Introduction aux processus stochastiques

Examen 2e session — 2007

Durée : 3h30.

Pas de gsm ni de calculatrice.

1 Chaînes de Markov

Suites et sous-suites de nombres.

Soit une chaîne de Markov invariante dans le temps définie sur 5 états.
La chaîne a une distribution initiale π et une matrice de transition Π.
Soit $\mathcal{X}_t$ la séquence des états de la chaîne, $t = 1, 2, \ldots$.
En notant les numéros correspondant aux 10 premiers états réalisés de la chaîne, on obtient une suite de 10 nombres $X_1, \ldots, X_{10}$, avec $1 \leq X_t \leq 5$. On demande :

1. La probabilité que le dernier nombre soit 2.
2. La probabilité de l'événement "Le 7e nombre est impair."

Ensuite, on considère une seconde suite $Y_1, \ldots, Y_{10}$ de 10 nombres, obtenue en extrayant des 20 premiers états $X_1, \ldots, X_{20}$ de la chaîne ceux des instants pairs. Autrement dit,

$$Y_1 = X_2, Y_2 = X_4, \ldots, Y_{10} = X_{20}.$$

On demande, pour la suite des Y :

3. La probabilité de l'événement "Le 3e nombre est 4 et le 6e est strictement inférieur à 3."

Enfin, on considère une troisième suite de 10 nombres $Z_1, \ldots, Z_{10}$ obtenue en notant uniquement les états 2, 3, 4, 5 de la suite des états de la chaîne $X_1, X_2, \ldots$
Ainsi on a $2 \leq Z_t \leq 5$, et la chaîne de Markov doit effectuer des transitions jusqu'à ce que la suite des Z compte 10 éléments.
Par exemple, au départ de 1 4 3 5 4 4 2 1 1 3 1 3 2 1 5 on tire la suite 4 3 5 4 4 2 3 3 2 5.
On demande :

4. De montrer que la suite des Z peut être obtenue en notant la séquence des états d'une chaîne de Markov invariante dans le temps (π', Π'). Pour ce faire, donner la distribution initiale π' et la matrice de transition Π' en fonction de π et Π.
   Travailler élément par élément si nécessaire (i.e. définir π'_i et Π'_{ij} pour $1 \leq i, j \leq 5$.)

2 Chaînes de Markov cachées

Inférence à partir d'observations d'état bruitées.

Soit une chaîne de Markov cachée invariante dans le temps (π, Π, Σ) à 3 états S_1, S_2, S_3 et 3 observations possibles E_1, E_2, E_3.
On note $\mathcal{X}_t$ la séquence des états cachés et $\mathcal{Y}_t$ la séquence des observations, pour t = $1, 2, \ldots, N$.
Pour rappel,

$$\pi_i = P(\mathcal{X}_1 = S_i) \qquad \Pi_{ij} = P(\mathcal{X}_{t+1} = S_j | \mathcal{X}_t = S_i) \qquad \Sigma_{ij} = P(\mathcal{Y}_t = E_j | \mathcal{X}_t = S_i).$$

On considère le cas particulier

$$\pi = \begin{bmatrix} 1 & 1 & 1 \end{bmatrix}/3 \qquad \Pi = \begin{bmatrix} .7 & .3 & 0 \\ 0 & .9 & .1 \\ .4 & 0 & .6 \end{bmatrix} \qquad \Sigma = \begin{bmatrix} .8 & .1 & .1 \\ .1 & .8 & .1 \\ .1 & .1 & .8 \end{bmatrix}.$$

On dispose d'une séquence $Y_1\, Y_2 \ldots Y_N$ d'observations, où chaque $Y_i \in \{E_1, E_2, E_3\}$.
On considère le cas particulier où $N = 3$ et la séquence observée est

$$E_3\; E_2\; E_1.$$

Autrement dit, on a $Y_1 = E_3$, $Y_2 = E_2$ et $Y_3 = E_1$.

Dans les questions qui viennent, la résolution numérique n'est pas demandée.

1. Dessiner le diagramme de transition d'états de la chaîne de Markov (π, Π).
2. Dessiner le réseau bayesien de la chaîne de Markov cachée (π, Π, Σ) pour les 3 premiers instants.
3. *Propagation α/Estimation d'état.*
   Calculer de façon efficace les probabilités $P(\mathcal{X}_N = S_i | \mathcal{Y}_1 = Y_1, \ldots, \mathcal{Y}_N = Y_N)$, pour $i \in \{1, 2, 3\}$ et $N = 3$.
   En notation abrégée, la question revient à demander $P(\mathcal{X}_N | Y_1 \ldots Y_N)$ pour $N = 3$.
   Suggestion : se ramener au calcul de $P(\mathcal{X}_t = S_i, \mathcal{Y}_1 = Y_1, \ldots, \mathcal{Y}_t = Y_t)$, c'est-à-dire en notation abrégée $P(\mathcal{X}_t, Y_1 \ldots Y_t)$, pour t valant successivement 1,2,3.
4. *Propagation β/Filtrage.*
   Calculer de façon efficace $P(\mathcal{X}_t | Y_1 \ldots Y_N)$ pour $t = 1$ et $t = 2$.
   Suggestion : se ramener au calcul de $P(\mathcal{Y}_{t+1} = Y_{t+1}, \ldots, \mathcal{Y}_N = Y_N | \mathcal{X}_t = S_i)$, c'est-à-dire $P(Y_{t+1} \ldots Y_N | \mathcal{X}_t)$, et exploiter le calcul de $P(\mathcal{X}_t, Y_1 \ldots Y_t)$ du point précédent.
5. *Prédiction.*
   Calculer $P(\mathcal{X}_t | Y_1 \ldots Y_N)$ pour $N = 3$ et $t \in \{4, 5\}$.
6. Supposer à présent qu'on n'observe que $Y_1 = E_3$ et $Y_3 = E_1$.
   Calculer $P(\mathcal{X}_2 | Y_1, Y_3)$.
   Suggestion : Modifier les algorithmes de propagation établis aux points précédents.

3 Estimation statique

Cascade de senseurs.

Soit x un vecteur source de dimension n, distribué selon $\mathcal{N}(\bar{x}, \Sigma_x)$.
Soit $y = Cx + v$ un premier vecteur de senseurs de dimension m_y,
Soit $z = Fx + Gy + w$ un second vecteur de senseurs de dimension m_z.
Le vecteur de bruit v suit une normale $\mathcal{N}(0, V)$, le vecteur de bruit w suit une normale $\mathcal{N}(0, W)$.
Les vecteurs x,v,w sont mutuellement indépendents.
On demande de déterminer :

1. L'estimateur $\hat{x}_{(1)} = \mathbf{E}\{x|y\}$
   et l'espérance de l'erreur quadratique d'estimation $\mathbf{E}\{||(x - \hat{x}_{(1)})||^2\}$.
2. L'estimateur $\hat{x}_{(2)} = \mathbf{E}\{x|y, z\}$
   et l'espérance de l'erreur quadratique d'estimation $\mathbf{E}\{||(x - \hat{x}_{(2)})||^2\}$.
3. L'estimateur $\hat{z} = \mathbf{E}\{z|y\}$
   et l'espérance de l'erreur quadratique d'estimation $\mathbf{E}\{||(z - \hat{z})||^2\}$.

Solutions

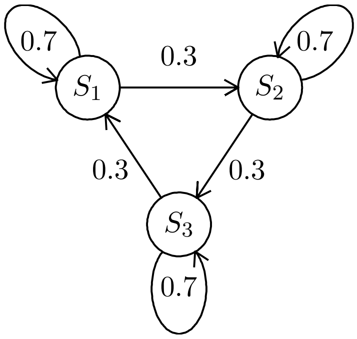
Suites et sous-suites de nombres.

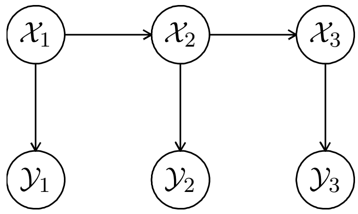
1. La probabilité que le dernier nombre soit 2 :
   $\pi\Pi^9 \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \end{bmatrix}^T$.
2. La probabilité de l'événement "Le 7e nombre est impair" :
   $\pi\Pi^6 \begin{bmatrix} 1 & 0 & 1 & 0 & 1 \end{bmatrix}^T$.
3. La probabilité que 3e nombre soit 4 et le 6e nombre strictement inférieur à 3 :
   $(\pi\Pi)(\Pi^2)^2 \begin{bmatrix} 0 & 0 & 0 & 1 & 0 \end{bmatrix}^T \begin{bmatrix} 0 & 0 & 0 & 1 & 0 \end{bmatrix} (\Pi^2)^3 \begin{bmatrix} 1 & 1 & 0 & 0 & 0 \end{bmatrix}^T$.
4. On a $\pi'_1 = 0$ et $\pi'_j = \pi_j + \dfrac{\pi_1}{1-\Pi_{11}}\Pi_{1j}$ pour $2 \leq j \leq 5$;
   $\Pi'_{i1} = 0$ pour $2 \leq i \leq 5$; $\Pi'_{ij} = \Pi_{ij} + \dfrac{\Pi_{i1}}{1-\Pi_{i1}}\Pi_{1j}$ pour $2 \leq i, j \leq 5$.
   On peut définir Π_{1j} assez librement car c'est sans incidence sur la chaîne.
   Par exemple, $\Pi_{11} = 1$ et $\Pi_{1j} = 0$ pour $2 \leq j \leq 5$.
   On vérifie que

$$\begin{aligned}\sum_{j=1}^{5} \pi'_j = \sum_{j=2}^{5} \pi'_j = \sum_{j=2}^{5} \pi_j + \frac{\pi_1}{1-\Pi_{11}} \sum_{j=2}^{5} \Pi_{1j} \\ = \sum_{j=2}^{5} \pi_j + \frac{\pi_1}{1-\Pi_{11}} (\sum_{j=1}^{5} \Pi_{1j} - \Pi_{11}) = \sum_{j=2}^{5} \pi_j + \frac{\pi_1}{1-\Pi_{11}} (1-\Pi_{11}) = 1.\end{aligned}$$

   De façon similaire, $\sum_{j=1}^{5} \Pi'_{ij} = 1$ pour $2 \leq i \leq 5$.

Inférence à partir d'observations d'état bruitées.

1. 

2. 

3. Comme

$$P(\mathcal{X}_t = S_i | Y_1 \dots Y_N) = \frac{P(\mathcal{X}_t = S_i, Y_1 \dots Y_N)}{\sum_{k=1}^3 P(\mathcal{X}_t = S_k, Y_1 \dots Y_N)},$$

tout se ramène bien à calculer efficacement $P(\mathcal{X}_N = S_i, Y_1 \dots Y_N)$.
Pour y parvenir, on va partir de l'état 1 et ajouter les observations séquentiellement.
Sans observation,

$$\hat{\alpha}_1 = \left[\begin{array}{ccc} P(\mathcal{X}_1 = S_1) & P(\mathcal{X}_1 = S_2) & P(\mathcal{X}_1 = S_3) \end{array}\right] = \pi.$$

Avec Y_1,

$$\begin{aligned}\alpha_1 &= \left[\begin{array}{ccc} P(\mathcal{X}_1 = S_1, Y_1) & P(\mathcal{X}_1 = S_2, Y_1) & P(\mathcal{X}_1 = S_3, Y_1) \end{array}\right] \\ &= \left[\begin{array}{ccc} P(\mathcal{X}_1 = S_1)P(Y_1|\mathcal{X}_1 = S_1) & \dots & P(\mathcal{X}_1 = S_3)P(Y_1|\mathcal{X}_1 = S_3) \end{array}\right] \\ &= \hat{\alpha}_1 * (\Sigma \mathbf{e}_{Y_1})^T\end{aligned}$$

où $*$ est le produit de Hadamard (produit élément par élément) et $\mathbf{e}_{Y_1}$ un vecteur colonne dont tous les éléments sont nuls, sauf le Y_1-ème qui vaut 1. $(\Sigma \mathbf{e}_{Y_1})$ est la Y_1-ème colonne de Σ.
Etat suivant sans observation supplémentaire :

$$\begin{aligned}\hat{\alpha}_2 &= \left[\begin{array}{ccc} P(\mathcal{X}_2 = S_1, Y_1) & P(\mathcal{X}_2 = S_2, Y_1) & P(\mathcal{X}_2 = S_3, Y_1) \end{array}\right] \\ &= \left[\begin{array}{ccc} \sum_{k=1}^3 P(\mathcal{X}_1 = S_k, \mathcal{X}_2 = S_1, Y_1) & \dots & \sum_{k=1}^3 P(\mathcal{X}_1 = S_k, \mathcal{X}_2 = S_3, Y_1) \end{array}\right] \\ &= \left[\begin{array}{ccc} \sum_{k=1}^3 P(\mathcal{X}_1 = S_k, Y_1)P(\mathcal{X}_2 = S_1|\mathcal{X}_1 = S_k) & \dots & \sum_{k=1}^3 P(\mathcal{X}_1 = S_k, Y_1)P(\mathcal{X}_2 = S_3|\mathcal{X}_1 = S_k) \end{array}\right] \\ &= \alpha_1 \Pi.\end{aligned}$$

Ajout de l'observation Y_2 :

$$\begin{aligned}\alpha_2 &= \left[\begin{array}{ccc} P(\mathcal{X}_2 = S_1, Y_1, Y_2) & P(\mathcal{X}_2 = S_2, Y_1, Y_2) & P(\mathcal{X}_2 = S_3, Y_1, Y_2) \end{array}\right] \\ &= \left[\begin{array}{ccc} P(\mathcal{X}_2 = S_1, Y_1)P(Y_2|\mathcal{X}_2 = S_1) & \dots & P(\mathcal{X}_2 = S_3, Y_1)P(Y_2|\mathcal{X}_2 = S_3). \end{array}\right] \\ &= \hat{\alpha}_2 * (\Sigma \mathbf{e}_{Y_2})^T\end{aligned}$$

Etat suivant sans observation supplémentaire :

$$\hat{\alpha}_3 = \left[\begin{array}{ccc} P(\mathcal{X}_3 = S_1, Y_1, Y_2) & \dots & P(\mathcal{X}_3 = S_3, Y_1, Y_2) \end{array}\right] = \alpha_2 \Pi.$$

Ajout de l'observation Y_3 :

$$\alpha_3 = \left[\begin{array}{ccc} P(\mathcal{X}_3 = S_1, Y_1, Y_2, Y_3) & \ldots & P(\mathcal{X}_3 = S_3, Y_1, Y_2, Y_3) \end{array} \right] = \hat{\alpha}_3 * (\Sigma \mathbf{e}_{Y_3})^T.$$

Finalement,

$$p_3 = \left[\begin{array}{ccc} P(\mathcal{X}_3 = S_1 | Y_1, Y_2, Y_3) & \ldots & P(\mathcal{X}_3 = S_3 | Y_1, Y_2, Y_3) \end{array} \right] = \alpha_3 / (\alpha_3 \mathbf{1}),$$

où $\mathbf{1} = [\begin{array}{ccc} 1 & 1 & 1 \end{array}]^T$.

Avec les observations particulières de l'énoncé, on aboutit à l'algorithme suivant :

$$\begin{aligned}
\hat{\alpha_1} &= \pi = [1\ 1\ 1]/3, \\
\alpha_1 &= \hat{\alpha_1} * (\Sigma \mathbf{e}_3)^T \\
\hat{\alpha}_2 &= \alpha_1 \Pi \\
\alpha_2 &= \hat{\alpha}_2 * (\Sigma \mathbf{e}_2)^T \\
\hat{\alpha}_3 &= \alpha_2 \Pi \\
\alpha_3 &= \hat{\alpha}_3 * (\Sigma \mathbf{e}_1)^T \\
p_3 &= \alpha_3 / (\alpha_3 \mathbf{1}),
\end{aligned}$$

où $\mathbf{e}_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \mathbf{e}_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \mathbf{e}_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}, \mathbf{1} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}.$

4. Comme

$$\begin{aligned}
P(\mathcal{X}_t = S_i | Y_1 \ldots Y_N) &= \frac{P(\mathcal{X}_t = S_i, Y_1 \ldots Y_N)}{\sum_{k=1}^3 P(\mathcal{X}_t = S_k, Y_1 \ldots Y_N)}, \\
P(\mathcal{X}_t = S_i, Y_1 \ldots Y_N) &= P(\mathcal{X}_t = S_i, Y_1 \ldots Y_t) P(Y_{t+1} \ldots Y_N | \mathcal{X}_t = S_i),
\end{aligned}$$

et comme $P(\mathcal{X}_t = S_i, Y_1 \ldots Y_t)$ est l'élément i de α_t, il reste à calculer $P(Y_{t+1} \ldots Y_N | \mathcal{X}_t = S_i)$.

On part de la dernière observation Y_3 :

$$\hat{\beta}_3 = \left[\begin{array}{ccc} P(Y_3 | \mathcal{X}_3 = S_1) & P(Y_3 | \mathcal{X}_3 = S_2) & P(Y_3 | \mathcal{X}_3 = S_3) \end{array} \right] = (\Sigma \mathbf{e}_{Y_3})^T.$$

On fait remonter le conditionnement à l'état précédent :

$$\begin{aligned}
\beta_2 &= \left[\begin{array}{ccc} P(Y_3 | \mathcal{X}_2 = S_1) & P(Y_3 | \mathcal{X}_2 = S_2) & P(Y_3 | \mathcal{X}_2 = S_3) \end{array} \right] \\
&= \left[\begin{array}{ccc} \sum_{k=1}^3 P(\mathcal{X}_3 = S_k, Y_3 | \mathcal{X}_2 = S_1) & \ldots & \sum_{k=1}^3 P(\mathcal{X}_3 = S_k, Y_3 | \mathcal{X}_2 = S_3) \end{array} \right] \\
&= \left[\begin{array}{ccc} \sum_{k=1}^3 P(\mathcal{X}_3 = S_k | \mathcal{X}_2 = S_1) P(Y_3 | \mathcal{X}_3 = S_k) & \ldots & \sum_{k=1}^3 P(\mathcal{X}_3 = S_k | \mathcal{X}_2 = S_3) P(Y_3 | \mathcal{X}_3 = S_k) \end{array} \right] \\
&= \hat{\beta}_3 \Pi^T.
\end{aligned}$$

On ajoute l'observation Y_2 :

$$\begin{aligned}
\hat{\beta}_2 &= \left[\begin{array}{ccc} P(Y_2, Y_3 | \mathcal{X}_2 = S_1) & \ldots & P(Y_2, Y_3 | \mathcal{X}_2 = S_3) \end{array} \right] \\
&= \left[\begin{array}{ccc} P(Y_3 | \mathcal{X}_2 = S_1) P(Y_2 | \mathcal{X}_2 = S_1) & \ldots & P(Y_3 | \mathcal{X}_2 = S_3) P(Y_2 | \mathcal{X}_2 = S_3) \end{array} \right] \\
&= \beta_2 * (\Sigma \mathbf{e}_{Y_2})^T.
\end{aligned}$$

On fait remonter le conditionnement à l'état précédent :

$$\beta_1 = \left[\begin{array}{ccc} P(Y_2, Y_3|\mathcal{X}_1 = S_1) & \ldots & P(Y_2, Y_3|\mathcal{X}_1 = S_3) \end{array}\right] = \hat{\beta}_2 \Pi^T.$$

On peut alors calculer

$$\begin{aligned}
\hat{p}_t &= \left[\begin{array}{ccc} P(\mathcal{X}_t = S_1, Y_1 \ldots Y_N) & \ldots & P(\mathcal{X}_t = S_3, Y_1 \ldots Y_N) \end{array}\right] = \alpha_t * \beta_t \\
p_t &= \left[\begin{array}{ccc} P(\mathcal{X}_t = S_1|Y_1 \ldots Y_N) & \ldots & P(\mathcal{X}_t = S_3|Y_1 \ldots Y_N) \end{array}\right] = \hat{p}_t/(\hat{p}_t \mathbf{1}).
\end{aligned}$$

Avec les observations particulières de l'énoncé, on aboutit à l'algorithme suivant :

$$\begin{aligned}
&\hat{\beta}_3 = (\Sigma \mathbf{e}_1)^T \\
&\beta_2 = \hat{\beta}_3 \Pi^T \\
&\hat{\beta}_2 = \beta_2 * (\Sigma \mathbf{e}_2)^T \\
&\beta_1 = \hat{\beta}_2 \Pi^T \\
&\hat{p}_1 = \alpha_1 * \beta_1 && \hat{p}_2 = \alpha_2 * \beta_2 \\
&p_1 = \hat{p}_1/(\hat{p}_1 \mathbf{1}) && p_2 = \hat{p}_2/(\hat{p}_2 \mathbf{1}),
\end{aligned}$$

avec les α_1, α_2 obtenus au point précédent.

5. On utilise la propagation avant, sans ajout d'observations, en partant de p_N.
   En général, pour $t > N$,

$$p_t = [\; P(\mathcal{X}_t = S_1|Y_1 \ldots Y_N) \quad \ldots \quad P(\mathcal{X}_t = S_3|Y_1 \ldots Y_N)] = p_N \Pi^{t-N}.$$

   Ici,

$$p_4 = p_3 \Pi \qquad\qquad p_5 = p_3 \Pi^2,$$

   avec p_3 calculé au point 3.

6. Il suffit de sauter dans les algorithmes de propagation l'étape correspondant à l'observation de Y_2.
   L'algorithme résultant est le suivant :

$$\begin{aligned}
&\hat{\alpha}_1 = \pi = [1\ 1\ 1]/3 && \hat{\beta}_3 = (\Sigma \mathbf{e}_1)^T \\
&\alpha_1 = \hat{\alpha}_1 * (\Sigma \mathbf{e}_3)^T && \beta_2 = \hat{\beta}_3 \Pi^T \\
&\hat{\alpha}_2 = \alpha_1 \Pi \\
&\alpha'_2 = \hat{\alpha}_2 \\
&\hat{p}'_2 = \alpha'_2 * \beta_2
\end{aligned}$$

et $p'_2 = [\; P(\mathcal{X}_2 = S_1, Y_1, Y_3) \quad P(\mathcal{X}_2 = S_2, Y_1, Y_3) \quad P(\mathcal{X}_2 = S_3, Y_1, Y_3)\;] = \hat{p}'_2/(\hat{p}'_2 \mathbf{1})$.

Cascade de senseurs.

1. On a $\hat{x}_{(1)} = \mathbf{E}\{x|y\} = \bar{x} + \Sigma_{xy}\Sigma_y^{-1}(y - \bar{y}) = \bar{x} + \Sigma_x C^T(C\Sigma_x C^T + V)(y - C\bar{x})$
   et $\mathbf{E}\{(x - \hat{x}_{(1)})(x - \hat{x}_{(1)})^T\} = \Sigma_x - \Sigma_{xy}\Sigma_y^{-1}\Sigma_{xy}^T = \Sigma_x - \Sigma_x C^T(C\Sigma_x C^T + V)^{-1}C\Sigma_x$.
   L'espérance de l'erreur quadratique d'estimation vaut la trace de $\mathbf{E}\{(x - \hat{x}_{(1)})(x - \hat{x}_{(1)})^T\}$.
2. Par substitution de l'expression de y dans z, on a

$$z = Fx + G(Cx + v) + w = (F + GC)x + Gv + w.$$

   On pose $A = F + GC$ et $u = Gv + w$, de sorte que $z = Ax + u$.
   On pose ensuite

$$y' = \begin{bmatrix} y \\ z \end{bmatrix} = \begin{bmatrix} C \\ A \end{bmatrix} x + \begin{bmatrix} v \\ u \end{bmatrix}, \qquad C' = \begin{bmatrix} C \\ A \end{bmatrix}, \qquad v' = \begin{bmatrix} v \\ u \end{bmatrix},$$

   de sorte que $y' = C'x + v'$.
   Le bruit v' suit $\mathcal{N}(0, V')$ avec $V' = \begin{bmatrix} V & VG^T \\ GV^T & GVG^T + W \end{bmatrix}$.
   On peut à présent recycler les formules du point précédent en remplaçant C par C' et V par V'.
   On a ainsi

$$\begin{aligned}
\hat{x}_{(2)} &= \mathbf{E}\{x|y,z\} = \mathbf{E}\{x|y'\} = \bar{x} + \Sigma_{xy'}\Sigma_{y'}^{-1}(y' - \bar{y}') \\
&= \bar{x} + \Sigma_x C'^T(C'\Sigma_x C'^T + V')^{-1}(y' - C'\bar{x}) \\
&= \bar{x} + \Sigma_x \begin{bmatrix} C^T & A^T \end{bmatrix} \left(\begin{bmatrix} C \\ A \end{bmatrix} \Sigma_x \begin{bmatrix} C^T & A^T \end{bmatrix} + \begin{bmatrix} V & VG^T \\ GV^T & GVG^T + W \end{bmatrix} \right)^{-1} \left(\begin{bmatrix} y \\ z \end{bmatrix} - \begin{bmatrix} C \\ A \end{bmatrix} \bar{x} \right) \\
&= \bar{x} + \begin{bmatrix} \Sigma_x C^T & \Sigma_x A^T \end{bmatrix} \begin{bmatrix} C\Sigma_x C^T + V & C\Sigma_x A^T + VG^T \\ A\Sigma_x C^T + GV^T & A\Sigma_x A^T + GVG^T + W \end{bmatrix}^{-1} \left(\begin{bmatrix} y \\ z \end{bmatrix} - \begin{bmatrix} C \\ A \end{bmatrix} \bar{x} \right)
\end{aligned}$$

   avec $A = F + GC$.
   L'espérance de l'erreur moyenne quadratique est la trace de

$$\mathbf{E}\{(x - \hat{x}_{(2)})(x - \hat{x}_{(2)})^T\} = \Sigma_x - \Sigma_x C'^T(C'\Sigma_x C'^T + V')^{-1}C'\Sigma_x$$

   avec substitutions similaires.
3. On a $\hat{z} = \mathbf{E}\{z|y\} = \bar{z} + \Sigma_{zy}\Sigma_y^{-1}(y - \bar{y})$
   et $\mathbf{E}\{||z - \hat{z}||^2\} = \text{tr}\{\Sigma_z - \Sigma_{zy}\Sigma_y^{-1}\Sigma_{yz}\}$.
   Les matrices Σ_y, Σ_{zy} et Σ_z peuvent être déduites du point précédent en remarquant que $\Sigma_{y'}$ peut être partitionnée :

$$\Sigma_{y'} = \begin{bmatrix} \Sigma_y & \Sigma_{yz} \\ \Sigma_{zy} & \Sigma_z \end{bmatrix},$$

   et par identification on tire

$$\Sigma_y = C\Sigma_x C^T + V \qquad \Sigma_{zy} = A\Sigma_x C^T + GV^T \qquad \Sigma_z = A\Sigma_x A^T + GVG^T + W.$$

   Ainsi,

$$\begin{aligned}
\hat{z} &= \mathbf{E}\{z|y\} = \bar{z} + \Sigma_{zy}\Sigma_y^{-1}(y - \bar{y}) \\
&= \bar{z} + (A\Sigma_x C^T + GV^T)(C\Sigma_x C^T + V)^{-1}(y - C\bar{x}),
\end{aligned}$$

$$
\begin{aligned}
\mathbf{E}\{||z-\hat{z}||^2\} &= \mathrm{tr}\{\Sigma_z - \Sigma_{zy}\Sigma_y^{-1}\Sigma_{yz}\} \\
&= \mathrm{tr}\{A\Sigma_x A^T + GVG^T + W - (A\Sigma_x C^T + GV^T)(C\Sigma_x C^T + V)^{-1}(C\Sigma_x A^T + VG^T)\}.
\end{aligned}
$$

Une variante consiste à partir de l'expression de y, z en fonction de x, v, w

$$
y' = \begin{bmatrix} y \\ z \end{bmatrix} = \begin{bmatrix} C & I_{m_y} & 0 \\ A & G & I_{m_z} \end{bmatrix} \begin{bmatrix} x \\ v \\ w \end{bmatrix},
$$

puis à calculer directement $\Sigma_{y'}$ à partir de sa définition :

$$
\begin{aligned}
\mathbf{E}\left\{ \begin{bmatrix} y-\bar{y} \\ z-\bar{z} \end{bmatrix} \begin{bmatrix} y-\bar{y} \\ z-\bar{z} \end{bmatrix}^T \right\} &= \begin{bmatrix} C & I_{m_y} & 0 \\ A & G & I_{m_z} \end{bmatrix} \begin{bmatrix} \Sigma_x & 0 & 0 \\ 0 & V & 0 \\ 0 & 0 & W \end{bmatrix} \begin{bmatrix} C^T & A^T \\ I_{m_y} & G^T \\ 0^T & I_{m_z} \end{bmatrix} \\
&= \begin{bmatrix} C\Sigma_x C^T + V & C\Sigma_x A^T + VG^T \\ A\Sigma_x C^T + GV^T & A\Sigma_x A^T + GVG^T + W \end{bmatrix}.
\end{aligned}
$$