ISCA Archive
http://www.isca-speech.org/archive

6th International Conference on Spoken Language Processing (ICSLP 2000)
Beijing, China
October 16-20, 2000

Extended Maximum A Posterior Linear Regression (EMAPLR) Model Adaptation for Speech Recognition

Wu Chou, Olivier Siohan, Tor Andr'e Myrvoll and Chin-Hui Lee

Bell Labs., Lucent Technologies, 600 Mountain Ave., Murray Hill, NJ 07974,
U.S.A., {wuchou,siohan,chl}@research.bell-labs.com

ABSTRACT

In this paper, a new approach for model adaptation, extended maximum a posterior linear regression (EMAPLR), is described and studied. EMAPLR is an extension of maximum a posterior linear regression (MAPLR) for transform based model adaptation. The proposed approach has a close form solution under the elliptic symmetric matrix variate priors, and it is effective in our speech recognition experiments. EMAPLR is based on a direct MAPLR solution of the transform image $W\xi_s$ without explicitly solving the transformation matrix W. This is fundamentally different from conventional MAPLR and MLLR. Moreover, the proposed EMAPLR approach is incorporated with the structured prior evolution which significantly improves the algorithm efficiency and robustness. The structure of prior evolution in MAPLR is studied and it is shown that under the structured prior evolution, the priors in MAPLR follows a recursive formulation. Experimental results on WSJ (Spoke 3) non-native speaker adaptation task indicates that significant gain over MLLR and MAPLR can be obtained with same amount of adaptation data.

1. INTRODUCTION

Model adaptation based on linear transformation has become increasingly popular in speech recognition. In maximum likelihood linear regression (MLLR) based approach [1], a group of linear transformation matrices is estimated based on the principle of maximizing the likelihood on the adaptation data. Model parameters are then adapted according to the estimated linear transforms. It is known that when adaptation data are sparse, maximum likelihood estimation often leads to biased estimate. Maximum a posterior (MAP) based adaptation is a powerful approach which can be applied to improve the adaptation performance [9]. In MAP estimation, an appropriate prior is used and the MAP estimate is given by

$$\lambda_{MAP} = \underset{\lambda}{\text{argmax}}\, f(x \mid \lambda) g(\lambda). \qquad (1)$$

where the prior distribution $g(\lambda)$ characterizes the distribution of the model parameter λ. The use of priors in MAP estimation provides a way to incorporate prior knowledge into the model parameter estimation process. The relation between ML and MAP estimation is based on the Bayes' theorem where the posterior distribution $p(\lambda \mid x) \propto f(x \mid \lambda)g(\lambda)$ and $f(x \mid \lambda)$ is the likelihood function. In linear transform based model adaptation, the prior distribution $g(\lambda)$ is a matrix variate distribution. Maximum a posterior linear regression (MAPLR) [3, 5, 6] estimates the linear transforms for model adaptation based on the principle of MAP as it differs from maximum likelihood based MLLR formulation. MAPLR has many interesting properties. In particular, there is no conjugate prior which can be specified for MAPLR. The use of normal-Washart distribution leads to close form solution for joint estimation of mean and covariance matrix in direct MAP adaptation [9], but it will not lead to close form solution in MAPLR. Therefore, the exact effects of prior in linear regression based model adaptation need to be derived and characterized.

Although numerical methods can be used to find approximate MAPLR solution, close form solution in MAPLR is more desirable. However, the use of prior makes the problem of MAPLR much more difficult, and the solution is strongly dependent on the form of the prior being used. In [3], it is proved that MAPLR has a close form solution for mixture Gaussian distribution under the elliptic symmetric matrix variate priors. From the close form MAPLR solution, the effects of such priors in MAPLR are explicitly characterized. Since then, the close form solution to MAPLR has been extended to matrix contour distribution [5], prior distribution from Markov random field [6], etc. The existence of close form solution under selected informative priors makes MAPLR applicable to large vocabulary speech recognition with efficiency.

In this paper, we study prior evolution in MAPLR with elliptic symmetric matrix variate priors, and a new approach for model adaptation, extended maximum a posterior linear regression (EMAPLR), is described and studied. The novel contributions of this paper are:

- MAPLR theory has been extended from cluster based linear transform model adaptation to individual component based MAPLR adaptation.
- EMAPLR is derived, which is based on a direct MAPLR solution of the transform image $W\xi_s$, without explicitly solving the transform matrix W as is typical in MLLR or MAPLR.
- The close form solution of EMAPLR is obtained for elliptic matrix variate priors utilizing the full structure of linear transform based MAPLR formulation.
- EMAPLR is integrated with a general structured prior evolution scheme, and in the special case of evolving priors along a tree, EMAPLR is an extension of SMAPLR [4].
- Prior evolution with elliptic symmetric matrix variate prior in MAPLR is studied and a close form recursive formulation is derived.
- An efficient EMAPLR implementation is described and experimental results are given which indicate the efficacy of the proposed approach.

2. MAXIMUM A POSTERIOR LINEAR REGRESSION (MAPLR)

In MAPLR framework, the group of linear transformations are estimated according to equation (1). We consider the problem of linear transformation based adaptation of the mean vectors in HMMs. With mixture Gaussian observation densities, the adapted mean in mixture component s in the mixture Gaussian density is given by

$$\hat{\mu_s} = W_s \xi_s \quad (2)$$

where W_s is an $n \times (n+1)$ matrix, and ξ_s is an extended mean vector (we consider general offset case) $\xi_s = [1, \mu_{s_1}, ..., \mu_{s_n}]'$. The auxiliary Q-function under EM algorithm for MAPLR with prior matrix distribution $p(W_s)$ is given by

$$Q(\lambda, \bar{\lambda}) = a + \log p(W_s) + \sum_{t,n,m} \gamma_t(n, m) \log(o_t \mid W_s, \mu_{n,m}, \Sigma^{-1}_{n,m}), \quad (3)$$

where a is a term not relevant to W, $\gamma_t(n, m) = P(s_t = n, l_t = m \mid O, \lambda)$ is the probability of being in state n and mixture m at time t given observation sequence $O = \{o_1, ..., o_T\}$. To reduce the complexity of notation, we consider the following relevant part in Q-function:

$$Q(\lambda, \bar{\lambda}) = logp(W_s) + \sum_{t,n,m} \gamma_t(n, m)(o_t - W_s\xi_s)'\Sigma_s^{-1}(o_t - W_s\xi_s), \quad (4)$$

where s is a generic index and a function of (t, n, m).

Let $X = (X_{i,j}) = (X_1, ..., X_n)'$ be an $N \times P$ random matrix with $E(X_i) = \mu_i = (\mu_{i,1}, ..., \mu_{i,p})$. A random matrix X with values $x \in R^{N\times P}$ has a distribution belonging to the family of elliptically symmetric matrix variate distribution with location parameter $\bar{U} = (\bar{\mu}_1, ..., \bar{\mu}_N)'$ and scale matrix $\Delta = diag(\Sigma_1, ...\Sigma_N)$, if its probability density function can be written as $f_X(x) = (det\Delta)^{-1/2} q(\sum_{i=1}^N (x_i - \mu_i)'\Sigma_i^{-1}(x_i - \mu_i))$, where q is a exponential function in $[0, \infty)$ of the sum of N quadratic forms $(x_i - \mu_i)'\Sigma_i^{-1}(x_i - \mu_i))$, $i = 1, ..., N$ [7]. For a given matrix W, we denote the i-th row of W by $W(\underline{i})$ and the j-th column of W by $W(\mid j)$. Elliptic symmetric matrix variate distribution can also be defined through column vectors. It corresponds to apply the current definition to the transpose of the matrix X.

MAPLR solution can be derived from (4) as a maximum of $Q(\lambda, \bar{\lambda})$. To maximize $Q(\lambda, \bar{\lambda})$ with respect to W, it is necessary to differentiate $Q(\lambda, \bar{\lambda})$ with respect to W and solve equation

$$\frac{dQ(\lambda, \bar{\lambda})}{dW} = 0, \quad (5)$$

where the solution involves both parameters of the model and parameters of the priors. These additional parameters of the priors are called hyperparameters and must be specified in MAPLR model adaptation.

3. STRUCTURED PRIOR EVOLUTION IN MAPLR

One critical issue in MAPLR is how to construct informative priors during model adaptation. This problem becomes acute when adapting model using a large regression tree or a large directed graph. In MAPLR, the purpose of using a large regression tree is to correlate model parameters and form cluster based transform matrix W. The appropriate priors need to be specified for each cluster node in order to form MAPLR estimation of W. In [8], a structured maximum a posterior (SMAP) algorithm was developed for direct MAP adaptation of model parameters. In this paper, we consider a general structured prior evolution for MAPLR which is only based on the hierarchical prior approximation, not on the tree structure nor on the existence of a cluster Gaussian. Therefore, it can be applied to various structures in MAPLR. The idea behind this approach is to allow priors to evolve in a structurally controlled fashion along a selected path in a directed graph, where tree structure is one special case. Consider $t_0 \rightarrow t_1 \rightarrow t_2 ... \rightarrow t_l$ is a path from node t_0 to node t_l, and let $W^{(0)}, W^{(1)}, ..., W^{(l-1)}, W^{(l)}$ be the MAPLR solution of the transform matrices on the path from t_0 to t_l. The structured prior evolution in MAPLR is a method of deriving priors from t_0 to all following nodes on the path. In this approach, the mode of the prior distribution at each node is taken to be the MAPLR solution of its predecessor node. In the case of elliptic symmetric matrix variate prior distribution and t_0 to t_l is a directed path, the location parameter matrix of prior at node t_l has the following form:

$$\bar{U}^{(l)} = W^{(l-1)}, \quad (6)$$

where $W^{(l-1)}$ is the MAPLR solution at the predecessor node t_{l-1}. Once the prior distribution is specified for the path root node, priors at subsequent nodes can be derived through this prior evolution process. In addition to other interpretations used in [4], such a structured prior evolution is meaningful in that it introduces a subjective knowledge that the transformation at the child node should be close to the MAPLR transformation at its parent node, and the data distribution at the child node is smoothed by this informative prior information embedded in the transforms of its predecessors along the evolution path. For the rest of this section, we focus on the special prior structure with elliptic symmetric matrix variate priors in MAPLR.

It is proved [3] that with elliptical symmetric matrix variate priors, the close form solution of MAPLR is given by:

$$W(\underline{i}) = (G^{(i)} + \Sigma_i^{-1})^{-1}(Z(\underline{i})) + \mu(\underline{i})\Sigma_i^{-1}), \quad (7)$$

where $W(\underline{i})$ is the i-th row of the transform matrix W, $\mu(\underline{i})$ is the i-th row in the location parameter matrix and Σ_i^{-1} is the corresponding scale matrix in the elliptic symmetric matrix variate distribution, $G^{(i)}$ and $Z^{(i)}$ are corresponding matrices used in MLLR formulation for $W(\underline{i})$. Let $A_l = (G_l^{(i)} + \Sigma_i^{-1}(l))^{-1}$, $b_l = Z_l(\underline{i})$, $c_l = \mu_l(\underline{i})$ and $D_l = \Sigma_i^{-1}(l)$ with l indexing the tree level. With structured prior evolution and from (6), $c_l = \mu_l(\underline{i}) = W(\underline{i})^{(l-1)}$. The MAPLR solution for i-th row of the transformation matrix at l-th level tree node can be represented using the following recursive relation:

$$\begin{aligned} W(\underline{i}) &= A_l(b_l + c_l D_l) \\ &= A_l b_l + A_l(A_{l-1}(b_{l-1} + c_{l-1}D_{l-1}))D_l \\ &= A_l b_l + A_l A_{l-1} b_{l-1} D_l + A_l A_{l-1} c_{l-1} D_{l-1} D_l \\ &= \sum_{k=1}^{l} P_k b_k R_k + A_l ... A_0 c_0 D_0 ... D_l \end{aligned} \quad (8)$$

where $P_k = \prod_{i=k}^{l} A_i$ and $R_k = \prod_{i=k+1}^{l} D_i$. This recursive relation is reminiscent to the recursive relation derived in SMAP for direct MAP model parameter adaption[9]. It shows that with structured prior evolution, the MAPLR solution is a combination of information from the data and the information from priors on the prior evolution path.

4. EXTENDED MAXIMUM A POSTERIOR LINEAR REGRESSION (EMAPLR)

As pointed in [1], the reason in MLLR that we do not initiate a matrix W for each Gaussian distribution is because any unseen mixture is never updated, and if enough data is available to estimate W then there is probably more than enough data to re-estimate the model parameters, since there are usually more free parameters in W then in the mixture itself. MLLR and MAPLR are all based on the estimation of transformation matrix W and then apply the estimated transformation matrix to adapt model parameters according to $\mu_{new} = W\xi_s$.

This is an indirect estimation of the transformed model parameters, and numerical errors in estimating W will carry over to the final results. Moreover, there are $N \times (N+1)$ free parameters in W. Even with certain amount of data, W at lower level tree node is often ill-conditioned and may not be solvable at the mixture component level. Therefore, the linear transform based adaptation is usually based on cluster of Gaussians in which different Gaussians in the cluster share the same transform. It is an effective solution to the unseen data problem, since all Gaussian distributions can be adapted at the same time sharing the cluster pooled adaptation data. When adaptation data is limited, cluster based model adaptation has the performance advantage over direct adaptation of model parameters. However, one drawback with cluster based adaptation is that the performance improvement can quickly saturate when the amount of adaptation data increases. On the other hand, direct adaptation can take advantage of more adaptation data, in which the adaptation is tailored to each individual model component and becomes more effective. This is a problem which plagues the effectiveness of cluster based model adaptation.

The key idea behind the proposed extended maximum a posteriori linear regression (EMAPLR) approach is to derive an exact estimate of $\mu_{new} = W\xi_s$ for each mixture component without explicitly solving the transformation matrix W. The adapted model parameters (i.e. the new mean vector $W\xi_s$) are solved directly through the EM formulation for MAPLR without explicitly solving the transformation matrix W as is typical. Therefore, mixture specific transform is applied for better adaptation efficiency. This is fundamentally different from the conventional MLLR or MAPLR which is based on solving the transformation matrix W first and then obtaining the image $W\xi_s$ for each mixture component. With elliptic symmetric matrix variate priors, the close form solution of EMAPLR is obtained:

Theorem: Under certain conditions and elliptic symmetric matrix variate priors, the EMAPLR solution for each mixture component has the following form

$$W\xi_s = (\tau I + \sum \gamma_t(n,m)I)^{-1}(\tau \bar{U}\xi_s + \sum \gamma_t(n,m)o_t) \tag{9}$$

where $\bar{U}$ is the location parameter matrix of the elliptic symmetric matrix variate distribution, τ is a vector of positive constants, $\gamma_t(n,m)$ is the posterior probability of being in state n and mixture m at time t given the observation $O = \{o_1, ..., o_T\}$. Proof to the theorem is given in the Appendix.

Under the structured prior evolution discussed in Section 3, $\bar{U}\xi_s = W^{(l-1)}\xi_s$ where $W^{(l-1)}$ is the MAPLR solution of the predecessor node, the EMAPLR formulation (9) becomes:

$$W\xi_s = (\tau I + \sum \gamma_t(n,m)I)^{-1}(\tau W^{(l-1)}\xi_s + \sum \gamma_t(n,m)o_t) \tag{10}$$

It is important to note that for unseen mixture component, Eq. (10) reduces to MAPLR, and the model parameters are transformed by the cluster based MAPLR transform $W^{(l-1)}$. For seen mixture components, EMAPLR introduces a mixture component dependent MAPLR transform where the cluster based transform $W^{(l-1)}$ is incorporated as the mode of the prior. EMAPLR maintains the advantage of cluster based linear transform adaptation where each individual Gaussian is adapted through cluster correlation, either directly by the cluster based transform or by the MAPLR transform which incorporates the cluster based transform as its prior mode. The use of MAPLR is instrumental in EMAPLR, and cluster based transforms are integrated in the individual component based transform through the theoretical framework of MAPLR. EMAPLR inherits the good property of direct model parameter adaptation, and linear transform for each individual Gaussian will be different once the specific adaptation data for that Gaussian becomes available. The adapted mean vector $\mu_{new} = W\xi_s$ is derived directly from the EM formulation of MAPLR without explicitly solving W which may not be possible for low sample counts. For a fixed number of cluster transforms, it is shown that MAPLR converges to MLLR as the amount of data at each cluster increases. This is not the case for EMAPLR. EMAPLR takes the cluster based transforms from MAPLR as prior modes and performs a direct mixture component dependent MAPLR transform. This extension to MAPLR is significant and always maintained even the data at each cluster increases.

	% Word Error	Reduction over Baseline
Baseline (SI-84)	27.45%	N/A
MLLR	14.55%	47%
SMAPLR	14.39%	48%
EMAPLR	12.76%	53.5%

5. EXPERIMENTAL RESULTS

The speech recognition experiments were performed on the Wall Street Journal speaker adaptation task using the official 1993 Spoke 3 (S3) non-native speaker adaptation and evaluation data. The standard 5k trigram specified for the evaluation was used. The speech feature vector is MFCC based with standard 39 dimensions. There are 10 speakers in the data base with 40 adaptation sentences for each speaker. The speaker independent models were trained on the standard speaker independent WSJ training corpus (SI-84 and SI-284). The baseline speaker independent models were obtained using decision tree state tying. It is based on continuous mixture Gaussian observation densities and context dependent triphones with position dependent cross-word context dependency.

A broad phonetic class tree was derived from the decision tree state tying and used to construct a group of linear transforms [1]. Two sample counts were used. One was for generating a true MAPLR linear transform and another one is significantly lower for prior scale factor estimation [3]. In order to reduce the number of hyperparameters, diagonal scale factor matrices Σ_i were used. The prior estimation was based on the structured prior evolution described in the previous section. At each cluster node, the location parameter matrix of the prior is derived from the structured prior evolution and the prior scale factor is estimated from the transforms with lower sample count. Due to the location-scale nature of elliptic symmetric matrix variate distribution, the number of data sample points needed for prior scale factor estimation can be set quite low. A lower bound threshold of 10 was used in the experiments. If the number of lower sample count transformations under that tree node is below the threshold, the scale factor matrix at the parent node will be used. The sample count for a MAPLR transform was set to 1000 and lower sample count to generate extra transforms for prior esti-

Adaptation	% Word Error	Reduction over Baseline
Baseline (SI-284)	21.84%	N/A
MLLR	12.19%	44.2%
SMAPLR	12.21%	44.1%
EMAPLR	11.15%	49%

mation was set to 200. The weighting parameter τ was set to 2. Experimental results were of batch mode supervised adaptation using the 40 enrollment sentences in S3 non-native speaker adaptation corpus. Two speaker independent seed models were used. One was based on a seed speaker independent model obtained from SI-84 WSJ training data set and another one was based on a seed speaker independent model obtained from SI-284 training data set. MLLR, SMAPLR and EMAPLR in the experiments were based on the same phonetic class tree and with the same number of cluster transforms. Moreover, EMAPLR was identical to SMAPLR up to the cluster level and differed only in the extension part described in Section 4.

The baseline SI model obtained from the SI-84 training data had a word error rate of 27.45% on S3 non-native speaker test data. After 40 sentence adaptation using EMAPLR, the word error rate reduced to 12.76% which is a 53.5% error reduction from the baseline model. In contrast, after 40 sentence adaptation using SMAPLR, the word error rate reduced to 14.39%. The performance improvement due to the extension from SMAPLR to EMAPLR is significant. Similar performance improvement of EMAPLR over MLLR is also observed. For baseline SI model from SI-284 training data, the word error rate on S3 non-native speaker test data is 21.84%. After 40 sentence adaptation using EMAPLR, the error rate reduced to 11.15% which is a 49% error rate reduction from the baseline model. The contrast MLLR and SMAPLR had a word error rate around 12.2% after 40 sentence adaptation. In these experiments, the proposed EMAPLR outperforms MLLR and SMAPLR with an additional ($\approx$ 10%) error rate reduction.

6. SUMMARY

In this paper, an approach in model adaptation based on extended maximum a posterior linear regression (EMAPLR) was proposed and studied. It is a further extension of the MAPLR theory in which the cluster based transform adaptation is extended to individual Gaussian component based MAPLR transform. Under the selected elliptic symmetric matrix variate priors, the close form solution of EMAPLR was derived. It is based on a direct MAPLR solution of the transform image $W\xi_s$ without explicitly solving the transformation matrix W. This is fundamentally different from conventional MAPLR and MLLR. Moreover, the transform parameters $W\xi_s$ was solved explicitly utilizing the full structure of linear transform based MAPLR formulation. This is critical, since the transform matrix W at lower level tree node is often ill conditioned and may not be solvable. The proposed EMAPLR approach is combined with the structured prior evolution which significantly improves the algorithm efficiency and robustness. A general prior evolution scheme was also described which can be applied to various structures including the tree structure as a special case. This opens the door to other flexible implementations in which the one used in this paper is just one such instance. The structure of prior evolution in MAPLR was studied and it was shown that the prior evolution in MAPLR had a close form recursive formulation. Experiments were performed and significant performance improvement of EMAPLR over MLLR and SMAPLR was observed.

7. REFERENCES

1. C. J. Leggetter and P. C. Woodland (1995), "Maximum Likelihood Linear Regression for Speaker Adaptation of Continuous Density Hidden Markov Models," *Speech Communication*, **Vol. 9**, pp. 171-185.
2. Q. Huo N. Smith and B. Ma, "Efficient ML Training of CDHMM Parameters Based on Prior Evolution, Posterior Intervention and Feedback," Proc. ICASSP'2000, vol2, pp. 1001-1004.
3. W. Chou "Maximum A Posterior Linear Regression with Elliptically Symmetric Matrix Variate Priors" Proc. EuroSpeech'99, vol.1, pp 1 – 4, 1999.
4. T. -A. Myrvoll, O. Siohan, C.-H. Lee and W. Chou "Structural Maximum a Posteriori Linear Regression for Unsupervised Speaker Adaptation" Proc. ICSLP'2000, Beijing, China.
5. C. Chesta, O. Siohan and C.-H. Lee, "Maximum A Posterior Linear Regression for Hidden Markov Model Adaptation", Proc. EuroSpeech'99, vol1. pp211-214.
6. X. Wu and Y. Yan "Linear Regression Under Maximum A Posteriori Criterion With Markov Random Field Prior", Proc. ICASSP'2000, vol2., pp 997-1000.
7. N. Giri, "Multivariate Statistical Analysis", em Marcel Dekker, 1996.
8. K. Shinoda and C.-H. Lee (1997), "Structural MAP Speaker Adaptation Using Hierarchical Priors", *Proc. 1997 IEEE Workshop on Automatic Speech Recognition and Understanding*, pp. 381-388, Santa Barbara, 1997.
9. J.-L. Gauvain, C.-H. Lee, *'Maximum A-Posteriori Estimation for Multivariate Gaussian Mixture Observations of Markov Chains'*, IEEE Trans. on Speech and Audio Processing, Vol. 2, No. 2, pp. 291-298, 1994.

Appendix

Consider elliptic symmetric matrix variate distribution defined on the column of matrix W. If assume that the scale matrices $\Sigma_1^{-1} = \Sigma_2^{-1} = \ldots = \Sigma^{-1}$, then $\frac{\partial \log P(W)}{\partial W} = \Sigma^{-1}(W-U)$, From EM equation and set $\frac{\partial Q}{\partial W} = 0$,

$$\sum_{t,n,m} \gamma_t(n,m)\Sigma_{n,m}^{-1} o_t \xi_s' + \Sigma^{-1}U = \sum_{t,n,m} \gamma_t(n,m)\Sigma_{n,m}^{-1} W\xi_s\xi_s' + \Sigma^{-1}W \quad (11)$$

To obtain MAPLR transform for one particular mixture component (i.e. ξ_s), $\Sigma_{n,m}^{-1}$ is a fixed covariance matrix. Left multiply $\Sigma_{n,m}$ and then right multiply ξ_s on both sides of Eq. (11)

$$\sum_{t,n,m} \gamma_t(n,m) o_t \xi_s'\xi_s + \Sigma_{n,m}\Sigma^{-1}U\xi_s = \sum_{t,n,m} \gamma_t(n,m) W\xi_s\xi_s'\xi_s + \Sigma_{n,m}\Sigma^{-1}W\xi_s \quad (12)$$

Note that $\xi_s'\xi_s = \| \xi_s \|_2^2$ is a scalar and let $\Sigma^{-1} = \tau \| \xi_s \|_2^2 \Sigma_{n,m}^{-1}$,

$$W\xi_s = [\tau I + \sum_{t,n,m} \gamma_t(n,m) I]^{-1} [\tau U\xi_s + \sum_{t,n,m} \gamma_t(n,m) o_t]. \quad (13)$$

With structural prior, $U\xi_s = W^{(l-1)}\xi_s$ and

$$W\xi_s = [\tau I + \sum_{t,n,m} \gamma_t(n,m)]^{-1} [\tau W^{(l-1)}\xi_s + \sum_{t,n,m} \gamma_t(n,m) o_t]. \quad (14)$$