Guess consistency for multi-incremental minimization

Benjamin Ménétrier

Last update: October 29, 2024

Contents

1 Problem linearization

1.1 Full cost function

The full cost function to minimize is defined as:

$$\mathcal{J}(\mathbf{x}) = \frac{1}{2} \left(\mathbf{x} - \mathbf{x}^b\right)^{\mathrm{T}} \mathbf{B}^{-1} \left(\mathbf{x} - \mathbf{x}^b\right) + \frac{1}{2} \left(\mathbf{y}^o - \mathcal{H}(\mathbf{x})\right)^{\mathrm{T}} \mathbf{R}^{-1} \left(\mathbf{y}^o - \mathcal{H}(\mathbf{x})\right) \tag{1}$$

where:

- $\mathbf{x} \in \mathbb{R}^n$ is the state in model space,

- $\mathbf{x}^b \in \mathbb{R}^n$ is the background state,

- $\mathbf{R} \in \mathbb{B}^{n \times n}$ is the background error covariance matrix,

- $\mathbf{y}^o \in \mathbb{R}^p$ is the observation vector,

- $\mathbf{R} \in \mathbb{R}^{p \times p}$ is the observation error covariance matrix,

- $\mathcal{H} : \mathbb{R}^n \to \mathbb{R}^p$ is the nonlinear observation operator.

1.2 Quadratic cost function

Instead of minimizing the full cost function $\mathcal{J}(\mathbf{x})$, we minimize successive quadratic approximations around a guess state $\mathbf{x}_k^g$, for $k \in [1, K]$. The increment $\delta\mathbf{x}_k \in \mathbb{R}^n$ is defined as:

$$\delta\mathbf{x}_k = \mathbf{x} - \mathbf{x}_k^g \tag{2}$$

and is used to linearize the observation operator for $\mathbf{x} \approx \mathbf{x}_k^g$:

$$\mathcal{H}(\mathbf{x}) \approx \mathcal{H}(\mathbf{x}_k^g) + \mathbf{H}_k \delta\mathbf{x}_k \tag{3}$$

where $\mathbf{H}_k \in \mathbb{R}^{p \times n}$ is the observation operator linearized around $\mathbf{x}_k^g$:

$$H_{k,ij} = \left. \frac{\partial \mathcal{H}_i}{\partial x_j} \right|_{\mathbf{x} = \mathbf{x}_k^g} \tag{4}$$

Thus, the quadratic cost function is given by:

$$J\left(\delta\mathbf{x}_k\right) = \frac{1}{2} \left(\delta\mathbf{x}_k - \delta\mathbf{x}_k^b\right)^{\mathrm{T}} \mathbf{B}^{-1} \left(\delta\mathbf{x}_k - \delta\mathbf{x}_k^b\right) + \frac{1}{2} \left(\mathbf{d}_k - \mathbf{H}_k \delta\mathbf{x}_k\right)^{\mathrm{T}} \mathbf{R}^{-1} \left(\mathbf{d}_k - \mathbf{H}_k \delta\mathbf{x}_k\right) \tag{5}$$

where:

- $\delta\mathbf{x}_k^b \in \mathbb{R}^n$ is the background increment:

$$\delta\mathbf{x}_k^b = \mathbf{x}^b - \mathbf{x}_k^g \tag{6}$$

- $\mathbf{d}_k \in \mathbb{R}^p$ is the innovation vector:

$$\mathbf{d}_k = \mathbf{y}^o - \mathcal{H}(\mathbf{x}_k^g) \tag{7}$$

The minimization of each quadratic approximation is called "outer iteration", and should be distinguished from the "inner iterations" required in classical algorithms (conjugate gradient, Lanczos method) to find the minimum of a quadratic approximation.

1.3 Linear system

Setting the gradient of $J(\delta\mathbf{x}_k)$ to zero gives the analysis increment $\delta\mathbf{x}_k^a$:

$$\mathbf{B}^{-1}\left(\delta\mathbf{x}_k^a - \delta\mathbf{x}_k^b\right) - \mathbf{H}_k^{\mathrm{T}}\mathbf{R}^{-1}\left(\mathbf{d}_k - \mathbf{H}_k\delta\mathbf{x}_k^a\right) = 0$$

$$\Leftrightarrow \left(\mathbf{B}^{-1} + \mathbf{H}_k^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{H}_k\right)\delta\mathbf{x}_k^a = \mathbf{B}^{-1}\delta\mathbf{x}_k^b + \mathbf{H}_k^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{d}_k$$

$$\Leftrightarrow \boxed{\mathbf{A}_k^{\mathbf{x}}\delta\mathbf{x}_k^a = \mathbf{b}_k^{\mathbf{x}}} \tag{8}$$

with $\mathbf{A}_k^{\mathbf{x}} \in \mathbb{R}^{n \times n}$ and $\mathbf{b}_k^{\mathbf{x}} \in \mathbb{R}^n$ defined as:

$$\mathbf{A}_k^{\mathbf{x}} = \mathbf{B}^{-1} + \mathbf{H}_k^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{H}_k \tag{9}$$

$$\mathbf{b}_k^{\mathbf{x}} = \mathbf{B}^{-1}\delta\mathbf{x}_k^b + \mathbf{H}_k^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{d}_k \tag{10}$$

1.4 Square-root B preconditioning

Since $\mathbf{B}$ is positive definite, there is an infinity of square-roots $\mathbf{U} \in \mathbb{R}^{n \times m}$ with $m \geq n$ verifying:

$$\mathbf{B} = \mathbf{U}\mathbf{U}^{\mathrm{T}} \tag{11}$$

1.4.1 Classic variable change

A new variable $\delta\mathbf{v}_k \in \mathbb{R}^m$ is defined as:

$$\delta\mathbf{v}_k = \mathbf{U}^{\mathrm{T}}\mathbf{B}^{-1}\delta\mathbf{x}_k \tag{12}$$

leading to

$$\boxed{\delta\mathbf{x}_k = \mathbf{U}\delta\mathbf{v}_k} \tag{13}$$

Using variable change (13) and applying $\mathbf{U}^{\mathrm{T}}$ on both sides, the linear system (8) can be transformed into:

$$\boxed{\mathbf{A}_k^{\mathbf{v}}\delta\mathbf{v}_k^a = \mathbf{b}_k^{\mathbf{v}}} \tag{14}$$

with $\mathbf{A}_k^{\mathbf{v}} \in \mathbb{R}^{m \times m}$ and $\mathbf{b}_k^{\mathbf{v}} \in \mathbb{R}^m$ defined as:

$$\mathbf{A}_k^{\mathbf{v}} = \mathbf{I}_m + \mathbf{U}^{\mathrm{T}} \mathbf{H}_k^{\mathrm{T}} \mathbf{R}^{-1} \mathbf{H}_k \mathbf{U} \tag{15}$$

$$\mathbf{b}_k^{\mathbf{v}} = \delta \mathbf{v}_k^b + \mathbf{U}^{\mathrm{T}} \mathbf{H}_k^{\mathrm{T}} \mathbf{R}^{-1} \mathbf{d}_k \tag{16}$$

with

$$\delta \mathbf{v}_k^b = \mathbf{U}^{\mathrm{T}} \mathbf{B}^{-1} \delta \mathbf{x}_k^b \tag{17}$$

1.4.2 Non-classic variable change

Another new variable $\delta \mathbf{w}_k \in \mathbb{R}^m$ can be defined as:

$$\delta \mathbf{w}_k = \mathbf{U}^{\mathrm{T}} \mathbf{B}^{-1} \left(\mathbf{x}_k - \mathbf{x}^b \right)$$
$$= \mathbf{U}^{\mathrm{T}} \mathbf{B}^{-1} \left(\delta \mathbf{x}_k - \delta \mathbf{x}_k^b \right) \tag{18}$$

leading to

$$\boxed{\delta \mathbf{x}_k = \mathbf{U} \delta \mathbf{w}_k + \delta \mathbf{x}_k^b} \tag{19}$$

The link between $\delta \mathbf{w}_k$ and $\delta \mathbf{v}_k$ is straightforward:

$$\delta \mathbf{v}_k = \delta \mathbf{w}_k + \delta \mathbf{v}_k^b \tag{20}$$

Using variable change (19) and applying $\mathbf{U}^{\mathrm{T}}$ on both sides, the linear system (8) can be transformed into:

$$\boxed{\mathbf{A}_k^{\mathbf{v}} \delta \mathbf{w}_k^a = \mathbf{b}_k^{\mathbf{w}}} \tag{21}$$

with $\mathbf{b}_k^{\mathbf{w}} \in \mathbb{R}^m$ defined as:

$$\mathbf{b}_k^{\mathbf{w}} = \mathbf{U}^{\mathrm{T}} \mathbf{H}_k^{\mathrm{T}} \mathbf{R}^{-1} \left(\mathbf{d}_k - \mathbf{H}_k \delta \mathbf{x}_k^b \right) \tag{22}$$

The obvious advantage of this second option is to get rid of $\mathbf{B}^{-1}$ in the right-hand side $\mathbf{b}_k^{\mathbf{w}}$, whereas it is required in the right-hand side $\mathbf{b}_k^{\mathbf{v}}$ for the term $\delta \mathbf{v}_k^b$.

1.5 Minimizer starting state

A starting state $\mathbf{x}_k^s$ has to be defined for each outer iteration, and converted into control space:

- For the classic variable change, the starting increment in control space is:

$$\delta \mathbf{v}_k^s = \mathbf{U}^{\mathrm{T}} \mathbf{B}^{-1} \left(\mathbf{x}_k^s - \mathbf{x}_k^g \right) \tag{23}$$

- For the non-classic variable change, the starting increment in control space is:

$$\delta \mathbf{w}_k^s = \mathbf{U}^{\mathrm{T}} \mathbf{B}^{-1} \left(\mathbf{x}_k^s - \mathbf{x}^b \right) \tag{24}$$

The most obvious choice for the starting state $\mathbf{x}_k^s$ is the guess $\mathbf{x}_k^g$, since it is the best estimate of the final analysis we have at a given outer iteration. In this case:

- For the classic variable change:

$$\delta \mathbf{v}_k^s = 0 \qquad (25)$$

- For the non-classic variable change:

$$\begin{aligned}
\delta \mathbf{w}_k^s &= -\mathbf{U}^{\mathrm{T}} \mathbf{B}^{-1} \delta \mathbf{x}_k^b \\
&= -\delta \mathbf{v}_k^b \qquad (26)
\end{aligned}$$

It should be noted that $\mathbf{B}^{-1}$ is absent from the right-hand side $\mathbf{b}_k^{\mathbf{w}}$ but it is required in the starting increment $\delta \mathbf{w}_k^s$.

Another possible starting state $\mathbf{x}_k^s$ for the non-classic variable change is the background $\mathbf{x}^b$, leading to a starting increment $\delta \mathbf{w}_k^s = 0$. This option removes the need for $\mathbf{B}^{-1}$ completely, but is not explored further in this document.

2 Practical computations

2.1 Getting rid of $\mathbf{B}^{-1}$

In practice, the inverse of the background error covariance matrix $\mathbf{B}^{-1}$ is not available, even if it is needed in general to compute the background increment $\delta \mathbf{v}_k^b$, required for the right-hand side $\mathbf{b}_k^{\mathbf{v}}$ or for the starting increment $\delta \mathbf{w}_k^s$. However, the analysis is defined as:

$$\mathbf{x}_k^a = \mathbf{x}_k^g + \delta \mathbf{x}_k^a \qquad (27)$$

and it seems logical to define the guess as the analysis of the previous outer iteration:

- for $k = 1$: $\mathbf{x}_1^g = \mathbf{x}^b$,
- for $k > 1$: $\mathbf{x}_k^g = \mathbf{x}_{k-1}^a$.

Thus:

- for $k = 1$:

$$\delta \mathbf{x}_1^b = \mathbf{x}_1^g - \mathbf{x}^b = 0 \qquad (28)$$

- for $k > 1$:

$$\begin{aligned}
\delta \mathbf{x}_k^b &= \mathbf{x}^b - \mathbf{x}_k^g \\
&= \mathbf{x}^b - \mathbf{x}_{k-1}^a \\
&= \mathbf{x}^b - \left(\mathbf{x}_{k-1}^g + \delta \mathbf{x}_{k-1}^a \right) \\
&= \delta \mathbf{x}_{k-1}^b - \delta \mathbf{x}_{k-1}^a \qquad (29)
\end{aligned}$$

which can be combined recursively to yield:

$$\delta \mathbf{x}_k^b = -\sum_{i=1}^{k-1} \delta \mathbf{x}_i^a \tag{30}$$

Using equation (12) and applying $\mathbf{U}^{\mathrm{T}}\mathbf{B}^{-1}$ on both sides of equation (30), we get:

$$\delta \mathbf{v}_k^b = -\sum_{i=1}^{k-1} \delta \mathbf{v}_i^a \tag{31}$$

Using equation (20), we get also:

$$\delta \mathbf{v}_k^b = -\sum_{i=1}^{k-1} \left(\delta \mathbf{w}_i^a + \delta \mathbf{v}_i^b \right) \tag{32}$$

So for $k > 1$:

$$\begin{aligned} \delta \mathbf{v}_k^b &= -\left(\delta \mathbf{w}_{k-1}^a + \delta \mathbf{v}_{k-1}^b + \sum_{i=1}^{k-2} \left(\delta \mathbf{w}_i^a + \delta \mathbf{v}_i^b \right) \right) \\ &= -\left(\delta \mathbf{w}_{k-1}^a + \delta \mathbf{v}_{k-1}^b - \delta \mathbf{v}_{k-1}^b \right) \end{aligned} \tag{33}$$

leading to:

$$\delta \mathbf{v}_k^b = -\delta \mathbf{w}_{k-1}^a \tag{34}$$

Equation (31) and (34) can be used to compute $\mathbf{b}_k^{\mathbf{v}}$ and $\delta \mathbf{w}_k^s$ respectively, without requiring $\mathbf{B}^{-1}$.

2.2 Changing the resolution

For computational efficiency, it is common to start the outer iterations at a lower grid resolution, and to increase it at each outer iteration k. At resolution $\mathcal{R}_k$, the model space size is denoted n_k and the control space size m_k. It is assumed hereafter that:

- resolutions are stricly increasing: for $i < k$, $n_i < n_k$ and $m_i < m_k$,

- the full resolution is obtained at the last iteration K.

Obviously in this case, $\mathbf{B}$ also depends on iteration k. Hereafter, it is denoted $\mathbf{B}_k$, and its square-root is denoted $\mathbf{U}_k$. The previous trick to get rid of $\mathbf{B}^{-1}$ no longer works, since the control vector size changes at each outer iteration.

2.3 Interpolators

We define two linear interpolators from resolution $\mathcal{R}_i$ to resolution $\mathcal{R}_k$:

- $\mathbf{T}^{\mathbf{x}}_{i \to k} \in \mathbb{R}^{n_k \times n_i}$ in model space,

- $\mathbf{T}^{\mathbf{v}}_{i \to k} \in \mathbb{R}^{m_k \times m_i}$ in control space,

Interpolators can be applied on "states" (i.e. physical quantities) like the background or the guess, or on "increments" (i.e. differences of physical quantities), which are denoted with a δ in this document. For downscaling, it seems reasonable to apply interpolators on either states or increments. However for upscaling, it seems better to apply interpolators on increments only and to add these increments to high resolution states, rather than to upscale states directly. Indeed, high resolution states already have meaningful small scale features, whereas upscaled states do not.

2.4 Analysis at high resolution

As a consequence, the multi-resolution problem should be solved with the following requirements in mind:

- The background $\mathbf{x}^b$ is provided at full resolution, but it can be simplified at resolution $\mathcal{R}_k$:

$$\boxed{\mathbf{x}^b_k = \mathbf{T}^{\mathbf{x}}_{K \to k} \mathbf{x}^b} \tag{35}$$

- A full resolution guess denoted $\mathbf{x}^{g+}_k$ has to be computed at each outer iteration to run model trajectories used in the operators linearization. This full resolution guess can be simplified at resolution $\mathcal{R}_k$ to give the actual guess of outer iteration k:

$$\boxed{\mathbf{x}^g_k = \mathbf{T}^{\mathbf{x}}_{K \to k} \mathbf{x}^{g+}_k} \tag{36}$$

- Only increments should be interpolated to higher resolution, and then possibly added to states at high resolution.

At each outer iteration, the analysis increment $\delta\mathbf{x}^a_k$ is given at resolution $\mathcal{R}_k$ for the classic variable change by:

$$\delta\mathbf{x}^a_k = \mathbf{U}_k \delta\mathbf{v}^a_k \tag{37}$$

or for the non-classic variable change by:

$$\delta\mathbf{x}^a_k = \mathbf{U}_k \delta\mathbf{w}^a_k + \delta\mathbf{x}^b_k \tag{38}$$

and it can be interpolated at full resolution:

$$\delta\mathbf{x}^{a+}_k = \mathbf{T}^{\mathbf{x}}_{k \to K} \delta\mathbf{x}^a_k \tag{39}$$

to give the full resolution analysis:

$$\mathbf{x}^{a+}_k = \mathbf{x}^{g+}_k + \delta\mathbf{x}^{a+}_k \tag{40}$$

3 Guess consistency

3.1 Problem clarification

In the linear system solved at each outer iteration, the full resolution guess $\mathbf{x}_k^{g+}$ appears for two distinct purposes.

1. It is used explicitly as the linearization state for the nonlinear observation operator in equation (4), and also to compute the innovation $\mathbf{d}_k$ in equation (7).

2. It is needed implicitly in the background increment $\delta\mathbf{v}_k^b$, required for the right-hand side $\mathbf{b}_k^{\mathbf{v}}$ or for the starting increment $\delta\mathbf{w}_k^s$:

$$
\begin{aligned}
\delta\mathbf{v}_k^b &= \mathbf{U}_k^{\mathrm{T}}\mathbf{B}_k^{-1}\delta\mathbf{x}_k^b \\
&= \mathbf{U}_k^{\mathrm{T}}\mathbf{B}_k^{-1}\left(\mathbf{x}_k^b - \mathbf{x}_k^g\right) \\
&= \mathbf{U}_k^{\mathrm{T}}\mathbf{B}_k^{-1}\mathbf{T}_{K\to k}^{\mathbf{x}}\left(\mathbf{x}^b - \mathbf{x}_k^{g+}\right)
\end{aligned}
\tag{41}
$$

For the first outer iteration, the full resolution guess is always taken as the background state (also provided at full resolution):

$$
\mathbf{x}_1^{g+} = \mathbf{x}^b
\tag{42}
$$

and the background increment is zero, which is consistent. The goal of this section is to define the possible strategies to maintain this consistency between the two purposes of the high resolution guess in subsequent outer iterations.

3.2 Theoretical method

In the theoretical method, $\mathbf{B}_k^{-1}$ is supposed to be available and can be used to compute $\delta\mathbf{v}_k^b$ explicitly from $\mathbf{x}_k^{g+}$ using equation (41), so the guess consistency is always maintained.

For outer iterations $k > 1$, the full resolution guess is simply taken as the previous analysis at full resolution:

$$
\mathbf{x}_k^{g+} = \mathbf{x}_{k-1}^{a+}
\tag{43}
$$

3.3 Practical methods

3.3.1 Approximations

In practical methods, $\mathbf{B}_k^{-1}$ is not available and the background increment $\delta\mathbf{v}_k^b$ is computed separately, using an updated version of equation (31) with appropriate interpolations:

- For the classic variable change:

$$
\boxed{\delta\mathbf{v}_k^b = -\sum_{i=1}^{k-1}\mathbf{T}_{i\to k}^{\mathbf{v}}\delta\mathbf{v}_i^a}
\tag{44}
$$

- For the non-classic variable change:

$$\boxed{\delta \mathbf{v}_k^b = -\mathbf{T}_{k-1 \to k}^{\mathbf{v}} \delta \mathbf{w}_{k-1}^a} \tag{45}$$

3.3.2 Consistent method

In the consistent method, the background increment $\delta \mathbf{v}_k^b$ is obtained from equation (44) or equation (45). From this, the background increment $\delta \mathbf{x}_k^b$ is given by:

$$\delta \mathbf{x}_k^b = \mathbf{U}_k \delta \mathbf{v}_k^b \tag{46}$$

It is then interpolated at full resolution:

$$\delta \mathbf{x}_k^{b+} = \mathbf{T}_{k \to K}^{\mathbf{x}} \delta \mathbf{x}_k^b \tag{47}$$

and the guess at full resolution is deduced as:

$$\mathbf{x}_k^{g+} = \mathbf{x}^b - \delta \mathbf{x}_k^{b+} \tag{48}$$

In this method, the guess at full resolution **is not equal** to the analysis at full resolution of the previous outer iteration, but the guess consistency is always maintained by construction.

3.3.3 Standard method

In the standard method, the full resolution guess is explicitly computed from the analysis of the previous outer iteration. However, contrary to the theoretical method, a correction has to be applied in order to maintain the guess consistency. This correction is neglected in operational data assimilation systems, and the goal of this section is to compute it explicitly.

For outer iterations $k > 1$, the full resolution guess is now given by:

$$\mathbf{x}_k^{g+} = \mathbf{x}_{k-1}^{a+} + \delta \mathbf{x}_k^{c+} \tag{49}$$

where $\delta \mathbf{x}_k^{c+}$ is the correction. Merging equations (40), (42) and (49), the guess at full resolution can be expressed from the previous iteration:

$$\begin{aligned}
\mathbf{x}_k^{g+} &= \mathbf{x}_{k-1}^{a+} + \delta \mathbf{x}_k^{c+} \\
&= \mathbf{x}_{k-1}^{g+} + \delta \mathbf{x}_{k-1}^{a+} + \delta \mathbf{x}_k^{c+}
\end{aligned} \tag{50}$$

and recursively backward to the first outer iteration:

$$\mathbf{x}_k^{g+} = \mathbf{x}^b + \sum_{i=1}^{k-1} \delta \mathbf{x}_i^{a+} + \sum_{i=2}^{k} \delta \mathbf{x}_i^{c+} \tag{51}$$

Thus, the background increment is given by:

$$\begin{aligned}
\delta\mathbf{x}_k^b &= \mathbf{x}_k^b - \mathbf{x}_k^g \\
&= \mathbf{T}_{K\to k}^{\mathbf{x}} \left(\mathbf{x}^b - \mathbf{x}_k^{g+} \right) \\
&= -\mathbf{T}_{K\to k}^{\mathbf{x}} \left(\sum_{i=1}^{k-1} \delta\mathbf{x}_i^{a+} + \sum_{i=2}^{k} \delta\mathbf{x}_i^{c+} \right) \\
&= -\left[\mathbf{T}_{K\to k}^{\mathbf{x}} \left(\sum_{i=1}^{k-1} \delta\mathbf{x}_i^{a+} + \sum_{i=2}^{k-1} \delta\mathbf{x}_i^{c+} \right) + \mathbf{T}_{K\to k}^{\mathbf{x}} \delta\mathbf{x}_k^{c+} \right]
\end{aligned} \tag{52}$$

which is equivalent to:

$$\mathbf{T}_{K\to k}^{\mathbf{x}} \delta\mathbf{x}_k^{c+} = -\left[\mathbf{T}_{K\to k}^{\mathbf{x}} \left(\sum_{i=1}^{k-1} \delta\mathbf{x}_i^{a+} + \sum_{i=2}^{k-1} \delta\mathbf{x}_i^{c+} \right) + \delta\mathbf{x}_k^b \right] \tag{53}$$

In the general case, it is not possible to compute the correction $\delta\mathbf{x}_k^{c+}$ explicitly.

3.3.4 Special case of transitive interpolators

A special class of interpolators called "transitive interpolators" has three extra properties:

- Upscaling transitivity: for $n_i < n_j$ and $n_i < n_k$:

$$\mathbf{T}_{j\to k}^{\mathbf{x}} \mathbf{T}_{i\to j}^{\mathbf{x}} = \mathbf{T}_{i\to k}^{\mathbf{x}} \tag{54}$$

- Downscaling transitivity: for $n_i < n_j < n_k$:

$$\mathbf{T}_{j\to i}^{\mathbf{x}} \mathbf{T}_{k\to j}^{\mathbf{x}} = \mathbf{T}_{k\to i}^{\mathbf{x}} \tag{55}$$

- Known right-inverse: for $n_i < n_k$

$$\mathbf{T}_{k\to i}^{\mathbf{x}} \mathbf{T}_{i\to k}^{\mathbf{x}} = \mathbf{I}_{n_i} \tag{56}$$

and similarly for $\mathbf{T}_{i\to k}^{\mathbf{v}}$ in control space, replacing n with m.

For transitive interpolators, the right inverse of $\mathbf{T}_{k\to K}^{\mathbf{x}}$ is $\mathbf{T}_{K\to k}^{\mathbf{x}}$ so we can get an explicit expression for the correction:

$$\boxed{\delta\mathbf{x}_k^{c+} = -\mathbf{T}_{k\to K}^{\mathbf{x}} \left[\mathbf{T}_{K\to k}^{\mathbf{x}} \left(\sum_{i=1}^{k-1} \delta\mathbf{x}_i^{a+} + \sum_{i=2}^{k-1} \delta\mathbf{x}_i^{c+} \right) + \delta\mathbf{x}_k^b \right]} \tag{57}$$

3.3.5 Special case of a projective B family

Associated with the different resolutions $\left\{ \mathcal{R}_1, \ldots, \mathcal{R}_K \right\}$, we characterize a $\mathbf{B}$ family $\left\{ \mathbf{B}_1, \ldots, \mathbf{B}_K \right\}$ as "projective" if a low resolution element can be defined as the projection of a high resolution one, using transitive interpolators. For $n_i < n_k$, this means that:

$$\mathbf{B}_k \mathbf{T}^{\mathbf{x}}_{i \to k} = \mathbf{T}^{\mathbf{x}}_{i \to k} \mathbf{B}_i \tag{58}$$

and for the square-root of $\mathbf{B}$:

$$\mathbf{U}_k \mathbf{T}^{\mathbf{v}}_{i \to k} = \mathbf{T}^{\mathbf{x}}_{i \to k} \mathbf{U}_i \tag{59}$$

For transitive interpolators, it should be noted that:

$$
\begin{aligned}
\mathbf{T}^{\mathbf{x}}_{K \to k} \sum_{i=1}^{k-1} \delta \mathbf{x}_i^{a+} &= \sum_{i=1}^{k-1} \mathbf{T}^{\mathbf{x}}_{K \to k} \mathbf{T}^{\mathbf{x}}_{i \to K} \delta \mathbf{x}_i^a \\
&= \sum_{i=1}^{k-1} \mathbf{T}^{\mathbf{x}}_{i \to k} \delta \mathbf{x}_i^a
\end{aligned}
\tag{60}
$$

and for a projective $\mathbf{B}$ family, equation (44) gives:

$$
\begin{aligned}
\delta \mathbf{x}_k^b &= \mathbf{U}_k \delta \mathbf{v}_k^b \\
&= -\mathbf{U}_k \sum_{i=1}^{k-1} \mathbf{T}^{\mathbf{v}}_{i \to k} \delta \mathbf{v}_i^a \\
&= -\sum_{i=1}^{k-1} \mathbf{U}_k \mathbf{T}^{\mathbf{v}}_{i \to k} \delta \mathbf{v}_i^a \\
&= -\sum_{i=1}^{k-1} \mathbf{T}^{\mathbf{x}}_{i \to k} \mathbf{U}_i \delta \mathbf{v}_i^a \\
&= -\sum_{i=1}^{k-1} \mathbf{T}^{\mathbf{x}}_{i \to k} \delta \mathbf{x}_i^a
\end{aligned}
\tag{61}
$$

Thus, these two terms vanish and equation (57) becomes:

$$\delta \mathbf{x}_k^{c+} = -\mathbf{T}^{\mathbf{x}}_{k \to K} \mathbf{T}^{\mathbf{x}}_{K \to k} \sum_{i=2}^{k-1} \delta \mathbf{x}_i^{c+} \tag{62}$$

which is equivalent to:

$$\boxed{\delta \mathbf{x}_k^{c+} = 0} \tag{63}$$

For a projective $\mathbf{B}$ family, no correction $\delta \mathbf{x}_k^{c+}$ is required to maintain the guess concistency.

3.4 Summary of the different cases

In the following tables, a green thick ✓ indicates that methods give the same results, and a red cross ✗ that they do not. For the first outer iteration, all methods give the same results, so we are only interested in the subsequent outer iterations ($k > 1$).

3.4.1 Projective B family

- Correction $\delta\mathbf{x}_k^{c+}$ for the standard method is zero.

- All methods give the same result.

		Theoretical	Consistent	Standard	
				w/ corr.	w/o corr.
Theoretical		✓	✓	✓	✓
Consistent		✓	✓	✓	✓
Standard	w/ corr.	✓	✓	✓	✓
	w/o corr.	✓	✓	✓	✓

3.4.2 Non-projective B family but transitive interpolators

- Correction $\delta\mathbf{x}_k^{c+}$ for the standard method is non-zero but can be explicitly computed.

- The consistent and standard methods give the same result.

- The theoretical method gives different results, since it computes the background increment $\delta\mathbf{v}_k^b$ differently.

		Theoretical	Consistent	Standard	
				w/ corr.	w/o corr.
Theoretical		✓	✗	✗	✗
Consistent		✗	✓	✓	✗
Standard	w/ corr.	✗	✓	✓	✗
	w/o corr.	✗	✗	✗	✓

3.4.3 Non-projective B family and non-transitive interpolators

- Correction $\delta\mathbf{x}_k^{c+}$ for the standard method cannot be computed, so the standard method cannot ensure the guess consistency.

- Theoretical and consistent method both ensure the guess consistency, but give different results.
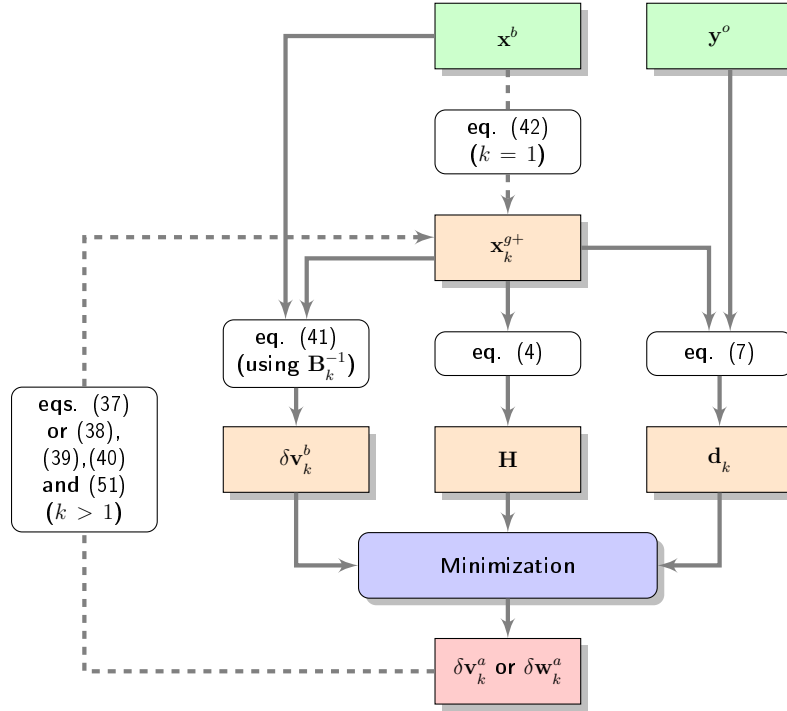
		Theoretical	Consistent	Standard	
				w/ corr.	w/o corr.
Theoretical		✓	✗	✗	✗
Consistent		✗	✓	✗	✗
Standard	w/ corr.	✗	✗	✗	✗
	w/o corr.	✗	✗	✗	✓

The black cross ✗ indicates that this option is not available.
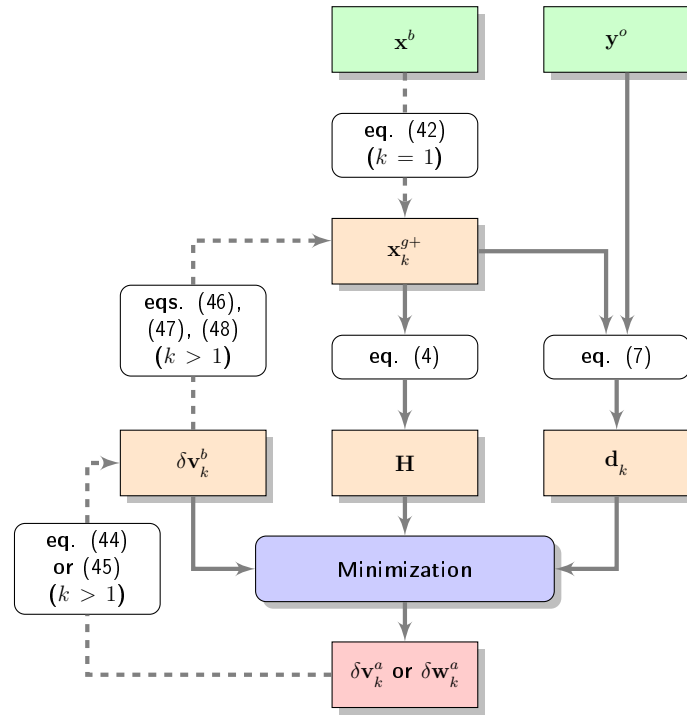
3.5 Workflows

3.5.1 Theoretical method

For the theoretical method, the background and the full resolution guess are used at each outer iteration to compute $\delta \mathbf{v}_k^b$, using $\mathbf{B}_k^{-1}$:


3.5.2 Consistent method

Because of its closed cycle form (no fork), the consistent method ensures the guess consistency, but it does not require $\mathbf{B}_k^{-1}$.

3.5.3 Standard method

The potential guess inconsistency of the standard method comes from the fork in the use of the minimization output to compute the full resolution guess and background increment in control space independently. If the appropriate correction is used to defined the full resolution guess (yellow box), then the guess consistency can be maintained despite the fork.
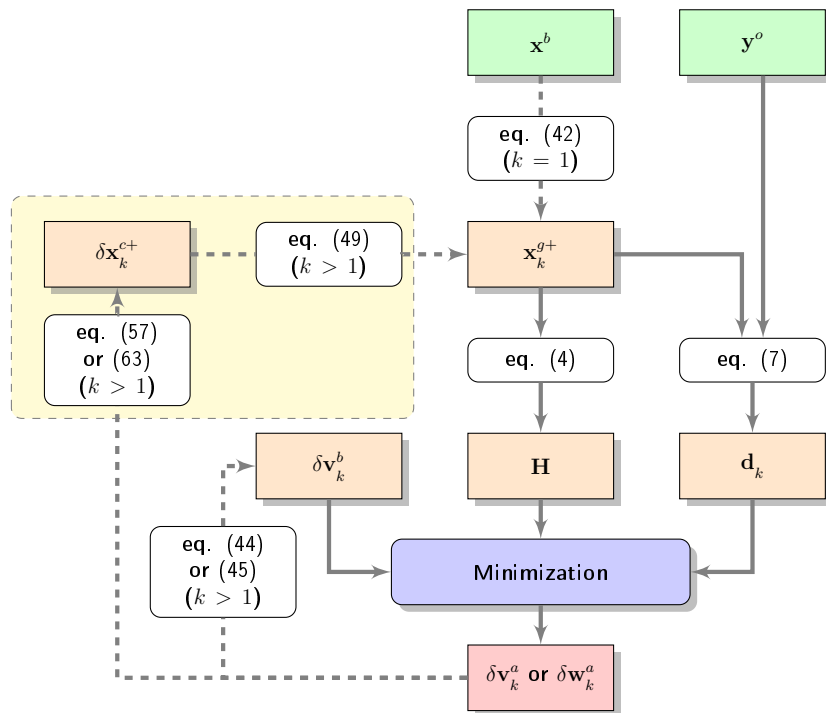
4 Examples

4.1 Transitive interpolators

A generic family of transitive interpolators is based on a zero-padding operator surrounded by an orthogonal transform. For instance in model space:

$$\mathbf{T}^{\mathbf{x}}_{i \to k} = \mathbf{S}^{\mathrm{T}}_k \mathbf{\Delta}_{i \to k} \mathbf{S}_i \tag{64}$$

where

- $\mathbf{\Delta}_{i \to k} \in \mathbb{R}^{n_k \times n_i}$ is a zero-padding operator:

$$\Delta_{i \to k, \alpha\beta} = \left\{ \begin{array}{ll} 1 & \text{if} \quad \alpha = \beta \\ 0 & \text{if} \quad \alpha \neq \beta \end{array} \right. \tag{65}$$

- $\mathbf{S}_k \in \mathbb{R}^{n_k \times n_k}$ is any orthogonal transform:

$$\mathbf{S}_k \mathbf{S}^{\mathrm{T}}_k = \mathbf{S}^{\mathrm{T}}_k \mathbf{S}_k = \mathbf{I}_{n_k} \tag{66}$$

It is easy to check the three required properties:

- Upscaling transitivity: for $n_i \leq n_j$ and $n_i \leq n_k$:

$$\begin{aligned} \mathbf{T}^{\mathbf{x}}_{j \to k} \mathbf{T}^{\mathbf{x}}_{i \to j} &= \mathbf{S}^{\mathrm{T}}_k \mathbf{\Delta}_{j \to k} \mathbf{S}_j \mathbf{S}^{\mathrm{T}}_j \mathbf{\Delta}_{i \to j} \mathbf{S}_i \\ &= \mathbf{S}^{\mathrm{T}}_k \mathbf{\Delta}_{j \to k} \mathbf{\Delta}_{i \to j} \mathbf{S}_i \\ &= \mathbf{S}^{\mathrm{T}}_k \mathbf{\Delta}_{i \to k} \mathbf{S}_i \\ &= \mathbf{T}^{\mathbf{x}}_{i \to k} \end{aligned} \tag{67}$$

- Downscaling transitivity: for $n_i \leq n_j \leq n_k$:

$$\begin{aligned} \mathbf{T}^{\mathbf{x}}_{j \to i} \mathbf{T}^{\mathbf{x}}_{k \to j} &= \mathbf{S}^{\mathrm{T}}_i \mathbf{\Delta}_{j \to i} \mathbf{S}_j \mathbf{S}^{\mathrm{T}}_j \mathbf{\Delta}_{k \to j} \mathbf{S}_k \\ &= \mathbf{S}^{\mathrm{T}}_i \mathbf{\Delta}_{j \to i} \mathbf{\Delta}_{k \to j} \mathbf{S}_k \\ &= \mathbf{S}^{\mathrm{T}}_i \mathbf{\Delta}_{k \to i} \mathbf{S}_k \\ &= \mathbf{T}^{\mathbf{x}}_{k \to i} \end{aligned} \tag{68}$$

- Known right-inverse: for $n_i \leq n_k$

$$\begin{aligned} \mathbf{T}^{\mathbf{x}}_{k \to i} \mathbf{T}^{\mathbf{x}}_{i \to k} &= \mathbf{S}^{\mathrm{T}}_i \mathbf{\Delta}_{k \to i} \mathbf{S}_k \mathbf{S}^{\mathrm{T}}_k \mathbf{\Delta}_{i \to k} \mathbf{S}_i \\ &= \mathbf{S}^{\mathrm{T}}_i \mathbf{\Delta}_{k \to i} \mathbf{\Delta}_{i \to k} \mathbf{S}_i \\ &= \mathbf{S}^{\mathrm{T}}_i \mathbf{S}_i \\ &= \mathbf{I}_{n_i} \end{aligned} \tag{69}$$

and similarly in control space.

4.2 Illustration: upscaling transitivity

The spectral decomposition is an orthogonal operator, which can be associated with a zero padding operator to build transitive interpolators. Figure 1 shows an example of upscaling with a spectral interpolator.
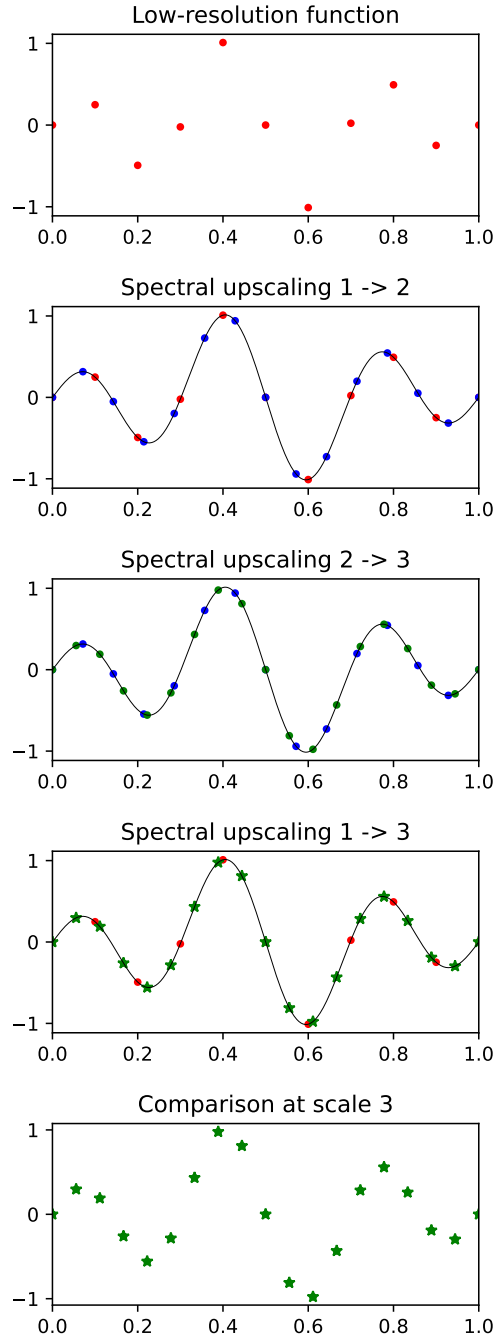

Figure 1: Transitivity upscaling for a spectral interpolator. Initial low resolution (red dots), intermediate resolution (blue dots) and high resolution interpolated from intermediate or low resolution (green dots and stars, respectively). The black lines indicate the continuous interpolation function. Successive grids are not colocated.

Usual grid-points interpolators like the (bi-)linear interpolator are not transitive in general, as illustrated in figure 2.
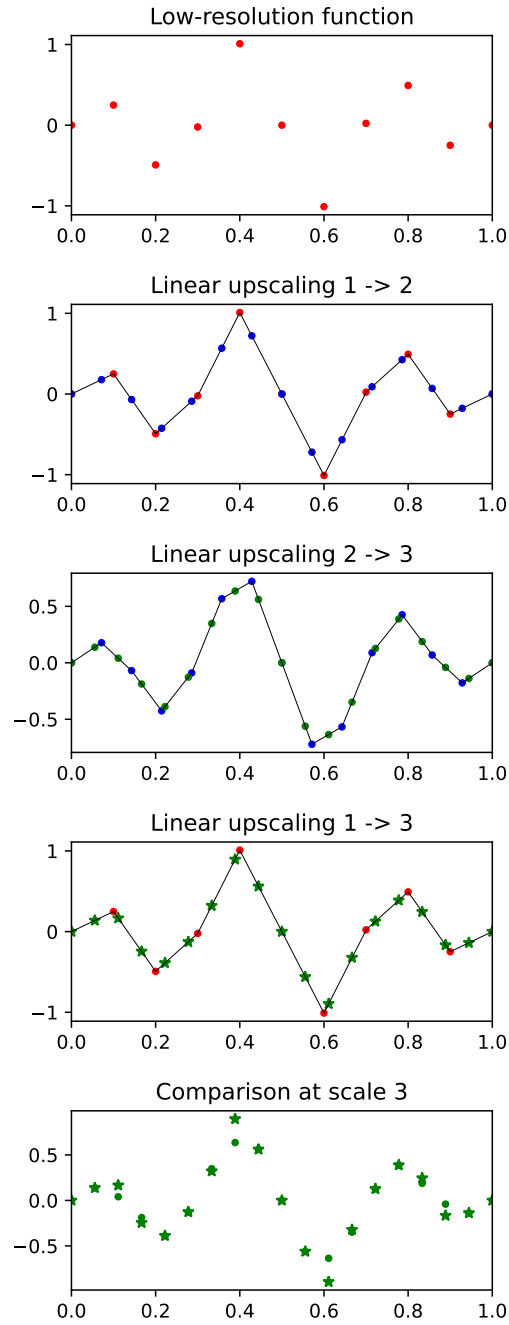

Figure 2: Transitivity upscaling for a linear interpolator. Initial low resolution (red dots), intermediate resolution (blue dots) and high resolution interpolated from intermediate or low resolution (green dots and stars, respectively). The black lines indicate the continuous interpolation functions. Successive grids are not colocated.

An important exception is the (bi-)linear interpolator when coarse grid points are colocated with fine grid points, as shown in figure 3.
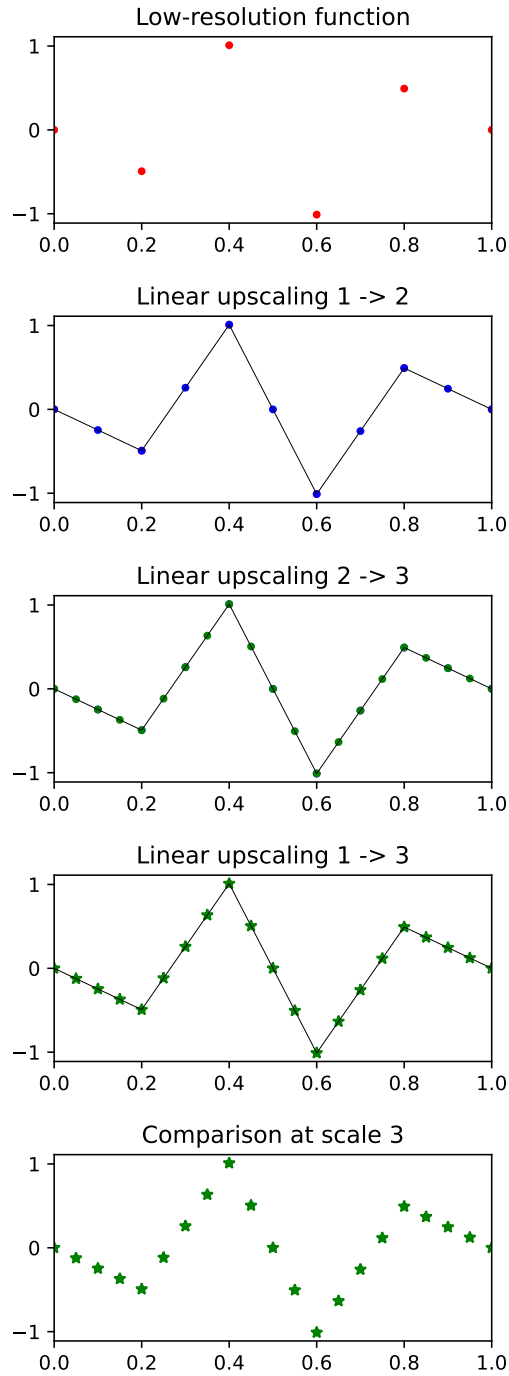

Figure 3: Transitivity upscaling for a linear interpolator. Initial low resolution (red dots), intermediate resolution (blue dots) and high resolution interpolated from intermediate or low resolution (green dots and stars, respectively). The black lines indicate the continuous interpolation functions. Successive grids are colocated.

4.3 Projective B family

A vector $\boldsymbol{\gamma} \in \mathbb{R}^{n_K}$ of positive coefficients can define the diagonal matrices $\boldsymbol{\Gamma}_k \in \mathbb{R}^{n_k \times n_k}$:

$$\Gamma_{k,\alpha\beta} = \left\{ \begin{array}{ll} \gamma_i & \text{if} \quad \alpha = \beta \\ 0 & \text{if} \quad \alpha \neq \beta \end{array} \right. \tag{70}$$

At resolution $\mathcal{R}_k$, the background error covariance matrix $\mathbf{B}_k$ is defined as:

$$\mathbf{B}_k = \mathbf{S}_k \boldsymbol{\Gamma}_k \mathbf{S}_k^{-1} \tag{71}$$

A square-root $\mathbf{U}_k$ is simply:

$$\mathbf{U}_k = \mathbf{S}_k \boldsymbol{\Gamma}_k^{1/2} \mathbf{G}_k \tag{72}$$

where $\mathbf{G}_k \in \mathbb{R}^{n_k \times m_k}$ can be any orthogonal matrix.

Thus:

$$\begin{aligned} \mathbf{U}_k^T \mathbf{T}_{i \to k}^{\mathbf{x}} &= \mathbf{G}_k^T \boldsymbol{\Gamma}_k^{1/2} \mathbf{S}_k^T \mathbf{S}_k \boldsymbol{\Delta}_{i \to k} \mathbf{S}_i^T \\ &= \mathbf{G}_k^T \boldsymbol{\Gamma}_k^{1/2} \boldsymbol{\Delta}_{i \to k} \mathbf{S}_i^T \end{aligned} \tag{73}$$

and

$$\begin{aligned} \mathbf{T}_{i \to k}^{\mathbf{v}} \mathbf{U}_i^T &= \mathbf{G}_k^T \boldsymbol{\Delta}_{i \to k} \mathbf{G}_i \mathbf{G}_i^T \boldsymbol{\Gamma}_i^{1/2} \mathbf{S}_i^T \\ &= \mathbf{G}_k^T \boldsymbol{\Delta}_{i \to k} \boldsymbol{\Gamma}_i^{1/2} \mathbf{S}_i^T \end{aligned} \tag{74}$$

Since $\boldsymbol{\Gamma}_k^{1/2} \boldsymbol{\Delta}_{i \to k} = \boldsymbol{\Delta}_{i \to k} \boldsymbol{\Gamma}_i^{1/2}$, the projective $\mathbf{B}$ family condition (59) is verified.