Robust methods for Geophysical Imaging and Kalman smoothing

Aleksandr Y. Aravkin
Earth and Ocean Sciences & Computer Science
University of British Columbia
saravkin@eos.ubc.ca

Kalman smoothing
James V. Burke (UW)
Bradley M. Bell (UW)
Gianluigi Pillonetto (Padova)

Geophysical Imaging
Tristan van Leeuwen (UBC)
Michael P. Friedlander (UBC)
Felix J. Herrmann (UBC)
Henri Calandra (Total, Pau)

WID-DOW, March 5, 2012

Robust methods for Geophysical Imaging and Kalman smoothing

Aleksandr Y. Aravkin Earth and Ocean Sciences & Computer Science University of British Columbia saravkin@eos.ubc.ca

Kalman smoothing Geophysical Imaging James V. Burke (UW) Tristan van Leeuwen (UBC) Bradley M. Bell (UW) Michael P. Friedlander (UBC) Gianluigi Pillonetto (Padova) Felix J. Herrmann (UBC)

Henri Calandra (Total, Pau)

WID-DOW, March 5, 2012

A general statistical framework for the study of inverse problems

Geophysical Imaging Kalman smoothing

Robust methods

Densities, penalties, and influence functions Common convex penalties: i

2

, i

1

, Vapnik, Huber PLQ (log-concave) densities and their dual representation Beyond log-concave: heavy tailed densities

A general statistical framework for the study of inverse problems

Geophysical Imaging Kalman smoothing

Robust methods

Densities, penalties, and influence functions Common convex penalties: i

2

, i

1

, Vapnik, Huber PLQ (log-concave) densities and their dual representation Beyond log-concave: heavy tailed densities

Robust seismic imaging

Student’s t formulation Fast optimization by sampling sources Numerical results (UBC and Total)

A general statistical framework for the study of inverse problems

Geophysical Imaging Kalman smoothing

Robust methods

Densities, penalties, and influence functions Common convex penalties: i

2

, i

1

, Vapnik, Huber PLQ (log-concave) densities and their dual representation Beyond log-concave: heavy tailed densities

Robust seismic imaging

Student's t formulation Fast optimization by sampling sources Numerical results (UBC and Total)

Robust Kalman smoothers

PLQ Kalman smoothers using interior point methods Student's t smothers Numerical results

Geophysical Inverse Problems

Data are acquired in the field using thousands of source experiments. For every source experiment, time traces are obtained by an array of receivers. Seismic surveys take months to complete, and full data volume is on the order of terabytes.

After collecting the data, we want to infer the sound velocity parameters on 2D or 3D grid. Seismic images can be used for exploration, monitoring, sequestration, policy decisions, etc. The smallest (academic) seismic 2D model is on a grid of 1000 × 1000, so we have 106 unknowns.

0.5

1) m K (h t p e D

1.5

2

2.5

3

0

1 2 3 4 5 6 7 8 9

] m k [z

1

2

0 2 4 6

Lateral distance (Km)

x [km]

5500

5000

4500

4000

3500

3000

2500

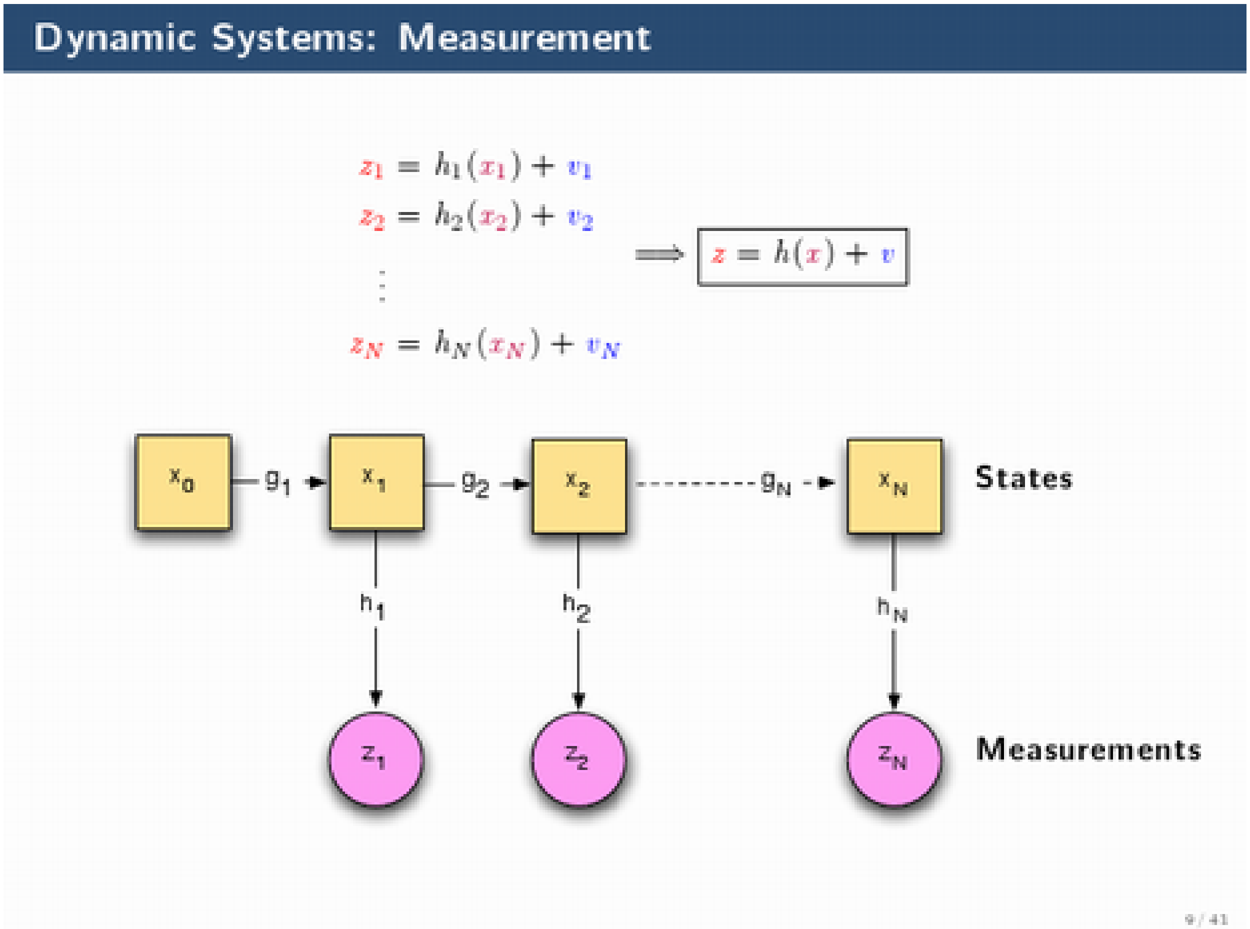
2000

4

3

2

] s / m k [



$z_1 = h_1(x_1) + v_1$ $z_2 = h_2(x_2) + v_2 \ldots z$

N

=⇒ z = h(x) + v

= h

N

(x

N

) + v

N

g X

0

1

Z

1

X

1

h

1

g

2

h

2

Z

2

X

2

g

N

X

N

h

N

Z

N

States

Measurements

Application Stat model Assumptions Inverse Problem

Seismic Imaging Kalman

$z = h(x) + v$ $v \sim N(0,\sigma^2 I)$ $\min_x \|z - h(x)\|_2^2$

Smoothing

$\|\mu - g(x)\|^2_{Q^{-1}} + \|z - h(x)\|^2_{R^{-1}}$

The model

$\mu = g(x) - w$

$\min_x$ $z = h(x) + v$

$w \sim N(0,Q)$ $v \sim N(0,R)$

$$\begin{bmatrix} \mu = g(x) - w \\ z = h(x) + v \end{bmatrix}$$

is very general.

We discuss density modeling for v,w, corresponding inverse problems, and present results for the motivating applications.

Two major themes:

Density modeling Exploiting application structure

Robust Methods Via Error Modeling

Goal: good results in the face of large measurement errors and artifacts in the data unexplained by the model h. We can design robust methods by changing the statistical model for v.

Goal: good results in the face of large measurement errors and artifacts in the data unexplained by the model h. We can design robust methods by changing the statistical model for v.

Gaussian

2 Laplace

Student’s t

y

x L

2

V (x) = 1x2 1 2

2

= sup u∈R x

(u, x) −

1 2

u

2

y

x L

2

V (x) = 1x2 1 2

x

2

= u∈R sup

(u, x) −

1 2

u

2

y

x i

1

V (x) = |x| |x| = sup

u∈[−1,1]

(u,x)

y

x L

2

V (x) = 1x2 1 2

x

2

= u∈R sup

(u, x) −

1 2

u

2

y

x

V (x) = |x|

−κ κ

x

i

1

Huber

V (x) = −κx − 1κ2; x < −κ

|x| = sup

ρ

H u∈[−1,1]

(u,x)

(x) = sup

u∈[−k,k]

1

2 2

12 / 41 (u, x) −

u

y

x

V (x) = 1x2

y

L

2

−∈ Vapnik

∈

x

V (x) = −x − ∈; x < −∈

1

ρ

V 2

〈[

u

1

[

−*ε*

]〉

u

2

−*ε*

−κ κ

]

,

[

l −l

]

x

2

= u∈R sup

(u, x) −

1 2

u

2

(x) = sup

u

1

x +

y

x

V (x) = |x|

,u

2

∈[0,1]

x

i

1

Huber

V (x) = −κx − 1κ2; x < −κ

|x| = sup

ρ

H u∈*[−1,1]*

(u,x)

(x) = sup

u∈*[−k,k]*

(u, x) −

1 2

u

2

Li

Define $\rho(U,M,b,B;\cdot) : R^n \to R$ as

$\rho(U,M,b,B;y) = \sup_{u \in U}$

{

$(u,b + By) -$

}

1 2

(u,Mu)

$U \subset R^m$ is a nonempty polyhedral set

$M \in R^{m \times m}$ is a symmetric positive semidefinite matrix

$b + By$ is an injective affine transformation with $B \in R^{m \times n}$.

The L

2

, l

1

, Huber, and Vapnik are all special cases. Class is closed under summation. Even though some PLQ penalties are non-smooth, all can be solved using a smooth reformulation (efficiently, for Kalman smoothing).

Robust Imaging

- Earth does not follow the acoustic model, and this drives the development of sophisticated modeling: elastic, anisotropic, ...
- Instead, **knowing the predictive model is wrong**, we try to recover the velocity consistent with the portion of data that we can explain.
- **robust Student's t** formulation for seismic inversion:

$$\min_x f(x) := \sum_{q,\omega,i} \log(k + (z^i_{q,\omega} - h^i_{q,\omega}(x))^2)$$

- It is **very easy** to modify existing first order methods to solve it.

$$\nabla f(x) := 2 \sum_{q,\omega,i} \frac{\nabla h^i_{q,\omega}(x)(z^i_{q,\omega} - h^i_{q,\omega}(x))}{k + (z^i_{q,\omega} - h^i_{q,\omega}(x))^2}$$

- But the least squares structure is now gone!

17 / 41

Earth does not follow the acoustic model, and this drives the development of sophisticated modeling: elastic, anisotropic, ... Instead, knowing the predictive model is wrong, we try to recover the velocity consistent with the portion of data that we can explain. robust Student's min

x

f(x) t formulation :=

∑

log(k for + seismic (zi q,ω

− inversion:

hi

q,ω

(x))2)

q,ω,i

It is very easy to modify existing first order methods to solve it.

∇f(x) := 2

∑

q,ω,i

∇h_i

q,ω

$(x)(z_i$

q,ω

$- h_i$

q,ω

$(x)) k + (z_i$

q,ω

$- h_i$

q,ω

$(x))^2$

But the least squares structure is now gone!

LS and Student's t both have form $f(x) = \frac{1}{M}\sum_{i=1}^{M} f_i(x)$. Sampling of the f_i's can be used to approximate $\nabla f(x)$, translating into potentially large savings per iteration.

Friedlander and Schmidt, 2011, suggested slowly increasing the sample size:

$$x^{\nu+1} = x^{\nu} - \eta^{\nu} g^{\nu}$$

where $g^{\nu} = \frac{1}{K}\sum_{j=1}^{K} \nabla f_j(x^{\nu})$, and K increases as we iterate.

LS and Student's t both have form $f(x) = \frac{1}{M}\sum_{i=1}^{M} f_i(x)$. Sampling of the f_i's can be used to approximate $\nabla f(x)$, translating into potentially large savings per iteration.

Friedlander and Schmidt, 2011, suggested slowly increasing the sample size:

$$x^{\nu+1} = x^{\nu} - \eta_{\nu} g^{\nu}$$

where $g^{\nu} = \frac{1}{K}\sum_{j=1}^{K} \nabla f_j(x^{\nu})$, and K increases as we iterate.

For seismic imaging, this translates to taking larger samples of random sequential shots as the iterations proceed.

The proposed strategy allows fast robust imaging, and also fast imaging for marine acquisition.

We have demonstrated 5-fold and 10-fold savings using this approach.

We take 50% of data values and set them to 0. The model does not know which values are ‘wrong’. Large scale data corruption scenario.

0

0.1

5

10

15

20

0 5 10 15 20 25 30

−3 0

0 3

Least-Squares fit is terrible. Note that residual histogram looks Gaussian!

0

0.1

5

10

15

20

0 5 10 15 20 25 30

−3 0

0 3

Laplace fit is better, though still noisy. Note that residual histogram has the Laplace shape.

0

0.1

5

10

15

20

0 5 10 15 20 25 30

−3 0

0 3

Student’s t fit is even better. Note histogram looks nothing like Student’s t, but instead matches the strange residual histogram.

0

0.1

5

10

15

20

0 5 10 15 20 25 30

−3 0

0 3

Looked at partial inversion results at a single frequency (4Hz). Replaced every entry of every 5th shot with the value 1000.

Initial Model

T, bad data

LS, good data

LS, bad data

Robust Kalman Smoothing

g
N
We consider the entire class of PLQ smoothers
]
,
where both w and v PLQ densities. This corresponds to the problem
min x
[z − h(x)] .
where ρ
w
ρ
w
[μ − g(x)] + ρ
v
and ρ
v
are (convex) PLQ penalties.

g X
0
1
Z
1
X
1
h
1
g
2
h
2
Z
2
X
2
[

$$\mu = g(x) - w \quad z = h(x) + v$$

X

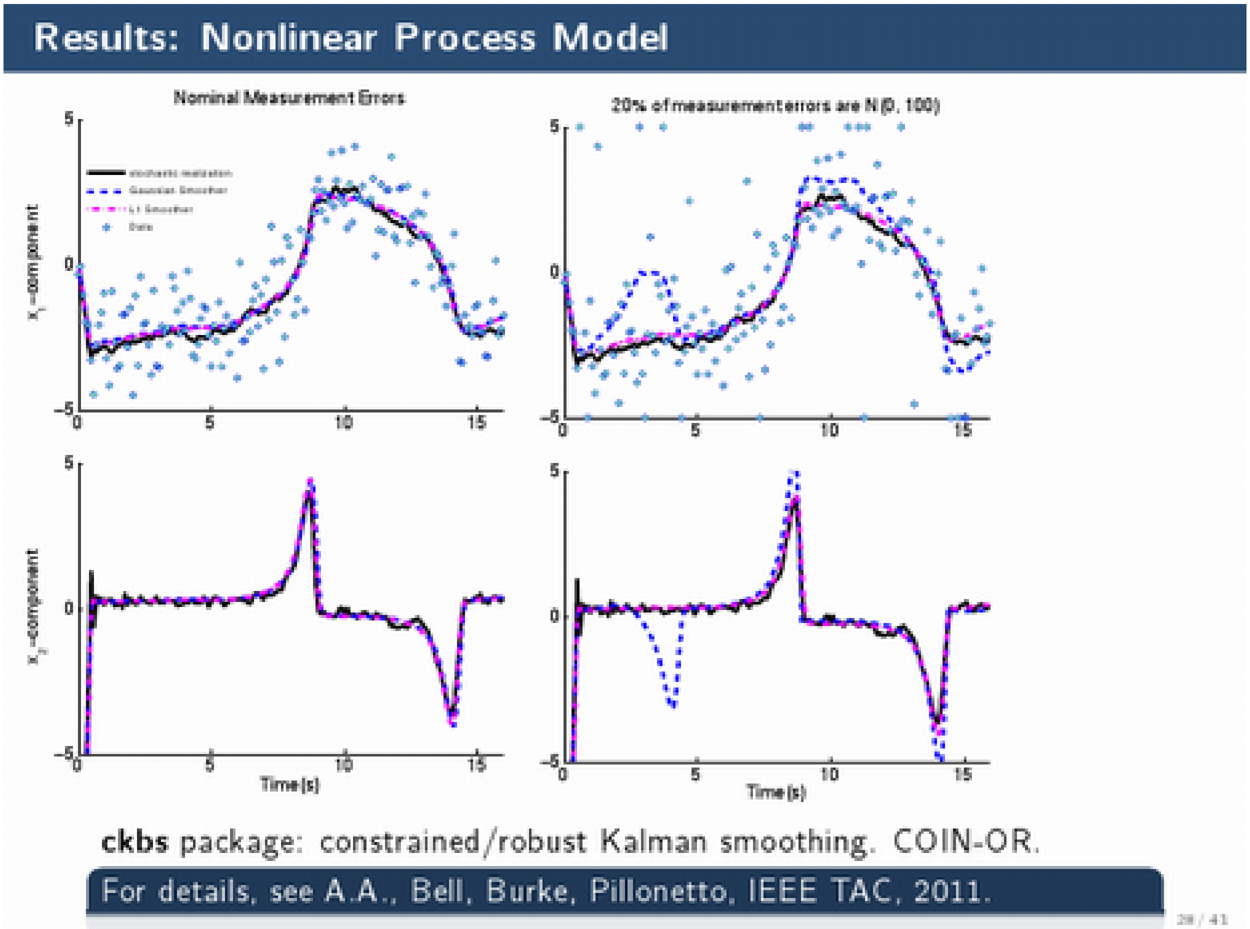
N

h

N

Z

N



5

Nominal Measurement Errors

5

5

2

20% of measurement errors are N(0, 100)

stochastic realization

Gaussian Smoother t n e n o p m o c

L1 Smoother

Data

0

0

0

x

ï5

0 5 10 15

15

ï5

ï5 0

0 5 10 15

5

5

5

0

0

0

x

ï5

0 5 10 Time(s)

15

15

ï5

ï5 0

0 5 10 15

Goal: to recover the function exp(8 sin(t)) from noisy measurements. The unknown function is modeled as integrated brownian noise. For $\Delta t = 1/2000$,

$G_k \begin{bmatrix} 1 & \Delta t \\ 0 & 1 \end{bmatrix} x_{k-1}$

$= \lambda^2 (x_{k-1}) =$

$, Q_k \begin{bmatrix} \Delta t & \Delta t^2/2 \\ \Delta t^2/2 & \Delta t^3/3 \end{bmatrix}$

.

where λ^2 is an unknown scale factor to be estimated from the data by cross-validation (efficiency essential!) Direct observation of function values: $h_k(x_k) = x_{2,k}$

. In the smoother, we model w as Gaussian, and v as Vapnik with unknown ε. Vapnik plays two important roles:

Measurements are contaminated by large N(0,25) outliers and The function we recover has a sparser representation in terms of the data, since only ‘active’ data points are used to evaluate the function.

Functional Kalman Recovery smoother with L

2

loss

Kalman smoother with Vapnik loss

4

L2 Result

1300 support vectors

0 0.2 0.4 0.6 0.8 1

0 0.2 0.4 0.6 0.8 1 Time

4

Vapnik Result

3

3

2

2

1

1

0

0

!1

!1

400 support vectors

!2

!2

Time

Average prediction error on the validation set

1.25

1.248

1.246

1.244

1.242

1.24

1.238

1.236

1

0.8

10000 0.6

8000

0.4

0.2

2000 ε

0

λ

 4000

6000

0

2

We can also model the measurement or process noise using the Student's t distribution. Two interesting and useful smoothers we get this way are

T-Robust:

1 2

[

s

k N∑

(s

k

+ *m*

k

) log

+ ‖*z* − *h(x)*‖2

R

]

+ ‖*μ* − *g(x)*‖2

Q

,

k=1

T-Trend:

−1 k

−1 k

1 2

[

r

k

] N∑

(r

k

+ n) log

k=1

+ ‖*μ* − *g(x)*‖2

Q

k −1

+ ‖*z* − *h(x)*‖2

R

−1 k

.

We can optimize these objectives, exploiting what we can of the convex-composite structure, and preserving the computational properties per iteration.

5

5

0

0

ï5

0 5 10 15

ï5

0 5 10 15

T Kalman smoother under nominal (top) and perturbed (bottom) conditions.

Nominal conditions 2

1.5

1

0.5

0

ï0.5

ï1

ï1.5

1 2 3 4 5 6

Perturbed conditions 12

10 s t i n u n o i t

1 2 3 4 5 6

Time 8

6

c n u

4

F

2

0

ï2

truction of a sudden change in state: results from a Monte Carlo run under nominal (top) and perturbed (bot

Work presented:

Robust modeling: PLQ densities, Student's t. Seismic imaging, dimensionality reduction techniques. Kalman smoothing, special block tridiagonal structure.

Work presented:

Robust modeling: PLQ densities, Student’s t. Seismic imaging, dimensionality reduction techniques. Kalman smoothing, special block tridiagonal structure. Related ongoing work:

Sparsity promotion in geophysical imaging (Herrmann, Li, van Leeuwen) Nonconvex approaches to group sparsity (Chiuso, Pillonetto, Burke) Variance/meta-parameter estimation (van Leeuwen) Robust & sparse inversion (Burke, Friedlander). Robust 3D reconstruction (NASA Ames).

Work presented:

Robust modeling: PLQ densities, Student's t. Seismic imaging, dimensionality reduction techniques. Kalman smoothing, special block tridiagonal structure. Related ongoing work:

Sparsity promotion in geophysical imaging (Herrmann, Li, van Leeuwen) Nonconvex approaches to group sparsity (Chiuso, Pillonetto, Burke) Variance/meta-parameter estimation (van Leeuwen) Robust & sparse inversion (Burke, Friedlander). Robust 3D reconstruction (NASA Ames). Future plans:

Continue ongoing work in geophysical imaging, (SLIM, TOTAL, ...) Strengthen connections between SLIM and UW-Madison (exploiting low-rank structure in geophysics) Work on problems in medical imaging (robust & sparse algorithms) Build and explore Kalman smoothing connections: PK/PD, weather prediction, robust online algorithms, large systems

Interior Point Methods

Define the relaxed optimality conditions F_γ by

$$F_\gamma(s, q, u, y) = \begin{bmatrix} s + A^T u - a \\ D(q)D(s)\mathbf{1} - \gamma \mathbf{1} \\ By - Mu - Aq + b \\ B^T u \end{bmatrix},$$

Note that F_0 is equivalent to the original optimality conditions.

- Interior point methods solve for an update of s, q, u, y by forcing $F_\gamma = 0$ using Newton iterations. γ is forced to 0, so final solution satisfies optimality criteria.
- Under some assumptions, these methods work on the entire PLQ class of problems. In practice, no more than 10-20 IP iterations are required to converge.
- For Kalman smoothing, IP iterations **preserve the structure and efficiency of the Gaussian case**.

See A., Burke, Pillonetto, 2012 for details.

Define the relaxed optimality conditions F
γ
by
F
γ
(s,q,u,y) =

By D(q)D(s)1 s − + Mu ATu − Aq − −
a
γ1 + b

,
BTu
Note that F

0

is equivalent to the original optimality conditions.

Interior point methods solve for an update of s,q,u,y by forcing F

γ

= 0 using Newton iterations. γ is forced to 0, so final solution satisfies optimality criteria. Under some assumptions, these methods work on the entire PLQ class of problems. In practice, no more than 10-20 IP iterations are required to converge. For Kalman smoothing, IP iterations preserve the structure and efficiency of the Gaussian case.

Application Stat model Assumptions Inverse Problem

Seismic Imaging Kalman

$z = h(x) + v$ $v \sim N(0,\sigma^2 I)$ min

x

$\|z - h(x)\|_2^2$

Smoothing

$\mu = g(x) - w$ $z = h(x) + v$

min x

$\|\mu - g(x)\|^2$

$Q^{-1} + \|z - h(x)\|^2$

R^{-1}

We introduced an entire (log-concave) class of PLQ densities for v,w, showing how to build densities with prescribed moments from small building blocks, and discussed heavy-tailed distributions. For Seismic imaging, we demonstrated how a robust Student's t approach can yield results superior to those given by convex methods in situations with extreme outliers. We also showed how to use sampling to cut down the computation. For Kalman smoothing, we showed how interior point methods can be used for a large class of PLQ smoothing formulations. We also discussed convex-composite structure, allowing us to consider nonlinear applications.

 $w \sim N(0,Q)$ $v \sim N(0,R)$