

# Variational Bayes

Xinwei Qu

2025 年 11 月 19 日

# 目录

<b>1 Common Distributions</b>	<b>3</b>
1.1 Gamma Distribution . . . . .	3
1.2 Complex Gaussian Distribution . . . . .	3
<b>2 Variational Bayes</b>	<b>3</b>
2.1 Bayes Interference . . . . .	3
2.2 Variational Interference . . . . .	4
2.2.1 ELBO . . . . .	4
2.3 Mean Field . . . . .	5
2.3.1 Coordinate Ascent Optimization . . . . .	5
2.4 Algorithm Structure . . . . .	6
<b>3 System Model</b>	<b>6</b>
3.1 OTFS System . . . . .	6
3.2 OTFS Frame Structure . . . . .	7
<b>4 Variational Bayes in QJCHESD</b>	<b>8</b>
4.1 OTFS Channel Estimation using VB . . . . .	8
4.1.1 Simplified OTFS Channel Estimation using VB . . . . .	11
4.2 Symbol Detection . . . . .	11
4.3 Joint Method . . . . .	12

# 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

$$\sigma_x^2 = \frac{a}{b^2} \quad (1)$$

## 1.2 Complex Gaussian Distribution

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

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

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

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

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

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

where  $\Sigma$  is the covariance matrix.

# 2 Variational Bayes

For a telecommunication system, we have

$$y = Hx + z, \quad (4)$$

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

$$x = x_p + x_d, \quad (5)$$

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  $p(y|x)$ ,

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

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

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

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.,

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

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

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

$$= \int_x q(x) \ln q(x) dx - \int_x q(x) \ln p(x|y) dx \quad (10)$$

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

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

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

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

$$= -ELBO(q) + \ln p(y) \quad (15)$$

### 2.2.1 ELBO

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

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

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

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

(15) can be rewritten as,

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

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

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

## 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.,

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

### 2.3.1 Coordinate Ascent Optimization

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

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

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

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

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

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  $j$ ),

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

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

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

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

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

Here, we define a new distribution

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

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

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

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

$$= -KL(q_j(x_j) || \tilde{p}_j(x_j, y)) \quad (33)$$

The KL divergence reaches the minimum when

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

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

## 2.4 Algorithm Structure

The structure is given as below:

---

```

1: initialize  $q_j(x_j)$  for  $j \in 1, \dots, m$ 
2: while ELBO not converge do
3:   for  $j \in 1, \dots, m$  do
4:      $q_{x_j}^* = \frac{\exp\{\mathbb{E}_{q_{-j}}[\ln p(x, y)]\}}{\int_{x_j} \exp\{\mathbb{E}_{q_{-j}}[\ln p(x, y)]\} dx_j}$ 
5:   end for
6:   ELBO(q)=  $\mathbb{E}_q[\ln p(x, y)] - \mathbb{E}_q[\ln q(x)]$ 
7: end while
8: 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

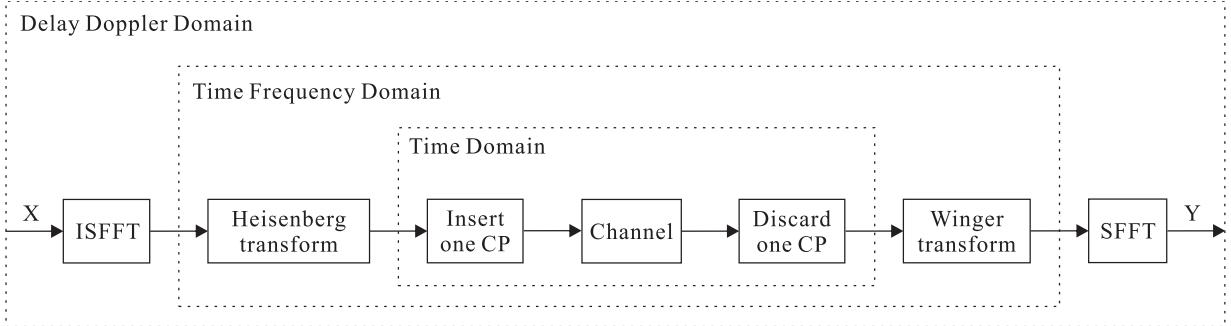


图 1: OTFS mode/demod

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  $k = 0, \dots, K - 1$  and  $l = 0, \dots, L - 1$  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  $\Delta f$  (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],

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

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,

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

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.,

$$\mathbf{X} = \mathbf{X}_d + \mathbf{X}_p, \quad (37)$$

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.,

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

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  $i = 0, \dots, N_p - 1$  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.

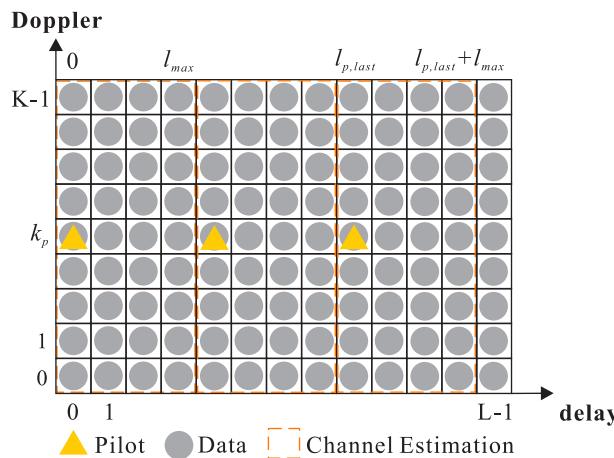


图 2: OTFS frame structure with the last pilot delay index at  $l_{p,\text{last}} = (N_p - 1)(l_{\max} + 1)$

## 4 Variational Bayes in QJCHESD

### 4.1 OTFS Channel Estimation using VB

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

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

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

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

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

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

where  $\alpha$  obeys the Gamma distribution, i.e.,

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

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.,

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

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

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

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

$$q(\Theta) = q(h)q(\gamma)q(\alpha) \quad (45)$$

Therefore, we can update the probability functions as follows:

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

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

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

The update is computed as follows

1) Update  $q(\alpha)$

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

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

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

Here,

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

$$= \frac{1}{\pi^Z \det(\alpha^{-1} \mathbf{I})} e^{-(\mathbf{y}_p - \Phi_p \mathbf{h})^H (\alpha^{-1} \mathbf{I})^{-1} (\mathbf{y}_p - \Phi_p \mathbf{h})} \quad (52)$$

$$= \frac{1}{\pi^Z \det(\alpha^{-1} \mathbf{I})} e^{-\alpha (\mathbf{y}_p - \Phi_p \mathbf{h})^H (\mathbf{y}_p - \Phi_p \mathbf{h})} \quad (53)$$

$$= \frac{1}{\pi^Z \det(\alpha^{-1} \mathbf{I})} e^{-\alpha \|\mathbf{y}_p - \Phi_p \mathbf{h}\|^2} \quad (54)$$

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

$$\det(\alpha^{-1} \mathbf{I}) = (\alpha^{-1})^Z = \alpha^{-Z} \quad (55)$$

Therefore,

$$\ln p(y|h, \alpha) \propto Z \ln(\alpha) - \alpha \|\mathbf{y}_p - \Phi_p \mathbf{h}\|^2 \quad (56)$$

$$\mathbb{E}_{q_h^{(t)}} \{\ln p(y|h, \alpha)\} \propto \mathbb{E}_{q_h^{(t)}} \{Z \ln(\alpha)\} - \mathbb{E}_{q_h^{(t)}} \{\alpha \|\mathbf{y}_p - \Phi_p \mathbf{h}\|^2\} \quad (57)$$

$$\mathbb{E}_{q_h^{(t)}} \{\ln p(y|h, \alpha)\} \propto Z \ln(\alpha) - \alpha \mathbb{E}_{q_h^{(t)}} \{\|\mathbf{y}_p - \Phi_p \mathbf{h}\|^2\} \quad (58)$$

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

$$\|\mathbf{y}_p - \Phi_p \mathbf{h}\|^2 = \mathbf{y}_p^H \mathbf{y}_p - \mathbf{y}_p^H \Phi_p \mathbf{h} - \mathbf{h}^H \Phi_p^H \mathbf{y}_p + \mathbf{h}^H \Phi_p^H \Phi_p \mathbf{h} \quad (59)$$

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

Here, as in [5],

$$\mathbb{E}_{q_h^{(t)}} \{\mathbf{h}^H \Phi_p^H \Phi_p \mathbf{h}\} = \boldsymbol{\mu}_p^H \Phi_p^H \Phi_p \boldsymbol{\mu}_p + \text{tr}(\Phi_p^H \Phi_p \boldsymbol{\Sigma}_h) \quad (61)$$

$$= \boldsymbol{\mu}_p^H \Phi_p^H \Phi_p \boldsymbol{\mu}_p + \text{tr}(\Phi_p \boldsymbol{\Sigma}_h \Phi_p^H) \quad (62)$$

Therefore,

$$\mathbb{E}_{q_h^{(t)}} \{\|\mathbf{y}_p - \Phi_p \mathbf{h}\|^2\} = \mathbf{y}_p^H \mathbf{y}_p - \mathbf{y}_p^H \Phi_p \boldsymbol{\mu}_h - \boldsymbol{\mu}_h^H \Phi_p^H \mathbf{y}_p + \boldsymbol{\mu}_p^H \Phi_p^H \Phi_p \boldsymbol{\mu}_p + \text{tr}(\Phi_p^H \Phi_p \boldsymbol{\Sigma}_h) \quad (63)$$

$$= \|\mathbf{y}_p - \Phi_p \boldsymbol{\mu}_h\|^2 + \text{tr}(\Phi_p^H \Phi_p \boldsymbol{\Sigma}_h) \quad (64)$$

Therefore,

$$\ln q^{(t+1)}(\alpha) \propto Z \ln(\alpha) - \alpha (\|\mathbf{y}_p - \Phi_p \boldsymbol{\mu}_h\|^2 + \text{tr}(\Phi_p^H \Phi_p \boldsymbol{\Sigma}_h)) + \ln p(\alpha) \quad (65)$$

$$\propto Z \ln(\alpha) - \alpha (\|\mathbf{y}_p - \Phi_p \boldsymbol{\mu}_h\|^2 + \text{tr}(\Phi_p^H \Phi_p \boldsymbol{\Sigma}_h)) + (a-1) \ln(\alpha) - b \alpha \quad (66)$$

$$\propto (a+Z-1) \ln(\alpha) - \alpha (b + \|\mathbf{y}_p - \Phi_p \boldsymbol{\mu}_h\|^2 + \text{tr}(\Phi_p^H \Phi_p \boldsymbol{\Sigma}_h)) \quad (67)$$

Therefore,

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

$$b^{(t+1)} = b^{(t)} + \|\mathbf{y}_p - \Phi_p \boldsymbol{\mu}_h^{(t)}\|^2 + \text{tr}(\Phi_p^H \Phi_p \boldsymbol{\Sigma}_h^{(t)}) \quad (69)$$

where  $\boldsymbol{\Sigma}_h^{(t)}$  and  $\boldsymbol{\mu}_h^{(t)}$  are the posterior covariance matrix and the posterior mean vector of  $h^{(t)}$ , which are both adjusted after updating  $q(h)$ . The mean of  $\alpha$  is

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

2) Update  $q(h)$

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

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

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

$$\propto -\mathbb{E}_{q_{-h}^{(t)}} \{ \alpha \} \| \mathbf{y}_p - \Phi_p \mathbf{h} \|^2 + \mathbb{E}_{q_{-h}^{(t)}} \{ \gamma^{(t)} \} \| \mathbf{h} \|^2 \quad (74)$$

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

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

$$\propto \hat{\alpha}^{(t+1)} \mathbf{h}^H \Phi_p^H \mathbf{y}_p + \hat{\alpha}^{(t+1)} \mathbf{y}_p^H \Phi_p \mathbf{h} - \hat{\alpha}^{(t+1)} \mathbf{h}^H \Phi_h^H \Phi_h \mathbf{h} - \mathbf{h}^H \text{diag}(\gamma^{(t)}) \mathbf{h} + \text{const} \quad (77)$$

$$\propto \hat{\alpha}^{(t+1)} \mathbf{h}^H \Phi_p^H \mathbf{y}_p + \hat{\alpha}^{(t+1)} \mathbf{y}_p^H \Phi_p \mathbf{h} - \mathbf{h}^H (\hat{\alpha}^{(t+1)} \Phi_h^H \Phi_h + \text{diag}(\gamma^{(t)})) \mathbf{h} + \text{const} \quad (78)$$

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

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

$$\ln q^{(t+1)}(\mathbf{h}) \propto -(\mathbf{h} - \boldsymbol{\mu}_h^{(t+1)})^H \boldsymbol{\Sigma}_h^{(t+1)^{-1}} (\mathbf{h} - \boldsymbol{\mu}_h^{(t+1)}) \quad (80)$$

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

Here, we can see that the covariance is

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

Therefore, (78) can be written as

$$\begin{aligned} \ln q^{(t+1)}(\mathbf{h}) &\propto \mathbf{h}^H \boldsymbol{\Sigma}_h^{(t+1)^{-1}} \left( \hat{\alpha}^{(t+1)} \boldsymbol{\Sigma}_h^{(t+1)} \Phi_p^H \mathbf{y}_p \right) + \\ &\quad \left( \hat{\alpha}^{(t+1)} \boldsymbol{\Sigma}_h^{(t+1)} \Phi_p^H \mathbf{y}_p \right)^H \boldsymbol{\Sigma}_h^{(t+1)^{-1}} \mathbf{h} - \\ &\quad \mathbf{h}^H \boldsymbol{\Sigma}_h^{(t+1)^{-1}} \mathbf{h} + \text{const} \end{aligned} \quad (83)$$

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

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

### 3) Update $q(\gamma)$

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

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

$$\propto \mathbb{E}_{q_{-\gamma}^{(t)}} \{ \ln p(h|\gamma) + \ln p(\gamma) \} \quad (86)$$

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

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

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

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

### 4) Initial values

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

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

#### 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  $q(h)$

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

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

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

$$\propto \alpha \mathbf{h}^H \Phi_p^H \mathbf{y}_p + \alpha \mathbf{y}_p^H \Phi_p \mathbf{h} - \mathbf{h}^H (\alpha \Phi_h^H \Phi_h + \text{diag}(\boldsymbol{\gamma}^{(t)}) \mathbf{h} + \text{const}) \quad (95)$$

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

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

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

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

Here, we can see that the covariance is

$$\boldsymbol{\Sigma}_h^{(t+1)} = (\alpha \Phi_h^H \Phi_h + \text{diag}(\boldsymbol{\gamma}^{(t)}))^{-1} \quad (99)$$

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

2) Update  $q(\gamma)$

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

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

$$\propto \mathbb{E}_{q_{-\gamma}^{(t)}}\{\ln p(h|\gamma) + \ln p(\gamma)\} \quad (103)$$

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

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

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

$$\gamma = \text{diag}\left(\frac{c_0^{t+1}}{d_0^{(t+1)}}, \dots, \frac{c_{P_{\max}}^{t+1}}{d_{P_{\max}}^{(t+1)}}\right) \quad (106)$$

## 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.,

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

Here, we use a

### 4.3 Joint Method

We shall look at some examples to solve this problem

$$\|y_p - \phi_p h\|^2 = (y_p - \phi_p h)^H (y_p - \phi_p h) \quad (108)$$

$$= y_p^H y_p - y_p^H \phi_p h - (\phi_p h)^H y_p + (\phi_p h)^H (\phi_p h) \quad (109)$$

$$= y_p^H y_p - y_p^H \phi_p h - (\phi_p h)^H y_p + h^H \phi_p^H \phi_p h \quad (110)$$

Here,

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

Therefore,

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

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

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

$$<\|y_p - \phi_p h\|^2>_h = y_p^H y_p - 2\text{Re}\{y_p^H \phi_p u_h\} + < h^H \phi_p^H \phi_p h >_h \quad (111)$$

$$= y_p^H y_p - 2\text{Re}\{y_p^H \phi_p u_h\} + < \text{tr}(h^H \phi_p^H \phi_p h) >_h \quad (112)$$

$$= y_p^H y_p - 2\text{Re}\{y_p^H \phi_p u_h\} + < \text{tr}(\phi_p^H \phi_p h h^H) >_h \quad (113)$$

$$= y_p^H y_p - 2\text{Re}\{y_p^H \phi_p u_h\} + \text{tr}(\phi_p^H \phi_p < h h^H >_h) \quad (114)$$

$$= y_p^H y_p - 2\text{Re}\{y_p^H \phi_p u_h\} + \Sigma_h \quad (115)$$

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,

$$<\|y_p - \phi_p h\|^2>_h = y_p^H y_p - 2\text{Re}\{y_p^H \phi_p u_h\} + \text{tr}(\phi_p^H \phi_p (\Sigma_h + u_h u_h^H)) \quad (116)$$

$$= y_p^H y_p - 2\text{Re}\{y_p^H \phi_p u_h\} + \text{tr}(\phi_p^H \phi_p \Sigma_h) + \text{tr}(\phi_p^H \phi_p u_h u_h^H) \quad (117)$$

$$= y_p^H y_p - 2\text{Re}\{y_p^H \phi_p u_h\} + \text{tr}(\phi_p^H \phi_p \Sigma_h) + \text{tr}(u_h^H \phi_p^H \phi_p u_h) \quad (118)$$

$$= y_p^H y_p - 2\text{Re}\{y_p^H \phi_p u_h\} + u_h^H \phi_p^H \phi_p u_h + \text{tr}(\phi_p^H \phi_p \Sigma_h) \quad (119)$$

$$= \|y_p - \phi_p u_h\|^2 + \text{tr}(\phi_p \Sigma_h \phi_p^H) \quad (120)$$

## 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 offs,” 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\\_\(statistics\)](https://en.wikipedia.org/wiki/Quadratic_form_(statistics))