_n \left( \frac{q_n(x)}{q(x)} \right).$$
Corrections for models with pairwise couplings

Models defined by

- the 'Prior'

\[ f_0 = \exp \left[ \frac{1}{2} x^\top J x \right] \]

- 'likelihood' terms \( f_n = f_n(x_n) \) depending on single variable

- 'likelihood approximations'

\[ g_n(x) \propto \exp \left[ \gamma_n x_n - \frac{1}{2} \lambda_n x_n^2 \right] \]

- Multivariate Gaussian approximation

\[ q(x) \propto \exp \left[ -\frac{1}{2} (x - \mu)^\top \Sigma^{-1} (x - \mu) \right] \]
Correction to marginal likelihood (partition function)

\[ R = \int dx \, q(x) \prod_n \frac{q_n(x_n)}{q(x_n)} \]

where \( q^{(n)}(x_n) \) and \( q_n^{(n)}(x_n) \) are the marginals of \( q \) and \( q_n \)

We set

\[ \prod_n \frac{q_n(x_n)}{q(x_n)} = \frac{p_1(x)}{p_0(x)} \]

- EP accurate if \( p_1 \) close to \( p_0 \). Measure closeness by cumulants!

- \( p_0 \) Gaussian (only 1. & 2. cumulants)

- \( p_1 \) and \( p_0 \) agree on 1. & 2. cumulants!
Characteristic function & cumulants

• Characteristic function

\[ \chi(k) = \int dx \, e^{ik^\top x} q(x) = \langle e^{ik^\top x} \rangle \]

\[ q(x) = \int \frac{dk}{(2\pi)^N} \, e^{-ik^\top x} \chi(k) \]

• Cumulants \( c_l \)

\[ \ln \chi(k) = \sum_l (i)^l \frac{c_l}{l!} k^l \]

• Express densities by cumulants

\[ p_0(x) = \int \frac{dk}{(2\pi)^N} \, e^{-ik^\top x} \chi_0(k) \quad \text{with} \quad \chi_0(k) = e^{ik^\top \mu} \prod_n e^{-\frac{1}{2} \Sigma_{nn} k_n^2} \]

\[ p_1(x) = \int \frac{dk}{(2\pi)^N} \, e^{-ik^\top x} \chi_0(k) \, e^{\sum_n g_n(k_n)} \]

\( g_n \) contains only the higher order cumulants \( c_{ln} \) for \( l \geq 3 \)
Express the ratio by cumulants

Use shift of variables $k_j \rightarrow \eta_j = k_j + i \frac{(x_j - \mu_j)}{\Sigma_{jj}}$ we get

$$\frac{p_1(x)}{p_0(x)} = \int \prod_n \left( d\eta_n \sqrt{\frac{\Sigma_{nn}}{2\pi}} \right) e^{-\sum_n \frac{\Sigma_{nn}\eta_n^2}{2}} \exp \left[ \sum_n g_n \left( \eta_n - i \frac{(x_n - \mu_n)}{\Sigma_{nn}} \right) \right]$$
Performing the average

Introduce complex Gaussian random variable $z_n \doteq \eta_n - i \frac{x_n - m_n}{\sum_{nn}}$

$$R = \int dx \frac{q(x)}{p_0(x)} \frac{p_1(x)}{p_0(x)} = \left\langle \exp \left[ \sum_n g_n(z_n) \right] \right\rangle_z$$

$z$ has the covariance

$$\left\langle z_i z_j \right\rangle_z = -\frac{\sum_{ij}}{\sum_{ii} \sum_{jj}} \quad i \neq j$$

$$\left\langle z_i^2 \right\rangle_z = 0$$

The last equation has some nice consequences for the surviving terms of corrections!
Perturbation expansion to Free energy

Assuming that the $g_n$ are small, introduce formal expansion parameter $\lambda$, set $\lambda = 1$ at the end.

\[
\ln R = \ln \langle \exp \left[ \lambda \sum_n g_n (z_n) \right] \rangle_z = \\
\lambda \sum_n \langle g_n \rangle + \frac{\lambda^2}{2} \left\{ \langle \left( \sum_n g_n \right)^2 \rangle_z - \left( \sum_n \langle g_n \rangle_z \right)^2 \right\} \pm \ldots \\
= \frac{\lambda^2}{2} \sum_{m \neq n} \langle g_m g_n \rangle_z \pm \ldots
\]

Single marginal terms vanish!
Gaussian averages & Feynman graphs

Let $\langle y_i \rangle = 0$. Then

$$\langle y_1 \cdots y_{2k} \rangle = \sum \text{pairings} \langle y_{i_1} y_{j_1} \rangle \cdots \langle y_{i_k} y_{j_k} \rangle$$

Example:

$$\langle y_1 \cdot y_2 \cdot y_3 \cdot y_4 \rangle =$$

\begin{center}
\begin{tikzpicture}
\draw (0,0) -- (2,0); \draw (2,0) -- (4,0);
\draw (0,2) -- (2,2); \draw (2,2) -- (4,2);
\draw (0,0) -- (0,2); \draw (2,0) -- (2,2); \draw (4,0) -- (4,2);
\end{tikzpicture}
\end{center}
For

\[ g_n(k) = \sum_{l \geq 3} i^l \frac{c_{ln}}{l!} k^l \]

we get

\[ \langle g_n(z_n)g_m(z_m) \rangle_z = \sum_{l,s \geq 3} i^{l+s} \frac{c_{ln}c_{sm}}{l!s!} \langle z_n^l z_m^s \rangle \]

\[ = \sum_{l \geq 3} i^{2l} l! \frac{c_{ln}c_{lm}}{(l!)^2} \langle z_n z_m \rangle^l \]

\[ = \sum_{l \geq 3} \frac{c_{ln}c_{lm}}{l!} \left( \frac{\sum_{nm}}{\sum_{nn} \sum_{mm}} \right)^l \]

In practice, truncate after \( l = 4 \) or \( l = 5 \).
Conjecture: EP is fairly accurate if:

- the cumulants $c_{ln}$ are small. This holds for GP classification, when posterior variance small compared to the mean.

- if posterior covariances $\Sigma_{ij}$ small for $i \neq j$. 
Gaussian process classification

Prior

\[ f_0(x) \propto \exp\left[-\frac{1}{2} x^\top K^{-1} x\right] \]

Likelihood

\[ f_n(x) = \theta(x_n \cdot \text{label}_n) \quad \text{(unit step function)} \]

The marginal \( q^{(n)}(x) \) might look like this

![Graph of a Gaussian process classification](image-url)
The cumulants

Cumulant $c_3$ with $m = 0$ and $v = 0$ (step function)

Cumulant $c_4$ with $m = 0$ and $v = 0$ (step function)
Correction to log partition function

Correction to the log marginal likelihood.

\[
\text{Varying } x_N, \text{ of } \circ \text{ class, with } D_{1:N-1}\text{ fixed}
\]
log partition function + correction

Correction to the log marginal likelihood.

\[ \log Z_{EP} + \log R, \text{ second order with } \ell = 3 \text{ and } \ell = 4 \]

Varying \( x_N, \) of \( \circ \) class, with \( D_{1:N-1} \) fixed

- \( \log Z_{EP} \)
- \( \log Z_{EP} + \log R \) (using c3 and c4 in 2nd order)

- Class 1
- Class -1
Correcting the posterior mean

Scaled correction to mean of $q(f)$: $\text{corr}_n/\mu_n$
A toy Ising case

Ising variables \( x_{1/2} = \pm 1 \) with

\[
p(x) = \frac{1}{Z} e^{Jx_1x_2} \quad f_n(x) = \delta(x - 1) + \delta(x + 1)
\]

with

\[
\ln Z = \ln 4 + \ln \cosh(J) = \ln 4 + \frac{J^2}{2} - \frac{J^4}{12} \pm \ldots
\]

One can show that

\[
\ln Z_{EC} = \ln 4 + \frac{J^2}{2} - \frac{J^4}{4} \pm \ldots
\]

- EC gives correct \( J^2 \) coefficient but \( J^4 \) comes out wrong.

- Adding correction from \( c_4 \) makes \( J^4 \) exact.
Random Ising networks

Ising variables $x_{1/2} = \pm 1$ with

$$p(x) = \frac{1}{Z} \exp \left[ \frac{1}{2} x^\top J x + \gamma^\top x \right]$$

$J \sim \mathcal{N}(0, \beta^2/N)$ and $N = 10$
Outlook

• Systematic expansion w.r.t. $\Sigma_{ij}$

• Develop sanity check

• Similar expansions for power EP ?

• Non Gaussian models