

## 4. 3D-VAR ET INTERPOLATION OPTIMALE

Le 3D-VAR et l'interpolation optimale sont deux méthodes d'assimilation de données très basiques, reposant toutes deux sur l'estimateur moindres carrés qui a été construit au chapitre précédent. On a vu l'équivalence entre l'analyse du BLUE, et la minimisation d'une fonction coût. Ces deux approches servent de base à l'interpolation optimale et au 3D-VAR respectivement. Comme nous allons le voir, la difficulté réside dans le calcul effectif de la solution, surtout en grande dimension.

**4.1. Interpolation optimale.** L'interpolation optimale est une méthode consistant à calculer effectivement les formules de l'analyse du BLUE :

$$(9) \quad x_a = x_b + K(y - Hx_b), \quad K = BH^T(HBH^T + R)^{-1}.$$

En effet, dans un contexte météorologique ou océanographique, la taille du vecteur d'état (et donc des matrices correspondantes) peut atteindre  $n = 10^9$ . Le produit matriciel ou le calcul de matrices inverses peut s'avérer délicat dans des espaces de telle dimension.

L'écriture de l'analyse comme étant l'ébauche corrigée par un incrément, peut être vue comme un système d'équations scalaires, une pour chaque variable du modèle. Pour chaque variable, l'incrément d'analyse est donné par la ligne correspondante de la matrice  $K$ , multipliée par le vecteur d'innovation  $y - Hx_b$ .

L'hypothèse fondamentale de l'interpolation optimale est que pour chaque élément du vecteur d'état, très peu d'observations sont importantes pour déterminer l'incrément d'analyse. L'algorithme est alors le suivant :

- (1) Pour chaque composante  $x(i)$  du vecteur d'état, sélectionner un *petit* nombre  $p_i$  d'observations, jugées significatives pour corriger  $x(i)$ .
- (2) Calculer les composantes correspondantes du vecteur d'innovation  $(y - Hx_b)_i$ , ainsi que les covariances d'erreur entre les  $x(i)$  et le vecteur d'état interpolé aux points d'observation (i.e. les  $p_i$  coefficients significatifs correspondants de la ligne  $i$  de  $BH^T$ ), et les sous-matrices  $p_i \times p_i$  correspondantes des matrices de covariance d'erreur d'ébauche et d'observation (formées par les restrictions de  $HBH^T$  et  $R$  aux observations sélectionnées).
- (3) Inverser la matrice  $p_i \times p_i$ , définie positive, restriction de  $(HBH^T + R)$  aux observations sélectionnées (par exemple par une méthode LU).
- (4) Multiplier cet inverse par la ligne  $i$  de  $BH^T$  pour obtenir la ligne correspondante de  $K$ .

On peut gagner encore en temps de calcul en ne calculant pas l'inverse de la matrice, mais en résolvant le système linéaire correspondant.

Le point crucial réside dans le choix des observations considérées comme significatives. Généralement, la zone d'influence des observations sur l'analyse est assez restreinte spatialement. On choisit en pratique les observations disponibles dans un certain voisinage du point considéré.

**4.2. 3D-VAR.** Le principe du 3D-VAR est de considérer le problème de minimisation de la fonction coût suivante :

$$(10) \quad J(x) = (x - x_b)^T B^{-1}(x - x_b) + (y - Hx)^T R^{-1}(y - Hx),$$

qui, comme nous l'avons vu, fournira la même analyse. Mais comme les matrices (dont celle de gain  $K$ ) peuvent être délicates à manipuler en grande dimension, on préfère résoudre le problème de minimisation en utilisant une méthode de gradient.

Le gradient de la fonction coût est

$$\nabla J(x) = 2B^{-1}(x - x_b) - 2H^T R^{-1}(y - Hx).$$

L'approximation tient ici dans le fait que seul un nombre fini (et faible) d'itérations seront réalisées dans l'algorithme de minimisation. On arrêtera la minimisation soit après un nombre maximal d'itérations, pour garantir un coût de calcul maximal, soit lorsque la norme du gradient de la fonctionnelle aura suffisamment diminué.

En pratique, on utilise l'ébauche  $x_b$  comme point de départ de la minimisation. Et on utilise un algorithme de type Newton pour l'optimisation.

**4.3. Information contenue dans la Hessienne.** On a vu que dans le cas de l'interpolation optimale, si la matrice de gain  $K$  a été calculée, alors la matrice de covariance d'erreurs d'analyse est donnée par

$$A = (I - KH)B(I - KH)^T + KRK^T,$$

qui se résume à  $A = (I - KH)B$  dans le cas où la matrice  $K$  est la matrice de gain optimal.

Dans le cas variationnel, les covariances d'erreur d'analyse peuvent être obtenues grâce à la dérivée seconde de la fonction coût.

**Théorème 4.1.** *La Hessienne de la fonction coût du 3D-VAR donne une information sur les statistiques d'erreur de l'analyse :*

$$(11) \quad A = \left( \frac{1}{2} \nabla^2 J \right)^{-1}.$$

*Démonstration.* On doit calculer la dérivée seconde de la fonction coût :

$$\begin{aligned} J(x) &= (x - x_b)^T B^{-1}(x - x_b) + (y - Hx)^T R^{-1}(y - Hx), \\ \nabla J(x) &= 2B^{-1}(x - x_b) - 2H^T R^{-1}(y - Hx), \\ \nabla^2 J(x) &= 2(B^{-1} + H^T R^{-1}H). \end{aligned}$$

À l'optimum, le gradient est nul, ce qui implique que

$$\begin{aligned} 0 &= B^{-1}(x_a - x_b) - H^T R^{-1}(y - Hx_a) \\ &= B^{-1}(x_a - x_t + x_t - x_b) - H^T R^{-1}(y - Hx_t + H(x_t - x_a)) \\ &= B^{-1}(x_a - x_t) - B^{-1}(x_b - x_t) - H^T R^{-1}(y - Hx_t) - H^T R^{-1}H(x_t - x_a) \end{aligned}$$

et donc

$$(B^{-1} + H^T R^{-1}H)(x_a - x_t) = B^{-1}(x_b - x_t) + H^T R^{-1}(y - Hx_t).$$

Si on multiplie à droite par la transposée de ces vecteurs, et qu'on prend l'espérance mathématique, les termes croisés  $E((x_b - x_t)(y - Hx_t)^T)$  sont nuls, par décorrélation des erreurs. On obtient donc, en utilisant la symétrie des matrices :

$$(B^{-1} + H^T R^{-1}H)A(B^{-1} + H^T R^{-1}H) = B^{-1}BB^{-1} + H^T R^{-1}RR^{-1}H = B^{-1} + H^T R^{-1}H,$$

ce qui prouve le résultat en divisant par  $(B^{-1} + H^T R^{-1}H)$  (qui est inversible).  $\square$