Θ - a set of unknown parameters
(e.g. initial conditions of an ODE or PDE)

$M(\Theta)$ - M is a computational / mathematical model.
(e.g.: solve the ODE or PDE)

y - data (formerly called observations)

We <u>assume</u> that:

$$\underbrace{y = M(\Theta) + \text{noise}}$$

-> computational model (with "right" choice of Θ) can reproduce data up to "noise".

-> Note that there is no "H" matrix, it can be absorbed in the function M.

Further assume that the noise is Gaussian with known covariance R

$$y = M(\Theta) + \eta, \quad \eta \sim N(0,R)$$

and that we have prior information about the parameters Θ.

$p_0(\Theta)$ is given (e.g. $N(\mu, P)$).

$\Rightarrow$ We have a prior: $p_0(\Theta) = N(\mu, P)$

— " — likelihood: $p(y|\Theta) = N(y - M(\Theta), R)$

↳ posterior distribution:

$$p(\Theta|y) = p_0(\Theta)\,p(y|\Theta).$$

Goal in parameter estimation: find posterior $p(\theta|y)$.

Special Case: Gaussian prior, linear M

→ this gives you formulas similar to KF / variational methods

$$p(\theta|y) \propto \exp(-F(\theta))$$

$$F(\theta) = \frac{1}{2}(y - M\theta)^T R^{-1}(y - M\theta) + \frac{1}{2}(\theta - \mu)^T P^{-1}(\theta - \mu)$$

Do the usual thing: solve $\nabla F = 0$ and find $\nabla^2 F$

$$\nabla^2 F = M^T R^{-1} M + P^{-1}$$ is $\overset{\text{inverse of}}{\text{posterior}}$ covariance

↳ $P_{post} = (P^{-1} + M^T R^{-1} M)^{-1} = (I - KH)P$

$K = PM^T(MPM^T + R)^{-1}$

$\nabla F = 0$ gives posterior mean:

$$M^T R^{-1}(M\theta - y) + P^{-1}(\theta - \mu) = 0$$

$$(M^T R^{-1} M) + P^{-1})\theta = P^{-1}\mu + M^T R^{-1} y$$

$$\vdots$$

$$\Rightarrow \underline{\theta = \mu + K(y - M\mu)}$$

→ You can think of inverse problems as one step of a DA problem, but typically with not so informative priors.

→ In DA, priors are "sharpened" over time as we sequentially assimilate data. This does not (automatically) happen in inverse problems.

Or you could have a nonlinear model:

$$-\log p(\theta|y) = \frac{1}{2}(\theta-\mu)P^{-1}(\theta-\mu) + \frac{1}{2}(M(\theta)-y)^T R^{-1}(M(\theta)-y)$$

↳ You can use Gauss-Newton optimization + Hessian approximations ("4D-Var").

Example:

Find initial conditions of ODE/PDE

$M(\theta)$ → simulate ODE/PDE up to time T

y → observations at some grid points / of some variables at time T.

prior → use "climatology".

⇒ We did this for L95 in HW!

↳ You could also use importance sampling.

Perhaps due to scalability issues this is not very popular.

↳ Many people use MCMC

We will talk about MCMC next.

MCMC has similar but perhaps not so severe scalability issues and these are less well known which lulls people into thinking it works.

Reality check

What does a real problem look like and who wants to solve it?

Examples:

DA

NWP → the biggest user in DA

Oceans

Hurricanes

Climate ?!? → DA is in infancy.

Space weather

Geomag → also just starting out

* geomag over long time-scales *

* Satellite tracking *

* forecasting for solar energy *

Inverse problems

Reservoir modelling / forecasting

* Combustion modelling / simulation *

* Image deblurring *