Appendix: parameter estimation for PAT

Zhijian Ou
ozj@tsinghua.edu.cn

Yang Zhang
zhangyangbill@gmail.com

Department of Electronic Engineering, Tsinghua University, Beijing 100084, China

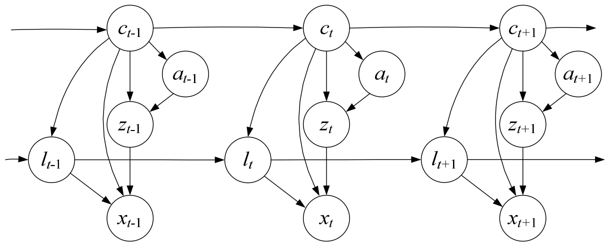

Figure 1: Probabilistic graphical model of PAT

As has been mentioned in Section 2.4, the parameters of PAT

$$\Theta \triangleq \{\mu_c, \Phi_c, \Psi, m_c, \sigma_c^2\}$$

can be estimated using the EM algorithm. In this appendix, we provide the details of parameter estimation for PAT.

1 Notations

Operators and constants used in this appendix are defined in Table 1.

Table 1: Notations

Notation	Definition
$p(\cdot\,\vert\Theta)$	probability density function given parameters Θ
$\wedge$	re-estimated parameters
$'$	matrix transpose
$\langle\cdot\rangle_p$	the expectation operator with respect to the distribution p as specified in the subscript
$\bar{\bar{\cdot}}$	two bars above means $\langle\cdot\rangle_{p(z_t,a_t\vert x_t,c_t,l_t)}$
diag[$\cdot$]	extracting the diagonal elements of a matrix to form a diagonal matrix
$tr[\cdot]$	the trace of a matrix
$\vert\cdot\vert$	the determinant of a matrix
T	the total number of frames
N	the dimension of observation vector x_t
M	the dimension of DCT vector z_t

2 Model Review

The probabilistic graphical model is shown in Figure 1. For detailed explanation, please refer to Section 2 in the paper. The joint probability distribution of a T-frame utterance is given as follows, where we use the Matlab notation to represent a set of variables, e.g. $x_{1:T} \triangleq x_1, \cdots, x_T$:

$$p(c_{1:T},l_{1:T},z_{1:T},a_{1:T},x_{1:T}) = \prod_t p(c_t|c_{t-1})p(l_t|l_{t-1},c_t)\cdot p(a_t|c_t)p(z_t|c_t,a_t)p(x_t|l_t,c_t,z_t)$$
$$= \prod_t p(c_t|c_{t-1})p(l_t|l_{t-1},c_t)\cdot \mathcal{N}\left(a_t;m_{c_t},\sigma_{c_t}^2\right)\mathcal{N}\left(z_t;a_t\mu_{c_t},m_{c_t}^2\Phi_{c_t}\right)\mathcal{N}\left(x_t;\Gamma_{l_t}z_t,m_{c_t}^2\Psi\right) \quad (1)$$

3 Auxiliary Function

The EM algorithm consists of two steps. In the E-step, the following auxiliary function is computed:

$$Q(\Theta|\Theta_{old}) = \sum_{c_{1:T},l_{1:T}}\int_{z_{1:T},a_{1:T}} dz_{1:T}da_{1:T}p(c_{1:T},l_{1:T},z_{1:T},a_{1:T}|x_{1:T},\Theta_{old})\cdot\log p(c_{1:T},l_{1:T},z_{1:T},a_{1:T},x_{1:T}|\Theta) \quad (2)$$

where Θ_{old} is the old parameters estimated in the previous iteration. The auxiliary function is essentially the expectation of the complete log-likelihood

$$L(\Theta) = p(c_{1:T},l_{1:T},z_{1:T},a_{1:T},x_{1:T}|\Theta) \quad (3)$$

over all hidden variables, with respect to the posteriori distribution $p(c_{1:T},l_{1:T},z_{1:T},a_{1:T}|x_{1:T},\Theta_{old})$.

In the M-step, the auxiliary function is maximized to update the parameters, namely

$$\widehat{\Theta} = \mathrm{argmax}_{\Theta}\, Q(\Theta|\Theta_{\mathrm{old}}) \quad (4)$$

subject to the normalizing constraint

$$\mu_c'\mu_c = 1 \quad (5)$$

In the following two sections, we introduce the details of the two steps. To make the results clearer and more straightforward, the M-step will be introduced first.

4 M-step

As is shown in Equ. (4) and (5), the M-step deals with a constrained maximization problem, which can be solved using Lagrange function:

$$J(\Theta|\Theta_{\mathrm{old}}) = Q(\Theta|\Theta_{\mathrm{old}}) + \sum_c \lambda_c(\mu_c'\mu_c - 1) \quad (6)$$

where λ_c is Lagrange multiplier. Maximizing Equ. (6) gives the re-estimation formula for all parameters. The result is given as follows:

$$\hat{\mu}_c = \left(\sum_t \gamma_{t,c}\langle a_t^2\rangle_{p(a_t|\,x_{1:T},c_t=c)}\,\mathbb{I} + \;2\lambda_c\Phi_c\mathbf{1}'\mathbf{1}\right)^{-1}\cdot\sum_t \gamma_{t,c}\langle a_t z_t\rangle_{p(a_t,z_t|x_{1:T},c_t=c)} \quad (7)$$

where $\gamma_{t,c} \triangleq p(c_t = c|x_{1:T})$, $\mathbf{1} \triangleq \{1\}_{M\times 1}$, $\mathbb{I}$ denotes the identity matrix. λ_c can be solved by Equ. (7) together with the following constraint:

$$\hat{\mu}_c'\hat{\mu}_c = 1 \quad (8)$$

$$\widehat{\Phi}_c = \frac{\sum_t \gamma_{t,c}\langle(z_t - a_t\hat{\mu}_c)(z_t - a_t\hat{\mu}_c)'\rangle_{p(z_t,a_t|x_{1:T},c_t=c)}}{\sum_t \gamma_{t,c}} \quad (9)$$

$$\widehat{\Psi} = \frac{1}{T}\sum_t \mathrm{diag}\left[\langle\widehat{m}_{c_t}^{-2}\left(x_t - \Gamma_{l_t}z_t\right)\left(x_t - \Gamma_{l_t}z_t\right)'\rangle_{p(z_t,c_t,l_t|x_{1:T})}\right] \quad (10)$$

m_c is re-estimated by solving the following quartic equation:

$$0 = \sum_t\langle\hat{\sigma}_c^{-2}(a_t - \widehat{m}_c)\rangle_{p(c_t=c,a_t|x_{1:T})} - \left(M\sum_t\gamma_{t,c}\right)\widehat{m}_c^{-1} + \sum_t\langle\widehat{m}_c^{-3}tr\left[\left(x_t - \Gamma_{l_t}z_t\right)\left(x_t - \Gamma_{l_t}z_t\right)'\widehat{\Psi}^{-1}\right]\rangle_{p(z_t,c_t=c,l_t|x_{1:T})} \quad (11)$$

$$\hat{\sigma}_c^2 = \frac{\sum_t \gamma_{t,c}\langle a_t^2 - 2a_t\widehat{m}_c\rangle_{p(a_t|\,x_{1:T},c_t=c)}}{\sum_t\gamma_{t,c}} + \widehat{m}_c^2 \quad (12)$$

5 E-step

It can be derived from Equ. (1) that the conditional distributions $p(z_t|x_t, c_t, l_t, a_t)$, $p(x_t|c_t, l_t, a_t)$ and $p(a_t|c_t, l_t, x_t)$ are all Gaussian distributions, whose conditional expectation and conditional covariance are given as follows:

$$Cov(z_t|x_t, c_t, l_t, a_t) = \left(\Phi_{c_t}^{-1} + \Gamma_{l_t}^T \Psi^{-1} \Gamma_{l_t}\right)^{-1} \triangleq \Omega_{cl} \tag{13}$$

$$E(z_t|x_t, c_t, l_t, a_t) = \Omega_{cl}\left(\Gamma_{l_t}' \Psi^{-1} x_t + a_t \Phi_{c_t}^{-1} \mu_{c_t}\right) \triangleq \zeta_{cl} \tag{14}$$

$$Cov(x_t|c_t, l_t, a_t) = \Gamma_{l_t} \Phi_{c_t} \Gamma_{l_t}' + \Psi \triangleq G_{cl} \tag{15}$$

$$E(x_t|c_t, l_t, a_t) = \Gamma_{l_t} \mu_{c_t} \triangleq f_{cl} \tag{16}$$

$$Cov(a_t|c_t, l_t, x_t) = (f_{cl}' G_{cl}^{-1} f_{cl} + \sigma_c^{-2})^{-1} \triangleq K_{cl} \tag{17}$$

$$E(a_t|c_t, l_t, x_t) = K_{cl}(f_{cl}' G_{cl}^{-1} x_t + \sigma_c^{-2} m_c) \triangleq j_{cl} \tag{18}$$

The posteriori of c_t and l_t can be calculated using forward-backward algorithm:

$$p(c_t, l_t|x_{1:T}) \propto \alpha(c_t, l_t)\beta(c_t, l_t) \tag{19}$$

where

$$\alpha(c_t, l_t) \triangleq p(c_t, l_t, x_{1:t}) \tag{20}$$

$$\beta(c_t, l_t) \triangleq p(x_{t+1:T}|c_t, l_t) \tag{21}$$

$\alpha(c_t, l_t)$ and $\beta(c_t, l_t)$ can be derived recursively. To reduce computation cost, we perform Viterbi approximation:

$$\alpha(c_t, l_t) = \max_{c_{t-1}, l_{t-1}} \alpha(c_{t-1}, l_{t-1}) p(l_t|c_t, l_{t-1}) p(c_t|c_{t-1}) p(x_t|c_t, l_t) \tag{22}$$

$$\beta(c_t, l_t) = \max_{c_{t+1}, l_{t+1}} \beta(c_{t+1}, l_{t+1}) p(l_{t+1}|c_{t+1}, l_t) \cdot p(c_{t+1}|c_t) p(x_{t+1}|c_{t+1}, l_{t+1}) \tag{23}$$

where $p(x_t|c_t, l_t)$ can be solved using Equ. (15)(16):

$$p(x_t|c_t, l_t) = \int_{a_t} da_t \, p(a_t|c_t) p(x_t|c_t, l_t, a_t)$$

$$= \sigma_{c_t}^{-1} K_{cl}^{1/2} (2\pi)^{-N/2} |G_{cl}|^{-1/2} \exp\left[\frac{1}{2}\left(K_{cl}^{-1} j_{cl}^2 - \sigma_{c_t}^{-2} m_{c_t}^2 - x_t' G_{cl}^{-1} x_t\right)\right] \tag{24}$$

The recursions as shown in Equ. (22) and (23) are initialized as follows:

$$\alpha(c_1, l_1) = p(l_1|c_1) p(c_1) p(x_1|c_1, l_1) \tag{25}$$

$$\beta(c_T, l_T) = 1 \tag{26}$$

With these probability distributions, we can first compute several basic statistics, which is then used to derive the statistics for the M-step.

$$\bar{\bar{a}}_t = j_{cl} \tag{27}$$

$$\overline{\overline{a_t^2}} = j_{cl}^2 + K_{cl} \tag{28}$$

$$\bar{\bar{z}}_t = \Omega_{cl}\left(\Gamma_{l_t}' \Psi^{-1} x_t + j_{cl} \Phi_{c_t}^{-1} \mu_{c_t}\right) \tag{29}$$

$$\overline{\overline{z_t z_t'}} = \Omega_{cl} + \Omega_{cl}\left\{K_{cl}\left(\Phi_{c_t}^{-1}\mu_{c_t}\right)\left(\Phi_{c_t}^{-1}\mu_{c_t}\right)'\right\}\Omega_{cl} \tag{30}$$

$$\overline{\overline{a_t z_t}} = \Omega_{cl}\left[j_{cl}\Gamma_{l_t}^T\Psi^{-1}x_t + (j_{cl}^2 + K_{cl})\Phi_{c_t}^{-1}\mu_{c_t}\right] \tag{31}$$

Now we can calculate the statistics needed in M-step. Expectations in Equ. (7) are calculated as

$$\gamma_{t,c}\langle a_t^2\rangle_{p(a_t|\, x_{1:T},c_t=c)} = \sum_l p(c_t = c, l_t = l|x_{1:T})\overline{\overline{a_t^2}} \tag{32}$$

$$\gamma_{t,c}\langle a_t z_t\rangle_{p(a_t,z_t|x_{1:T},c_t=c)} = \sum_l p(c_t = c, l_t = l|x_{1:T})\overline{\overline{a_t z_t}} \tag{33}$$

Expectation in Equ. (9) is calculated as

$$\begin{aligned}&\gamma_{t,c}\langle(z_t - a_t\hat{\mu}_c)(z_t - a_t\hat{\mu}_c)'\rangle_{p(z_t,a_t|x_{1:T},c_t=c)}\\ &= \sum_l p(c_t = c, l_t = l|x_{1:T})\left(\overline{\overline{z_t z_t'}} - \hat{\mu}_c\overline{\overline{a_t z_t}}' - \overline{\overline{a_t z_t}}\hat{\mu}_c' + \overline{\overline{a_t^2}}\hat{\mu}_c\hat{\mu}_c'\right)\end{aligned} \tag{34}$$

Expectation in Equ. (10) is calculated as

$$\begin{aligned}&\langle\widehat{m}_{c_t}^{-2}\left(x_t - \Gamma_{l_t}z_t\right)\left(x_t - \Gamma_{l_t}z_t\right)'\rangle_{p(z_t,c_t,l_t|x_{1:T})}\\ &= \sum_{c,l} p(c_t = c, l_t = l|x_{1:T})m_c^{-2}\left\{x_t x_t' - \Gamma_l\overline{\overline{z}}_t x_t' - x_t(\Gamma_l\overline{\overline{z}}_t)' + \Gamma_l\overline{\overline{z_t z_t'}}\Gamma_l'\right\}\end{aligned} \tag{35}$$

Expectations in Equ. (11) are calculated as

$$\langle\hat{\sigma}_c^{-2}(a_t - \widehat{m}_c)\rangle_{p(c_t=c,a_t|x_{1:T})} = \sum_l p(c_t = c, l_t = l|x_{1:T})\hat{\sigma}_c^{-2}(\overline{\overline{a}}_t - \widehat{m}_c) \tag{36}$$

$$\begin{aligned}&\langle\widehat{m}_c^{-3}tr\left[\left(x_t - \Gamma_{l_t}z_t\right)\left(x_t - \Gamma_{l_t}z_t\right)'\widehat{\Psi}^{-1}\right]\rangle_{p(z_t,c_t=c,l_t|x_{1:T})}\\ &= \sum_l p(c_t = c, l_t = l|x_{1:T})m_c^{-3}\widehat{\Psi}^{-1}\left\{x_t x_t' - \Gamma_l\overline{\overline{z}}_t x_t' - x_t(\Gamma_l\overline{\overline{z}}_t)' + \Gamma_l\overline{\overline{z_t z_t'}}\Gamma_l'\right\}\end{aligned} \tag{37}$$

Expectation in Equ. (12) is calculated as

$$\gamma_{t,c}\langle a_t^2 - 2a_t\widehat{m}_c\rangle_{p(a_t|\, x_{1:T},c_t=c)} = \sum_l p(c_t = c, l_t = l|x_{1:T})\hat{\sigma}_c^{-2}\left(\overline{\overline{a_t^2}} - 2\overline{\overline{a}}_t\widehat{m}_c\right) \tag{38}$$