3

Analysis scheme

This chapter discusses the problem of how to combine a model prediction of a state variable at a given time with a set of measurements available at this particular time. It is assumed that error statistics of the model prediction as well as the measurements are known and characterized by the respective error covariances. Based on this information the so-called analysis scheme used in linear data assimilation methods is presented in some detail. First the theory is derived for the scalar case and then it is extended to the case with a spatial dimension. An extensive analysis of the properties of the analysis scheme is given and this introduces notation and concepts which are also valid for the time dependent problems treated in the following chapters.

3.1 Scalar case

We start by deriving the optimal linear and unbiased estimator for a scalar state variable combined with a single measurement.

3.1.1 State-space formulation

Given two different estimates of the true state ψ^{t} (e.g. a temperature at a particular location and time):

$$\psi^{\mathrm{f}} = \psi^{\mathrm{t}} + p^{\mathrm{f}}, \tag{3.1}$$

$$d = \psi^{\mathrm{t}} + \epsilon, \tag{3.2}$$

where ψ^{f} may be a model forecast or a first-guess estimate and d is a measurement of ψ^{t}. The term p^{f} denotes the unknown error in the forecast and ϵ is the unknown measurement error. The problem is now, to find an improved analyzed estimate ψ^{a} of ψ^{t}. Thus, additional information about the error terms must be supplied and we make the following assumptions:

$$\begin{aligned} \overline{p^{\mathrm{f}}} &= 0, & \overline{(p^{\mathrm{f}})^2} &= C^{\mathrm{f}}_{\psi\psi}, \\ \overline{\epsilon} &= 0, & \overline{(\epsilon)^2} &= C_{\epsilon\epsilon}, \\ \overline{\epsilon p^{\mathrm{f}}} &= 0. \end{aligned} \tag{3.3}$$

Here the overbar denotes ensemble averaging or expected value.

We now seek a linear estimator

$$\psi^{\mathrm{a}} = \psi^{\mathrm{t}} + p^{\mathrm{a}} = \alpha_1 \psi^{\mathrm{f}} + \alpha_2 d, \tag{3.4}$$

where we define

$$\overline{p^{\mathrm{a}}} = 0, \qquad \overline{(p^{\mathrm{a}})^2} = C^{\mathrm{a}}_{\psi\psi}. \tag{3.5}$$

The definition (3.5) means that we assume that the error p^{a}, in the analyzed estimate is unbiased. Thus, the analyzed estimate itself becomes an unbiased estimate of the true state ψ^{t}, i.e. $\overline{\psi^{\mathrm{a}}} = \psi^{\mathrm{t}}$.

Inserting the estimates (3.1) and (3.2) in (3.4) we get

$$\psi^{\mathrm{t}} + p^{\mathrm{a}} = \alpha_1(\psi^{\mathrm{t}} + p^{\mathrm{f}}) + \alpha_2(\psi^{\mathrm{t}} + \epsilon). \tag{3.6}$$

The expectation of this equation is

$$\psi^{\mathrm{t}} = \alpha_1 \psi^{\mathrm{t}} + \alpha_2 \psi^{\mathrm{t}} = (\alpha_1 + \alpha_2)\psi^{\mathrm{t}}. \tag{3.7}$$

Thus, we must have

$$\alpha_1 + \alpha_2 = 1, \qquad \text{or} \qquad \alpha_1 = 1 - \alpha_2, \tag{3.8}$$

and a linear unbiased estimator for ψ^{t} is given as

$$\begin{aligned} \psi^{\mathrm{a}} &= (1 - \alpha_2)\psi^{\mathrm{f}} + \alpha_2 d \\ &= \psi^{\mathrm{f}} + \alpha_2(d - \psi^{\mathrm{f}}). \end{aligned} \tag{3.9}$$

Using (3.1), (3.2) and (3.4) in this equation gives an expression for the error in the analysis

$$p^{\mathrm{a}} = p^{\mathrm{f}} + \alpha_2(\epsilon - p^{\mathrm{f}}). \tag{3.10}$$

The error variance is then using (3.3)

$$\begin{aligned} \overline{(p^{\mathrm{a}})^2} = C^{\mathrm{a}}_{\psi\psi} &= \overline{(p^{\mathrm{f}} + \alpha_2(\epsilon - p^{\mathrm{f}}))^2} \\ &= \overline{(p^{\mathrm{f}})^2} + 2\alpha_2 \overline{p^{\mathrm{f}}(\epsilon - p^{\mathrm{f}})} + \alpha_2^2 \overline{\epsilon^2 - 2\epsilon p^{\mathrm{f}} + (p^{\mathrm{f}})^2} \\ &= C^{\mathrm{f}}_{\psi\psi} - 2\alpha_2 C^{\mathrm{f}}_{\psi\psi} + \alpha_2^2(C_{\epsilon\epsilon} + C^{\mathrm{f}}_{\psi\psi}), \end{aligned} \tag{3.11}$$

and the minimum variance is defined by

$$\frac{dC^{\mathrm{a}}_{\psi\psi}}{d\alpha_2} = -2C^{\mathrm{f}}_{\psi\psi} + 2\alpha_2(C_{\epsilon\epsilon} + C^{\mathrm{f}}_{\psi\psi}) = 0. \tag{3.12}$$

Solving for α_2 gives

$$\alpha_2 = \frac{C^{\mathrm{f}}_{\psi\psi}}{C_{\epsilon\epsilon} + C^{\mathrm{f}}_{\psi\psi}}, \tag{3.13}$$

and the analyzed estimate becomes

$$\psi^{\mathrm{a}} = \psi^{\mathrm{f}} + \frac{C^{\mathrm{f}}_{\psi\psi}}{C_{\epsilon\epsilon} + C^{\mathrm{f}}_{\psi\psi}} \left(d - \psi^{\mathrm{f}}\right). \tag{3.14}$$

Further, the error variance of the analyzed estimate is now from (3.11) and (3.13)

$$\begin{aligned} C^{\mathrm{a}}_{\psi\psi} &= C^{\mathrm{f}}_{\psi\psi} - 2\frac{C^{\mathrm{f}}_{\psi\psi}}{C_{\epsilon\epsilon} + C^{\mathrm{f}}_{\psi\psi}} C^{\mathrm{f}}_{\psi\psi} + \left(\frac{C^{\mathrm{f}}_{\psi\psi}}{C_{\epsilon\epsilon} + C^{\mathrm{f}}_{\psi\psi}}\right)^2 \left(C_{\epsilon\epsilon} + C^{\mathrm{f}}_{\psi\psi}\right) \\ &= C^{\mathrm{f}}_{\psi\psi} - \frac{(C^{\mathrm{f}}_{\psi\psi})^2}{C_{\epsilon\epsilon} + C^{\mathrm{f}}_{\psi\psi}} = C^{\mathrm{f}}_{\psi\psi}\left(1 - \frac{C^{\mathrm{f}}_{\psi\psi}}{C_{\epsilon\epsilon} + C^{\mathrm{f}}_{\psi\psi}}\right). \end{aligned} \tag{3.15}$$

3.1.2 Bayesian formulation

Given a probability density function $f(\psi)$ for the first-guess estimate ψ^{f}, and a likelihood function $f(d|\psi)$ for the measurement d; then, from Chap. 2 we have Bayes' theorem

$$f(\psi|d) \propto f(\psi) f(d|\psi). \tag{3.16}$$

Thus, the posterior density for ψ given the measurement d, is proportional to the product of the prior density for ψ times the likelihood function for the measurement d.

Again consider the two estimates (3.1) and (3.2) of the true state ψ^{t}. In the case with Gaussian statistics we can define the prior and likelihood as

$$f(\psi) \propto \exp\Big(-\frac{1}{2}\left(\psi - \psi^{\mathrm{f}}\right)\left(C^{\mathrm{f}}_{\psi\psi}\right)^{-1}\left(\psi - \psi^{\mathrm{f}}\right)\Big) \tag{3.17}$$

and

$$f(d|\psi) \propto \exp\Big(-\frac{1}{2}(\psi - d) C^{-1}_{\epsilon\epsilon}(\psi - d)\Big). \tag{3.18}$$

Thus, the posterior density can be written as

$$f(\psi|d) \propto \exp\Big(-\frac{1}{2}\mathcal{J}[\psi]\Big), \tag{3.19}$$

where

$$\mathcal{J}[\psi] = \left(\psi - \psi^{\mathrm{f}}\right)\left(C^{\mathrm{f}}_{\psi\psi}\right)^{-1}\left(\psi - \psi^{\mathrm{f}}\right) + (\psi - d) C^{-1}_{\epsilon\epsilon}(\psi - d). \tag{3.20}$$

The least squares solution ψ^{a}, that gives a minimum for $\mathcal{J}$, also gives a maximum of $f(\psi|d)$, i.e. it is the maximum likelihood estimate. This will always be true as long as all the error terms are normally distributed.

The minimum value of $\mathcal{J}$ is found from

$$\frac{d\mathcal{J}}{d\psi} = 2\left(\psi - \psi^{\mathrm{f}}\right)\left(C^{\mathrm{f}}_{\psi\psi}\right)^{-1} + 2\left(\psi - d\right)C^{-1}_{\epsilon\epsilon} = 0. \tag{3.21}$$

Solving for ψ gives again the result ψ^{a} in (3.14), thus, the minimum variance estimate is also the maximum likelihood estimate in the case with Gaussian priors.

3.2 Extension to spatial dimensions

Now we extend the discussion to involve a variable $\psi^{\mathrm{f}}(\boldsymbol{x})$, with a spatial dimension which may be one or larger, e.g. $\boldsymbol{x} = (x, y, z)$ for a three dimensional space. In the following discussion we adopt the notation used by *Bennett* (1992) who gave a similar derivation for the time dependent problem.

3.2.1 Basic formulation

Assume now a multidimensional variable (e.g. a temperature field), and a vector of measurements $\boldsymbol{d} \in \Re^M$, which is related to the true state through the measurement functional $\boldsymbol{\mathcal{M}} \in \Re^M$, with M being the number of measurements:

$$\psi^{\mathrm{f}}(\boldsymbol{x}) = \psi^{\mathrm{t}}(\boldsymbol{x}) + p^{\mathrm{f}}(\boldsymbol{x}), \tag{3.22}$$

$$\boldsymbol{d} = \boldsymbol{\mathcal{M}}[\psi^{\mathrm{t}}(\boldsymbol{x})] + \boldsymbol{\epsilon}. \tag{3.23}$$

The term $p^{\mathrm{f}}(\boldsymbol{x})$ is the error in the first-guess field $\psi^{\mathrm{f}}(\boldsymbol{x})$, relative to the truth $\psi^{\mathrm{t}}(\boldsymbol{x})$. Further, we have defined the vector of measurement errors $\boldsymbol{\epsilon} \in \Re^M$. The measurement errors may be a composite of errors introduced when measuring the variable and additional representation errors introduced when constructing the measurement functional. This will be discussed in more detail in the following chapters.

As an example of a measurement functional, a direct measurement would be represented by a functional of the form

$$\mathcal{M}_i[\psi(\boldsymbol{x})] = \int_{\mathcal{D}} \psi(\boldsymbol{x})\delta(\boldsymbol{x} - \boldsymbol{x}_i)d\boldsymbol{x} = \psi(\boldsymbol{x}_i), \tag{3.24}$$

where $\boldsymbol{x}_i$ is the measurement location, $\delta(\boldsymbol{x}-\boldsymbol{x}_i)$ is the Dirac delta function, and the subscript i denotes the component i of the measurement functional. Note that in some of the following equations we will use a subscript on the vector form of the measurement functional, e.g. $\boldsymbol{\mathcal{M}}_{(3)}[\delta\psi(\boldsymbol{x}_3)]$ which just denote that the integration is performed on the dummy variable $\boldsymbol{x}_3$ rather than $\boldsymbol{x}$ as is used in (3.24).

The actual values of the errors $p^{\mathrm{f}}(\boldsymbol{x})$ and $\boldsymbol{\epsilon}$ are not known. Thus, to make progress, a statistical hypothesis must be used, and we make the following assumptions:

$$\begin{aligned} \overline{p^{\mathrm{f}}(\boldsymbol{x})} &= 0, & \overline{p^{\mathrm{f}}(\boldsymbol{x}_1)p^{\mathrm{f}}(\boldsymbol{x}_2)} &= C^{\mathrm{f}}_{\psi\psi}(\boldsymbol{x}_1,\boldsymbol{x}_2),\\ \overline{\boldsymbol{\epsilon}} &= \boldsymbol{0}, & \overline{\boldsymbol{\epsilon}\boldsymbol{\epsilon}^{\mathrm{T}}} &= \boldsymbol{C}_{\epsilon\epsilon},\\ \overline{p^{\mathrm{f}}(\boldsymbol{x})\boldsymbol{\epsilon}} &= \boldsymbol{0}. \end{aligned} \tag{3.25}$$

Thus, the means of the errors in the first-guess and the measurements are zero, and there are no cross correlations between these error terms. Further, we have knowledge of the forecast or first-guess error covariance between two points in space $C^{\mathrm{f}}_{\psi\psi}(\boldsymbol{x}_1,\boldsymbol{x}_2)$, and the observation error covariance matrix $\boldsymbol{C}_{\epsilon\epsilon} \in \Re^{M\times M}$. Note that the error covariance differs from the sample covariance as defined in (2.22) by referring to the true (unknown) state rather than the sample average.

We are now defining a variational functional

$$\begin{aligned} \mathcal{J}[\psi] = &\iint_{\mathcal{D}} \big(\psi^{\mathrm{f}}(\boldsymbol{x}_1)-\psi(\boldsymbol{x}_1)\big)W^{\mathrm{f}}_{\psi\psi}(\boldsymbol{x}_1,\boldsymbol{x}_2)\big(\psi^{\mathrm{f}}(\boldsymbol{x}_2)-\psi(\boldsymbol{x}_2)\big)d\boldsymbol{x}_1 d\boldsymbol{x}_2\\ &+\big(\boldsymbol{d}-\boldsymbol{\mathcal{M}}_{(3)}[\psi_3]\big)^{\mathrm{T}}\boldsymbol{W}_{\epsilon\epsilon}\big(\boldsymbol{d}-\boldsymbol{\mathcal{M}}_{(4)}[\psi_4]\big), \end{aligned} \tag{3.26}$$

where $W^{\mathrm{f}}_{\psi\psi}(\boldsymbol{x}_1,\boldsymbol{x}_2)$ is defined as a functional inverse of $C^{\mathrm{f}}_{\psi\psi}(\boldsymbol{x}_1,\boldsymbol{x}_2)$ from

$$\int_{\mathcal{D}} C^{\mathrm{f}}_{\psi\psi}(\boldsymbol{x}_1,\boldsymbol{x}_2)W^{\mathrm{f}}_{\psi\psi}(\boldsymbol{x}_2,\boldsymbol{x}_3)d\boldsymbol{x}_2 = \delta(\boldsymbol{x}_1-\boldsymbol{x}_3), \tag{3.27}$$

and $\boldsymbol{W}_{\epsilon\epsilon}$ is the inverse of the measurement error covariance matrix $\boldsymbol{C}_{\epsilon\epsilon}$. Here we have used subscripts on the measurement operator and its argument, e.g. $\boldsymbol{\mathcal{M}}_{(3)}[\psi_3]$ indicating that the dummy variable for the integration is $\boldsymbol{x}_3$. This has no implications in this expression but it will be useful in the following derivation.

The variational functional (3.26) measures, in a weighted sense, the distance between an estimate $\psi(\boldsymbol{x})$ and the forecast or first-guess $\psi^{\mathrm{f}}(\boldsymbol{x})$, plus the distance between the estimate and the observations $\boldsymbol{d}$. The field $\psi(\boldsymbol{x})$ which minimizes (3.26) is named $\psi^{\mathrm{a}}(\boldsymbol{x})$. The use of inverses of the error covariances as weights, ensures that the variance minimizing estimate becomes equal to the maximum likelihood estimate in the case with Gaussian error statistics.

3.2.2 Euler–Lagrange equation

To minimize the variational functional, (3.26), we can calculate the variational derivative of $\mathcal{J}[\psi]$ and require that it approaches zero when the arbitrary perturbation $\delta\psi(\boldsymbol{x})$ goes to zero. Thus, we have

$$\delta\mathcal{J} = \mathcal{J}[\psi+\delta\psi] - \mathcal{J}[\psi] = \mathcal{O}(\delta\psi^2). \tag{3.28}$$

Evaluating (3.28) gives

$$\begin{aligned}\delta\mathcal{J} = -2\iint_{\mathcal{D}} &\delta\psi(\boldsymbol{x}_1)W^{\mathrm{f}}_{\psi\psi}(\boldsymbol{x}_1,\boldsymbol{x}_2)\big(\psi^{\mathrm{f}}(\boldsymbol{x}_2)-\psi(\boldsymbol{x}_2)\big)d\boldsymbol{x}_1 d\boldsymbol{x}_2\\ &- 2\boldsymbol{\mathcal{M}}_{(3)}[\delta\psi(\boldsymbol{x}_3)]^{\mathrm{T}}\boldsymbol{W}_{\epsilon\epsilon}\big(\boldsymbol{d}-\boldsymbol{\mathcal{M}}_{(4)}[\psi(\boldsymbol{x}_4)]\big)\\ &+\mathcal{O}(\delta\psi^2)=\mathcal{O}(\delta\psi^2).\end{aligned} \tag{3.29}$$

Thus, to have an extrema of $\mathcal{J}$ we must have

$$\begin{aligned}\iint_{\mathcal{D}} &\delta\psi(\boldsymbol{x}_1)W^{\mathrm{f}}_{\psi\psi}(\boldsymbol{x}_1,\boldsymbol{x}_2)\big(\psi^{\mathrm{f}}(\boldsymbol{x}_2)-\psi^{\mathrm{a}}(\boldsymbol{x}_2)\big)d\boldsymbol{x}_1 d\boldsymbol{x}_2\\ &+\boldsymbol{\mathcal{M}}_{(3)}[\delta\psi(\boldsymbol{x}_3)]^{\mathrm{T}}\boldsymbol{W}_{\epsilon\epsilon}\big(\boldsymbol{d}-\boldsymbol{\mathcal{M}}_{(4)}[\psi^{\mathrm{a}}(\boldsymbol{x}_4)]\big)=0.\end{aligned} \tag{3.30}$$

To proceed we need to get the second term in under the integral and both terms need to be proportional to $\delta\psi$. We will now show that

$$\boldsymbol{\mathcal{M}}_{(3)}[\delta\psi(\boldsymbol{x}_3)]^{\mathrm{T}} = \int_{\mathcal{D}}\delta\psi(\boldsymbol{x}_1)\boldsymbol{\mathcal{M}}^{\mathrm{T}}_{(3)}[\delta(\boldsymbol{x}_1-\boldsymbol{x}_3)]d\boldsymbol{x}_1. \tag{3.31}$$

We start by writing out the measurement of a Dirac delta function, $\delta(\boldsymbol{x}_1-\boldsymbol{x}_3)$, as

$$\mathcal{M}_{i(3)}[\delta(\boldsymbol{x}_1-\boldsymbol{x}_3)] = \int_{\mathcal{D}}\delta(\boldsymbol{x}_1-\boldsymbol{x}_3)\delta(\boldsymbol{x}_3-\boldsymbol{x}_i)d\boldsymbol{x}_3 = \delta(\boldsymbol{x}_1-\boldsymbol{x}_i), \tag{3.32}$$

for $i=1,\ldots,M$ where M is the number of measurements. The subscript (3) on $\mathcal{M}_i$ defines the variable the functional is operating on, thus, the integration variable is $\boldsymbol{x}_3$. Multiplying this equation with $\delta\psi(\boldsymbol{x}_1)$ and integrating in $\boldsymbol{x}_1$ now gives

$$\begin{aligned}\int_{\mathcal{D}}\delta\psi(\boldsymbol{x}_1)\mathcal{M}_{i(3)}[\delta(\boldsymbol{x}_1-\boldsymbol{x}_3)]d\boldsymbol{x}_1 &= \int_{\mathcal{D}}\delta\psi(\boldsymbol{x}_1)\delta(\boldsymbol{x}_1-\boldsymbol{x}_i)d\boldsymbol{x}_1\\ &=\mathcal{M}_{i(1)}[\delta\psi(\boldsymbol{x}_1)]\\ &=\mathcal{M}_{i(3)}[\delta\psi(\boldsymbol{x}_3)].\end{aligned} \tag{3.33}$$

where in the last line, we changed the dummy variable for the integration to $\boldsymbol{x}_3$. Thus, we have obtained (3.31).

We also have that

$$\begin{aligned}\int_{\mathcal{D}} C^{\mathrm{f}}_{\psi\psi}(\boldsymbol{x}_1,\boldsymbol{x}_2)\mathcal{M}^{\mathrm{T}}_{i(3)}[\delta(\boldsymbol{x}_2-\boldsymbol{x}_3)]d\boldsymbol{x}_2 &= C^{\mathrm{f}}_{\psi\psi}(\boldsymbol{x}_1,\boldsymbol{x}_i)\\ &=\mathcal{M}_{i(2)}[C^{\mathrm{f}}_{\psi\psi}(\boldsymbol{x}_1,\boldsymbol{x}_2)].\end{aligned} \tag{3.34}$$

Note that the second term of (3.30), i.e. the measurement term, is constant in the integration with respect to $\boldsymbol{x}_2$. Equations (3.32–3.34) are verified for $i=1,\ldots,M$, and their results can be generalized and substituted into (3.30) which then leads to

$$\begin{aligned}\iint_{\mathcal{D}} \delta\psi(\boldsymbol{x}_1)\Big(&W^{\mathrm{f}}_{\psi\psi}(\boldsymbol{x}_1,\boldsymbol{x}_2)\big(\psi^{\mathrm{f}}(\boldsymbol{x}_2)-\psi^{\mathrm{a}}(\boldsymbol{x}_2)\big)\\ &+\boldsymbol{\mathcal{M}}^{\mathrm{T}}_{(3)}[\delta(\boldsymbol{x}_1-\boldsymbol{x}_3)]\boldsymbol{W}_{\epsilon\epsilon}\big(\boldsymbol{d}-\boldsymbol{\mathcal{M}}_{(4)}[\psi^{\mathrm{a}}(\boldsymbol{x}_4)]\big)\Big)d\boldsymbol{x}_1 d\boldsymbol{x}_2=0,\end{aligned} \tag{3.35}$$

or since this must be true for all $\delta\psi$ we must have

$$\begin{aligned}W^{\mathrm{f}}_{\psi\psi}(\boldsymbol{x}_1,\boldsymbol{x}_2)&\big(\psi^{\mathrm{f}}(\boldsymbol{x}_2)-\psi^{\mathrm{a}}(\boldsymbol{x}_2)\big)\\ &+\boldsymbol{\mathcal{M}}^{\mathrm{T}}_{(3)}[\delta(\boldsymbol{x}_1-\boldsymbol{x}_3)]\boldsymbol{W}_{\epsilon\epsilon}\big(\boldsymbol{d}-\boldsymbol{\mathcal{M}}_{(4)}[\psi^{\mathrm{a}}(\boldsymbol{x}_4)]\big)=0.\end{aligned} \tag{3.36}$$

This is the Euler–Lagrange equation for the variational problem, of which the solution ψ^{a} must be a minimum of $\mathcal{J}$.

Now multiply (3.36) with $C^{\mathrm{f}}_{\psi\psi}(\boldsymbol{x},\boldsymbol{x}_1)$ and integrate with respect to $\boldsymbol{x}_1$. Using the definition (3.27) and the identity (3.34) we get the Euler–Lagrange equation of the form

$$\psi^{\mathrm{a}}(\boldsymbol{x})-\psi^{\mathrm{f}}(\boldsymbol{x})=\boldsymbol{\mathcal{M}}^{\mathrm{T}}_{(3)}[C^{\mathrm{f}}_{\psi\psi}(\boldsymbol{x},\boldsymbol{x}_3)]\boldsymbol{W}_{\epsilon\epsilon}\big(\boldsymbol{d}-\boldsymbol{\mathcal{M}}_{(4)}[\psi^{\mathrm{a}}_4]\big). \tag{3.37}$$

3.2.3 Representer solution

A problem with the Euler–Lagrange equation (3.37) is that ψ^{a} is contained on both sides of the equality sign. To resolve this we first define the vector $\boldsymbol{b}\in\Re^M$ as

$$\boldsymbol{b}=\boldsymbol{W}_{\epsilon\epsilon}\big(\boldsymbol{d}-\boldsymbol{\mathcal{M}}_{(4)}[\psi^{\mathrm{a}}_4]\big), \tag{3.38}$$

and then seek a solution of the form

$$\psi^{\mathrm{a}}(\boldsymbol{x})=\psi^{\mathrm{f}}(\boldsymbol{x})+\boldsymbol{b}^{\mathrm{T}}\boldsymbol{r}(\boldsymbol{x}), \tag{3.39}$$

where we have introduced the vector of representers $\boldsymbol{r}(\boldsymbol{x})\in\Re^M$.

Inserting this into (3.37) gives

$$\psi^{\mathrm{f}}(\boldsymbol{x})-\psi^{\mathrm{f}}(\boldsymbol{x})+\boldsymbol{b}^{\mathrm{T}}\boldsymbol{r}(\boldsymbol{x})=\boldsymbol{\mathcal{M}}^{\mathrm{T}}_{(3)}[C^{\mathrm{f}}_{\psi\psi}(\boldsymbol{x},\boldsymbol{x}_3)]\boldsymbol{b}, \tag{3.40}$$

Thus, we get the influence functions or representers $\boldsymbol{r}(\boldsymbol{x})$ defined as

$$\boldsymbol{r}(\boldsymbol{x})=\boldsymbol{\mathcal{M}}_{(3)}[C^{\mathrm{f}}_{\psi\psi}(\boldsymbol{x},\boldsymbol{x}_3)]. \tag{3.41}$$

Now using (3.39) in (3.38) gives

$$\begin{aligned}\boldsymbol{b}&=\boldsymbol{W}_{\epsilon\epsilon}\big(\boldsymbol{d}-\boldsymbol{\mathcal{M}}_{(4)}[\psi^{\mathrm{f}}_4+\boldsymbol{b}^{\mathrm{T}}\boldsymbol{r}_4]\big)\\ &=\boldsymbol{W}_{\epsilon\epsilon}\big(\boldsymbol{d}-\boldsymbol{\mathcal{M}}_{(4)}[\psi^{\mathrm{f}}_4]\big)-\boldsymbol{W}_{\epsilon\epsilon}\boldsymbol{\mathcal{M}}_{(4)}[\boldsymbol{b}^{\mathrm{T}}\boldsymbol{r}_4]\\ &=\boldsymbol{W}_{\epsilon\epsilon}\big(\boldsymbol{d}-\boldsymbol{\mathcal{M}}_{(4)}[\psi^{\mathrm{f}}_4]\big)-\boldsymbol{W}_{\epsilon\epsilon}\boldsymbol{b}^{\mathrm{T}}\boldsymbol{\mathcal{M}}_{(4)}[\boldsymbol{r}_4],\end{aligned} \tag{3.42}$$

because of the linearity of $\boldsymbol{\mathcal{M}}$. Rearranging gives

$$\boldsymbol{b}+\boldsymbol{W}_{\epsilon\epsilon}\boldsymbol{b}^{\mathrm{T}}\boldsymbol{\mathcal{M}}_{(4)}[\boldsymbol{r}_4])=\boldsymbol{W}_{\epsilon\epsilon}\big(\boldsymbol{d}-\boldsymbol{\mathcal{M}}_{(4)}[\psi^{\mathrm{f}}_4]\big), \tag{3.43}$$

and, multiplying from the left with $\boldsymbol{C}_{\epsilon\epsilon}$, we obtain

$$\boldsymbol{C}_{\epsilon\epsilon}\boldsymbol{b} + \boldsymbol{b}^{\mathrm{T}}\boldsymbol{\mathcal{M}}_{(4)}[\boldsymbol{r}_4] = \boldsymbol{d} - \boldsymbol{\mathcal{M}}_{(4)}[\psi_4^{\mathrm{f}}], \tag{3.44}$$

or

$$\left(\boldsymbol{\mathcal{M}}_{(4)}^{\mathrm{T}}[\boldsymbol{r}_4] + \boldsymbol{C}_{\epsilon\epsilon}\right)\boldsymbol{b} = \boldsymbol{d} - \boldsymbol{\mathcal{M}}_{(4)}[\psi_4^{\mathrm{f}}], \tag{3.45}$$

which is a linear system of equations for $\boldsymbol{b}$. Rewriting by using (3.41) the equation becomes

$$\left(\boldsymbol{\mathcal{M}}_{(3)}\boldsymbol{\mathcal{M}}_{(4)}^{\mathrm{T}}[C_{\psi\psi}^{\mathrm{f}}(\boldsymbol{x}_3,\boldsymbol{x}_4)] + \boldsymbol{C}_{\epsilon\epsilon}\right)\boldsymbol{b} = \boldsymbol{d} - \boldsymbol{\mathcal{M}}_{(4)}[\psi^{\mathrm{f}}(\boldsymbol{x}_4)]. \tag{3.46}$$

A solution can now be found from the equations (3.39), (3.41) and (3.45).

3.2.4 Representer matrix

Note that with direct measurements as given in (3.24), we have

$$\mathcal{M}_{i(3)}\mathcal{M}_{j(4)}^{\mathrm{T}}[C_{\psi\psi}^{\mathrm{f}}(\boldsymbol{x}_3,\boldsymbol{x}_4)] = C_{\psi\psi}^{\mathrm{f}}(\boldsymbol{x}_i,\boldsymbol{x}_j). \tag{3.47}$$

The matrix $C_{\psi\psi}^{\mathrm{f}}(\boldsymbol{x}_i,\boldsymbol{x}_j)$ is often called the representer matrix and with direct measurements it describes the covariances of the first-guess between the two locations $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$.

3.2.5 Error estimate

It is possible to derive an error estimate for the analysis (3.39). The simplest is to use the procedure as derived by *Bennett* (1992) for the time dependent problem. From the definition of the error covariance in (3.25) we can write

$$C_{\psi\psi}^{\mathrm{a}}(\boldsymbol{x}_1,\boldsymbol{x}_2) = \overline{(\psi^{\mathrm{t}}(\boldsymbol{x}_1) - \psi^{\mathrm{a}}(\boldsymbol{x}_1))(\psi^{\mathrm{t}}(\boldsymbol{x}_2) - \psi^{\mathrm{a}}(\boldsymbol{x}_2))}, \tag{3.48}$$

and insert the equation for the analysis to get

$$\begin{aligned} C_{\psi\psi}^{\mathrm{a}}(\boldsymbol{x}_1,\boldsymbol{x}_2) &= \overline{(\psi_1^{\mathrm{t}} - \psi_1^{\mathrm{f}} - \boldsymbol{b}^{\mathrm{T}}\boldsymbol{r}_1)(\psi_2^{\mathrm{t}} - \psi_2^{\mathrm{f}} - \boldsymbol{b}^{\mathrm{T}}\boldsymbol{r}_2)} \\ &= \overline{(\psi_1^{\mathrm{t}} - \psi_1^{\mathrm{f}})(\psi_2^{\mathrm{t}} - \psi_2^{\mathrm{f}})} - 2\overline{(\psi_1^{\mathrm{t}} - \psi_1^{\mathrm{f}})\boldsymbol{b}^{\mathrm{T}}}\boldsymbol{r}_2 + \boldsymbol{r}_1^{\mathrm{T}}\overline{\boldsymbol{b}\boldsymbol{b}^{\mathrm{T}}}\boldsymbol{r}_2. \end{aligned} \tag{3.49}$$

We have used that $\boldsymbol{b}$ is a function of ψ and the representers $\boldsymbol{r}$, are functions of the covariance matrix and then $\overline{\psi}$. Further, we used the property $(\boldsymbol{A}\boldsymbol{B})^{\mathrm{T}} = \boldsymbol{B}^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}$ for matrices $\boldsymbol{A}$ and $\boldsymbol{B}$, and that the covariance is symmetrical in $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$.

The first term is just $C_{\psi\psi}^{\mathrm{f}}$ while the two other terms will be treated next and we now define for convenience

$$\boldsymbol{\mathcal{P}} = \boldsymbol{\mathcal{M}}_{(3)}\boldsymbol{\mathcal{M}}_{(4)}^{\mathrm{T}}[C_{\psi\psi}^{\mathrm{f}}(\boldsymbol{x}_3,\boldsymbol{x}_4)] + \boldsymbol{C}_{\epsilon\epsilon}, \tag{3.50}$$

and the residual or innovation

$$\boldsymbol{h} = \boldsymbol{d} - \boldsymbol{\mathcal{M}}_{(4)}[\psi_4^{\mathrm{f}}]. \tag{3.51}$$

Using (3.41), (3.50) and (3.51) in (3.45) gives $\boldsymbol{b} = \boldsymbol{\mathcal{P}}^{-1}\boldsymbol{h}$. Furthermore, by using (3.23), (3.25), (3.41) and (3.45), in addition to the two definitions above, the second term in (3.49) becomes

$$\begin{aligned}
&-2\overline{(\psi_1^{\mathrm{t}} - \psi_1^{\mathrm{f}})\boldsymbol{b}^{\mathrm{T}}}\boldsymbol{r}_2 \\
&\quad = -2\overline{(\psi_1^{\mathrm{t}} - \psi_1^{\mathrm{f}})(\boldsymbol{\mathcal{P}}^{-1}\boldsymbol{h})^{\mathrm{T}}}\boldsymbol{r}_2 \\
&\quad = -2\overline{(\psi_1^{\mathrm{t}} - \psi_1^{\mathrm{f}})\left(\boldsymbol{\mathcal{P}}^{-1}(\boldsymbol{d} - \boldsymbol{\mathcal{M}}_{(4)}[\psi_4^{\mathrm{f}}])\right)^{\mathrm{T}}}\boldsymbol{r}_2 \\
&\quad = -2\overline{(\psi_1^{\mathrm{t}} - \psi_1^{\mathrm{f}})\left(\boldsymbol{\mathcal{P}}^{-1}(\boldsymbol{\mathcal{M}}_{(4)}[\psi_4^{\mathrm{t}}] + \boldsymbol{\epsilon} - \boldsymbol{\mathcal{M}}_{(4)}[\psi_4^{\mathrm{f}}])\right)^{\mathrm{T}}}\boldsymbol{r}_2 \\
&\quad = -2\overline{(\psi_1^{\mathrm{t}} - \psi_1^{\mathrm{f}})\boldsymbol{\mathcal{M}}_{(4)}^{\mathrm{T}}[\psi_4^{\mathrm{t}} - \psi_4^{\mathrm{f}}]}\boldsymbol{\mathcal{P}}^{-1}\boldsymbol{r}_2 + 0 \\
&\quad = -2\boldsymbol{\mathcal{M}}_{(4)}^{\mathrm{T}}[\overline{(\psi_1^{\mathrm{t}} - \psi_1^{\mathrm{f}})(\psi_4^{\mathrm{t}} - \psi_4^{\mathrm{f}})}]\boldsymbol{\mathcal{P}}^{-1}\boldsymbol{r}_2 \\
&\quad = -2\boldsymbol{\mathcal{M}}_{(4)}^{\mathrm{T}}[C_{\psi\psi}^{\mathrm{f}}(\boldsymbol{x}_1, \boldsymbol{x}_4)]\boldsymbol{\mathcal{P}}^{-1}\boldsymbol{r}_2 \\
&\quad = -2\boldsymbol{r}_1^{\mathrm{T}}\boldsymbol{\mathcal{P}}^{-1}\boldsymbol{r}_2.
\end{aligned} \tag{3.52}$$

Here we have also used that $\overline{\boldsymbol{\epsilon}} = 0$ from (3.25), and that $\boldsymbol{\mathcal{P}}$ is a symmetrical function of the covariance and can be moved outside the averaging.

Further, using $(\boldsymbol{\mathcal{P}}^{-1}\boldsymbol{h})^{\mathrm{T}} = \boldsymbol{h}^{\mathrm{T}}\boldsymbol{\mathcal{P}}^{-1}$, the last term becomes

$$\begin{aligned}
&\boldsymbol{r}_1^{\mathrm{T}}\overline{\boldsymbol{b}\boldsymbol{b}^{\mathrm{T}}}\boldsymbol{r}_2 \\
&\quad = \boldsymbol{r}_1^{\mathrm{T}}\boldsymbol{\mathcal{P}}^{-1}\overline{\boldsymbol{h}\boldsymbol{h}^{\mathrm{T}}}\boldsymbol{\mathcal{P}}^{-1}\boldsymbol{r}_2 \\
&\quad = \boldsymbol{r}_1^{\mathrm{T}}\boldsymbol{\mathcal{P}}^{-1}\overline{(\boldsymbol{d} - \boldsymbol{\mathcal{M}}_{(1)}[\psi_1^{\mathrm{f}}])(\boldsymbol{d} - \boldsymbol{\mathcal{M}}_{(2)}[\psi_2^{\mathrm{f}}])^{\mathrm{T}}}\boldsymbol{\mathcal{P}}^{-1}\boldsymbol{r}_2 \\
&\quad = \boldsymbol{r}_1^{\mathrm{T}}\boldsymbol{\mathcal{P}}^{-1}\overline{(\boldsymbol{\mathcal{M}}_{(1)}[\psi_1^{\mathrm{t}}] + \boldsymbol{\epsilon} - \boldsymbol{\mathcal{M}}_{(1)}[\psi_1^{\mathrm{f}}])(\boldsymbol{\mathcal{M}}_{(2)}[\psi_2^{\mathrm{t}}] + \boldsymbol{\epsilon} - \boldsymbol{\mathcal{M}}_{(1)}[\psi_2^{\mathrm{f}}])^{\mathrm{T}}}\boldsymbol{\mathcal{P}}^{-1}\boldsymbol{r}_2 \\
&\quad = \boldsymbol{r}_1^{\mathrm{T}}\boldsymbol{\mathcal{P}}^{-1}\overline{(\boldsymbol{\mathcal{M}}_{(1)}[\psi_1^{\mathrm{t}} - \psi_1^{\mathrm{f}}] + \boldsymbol{\epsilon})(\boldsymbol{\mathcal{M}}_{(2)}[\psi_2^{\mathrm{t}} - \psi_2^{\mathrm{f}}] + \boldsymbol{\epsilon})^{\mathrm{T}}}\boldsymbol{\mathcal{P}}^{-1}\boldsymbol{r}_2 \\
&\quad = \boldsymbol{r}_1^{\mathrm{T}}\boldsymbol{\mathcal{P}}^{-1}(\boldsymbol{\mathcal{M}}_{(1)}\boldsymbol{\mathcal{M}}_{(2)}^{\mathrm{T}}[\overline{(\psi_1^{\mathrm{t}} - \psi_1^{\mathrm{f}})(\psi_2^{\mathrm{t}} - \psi_2^{\mathrm{f}})}] + \overline{\boldsymbol{\epsilon}\boldsymbol{\epsilon}^{\mathrm{T}}})\boldsymbol{\mathcal{P}}^{-1}\boldsymbol{r}_2 \\
&\quad = \boldsymbol{r}_1^{\mathrm{T}}\boldsymbol{\mathcal{P}}^{-1}\boldsymbol{\mathcal{P}}\boldsymbol{\mathcal{P}}^{-1}\boldsymbol{r}_2 \\
&\quad = \boldsymbol{r}_1^{\mathrm{T}}\boldsymbol{\mathcal{P}}^{-1}\boldsymbol{r}_2.
\end{aligned} \tag{3.53}$$

Thus, an error estimate is given as

$$\begin{aligned}
C_{\psi\psi}^{\mathrm{a}}(\boldsymbol{x}_1, \boldsymbol{x}_2) &= C_{\psi\psi}^{\mathrm{f}}(\boldsymbol{x}_1, \boldsymbol{x}_2) \\
&\quad - \boldsymbol{r}^{\mathrm{T}}(\boldsymbol{x}_1)\left(\boldsymbol{\mathcal{M}}_{(3)}\boldsymbol{\mathcal{M}}_{(4)}^{\mathrm{T}}[C_{\psi\psi}^{\mathrm{f}}(\boldsymbol{x}_3, \boldsymbol{x}_4)] + \boldsymbol{C}_{\epsilon\epsilon}\right)^{-1}\boldsymbol{r}(\boldsymbol{x}_2).
\end{aligned} \tag{3.54}$$

where the definition for $\boldsymbol{\mathcal{P}}$ has been used.

3.2.6 Uniqueness of the solution

By expressing the solution as in (3.39) not all arbitrary functions can be represented. To show that the solution (3.39) is the unique variance minimizing linear solution we proceed with the following argumentation using a geometrical formulation, identical to the formulation used for the time dependent problem by *Bennett* (1992). First define the inner product

$$< f(\boldsymbol{x}_1), g(\boldsymbol{x}_2) >= \iint_{\mathcal{D}} f(\boldsymbol{x}_1) W^{\mathrm{f}}_{\psi\psi}(\boldsymbol{x}_1, \boldsymbol{x}_2) g(\boldsymbol{x}_2) d\boldsymbol{x}_1 d\boldsymbol{x}_2. \tag{3.55}$$

Note that

$$\begin{aligned} &< C^{\mathrm{f}}_{\psi\psi}(\boldsymbol{x}_3, \boldsymbol{x}_1), \psi(\boldsymbol{x}_2) > \\ &\quad = \iint_{\mathcal{D}} C^{\mathrm{f}}_{\psi\psi}(\boldsymbol{x}_3, \boldsymbol{x}_1) W^{\mathrm{f}}_{\psi\psi}(\boldsymbol{x}_1, \boldsymbol{x}_2) \psi(\boldsymbol{x}_2) d\boldsymbol{x}_1 d\boldsymbol{x}_2 \\ &\quad = \psi(\boldsymbol{x}_3), \end{aligned} \tag{3.56}$$

thus, $C^{\mathrm{f}}_{\psi\psi}(\boldsymbol{x}_3, \boldsymbol{x}_1)$ is a "reproducing kernel" for the inner product (3.55) and the expression (3.56) is true for every field ψ in any point $\boldsymbol{x}$.

Recalling the definition of the representer (3.41) we get

$$\begin{aligned} < \boldsymbol{r}(\boldsymbol{x}_1), \psi(\boldsymbol{x}_2) > &=< \boldsymbol{\mathcal{M}}_{(1)}[C^{\mathrm{f}}_{\psi\psi}(\boldsymbol{x}_3, \boldsymbol{x}_1)], \psi(\boldsymbol{x}_2) > \\ &= \boldsymbol{\mathcal{M}}_{(1)}[< C^{\mathrm{f}}_{\psi\psi}(\boldsymbol{x}_3, \boldsymbol{x}_1), \psi(\boldsymbol{x}_2) >] \\ &= \boldsymbol{\mathcal{M}}_{(1)}[\psi(\boldsymbol{x}_1)] \end{aligned} \tag{3.57}$$

Thus, the measurement of a field $\psi(\boldsymbol{x})$ is equivalent to projecting the field onto the representer using the inner product (3.55).

The penalty function (3.26) can now be written entirely in terms of inner products as

$$\mathcal{J}[\psi] =< \psi^{\mathrm{f}} - \psi, \psi^{\mathrm{f}} - \psi > +(\boldsymbol{d} - < \psi, \boldsymbol{r} >)^{\mathrm{T}} \boldsymbol{W}_{\epsilon\epsilon} (\boldsymbol{d} - < \psi, \boldsymbol{r} >). \tag{3.58}$$

Assume now that the minimizing solution is expressed as

$$\psi^{\mathrm{a}}(\boldsymbol{x}) = \psi^{\mathrm{f}}(\boldsymbol{x}) + \boldsymbol{b}^{\mathrm{T}} \boldsymbol{r}(\boldsymbol{x}) + g(\boldsymbol{x}), \tag{3.59}$$

where $g(\boldsymbol{x})$ is an arbitrary function orthogonal to the representers, i.e.

$$< g, \boldsymbol{r} >= \boldsymbol{0}. \tag{3.60}$$

Because of this identity the field g may be regarded as unobservable. Substituting (3.59) into (3.58) gives

$$\begin{aligned} \mathcal{J}[\psi^{\mathrm{a}}] &=< \boldsymbol{r}^{\mathrm{T}}\boldsymbol{b} + g, \boldsymbol{r}^{\mathrm{T}}\boldsymbol{b} + g > \\ &\quad + (\boldsymbol{d} - < \psi^{\mathrm{a}}, \boldsymbol{r} >)^{\mathrm{T}} \boldsymbol{W}_{\epsilon\epsilon} (\boldsymbol{d} - < \psi^{\mathrm{a}}, \boldsymbol{r} >) \\ &\quad = \boldsymbol{b}^{\mathrm{T}} < \boldsymbol{r}, \boldsymbol{r}^{\mathrm{T}} > \boldsymbol{b} + \boldsymbol{b}^{\mathrm{T}} < \boldsymbol{r}, g > + < g, \boldsymbol{r}^{\mathrm{T}} > \boldsymbol{b} + < g, g > \\ &\quad + (\boldsymbol{d} - < \psi^{\mathrm{f}}, \boldsymbol{r} > - \boldsymbol{b}^{\mathrm{T}} < \boldsymbol{r}, \boldsymbol{r}^{\mathrm{T}} > - < g, \boldsymbol{r} >)^{\mathrm{T}} \\ &\qquad \times \boldsymbol{W}_{\epsilon\epsilon} (\boldsymbol{d} - < \psi^{\mathrm{f}}, \boldsymbol{r} > - \boldsymbol{b}^{\mathrm{T}} < \boldsymbol{r}, \boldsymbol{r}^{\mathrm{T}} > - < g, \boldsymbol{r} >). \end{aligned} \tag{3.61}$$

Defining the residual

$$\boldsymbol{h} = \boldsymbol{d} - < \psi^{\mathrm{f}}, \boldsymbol{r} >, \tag{3.62}$$

and using the definition of the representer matrix,

$$\boldsymbol{R} =< \boldsymbol{r}_3, \boldsymbol{r}_4^{\mathrm{T}} >= \boldsymbol{\mathcal{M}}_{(3)}[\boldsymbol{r}_3^{\mathrm{T}}], \tag{3.63}$$

and (3.41) and (3.47), we get the penalty function of the form

$$\mathcal{J}[\psi^{\mathrm{a}}] = \boldsymbol{b}^{\mathrm{T}}\boldsymbol{R}\boldsymbol{b} + < g, g > + (\boldsymbol{h} - \boldsymbol{R}\boldsymbol{b})^{\mathrm{T}}\boldsymbol{W}_{\epsilon\epsilon}(\boldsymbol{h} - \boldsymbol{R}\boldsymbol{b}). \tag{3.64}$$

The original penalty function (3.26) has now been reduced to a compact form where the disposable parameters are $\boldsymbol{b}$ and $g(\boldsymbol{x})$. If $\boldsymbol{\psi}$ minimizes $\mathcal{J}$ then clearly $< g, g >= 0$ and thus

$$g(\boldsymbol{x}) \equiv 0. \tag{3.65}$$

The unobservable field g must be discarded, reducing $\mathcal{J}$ from the infinite dimensional quadratic form (3.26) to the finite dimensional quadratic form

$$\mathcal{B}[\boldsymbol{b}] = \boldsymbol{b}^{\mathrm{T}}\boldsymbol{R}\boldsymbol{b} + (\boldsymbol{h} - \boldsymbol{R}\boldsymbol{b})^{\mathrm{T}}\boldsymbol{W}_{\epsilon\epsilon}(\boldsymbol{h} - \boldsymbol{R}\boldsymbol{b}), \tag{3.66}$$

where $\mathcal{B}[\boldsymbol{b}] = \mathcal{J}[\psi^{\mathrm{a}}]$.

3.2.7 Minimization of the penalty function

The minimizing solution for $\boldsymbol{b}$ can again be found by setting the variational derivative of (3.66) with respect to $\boldsymbol{b}$ equal to zero,

$$\mathcal{B}[\boldsymbol{b}+\delta\boldsymbol{b}] - \mathcal{B}[\boldsymbol{b}] = 2\delta\boldsymbol{b}^{\mathrm{T}}\boldsymbol{R}\boldsymbol{b} + 2\delta\boldsymbol{b}^{\mathrm{T}}\boldsymbol{R}\boldsymbol{W}_{\epsilon\epsilon}(\boldsymbol{R}\boldsymbol{b} - \boldsymbol{h}) + \mathcal{O}(\delta\boldsymbol{b}^2) = \mathcal{O}(\delta\boldsymbol{b}^2), \tag{3.67}$$

which gives

$$\delta\boldsymbol{b}^{\mathrm{T}}\big(\boldsymbol{R}\boldsymbol{b} + \boldsymbol{R}\boldsymbol{W}_{\epsilon\epsilon}(\boldsymbol{R}\boldsymbol{b} - \boldsymbol{h})\big) = 0, \tag{3.68}$$

or

$$\boldsymbol{R}\boldsymbol{b} + \boldsymbol{R}\boldsymbol{W}_{\epsilon\epsilon}(\boldsymbol{R}\boldsymbol{b} - \boldsymbol{h}) = 0, \tag{3.69}$$

since $\delta\boldsymbol{b}$ is arbitrary. This equation can be written as

$$\boldsymbol{R}(\boldsymbol{b} + \boldsymbol{W}_{\epsilon\epsilon}\boldsymbol{R}\boldsymbol{b} - \boldsymbol{W}_{\epsilon\epsilon}\boldsymbol{h}) = 0, \tag{3.70}$$

which leads to the standard linear system of equations

$$(\boldsymbol{R} + \boldsymbol{C}_{\epsilon\epsilon})\boldsymbol{b} = \boldsymbol{h}, \tag{3.71}$$

or

$$\boldsymbol{b} = \boldsymbol{\mathcal{P}}^{-1}\boldsymbol{h}, \tag{3.72}$$

as the solution for $\boldsymbol{b}$. Note that we have used that $\boldsymbol{R} = \mathcal{M}_{(i)}[\boldsymbol{r}_i]$ for all i.

3.2.8 Prior and posterior value of the penalty function

Inserting the first-guess value ψ^{f}, into the penalty function (3.58) gives

$$\mathcal{J}[\psi^{\mathrm{f}}] = (\boldsymbol{d}- < \psi^{\mathrm{f}}, \boldsymbol{r} >)^{\mathrm{T}} \boldsymbol{W}_{\epsilon\epsilon} (\boldsymbol{d}- < \psi^{\mathrm{f}}, \boldsymbol{r} >) = \boldsymbol{h}^{\mathrm{T}} \boldsymbol{W}_{\epsilon\epsilon} \boldsymbol{h}. \tag{3.73}$$

This is known as the prior value of the penalty function.

Similarly by inserting the minimizing solution (3.72) into the penalty function (3.66) we get the following,

$$\begin{aligned}
\mathcal{J}[\boldsymbol{\mathcal{P}}^{-1}\boldsymbol{h}] &= (\boldsymbol{\mathcal{P}}^{-1}\boldsymbol{h})^{\mathrm{T}} \boldsymbol{R} (\boldsymbol{\mathcal{P}}^{-1}\boldsymbol{h}) + (\boldsymbol{h} - \boldsymbol{R}\boldsymbol{\mathcal{P}}^{-1}\boldsymbol{h})^{\mathrm{T}} \boldsymbol{W}_{\epsilon\epsilon} (\boldsymbol{h} - \boldsymbol{R}\boldsymbol{\mathcal{P}}^{-1}\boldsymbol{h}) \\
&= \boldsymbol{h}^{\mathrm{T}} \boldsymbol{\mathcal{P}}^{-1} \boldsymbol{R} \boldsymbol{\mathcal{P}}^{-1} \boldsymbol{h} + \boldsymbol{h}^{\mathrm{T}} (\boldsymbol{R}\boldsymbol{\mathcal{P}}^{-1} - \boldsymbol{I}) \boldsymbol{W}_{\epsilon\epsilon} (\boldsymbol{R}\boldsymbol{\mathcal{P}}^{-1} - \boldsymbol{I}) \boldsymbol{h} \\
&= \boldsymbol{h}^{\mathrm{T}} \left\{ \boldsymbol{\mathcal{P}}^{-1} \boldsymbol{R} \boldsymbol{\mathcal{P}}^{-1} + (\boldsymbol{R}\boldsymbol{\mathcal{P}}^{-1} - \boldsymbol{I}) \boldsymbol{W}_{\epsilon\epsilon} (\boldsymbol{R}\boldsymbol{\mathcal{P}}^{-1} - \boldsymbol{I}) \right\} \boldsymbol{h} \\
&= \boldsymbol{h}^{\mathrm{T}} \left\{ \boldsymbol{\mathcal{P}}^{-1} \boldsymbol{R} \boldsymbol{\mathcal{P}}^{-1} + \boldsymbol{\mathcal{P}}^{-1} (\boldsymbol{R} - \boldsymbol{\mathcal{P}}) \boldsymbol{W}_{\epsilon\epsilon} (\boldsymbol{R} - \boldsymbol{\mathcal{P}}) \boldsymbol{\mathcal{P}}^{-1} \right\} \boldsymbol{h} \\
&= \boldsymbol{h}^{\mathrm{T}} \boldsymbol{\mathcal{P}}^{-1} \left\{ \boldsymbol{R} + (\boldsymbol{R} - \boldsymbol{\mathcal{P}}) \boldsymbol{W}_{\epsilon\epsilon} (\boldsymbol{R} - \boldsymbol{\mathcal{P}}) \right\} \boldsymbol{\mathcal{P}}^{-1} \boldsymbol{h} \\
&= \boldsymbol{h}^{\mathrm{T}} \boldsymbol{\mathcal{P}}^{-1} \left\{ \boldsymbol{R} + \boldsymbol{C}_{\epsilon\epsilon} \right\} \boldsymbol{\mathcal{P}}^{-1} \boldsymbol{h} \\
&= \boldsymbol{h}^{\mathrm{T}} \boldsymbol{\mathcal{P}}^{-1} \boldsymbol{\mathcal{P}} \boldsymbol{\mathcal{P}}^{-1} \boldsymbol{h} \\
&= \boldsymbol{h}^{\mathrm{T}} \boldsymbol{\mathcal{P}}^{-1} \boldsymbol{h} \\
&= \boldsymbol{h}^{\mathrm{T}} \boldsymbol{b},
\end{aligned} \tag{3.74}$$

as long as $\boldsymbol{b}$ is given from (3.72). This is known as the posterior value of the penalty function.

It is explained by *Bennett* (2002, section 2.3) that the reduced penalty function is a $\mathcal{X}^2_M$ variable. Thus, we have a mean to test the validity of our statistical assumptions, by checking if the value of reduced penalty function is a Gaussian variable with mean equal to M and variance equal to $2M$. This could be done rigorously by repeated minimizations of the penalty function using different data sets.

3.3 Discrete form

When discretized on a numerical grid, (3.22–3.23) are written as

$$\boldsymbol{\psi}^{\mathrm{f}} = \boldsymbol{\psi}^{\mathrm{t}} + \boldsymbol{p}^{\mathrm{f}}, \tag{3.75}$$

$$\boldsymbol{d} = \boldsymbol{M}\boldsymbol{\psi}^{\mathrm{t}} + \boldsymbol{\epsilon}, \tag{3.76}$$

where $\boldsymbol{M}$, now called the measurement matrix, is the discrete representation of $\boldsymbol{\mathcal{M}}$.

The statistical null hypothesis $\mathcal{H}_0$ is then

$$\begin{aligned}
\overline{\boldsymbol{p}^{\mathrm{f}}} &= \boldsymbol{0}, & \overline{\boldsymbol{p}^{\mathrm{f}}(\boldsymbol{p}^{\mathrm{f}})^{\mathrm{T}}} &= \boldsymbol{C}^{\mathrm{f}}_{\psi\psi}, \\
\overline{\boldsymbol{\epsilon}} &= \boldsymbol{0}, & \overline{\boldsymbol{\epsilon}\boldsymbol{\epsilon}^{\mathrm{T}}} &= \boldsymbol{C}_{\epsilon\epsilon}, \\
\overline{\boldsymbol{p}^{\mathrm{f}}\boldsymbol{\epsilon}^{\mathrm{T}}} &= \boldsymbol{0}.
\end{aligned} \tag{3.77}$$

By using the same statistical procedure as in Sect. 3.1, or alternatively by minimizing the variational functional

$$\mathcal{J}[\boldsymbol{\psi}^{\mathrm{a}}] = (\boldsymbol{\psi}^{\mathrm{f}} - \boldsymbol{\psi}^{\mathrm{a}})^{\mathrm{T}} (\boldsymbol{C}^{\mathrm{f}}_{\psi\psi})^{-1} (\boldsymbol{\psi}^{\mathrm{f}} - \boldsymbol{\psi}^{\mathrm{a}}) + (\boldsymbol{d} - \boldsymbol{M}\boldsymbol{\psi}^{\mathrm{a}})^{\mathrm{T}} \boldsymbol{W}_{\epsilon\epsilon} (\boldsymbol{d} - \boldsymbol{M}\boldsymbol{\psi}^{\mathrm{a}}), \tag{3.78}$$

with respect to $\boldsymbol{\psi}^{\mathrm{a}}$, one get,

$$\boldsymbol{\psi}^{\mathrm{a}} = \boldsymbol{\psi}^{\mathrm{f}} + \boldsymbol{r}^{\mathrm{T}} \boldsymbol{b}, \tag{3.79}$$

where the influence functions (e.g. error covariance functions for direct measurements) are given as

$$\boldsymbol{r} = \boldsymbol{M}\boldsymbol{C}^{\mathrm{f}}_{\psi\psi}, \tag{3.80}$$

i.e. “measurements” of the error covariance matrix $\boldsymbol{C}^{\mathrm{f}}_{\psi\psi}$. Thus, $\boldsymbol{r}$ is a matrix where each row contains a representer for a particular measurement. The coefficients $\boldsymbol{b}$ are determined from the system of linear equations

$$(\boldsymbol{M}\boldsymbol{C}^{\mathrm{f}}_{\psi\psi}\boldsymbol{M}^{\mathrm{T}} + \boldsymbol{C}_{\epsilon\epsilon})\boldsymbol{b} = \boldsymbol{d} - \boldsymbol{M}\boldsymbol{\psi}^{\mathrm{f}}. \tag{3.81}$$

In addition the error estimate (3.54) becomes

$$\boldsymbol{C}^{\mathrm{a}}_{\psi\psi} = \boldsymbol{C}^{\mathrm{f}}_{\psi\psi} - \boldsymbol{r}^{\mathrm{T}} \left(\boldsymbol{M}\boldsymbol{C}^{\mathrm{f}}_{\psi\psi}\boldsymbol{M}^{\mathrm{T}} + \boldsymbol{C}_{\epsilon\epsilon} \right)^{-1} \boldsymbol{r}. \tag{3.82}$$

Thus, the inverse estimate $\boldsymbol{\psi}^{\mathrm{a}}$, is given by the first-guess $\boldsymbol{\psi}^{\mathrm{f}}$, plus a linear combination of influence functions $\boldsymbol{r}^{\mathrm{T}}\boldsymbol{b}$, one for each of the measurements. The coefficients $\boldsymbol{b}$ are clearly small if the first-guess is close to the data, and large if the residual between the data and the first-guess is large.

Note that a more common way of writing the previous equations is the following:

$$\boldsymbol{\psi}^{\mathrm{a}} = \boldsymbol{\psi}^{\mathrm{f}} + \boldsymbol{K}(\boldsymbol{d} - \boldsymbol{M}\boldsymbol{\psi}^{\mathrm{f}}), \tag{3.83}$$

$$\boldsymbol{C}^{\mathrm{a}}_{\psi\psi} = (\boldsymbol{I} - \boldsymbol{K}\boldsymbol{M})\boldsymbol{C}^{\mathrm{f}}_{\psi\psi}, \tag{3.84}$$

$$\boldsymbol{K} = \boldsymbol{C}^{\mathrm{f}}_{\psi\psi}\boldsymbol{M}^{\mathrm{T}}(\boldsymbol{M}\boldsymbol{C}^{\mathrm{f}}_{\psi\psi}\boldsymbol{M}^{\mathrm{T}} + \boldsymbol{C}_{\epsilon\epsilon})^{-1}, \tag{3.85}$$

where the matrix $\boldsymbol{K}$ is often called the Kalman gain. This can be derived directly from (3.79)–(3.82) by rearranging terms, and it is the standard way of writing the analysis equations for the Kalman filter to be discussed in Chap. 4. The numerical evaluation of these equations, however, is simpler and more efficient using the form (3.79)–(3.82)