Variational Bayes

Xinwei Qu

2025 年 11 月 10 日

目录

1 Common Distributions

1.1 Gamma Distribution

Supposing we have a gamma distribution $p(x; a, b)$. a表示事件次数，b表示每次发生的概率。The probability density function (PDF) of gamma distribution is

$$p(x) = \frac{x^{(a-1)} b^a e^{-bx}}{\Gamma(a)}$$

$$\ln p(x) \propto (a-1) ln(x) - bx$$

The mean is

$$\mu_x = \frac{a}{b}$$

The variance is

$$\sigma_x^2 = \frac{a}{b^2} \tag{1}$$

1.2 Complex Gaussian Distribution

- For a complex Gaussian distributed variable x, its PDF is

$$p(x) = \frac{1}{\pi\sigma^2} e^{-\frac{|x-\mu|^2}{\sigma^2}}$$

$$ln p(x) \propto -ln(\sigma^2) - \sigma^{-2}|x-\mu|^2 \tag{2}$$

- For a complex Gaussian distributed variable vector $x = [x_0, \cdots, x_{N-1}]$, its PDF is

$$p(x) = \frac{1}{\pi^N det(\Sigma)} e^{-(x-\mu)^H \Sigma^{-1}(x-\mu)}$$

$$ln p(x) \propto -ln det(\Sigma) - (x-\mu)^H \Sigma^{-1}(x-\mu) \tag{3}$$

  where Σ is the covariance matrix.

2 Variational Bayes

For a telecommunication system, we have

$$y = Hx + z, \tag{4}$$

where y is the received signal, x is the transmitted signal and $z \in \mathcal{CN}(0, \sigma^2)$. Please note that,

$$x = x_p + x_d, \tag{5}$$

where x_p is the pilot and x_d is data.

2.1 Bayes Interference

In the Rx, given the prior $p(y)$, we compute posterior distribution $p(y|x)$,

$$p(x|y) = \frac{p(x,y)}{p(y)} = \frac{\overbrace{p(y|x)}^{likelihood}\ \overbrace{p(x)}^{prior}}{\underbrace{p(y)}_{evidence}} = \frac{p(y|x)p(x)}{\int_y p(x,y)dy} \tag{6}$$

Usually, we assume the evidence is 100%, i.e., $p(y) = 1$. Hence,

$$p(x|y) \propto \overbrace{p(y|x)}^{likelihood}\ \overbrace{p(x)}^{prior} \tag{7}$$

Here, we need to choose the likelihood and the prior.

2.2 Variational Interference

However, the posterior may have no closed form, i.e., computing $p(x|y)$ is not feasible. Instead, we use a distribution Q over the symbols x to approximate $p(x|y)$, i.e.,

$$\begin{aligned} q^*(x) &= \arg\min_{q(x)\in Q} KL(q(x)||p(x|y)) \\ &= \arg\min_{q(x)\in Q} \int_x q(x)\ln\frac{q(x)}{p(x|y)}dx \\ &= \arg\min_{q(x)\in Q} -\int_x q(x)\ln\frac{p(x|y)}{q(x)}dx \end{aligned} \tag{8}$$

Here, $q^*(x)$ is the optimal $q(x)$ but $p(x|y)$ is unknown. Herefore,

$$KL(q(x)||p(x|y)) = -\int_x q(x)\ln\frac{p(x|y)}{q(x)}dx \tag{9}$$

$$= \int_x q(x)\ln q(x)dx - \int_x q(x)\ln p(x|y)dx \tag{10}$$

$$= \int_x q(x)\ln q(x)dx - \int_x q(x)\ln\frac{p(x,y)}{p(y)}dx \tag{11}$$

$$= \int_x q(x)\ln q(x)dx - \int_x q(x)\ln p(x,y)dx + \int_x q(x)\ln p(y)dx \tag{12}$$

$$= \int_x q(x)\ln q(x)dx - \int_x q(x)\ln p(x,y)dx + \ln p(y)\int_x q(x)dx \tag{13}$$

$$= \underbrace{\mathbb{E}_q[\ln q(x)] - \mathbb{E}_q[\ln p(x,y)]}_{-ELBO} + \ln p(y) \tag{14}$$

$$= -ELBO(q) + \ln p(y) \tag{15}$$

2.2.1 ELBO

Here, ELBO is Evidence Lower Bound, i.e.,

$$ELBO(q) = \mathbb{E}_q[\ln p(x,y)] - \mathbb{E}_q[\ln q(x)] \tag{16}$$

$$= \int_x q(x)\ln\frac{\overbrace{p(x,y)}^{known}}{q(x)}dx \tag{17}$$

$$= \int_x q(x)\ln\frac{p(y|x)p(x)}{q(x)}dx \tag{18}$$

(15) can be rewritten as,

$$\begin{aligned} \underbrace{\ln p(y)}_{\text{CONST}} &= ELBO(q) + \underbrace{KL(q(x)||p(x|y))}_{\geq 0} \\ &\geq ELBO(q) \end{aligned} \tag{19}$$

The minimizing KL can be taken as the maximizing ELBO, i.e.,

$$\begin{aligned} q^*(x) &= \arg\min_{q(x)\in Q} KL(q(x)||p(x|y)) \\ &= \arg\max_{q(x)\in Q} ELBO(q) \end{aligned} \tag{20}$$

2.3 Mean Field

Now, we know the problem has been simplified as the maximizing ELBO. Here, we use the mean-field assumption, i.e.,

$$\begin{aligned} q(x) &= \prod_{i=1}^{m} q_i(x_i) \\ \ln q(x) &= \sum_{i=1}^{m} \ln q_i(x_i) \\ \mathbb{E}_q[\ln q(x)] &= \sum_{i=1}^{m} \mathbb{E}_{q_i}[\ln q_i(x_i)] \end{aligned} \tag{21}$$

2.3.1 Coordinate Ascent Optimization

In $q = [q_1, q_2, \cdots, q_j, \cdots, q_m]$, we fix others to update q_j, i.e.,

$$\begin{aligned} q_j^*(x_j) &= \arg\min_{q_j} ELBO(q_j) \\ &= \frac{\exp\{\mathbb{E}_{q_{-j}}[\ln p(x, y)]\}}{\int_{x_j} \exp\{\mathbb{E}_{q_{-j}}[\ln p(x, y)]\} dx_j} \end{aligned} \tag{22}$$

To prove (22), we need to load (21) into (16),

$$ELBO(q) = \mathbb{E}_q[\ln p(x, y)] - \mathbb{E}_q[\ln q(x)] \tag{23}$$

$$= \int_x q(x)\ln p(x, y)dx - \left[\mathbb{E}_{q_j}[\ln q_j(x_j)] + \sum_{i\neq j} \mathbb{E}_{q_i}[\ln q_i(x_i)]\right] \tag{24}$$

Here, $\sum_{i\neq j} \mathbb{E}_{q_i}[\ln q_i(x_i)]$ can be seen as a constant because it is not related to q_j. Therefore, (24) can be simplified as ($*_{-j}$ represents the other elements except j),

$$ELBO(q) = \int_x q(x)\ln p(x, y)dx - \mathbb{E}_{q_j}[\ln q_j(x_j)] + \text{const} \tag{25}$$

$$= \int_x q(x)\ln p(x, y)dx - \int_{x_j} q(x_j)\ln q_j(x_j)dx_j + \text{const} \tag{26}$$

$$= \int_{x_j}\int_{x_{-j}} q(x_j)q(x_{-j})\ln p(x, y)dx_j dx_{-j} - \int_{x_j} q(x_j)\ln q_j(x_j)dx_j + \text{const} \tag{27}$$

$$= \int_{x_j} q(x_j)\left[\int_{x_{-j}} q(x_{-j})\ln p(x, y)dx_{-j}\right] dx_j - \int_{x_j} q(x_j)\ln q_j(x_j)dx_j + \text{const} \tag{28}$$

$$= \int_{x_j} q(x_j)\mathbb{E}_{q_{-j}}[\ln p(x, y)]dx_j - \int_{x_j} q(x_j)\ln q_j(x_j)dx_j + \text{const} \tag{29}$$

Here, we define a new distribution

$$\begin{aligned} \ln\tilde{p_j}(x_j, y) &= \mathbb{E}_{q_{-j}}[\ln p(x, y)] + const \\ \tilde{p_j}(x_j, y) &\propto \exp\{\mathbb{E}_{q_{-j}}[\ln p(x, y)]\} \end{aligned} \tag{30}$$

Here, we load (30) into (29),

$$ELBO(q) = \int_{x_j} q(x_j)\ln\tilde{p_j}(x_j, y)dx_j - \int_{x_j} q(x_j)\ln q_j(x_j)dx_j + \text{const} \tag{31}$$

$$= \int_{x_j} q(x_j)\ln\frac{\tilde{p_j}(x_j, y)}{\ln q_j(x_j)}dx_j + \text{const} \tag{32}$$

$$= -KL(q_j(x_j)||\tilde{p_j}(x_j, y)) \tag{33}$$

The KL divergence reaches the minimum when

$$\begin{aligned} q^*_{x_j} &= \tilde{p_j}(x_j, y) \\ &\propto \exp\{\mathbb{E}_{q_{-j}}[\ln p(x,y)]\} \\ &= \frac{\exp\{\mathbb{E}_{q_{-j}}[\ln p(x,y)]\}}{\int_{x_j} \exp\{\mathbb{E}_{q_{-j}}[\ln p(x,y)]\} dx_j} \end{aligned} \tag{34}$$

$\int_{x_j} \exp\{\mathbb{E}_{q_{-j}}[\ln p(x,y)]\} dx_j$是为了让总体概率为1

2.4 Algorithm Structure

The structure is given as below:

initialize $q_j(x_j)$ for $j \in 1, \cdots, m$
while ELBO not converge **do**
    for $j \in 1, \cdots, m$ **do**
        $q^*_{x_j} = \frac{\exp\{\mathbb{E}_{q_{-j}}[\ln p(x,y)]\}}{\int_{x_j} \exp\{\mathbb{E}_{q_{-j}}[\ln p(x,y)]\} dx_j}$
    end for
    ELBO(q)= $\mathbb{E}_q[\ln p(x,y)] - \mathbb{E}_q[\ln q(x)]$
end while
return $q(x)$

3 System Model

We first introduce the OTFS mod/demod and its frame structure. Subsequently, we derive the input-output relation in the delay-Doppler (DD) domain for the two most widely adopted pulse-shaping waveforms.

3.1 OTFS System

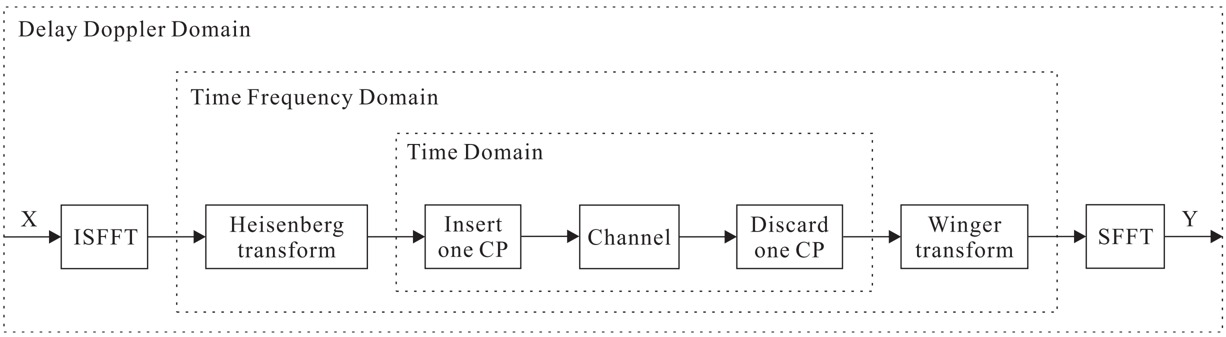

图 1: OTFS mode/demod

We consider a single input single output (SISO) OTFS system as illustrated in Fig. 1. The transmitter operates an OTFS frame (detailed in 3.2), $\mathbf{X}[k,l] \in \mathbb{C}^{K \times L}$, with $k = 0, \cdots, K-1$ and $l = 0, \cdots, L-1$ indexing discretized Doppler and delay shifts, respectively. After transposition. the frame is converted to the time-frequency (TF) domain via the inverse symplectic finite Fourier transform (ISFFT), mapping the data on $L \times K$ grids with uniform intervals Δf (Hz) and $T = 1/\Delta f$ (seconds). The time-domain signal is synthesized using (discrete) Heisenberg transform with a pulse-shaping waveform employing a single initial cyclic prefix spanning the full OTFS frame duration. The time-domain signal is transmitted over a time-varying wireless channel characterized by the delay-Doppler impulse response $h(\tau, v)$ as [1],

$$h(\tau, v) = \sum_{i=1}^{P} h_i \delta(\tau - \tau_i)\delta(v - v_i), \tag{35}$$

where $\delta(\cdot)$ denotes the Dirac delta function, $h_i \sim \mathcal{N}(0, \frac{1}{P})$ is the gain of the i-th propagation path, and P represents the total number of paths. Each path is characterized by distinct delay and/or Doppler shifts, modeling the channel response between the receiver and either moving reflectors or the transmitting source. The delay and Doppler shifts are given as,

$$\tau_i = l_i \frac{T}{L}, v_i = k_i \frac{\Delta f}{K}, \tag{36}$$

respectively. Let the integers $l_i \in [0, l_{\max}]$ and $k_i \in [-k_{\max}, k_{\max}]$ represent the delay and Doppler shift indices, respectively, where $l_{\max}$ and $k_{\max}$ denote the maximum delay index and maximum Doppler shift index across all propagation paths. Note that we restrict our consideration to integer-valued indices, as fractional delay and Doppler shifts can be equivalently represented through virtual integer taps in the delay-Doppler domain using the techniques described in [2–4].

3.2 OTFS Frame Structure

As illustrated in Fig. 2, a superimposed OTFS frame structure is considered, where pilot and data symbols are jointly embedded over delay-Doppler grids, i.e.,

$$\mathbf{X} = \mathbf{X}_d + \mathbf{X}_p, \tag{37}$$

where $\mathbf{X}_d[k, l] \in \mathbb{C}^{K \times L}$ denotes the data frame composed of quadrature amplitude modulation (QAM) symbols drawn from a constellation $\mathcal{A}$ with average energy E_d. The pilot frame $\mathbf{X}_p[k, l]$ contains nonzero elements only at designated positions, i.e.,

$$\mathbf{X}_p[k, l] = \begin{cases} x_p, & k = k_p,\ l = l_p, \\ 0, & \text{otherwise}, \end{cases} \tag{38}$$

where x_p is the pilot symbol with energy E_p, $k_p = \lfloor (K-1)/2 \rfloor$ is the Doppler index of all pilots, and $l_p = i(l_{\max} + 1)$ for $i = 0, \ldots, N_p - 1$ are their delay indices. Here, $N_p = \lfloor L/(l_{\max} + 1) \rfloor$ denotes the total number of pilots. Each pilot facilitates channel estimation over a region of size $K \times (l_{\max} + 1)$ in the DD domain.

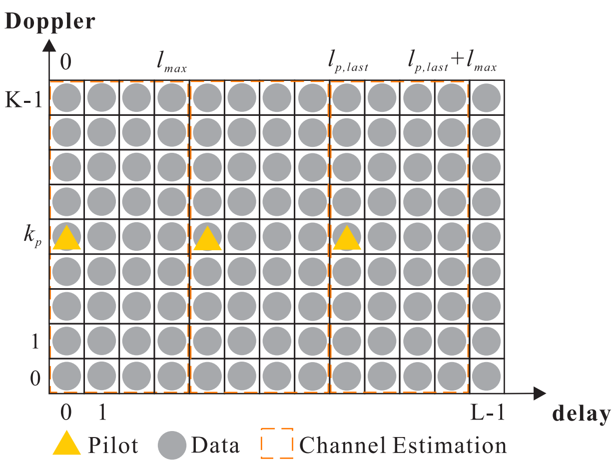

图 2: OTFS frame structure with the last pilot delay index at $l_{p,\text{last}} = (N_p - 1)(l_{\max} + 1)$

4 Variational Bayes in QJCHESD

4.1 OTFS Channel Estimation using VB

We assume the channel follows the Gaussian distribution, i.e.,

$$p(h|\gamma) = \prod_{i=0}^{P_{\max}-1} p(h_i|\gamma_i) = \prod_{i=0}^{P_{\max}-1} \mathcal{CN}(h_i; 0, \gamma_i^{-1}), \tag{39}$$

where $P_{\max} = (l_{\max} + 1)(2k_{\max} + 1)$, $\gamma = [\gamma_0, \gamma_1, \cdots, \gamma_{P_{\max}-1}]$ is the precision vector of h. γ follows the Gamma distribution, i.e.,

$$p(\gamma) = \text{Gamma}(\gamma; a, b) \tag{40}$$

where a is the shape parameter and b is the inverse scale parameter. Also, we assume the noise obeys the Gaussian distribution, i.e.,

$$p(z) = \mathbf{z} \sim \mathcal{CN}(0, \alpha^{-1}\boldsymbol{I}) \tag{41}$$

where α obeys the Gamma distribution, i.e.,

$$p(\alpha) = \text{Gamma}(\alpha; c, d) \tag{42}$$

where c is the shape parameter and d is the inverse scale parameter. For all parameters, we define $\Theta = \{h, \gamma, \alpha\}$.

Here, we use the mean field assumption to estimate the channel, i.e.,

$$\begin{aligned} p(y, \Theta) &= p(y|h, \alpha)p(h|\gamma)p(\gamma)p(\alpha) \\ lnp(y, \Theta) &= lnp(y|h, \alpha) + lnp(h|\gamma) + lnp(\gamma) + lnp(\alpha) \end{aligned} \tag{43}$$

Here, we use a distribution family Q over Θ to approximate $p(\Theta|y)$, i.e.,

$$\begin{aligned} p(\Theta|y) &= \underset{q(\Theta)\in Q}{\arg\min}\, KL(q(\Theta)||p(\Theta|y)) \\ &= \underset{q(\Theta)\in Q}{\arg\min} - \int_{\Theta} q(\Theta)\ln\frac{p(\Theta|y)}{q(\Theta)} d\Theta \end{aligned} \tag{44}$$

where $q(\Theta)$ follows the mean field assumption, i.e.,

$$q(\Theta) = q(h)q(\gamma)q(\alpha) \tag{45}$$

Therefore, we can update the probability functions as follows:

$$q^{(t+1)}(\alpha) \propto exp(\mathbb{E}_{q_{-\alpha}^{(t)}}[\ln p(y, \Theta)]\}) \tag{46}$$

$$q^{(t+1)}(h) \propto exp(\mathbb{E}_{q_{-h}^{(t)}}[\ln p(y, \Theta)]\}) \tag{47}$$

$$q^{(t+1)}(\gamma) \propto exp(\mathbb{E}_{q_{-\gamma}^{(t)}}[\ln p(y, \Theta)]\}) \tag{48}$$

The update is computed as follows

1) Update $q(\alpha)$

$$q^{(t+1)}(\alpha) \propto exp(\mathbb{E}_{q_{-\alpha}^{(t)}}[\ln p(y, \Theta)]\}) \tag{49}$$

$$lnq^{(t+1)}(\alpha) \propto \mathbb{E}_{q_{-\alpha}^{(t)}}[\ln p(y, \Theta)] \tag{50}$$

$$lnq^{(t+1)}(\alpha) \propto \mathbb{E}_{q_h^{(t)}}[\ln p(y|h, \alpha)] + lnp(\alpha) \tag{51}$$

Here,

$$p(y|h, \alpha) = \mathcal{CN}(\boldsymbol{y}_p; \boldsymbol{\Phi}_p\boldsymbol{h}, \alpha^{-1}\boldsymbol{I})$$

$$= \frac{1}{\pi^Z det(\alpha^{-1}\boldsymbol{I})} e^{-(\boldsymbol{y}_p - \boldsymbol{\Phi}_p\boldsymbol{h})^H(\alpha^{-1}\boldsymbol{I})^{-1}(\boldsymbol{y}_p - \boldsymbol{\Phi}_p\boldsymbol{h})} \tag{52}$$

$$= \frac{1}{\pi^Z det(\alpha^{-1}\boldsymbol{I})} e^{-\alpha(\boldsymbol{y}_p - \boldsymbol{\Phi}_p\boldsymbol{h})^H(\boldsymbol{y}_p - \boldsymbol{\Phi}_p\boldsymbol{h})} \tag{53}$$

$$= \frac{1}{\pi^Z det(\alpha^{-1}\boldsymbol{I})} e^{-\alpha\|\boldsymbol{y}_p - \boldsymbol{\Phi}_p\boldsymbol{h}\|^2} \tag{54}$$

where Z is the dimension of $\boldsymbol{y}_p$. Here,

$$det(\alpha^{-1}\boldsymbol{I}) = (\alpha^{-1})^Z = \alpha^{-Z} \tag{55}$$

Therefore,

$$lnp(y|h,\alpha) \propto Zln(\alpha) - \alpha\|\boldsymbol{y}_p - \boldsymbol{\Phi}_p\boldsymbol{h}\|^2 \tag{56}$$

$$\mathbb{E}_{q_h^{(t)}}\{lnp(y|h,\alpha)\} \propto \mathbb{E}_{q_h^{(t)}}\{Zln(\alpha)\} - \mathbb{E}_{q_h^{(t)}}\{\alpha\|\boldsymbol{y}_p - \boldsymbol{\Phi}_p\boldsymbol{h}\|^2\} \tag{57}$$

$$\mathbb{E}_{q_h^{(t)}}\{lnp(y|h,\alpha)\} \propto Zln(\alpha) - \alpha\mathbb{E}_{q_h^{(t)}}\{\|\boldsymbol{y}_p - \boldsymbol{\Phi}_p\boldsymbol{h}\|^2\} \tag{58}$$

Now, we need to get $\mathbb{E}_{q_h^{(t)}}\{\|\boldsymbol{y}_p - \boldsymbol{\Phi}_p\boldsymbol{h}\|^2$,i.e.,

$$\|\boldsymbol{y}_p - \boldsymbol{\Phi}_p\boldsymbol{h}\|^2 = \boldsymbol{y}_p^H\boldsymbol{y}_p - \boldsymbol{y}_p^H\boldsymbol{\Phi}_p\boldsymbol{h} - \boldsymbol{h}^H\boldsymbol{\Phi}_p^H\boldsymbol{y}_p + \boldsymbol{h}^H\boldsymbol{\Phi}_p^H\boldsymbol{\Phi}_p\boldsymbol{h} \tag{59}$$

$$\mathbb{E}_{q_h^{(t)}}\{\|\boldsymbol{y}_p - \boldsymbol{\Phi}_p\boldsymbol{h}\|^2 = \boldsymbol{y}_p^H\boldsymbol{y}_p - \boldsymbol{y}_p^H\boldsymbol{\Phi}_p\boldsymbol{\mu}_h - \boldsymbol{\mu}_h^H\boldsymbol{\Phi}_p^H\boldsymbol{y}_p + \mathbb{E}_{q_h^{(t)}}\{\boldsymbol{h}^H\boldsymbol{\Phi}_p^H\boldsymbol{\Phi}_p\boldsymbol{h}\} \tag{60}$$

Here, as in [5],

$$\mathbb{E}_{q_h^{(t)}}\{\boldsymbol{h}^H\boldsymbol{\Phi}_p^H\boldsymbol{\Phi}_p\boldsymbol{h}\} = \boldsymbol{\mu}_p^H\boldsymbol{\Phi}_p^H\boldsymbol{\Phi}_p\boldsymbol{\mu}_p + \mathrm{tr}(\boldsymbol{\Phi}_p^H\boldsymbol{\Phi}_p\boldsymbol{\Sigma}_h) \tag{61}$$

$$= \boldsymbol{\mu}_p^H\boldsymbol{\Phi}_p^H\boldsymbol{\Phi}_p\boldsymbol{\mu}_p + \mathrm{tr}\left(\boldsymbol{\Phi}_p\boldsymbol{\Sigma}_h\boldsymbol{\Phi}_p^H\right) \tag{62}$$

Therefore,

$$\mathbb{E}_{q_h^{(t)}}\{\|\boldsymbol{y}_p - \boldsymbol{\Phi}_p\boldsymbol{h}\|^2 = \boldsymbol{y}_p^H\boldsymbol{y}_p - \boldsymbol{y}_p^H\boldsymbol{\Phi}_p\boldsymbol{\mu}_h - \boldsymbol{\mu}_h^H\boldsymbol{\Phi}_p^H\boldsymbol{y}_p + \boldsymbol{\mu}_p^H\boldsymbol{\Phi}_p^H\boldsymbol{\Phi}_p\boldsymbol{\mu}_p + \mathrm{tr}(\boldsymbol{\Phi}_p^H\boldsymbol{\Phi}_p\boldsymbol{\Sigma}_h) \tag{63}$$

$$= \|\boldsymbol{y}_p - \boldsymbol{\Phi}_p\boldsymbol{\mu}_h\|^2 + \mathrm{tr}(\boldsymbol{\Phi}_p^H\boldsymbol{\Phi}_p\boldsymbol{\Sigma}_h) \tag{64}$$

Therefore,

$$lnq^{(t+1)}(\alpha) \propto Zln(\alpha) - \alpha\left(\|\boldsymbol{y}_p - \boldsymbol{\Phi}_p\boldsymbol{\mu}_h\|^2 + \mathrm{tr}(\boldsymbol{\Phi}_p^H\boldsymbol{\Phi}_p\boldsymbol{\Sigma}_h)\right) + lnp(\alpha) \tag{65}$$

$$\propto Zln(\alpha) - \alpha(\|\boldsymbol{y}_p - \boldsymbol{\Phi}_p\boldsymbol{\mu}_h\|^2 + \mathrm{tr}(\boldsymbol{\Phi}_p^H\boldsymbol{\Phi}_p\boldsymbol{\Sigma}_h)) + (a-1)ln(\alpha) - b\alpha \tag{66}$$

$$\propto (a+Z-1)ln(\alpha) - \alpha(b + \|\boldsymbol{y}_p - \boldsymbol{\Phi}_p\boldsymbol{\mu}_h\|^2 + \mathrm{tr}(\boldsymbol{\Phi}_p^H\boldsymbol{\Phi}_p\boldsymbol{\Sigma}_h)) \tag{67}$$

Therefore,

$$a^{(t+1)} = a^{(t)} + z \tag{68}$$

$$b^{(t+1)} = b^{(t)} + \|\boldsymbol{y}_p - \boldsymbol{\Phi}_p\boldsymbol{\mu}_h^{(t)})\|^2 + \mathrm{tr}(\boldsymbol{\Phi}_p^H\boldsymbol{\Phi}_p\boldsymbol{\Sigma}_h^{(t)}) \tag{69}$$

where $\Sigma_h^{(t)}$ and μ_h^t are the posterior covariance matrix and the posterior mean vector of $h^{(t)}$, which are both adjusted after updating $q(h)$. The mean of α is

$$\hat{\alpha}^{(t+1)} = \frac{a^{(t+1)}}{b^{(t+1)}} \tag{70}$$

2) Update $q(h)$

$$q^{(t+1)}(h) \propto exp(\mathbb{E}_{q_{-h}^{(t)}}[\ln p(y,\Theta)]\}) \tag{71}$$

$$lnq^{(t+1)}(h) \propto \mathbb{E}_{q_{-h}^{(t)}}\{\ln p(y,\Theta)\} \tag{72}$$

$$\propto \mathbb{E}_{q_{-h}^{(t)}}\{lnp(y|h,\alpha)\} + \mathbb{E}_{q_{-h}^{(t)}}\{lnp(h|\gamma^{(t)})\} \tag{73}$$

$$\propto -\mathbb{E}_{q_{-h}^{(t)}}\{\alpha\}\|\boldsymbol{y}_p - \boldsymbol{\Phi}_p\boldsymbol{h}\|^2 + \mathbb{E}_{q_{-h}^{(t)}}\{\boldsymbol{\gamma}^{(t)}\}\|\boldsymbol{h}\|^2 \tag{74}$$

$$\propto -\hat{\alpha}^{(t+1)}\|\boldsymbol{y}_p - \boldsymbol{\Phi}_p\boldsymbol{h}\|^2 - \boldsymbol{h}^H\text{diag}(\boldsymbol{\gamma}^{-1^{(t)}})^{-1}\boldsymbol{h} \tag{75}$$

$$\propto -\hat{\alpha}^{(t+1)}\|\boldsymbol{y}_p - \boldsymbol{\Phi}_p\boldsymbol{h}\|^2 - \boldsymbol{h}^H\text{diag}(\boldsymbol{\gamma}^{(t)})\boldsymbol{h} \tag{76}$$

$$\propto \hat{\alpha}^{(t+1)}\boldsymbol{h}^H\boldsymbol{\Phi}_p^H\boldsymbol{y}_p + \hat{\alpha}^{(t+1)}\boldsymbol{y}_p^H\boldsymbol{\Phi}_p\boldsymbol{h} - \hat{\alpha}^{(t+1)}\boldsymbol{h}^H\boldsymbol{\Phi}_h^H\boldsymbol{\Phi}_h\boldsymbol{h} - \boldsymbol{h}^H\text{diag}(\boldsymbol{\gamma}^{(t)})\boldsymbol{h} + \text{const} \tag{77}$$

$$\propto \hat{\alpha}^{(t+1)}\boldsymbol{h}^H\boldsymbol{\Phi}_p^H\boldsymbol{y}_p + \hat{\alpha}^{(t+1)}\boldsymbol{y}_p^H\boldsymbol{\Phi}_p\boldsymbol{h} - \boldsymbol{h}^H(\hat{\alpha}^{(t+1)}\boldsymbol{\Phi}_h^H\boldsymbol{\Phi}_h + \text{diag}(\boldsymbol{\gamma}^{(t)})\boldsymbol{h} + \text{const} \tag{78}$$

As in the assumption, $\boldsymbol{h}$ follows Guassian distribution, i.e.,

$$q^{(t+1)}(\boldsymbol{h}) = \mathcal{CN}\left(\boldsymbol{h}|\boldsymbol{\mu}_h^{(t+1)}, \boldsymbol{\Sigma}_h^{(t+1)}\right) \tag{79}$$

$$lnq^{(t+1)}(\boldsymbol{h}) \propto -(\boldsymbol{h} - \boldsymbol{\mu}_h^{(t+1)})^H\boldsymbol{\Sigma}_h^{(t+1)^{-1}}(\boldsymbol{h} - \boldsymbol{\mu}_h^{(t+1)}) \tag{80}$$

$$\propto \underbrace{\boldsymbol{h}^H\boldsymbol{\Sigma}_h^{(t+1)^{-1}}\boldsymbol{\mu}_h^{(t+1)} + \boldsymbol{\mu}_h^{(t+1)^H}\boldsymbol{\Sigma}_h^{(t+1)^{-1}}\boldsymbol{h}}_{linear} + \underbrace{\boldsymbol{h}^H\boldsymbol{\Sigma}_h^{(t+1)^{-1}}\boldsymbol{h}}_{quadratic} + \text{const} \tag{81}$$

Here, we can see that the covariance is

$$\boldsymbol{\Sigma}_h^{(t+1)} = \hat{\alpha}^{(t+1)}\boldsymbol{\Phi}_h^H\boldsymbol{\Phi}_h + \text{diag}(\boldsymbol{\gamma}^{(t)}) \tag{82}$$

Therefore, (78) can be written as

$$\begin{aligned} lnq^{(t+1)}(h) \propto \boldsymbol{h}^H\boldsymbol{\Sigma}_h^{(t+1)^{-1}}\left(\hat{\alpha}^{(t+1)}\boldsymbol{\Sigma}_h^{(t+1)}\boldsymbol{\Phi}_p^H\boldsymbol{y}_p\right) + \\ \left(\hat{\alpha}^{(t+1)}\boldsymbol{\Sigma}_h^{(t+1)}\boldsymbol{\Phi}_p^H\boldsymbol{y}_p\right)^H\boldsymbol{\Sigma}_h^{(t+1)^{-1}}\boldsymbol{h} - \\ \boldsymbol{h}^H\boldsymbol{\Sigma}_h^{(t+1)^{-1}}\boldsymbol{h} + \text{const} \end{aligned} \tag{83}$$

Please note that the Hermitian matrix $\boldsymbol{\Sigma}_h^{(t+1)^H} = \boldsymbol{\Sigma}_h^{(t+1)}$. Therefore, the mean is

$$\boldsymbol{\mu}^{(t+1)} = \hat{\alpha}^{(t+1)}\boldsymbol{\Sigma}_h^{(t+1)}\boldsymbol{\Phi}_p^H\boldsymbol{y}_p$$

3) Update $q(\gamma)$

$$\text{ln}q^{(t+1)}(\gamma) \propto \mathbb{E}_{q_{-\gamma}^{(t)}}\{\text{ln}p(y,\Theta)\} \tag{84}$$

$$\propto \mathbb{E}_{q_{-\gamma}^{(t)}}\{lnp(y|h,\alpha) + lnp(h|\gamma) + lnp(\gamma) + lnp(\alpha)\} \tag{85}$$

$$\propto \mathbb{E}_{q_{-\gamma}^{(t)}}\{lnp(h|\gamma) + lnp(\gamma)\} \tag{86}$$

$$\propto \sum_{l=0}^{P_{\max}-1}\left(ln(\gamma_l) - \gamma_l\|h_l^{(t+1)}\|^2 + (c-1)ln\gamma_l - d\gamma_l\right) \tag{87}$$

$$\propto \sum_{l=0}^{P_{\max}-1}\left(\underbrace{(c+1-1)}_{c^{t+1}}ln(\gamma_l) - \gamma_l\underbrace{(d + \|h_l^{(t+1)}\|^2)}_{d^{(t+1)}}\right) \tag{88}$$

where $\|h_l^{(t+1)}\|^2 = \Sigma_{h_l}^{t+1} + |\mu_{h_l}^{t+1}|^2$. Therefore,

$$\gamma = \text{diag}\left(\frac{c_0^{t+1}}{d_0^{(t+1)}}, \cdots, \frac{c_{P_{\max}}^{t+1}}{d_{P_{\max}}^{(t+1)}}\right) \tag{89}$$

4) Initial values

$$a^{(0)} = b^{(0)} = 1 \tag{90}$$

$$c^{(0)} = d^{(0)} = 1 \tag{91}$$

4.1.1 Simplified OTFS Channel Estimation using VB

If we assume the noise is known to us, the estimation process can be simplified as,

1) Update $q(h)$

$$q^{(t+1)}(h) \propto exp(\mathbb{E}_{q_{-h}^{(t)}}[\ln p(y,\Theta)]\}) \tag{92}$$

$$ln q^{(t+1)}(h) \propto \mathbb{E}_{q_{-h}^{(t)}}\{\ln p(y,\Theta)\} \tag{93}$$

$$\propto \mathbb{E}_{q_{-h}^{(t)}}\{lnp(y|h,\alpha)\} + \mathbb{E}_{q_{-h}^{(t)}}\{lnp(h|\gamma^{(t)})\} \tag{94}$$

$$\propto \alpha \boldsymbol{h}^H \boldsymbol{\Phi}_p^H \boldsymbol{y}_p + \alpha \boldsymbol{y}_p^H \boldsymbol{\Phi}_p \boldsymbol{h} - \boldsymbol{h}^H(\alpha \boldsymbol{\Phi}_h^H \boldsymbol{\Phi}_h + \mathrm{diag}(\boldsymbol{\gamma}^{(t)})\boldsymbol{h} + \mathrm{const} \tag{95}$$

As in the assumption, $\boldsymbol{h}$ follows Guassian distribution, i.e.,

$$q^{(t+1)}(\boldsymbol{h}) = \mathcal{CN}\left(\boldsymbol{h}|\boldsymbol{\mu}_h^{(t+1)}, \boldsymbol{\Sigma}_h^{(t+1)}\right) \tag{96}$$

$$ln q^{(t+1)}(\boldsymbol{h}) \propto -(\boldsymbol{h} - \boldsymbol{\mu}_h^{(t+1)})^H {\boldsymbol{\Sigma}_h^{(t+1)}}^{-1}(\boldsymbol{h} - \boldsymbol{\mu}_h^{(t+1)}) \tag{97}$$

$$\propto \underbrace{\boldsymbol{h}^H {\boldsymbol{\Sigma}_h^{(t+1)}}^{-1} \boldsymbol{\mu}_h^{(t+1)} + {\boldsymbol{\mu}_h^{(t+1)}}^H {\boldsymbol{\Sigma}_h^{(t+1)}}^{-1} \boldsymbol{h}}_{linear} + \underbrace{\boldsymbol{h}^H {\boldsymbol{\Sigma}_h^{(t+1)}}^{-1} \boldsymbol{h}}_{quadratic} + \mathrm{const} \tag{98}$$

Here, we can see that the covariance is

$$\boldsymbol{\Sigma}_h^{(t+1)} = \alpha \boldsymbol{\Phi}_h^H \boldsymbol{\Phi}_h + \mathrm{diag}(\boldsymbol{\gamma}^{(t)}) \tag{99}$$

$$\boldsymbol{\mu}^{(t+1)} = \alpha \boldsymbol{\Sigma}_h^{(t+1)} \boldsymbol{\Phi}_p^H \boldsymbol{y}_p \tag{100}$$

2) Update $q(\gamma)$

$$\ln q^{(t+1)}(\gamma) \propto \mathbb{E}_{q_{-\gamma}^{(t)}}\{\ln p(y,\Theta)\} \tag{101}$$

$$\propto \mathbb{E}_{q_{-\gamma}^{(t)}}\{lnp(y|h,\alpha) + lnp(h|\gamma) + lnp(\gamma) + lnp(\alpha)\} \tag{102}$$

$$\propto \mathbb{E}_{q_{-\gamma}^{(t)}}\{lnp(h|\gamma) + lnp(\gamma)\} \tag{103}$$

$$\propto \sum_{l=0}^{P_{\max}-1}\left(ln(\gamma_l) - \gamma_l \|h_l^{(t+1)}\|^2 + (c-1)ln\gamma_l - d\gamma_l\right) \tag{104}$$

$$\propto \sum_{l=0}^{P_{\max}-1}\left(\underbrace{(c+1-1)}_{c^{t+1}} ln(\gamma_l) - \gamma_l \underbrace{(d + \|h_l^{(t+1)}\|^2)}_{d^{(t+1)}}\right) \tag{105}$$

where $\|h_l^{(t+1)}\|^2 = \Sigma_{h_l}^{t+1} + |\mu_{h_l}^{t+1}|^2$. Therefore,

$$\gamma = \mathrm{diag}(\frac{c_0^{t+1}}{d_0^{(t+1)}}, \cdots, \frac{c_{P_{\max}}^{t+1}}{d_{P_{\max}}^{(t+1)}}) \tag{106}$$

4.2 Symbol Detection

In this section, we assume the channel is known. The target is to find the channel and the symbol to maximize the posterior, i.e.,

$$p(x; y, H, \sigma^2) = \arg\max_{x \in \Omega} p(y|x; H, \sigma^2) p(x) \tag{107}$$

Here, we use a

4.3 Joint Method

We shall look at some examples to solve this problem

$$||y_p - \phi_p h||^2 = (y_p - \phi_p h)^H (y_p - \phi_p h) \tag{108}$$
$$= y_p^H y_p - y_p^H \phi_p h - (\phi_p h)^H y_p + (\phi_p h)^H (\phi_p h) \tag{109}$$
$$= y_p^H y_p - y_p^H \phi_p h - (\phi_p h)^H y_p + h^H \phi_p^H \phi_p h \tag{110}$$

Here,

$$(y_p^H \phi_p h)^H = (\phi_p h)^H y_p$$

Therefore,

$$y_p^H \phi_p h + (\phi_p h)^H y_p = 2Re\{y_p^H \phi_p h\}$$
$$||y_p - \phi_p h||^2 = y_p^H y_p - 2Re\{y_p^H \phi_p h\} + h^H \phi_p^H \phi_p h$$

Now, we do the expectation for $||y_p - \phi_p h||^2$ on h,

$$< ||y_p - \phi_p h||^2 >_h = y_p^H y_p - 2Re\{y_p^H \phi_p u_h\} + < h^H \phi_p^H \phi_p h >_h \tag{111}$$
$$= y_p^H y_p - 2Re\{y_p^H \phi_p u_h\} + < tr(h^H \phi_p^H \phi_p h) >_h \tag{112}$$
$$= y_p^H y_p - 2Re\{y_p^H \phi_p u_h\} + < tr(\phi_p^H \phi_p h h^H) >_h \tag{113}$$
$$= y_p^H y_p - 2Re\{y_p^H \phi_p u_h\} + tr(\phi_p^H \phi_p < h h^H >_h) \tag{114}$$
$$\tag{115}$$

The covariance of h is,

$$\Sigma_h = < h h^H > - u_h u_h^H$$
$$< h h^H > = \Sigma_h + u_h u_h^H$$

Therefore,

$$< ||y_p - \phi_p h||^2 >_h = y_p^H y_p - 2Re\{y_p^H \phi_p u_h\} + tr(\phi_p^H \phi_p (\Sigma_h + u_h u_h^H)) \tag{116}$$
$$= y_p^H y_p - 2Re\{y_p^H \phi_p u_h\} + tr(\phi_p^H \phi_p \Sigma_h) + tr(\phi_p^H \phi_p u_h u_h^H) \tag{117}$$
$$= y_p^H y_p - 2Re\{y_p^H \phi_p u_h\} + tr(\phi_p^H \phi_p \Sigma_h) + tr(u_h^H \phi_p^H \phi_p u_h) \tag{118}$$
$$= y_p^H y_p - 2Re\{y_p^H \phi_p u_h\} + u_h^H \phi_p^H \phi_p u_h + tr(\phi_p^H \phi_p \Sigma_h) \tag{119}$$
$$= ||y_p - \phi_p u_h||^2 + tr(\phi_p \Sigma_h \phi_p^H) \tag{120}$$

References

[1] R. Hadani, S. Rakib, M. Tsatsanis, A. Monk, A. J. Goldsmith, A. F. Molisch, and R. Calderbank, "Orthogonal time frequency space modulation," in 2017 IEEE Wireless Communications and Networking Conference (WCNC), 2017, pp. 1–6.

[2] A. Fish, S. Gurevich, R. Hadani, A. M. Sayeed, and O. Schwartz, "Delay-doppler channel estimation in almost linear complexity," IEEE Transactions on Information Theory, vol. 59, no. 11, pp. 7632–7644, 2013.

[3] P. Raviteja, K. T. Phan, Q. Jin, Y. Hong, and E. Viterbo, "Low-complexity iterative detection for orthogonal time frequency space modulation," in 2018 IEEE Wireless Communications and Networking Conference (WCNC), 2018, pp. 1–6.

[4] P. Raviteja, Y. Hong, E. Viterbo, and E. Biglieri, "Practical pulse-shaping waveforms for reduced-cyclic-prefix otfs," IEEE Transactions on Vehicular Technology, vol. 68, no. 1, pp. 957–961, 2019.

[5] Wikipedia. Quadratic form (statistics). [Online]. Available: https://en.wikipedia.org/wiki/Quadratic_form_(statistics)