Information Geometry of Turbo Codes  
Shiro Ikeda  
Kyushu Inst. of Tech., & JST  
Toshiyuki Tanaka  
Tokyo Metropolitan Univ.  
Shun-ichi Amari  
RIKEN BSI

Abstract

The properties of the turbo decoding is studied from information geometrical viewpoint. Our study gives an intuitive understanding of the theoretical background, and a new framework for the analysis. Based on the framework, we reveal basic properties of the turbo decoding.

1 Introduction

Turbo codes[2] has been attracting a lot of interests because of its high performance of error correction. Although the thorough experimental results support the potential of the iterative decoding method, the mathematical background is not sufficiently understood.

The problem of the turbo decoding is equivalent to marginalizing an exponential family distribution. The distribution includes higher order correlations, and its direct marginalization is intractable. But the partial model with a part of the correlations (i.e., a constituent code), can be marginalized efficiently, by soft decoding. By collecting and exchanging the results of the partial models, the true decoding result is approximated.

Richardson[6] initiated a geometrical understanding of the turbo decoding. But further intuitive understanding seems to be necessary. We investigate the problem from information geometrical viewpoint[1]. It gives a new framework for analyzing the iterative methods, and shows an intuitive understanding. Also it reveals a lot of basic properties, such as characteristics of the equilibrium, the condition of stability, the cost function related to the iterative method, and the decoding error[3, 4].

2 Information Geometrical Framework

2.1 Turbo Decoding

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{turbo_codes.png}
\caption{Turbo codes}
\end{figure}

Let \( x \in \{ -1, +1 \}^N \) be the information bits, from which the turbo encoder generates two sets of parity bits, \( y_1 = (y_{11}, \cdots, y_{1L})^T \), and \( y_2 = (y_{21}, \cdots, y_{2L})^T \), \( y_{1j}, y_{2j} \in \{ -1, +1 \} \) (Fig.1). Each parity bit is expressed in the form \( \prod_{i \in \mathcal{L}_{rj}} x_i \), \( (r = 1, 2) \), where the product is taken through a subset \( \mathcal{L}_{rj} \in \{ 1, \cdots, N \} \).

The codeword \((x, y_1, y_2)\) is then transmitted over a noisy channel, which we assume a BSC (binary symmetric channel) with flipping probability \( \sigma < 1/2 \). The receiver observes their noisy version as \((\tilde{x}, \tilde{y}_1, \tilde{y}_2)\), \( \tilde{x}, \tilde{y}_{1j}, \tilde{y}_{2j} \in \{ -1, +1 \} \).

The ultimate goal of the turbo decoding is the MPM (maximization of the posterior marginals) decoding of \( x \) based on \( p(x|\tilde{x}, \tilde{y}_1, \tilde{y}_2) \). Since the channel is memoryless, the following relation holds

\[
p(x, \tilde{y}_1, \tilde{y}_2|x) = \exp(\beta \tilde{x} \cdot x + \beta \tilde{y}_1 \cdot y_1 + \beta \tilde{y}_2 \cdot y_2 - (N + 2L)\psi(\beta))
\]

\[
\beta > 0, \quad \sigma = \frac{1}{2} (1 - \tanh \beta), \quad \psi(\beta) \overset{\text{def}}{=} \ln(e^\beta + e^{-\beta}),
\]


Let us consider the family of all the probability distributions over $\mathbb{R}^2$. This distribution is derived from
\[ p(x) = \exp(\beta x \cdot y_1 + \beta y_2 \cdot y_2) \]
where, $\beta \cdot \cdot$ denotes the inner-product of vectors. By assuming the uniform prior on $x$, the posterior distribution is given as follows,
\[
p(x|\bar{x}, \tilde{y}_1, \tilde{y}_2) = \frac{p(\bar{x}, \tilde{y}_1, \tilde{y}_2 | x)}{\sum_x p(\bar{x}, y_1, y_2 | x)} = C \exp(\beta \bar{x} \cdot y_1 + \beta \tilde{y}_2 \cdot y_2)
\]
\[= C \exp(c_0(x) + c_1(x) + c_2(x)).\]
Here $C$ is the normalizing factor, and $c_0(x) = \beta \bar{x} \cdot x$, $c_r(x) = \beta \tilde{y}_r \cdot y_r$, $(r = 1, 2)$. The soft decoding is expressed as follows,
\[\bar{x} \triangleq \sum_x \exp(x|\bar{x}, \tilde{y}_1, \tilde{y}_2),\]
\[\tilde{x}_i \ (i = 1, \ldots, N)\] is the soft bit, and the sign of each is the MPM decoding result. Let $\Pi$ denote the operator of marginalization,
\[\Pi \circ p(x|\bar{x}, \tilde{y}_1, \tilde{y}_2) = \prod_{i=1}^N p(x_i|\bar{x}, \tilde{y}_1, \tilde{y}_2).\]
Since $\tilde{x}_i$ is directly calculated from $p(x_i|\bar{x}, \tilde{y}_1, \tilde{y}_2)$, calculation cost of soft bits and $\Pi \circ p(x|\bar{x}, \tilde{y}_1, \tilde{y}_2)$ are the same.

In the turbo decoding, direct calculation of the soft bits is not tractable. Turbo codes utilize two decoders. Each of them gives the soft decoding based on one of the two sets of the parity bits. For the soft decoding, the following $p_r(x; \xi) \ (r = 1, 2)$ is used.
\[p_r(x; \xi) = \exp(c_0(x) + c_r(x) + \xi \cdot x - \varphi_r(\xi)) \in \mathbb{R}^N, \quad \varphi_r(\xi) = \ln \sum_x \exp(c_0(x) + c_r(x) + \xi \cdot x) \quad (2)\]
This distribution is derived from $p(\bar{x}, \tilde{y}_r | x)$ and the prior of $x$ which has the form of
\[\omega(x; \xi) = \exp(\xi \cdot x - \psi(\xi)), \quad \psi(\xi) = \sum_i \psi(\xi_i).\]
Each $p_r(x; \xi)$ includes one of $\{c_1(x), c_2(x)\}$ in eq.(1), and additional parameter $\xi$ adjusts the linear term of $x$. In the turbo decoding, the marginalization of $p_r(x; \xi)$ is feasible. The soft decoding of $p(x|\bar{x}, \tilde{y}_1, \tilde{y}_2)$ is approximated by updating $\xi$ iteratively in “turbo” like way.

### 2.2 Information Geometrical View of MPM Decoding

Let us consider the family of all the probability distributions over $x$, which we call $S$. The manifold $S$ is equivalent to the family of distributions over $2^N$ atoms,
\[S = \left\{ p(x) | p(x) > 0, \sum_x p(x) = 1 \right\}.\]
We define $e$–flat and $m$–flat submanifolds of $S$.

**$e$–flat manifold:** A submanifold $M \in S$ is said to be $e$–flat, when the following $r(x; t)$ belongs to $M$ for all $q(x), p(x) \in M$. Here, $c(t)$ is the normalization factor,
\[\ln r(x; t) = (1 - t) \ln q(x) + t \ln p(x) + c(t), \quad t \in \mathbb{R}.\]

**$m$–flat manifold:** A submanifold $M \in S$ is said to be $m$–flat, when the following mixture $r(x; t)$ belongs to $M$ for all $q(x), p(x) \in M$,
\[r(x; t) = (1 - t)q(x) + tp(x), \quad t \in [0, 1].\]
Now, let us consider a submanifold of $p_0(x; \theta)$ defined as
\[M_0 = \{ p_0(x; \theta) = \exp(c_0(x) + \theta \cdot x - \varphi_0(\theta)) \mid \theta \in \mathbb{R}^N \}, \quad (3)\]
$\theta$ gives the coordinate system of $M_0$, and is called the natural parameter. Since $c_0(x) = \beta \bar{x} \cdot x$, every distribution of $M_0$ can be rewritten as follows,
\[p_0(x; \theta) = \exp(c_0(x) + \theta \cdot x - \varphi_0(\theta)) = \exp((\beta \bar{x} + \theta) \cdot x - \varphi_0(\theta)), \quad \varphi_0(\theta) = \psi(\beta \bar{x} + \theta).\]
It shows that every distribution of $M_0$ is decomposable, or factorizable. From the information geometry[1], we have the following theorem of $m$–projection.
Theorem 1. Let \( q(x) \in S \), and \( \hat{\theta} \) be the parameter of \( p_0(x; \theta) \) \( \in M_0 \), that minimizes the KL-divergence from \( q(x) \) to \( M_0 \),

\[
\hat{\theta} = \pi_{M_0} o q(x) \overset{\text{def}}{=} \arg \min_{\theta \in \mathbb{R}^N} D[q(x); p_0(x; \theta)].
\]

\( \hat{\theta} \) is called the \( m \)-projection of \( q(x) \) to \( M_0 \). The \( m \)-projection is unique.

Here, \( D[q(x); p_0(x; \theta)] \) is the Kullback-Leibler divergence, which is defined as,

\[
D[q(x); p_0(x; \theta)] = \sum_x q(x) \log \frac{q(x)}{p_0(x; \theta)}.
\]

Generally \( D[q(x); p_0(x; \theta)] \geq 0 \), and it is equal to 0 if and only if \( q(x) = p_0(x; \theta) \) holds for every \( x \). It is easy to show that the marginalization corresponds to the \( m \)-projection to \( M_0[7] \). Since the soft decoding and marginalization is equivalent, the soft decoding is also equivalent to the \( m \)-projection to \( M_0 \). Finally, we show that the following equation holds,

\[
\sum_x x q(x) = \sum_x x p_0(x; \hat{\theta}) \overset{\text{def}}{=} \eta_0(\hat{\theta}).
\]

\( \eta_0(\theta) \), which is called expectation parameter in information geometry[1], is the soft bits of \( p_0(x; \theta) \) and gives another coordinate system of \( M_0 \).

### 2.3 Information Geometry of Turbo Decoding

The turbo decoding process is written as follows,

1. Let \( \xi_1^t = 0 \) for \( t = 0 \), and \( t = 1 \).
2. Project \( p_2(x; \xi_1^t) \) onto \( M_0 \) and calculate \( \xi_2^{t+1} \) by
   \[
   \xi_2^{t+1} = \pi_{M_0} \circ p_2(x; \xi_1^t) - \xi_1^t.
   \]
3. Project \( p_1(x; \xi_2^{t+1}) \) onto \( M_0 \) and calculate \( \xi_1^{t+1} \) by
   \[
   \xi_1^{t+1} = \pi_{M_0} \circ p_1(x; \xi_2^{t+1}) - \xi_2^{t+1}.
   \]
4. If \( \pi_{M_0} \circ p_1(x; \xi_1^{t+1}) \neq \pi_{M_0} \circ p_2(x; \xi_1^t) \), go to step 2.

The turbo decoding approximates the estimated parameter \( \theta^* \), the projection of \( p(x|x, \hat{\theta}_1, \hat{\theta}_2) \) onto \( M_0 \), as \( \theta^* = \xi_1^t + \xi_2^t \), where the estimated distribution is

\[
p_0(x; \theta^*) = \exp(c_0(x) + \xi_1^t \cdot x + \xi_2^t \cdot x - \varphi_0(\xi_1^t + \xi_2^t)).
\]

An intuitive understanding of the turbo decoding is as follows. In step 2, \( (\xi_2 \cdot x) \) in eq.(4) is replaced with \( c_2(x) \). The distribution becomes \( p_2(x; \xi_1) \), and \( \xi_2 \) is estimated by projecting it onto \( M_0 \). In step 3, \( (\xi_1 \cdot x) \) in eq.(4) is replaced with \( c_1(x) \), and \( \xi_1 \) is estimated by \( m \)-projection of \( p_1(x; \xi_2) \).

We now define the submanifold corresponding to \( p_r(x; \xi) \), \( (r = 1, 2) \),

\[
M_r = \{ p_r(x; \xi) = \exp(c_0(x) + c_r(x) + \xi \cdot x - \varphi_r(\xi)) \mid \xi \in \mathbb{R}^N \}.
\]

\( \xi \) is the coordinate system of \( M_r \). \( M_r \) is also an \( e \)-flat submanifold. \( M_1 \neq M_2 \) and \( M_r \neq M_0 \) hold because \( c_r(x) \) includes cross terms of \( x \) and \( c_1(x) \neq c_2(x) \) in general. The information geometrical view of the turbo decoding is schematically shown in Fig.2.

### 3 The Properties of Turbo decoding

#### 3.1 Equilibrium

For the following discussion, we define the expectation parameters of \( p_r(x; \xi) \), \( r = 1, 2 \),

\[
\eta_r(\xi) \overset{\text{def}}{=} \sum_x x p_r(x; \xi) \quad r = 1, 2.
\]
When the turbo decoding converges, equilibrium solution defines three important distributions, $p_1(x; \xi^*_1)$, $p_2(x; \xi^*_2)$, and $p_0(x; \theta^*)$. They satisfy the following two conditions:

1. $\pi_{M_0} \circ p_1(x; \xi^*_1) = \pi_{M_0} \circ p_2(x; \xi^*_2) = p_0(x; \theta^*)$, in other word, $\eta_1(\xi^*_1) = \eta_1(\xi^*_2) = \eta_0(\theta^*)$. (5)

2. $\theta^* = \xi^*_1 + \xi^*_2$. (6)

For $p(x; \theta) \in M_0$, there exist $\xi_1$ and $\xi_2$ such that $\eta_1(\xi_1) = \eta_2(\xi_1) = \eta_0(\theta)$. We let $\xi_1(\theta)$ and $\xi_2(\theta)$ which satisfy this equation. The manifold $M(\theta)$ is defined by

$$M(\theta) = \{ p(x) \big| \sum x p(x) = \eta_0(\theta) \}.$$ 

This is an $m$–flat submanifold, which includes $p_1(x; \xi_2(\theta))$, $p_2(x; \xi_1(\theta))$, and $p_0(x; \theta)$. From its definition, for any $p(x) \in M(\theta)$, the expectation of $x$ is the same. Hence for any $p(x) \in M(\theta)$, its $m$–projection to $M_0$ coincides with $p_0(x; \theta)$. We call $M(\theta)$ an equimarginal submanifold.

Let us define an $e$–flat version of the submanifold as $E(\theta)$, which connects $p_0(x; \theta)$, $p_1(x; \xi_1)$, and $p_2(x; \xi_1)$ in log-linear manner

$$E(\theta) = \left\{ p(x) = C \eta_0(x; \theta)^{t_0} p_1(x; \xi_1)^{t_1} p_2(x; \xi_1)^{t_2} \bigg| \sum_{r=0}^2 t_r = 1 \right\}.$$ 

When eq. (6) holds, $p(x|x, \bar{y}_1, \bar{y}_2)$ is included in the $E(\theta)$. It can be proved by taking $t_0 = -1$, $t_1 = t_2 = 1$.

**Theorem 2.** When the turbo decoding procedure converges, the convergent probability distributions $p_0(x; \theta^*)$, $p_1(x; \xi^*_1)$, and $p_2(x; \xi^*_2)$ belong to equimarginal submanifold $M(\theta^*)$, while its $e$–flat version $E(\theta^*)$ includes these three distributions and also the posterior distribution $p(x|x, \bar{y}_1, \bar{y}_2)$ (Fig. 3). □

![Figure 2: Turbo decoding](image1)

![Figure 3: $M(\theta^*)$ and $E(\theta^*)$](image2)

If $M(\theta^*)$ includes $p(x|x, \bar{y}_1, \bar{y}_2)$, $p_0(x; \theta^*)$ is the true marginalization of $p(x|x, \bar{y}_1, \bar{y}_2)$. However, $M(\theta^*)$ does not necessarily include $p(x|x, \bar{y}_1, \bar{y}_2)$. This fact means that $p(x|x, \bar{y}_1, \bar{y}_2)$ and $p_0(x; \theta^*)$ are not necessarily equimarginal, which is the origin of the decoding error.

### 3.2 Condition of Stability

From the properties of exponential family distributions, the following relation holds for expectation parameters and $\varphi_0$ in eq. (3) and $\varphi_r$, $r = 1, 2$ in eq. (2)

$$\eta_0(\theta) = \delta_0 \varphi_0(\theta), \quad \eta_r(\xi) = \delta_\xi \varphi_r(\xi).$$

We give a sufficiently small perturbation $\delta$ to $\xi^*_1$ and apply one turbo decoding step. The $m$–projection from $p_2(x; \xi^*_1 + \delta)$ to $M_0$ gives

$$\eta_0(\theta^* + \Delta \theta) = \eta_2(\xi^*_1 + \delta) \quad \Delta \theta = G_0(\theta^*)^{-1} G_2(\xi^*_1) \delta.$$
$G_0(\theta)$ is the Fisher information matrix of $p_0(x; \theta)$, and $G_r(\xi)$ is that of $p_r(x; \xi)$, $(r = 1, 2)$, defined as,

$$G_0(\theta) = \partial_{\theta \theta^T} \varphi_0(\theta) = \partial_{\theta} \eta_0(\theta), \quad G_r(\xi) = \partial_{\xi \xi^T} \varphi_r(\xi) = \partial_{\xi} \eta_r(\xi), \quad r = 1, 2.$$

Note that $G_0(\theta)$ is a diagonal matrix. $\xi_2$ in step 2 will be,

$$\xi_2 = \xi_2^* + (G_0(\theta^*)^{-1}G_2(\xi_1^*) - I_N)\delta.$$

Here, $I_N$ is an identity matrix of size $N$. Following the same line for step 3, we derive the theorem which coincides with the result of Richardson[6].

**Theorem 3.** Let $\lambda_i$ be the eigenvalues of the matrix $T$ defined as

$$T = (G_0(\theta^*)^{-1}G_1(\xi_2^*) - I_N)(G_0(\theta^*)^{-1}G_2(\xi_1^*) - I_N).$$

When $|\lambda_i| < 1$ holds for all $i$, the equilibrium point is stable. \hfill \Box

### 3.3 Cost Function and Characteristics of Equilibrium

We give the cost function which plays an important role in turbo decoding.

$$\mathcal{F}(\xi_1, \xi_2) = \varphi_0(\theta) - (\varphi_1(\xi_2) + \varphi_2(\xi_1)).$$

Here, $\theta = \xi_1 + \xi_2$. This function is identical to the “free energy” defined in [5], implying close relationship between our approach and the statistical-mechanical one.

**Theorem 4.** The equilibrium state $\xi_1^*, \xi_2^*$ is a critical point of $\mathcal{F}$.

**Proof.** Direct calculation gives $\partial_{\xi_1} \mathcal{F} = \eta_0(\theta) - \eta_1(\xi_1), \partial_{\xi_2} \mathcal{F} = \eta_0(\theta) - \eta_1(\xi_2)$. For the equilibrium, $\eta_0(\theta^*) = \eta_1(\xi_2^*) = \eta_2(\xi_1^*)$ holds, and the proof is completed. \hfill \Box

When $(\xi_1^{t+1} - \xi_1^*)$ is small, the linear approximation of the mapping, defined by the one cycle of the turbo decoding, is derived as

$$\begin{pmatrix} \frac{\xi_1^{t+1}}{\xi_1^2} \\ \frac{\xi_2^{t+1}}{\xi_2^2} \end{pmatrix} = \left( \begin{array}{cc} O & G_0(\theta)^{-1} \\ G_0(\theta) & O \end{array} \right) \left( \begin{array}{c} \partial_{\xi_1} \mathcal{F} \\ \partial_{\xi_2} \mathcal{F} \end{array} \right).$$

This shows that the algorithm performs a “skewed” gradient ascent in the vicinity of an equilibrium. The Hessian of $\mathcal{F}$ is

$$\mathcal{H} = \begin{pmatrix} \partial_{\xi_1, \xi_1} \mathcal{F} & \partial_{\xi_1, \xi_2} \mathcal{F} \\ \partial_{\xi_2, \xi_1} \mathcal{F} & \partial_{\xi_2, \xi_2} \mathcal{F} \end{pmatrix} = \begin{pmatrix} G_0 - G_1 & G_0 \\ G_0 & G_0 - G_2 \end{pmatrix}.$$

And by transforming the variables as, $\theta = \xi_1 + \xi_2$ and $\nu = \xi_1 - \xi_2$, we have

$$\begin{pmatrix} \partial_{\theta \theta} \mathcal{F} & \partial_{\theta \nu} \mathcal{F} \\ \partial_{\nu \theta} \mathcal{F} & \partial_{\nu \nu} \mathcal{F} \end{pmatrix} = \frac{1}{4} \begin{pmatrix} (4G_0(\theta) - (G_1 + G_2) & (G_1 - G_2) \\ (G_1 - G_2) & -(G_1 + G_2) \end{pmatrix}.$$

Most probably, $\partial_{\theta \theta} \mathcal{F}$ is positive definite and $\partial_{\nu \nu} \mathcal{F}$ is always negative, thus, $\mathcal{F}$ is generally saddle at equilibrium.

### 3.4 Perturbation Analysis

For the following discussion, we define a distribution $p(x; \theta, v)$ as

$$p(x; \theta, v) = \exp(c_0(x) + \theta \cdot x + v \cdot c(x) - \varphi(\theta, v)),$$

$$\varphi(\theta, v) = \ln \sum_x \exp(c_0(x) + \theta \cdot x + v \cdot c(x)), \quad c(x) \stackrel{\text{def}}{=} (c_1(x), c_2(x))^T.$$

This distribution includes $p_0(x; \theta) (v = 0), p(x|\tilde{y}_1, \tilde{y}_2) (\theta = \theta, v = 1)$, and $p_r(x; \xi) (\theta = \xi, v = e_r)$, where $1 = (1, 1)^T, e_1 = (1, 0)^T$, and $e_2 = (0, 1)^T$. The expectation parameter $\eta(\theta, v)$ is defined as,

$$\eta(\theta, v) = \partial_{\theta} \varphi(\theta, v) = \sum_x xp(x; \theta, v).$$
Let us consider $M(\theta^*)$, where every distribution $p(x; \theta, v) \in M(\theta^*)$ has the same expectation parameter, that is, $\eta(\theta, v) = \eta(\theta^*)$ holds. Here, we define, $\eta(\theta^*) = \eta(\theta^*, o)$. From the Taylor expansion, we have,

$$
\eta(\theta, v) = \eta(\theta^*) + \sum_j \partial_j \eta(\theta^*) \Delta \theta_j + \frac{1}{2} \sum_{r,s} \partial_r \partial_s \eta(\theta^*) v_r v_s 
$$

$$
+ \sum_{j,s} \partial_j \partial_s \eta(\theta^*) v_r \Delta \theta_j + \frac{1}{2} \sum_{k,i,j,k,l} \partial_k \partial_i \eta(\theta^*) \Delta \theta_k \Delta \theta_l + O(\|v\|^3) + O(||\Delta \theta||^3).
$$

(7)

The indexes $\{i, j, k, l\}$ are for $\theta$, $\{r, s\}$ are for $v$, and $\Delta \theta \equiv \theta - \theta^*$. After adding some definitions, that is, $\eta(\theta, v) = \eta(\theta^*)$, and $\partial_j \eta(\theta^*) = g_{ij}(\theta^*)$, where $\{g_{ij}\}$ is the Fisher information matrix of $p(x; \theta^*, o)$ which is a diagonal matrix, we substitute $\Delta \theta_i$ with function of $v_r$ up to its 2nd order, and neglect the higher orders of $v_r$. And we have,

$$
\Delta \theta_i \approx -g_{ii}^{-1} \sum_r A_r^{i} v_r - \frac{1}{2} \sum_{r,s} \left( \partial_r - \sum_k g_{r,k} A_k^i \partial_k \right) \left( \partial_s - \sum_j g_{s,j} A_j^i \partial_j \right) \eta(\theta^*) v_r v_s.
$$

(8)

where, $g_{ii} = 1/g_{ii}$, and $A_r^i = \partial_i \eta(\theta^*)$. Let us consider, $p(x; \theta^*, o)$, $p(x; \xi_1, \delta e_2)$, $p(x; \xi_2, \delta e_1)$, and $p(x; o, \delta i)$. Let $p(x; \theta^*, o)$, $p(x; \xi_1, \delta e_2)$, and $p(x; \xi_2, \delta e_1)$ be included in $M(\theta^*)$, and $\theta^* = \xi_1 + \xi_2$ be satisfied. By putting $\delta = 1$, this coincides with the converged point of the turbo decoding. From the result of eq.(7) and eq.(8), we have the following theorem.

**Theorem 5.** The true expectation of $x$, which is $\eta(o, \sum_e e_r)$, is approximated as,

$$
\eta(o, \sum_e e_r) \simeq \eta(\theta^*) + \frac{1}{2} \sum_{r,s} \left( \partial_r - \sum_k g_{r,k} A_k^i \partial_k \right) \left( \partial_s - \sum_j g_{s,j} A_j^i \partial_j \right) \eta(\theta^*).
$$

(9)

Where $\eta(\theta^*)$ is the solution of the turbo decoding.

Equation (9) is related to the $m$-embedded–curvature of $E(\theta^*)$ (Fig.3). The result can be extended to general case where $K>2$ [8].

**4 Discussion**

We have shown a new framework for understanding and analyzing the turbo decoding. It elucidates the mathematical background, and reveals basic properties.

The information geometrical structure of the equilibrium is summarized in Theorem 2. It shows the $e$–flat submanifold $E(\theta^*)$ plays an important role. Furthermore, Theorem 5 shows that the relation between $E(\theta^*)$ and the $m$–flat submanifold $M(\theta^*)$ causes the decoding error, and the principal component of the error is the curvature of $E(\theta^*)$. Since the curvature strongly depends on the codeword, we can control it by the encoder design. This shows a room for improvement of the “near optimum error correcting code”.

This paper gives a first step to the information geometrical understanding of the belief propagation decoder. The main results are for the turbo decoding, but the mechanism is common with wider class, and the framework is also valid for them. We believe further study in this direction will lead us to better understanding of these methods.

**References**


