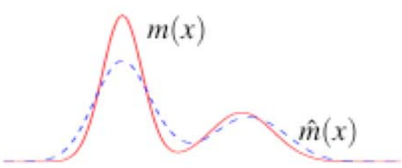
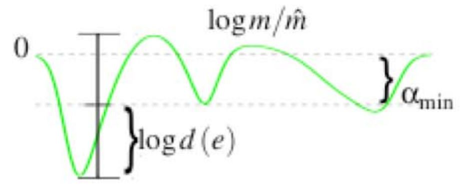
Loopy Belief Propagation:

Convergence and Effects of Message Errors

By

A.Ihler, J.Fisher III, A.Willsky

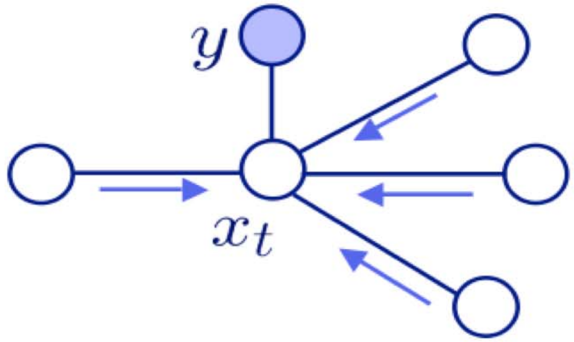




Presented by: Donglai Wei

Belief Propagation (Sum-Product)

BELIEFS: Posterior marginals (possibly approximate)

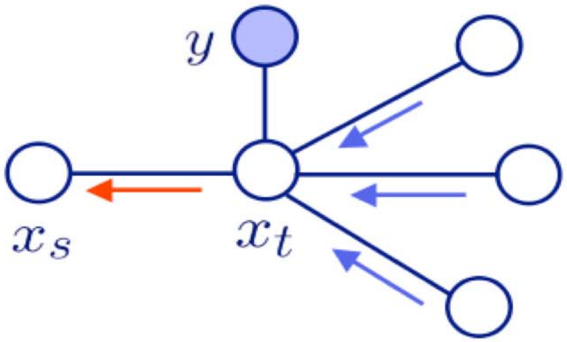


$$q_t(x_t \mid y) = \alpha \psi_t(x_t, y) \prod_{u \in \Gamma(t)} m_{ut}(x_t)$$

$\Gamma(t) \longrightarrow$ *neighborhood* of node t (adjacent nodes)

MESSAGES: Sufficient statistics (possibly approximate)

$$m_{ts}(x_s) = \alpha \sum_{x_t} \psi_{st}(x_s, x_t) \psi_t(x_t, y) \prod_{u \in \Gamma(t) \setminus s} m_{ut}(x_t)$$



I) Message Product
II) Message Propagation

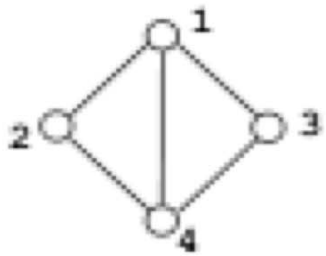
PROBLEM:

Approximation of the Messages

$$m_{ts}^{i}(x_s) \propto \int \psi_{ts}(x_t, x_s)\psi_t(x_t) \prod_{u \in \Gamma_t \setminus s} m_{ut}^{i-1}(x_t)\, dx_t\,.$$

Belief Propagation (BP) + **Further Approximation**

1) Loopy



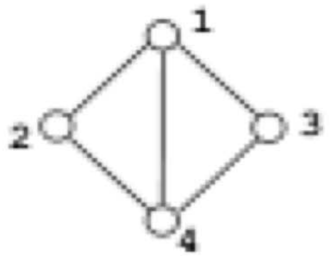
1) Discard low-likelihood state
2) Finite Parameterization
3) Edge Removal
4) Reduce Communication

GOAL: Understand the Error !!!

Belief Propagation (BP) + **Further Approximation**

1)When will the Loopy Belief Propagation Converge?

2)What is the distance between Multiple Fixed Points?



1) Additional error terms

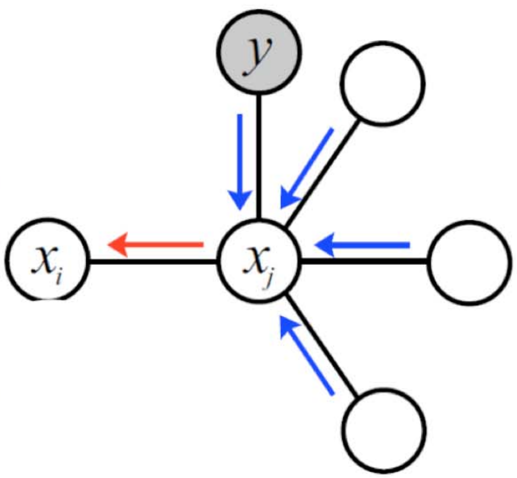
2) More realistic assumptions

1) What is the ERROR:

Notation:
m_{ts}: single message
e_{ts}: single message error
M_{ts}: product of message
E_{ts}: product of message error

$$\hat{m}^i_{ts}(x_s) = m_{ts}(x_s)e^i_{ts}(x_s).$$

$$\hat{m}^{i+1}_{ts}(x_s) \propto \int \psi_{ts}(x_s, x_t)\hat{M}^i_{ts}(x_t)dx_t \qquad \hat{M}^i_{ts}(x_t) \propto \psi_t(x_t) \prod_{u \in \Gamma_t \setminus s} \hat{m}^i_{ut}(x_t)$$



$$m_{ji}(x_i) \propto \int_{\mathcal{X}_j} \psi_{ij}(x_i, x_j)\, \psi_j(x_j, y) \prod_{k \in \Gamma(j) \setminus i} m_{kj}(x_j)\, dx_j$$

2)How to Measure the ERROR:

Metric 1: KL Divergence

$D(m_{ts} \,||\, \hat{m}_{ts})$

Metric 2: log (Dynamic Range)

$$d(e_{ts}) = \sup_{a,b} \sqrt{e_{ts}(a)/e_{ts}(b)}.$$

$$\log(d(e_{ts}))=\log\left(d\left(\frac{\hat{m}_{ts}}{m_{ts}}\right)\right) = f(\hat{m}_{ts}, m_{ts})$$

Part I: Dynamic Range

(discrete error term)

$$d(e_{ts}) = \sup_{a,b} \sqrt{e_{ts}(a)/e_{ts}(b)}.$$

a) “Metric” $\log(d(e_{ts}))$:

1) Non-negativity:

$d(e_{ts})=1 \Longleftrightarrow e_{ts}==1$

$$d(e_{ts}) = \sup_{a,b} \sqrt{e_{ts}(a)/e_{ts}(b)}. \quad >=1$$

2) Triangle Inequality:

$$d(e_1e_2)^2 = \sup_{a,b} \frac{e_1(a)e_2(a)}{e_1(b)e_2(b)} <= \sup_{a,b} \frac{e_1(a)}{e_1(b)} \sup_{a,b} \frac{e_2(a)}{e_2(b)} = d(e_1)^2 d(e_2)^2$$

$\log(d(e_1e_2)) <= \log(d(e_1)) + \log(d(e_2))$

3) Symmetry:

$$\log(d(e_{ts})) = f(\hat{m}_{ts}, \mathbf{m}_{ts}) = f(\mathbf{m}_{ts}, \hat{m}_{ts}) = \log(d(1/e_{ts}))$$

b) Basic Facts About d(e_{ts}):

1) Equivalence: point-wise log error

$$d\left(e_{ts}\right) = \sup_{a,b} \sqrt{e_{ts}(a)/e_{ts}(b)}.$$

log(d(e_{ts}))=log(sup(...))=sup(log(...))

=0.5 [sup(log(e_{ts}(a))-log(e_{ts}(b)))]

=0.5 [sup(log(e_{ts}(a))-inf(log(e_{ts}(b))]

$$\log d\left(e_{ts}\right) = \inf_{\alpha} \sup_{x} \left|\log \alpha m_{ts}(x) - \log \hat{m}_{ts}(x)\right| = \inf_{\alpha} \sup_{x} \left|\log \alpha - \log e_{ts}(x)\right|.$$

b) Basic Facts About $d(e_{ts})$:

2) Lower Bound

Thm 2:

$$\left|\log m_{ts}(x) - \log \hat{m}_{ts}(x)\right| \leq 2\log d\left(e_{ts}\right)$$

Lemma 4:

$$D\left(m_{ts} \| \hat{m}_{ts}\right) \leq 2\log d\left(e_{ts}\right)$$

b) Basic Facts About d(e_{ts}):

3) Upper Bound:

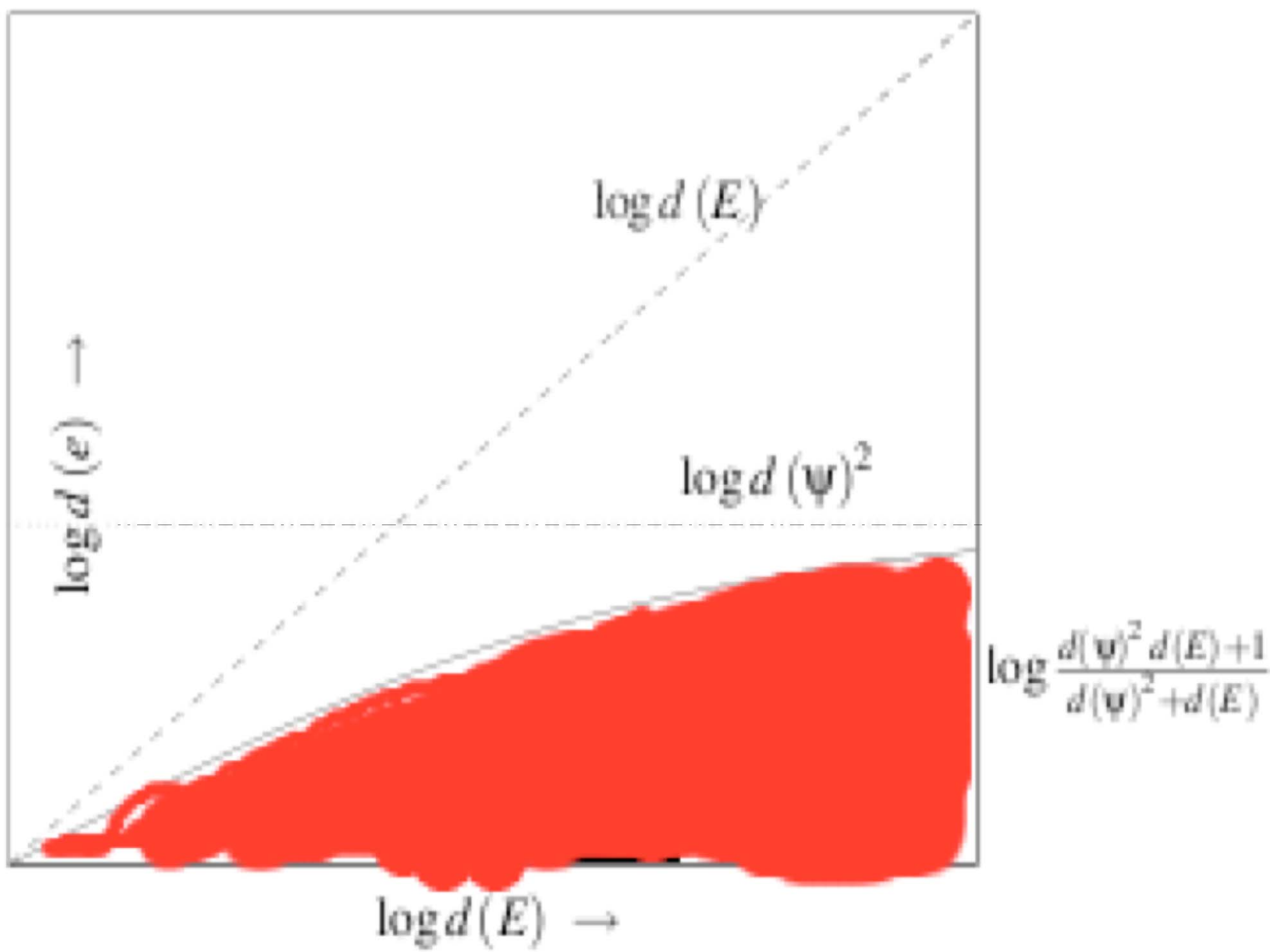
Thm8: $$d\left(e_{ts}^{i+1}\right) \leq \frac{d\left(\psi_{ts}\right)^2 d\left(E_{ts}^i\right)+1}{d\left(\psi_{ts}\right)^2+d\left(E_{ts}^i\right)}.$$

$$d\left(e_{ts}^{i+1}\right) \leq d\left(E_{ts}^i\right)$$

$$d\left(e_{ts}^{i+1}\right) \leq d\left(\psi_{ts}\right)^2.$$

$d\left(E_{ts}^i\right)$ **>=1**

$d\left(\psi_{ts}\right)^2$ **>=1**


log d (E)
log d (ψ)^2
log (d(ψ)^2 d(E)+1) / (d(ψ)^2+d(E))
log d (e) →
log d (E) →

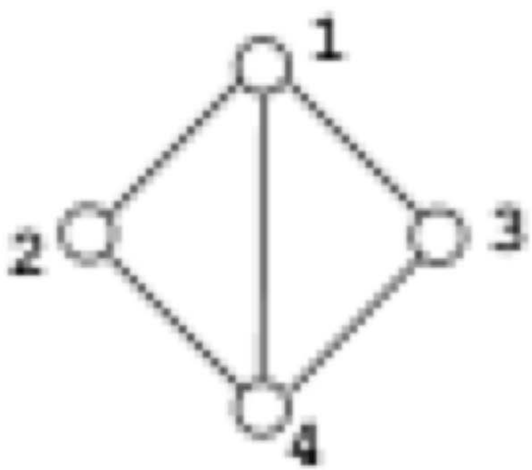
1) When Loopy Belief Propagation Converge?

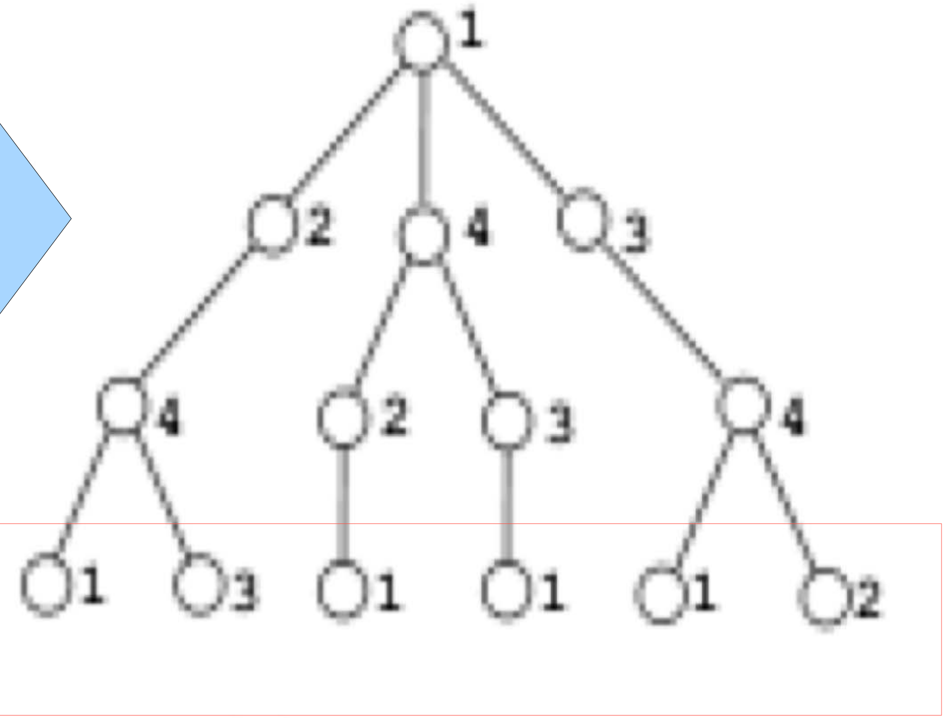
Theorem 10 (Simon's condition). *Loopy belief propagation is guaranteed to converge if*

$$\max_{t} \sum_{u \in \Gamma_t} \log d(\psi_{ut}) < 1.$$

$$\max_{(s,t) \in \mathcal{E}} \sum_{u \in \Gamma_t \setminus s} \frac{d(\psi_{ut})^2 - 1}{d(\psi_{ut})^2 + 1} < 1$$

Computation Tree





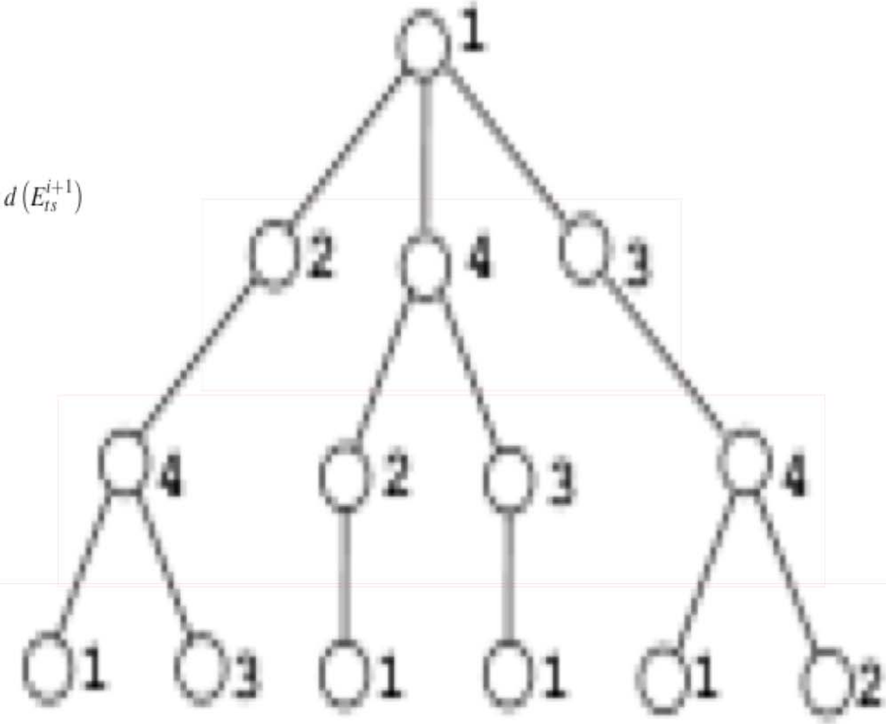
Iteration 4:

Contraction!!!

$\log d\left(E_{ts}^{i+1}\right)$

$$\log d\left(E_{ts}^{i+2}\right)$$

$$\log d\left(E_{ts}^{i+1}\right)$$



Induction:

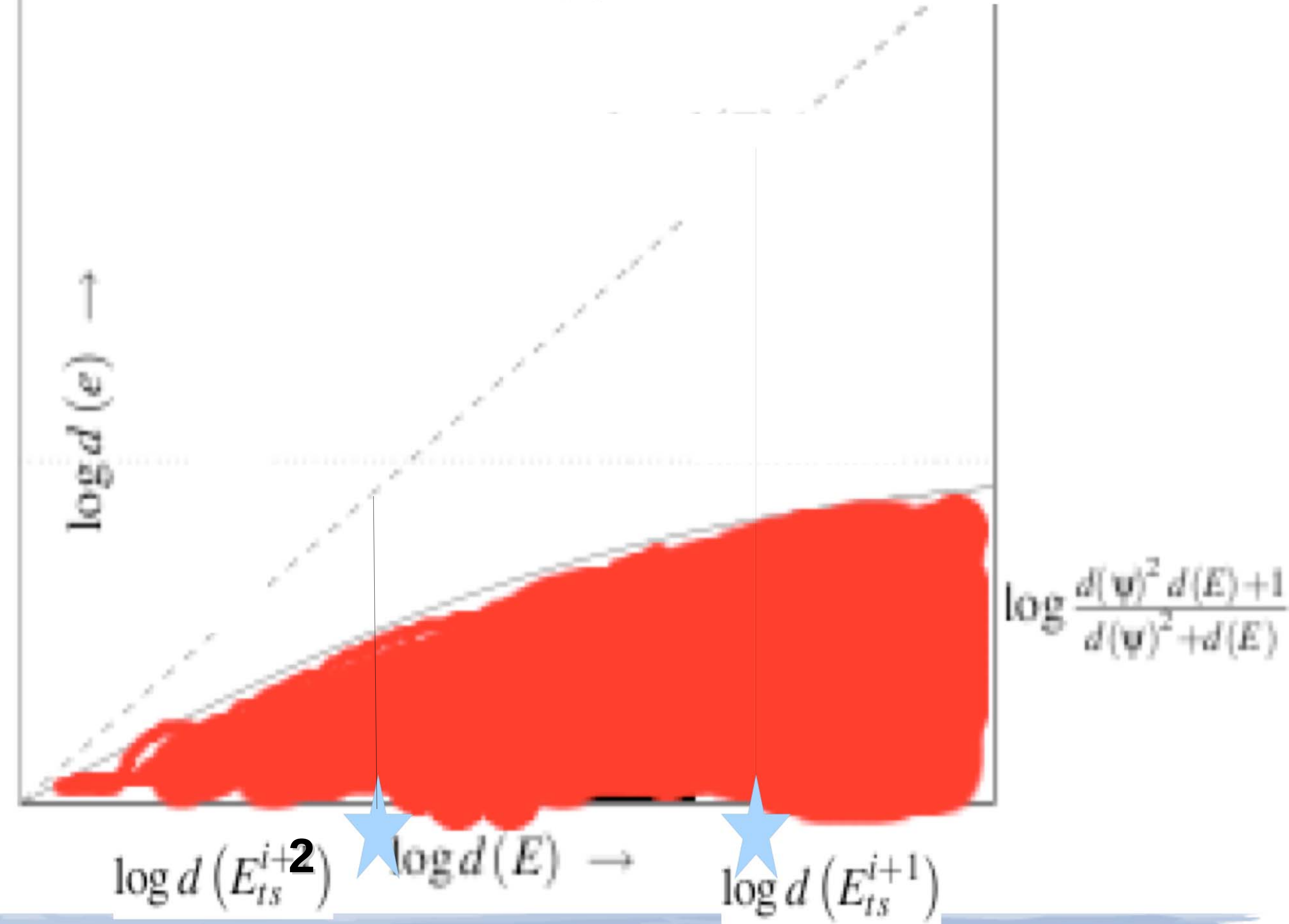
$$\log \varepsilon^1 = \max_t \sum_{u\in\Gamma_t} \log d\left(\psi_{ut}\right)^2,$$

Thm 8:

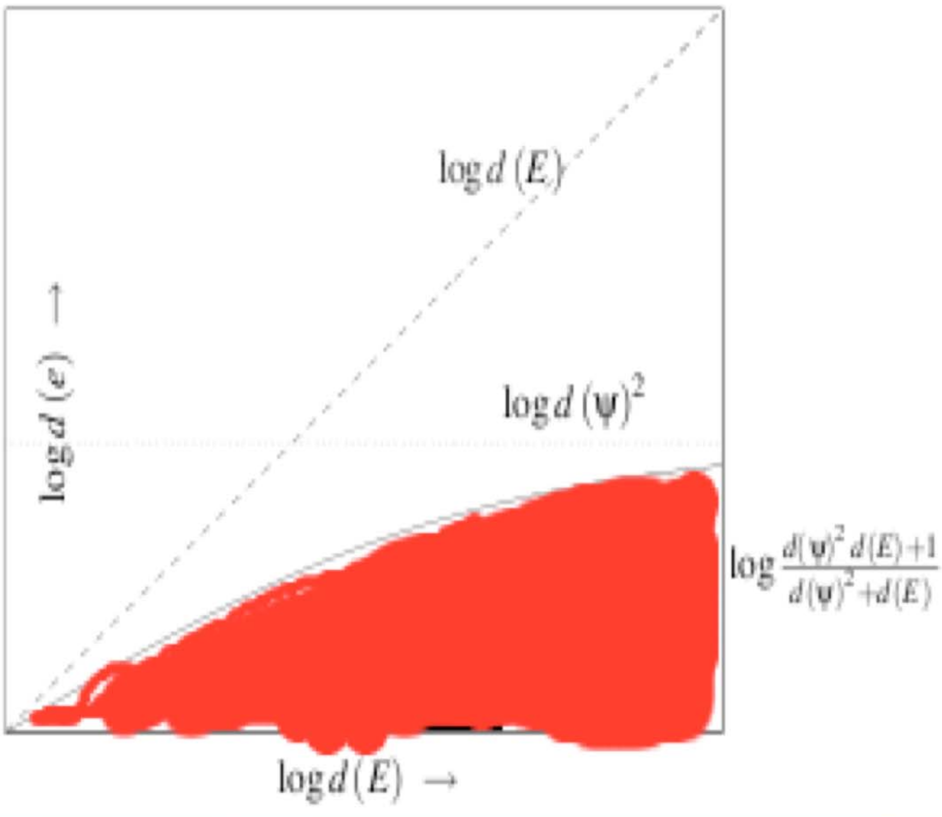
$$d\left(e_{ts}^{i+1}\right) \le \frac{d\left(\psi_{ts}\right)^2 d\left(E_{ts}^i\right)+1}{d\left(\psi_{ts}\right)^2 + d\left(E_{ts}^i\right)}.$$

$$\log d\left(E_{ts}^{i+1}\right) \le \sum_{u\in\Gamma_t\setminus s} \log \frac{d\left(\psi_{ut}\right)^2 \varepsilon^i + 1}{d\left(\psi_{ut}\right)^2+\varepsilon^i} = g(\log \epsilon)$$

$$\log d\left(E_{ts}^{i+1}\right) \le \sum_{u \in \Gamma_t \setminus s} \log \frac{d\left(\psi_{ut}\right)^2 \varepsilon^i + 1}{d\left(\psi_{ut}\right)^2 + \varepsilon^i} = g(\log \epsilon^i)$$



What We Want :



$$g_{ts}(0) = 0,$$

$$g'_{ts}(0) < 1,$$

$$g''_{ts}(z) \leq 0$$

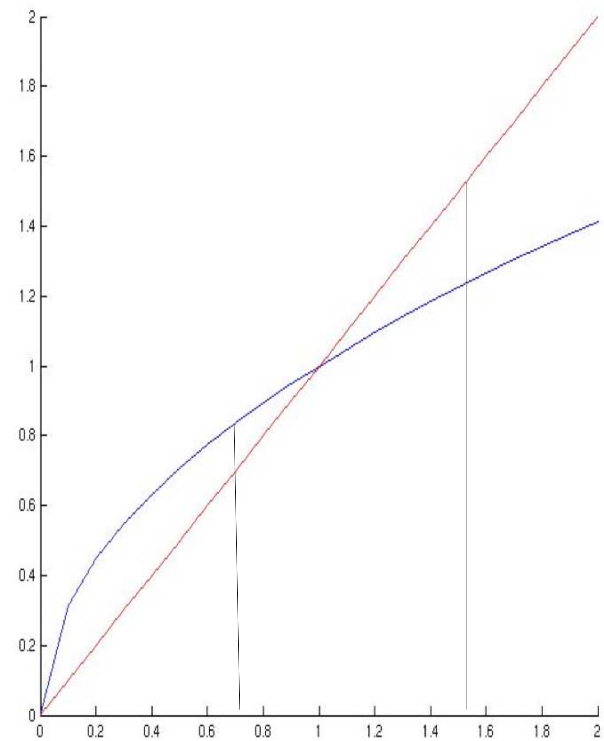
$$g_{ts}(0) = 0, \quad \sum_{u \in \Gamma_t \setminus s} \log \frac{d\,(\psi_{ut})^2\, \varepsilon^i + 1}{d\,(\psi_{ut})^2 + \varepsilon^i} = g(\log \varepsilon^i) = G(\varepsilon^i)$$

$$g'_{ts}(0) < 1, \quad \Longleftrightarrow \quad \max_{(s,t) \in \mathcal{E}} \sum_{u \in \Gamma_t \setminus s} \frac{d\,(\psi_{ut})^2 - 1}{d\,(\psi_{ut})^2 + 1} < 1$$

$$g''_{ts}(z) \le 0 \qquad g''_{ts}(z) = \varepsilon^2 G''_{ts}(\varepsilon) + \varepsilon G'_{ts}(\varepsilon)$$

2)What
is the distance between Multiple Fixed Points?

$$\max_{(s,t)\in\mathcal{E}} \sum_{u\in\Gamma_t\setminus s} \frac{d(\psi_{ut})^2-1}{d(\psi_{ut})^2+1} < 1$$



Directly From Thm 11~~

Thm 12: BP distance bound:

$$\log d\left(M_t / \hat{M}_t^n\right) \leq \sum_{u \in \Gamma_t} \log \frac{d\left(\psi_{ut}\right)^2 \varepsilon^{n-1} + 1}{d\left(\psi_{ut}\right)^2 + \varepsilon^{n-1}}$$

Initialization: $\varepsilon^1 = \max_{s,t} d\left(\psi_{st}\right)^2$

Iteration: $$\log \varepsilon^{i+1} = \max_{(s,t) \in \mathcal{E}} \sum_{u \in \Gamma_t \setminus s} \log \frac{d\left(\psi_{ut}\right)^2 \varepsilon^i + 1}{d\left(\psi_{ut}\right)^2 + \varepsilon^i}.$$

Iteration:

$$\log \varepsilon^{i+1} = \max_{(s,t)\in\mathcal{E}} \sum_{u\in\Gamma_t\setminus s} \log \frac{d\,(\psi_{ut})^2\,\varepsilon^i + 1}{d\,(\psi_{ut})^2 + \varepsilon^i}.$$

Key Observation:

$$\varepsilon^{i+1} < \varepsilon^i$$

$g''_{ts}(z) < 0$, ⇒ for z sufficiently large, $g_{ts}(z) < z$.

⇒ **Stationary:**

$$\log \varepsilon = \max_{(s,t)\in\mathcal{E}} G_{ts}(\varepsilon) = \max_{(s,t)\in\mathcal{E}} \sum_{u\in\Gamma_t\setminus s} \log \frac{d\,(\psi_{ut})^2\,\varepsilon + 1}{d\,(\psi_{ut})^2 + \varepsilon}.$$

Thm 13: Fixed-point distance bound:

$$\left|\log M_t(x)/\tilde{M}_t(x)\right| \leq 2\log d\left(M_t/\tilde{M}_t\right) \leq 2\sum_{u\in\Gamma_t}\log\frac{d\left(\psi_{ut}\right)^2\varepsilon+1}{d\left(\psi_{ut}\right)^2+\varepsilon}$$

Stationary:

$$\log\varepsilon = \max_{(s,t)\in\mathcal{E}} G_{ts}(\varepsilon) = \max_{(s,t)\in\mathcal{E}} \sum_{u\in\Gamma_t\setminus s}\log\frac{d\left(\psi_{ut}\right)^2\varepsilon+1}{d\left(\psi_{ut}\right)^2+\varepsilon}.$$

2.5) Relaxation ?

Theorem 14 (Non-uniform distance bound). *Let $\{M_t\}$ be any fixed point belief of loopy BP. Then, after $n \geq 1$ iterations of loopy BP resulting in beliefs $\{\hat{M}_t^n\}$, for any node t and for all x*

$$\left|\log M_t(x)/\hat{M}_t(x)\right| \leq 2\log d\left(M_t/\hat{M}_t^n\right) \leq 2\sum_{u\in\Gamma_t}\log \upsilon_{ut}^n$$

where υ_{ut}^i is defined by the iteration

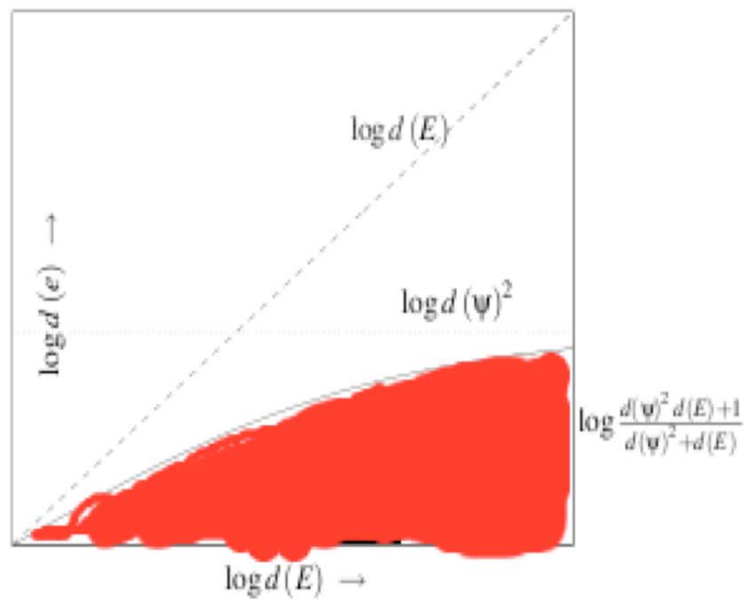
$$\log \upsilon_{ts}^{i+1} = \log\frac{d\left(\psi_{ts}\right)^2 \varepsilon_{ts}^i + 1}{d\left(\psi_{ts}\right)^2 + \varepsilon_{ts}^i} \qquad \log \varepsilon_{ts}^i = \sum_{u\in\Gamma_t\setminus s}\log \upsilon_{ut}^i \tag{16}$$

with initial condition $\upsilon_{ut}^1 = d\left(\psi_{ut}\right)^2$.

3)What about extra Error Terms?

Assumption:

additional distortion has maximum dynamic range at most δ,



$$\log d\left(E_{ts}^{i+1}\right) \le \sum_{u \in \Gamma_t \setminus s} \log \frac{d(\psi_{ut})^2 \varepsilon^i + 1}{d(\psi_{ut})^2 + \varepsilon^i}$$ **+log δ=g(log Є)+log δ**

3)What about extra Error Terms?

Theorem 15.

$$\log d\left(M_t/\hat{M}_t^n\right) \leq \sum_{u\in\Gamma_t} \log \upsilon_{ut}^n$$

Where:

$$\log \upsilon_{ts}^{i+1} = \log \frac{d(\psi_{ts})^2 \varepsilon_{ts}^i + 1}{d(\psi_{ts})^2 + \varepsilon_{ts}^i} + \log \delta \qquad \log \varepsilon_{ts}^i = \sum_{u\in\Gamma_t\setminus s} \log \upsilon_{ut}^i$$

4)Further Assumptions?

Assumption:

1) *errors* $\log e_{ts}$ *are random and uncorrelated,*

$$E\left[\log e^{i}_{st}(x) \cdot \log e^{i}_{ut}(x)\right] = 0,$$

2) Var(additional error) at most: $(\log \delta)^2.$

4)Further Assumptions?

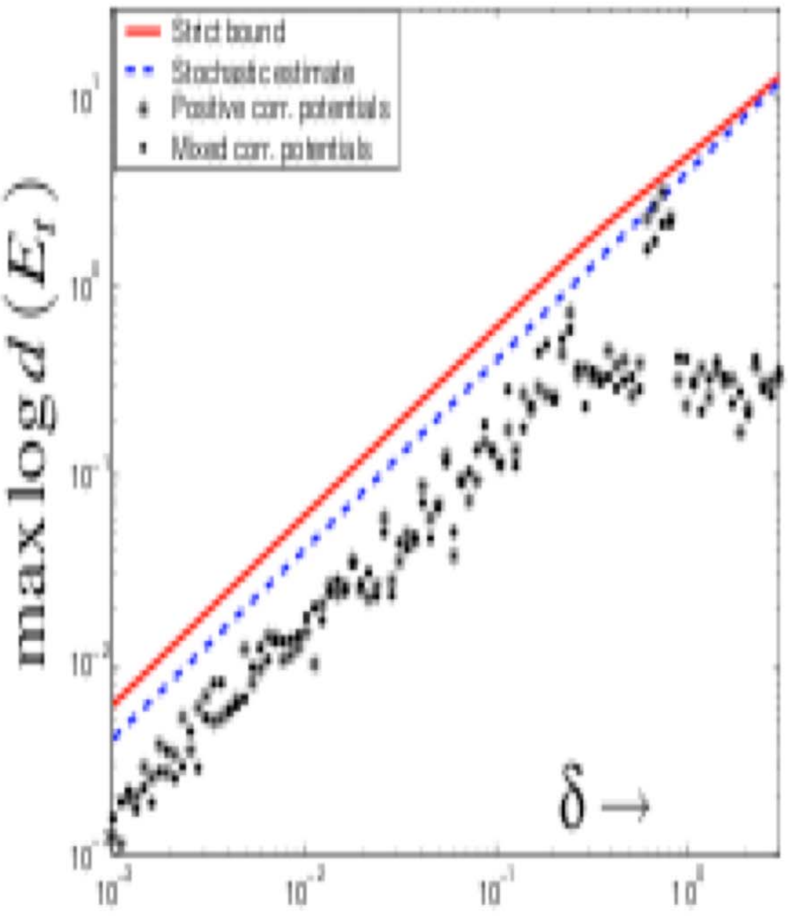
Proposition 17.

$$E\left[\left(\log d\left(E_t^i\right)\right)^2\right] \le \sum_{u \in \Gamma_t} \left(\sigma_{ut}^i\right)^2$$

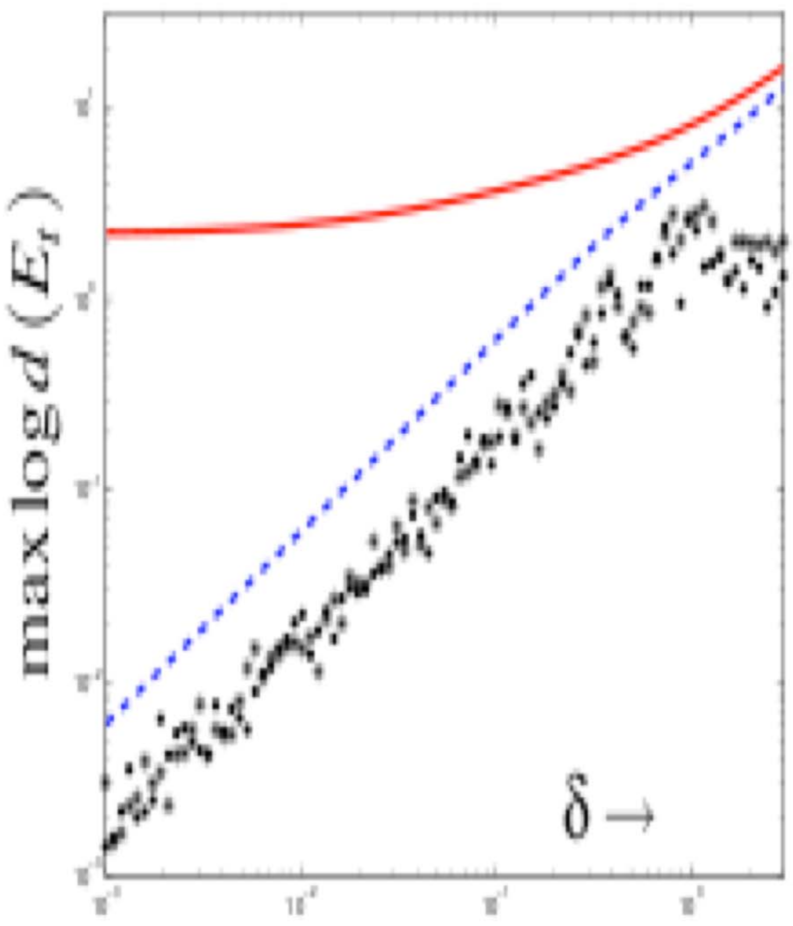
Where: $\sigma_{ts}^1 = \log d\left(\psi_{ts}\right)^2$ $\left(\log \lambda_{ts}^i\right)^2 = \sum_{u \in \Gamma_t \setminus s} \left(\sigma_{ut}^i\right)^2.$

$$\left(\sigma_{ts}^{i+1}\right)^2 = \left(\log \frac{d\left(\psi_{ts}\right)^2 \lambda_{ts}^i + 1}{d\left(\psi_{ts}\right)^2 + \lambda_{ts}^i}\right)^2 + (\log \delta)^2$$

Experiments:



(a) $\log d\,(\psi)^2 = .25$



(b) $\log d\,(\psi)^2 = 1$

Part II:
KL Divergence

(continuos error term)

To be Continued....