TRATAMIENTO DIGITAL DE SEÑALES
Ingeniería de Telecomunicación (4º, 2º c)

Unidad 4ª: Los problemas generales gaussianos

Aníbal R. Figueiras Vidal
Jesús Cid Sueiro
Ángel Navia Vázquez

Área de Teoría de la Señal y Comunicaciones
Universidad Carlos III de Madrid

Ejercicio de Ampliación

A: *Dos v. a. independientes,* s: $G(0, v_s)$ *y* r: $G(0, v_r)$*, dan lugar a la observación:*
$x = s+r$ *(señal gaussiana en ruido gaussiano).*

Determínese el estimador ms de s.

Se está buscando $\hat{s}_{ms} = E\{s \mid x\}$

y no se conoce directamente $p(s \mid x)$*; pero sí*

$$p(s): \quad G(0, v_s)$$

y además

$$p(x): \quad G(0, v_s + v_r)$$

$$p(x \mid s): \quad G(s, v_r)$$

con lo que se puede determinar

$$p(s \mid x) = \frac{p(x \mid s)\, p(s)}{p(x)}$$

que será gaussiana, al serlo las demás:

$$p(s\,|\,x) = \frac{1}{\sqrt{2\pi}} \frac{\frac{1}{\sqrt{v_r}} \frac{1}{\sqrt{v_s}}}{\frac{1}{\sqrt{v_s + v_r}}} \exp\left[-\frac{1}{2} \left(\frac{(x-s)^2}{v_r} + \frac{s^2}{v_s} - \frac{x^2}{v_s + v_r} \right) \right]$$

e identificando exponentes:

$$\frac{\left(s - E\{s\,|\,x\}\right)^2}{v_{s|x}} = \frac{(x-s)^2}{v_r} + \frac{s^2}{v_s} - \frac{x^2}{v_s + v_r}$$

- *términos en* s^2:

$$\frac{1}{v_{s|x}} = \frac{1}{v_r} + \frac{1}{v_s}; \quad v_{s|x} = \frac{v_s v_r}{v_s + v_r}$$

- *términos en* $-2s$:

$$\frac{E\{s\,|\,x\}}{v_{s|x}} = \frac{x}{v_r}; \qquad E\{s\,|\,x\} = \frac{v_{s|x}}{v_r} x = \frac{v_s}{v_s + v_r} x$$

(no es preciso seguir)

de modo que

$$\hat{s}_{ms} = \frac{v_s}{v_s + v_r} x$$

sobre lo que puede observarse que:

- *es lineal*
- *atribuye a la señal una parte de la observación de acuerdo con los valores de potencia de señal y ruido.*

Debe recordarse también que, dada la gaussianidad, el valor encontrado es también $\hat{s}_{ma}$ *y* $\hat{s}_{map}$.

Ejercicio de Ampliación

A: *El mismo problema en el caso multidimensional toma la forma general*

$$\mathbf{x} = \mathbf{H}\mathbf{s}+\mathbf{r}$$

donde $\mathbf{s}$ *(N×1) es* $G(\mathbf{0},\mathbf{V}_s)$, $\mathbf{r}$ *(M×1) es* $G(\mathbf{0},\mathbf{V}_r)$, *y* $\mathbf{H}$ *es una matriz M×N, que se supondrá conocida.*

Establecer $\hat{\mathbf{s}}_{ms}$

Notando que

$\mathbf{x}$: $G(\mathbf{0}, \mathbf{H}\mathbf{V}_s\mathbf{H}^T+\mathbf{V}_r)$

$\mathbf{x}|\mathbf{s}$: $G(\mathbf{H}\mathbf{s}, \mathbf{V}_r)$

y procediendo, como en el ejercicio anterior, a identificar exponentes:

$$\left[\mathbf{s} - E\{\mathbf{s} \mid \mathbf{x}\}\right]^T \mathbf{V}_{s|x}^{-1}\left[\mathbf{s} - E\{\mathbf{s} \mid \mathbf{x}\}\right]=$$

$$=\left(\mathbf{x}\text{-}\mathbf{H}\mathbf{s}\right)^T \mathbf{V}_r^{-1}\left(\mathbf{x}\text{-}\mathbf{H}\mathbf{s}\right)+\mathbf{s}^T\mathbf{V}_s^{-1}\mathbf{s} - \mathbf{x}^T\left(\mathbf{H}\mathbf{V}_s\mathbf{H}^T+\mathbf{V}_r\right)^{-1}\mathbf{x}$$

- *términos cuadráticos:*

$$\mathbf{s}^T \mathbf{V}_{s|x}^{-1} \mathbf{s} = (\mathbf{H}\mathbf{s})^T \mathbf{V}_r^{-1} \mathbf{H}\mathbf{s} + \mathbf{s}^T \mathbf{V}_s^{-1} \mathbf{s}$$

de donde:

$$\mathbf{V}_{s|x}^{-1} = \mathbf{H}^T \mathbf{V}_r^{-1} \mathbf{H} + \mathbf{V}_s^{-1}$$

$$\mathbf{V}_{s|x} = \left(\mathbf{H}^T \mathbf{V}_r^{-1} \mathbf{H} + \mathbf{V}_s^{-1}\right)^{-1}$$

- *términos lineales (sin -2):*

$$E^T\{\mathbf{s}|\mathbf{x}\} \mathbf{V}_{s|x}^{-1} = \mathbf{x}^T \mathbf{V}_r^{-1} \mathbf{H}$$

de donde:

$$E^T\{\mathbf{s} \mid \mathbf{x}\} = \mathbf{x}^T \mathbf{V}_r^{-1} \mathbf{H} \mathbf{V}_{s|x}$$

$$E\{\mathbf{s} \mid \mathbf{x}\} = \mathbf{V}_{s|x} \mathbf{H}^T \mathbf{V}_r^{-1} \mathbf{x}$$

y sustituyendo $\mathbf{V}_{s|x}$*:*

$$\hat{s}_{ms} = \left(\mathbf{H}^T \mathbf{V}_r^{-1} \mathbf{H} + \mathbf{V}_s^{-1}\right)^{-1} \mathbf{H}^T \mathbf{V}_r^{-1} \mathbf{x}$$

Puede notarse lo mismo que en el caso unidimensional: en particular, que el estimador es una arquitectura <u>lineal</u> (forma $\mathbf{w}^T\mathbf{x}$*).*

Es claro que, si **H=I** *(implica M=N), la solución tiene el mismo sentido de "reparto" que el en caso unidimensional.*

Ejercicio

E: *Analice y discuta el caso de (M) observaciones unidimensionales directas de variables independientes en ruido de valores independientes.*

Se trata de un caso con ***H=I***

con matrices de autocovarianza diagonales: $[\boldsymbol{D}_s]_{ii} = v_{si}$

$[\boldsymbol{D}_r]_{ii} = v_{ri}$

lo será $\boldsymbol{V}_{s|x}$: $\boldsymbol{D}_{s|x} = (\boldsymbol{D}_s^{-1} + \boldsymbol{D}_r^{-1})^{-1}$

$$\left[\boldsymbol{D}_{s|x}\right]_{ii} = \left(\frac{1}{v_{si}} + \frac{1}{v_{ri}}\right)^{-1} = \frac{v_{si}v_{ri}}{v_{si} + v_{ri}}$$

con lo que

$$\hat{\boldsymbol{s}}_{ms} = \boldsymbol{D}_{s|x}\boldsymbol{D}_r^{-1}\boldsymbol{x}$$

tiene como elementos

$$\hat{s}_{ms\,i} = \frac{v_{si}}{v_{si} + v_{ri}} x_i$$

como cabía esperar, extensión directa del caso unidimensional.

Ejercicio

E: *Analice y discuta el caso de (M) observaciones de la misma variable en ruido de valores independientes.*

En este caso, $\boldsymbol{H} = \boldsymbol{1}\ (M\times 1)$

$\boldsymbol{V}_s = v_s$

$[\boldsymbol{D}_r]_{ii} = v_{ri}$

$$V_{s|x} = \left(\boldsymbol{1}^T \boldsymbol{D}_r^{-1}\boldsymbol{1} + v_s^{-1}\right)^{-1} = \frac{1}{v_s^{-1} + \sum_i v_{ri}^{-1}}$$

$$\hat{s}_{ms} = \frac{1}{v_s^{-1} + \sum_i v_{ri}^{-1}}\ \boldsymbol{1}^T \boldsymbol{D}_r^{-1}\boldsymbol{x} = \frac{1}{v_s^{-1} + \sum_i v_{ri}^{-1}} \sum_i v_{ri}^{-1} x_i$$

que es una media muestral ponderada; si los ruidos son i. d., $v_{ri} = v_r$

$$\hat{s}_{ms} = \frac{1}{v_s^{-1} + M v_r^{-1}} v_r^{-1} \sum_i x_i = \frac{1}{M + v_r v_s^{-1}} \sum_i x_i$$

que es una media muestral modificada (según potencias: teniendo en cuenta la coherencia de la señal y la incoherencia del ruido).

Ejercicio de Ampliación

A: *La forma general del problema de decidir entre dos señales deterministas (conocidas) en presencia de ruido gaussiano (conocido) es*

$H_0 : \mathbf{x} = \mathbf{m}_0 + \mathbf{n}_0$

$H_1 : \mathbf{x} = \mathbf{m}_1 + \mathbf{n}_1$

con los ruidos $\mathbf{n}_i : G(\mathbf{0}, \mathbf{V}_i)$

Formule el correspondiente LRT.

Recordando que $$p(\mathbf{x} \mid H_i) = \frac{1}{(2\pi)^{N/2} |\mathbf{V}_i|^{1/2}} exp\left[-\frac{1}{2}(\mathbf{x}-\mathbf{m}_i)^T \mathbf{V}_i^{-1}(\mathbf{x}-\mathbf{m}_i)\right]$$

tomando neperianos y eliminando términos comunes

$$ln \frac{p(\mathbf{x} \mid H_1)}{p(\mathbf{x} \mid H_0)} = ln\left(\frac{|\mathbf{V}_0|}{|\mathbf{V}_1|}\right)^{1/2} - \frac{1}{2}\left[(\mathbf{x}-\mathbf{m}_1)^T \mathbf{V}_1^{-1}(\mathbf{x}-\mathbf{m}_1) - (\mathbf{x}-\mathbf{m}_0)^T \mathbf{V}_0^{-1}(\mathbf{x}-\mathbf{m}_0)\right] \underset{D_0}{\overset{D_1}{\gtrless}}$$

$$\underset{D_0}{\overset{D_1}{\gtrless}} ln \frac{(C_{10}-C_{00})Pr(H_0)}{(C_{01}-C_{11})Pr(H_1)} = ln\,\eta$$

que se puede reescribir

$$(\mathbf{x}-\mathbf{m}_0)^T \mathbf{V}_0^{-1}(\mathbf{x}-\mathbf{m}_0)-(\mathbf{x}-\mathbf{m}_1)^T \mathbf{V}_1^{-1}(\mathbf{x}-\mathbf{m}_1) \underset{D_0}{\overset{D_1}{\gtrless}} 2\ln\eta - \ln\frac{|\mathbf{V}_0|}{|\mathbf{V}_1|} = \eta_1$$

que supone una frontera (general) ***cuadrática*** *(e igual tipo de* ***discriminantes****).*

Las formas ponderadas $(\mathbf{x}-\mathbf{m}_i)^T \mathbf{V}_i^{-1}(\mathbf{x}-\mathbf{m}_i)$*, obviamente generalizaciones de los (cuadrados de) las distancias euclídeas entre* $\mathbf{x}$ *y* $\mathbf{m}_i$*, se denominan* ***distancias de Mahalanobis****: la decisión se realiza en función de ellas.*

Ejercicio de Ampliación

A: Analice y discuta el problema anterior si el ruido no depende de la señal presente.

Será $\mathbf{V_1} = \mathbf{V_0} = \mathbf{V}$; operando, desaparecen los términos cuadráticos y queda

$$-2\mathbf{m}_0^T\mathbf{V}^{-1}\mathbf{x} + \mathbf{m}_0^T\mathbf{V}^{-1}\mathbf{m}_0 + 2\mathbf{m}_1^T\mathbf{V}^{-1}\mathbf{x} - \mathbf{m}_1^T\mathbf{V}^{-1}\mathbf{m}_1 \underset{D_0}{\overset{D_1}{\gtrless}} \eta_1$$

$$(\mathbf{m}_1 - \mathbf{m}_0)^T\mathbf{V}^{-1}\mathbf{x} \underset{D_0}{\overset{D_1}{\gtrless}} \frac{1}{2}\left(\eta_1 + \mathbf{m}_1^T\mathbf{V}^{-1}\mathbf{m}_1 - \mathbf{m}_0^T\mathbf{V}^{-1}\mathbf{m}_0\right) = \eta_2$$

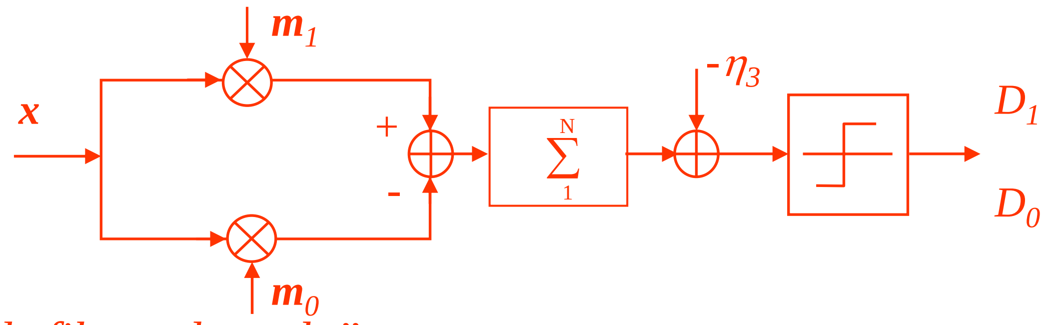
que es un decisor lineal $\quad \mathbf{w}^T\mathbf{x} \underset{D_0}{\overset{D_1}{\gtrless}} \eta_2$

Si además las componentes del ruido son iid: $\quad \mathbf{V} = v\mathbf{I}$

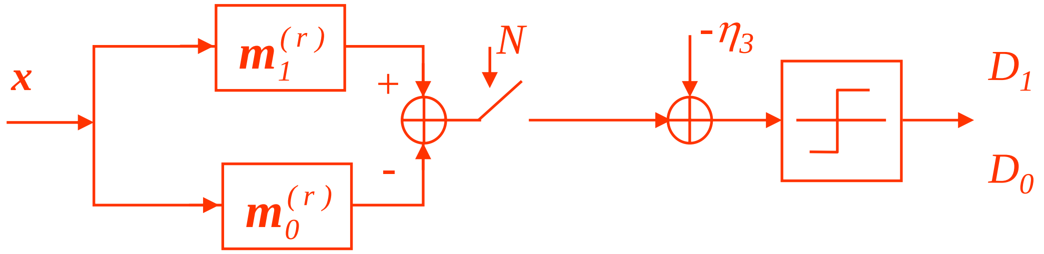
$$(\mathbf{m}_1 - \mathbf{m}_0)^T v^{-1}\mathbf{I}\mathbf{x} \underset{D_0}{\overset{D_1}{\gtrless}} \eta_2$$

$$(\mathbf{m}_1 - \mathbf{m}_0)^T\mathbf{x} \underset{D_0}{\overset{D_1}{\gtrless}} v\eta_2 = \eta_3 \qquad \left(= \frac{1}{2}\left[2v\ln\eta + \|m_1\|_2^2 - \|m_0\|_2^2\right]\right)$$

que, como se sabe, es realizable como "decisor de correlación"



o "de filtro adaptado"



Si, además $\mathbf{m}_0 = \mathbf{0}$ *(<u>detección</u> de* $\mathbf{m}_1$*): desaparecen las ramas correspondientes; el umbral pasa a* $\eta_4 = \frac{1}{2}\left(2\nu \ln \eta + \|\mathbf{m}_1\|_2^2\right)$

(en caso ML, queda $\frac{1}{2}\|\mathbf{m}_1\|_2^2$ *: se pide que* $\mathbf{m}_1^T \mathbf{x}$ *llegue a la mitad de energía de* $\mathbf{m}_1$*).*

Ejercicio de Ampliación

A: *Analice y discuta la decisión (binaria) gaussiana cuando se trata de la misma señal en distintos ruidos.*

$\mathbf{m}_1 = \mathbf{m}_0 = \mathbf{m}$: *siguen empleándose las distancias de Mahalanobis*

$$(\mathbf{x}-\mathbf{m})^T\left(\mathbf{V}_0^{-1}-\mathbf{V}_1^{-1}\right)(\mathbf{x}-\mathbf{m}) \underset{D_0}{\overset{D_1}{\gtrless}} \eta_1$$

- *Si además, los ruidos son iid:* $\mathbf{V}_i = v_i\mathbf{I}$

$$(\mathbf{x}-\mathbf{m})^T\left(\frac{1}{v_0}-\frac{1}{v_1}\right)\mathbf{I}\,(\mathbf{x}-\mathbf{m})$$

admitiendo $v_1>v_0$ *(si no, voltean las desigualdades)*

$$\|\mathbf{x}-\mathbf{m}\|_2^2 \underset{D_0}{\overset{D_1}{\gtrless}} \frac{v_0 v_1}{v_1-v_0}\eta_1$$

decisor por distancia euclídea

- *Si además* $\mathbf{m}=\mathbf{0}$:

$$\|\mathbf{x}\|_2^2 \underset{D_0}{\overset{D_1}{\gtrless}} \frac{v_0 v_1}{v_1-v_0}\eta_1$$

que decide por energía entre las dos opciones de varianza del ruido.

Estadísticos suficientes

Se han visto casos en los que no es preciso el conocimiento del conjunto de las observaciones (o de las variables) para realizar la estimación o la decisión:

- estimar m y v de una gaussiana requiere sólo $\sum_k x^{(k)}, \sum_k x^{(k)^2}$
- estimar una gaussiana en otra gaussiana requiere formas del tipo $\mathbf{w}^T\mathbf{x}$: lineales también
- decidir qué señal determinista es la que se encuentra en ruido gaussiano requiere formas cuadráticas (o lineales)

Estas formas se llaman **estadísticos suficientes**: conocerlos permite simplificar procesos de estimación o decisión.

Teorema de Factorización

El resultado de una transformación de las observaciones $\mathbf{t} = \mathbf{t}(\mathbf{x})$ es un estadístico suficiente para estimar (ML) el parámetro determinista **s** si y sólo si

$$p(\mathbf{x} \mid \mathbf{s}) = g(\mathbf{t}, \mathbf{s})\, h(\mathbf{x}, \mathbf{t})$$

(La descomposición no es única: $g'(\mathbf{t}, \mathbf{s}) = g(\mathbf{t}, \mathbf{s})\, f(\mathbf{t})$
$h'(\mathbf{x}, \mathbf{t}) = h(\mathbf{x}, \mathbf{t}) / f(\mathbf{t})$)

ya que

$$\arg[\max_{\mathbf{s}} p(\mathbf{x} \mid \mathbf{s})] = \arg[\max_{\mathbf{s}} g(\mathbf{t}, \mathbf{s})]$$

(y se comprende fácilmente la ventaja si g es sencilla); y no ocurre en otro caso.

Una formulación equivalente es:

$$p(\mathbf{x} \mid \mathbf{s}, \mathbf{t}) = p(\mathbf{x} \mid \mathbf{t})$$

Suficiente: si se da la equivalente,

$$p(\mathbf{x} \mid \mathbf{s}) = p(\mathbf{x} \mid \mathbf{s}, \mathbf{t})\, p(\mathbf{t} \mid \mathbf{s}) = p(\mathbf{x} \mid \mathbf{t})\, p(\mathbf{t} \mid \mathbf{s})$$

y se pueden identificar los factores con h y g, respectivamente.

Necesaria: suponemos que se cumple el Teorema de Factorización; como

$$p(\mathbf{x} \mid \mathbf{s}, \mathbf{t}) = \frac{p(\mathbf{x}, \mathbf{t} \mid \mathbf{s})}{p(\mathbf{t} \mid \mathbf{s})} = \frac{p(\mathbf{x}, \mathbf{t} \mid \mathbf{s})}{\int p(\mathbf{x}, \mathbf{t} \mid \mathbf{s})\, d\mathbf{x}}$$

y como $p(\mathbf{x}, \mathbf{t} \mid \mathbf{s})$ es $p(\mathbf{x} \mid \mathbf{s})$, al ser $\mathbf{t}$ una función de $\mathbf{x}$, se tiene

$$p(\mathbf{x} \mid \mathbf{s}, \mathbf{t}) = \frac{p(\mathbf{x} \mid \mathbf{s})}{\int p(\mathbf{x} \mid \mathbf{s})\, d\mathbf{x}}$$

aplicando el Teorema de Factorización

$$p(\mathbf{x} \mid \mathbf{s}, \mathbf{t}) = \frac{g(\mathbf{t}, \mathbf{s})\, h(\mathbf{x}, \mathbf{t})}{\int g(\mathbf{t}, \mathbf{s}) h(\mathbf{x}, \mathbf{t})\, d\mathbf{x}} = \frac{h(\mathbf{x}, \mathbf{t})}{\int h(\mathbf{x}, \mathbf{t})\, d\mathbf{x}}$$

que no depende de **s**, y que, por identificación, resulta ser p(**x** | **t**).

En ocasiones **t** se obtiene de una transformación uno a uno de los datos: $\mathbf{x} \to \mathbf{t}, \mathbf{i}$;   el conjunto **i** es un **estadístico irrelevante**.

Caso de estimación de parámetro aleatorio

Sigue siendo válido el Teorema de Factorización, ya que de p(**x** | **s**, **t**) = p(**x** | **t**) se sigue

$$p(\mathbf{s}\,|\,\mathbf{x}) = p(\mathbf{s}\,|\,\mathbf{x},\mathbf{t}) = \frac{p(\mathbf{s},\mathbf{x}\,|\,\mathbf{t})}{p(\mathbf{x}\,|\,\mathbf{t})} = \frac{p(\mathbf{x}\,|\,\mathbf{s},\mathbf{t})p(\mathbf{s}\,|\,\mathbf{t})}{p(\mathbf{x}\,|\,\mathbf{t})} = p(\mathbf{s}\,|\,\mathbf{t})$$

lo que indica claramente que, dado **t**, los datos originales no se necesitan para realizar cualquier estimación Bayesiana

$$\min_{\hat{s}} \int C\left(\mathbf{s},\hat{\mathbf{s}}\right) p(\mathbf{s}\,|\,\mathbf{x})\, d\mathbf{s}$$

Caso de decisión

Es obvio que la condición equivalente es

$$p(\mathbf{x} \mid H_i) = g(\mathbf{t}, H_i)\, h(\mathbf{x}, \mathbf{t})$$

o también

$$\Pr(H_i \mid \mathbf{x}) \left(= \Pr(H_i \mid \mathbf{x}, \mathbf{t})\right) = \Pr(H_i \mid \mathbf{t})$$

ya que basta $g(\mathbf{t}, H_i)$ o $\Pr(H_i \mid \mathbf{t})$ para la decisión Bayesiana.

<u>Ejercicio</u>

E: *Determinar un estadístico suficiente para la estimación del parámetro de una distribución exponencial unilateral (Maxwell), p(x) = a exp (-ax) u(x) (a > 0), mediante* K *observaciones independientes.*

$$p(\mathbf{x} \mid a) = \prod_{k=1}^{K} a\, exp\left(-ax^{(k)}\right), \qquad \{\mathbf{x}\} \geq 0$$

$$ln\, p(\mathbf{x} \mid a) = \sum_{k=1}^{K} \left[ln\, a - ax^{(k)} \right] = K\, ln\, a - a \sum_{k=1}^{K} x^{(k)}$$

y ya se observa que $\sum_{k=1}^{K} x^{(k)}$ *es un estadístico suficiente...*

$$\frac{\partial\, ln}{\partial a} = \frac{K}{a} - \sum_{k=1}^{K} x^{(k)}; \quad \left. \frac{\partial\, ln}{\partial a} \right|_{a=\hat{a}_{ml}} = 0 = \frac{K}{\hat{a}_{ml}} - \sum_{k=1}^{K} x^{(k)} \quad \left(\frac{\partial^2\, ln}{\partial a^2} = -\frac{K}{a^2} < 0 \right)$$

$$\hat{a}_{ml} = \frac{K}{\sum_{k=1}^{K} x^{(k)}}$$

Obsérvese que en este caso la propia $p(x \mid a)$ tiene la forma de $g(t, a)$, con

$$t = \sum_{k=1}^{K} x^{(k)}$$

lo que se ve claramente en las expresiones que resultan de tomar logaritmos (en que la factorización se convierte en separación de sumandos):

$$\ln p(\mathbf{x} \mid a) = K \ln a - a \sum_{k=1}^{K} x^{(k)} = K \ln a - at \qquad (= g(t, a))$$