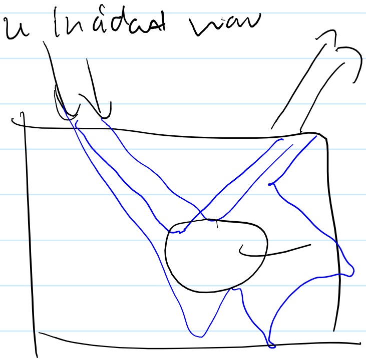
Lecture I

Monday, January 19, 2015 6:57 AM

motivation examples,

u incident wave

observation $\vec{y}_{obs}$



hidden object $\vec{m}$

Math model

$$y_{obs} = h(\vec{m}, u) \quad \text{Forward map}$$

$$\frac{\partial \vec{u}}{\partial t} + \nabla \cdot \vec{F}(u) = 0$$

$$u(0, x) = u_0$$

goal: reconstruct $\vec{m}$ given y_{obs}, $h(\vec{m}, u)$

$$\vec{m} = h^{-1}(y_{obs})$$

have problems here;

Hadamard's wellposedness:

1) Existence (surjectivity)
2) Uniqueness (injectivity)
3) Stability (continuity)

→ h does not satisfy these two conditions

Ex

$$y_{obs} = h(\vec{m}) = Am$$

Ex

$$y^{obs} = h(\vec{m}) = Am$$

$y^{obs} \in \mathbb{R}^n$, $Am \in \mathbb{R}^m$

$A \in \mathbb{R}^{n \times m}$; not square if $n \neq m$

$\Rightarrow N(A) \neq \{0\}$ not injective

In fact A is ill-conditioned $\Rightarrow$ not stable

$\Rightarrow$ use least squares to solve the inverse problem,

$$\min_m J = \frac{1}{2} \| y^{obs} - h(m,u) \|_2^2$$

$$h(m,u) = \int u \, ds$$

u solution of a PDE

$\Rightarrow$ J is not a nice parabola $\Rightarrow$ make it parabola

$$\min \tilde{J} = J + \frac{k}{2} \| R^{-1/2} m \|_2^2$$

s.t PDE contraints:

Ex:

$$y^{obs} = Am$$

$$\tilde{J} = \frac{1}{2} \| y^{obs} - Am \|^2 + \frac{k}{2} \| R^{-1/2} m \|^2$$

$$\frac{\partial \tilde{J}}{\partial m} = A^T (Am - y^{obs}) + k R^{-1} m = 0$$

$$\Downarrow$$

$$(A^T A + k R^{-1}) m = A^T y^{obs}$$

What about noise in y^{obs}

In practice

$$y^{obs} = h(m, u) + e$$

How to incorporate e in our inverse solution?

$\Downarrow$ many ways

1) Evidence theory
2) Fuzzy set
3) $\boxed{\text{probability theory}}$
4) etc . . . .

Some concepts about probability

Def: deterministic event: completely predictable

Def: random event: not deterministic, not completely predictable.

$\Downarrow$

choose to use probability to express unpredictability
or uncertainty:

Ex: $\mathbb{P}[\text{deterministic event}] := 1$

$0 \leq \mathbb{P}[\text{random event}] \leq 1$

Probability space: $(\Omega, \mathcal{F}, \mathbb{P})$

- Ω : sample space
- $\mathcal{F}$: σ-algebra : collection (measurable) events of information Ω

$\mathbb{P}$: probability measure : assign weight on $A \in \mathcal{F}$

Ex: toss a coin: $\Omega = \{\{head\}, \{tail\}\}$

$\mathcal{F} = \{\phi, \{head\}, \{tail\}, \Omega\}$

$\mathbb{P}[\{head\}] = \mathbb{P}[\{tail\}] = 1/2$

$\mathbb{P}[\phi] = 0$

$\mathbb{P}[\Omega] = 1$

Conditional probability

Def: Conditional probability of A given B

$$\mathbb{P}[A|B] = \frac{\mathbb{P}[A \cap B]}{\mathbb{P}[B]}$$

$\downarrow$

probability of the event A when B has already happened.

Ex: roll a dice:

$A :=$ getting face 6

$B =$ event of getting face bigger than 4:

$\mathbb{P}[A \mid B]$

Solution 1:

$A = \{6\}$

$B = \{5, 6\}$

$A \cap B = \{6\}$

$$\mathbb{P}[A \mid B] = \frac{\mathbb{P}[A \cap B]}{\mathbb{P}[B]} = \frac{\mathbb{P}[\{6\}]}{\mathbb{P}[\{5,6\}]} = \frac{1/6}{2/6} = \frac{1}{2}$$

Solution 2: $B = \{5, 6\} \Rightarrow \mathbb{P}[A \mid B] = 1/2$

Random variables

Def:

$$M : \Omega \ni \omega \longmapsto M(\omega) \in S$$

randomness: because of ω

$M(\omega)$

Def: probability distribution of a random variable m is defined a probability measure:

$$\mu_m(A) := \mathbb{P}\left[m^{-1}(A)\right] = \mathbb{P}\left[\{m \in A\}\right]$$

$$\{m \in A\} := \{\omega : m(\omega) \in A\}$$

Def: probability density : $S = \mathbb{R}^n$

$\pi_m(m)$

$$\mu_m(A) = \mathbb{P}\left[m^{-1}(A)\right] := \int_A \pi_m(m)\, dm$$

$\Rightarrow$ for set $(\Omega, \mathcal{F}, \mathbb{P})$

=) forget $(\Omega, \mathcal{F}, \mathbb{P})$

we focus on m, $\pi_M(m)$

Expectation :
$$E[m] := \int_S m\, \pi(m)\, dm$$

$$\mu_{M|y}(\{M \in A\} \mid Y=y) = \int_A \pi(m|y)\, dm$$

conditional density

Lemma :
$$\pi(m|y) = \frac{\pi(m,y)}{\pi(y)}$$

Proof:
$$\mathbb{P}[A \mid B] = \frac{\mathbb{P}[A \cap B]}{\mathbb{P}[B]}$$

Corollary : (Bayes' formula)
$$\pi(m|y) = \frac{\pi(y|m)\, \pi(m)}{\pi(y)}$$

Def: $\pi(m)$: prior distribution : distribution of m before making any observation y

Def: $\pi(y|m)$: likelihood : density of y given m
likelihood of m that generate y

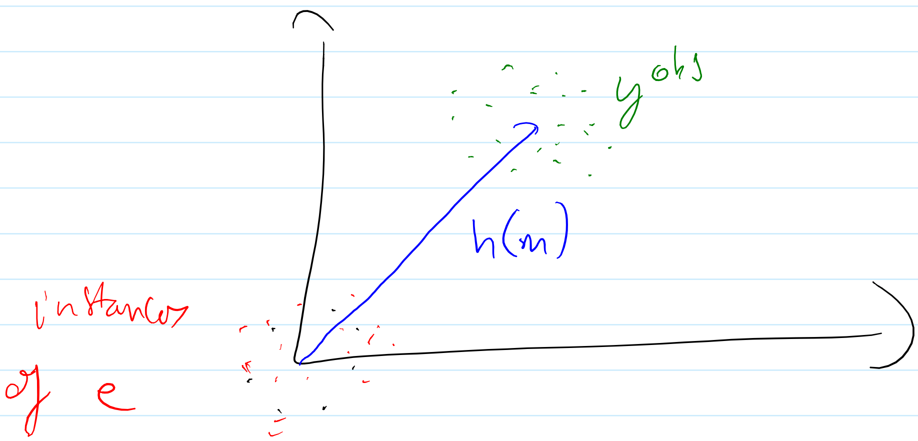
Def: $\pi(m|y)$: posterior distribution of m given y.

Construction of likelihood

additive noise model

$$y^{obs} = h(m) + e$$

$$e \sim \pi_E(e) \qquad \pi(y^{obs}|m)\ ??$$



$$\Rightarrow \pi(y^{obs}|m) = \pi_E\left(y^{obs} - h(m)\right)$$

more rigorously:

$$\mu_{y^{obs}|m}(A) \overset{def}{=} \mu_E\left(A - h(m)\right)$$

$$\mu_{y^{obs}|m}(A) \overset{def}{=} \int_A \pi(y^{obs}|m)\, dy^{obs}$$

$$\mu_E\left(A - h(m)\right) \overset{def}{=} \int_A \pi_E(e)\, de$$

|| change of variable $e := y^{obs} - h(m)$

$$\int \pi_E\left(y^{obs} - h(m)\right) dy^{obs}$$

$$\forall A \in \mathcal{S} \qquad \int_A \pi_E\left(y^{obs} - h(m)\right) dy^{obs}$$

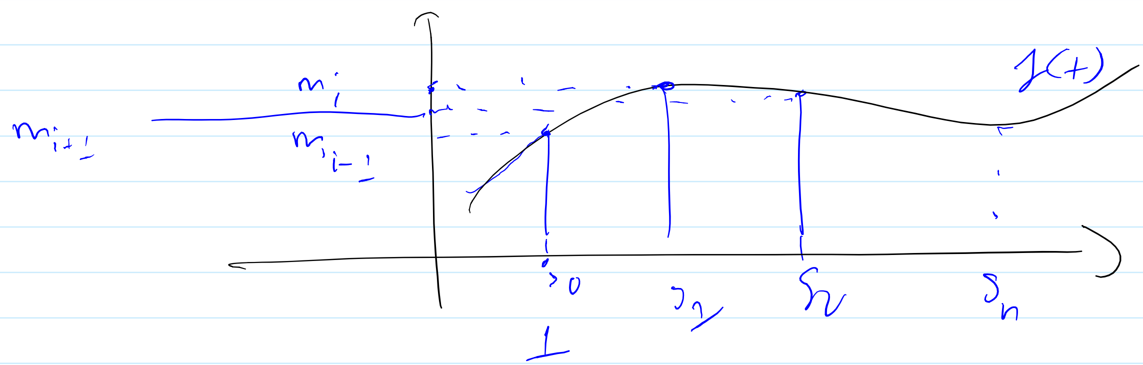
$$\Downarrow$$

$$\pi(y^{obs} \mid m) = \pi_E\left(y^{obs} - h(m)\right)$$

Construction priors :

linear

$$h(m) \overset{!}{=} A\,m$$



Ex :

$$y(s_i) = \int_0^1 g(s_i, t)\, f(t)\, dt + e_i$$

$\Downarrow$ discretize

$$y^{obs} = A\,m + E$$

$$m = [f(s_0), \ldots, f(s_n)]$$

belief: subjective : smooth (f)

$$m_i = \frac{1}{2}\left(m_{i-1} + m_{i+1}\right)$$

$\Downarrow$ uncertain

$$m_i = \frac{1}{2}(m_{i-1} + m_{i+1}) + W_i$$

$$W_i \sim N(0, \sigma^2)$$

$$\Downarrow$$

$$Lm = W$$

$$L = \begin{bmatrix} -1 & 2 & -1 & & & \\ & \ddots & \ddots & \ddots & & \\ & & & -1 & 2 & -1 \end{bmatrix} \in \mathbb{R}^{(n-1)\times(n+1)}$$

$$\Downarrow$$

$$m \sim \exp\left(-\frac{1}{2\sigma^2}\|Lm\|^2\right)$$

assume f has (almost) zero Dirichlet conditions

$$M_0 = \frac{1}{2}(M_1 + M_{-1}) + w_0$$

$$= \frac{1}{2}(M_1 + 0) + w_0$$

Same for M_n

$$\rightarrow \quad m \sim \exp\left(-\frac{1}{} \|L_n m\|^2\right)$$

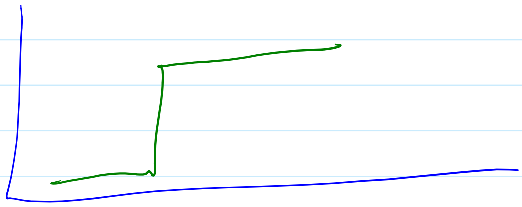
$$\Rightarrow \quad m \sim \exp\left(-\frac{1}{2\sigma^2}\|L_D m\|^2\right)$$

$$L_D = \begin{bmatrix} 2 & -1 & & \\ -1 & 2 & -1 & \\ & & \ddots & \\ & & -1 & 2 \end{bmatrix}$$

$$\|L_D m\|^2 \approx c\|\Delta m\|^2$$

second finite difference

non-smooth prior.



$$m_i = m_{i-1} + w_i$$

$$w_i \sim N(0, \sigma^2)$$

$$m \sim \exp\left(-\frac{1}{2\sigma^2}\|L_N m\|^2\right)$$

$$L_N = \begin{bmatrix} -1 & 1 & & \\ & \ddots & \ddots & \\ & & & -1 \end{bmatrix}$$

$$\| L_N m \|^2 \approx \| \nabla m \|^2$$

Recall

$$\pi_{post}(m \mid y^{obs}) \propto \pi_{like}(y^{obs} \mid m) \times \pi_{prior}(m)$$

Solution of the Bayesian inverse

* Linear forward model.

$$h(m) = Am$$

$$\pi_{post}(m \mid y^{obs}) \propto \exp(-\tilde{J})$$

$$\tilde{J} = \frac{1}{2\sigma^2} \| y^{obs} - Am \|^2 + \frac{\alpha}{2} \| R^{-1/2} m \|^2$$

Tihkonov regularization

+ Compute mean

*
1) complete square
2) derivative w.r.t m, set it to zero
then solve m

* Variance / Covariance
1) take the second derivative Σ
inverse of covariance

Def: Maximum a posteriori (MAP)
most likely point of a distribution
= point at which probability density is highest

MAP = mean for Gaussian

How about nonlinear $h(m)$

$$\pi_{post} \propto \exp(-J)$$

$$J = \frac{1}{2\sigma^2} \| y^{ob} - h(m) \|^2 + \frac{k}{2} \| R^{-1/2} m \|^2$$

MAP point is still valid

MAP = arg min J

m

How about mean, variance

$$\text{mean} = \mathbb{E}_{\pi_{post}}[m] := \int_{S = \mathbb{R}^n} m\, \pi_{post}(m|y^{obs})\, dm$$

take $S = [0, 1]$

$$\int_0^1 m\, \pi_{post}(m|y^{obs})\, dm$$

|| rectangular rule

$$\frac{1}{N}(m_1 + \dots + m_N)$$

~~not true~~

too expensive n large

What can we do here

Thm

Law of large numbers

$m_1, \dots, m_N$ independently identically distributed under some probability density

Then:

$$S = \frac{1}{N}(m_1 + \dots + m_N) \xrightarrow{a.s} \mathbb{E}[m]$$

$$\Rightarrow S_N = \frac{1}{N}(m_1 + \dots + m_N) \xrightarrow[\pi_{post}]{a.s} \mathbb{E}[m]$$

Central limit theorem

$\boxed{m_1, \dots, m_N \quad \text{i.i.d} \quad \pi}$ with mean $\bar{m}$

$\sigma^2 < \infty$

$$z_N = \frac{1}{\sigma\sqrt{N}}(m_1 + \dots + m_N) - \frac{\bar{m}}{\sigma}\sqrt{N}$$

$\downarrow$ weakly / distribution (weak *)

$$N(0, 1)$$

$$\mathbb{E}[(z_N - 0)^2] \approx 1$$

$\| \text{def}$

$$\int \left[\frac{1}{\sigma\sqrt{N}}(m_1 + \dots + m_N) - \frac{\bar{m}}{\sigma}\sqrt{N}\right]^2 dm$$

$$\|$$

$$\int \frac{N}{\sigma^2}\Big[\underbrace{\frac{m_1 + \dots + m_N}{N}}_{S_N} - \bar{m}\Big]^2 dm \overset{C.L.T}{\simeq} 1$$

$$\mathbb{E}\|S_N - \bar{m}\|^2 \approx \frac{\sigma^2}{N}$$

$$\| S_N - \bar{m} \| \approx \frac{\sigma}{\sqrt{N}}$$

$$\mathbb{E}[m] = \int_S m \, \pi_{post}(m) \, dm$$

$\Downarrow$ MCMC convergence rate

compute autocorrelation function

$\Downarrow$

every k sample I get 1 uncorrelated sample

$\Downarrow$?

every k sample I get 1 independent samples

$\tilde{N} = \frac{N}{k}$ effective no i.i.d samples

$$\| S_N - \bar{m} \| \approx \frac{\sigma}{\sqrt{\frac{N}{k}}}$$

Summary

1) if you can draw i.i.d good $\Rightarrow$ LLN, CLT

2) If not, resort to MCMC

How to improve MCMC

1) Make k as small as possible

* $\frac{\sigma}{\sqrt{\frac{N}{k}}}$ is conservative rate

How to make k small:

$$q(m, p) \propto \exp\left(-\frac{1}{2\gamma^2}\|p - m\|^2\right)$$

$\Updownarrow$ drawing

$$p = m + \sigma \mathcal{N}(0, I)$$

random walk

discretization $\Updownarrow$

forward Euler

with $\Delta t = \gamma^2$

$$dm(t) = dW(t)$$

Brownian motion

How big $\gamma^2 = \Delta t$ can you take?!

Shown: $\sigma^2 = \frac{\ell}{\quad}$ as $n \to \infty$

$$\sigma = \frac{\ell}{n} \quad \text{as } n \to \infty$$

dimension of m

In order for acceptance rate to be bounded away from

Can we improve $\sigma^2 = \Delta t$? $1/\infty$

$$dx(t) = \frac{1}{2} \nabla \log (\pi(x)) + dW(t)$$

(→ Langevin eqn.)

drift term

discretize using forward Euler $\Delta t = \sigma^2$

$$p = m + \frac{\sigma^2}{2} \nabla \log (\pi(m)) + \sigma N(0,I)$$

$$q(p) = N\left(m + \frac{\sigma^2}{2} \nabla \log (\pi(m)),\ \sigma^2 \right)$$

$$\Delta t = \frac{\ell}{n^{1/3}} \quad \text{as } n \to \infty$$

why Langevin is better?

$$\pi(m) = \exp(-J(m))$$

$$p = m - \underbrace{\frac{\sigma^2}{2} \nabla J(m)} + \sigma N(0, I)$$

$\Downarrow$ Newton: stochastic

$$p = m - \frac{\sigma^2}{2} \left(\nabla^2 J(m)\right)^{-1} \nabla J(m) + \sigma (N, [\nabla^2 J(m)]^{-1})$$

$\Downarrow$

Riemannian Hamiltonian MCMC

uses third order derivative

much

How to compute derivatives

$$\nabla J(m) \ ??$$

$$\nabla J = \nabla \left(\frac{1}{2} \| y^{ob} - h(m) \|^2 + \dots \right)$$

$$h = \int v \, dx$$

s.t: v statisfies a PDE(m)

$\Downarrow$

<u>chain rule</u>

$$J(\cdot,\cdot) : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$$

$$C(\cdot,\cdot) : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^r$$

$$\min J(x,m)$$

$$\text{s.t } (*) \; C(x,m) = 0$$

chain rule:

$$\frac{\partial J}{\partial m_i} = \frac{\partial J}{\partial x}\frac{\partial x}{\partial m_i} + \frac{\partial J}{\partial m_i}$$

taking derivative of $(*)$:

$$\frac{\partial c}{\partial x}\frac{\partial x}{\partial \alpha_i} + \frac{\partial c}{\partial \alpha_i} = 0$$

$\Downarrow$

$$\frac{\partial x}{\partial \alpha_i} = -\left(\frac{\partial c}{\partial x}\right)^{-1}\frac{\partial c}{\partial \alpha_i}$$

$$\Rightarrow \frac{\partial J}{\cdot} = -\frac{\partial J}{\cdot}\left(\frac{\partial c}{\cdot}\right)^{-1}\frac{\partial c}{\cdot} + \frac{\partial J}{\cdot}$$

$$\Rightarrow \quad \frac{\partial \hat{J}}{\partial m_i} = -\frac{\partial J}{\partial x}\left(\frac{\partial c}{\partial x}\right)^{-1}\frac{\partial c}{\partial m_i} + \frac{\partial J}{\partial m_i}$$

Direct sensitivity method

Adjoint

$$\frac{\partial \hat{J}}{\partial m_i} = \underbrace{-\frac{\partial J}{\partial x}\left(\frac{\partial c}{\partial x}\right)^{-1}}_{\lambda^T}\frac{\partial c}{\partial m_i} + \frac{\partial J}{\partial m_i}$$

lldey

Λ^T

then

$$\boxed{\left(\frac{\partial c}{\partial x}\right)^T \lambda = \frac{\partial J}{\partial x}^T}$$

adjoint eqn

Steihaug:

$$\nabla J(m) = 0$$

$\Downarrow$ Newton

$$\nabla^2 J(m_k)\,\Delta m = -\nabla J(m_k)$$

$$\nabla^2 J(m_k)\, d = \underline{\text{adjoint}}$$

solve 1 incremental forward PDE

+ 1 adjoint PDE

Bayesian in function space

Recall

$$\pi_{post}(m \mid y^{obs}) \propto \pi_{like}(y^{obs} \mid m)\, \pi_{prior}(m)$$

$\Downarrow$

valid if $m \in L^2(m)$

No: $\pi_{prior}(m)$ is the density of the prior distribution w.r.t Lebesgue measure

$\Downarrow$

no lebesgue measure

What is Bayes' in ∞.

Recall Radon - Nikodym:

if $\mu \ll \nu$

absolutely continuous

probability mean

$\exists \; f(m) \geq 0$ such that

$$\mu(A) = \int_A f(m)\, d\nu$$

$\Downarrow$ Radon-Nikodym:

$$\frac{d\mu}{d\nu} = f(m) \quad (\underline{a.s})$$

---

recall

$$\pi_{post}(m) = \pi_{like}(y_{obs} \mid m)\, \pi_{prior}(m)$$

$$\mu_{post}(A) = \int_A \pi_{post}(m)\, dm = \int_A \pi_{like}\, \underbrace{\pi_{prior}\, dm}_{\parallel \; d\mu_{prior}(m)}$$

$\Downarrow$

$$\mu_{post}(A) = \int_A \pi_{like}\, d\mu_{prior}(m)$$

$\Downarrow$ Radon – Nikodym

$$\frac{d\mu_{post}}{d\mu_{prior}} = \pi_{like}(y^{obs}/m)$$

SUMMARY:

$$y^{obs} = h(m, u) + e$$

m: unknown

PDE constraints $C(u, m) = 0$

1) Deterministic approach

$$\min_m \tilde{J} = \frac{1}{2\sigma^2} \| y^{obs} - h(m,u) \|^2 + \frac{k}{2} \| R^{-1/2} m \|^2$$

s.t.: $C(m, u) = 0$

$\Downarrow$

Tikhonov,

2) To account for noise, we probability

$\Downarrow$ Bayes

$$\pi_{post}(m|y^{obs}) \propto \pi_{like}(y^{obs}|m)\, \pi_{prior}(m)$$

$\Downarrow$

$$\pi_{post}(m) \propto \exp\left(-\tilde{J}\right)$$

MAP = deterministic solution

$\Updownarrow$ true

prior = Gaussian, $\pi_{prior} \propto \exp\left(-\frac{1}{2}\|R^{-1/2} m\|^2\right)$

Noise = additive + Gaussian

$\Downarrow$ explore posterior prediction + UQ

$h(m, u)$ = linear

prior = Gaussian

likelihood = Gauss + additive

YES $\leftarrow$

π_{post} is Gaussian

$\Downarrow$

Mean = MAP

+

covariance

NO $\Downarrow$

we need to compute

$\mathbb{E}_{\pi_{post}}[m]$

$\mathbb{V}_{\pi_{post}}[m]$

$\Downarrow$

Can we draw I.I.D samples from π_{post}

YES

LLN + CLT

NO

MCMC

Convergence of MCMC

$$\| S_N - \bar{m} \| = O\left(\frac{\sigma}{\sqrt{N/k}} \right)$$

where

N: # MCMC samples

k: for every k samples you get 1 i.i.d sample

RWMH

||

forward Euler of S.D.E

MCMC on function spaces

+

implicit discretization

forward Euler

of S.D.E

$\Delta t = \gamma^2 = O\left(\frac{1}{n}\right)$

as $n \to \infty$

optimization

Langevin

$\Delta t = \gamma^2 = O\left(\frac{1}{n^{1/3}}\right)$

stochastic Newton

$\Delta t = \gamma^2 = O\left(\frac{1}{n^{1/3}}\right)$

HMC

Third order derivative

$\Delta t = \gamma^2 = O\left(\frac{1}{n^{1/4}}\right)$

implicit discretization

S.D.E

Function space MCMC

method (Stuart et al)

$\Delta t \neq f(n)$