

Estimación y Predicción en Series Temporales

Modelos de estado no-lineales: EKF y UKF

Departamento de Procesamiento de Señales

Instituto de Ingeniería Eléctrica
Facultad de Ingeniería

2022

Biblio:

- Anderson and Moore, Optimal Filtering, 1979 (re-edición Dover 2005).
- Julier, S.J. and Uhlmann, J.K., *Unscented filtering and nonlinear estimation*, Proceedings of the IEEE , vol.92, no.3, pp.401,422, Mar 2004.
- Wan, E. A., and R. van der Merwe. *The unscented Kalman filter*, in S. Haykin, ed., Kalman Filtering and Neural Networks, Wiley, New York, Chapter 7, 2001.

Introducción - Motivación

Consideremos los modelos de estado en sus versiones general y lineal:

$$\begin{cases} \mathbf{x}_{k+1} = \mathbf{f}_{k+1}(\mathbf{x}_k, \mathbf{v}_{k+1}) \\ \mathbf{y}_k = \mathbf{h}_k(\mathbf{x}_k, \mathbf{n}_k) \end{cases} \quad \begin{cases} \mathbf{x}_{k+1} = \mathbf{F}_{k+1}\mathbf{x}_k + \mathbf{G}_{k+1}\mathbf{v}_{k+1} \\ \mathbf{y}_k = \mathbf{H}_k\mathbf{x}_k + \mathbf{U}_k\mathbf{n}_k \end{cases}$$

Hasta ahora hemos visto sobre el filtro de Kalman, que:

- Si los ruidos no son necesariamente Gaussianos, el filtro de Kalman es óptimo entre todos los filtros lineales, en el sentido del MSE.
- Si los ruidos son Gaussianos, el filtro de Kalman es universalmente óptimo (i.e. aún dentro del conjunto de los filtros no-lineales) en el sentido del MSE.
- Todo proceso ARMA se puede representar como un filtro de Kalman, pasando a una representación en variables de estado.

Hoy nos concentraremos en dos técnicas para modelos no-lineales:

- 1 *Extended Kalman Filter (EKF)*: básicamente una

El filtro de Kalman Extendido (EKF)

La linealización

Consideramos ahora un modelo en el espacio de estados no-lineal, de la forma:

$$\begin{cases} \mathbf{x}_{k+1} &= \mathbf{f}_{k+1}(\mathbf{x}_k, \mathbf{v}_{k+1}) \\ \mathbf{y}_k &= \mathbf{h}_k(\mathbf{x}_k, \mathbf{n}_k). \end{cases}$$

El filtro de Kalman extendido se basa en una aproximación de primer orden del modelo, seguido del loop de Kalman con ecuaciones lineales.

- La ecuación de estado se linealiza entorno a $(\mathbf{x}, \mathbf{v}) = (\mathbf{x}_{k|k}, 0)$.
- La ecuación de medida se linealiza entorno a $(\mathbf{x}, \mathbf{n}) = (\mathbf{x}_{k|k-1}, 0)$.

$$\begin{cases} \mathbf{x}_{k+1} &= \mathbf{f}_{k+1}(\mathbf{x}_{k|k}, 0) + \mathbf{F}_{k+1}(\mathbf{x}_k - \mathbf{x}_{k|k}) + \mathbf{G}_{k+1}\mathbf{v}_{k+1} \\ \mathbf{y}_k &= \mathbf{h}_k(\mathbf{x}_{k|k-1}, 0) + \mathbf{H}_k(\mathbf{x}_k - \mathbf{x}_{k|k-1}) + \mathbf{U}_k\mathbf{n}_k, \end{cases}$$

dónde

$$\frac{\partial \mathbf{f}_{k+1}}{\partial \mathbf{x}_k}$$

$$\frac{\partial \mathbf{f}_{k+1}}{\partial \mathbf{v}_{k+1}}$$

El filtro de Kalman Extendido (EKF)

Las iteraciones de Kalman

Las ecuaciones para el EKF quedan entonces:

- Inicialización: $\mathbf{x}_0, \mathbf{P}_0$.
- Para $k = 0, 1, 2, \dots$, calcular:

$$\hat{\mathbf{x}}_{k|k-1} = \mathbf{f}_k(\hat{\mathbf{x}}_{k-1|k-1}, 0)$$

$$\mathbf{P}_{k|k-1} = \mathbf{F}_k \mathbf{P}_{k-1|k-1} \mathbf{F}_k^T + \mathbf{G}_k \mathbf{Q}_k \mathbf{G}_k^T$$

$$\mathbf{K}_k = \mathbf{P}_{k|k-1} \mathbf{H}_k^T (\mathbf{H}_k \mathbf{P}_{k|k-1} \mathbf{H}_k^T + \mathbf{U}_k \mathbf{R}_k \mathbf{U}_k^T)^{-1}$$

$$\hat{\mathbf{x}}_{k|k} = \hat{\mathbf{x}}_{k|k-1} + \mathbf{K}_k (\mathbf{y}_k - \mathbf{h}_k(\hat{\mathbf{x}}_{k|k-1}, 0))$$

$$\mathbf{P}_{k|k} = (\mathbf{I} - \mathbf{K}_k \mathbf{H}_k) \mathbf{P}_{k|k-1}$$

Observaciones y comentarios:

- El problema del filtro EKF es que la linealización del modelo introduce un sesgo en $\hat{\mathbf{x}}_{k|k}$ y una covarianza errónea.
- La calidad de la aproximación debida a la linealización será mejor cuanto menor sean $\|\mathbf{x} - \hat{\mathbf{x}}_{k|k}\|^2$ y $\|\mathbf{x} - \hat{\mathbf{x}}_{k|k-1}\|^2$.
- Dependiendo de como se comporte el EKF se comportará mejor en

El *Unscented Kalman Filter* (UKF)

Julier & Uhlmann 1996

Este método busca dejar de lado las aproximaciones y trabajar directamente sobre las transformaciones no lineales, mapeando un conjunto de puntos bien elegidos.

La transformación *unscented*

- Sea $\mathbf{x} \sim \mathcal{N}(\bar{\mathbf{x}}, \mathbf{P}_{\mathbf{x}})$ un vector aleatorio de \mathbb{R}^n , $\mathbf{g} : \mathbb{R}^n \rightarrow \mathbb{R}^m$ una función, e $\mathbf{y} = \mathbf{g}(\mathbf{x}) \in \mathbb{R}^m$.
- La *unscented transform* busca calcular estimadores insesgados de $\bar{\mathbf{y}}$ y de $\mathbf{P}_{\mathbf{y}}$.
- En lugar de linealizar \mathbf{g} como en el EKF, la propuesta es evaluar $\bar{\mathbf{y}}$ y $\mathbf{P}_{\mathbf{y}}$ (1er y 2do momento de \mathbf{y}) mediante estimaciones empíricas ponderadas, calculadas a partir de la transformación exacta \mathbf{g} de un conjunto de puntos bien elegidos alrededor de $\bar{\mathbf{x}}$, denominados *σ -puntos*.

UKF: Los σ -puntos

Tenemos $\mathbf{x} \in \mathbb{R}^n$, $\mathbf{x} \sim \mathcal{N}(\bar{\mathbf{x}}, \mathbf{P}_{\mathbf{x}})$.

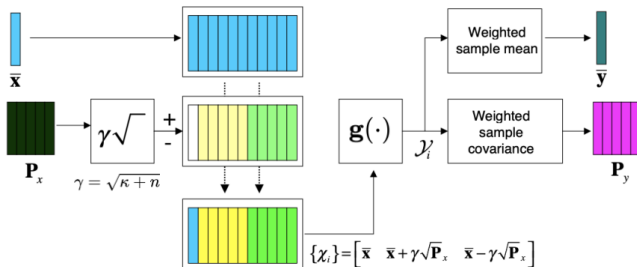
Los $2n + 1$ σ -puntos y sus pesos correspondientes se definen como:

$$\begin{aligned}\mathcal{X}_0 &= \bar{\mathbf{x}} & W_0 &= \frac{\kappa}{\kappa + n} \\ \mathcal{X}_i &= \bar{\mathbf{x}} + (\sqrt{(\kappa + n)\mathbf{P}_{\mathbf{x}}})_i & W_i &= \frac{1}{2(\kappa + n)}, \quad i = 1, \dots, n \\ \mathcal{X}_i &= \bar{\mathbf{x}} - (\sqrt{(\kappa + n)\mathbf{P}_{\mathbf{x}}})_{i-n} & W_i &= \frac{1}{2(\kappa + n)}, \quad i = n + 1, \dots, 2n.\end{aligned}$$

donde κ es un escalar y $(\sqrt{\mathbf{P}_{\mathbf{x}}})_i$ representa a la i -ésima columna de alguna factorización de $\mathbf{P}_{\mathbf{x}}$, e.g. Cholesky, i.e.

$$\mathbf{P}_{\mathbf{x}} = \sqrt{\mathbf{P}_{\mathbf{x}}} \times \sqrt{\mathbf{P}_{\mathbf{x}}}^T = \sum_{i=1}^n (\sqrt{\mathbf{P}_{\mathbf{x}}})_i (\sqrt{\mathbf{P}_{\mathbf{x}}})_i^T.$$

UKF: Los σ -puntos



Transformada *Unscented*. Adaptada de (Wan and Van der Merwe, 2001)

Justificación de la elección de los σ -puntos

Veamos porqué Julier & Uhlmann (1996) proponen esta elección.

A partir de los σ -puntos y sus transformados, se pueden calcular estadísticos empíricos de \mathbf{x} y de \mathbf{y} . Sea $\mathbf{y}_i = \mathbf{g}(\mathbf{x}_i)$, definimos

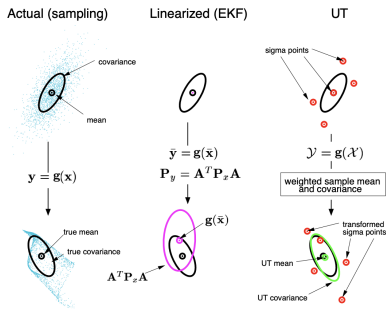
$$\bar{\mathbf{x}} = \sum_{i=0}^{2n} W_i \mathbf{x}_i$$

$$\mathbf{P}_{\mathbf{x}} = \sum_{i=0}^{2n} W_i (\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})^T$$

$$\bar{\mathbf{y}} = \sum_{i=0}^{2n} W_i \mathbf{y}_i$$

$$\mathbf{P}_{\mathbf{y}} = \sum_{i=0}^{2n} W_i (\mathbf{y}_i - \bar{\mathbf{y}})(\mathbf{y}_i - \bar{\mathbf{y}})^T$$

Ejercicio: Demostrar que media y covarianza empíricas de $\mathcal{X} = \{\mathbf{x}_i\}_{i=0}^{2n}$ coinciden con las de \mathbf{x} .



Propagación de media y matriz de covarianza: (a) Monte Carlo en la imagen de \mathbf{g} , (b) Linealización (UKF), (c) Transformada *Unscented*. Figura adaptada de (Wan and Van der Merwe, 2001)

Justificación de la elección de los σ -puntos (cont.)

Resta ver qué pasa con la media y covarianza empíricas de los $\{\mathcal{Y}_i\}_{i=1}^{2n+1}$ en relación a y .

Suponemos que $\mathbf{g} = [g_1, g_2, \dots, g_m]^T$, $g_i : \mathbb{R}^n \rightarrow \mathbb{R}$, admite un desarrollo de Taylor entorno a $\bar{\mathbf{x}}$. Entonces tenemos

$$\mathbf{g}(\bar{\mathbf{x}} + \boldsymbol{\delta}) = \mathbf{g}(\bar{\mathbf{x}}) + D_{\boldsymbol{\delta}}\mathbf{g}(\bar{\mathbf{x}}) + \frac{1}{2}D_{\boldsymbol{\delta}}^2\mathbf{g}(\bar{\mathbf{x}}) + \dots,$$

dónde

$$D_{\boldsymbol{\delta}}\mathbf{g}(\bar{\mathbf{x}}) = \nabla^T \mathbf{g}(\bar{\mathbf{x}})\boldsymbol{\delta}$$

$$\begin{aligned} D_{\boldsymbol{\delta}}^2\mathbf{g}(\bar{\mathbf{x}}) &= D_{\boldsymbol{\delta}}[\nabla^T \mathbf{g}(\bar{\mathbf{x}})\boldsymbol{\delta}] = \nabla^T [\nabla^T \mathbf{g}(\bar{\mathbf{x}})\boldsymbol{\delta}]\boldsymbol{\delta} \\ &= [\text{Tr}(\nabla^2 g_1(\bar{\mathbf{x}})\boldsymbol{\delta}\boldsymbol{\delta}^T), \dots, \text{Tr}(\nabla^2 g_m(\bar{\mathbf{x}})\boldsymbol{\delta}\boldsymbol{\delta}^T)]^T, \end{aligned}$$

con $\nabla \mathbf{g}$ la matriz Jacobiana de \mathbf{g} , y $\nabla^2 g_i$ la matriz Hessiana de g_i (Apéndice)

Justificación de la elección de los σ -puntos (cont.)

Luego,

$$\begin{aligned}\bar{\mathbf{y}} &= \mathbb{E}[\mathbf{g}(\mathbf{x})] = \mathbb{E}[\mathbf{g}(\bar{\mathbf{x}} + \boldsymbol{\delta})] \\ &= \mathbf{g}(\bar{\mathbf{x}}) + \mathbb{E}[\nabla^T \mathbf{g}(\bar{\mathbf{x}}) \boldsymbol{\delta}] + \frac{1}{2} \mathbb{E}[[\text{Tr}(\nabla^2 g_1(\bar{\mathbf{x}}) \boldsymbol{\delta} \boldsymbol{\delta}^T), \dots, \text{Tr}(\nabla^2 g_m(\bar{\mathbf{x}}) \boldsymbol{\delta} \boldsymbol{\delta}^T)]^T] + \dots\end{aligned}$$

Como $\mathbf{x} = \bar{\mathbf{x}} + \boldsymbol{\delta}$, tenemos $\boldsymbol{\delta} \sim \mathcal{N}(\mathbf{0}, \mathbf{P}_{\mathbf{x}})$, por lo que $\mathbb{E}[\boldsymbol{\delta}] = \mathbf{0}$ y $\mathbb{E}[\boldsymbol{\delta} \boldsymbol{\delta}^T] = \mathbf{P}_{\mathbf{x}}$:

$$\bar{\mathbf{y}} = \mathbf{g}(\bar{\mathbf{x}}) + \frac{1}{2} [\text{Tr}(\nabla^2 g_1(\bar{\mathbf{x}}) \mathbf{P}_{\mathbf{x}}), \dots, \text{Tr}(\nabla^2 g_m(\bar{\mathbf{x}}) \mathbf{P}_{\mathbf{x}})]^T + \dots$$

Por otra parte, considerando $\boldsymbol{\sigma}_i = (\sqrt{(\kappa + n)} \mathbf{P}_{\mathbf{x}})_i$,

$$\begin{aligned}\bar{\mathbf{y}} &= W_0 \mathbf{g}(\bar{\mathbf{x}}) + \sum_{i=1}^n W_i (\mathbf{g}(\bar{\mathbf{x}} + \boldsymbol{\sigma}_i) + \mathbf{g}(\bar{\mathbf{x}} - \boldsymbol{\sigma}_i)) \\ &= \mathbf{g}(\bar{\mathbf{x}}) + \mathbf{0} + \frac{1}{2} \mathbb{E}[[\text{Tr}(\nabla^2 g_1(\bar{\mathbf{x}}) \sum_{i=1}^{2n} \frac{\boldsymbol{\sigma}_i \boldsymbol{\sigma}_i^T}{2(\kappa + n)}, \dots, \text{Tr}(\nabla^2 g_m(\bar{\mathbf{x}}) \sum_{i=1}^{2n} \frac{\boldsymbol{\sigma}_i \boldsymbol{\sigma}_i^T}{2(\kappa + n)})]^T] + \\ &= \mathbf{g}(\bar{\mathbf{x}}) + \frac{1}{2} [\text{Tr}(\nabla^2 g_1(\bar{\mathbf{x}}) \mathbf{P}_{\mathbf{x}}), \dots, \text{Tr}(\nabla^2 g_m(\bar{\mathbf{x}}) \mathbf{P}_{\mathbf{x}})]^T + \dots\end{aligned}$$

En ambos casos, debido a la simetría de los σ -puntos entorno a $\bar{\mathbf{x}}$, los términos impares se anulan, por los que la media empírica de los $\{\mathcal{Y}_i\}_{i=0}^{2n}$ coincide con $\bar{\mathbf{y}}$ hasta el tercer orden.

Justificación de la elección de los σ -puntos (cont.)

Ahora vemos qué pasa con \mathbf{P}_y y \mathbf{P}_y .

$$\begin{aligned}\mathbf{P}_y &= \mathbb{E}[(\mathbf{g}(\bar{\mathbf{x}} + \boldsymbol{\delta}) - \mathbf{g}(\bar{\mathbf{x}}))(\mathbf{g}(\bar{\mathbf{x}} + \boldsymbol{\delta}) - \mathbf{g}(\bar{\mathbf{x}}))^T] \\ &= \mathbb{E}[(\nabla^T \mathbf{g}(\bar{\mathbf{x}})\boldsymbol{\delta})(\nabla^T \mathbf{g}(\bar{\mathbf{x}})\boldsymbol{\delta})^T] + \dots = \nabla^T \mathbf{g}(\bar{\mathbf{x}})\mathbb{E}[\boldsymbol{\delta}\boldsymbol{\delta}^T]\nabla \mathbf{g}(\bar{\mathbf{x}}) + \dots \\ &= \nabla^T \mathbf{g}(\bar{\mathbf{x}})\mathbf{P}_x \nabla \mathbf{g}(\bar{\mathbf{x}}) + \dots\end{aligned}$$

Por otro lado,

$$\begin{aligned}\mathbf{P}_y &= \sum_{i=0}^{2n} W_i (\mathcal{Y}_i - \bar{\mathcal{Y}})(\mathcal{Y}_i - \bar{\mathcal{Y}})^T \\ &= \frac{1}{2(n+\kappa)} \times 2 \sum_{i=1}^n (\nabla^T \mathbf{g}(\bar{\mathbf{x}})\boldsymbol{\sigma}_i + \dots)(\nabla^T \mathbf{g}(\bar{\mathbf{x}})\boldsymbol{\sigma}_i + \dots)^T \\ &= \nabla^T \mathbf{g}(\bar{\mathbf{x}})\mathbf{P}_x \nabla \mathbf{g}(\bar{\mathbf{x}}) + \dots\end{aligned}$$

Por los que ambas coinciden en el primer orden. Ajustando κ para que $n + \kappa = 3$, se puede mostrar que las medias pueden coincidir hasta orden 4. Sin embargo, la covarianza empírica ponderada puede volverse semidefinida negativa. Este problema fue resuelto por Julier en 2002, introduciendo la *scaled unscented transformation*.

Scaled Unscented Transform

- La idea es escalar la distancia entre los σ -puntos reemplazando \mathcal{X}_i por

$$\mathcal{X}'_i = \mathcal{X}_0 + \alpha(\mathcal{X}_i - \mathcal{X}_0), \quad \text{con } \alpha \text{ el factor de escala.}$$

- En este caso W_0 debe ser diferente. Llamando $\lambda = \alpha^2(\kappa + n) + n$, la *scaled unscented transform* se resume así:

$$\begin{aligned}\mathcal{X}_0 &= \bar{\mathbf{x}} & W_0^{(m)} &= \frac{\lambda}{\lambda + n}, & W_0^{(c)} &= \frac{\lambda}{\lambda + n} + (1 - \alpha^2) \\ \mathcal{X}_i &= \bar{\mathbf{x}} + (\sqrt{(\lambda + n)\mathbf{P}_{\mathbf{x}}})_i & W_i^{(m)} &= W_i^{(c)} = \frac{1}{2(\kappa + n)}, & i &= 1, \dots, n \\ \mathcal{X}_i &= \bar{\mathbf{x}} - (\sqrt{(\lambda + n)\mathbf{P}_{\mathbf{x}}})_{i-n} & W_i^{(m)} &= W_i^{(c)} = \frac{1}{2(\kappa + n)}, & i &= n + 1, \dots, 2n\end{aligned}$$

donde $0 \leq \alpha \leq 1$, κ puede ser seteado a 0, y $\beta = 2$ es el valor óptimo en el caso Gaussiano.

De esta forma, las medias y covarianzas empíricas de \mathbf{x} e \mathbf{y} quedan:

$$\bar{\mathcal{X}} = \sum_{i=0}^{2n} W_i^{(m)} \mathcal{X}_i$$

Unscented Kalman Filter (UKF)

El UKF se obtiene juntando las formulas del filtro de Kalman con las de la Unscented Transform. Considerando nuevamente el modelo no-lineal

$$\begin{cases} \mathbf{x}_{k+1} &= \mathbf{f}_{k+1}(\mathbf{x}_k, \mathbf{v}_{k+1}) \\ \mathbf{y}_k &= \mathbf{h}_k(\mathbf{x}_k, \mathbf{n}_k), \end{cases}$$

y usando una representación de estado aumentada $\mathbf{x}_k^a = [\mathbf{x}_k^T \mathbf{v}_k^T \mathbf{n}_k^T]^T$ con su correspondiente matriz de covarianza, el UKF se resume como:

1 Inicialización: $\mathbf{x}_{0|0}$, $\mathbf{P}_{0|0}$

2 Iteraciones: para $k = 1, 2, \dots$,

$$\mathbf{P}^a = \begin{bmatrix} \mathbf{P}_{k-1|k-1} & 0 & 0 \\ 0 & \mathbf{Q}_k & 0 \\ 0 & 0 & \mathbf{R}_k \end{bmatrix}$$

$$\mathcal{X}^a = \left[\mathbf{x}_{k-1|k-1}^a \mid \mathbf{x}_{k-1|k-1}^a \pm \sqrt{(n + \lambda)\mathbf{P}^a} \right] = \begin{bmatrix} \mathcal{X}_{k-1|k-1}^x \\ \mathcal{X}_k^v \\ \mathcal{X}_k^n \end{bmatrix}$$

$$\mathcal{X}_{k|k-1}^x = \mathbf{f}_k(\mathcal{X}_{k-1|k-1}^x, \mathcal{X}_k^v)$$

$$\mathbf{x}_{k|k-1} = \sum_{i=0}^{2n} W_i^{(m)} \mathcal{X}_{k|k-1,i}^x$$

$$\mathbf{P}_{k|k-1}^x = \sum_{i=0}^{2n} W_i^{(c)} (\mathcal{X}_{k|k-1,i}^x - \mathbf{x}_{k|k-1})(\mathcal{X}_{k|k-1,i}^x - \mathbf{x}_{k|k-1})^T$$

$$\mathcal{Y}_{k|k-1} = \mathbf{h}(\mathcal{X}_{k|k-1}^x, \mathcal{X}_k^n)$$

$$\mathbf{y}_{k|k-1} = \sum_{i=0}^{2n} W_i^{(m)} \mathcal{Y}_{k|k-1,i}$$

$$\mathbf{P}_{k|k-1}^y = \sum_{i=0}^{2n} W_i^{(c)} (\mathcal{Y}_{k|k-1,i} - \mathbf{y}_{k|k-1})(\mathcal{Y}_{k|k-1,i} - \mathbf{y}_{k|k-1})^T$$

$$\mathbf{P}_{k|k-1}^{xy} = \sum_{i=0}^{2n} W_i^{(c)} (\mathcal{X}_{k|k-1,i}^x - \mathbf{x}_{k|k-1})(\mathcal{Y}_{k|k-1,i} - \mathbf{y}_{k|k-1})^T$$

$$\mathbf{K}_k = \mathbf{P}_{k|k-1}^{xy} (\mathbf{P}_{k|k-1}^y)^{-1}$$

$$\mathbf{x}_{k|k} = \mathbf{x}_{k|k-1} + \mathbf{K}_k (\mathbf{y}_k - \mathbf{y}_{k|k-1})$$

$$\mathbf{P}_{k|k} = \mathbf{P}_{k|k-1} - \mathbf{K}_k \mathbf{P}_{k|k-1}^y \mathbf{K}_k^T.$$

Consideramos el caso $n = 2$, $m = 1$, i.e. $\mathbf{g}(\bar{\mathbf{x}}) = g_1(x_1, x_2)$.

$$D_{\delta} \mathbf{g}(\bar{\mathbf{x}}) = \nabla^T \mathbf{g}(\bar{\mathbf{x}}) \delta = \frac{\partial g_1}{\partial x_1} \delta_1 + \frac{\partial g_1}{\partial x_2} \delta_2$$

$$\begin{aligned} D_{\delta}^2 \mathbf{g}(\bar{\mathbf{x}}) &= D_{\delta} [\nabla^T \mathbf{g}(\bar{\mathbf{x}}) \delta] = \nabla^T [\nabla^T \mathbf{g}(\bar{\mathbf{x}}) \delta] \delta \\ &= \left[\frac{\partial^2 g_1}{\partial x_1^2} \delta_1 + \frac{\partial^2 g_1}{\partial x_1 \partial x_2} \delta_2, \frac{\partial^2 g_1}{\partial x_2 \partial x_1} \delta_1 + \frac{\partial^2 g_1}{\partial x_2^2} \delta_2 \right] \begin{bmatrix} \delta_1 \\ \delta_2 \end{bmatrix} \\ &= \frac{\partial^2 g_1}{\partial x_1^2} \delta_1^2 + \frac{\partial^2 g_1}{\partial x_1 \partial x_2} \delta_1 \delta_2 + \frac{\partial^2 g_1}{\partial x_2 \partial x_1} \delta_1 \delta_2 + \frac{\partial^2 g_1}{\partial x_2^2} \delta_2^2 \\ &= \text{Tr} \left(\begin{bmatrix} \frac{\partial^2 g_1}{\partial x_1^2} & \frac{\partial^2 g_1}{\partial x_1 \partial x_2} \\ \frac{\partial^2 g_1}{\partial x_1 \partial x_2} & \frac{\partial^2 g_1}{\partial x_2^2} \end{bmatrix} \begin{bmatrix} \delta_1^2 & \delta_1 \delta_2 \\ \delta_1 \delta_2 & \delta_2^2 \end{bmatrix} \right) = \text{Tr}(\nabla^2 g_1(\bar{\mathbf{x}}) \delta \delta^T) \end{aligned}$$