

5. MÉTHODES VARIATIONNELLES : CONTRÔLE OPTIMAL, ÉTAT ADJOINT

Nous allons étudier dans cette partie le 4D-VAR, qui est une généralisation du 3D-VAR pour des observations distribuées dans le temps. On tient donc compte désormais d'un modèle d'évolution (dans le temps) qui permettra de comparer l'état du système avec les observations à l'instant approprié.

La fenêtre d'assimilation est un intervalle de temps donné, l'analyse est réalisée à l'instant initial, et on suppose les observations distribuées sur m instants $(t_i)_{0 \leq i \leq m}$ dans l'intervalle. On notera $y(t_i)$ les observations, $x(t_i)$ l'état du système, et $x_t(t_i)$ l'état vrai du système, à l'instant t_i . La matrice de covariance des erreurs d'observations à l'instant t_i est notée R_i . L'opérateur d'observation correspondant est noté H_i . Par contre, la matrice de covariance d'erreur d'ébauche est toujours notée B , car elle n'est définie qu'à l'instant initial, l'ébauche x_b étant une estimation a priori de l'analyse, donc à l'instant initial.

5.1. 4D-VAR : assimilation variationnelle en dimension 4. Dans sa formulation générale, le principe du 4D-VAR est de considérer la minimisation de la fonction coût suivante :

$$(12) \quad J(x_0) = (x_0 - x_b)^T B^{-1} (x_0 - x_b) + \sum_{i=0}^m (y_i - H_i(x(t_i)))^T R_i^{-1} (y_i - H_i(x(t_i))).$$

La minimisation de cette fonction coût est réalisée sous contrainte. Il s'agit d'une contrainte forte de modèle, puisque l'écriture de $J(x)$ dépend des valeurs $x(t_i)$, qui elles-mêmes dépendent de la condition initiale x_0 .

D'un point de vue *continu*, on peut supposer que le modèle dynamique d'évolution s'écrit de la façon suivante :

$$(13) \quad \frac{dx}{dt} = F(x), \quad x(0) = x_0.$$

Mais on peut également écrire le problème de façon *discrète*, en supposant que la suite des états du modèle $x(t_i)$ s'écrit de la façon suivante :

$$(14) \quad x(t_i) = M_{0 \rightarrow t_i}(x_0),$$

où $M_{0 \rightarrow t_i}$ est l'opérateur modèle (ou la résolvante) permettant de passer de l'instant initial à l'instant t_i .

Le 4D-VAR est donc un problème compliqué de minimisation non linéaire sous contrainte. On peut le simplifier à l'aide des deux hypothèses suivantes :

- **Causalité** : le modèle d'évolution peut s'écrire comme une suite de modèles intermédiaires, permettant de passer d'un instant au suivant. On peut supposer que M_0 est l'identité (permettant de passer de l'instant 0 à lui-même), et si on note M_i la résolvante du modèle permettant de passer de l'instant t_{i-1} à t_i , alors $x_i = M_i x_{i-1}$, et donc par récurrence,

$$x_i = M_i M_{i-1} \dots M_1 x_0.$$

- **Approximation linéaire tangente** : la fonction coût peut être rendue quadratique en supposant, en plus de la linéarisation des opérateurs d'observation H_i , que le modèle M peut être linéarisé. On pourra alors remplacer le modèle par son approximation linéaire tangente (ou sa dérivée), et M sera alors le modèle linéaire tangent. Il s'agit du même procédé que pour les opérateurs d'observation H (voir chapitres précédents). Cette hypothèse est généralement valide si la période d'assimilation n'est pas trop longue.

Ces deux hypothèses permettent de se ramener à un problème d'optimisation linéaire et sans contrainte, ce qui est plus simple à résoudre. Le premier terme J_b de J est le même que dans le 3D-VAR, c'est un terme de rappel à l'ébauche, qui agit également comme une régularisation. Le deuxième terme J_o est par contre plus compliqué, puisqu'il inclut la résolution du modèle d'évolution.

5.2. Modèle adjoint et minimisation. Le calcul de $J(x_0)$ semble complexe à cause du modèle à intégrer, et celui de son gradient paraît encore plus délicat. Toutefois, il est possible de calculer très rapidement celui-ci.

Théorème 5.1. *L'évaluation de la fonction coût du 4D-VAR $J(x_0)$ et de son gradient $\nabla J(x_0)$ nécessite une intégration du modèle direct, de l'instant initial jusqu'à l'instant final, et une intégration du modèle adjoint.*

Démonstration. La première étape est d'intégrer le modèle direct, de l'instant initial avec x_0 jusqu'à l'instant final pour calculer $x(t_n)$, tout en calculant à chaque étape i intermédiaire :

- les états intermédiaires $x(t_i) = M_i x(t_{i-1})$;
- les vecteurs d'innovation normalisés $d_i = R_i^{-1}(y_i - H_i x(t_i))$ qui sont stockés ;
- les contributions au deuxième terme de la fonction coût $J_{o_i}(x_0) = (y_i - H_i x(t_i))^T d_i$;
- et finalement $J_o(x_0) = \sum_{i=0}^m J_{o_i}(x_0)$.

Il suffit d'ajouter le terme $J_b(x_0)$ qui ne pose aucune difficulté.

Pour calculer le gradient, il faut passer par une réécriture :

$$\begin{aligned} -\frac{1}{2}\nabla J_o(x_0) &= \sum_{i=0}^m M_1^T \dots M_i^T H_i^T d_i \\ &= H_0^T d_0 + M_1^T [H_1^T d_1 + M_2^T [H_2^T d_2 + \dots + M_n^T H_n^T d_n] \dots], \end{aligned}$$

ce qui peut se calculer de la façon suivante :

- initialiser l'état adjoint p à 0 à l'instant final : $p(t_n) = 0$;
- à chaque étape $i - 1$, l'état adjoint $p(t_{i-1})$ est déduit de la valeur de l'état adjoint à l'instant t_i en utilisant le modèle adjoint (qui est l'adjoint du modèle linéaire tangent) entre les instants i et $i - 1$: $p(t_{i-1}) = M_i^T (p(t_i) + H_i^T d_i)$;
- à la fin, on obtient à l'instant initial $-\frac{1}{2}\nabla J_o(x_0) = p(0)$.

Le gradient de J_b ne pose aucun problème. □

Une fois la fonction coût et son gradient calculés, la minimisation est réalisée à l'aide d'une méthode de gradient (du type Newton).

Le modèle adjoint est donc un modèle qui se résout de façon rétrograde en temps, et qui est forcé par les vecteurs d'innovation $H_i^T d_i$, qui dépendent de la distance entre les observations et la trajectoire du modèle direct. On voit notamment que si la trajectoire directe colle parfaitement avec les observations, le forçage devient nul, et comme le modèle adjoint est linéaire et qu'il est initialisé par 0, la valeur du gradient est donc nulle.

En comparaison avec le 3D-VAR, les caractéristiques principales du 4D-VAR sont :

- il fonctionne sous l'hypothèse que le modèle est exact, puisqu'il est vu comme une contrainte forte ;
- il nécessite le modèle adjoint (i.e. les opérateurs M_i^T) ;

- il faut attendre que toutes les observations soient disponibles tout au long de la période d'assimilation avant de commencer la minimisation ;
- une fois l'analyse réalisée (x_a sera le minimum de la fonction coût, ou une approximation après un nombre fini d'itérations dans la minimisation), la prévision est déduite en résolvant simplement le modèle direct, initialisé par l'analyse.

5.3. Lagrangien et système d'optimalité. On peut considérer le problème de façon continue, au lieu de discrétiser l'intervalle de temps en fonction des instants d'observation t_i . La minimisation de la fonction coût J se fait sous la contrainte (forte) de modèle (cf équations 12 et 13). On peut alors définir un Lagrangien pour tenir compte de cette contrainte dans l'optimisation, en réécrivant J comme une fonction à la fois de la condition initiale x_0 , mais également d'une trajectoire $x(t)$, a priori indépendant de x_0 , et on rajoute la contrainte forte que $x(t)$ soit solution du modèle dynamique, avec x_0 comme condition initiale. Plus précisément, on réécrit le problème sous forme de minimisation de la fonction coût suivante :

$$(15) \quad J(x_0, x) = (x_0 - x_b)^T B^{-1}(x_0 - x_b) + \sum_{i=1}^m (y_i - H_i(x(t_i)))^T R_i^{-1}(y_i - H_i(x(t_i)))$$

sous la contrainte que x soit solution du modèle dynamique

$$(16) \quad \frac{dx}{dt} = F(x),$$

initialisé par la condition $x(0) = x_0$.

Le Lagrangien associé à ce problème de minimisation sous contrainte est :

$$(17) \quad \mathcal{L}(x_0, x; p) = J(x_0, x) + \int_0^T p(t) \cdot \left(\frac{dx}{dt} - F(x) \right) dt.$$

On alors le résultat suivant pour faire le lien entre l'optimisation de J et celle du lagrangien \mathcal{L} .

Théorème 5.2.

$$\min_{x_0; x} J(x_0, x) \text{ (sous contrainte (16))} = \min_{x_0; x} \max_p \mathcal{L}(x_0, x; p).$$

Démonstration. Fixons x_0 et x . Soit la contrainte de modèle (16) n'est pas vérifiée par x , auquel cas le maximum sur tous les p du terme intégrale (second terme) de \mathcal{L} est $+\infty$. Soit la contrainte est vérifiée, auquel cas quel que soit p , le terme intégral est nul. Donc le maximum vaut soit $J(x_0, x)$ si la contrainte est satisfaite, soit $+\infty$ si la contrainte n'est pas satisfaite. Comme on prend ensuite le minimum sur x_0 et x , on voit donc que ce minimum vaut exactement le minimum de J pour tous les (x_0, x) qui vérifient la contrainte de modèle. \square

On cherche donc un point-selle $(x_0, x; p)$ du Lagrangien : maximum par rapport à p , minimum par rapport à (x_0, x) .

En faisant des hypothèses sur la convexité des fonctions (qui seront vérifiées notamment si on linéarise le modèle et les opérateurs d'observation, par exemple autour d'une trajectoire de référence), on peut alors chercher à caractériser le point-selle de \mathcal{L} : chacune des dérivées partielles du Lagrangien \mathcal{L} (par rapport à chacune des trois variables) est nulle, à l'optimum.

- Maximum par rapport à p : $\frac{\partial \mathcal{L}}{\partial p}(x_0^*, x^*; p^*) = 0$.

On obtient la dérivée partielle en perturbant le Lagrangien :

$$\lim_{\varepsilon \rightarrow 0} \frac{\mathcal{L}(x_0^*, x^*; p^* + \varepsilon q) - \mathcal{L}(x_0^*, x^*; p^*)}{\varepsilon} = \frac{\partial \mathcal{L}}{\partial p}(x_0^*, x^*; p^*) \cdot q$$

Par linéarité,

$$\mathcal{L}(x_0^*, x^*; p^* + \varepsilon q) - \mathcal{L}(x_0^*, x^*; p^*) = \varepsilon \int_0^T q(t) \cdot \left(\frac{dx}{dt} - F(x) \right) dt,$$

et donc par identification, la condition $\frac{\partial \mathcal{L}}{\partial p} = 0$ est équivalente à $\frac{dx}{dt} - F(x) = 0$ à chaque instant t . On retrouve donc la contrainte de modèle : $x(t)$ doit être une solution du modèle dynamique.

- Minimum par rapport à x : on doit réécrire le Lagrangien afin de faire apparaître la dérivée temporelle sur p et non plus sur x , pour faciliter la dérivée. En intégrant par parties, et en utilisant la condition $x(0) = x_0$, on obtient :

$$\mathcal{L}(x_0, x; p) = J(x_0, x) - \int_0^T \frac{dp}{dt} \cdot x(t) dt - \int_0^T p(t) \cdot F(x) dt + p(T) \cdot x(T) - p(0) \cdot x_0.$$

On calcule maintenant la dérivée par rapport à x :

$$\begin{aligned} \mathcal{L}(x_0, x + \varepsilon y; p) - \mathcal{L}(x_0, x; p) &= J(x_0, x + \varepsilon z) - J(x_0, x) \\ - \varepsilon \int_0^T \frac{dp}{dt} \cdot z(t) dt - \int_0^T p(t) \cdot (F(x + \varepsilon z) - F(x)) dt + \varepsilon p(T) \cdot z(T) - p(0) \cdot (x_0 - x_0). \end{aligned}$$

En utilisant la définition de J , les deux premiers termes valent

$$\sum_{i=1}^m [(y_i - H_i(x(t_i) + \varepsilon z(t_i)))^T R_i^{-1} (y_i - H_i(x(t_i))) - (y_i - H_i(x(t_i)))^T R_i^{-1} (y_i - H_i(x(t_i)))],$$

et en utilisant un développement de Taylor des opérateurs d'observation : $H_i(x(t_i) + \varepsilon z(t_i)) = H_i(x(t_i)) + \varepsilon H'_i(x(t_i))z(t_i) + \mathcal{O}(\varepsilon^2)$, si on divise par ε et qu'on fait tendre ε vers 0, on obtiendra à la limite : $-2 \sum_{i=1}^m (H'_i(x(t_i))z(t_i))^T R_i^{-1} (y_i - H_i(x(t_i)))$, ou encore

$$-2 \sum_{i=1}^m [H'_i(x(t_i))^T R_i^{-1} (y_i - H(x(t_i)))] \cdot z(t_i).$$

La deuxième intégrale va, de la même façon, tendre vers

$$- \int_0^T p(t) \cdot (F'(x)z) dt.$$

À la fin, on obtient :

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial x}(x_0^*, x^*; p^*) \cdot z &= -2 \sum_{i=1}^m [H'_i(x(t_i))^T R_i^{-1} (y_i - H(x(t_i)))] \cdot z(t_i) \\ &\quad - \int_0^T \frac{dp}{dt} \cdot z(t) dt - \int_0^T [F'(x)]^T p(t) \cdot z(t) dt + p(T) \cdot z(T). \end{aligned}$$

Si on impose que cette dérivée est nulle, quelle que soit la fonction z , on voit déjà (en imposant n'importe quelle valeur pour $z(T)$) qu'on doit avoir

$p(T) = 0$. Ensuite, comme on a toute latitude pour choisir $z(t)$ à n'importe quel instant t , on en déduit qu'à chaque instant t , on doit avoir :

$$-\frac{dp}{dt} - [F'(x)]^T p(t) - 2 \sum_{i=1}^m H'_i(x(t_i))^T R_i^{-1} (y_i - H_i(x(t_i))) \delta_{t_i}(t) = 0,$$

ou encore

$$(18) \quad -\frac{dp}{dt} = [F'(x)]^T p(t) + 2 \sum_{i=1}^m H'_i(x(t_i))^T R_i^{-1} (y_i - H_i(x(t_i))) \delta_{t_i}(t),$$

avec la condition finale $p(T) = 0$. L'équation (18) est appelée équation adjointe, et p est l'état adjoint. C'est une équation qu'on résout de façon rétrograde en temps, en partant de la condition finale $p(T) = 0$. Le terme de *modèle* n'est pas $F(x)$ comme dans le modèle direct, mais $[F'(x)]^T p$, qui est l'adjoint du linéarisé de F . Enfin, on force le modèle par un terme source provenant des écarts entre observations et quantités correspondantes, aux instants d'observation.

- En reprenant l'expression du Lagrangien réécrite au début de l'item précédent, la dérivée par rapport à x_0 est triviale :

$$(19) \quad \frac{\partial \mathcal{L}}{\partial x_0}(x_0^*, x^*; p^*) = 2B^{-1}(x_0 - x_b) - p(0).$$

On en déduit le système d'optimalité, qui caractérise la solution optimale :

Modèle direct :

$$\frac{dx}{dt} = F(x), \quad x(0) = x_0$$

Modèle adjoint :

$$-\frac{dp}{dt} = [F'(x)]^T p(t) + 2 \sum_{i=1}^m H'_i(x(t_i))^T R_i^{-1} (y_i - H_i(x(t_i))) \delta_{t_i}(t), \quad p(T) = 0$$

Condition d'optimalité :

$$2B^{-1}(x_0 - x_b) = p(0)$$

En pratique, la fonction coût est minimisée avec un algorithme de type gradient, la fonction coût étant calculée grâce à la résolution du modèle direct (qui permet de calculer la trajectoire $x(t)$), et le gradient est calculé grâce au modèle adjoint. En effet, si la contrainte de modèle est satisfaite, $J(x_0) = \mathcal{L}(x_0, x; p)$ et par conséquent, $\nabla J(x_0) = \frac{\partial \mathcal{L}}{\partial x_0} = 2B^{-1}(x_0 - x_b) - p(0)$ (cf équation (19)).

5.4. Estimation de paramètres. La méthode du 4D-VAR permet d'estimer des paramètres en plus (ou à la place) de la condition initiale. Supposons que le modèle F dépend non seulement de la solution x , mais également d'un jeu de paramètres u . Le modèle direct est alors :

$$\begin{cases} \frac{dx}{dt} = F(x, u), \\ x(0) = x_0. \end{cases}$$

On souhaite désormais identifier le jeu de paramètres u (en plus de la condition initiale x_0 , mais ce n'est pas obligatoire), qui minimise l'écart aux données. La

fonction coût s'écrit alors :

(20)

$$J(x_0, u) = (x_0 - x_b)^T B^{-1}(x_0 - x_b) + (u - u_b)^T Q^{-1}(u - u_b) + \sum_{i=0}^m (y_i - H_i(x(t_i)))^T R_i^{-1} (y_i - H_i(x(t_i))),$$

où u_b représente une ébauche (estimation a priori) de la valeur des paramètres u , Q est la matrice de covariance d'erreur sur cette ébauche. Le terme d'attache aux données dans la fonction coût (dernier terme) ne change pas par rapport à la situation où on cherche à identifier uniquement la condition initiale, mais désormais la contrainte est que $x(t)$ doit être une solution du modèle dépendant des paramètres.

Comme précédemment, on définit un Lagrangien pour réécrire le problème de minimisation sous contrainte :

$$(21) \quad \mathcal{L}(x_0, u, x; p) = J(x_0, u, x) + \int_0^T p(t) \cdot \left(\frac{dx}{dt} - F(x, u) \right) dt,$$

où (comme précédemment) $x(t)$ est désormais vu comme une trajectoire "quelconque", et $p(t)$ est le multiplicateur de Lagrange associé à la contrainte de modèle.

À l'optimum, les dérivées partielles du Lagrangien sont nulles, et on obtient le système d'optimalité suivant :

Modèle direct :

$$\frac{dx}{dt} = F(x, u), \quad x(0) = x_0$$

Modèle adjoint :

$$-\frac{dp}{dt} = \left[\frac{\partial F}{\partial x}(x, u) \right]^T p(t) + 2 \sum_{i=1}^m H'_i(x(t_i))^T R_i^{-1} (y_i - H_i(x(t_i))) \delta_{t_i}(t), \quad p(T) = 0$$

Condition d'optimalité pour la condition initiale :

$$\frac{\partial \mathcal{L}}{\partial x_0} = 2B^{-1}(x_0 - x_b) - p(0) = 0$$

Condition d'optimalité pour les paramètres : on obtient une nouvelle condition en exigeant que la dérivée partielle du Lagrangien par rapport à u s'annule :

$$\frac{\partial \mathcal{L}}{\partial u} = 2Q^{-1}(u - u_b) - \left[\frac{\partial F}{\partial u} \right]^T p = 0.$$

Lorsque la contrainte de modèle est satisfaite, $\mathcal{L}(x_0, u, x, p) = J(x_0, u)$, et donc $\frac{\partial \mathcal{L}}{\partial x_0} = \frac{\partial J}{\partial x_0}$ et $\frac{\partial \mathcal{L}}{\partial u} = \frac{\partial J}{\partial u}$. Le gradient de J par rapport aux paramètres u est donc fourni sans surcoût par l'état adjoint, ce qui permet alors de minimiser la fonction coût par rapport aux paramètres, afin d'identifier le jeu de paramètres optimal u^* .

Si le jeu de paramètres est constant (indépendant du temps), on a

$$\frac{\partial \mathcal{L}}{\partial u} \cdot v = \langle Q^{-1}(u - u_b), v \rangle - \int_0^T \langle [F'_u(x, u)]^T p(t), v \rangle dt = 0$$

pour tout v (vecteur indépendant du temps), ce qui implique que

$$Q^{-1}(u - u_b) - \int_0^T [F'_u(x, u)]^T p(t) dt = 0$$

à l'optimum.

5.5. Préconditionnement. Nous avons vu que dans le 4D-VAR, comme dans le 3D-VAR, nous pouvons calculer *facilement* la fonction coût et son gradient. Pour minimiser la fonctionnelle, on utilisera donc une méthode de descente, dont le but est de mettre à jour le point x en ajoutant une correction proportionnelle à $-\nabla J(x)$.

Mais la méthode de Newton n'est pas la plus efficace, et on utilise plutôt des algorithmes comme le gradient conjugué, ou les méthodes de quasi-Newton. Ces méthodes sont beaucoup plus appropriées quand la fonctionnelle n'est pas quadratique.

Différentes méthodes existent pour améliorer la minimisation : preconditionnement, construction d'une approximation de la matrice Hessienne (méthode de quasi-Newton), méthode incrémentale (voir section suivante), ...

5.5.1. Conditionnement. La principale caractéristique de la fonctionnelle, qui affecte notamment l'efficacité de la minimisation, est le conditionnement. Cette quantité mesure l'ellipticité des iso-surfaces de J , et décrit la difficulté du problème d'optimisation. En effet, moins les iso-surfaces sont circulaires, et plus le gradient ne pointe pas vers le minimum (qui est au centre), ralentissant considérablement la minimisation.

Le conditionnement est défini par le rapport entre la plus grande et la plus petite des valeurs propres de la Hessienne $\nabla^2 J$. Si le conditionnement est proche de 1, la Hessienne sera presque proportionnelle à la matrice identité, et dans ce cas, le minimum peut presque être trouvé en une seule itération car $-\nabla J(x_b)$ pointe directement vers le minimum. En général, J est elliptique, ce qui ralentit le procédé, mais il est possible d'utiliser un opérateur pour preconditionner le problème.

Théorème 5.3. *Si L est un opérateur inversible, alors le problème original :*

$$(22) \quad x_a = \arg \min J(x_0),$$

où on suppose qu'on sait construire $J(x_0)$ et $\nabla J(x_0)$, et en initialisant l'optimisation avec x_b , est équivalent au problème preconditionné suivant :

$$(23) \quad \chi_a = \arg \min \hat{J}(\chi_0),$$

avec $\hat{J}(\chi) = J(L\chi)$, $\nabla \hat{J}(\chi) = L^T \nabla J(L\chi)$, en initialisation l'optimisation avec $L^{-1}x_b$.

La solution est alors donnée par $x_a = L\chi_a$.

Dans le cas du 3D-VAR, un preconditionnement simple et efficace est d'utiliser la matrice de covariance d'erreur d'ébauche B : on décompose $B = LL^T$, et dans ce cas, la nouvelle fonction coût est

$$(24) \quad \hat{J}(\chi) = \chi^T \chi + J_o(L\chi),$$

et on voit donc que le premier terme \hat{J}_b est simplement le produit scalaire canonique. Un autre preconditionneur efficace est la racine carrée de la matrice hessienne, mais qui est beaucoup plus coûteuse à calculer.

5.6. Méthode incrémentale. La méthode incrémentale est une méthode relativement empirique, conçue pour réduire le coût de résolution d'une méthode variationnelle. Il s'agit de remplacer un grand problème compliqué par une série de problèmes plus petits et plus simples. L'idée générale est de rechercher une mise à jour *basse résolution* de l'ébauche *haute résolution*. C'est le même principe que

dans l'interpolation optimale, où on utilise l'influence souvent locale des observations pour corriger localement l'ébauche.

On note (J_h, H_h, x_h) la fonction coût, l'opérateur d'observation, et l'état du modèle, dans leur version à haute résolution. On cherche à minimiser $J_h(x_h)$, et on va résoudre des approximations successives de ce problème. On initialise avec $x_{h,i} = x_b$, et on cherche une mise à jour pour obtenir $x_{h,i+1}$. On passe alors à la boucle interne (à basse résolution) un certain nombre d'informations : les vecteurs d'innovation à haute résolution $d_{h,i} = y - H_h(x_{h,i})$, et une version basse résolution x_i de $x_{h,i}$. On utilise pour cela un opérateur de conversion (généralement d'interpolation/extrapolation) pour passer de la haute à basse résolution : $x_i = S_{h \rightarrow l}(x_{h,i})$.

En linéarisant l'opérateur d'observation à basse résolution H autour de x_i , on obtient H_i et les vecteurs d'innovation à basse résolution sont définis de la façon suivante :

$$y - H(x) = y - H(x_i) - H_i(x - x_i) \simeq y - [H_i(x - x_i) + H_h(x_{h,i})] = d_{h,i} - H_i(x - x_i)$$

de sorte que la fonction coût basse résolution à minimiser dans la boucle interne est

$$J_i(x) = (x - x_b)^T B^{-1}(x - x_b) + [d_{h,i} - H_i(x - x_i)]^T R^{-1}[d_{h,i} - H_i(x - x_i)]$$

qui est quadratique (puisque tout a été linéarisé autour de x_i). Son minimum est noté x_{i+1} . On met alors à jour l'état haute résolution en utilisant un opérateur de conversion basse vers haute résolution :

$$x_{h,i+1} = x_{h,i} + S_{l \rightarrow h}(x_{i+1}) - S_{l \rightarrow h}(x_i),$$

ce qui garantit que l'état à haute résolution (dans la boucle externe) n'est pas modifié si la boucle interne de minimisation ne change pas d'état.

L'essentiel de la minimisation a lieu à basse résolution, sur des approximations linéarisées des opérateurs, ce qui donne une fonction coût quadratique, et donc facile à minimiser. Les mises à jour, et le forçage par le vecteur d'innovation, se font à haute résolution. Dans les implémentations opérationnelles, cette méthode permet d'atteindre beaucoup plus rapidement et efficacement une bonne approximation du minimum de la fonctionnelle haute résolution.