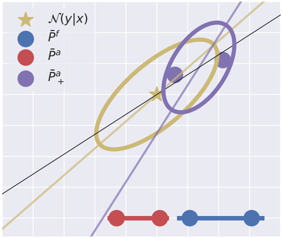
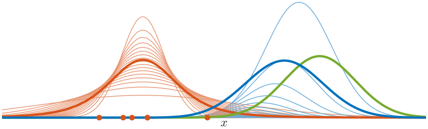
EnKF – FAQ

(Ensemble Kalman filter – Frequently asked questions)





Patrick N. Raanes, Geir Evensen, Andreas S. Stordal

Marc Bocquet, Alberto Carrassi

NORCE DIGIRES NERSC EmblA NordForsk cerea École des Ponts ParisTech EDF

Leeds, May 16, 2019

Revising the stochastic iterative ensemble smoother

Patrick N. Raanes*[1,2], Geir Evensen[1,2], and Andreas S. Stordal[1]

[1]NORCE, Pb. 22 Nygårdstangen, 5838 Bergen, Norway
[2]NERSC, Thormøhlens gate 47, 5006 Bergen, Norway

February 4, 2019

Abstract

Ensemble randomized maximum likelihood (EnRML) is an iterative (stochastic) ensemble smoother, used for large and nonlinear inverse problems, such as history matching and data assimilation. Its current formulation is overly complicated and has issues with computational costs, noise, and covariance localization, even causing some practitioners to omit crucial prior information. This paper resolves these difficulties and streamlines the algorithm, without changing its output. These simplifications are achieved through the careful treatment of the linearizations and subspaces. For example, it is shown (a) how ensemble linearizations relate to average sensitivity, and (b) that the ensemble does not loose rank during updates. The paper also draws significantly on the theory of the (deterministic) iterative ensemble Kalman smoother (IEnKS). Comparative benchmarks are obtained with the Lorenz-96 model with these two smoothers and the ensemble smoother using multiple data assimilation (ES-MDA).

1 Introduction

Ensemble (Kalman) smoothers are approximate methods used for data assimilation (state estimation in geoscience), history matching (parameter estimation for reservoirs), and other inverse problems constrained by partial differential equations. Iterative versions of the ensemble smoother, derived from optimization perspectives, have proven useful in improving the estimation accuracy when the forward operator is nonlinear. Ensemble randomized maximum likelihood (EnRML), also known as the iterative ensemble smoother (IES), is one such method. This paper fixes several issues with EnRML, described in the following. *Readers unfamiliar with EnRML may jump to the beginning of the derivation:*

linearization $\overline{\mathbf{M}}_i$ only appears in the product $\overline{\mathbf{M}}_i \overline{\mathbf{C}}_{x,i} \overline{\mathbf{M}}_i^\top$, which does not require $\mathbf{X}_i^+$. For the *prior increment*, on the other hand, the modification breaks its Kalman gain form. Meanwhile, the precision matrix form, i.e. their equation 10, is already invalid because it requires the inverse of $\overline{\mathbf{C}}_{x,i}$. Still, in their equation 15, the prior increment is formulated with an inversion in ensemble space, and also unburdened of the explicit computation of $\overline{\mathbf{M}}_i$. Intermediate explanations are lacking, but could be construed to involve approximate inversions. Another issue is that the pseudo-inverse of $\overline{\mathbf{C}}_x$ is now required (via $\mathbf{X}$), and covariance localization is further complicated.

An approximate version was therefore also proposed, where the prior mismatch term is omitted from the update formula altogether. This is not principled, and severely

Revising the stochastic iterative ensemble smoother

Patrick N. Raanes*[1,2], Geir Evensen[1,2], and Andreas S. Stordal[1]

[1]NORCE, Pb. 22 Nygårdstangen, 5838 Bergen, Norway
[2]NERSC, Thormøhlens gate 47, 5006 Bergen, Norway

February 4, 2019

Abstract

Ensemble randomized maximum likelihood (EnRML) is an iterative (stochastic) ensemble smoother, used for large and nonlinear inverse problems, such as history matching and data assimilation. Its current formulation is overly complicated and has issues with computational costs, noise, and covariance localization, even causing some practitioners to omit crucial prior information. This paper resolves these difficulties and streamlines the algorithm, without changing its output. These simplifications are achieved through the careful treatment of the linearizations and subspaces. For example, it is shown (a) how ensemble linearizations relate to average sensitivity, and (b) that the ensemble does not loose rank during updates. The paper also draws significantly on the theory of the (deterministic) iterative ensemble Kalman smoother (IEnKS). Comparative benchmarks are obtained with the Lorenz-96 model with these two smoothers and the ensemble smoother using multiple data assimilation (ES-MDA).

Also answered these questions about the EnKF:

1 Introduction

Ensemble (Kalman) smoothers are approximate methods used for data assimilation (state estimation in geoscience), history matching (parameter estimation for reservoirs), and other inverse problems constrained by partial differential equations. Iterative versions of the ensemble smoother, derived from optimization perspectives, have proven useful in improving the estimation accuracy when the forward operator is nonlinear. Ensemble randomized maximum likelihood (EnRML), also known as the iterative ensemble smoother (IES), is one such method. This paper fixes several issues with EnRML, described in the following. *Readers unfamiliar with EnRML may jump to the beginning of the derivation*

linearization $\bar{\mathbf{M}}_i$ only appears in the product $\bar{\mathbf{M}}_i \bar{\mathbf{C}}_{x,i} \bar{\mathbf{M}}_i^\top$, which does not require $\mathbf{X}_i^+$. For the *prior increment*, on the other hand, the modification breaks its Kalman gain form. Meanwhile, the precision matrix form, i.e. their equation 10, is already invalid because it requires the inverse of $\bar{\mathbf{C}}_{x,i}$. Still, in their equation 15, the prior increment is formulated with an inversion in ensemble space, and also unburdened of the explicit computation of $\bar{\mathbf{M}}_i$. Intermediate explanations are lacking, but could be construed to involve approximate inversions. Another issue is that the pseudo-inverse of $\bar{\mathbf{C}}_x$ is now required (via X), and covariance localization is further complicated.

An approximate version was therefore also proposed, where the prior mismatch term is omitted from the update formula altogether.

Revising the stochastic iterative ensemble smoother

Patrick N. Raanes*[1,2], Geir Evensen[1,2], and Andreas S. Stordal[1]

[1]NORCE, Pb. 22 Nygårdstangen, 5838 Bergen, Norway
[2]NERSC, Thormøhlens gate 47, 5006 Bergen, Norway

February 4, 2019

Abstract

Ensemble randomized maximum likelihood (EnRML) is an iterative (stochastic) ensemble smoother, used for large and nonlinear inverse problems, such as history matching and data assimilation. Its current formulation is overly complicated and has issues with computational costs, noise, and covariance localization, even causing some practitioners to omit crucial prior information. This paper resolves these difficulties and streamlines the algorithm, without changing its output. These simplifications are achieved through the careful treatment of the linearizations and subspaces. For example, it is shown (a) how ensemble linearizations relate to average sensitivity, and (b) that the ensemble does not loose rank during updates. The paper also draws significantly on the theory of the (deterministic) iterative ensemble Kalman smoother (IEnKS). Comparative benchmarks are obtained with the Lorenz-96 model with these two smoothers and the ensemble smoother using multiple data assimilation (ES-MDA).

Also answered these questions about the EnKF:

- About ensemble linearizations:
  - What exactly are they?
  - Why does this rarely get mentioned?
  - How does it relate to analytic derivatives?

Revising the stochastic iterative ensemble smoother

Patrick N. Raanes*[1,2], Geir Evensen[1,2], and Andreas S. Stordal[1]

[1]NORCE, Pb. 22 Nygårdstangen, 5838 Bergen, Norway
[2]NERSC, Thormøhlens gate 47, 5006 Bergen, Norway

February 4, 2019

Abstract

Ensemble randomized maximum likelihood (EnRML) is an iterative (stochastic) ensemble smoother, used for large and nonlinear inverse problems, such as history matching and data assimilation. Its current formulation is overly complicated and has issues with computational costs, noise, and covariance localization, even causing some practitioners to omit crucial prior information. This paper resolves these difficulties and streamlines the algorithm, without changing its output. These simplifications are achieved through the careful treatment of the linearizations and subspaces. For example, it is shown (a) how ensemble linearizations relate to average sensitivity, and (b) that the ensemble does not loose rank during updates. The paper also draws significantly on the theory of the (deterministic) iterative ensemble Kalman smoother (IEnKS). Comparative benchmarks are obtained with the Lorenz-96 model with these two smoothers and the ensemble smoother using multiple data assimilation (ES-MDA).

Also answered these questions about the EnKF:

- About ensemble linearizations:
  - What exactly are they?
  - Why does this rarely get mentioned?
  - How does it relate to analytic derivatives?
- Why do we prefer the Kalman gain "form"?

Received: 25 January 2018 | Revised: 03 July 2018 | Accepted: 18 July 2018

DOI: 10.1002/qj.3386

RESEARCH ARTICLE

Quarterly Journal of the Royal Meteorological Society RMetS

Adaptive covariance inflation in the ensemble Kalman filter by Gaussian scale mixtures

Patrick N. Raanes[1] | Marc Bocquet[2] | Alberto Carrassi[1]

[1]Nansen Environmental and Remote Sensing Center, Bergen, Norway

[2]CEREA, Joint laboratory of École des Ponts ParisTech and EDF R&D, Université Paris-Est, Champs-sur-Marne, France

Correspondence

Patrick N. Raanes, NERSC, Thormøhlensgate 47, Bergen N-5006, Norway.

Email: patrick.n.raanes@gmail.com

Funding information

EmblAUS project of the Nordic countries funding agency NordForsk, by DIGIRES sponsored by PETROMAKS2 of the Research Council of Norway and industry partners, and project REDDA of the Norwegian Research Council.,

This paper studies multiplicative inflation: the complementary scaling of the state covariance in the ensemble Kalman filter (EnKF). Firstly, error sources in the EnKF are catalogued and discussed in relation to inflation; nonlinearity is given particular attention as a source of sampling error. In response, the "finite-size" refinement known as the EnKF-*N* is re-derived via a Gaussian scale mixture, again demonstrating how it yields adaptive inflation. Existing methods for adaptive inflation estimation are reviewed, and several insights are gained from a comparative analysis. One such adaptive inflation method is selected to complement the EnKF-*N* to make a hybrid that is suitable for contexts where model error is present and imperfectly parametrized. Benchmarks are obtained from experiments with the two-scale Lorenz model and its slow-scale truncation. The proposed hybrid EnKF-*N* method of adaptive inflation is found to yield systematic accuracy improvements in comparison with the existing methods, albeit to a moderate degree.

KEYWORDS

adaptive filtering, Bayesian inference, covariance inflation, data assimilation, ensemble Kalman filter (EnKF), scale mixture

1 | INTRODUCTION

The EnKF computes an ensemble of N realizations, or

Received: 25 January 2018 | Revised: 03 July 2018 | Accepted: 18 July 2018

DOI: 10.1002/qj.3386

RESEARCH ARTICLE

Quarterly Journal of the Royal Meteorological Society RMetS

Adaptive covariance inflation in the ensemble Kalman filter by Gaussian scale mixtures

Patrick N. Raanes[1] | Marc Bocquet[2] | Alberto Carrassi[1]

[1]Nansen Environmental and Remote Sensing Center, Bergen, Norway

[2]CEREA, Joint laboratory of École des Ponts ParisTech and EDF R&D, Université Paris-Est, Champs-sur-Marne, France

Correspondence
Patrick N. Raanes, NERSC, Thormøhlensgate 47, Bergen N-5006, Norway.
Email: patrick.n.raanes@gmail.com

Funding information
EmblAUS project of the Nordic countries funding agency NordForsk, by DIGIRES sponsored by PETROMAKS2 of the Research Council of Norway and industry partners, and project REDDA of the Norwegian Research Council.,

This paper studies multiplicative inflation: the complementary scaling of the state covariance in the ensemble Kalman filter (EnKF). Firstly, error sources in the EnKF are catalogued and discussed in relation to inflation; nonlinearity is given particular attention as a source of sampling error. In response, the "finite-size" refinement known as the EnKF-N is re-derived via a Gaussian scale mixture, again demonstrating how it yields adaptive inflation. Existing methods for adaptive inflation estimation are reviewed, and several insights are gained from a comparative analysis. One such adaptive inflation method is selected to complement the EnKF-N to make a hybrid that is suitable for contexts where model error is present and imperfectly parametrized. Benchmarks are obtained from experiments with the two-scale Lorenz model and its slow-scale truncation. The proposed hybrid EnKF-N method of adaptive inflation is found to yield systematic accuracy improvements in comparison with the existing methods, albeit to a moderate degree.

KEYWORDS
adaptive filtering, Bayesian inference, covariance inflation, data assimilation, ensemble Kalman filter (EnKF), scale mixture

Also answered these questions about the EnKF:

Received: 25 January 2018 | Revised: 03 July 2018 | Accepted: 18 July 2018
DOI: 10.1002/qj.3386

RESEARCH ARTICLE

Quarterly Journal of the Royal Meteorological Society RMetS

Adaptive covariance inflation in the ensemble Kalman filter by Gaussian scale mixtures

Patrick N. Raanes[1] | Marc Bocquet[2] | Alberto Carrassi[1]

[1]Nansen Environmental and Remote Sensing Center, Bergen, Norway
[2]CEREA, Joint laboratory of École des Ponts ParisTech and EDF R&D, Université Paris-Est, Champs-sur-Marne, France

Correspondence
Patrick N. Raanes, NERSC, Thormøhlensgate 47, Bergen N-5006, Norway.
Email: patrick.n.raanes@gmail.com

This paper studies multiplicative inflation: the complementary scaling of the state covariance in the ensemble Kalman filter (EnKF). Firstly, error sources in the EnKF are catalogued and discussed in relation to inflation; nonlinearity is given particular attention as a source of sampling error. In response, the "finite-size" refinement known as the EnKF-N is re-derived via a Gaussian scale mixture, again demonstrating how it yields adaptive inflation. Existing methods for adaptive inflation estimation are reviewed, and several insights are gained from a comparative analysis. One such adaptive inflation method is selected to complement the EnKF-N to make a hybrid that is suitable for contexts where model error is present and imperfectly parametrized. Benchmarks are obtained from experiments with the two-scale Lorenz model and its slow-scale truncation. The proposed hybrid EnKF-N method of adaptive inflation is found to yield systematic accuracy improvements in comparison with the existing methods, albeit to a moderate degree.

KEYWORDS
adaptive filtering, Bayesian inference, covariance inflation, data assimilation, ensemble Kalman filter (EnKF), scale mixture

EmblAUS project of the Nordic countries funding agency NordForsk, by DIGIRES sponsored by PETROMAKS ... Norway and industry partners, and project REDDA of the Norwegian Research Council.

Also answered these questions about the EnKF:

- Why do we use $(N-1)$ in $\frac{1}{N-1}\sum_n (x_n - \bar{x})^2$?

Received: 25 January 2018 | Revised: 03 July 2018 | Accepted: 18 July 2018

DOI: 10.1002/qj.3386

RESEARCH ARTICLE

Quarterly Journal of the Royal Meteorological Society RMetS

Adaptive covariance inflation in the ensemble Kalman filter by Gaussian scale mixtures

Patrick N. Raanes[1] | Marc Bocquet[2] | Alberto Carrassi[1]

[1]Nansen Environmental and Remote Sensing Center, Bergen, Norway

[2]CEREA, Joint laboratory of École des Ponts ParisTech and EDF R&D, Université Paris-Est, Champs-sur-Marne, France

Correspondence
Patrick N. Raanes, NERSC, Thormøhlensgate 47, Bergen N-5006, Norway.
Email: patrick.n.raanes@gmail.com

This paper studies multiplicative inflation: the complementary scaling of the state covariance in the ensemble Kalman filter (EnKF). Firstly, error sources in the EnKF are catalogued and discussed in relation to inflation; nonlinearity is given particular attention as a source of sampling error. In response, the "finite-size" refinement known as the EnKF-N is re-derived via a Gaussian scale mixture, again demonstrating how it yields adaptive inflation. Existing methods for adaptive inflation estimation are reviewed, and several insights are gained from a comparative analysis. One such adaptive inflation method is selected to complement the EnKF-N to make a hybrid that is suitable for contexts where model error is present and imperfectly parametrized. Benchmarks are obtained from experiments with the two-scale Lorenz model and its slow-scale truncation. The proposed hybrid EnKF-N method of adaptive inflation is found to yield systematic accuracy improvements in comparison with the existing methods, albeit to a moderate degree.

Also answered these questions about the EnKF:

- Why do we use $(N-1)$ in $\frac{1}{N-1}\sum_n (x_n - \bar{x})^2$?
- About nonlinearity:
  - Why does it create sampling error?
  - Why does it cause divergence?

$\mathcal{N}$ despite nonlinearity

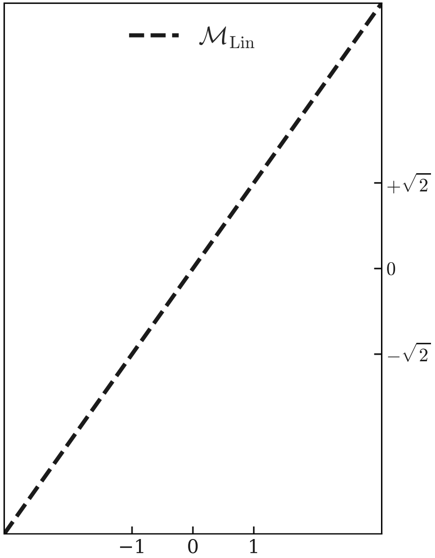
$$\mathcal{M}_{\mathsf{Lin}}(x) = \sqrt{2}x\,,$$

$$\mathcal{M}_{\mathsf{NonLin}}(x) = \sqrt{2}F_{\mathcal{N}}^{-1}\left(F_{\chi}(x^2)\right)$$

$\mathcal{N}$ despite nonlinearity

$$\mathcal{M}_{\mathsf{Lin}}(x) = \sqrt{2}x\,,$$

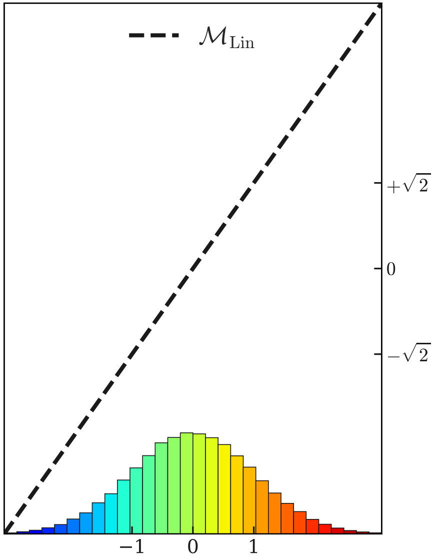
$$\mathcal{M}_{\mathsf{NonLin}}(x) = \sqrt{2}F_{\mathcal{N}}^{-1}\left(F_{\chi}(x^2)\right)$$



$\mathcal{N}$ despite nonlinearity

$$\mathcal{M}_{\mathsf{Lin}}(x) = \sqrt{2}x\,,$$

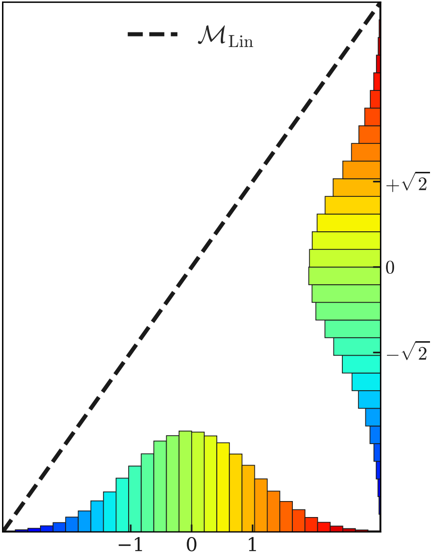
$$\mathcal{M}_{\mathsf{NonLin}}(x) = \sqrt{2}F_{\mathcal{N}}^{-1}\left(F_X(x^2)\right)$$



$\mathcal{N}$ despite nonlinearity

$$\mathcal{M}_{\mathsf{Lin}}(x) = \sqrt{2}x\,,$$

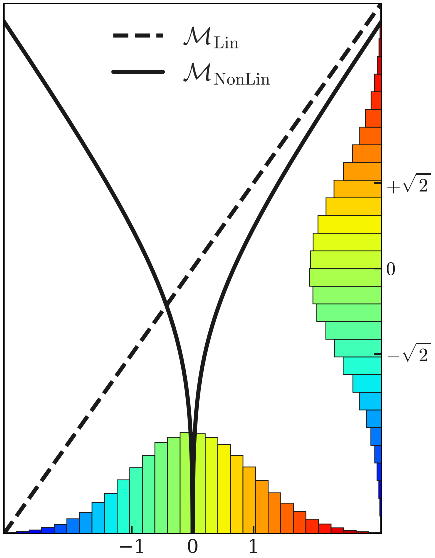
$$\mathcal{M}_{\mathsf{NonLin}}(x) = \sqrt{2}F_{\mathcal{N}}^{-1}\left(F_{\chi}(x^2)\right)$$



$\mathcal{N}$ despite nonlinearity

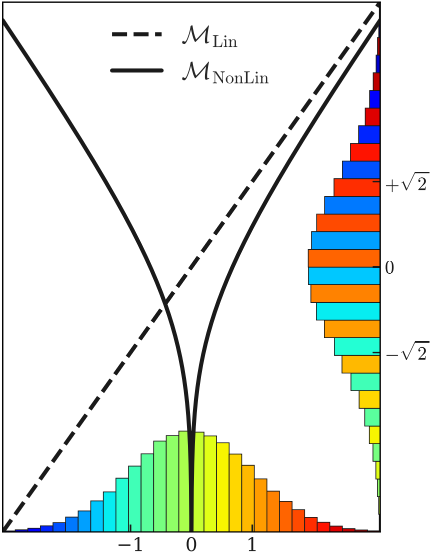
$$\mathcal{M}_{\mathsf{Lin}}(x) = \sqrt{2}x\,,$$

$$\mathcal{M}_{\mathsf{NonLin}}(x) = \sqrt{2}F_{\mathcal{N}}^{-1}(F_{\chi}(x^2))$$



$\mathcal{N}$ despite nonlinearity

$$\mathcal{M}_{\mathsf{Lin}}(x) = \sqrt{2}x\,,$$
$$\mathcal{M}_{\mathsf{NonLin}}(x) = \sqrt{2}F_{\mathcal{N}}^{-1}(F_{\chi}(x^2))$$



Two scalar filtering problems

Consider the problem with:

$$\begin{aligned} \text{prior} &= \mathcal{N}(x|0, \underline{\;2\;}), \\ \text{likelihood} &= \mathcal{N}(0|x, 2), \\ \implies \text{posterior} &= \mathcal{N}(x|\underline{\;0\;}, \underline{\;1\;}). \end{aligned}$$

Now apply a (square-root) EnKF to it.

Two scalar filtering problems

Consider the problem with:

$$\text{prior} = \mathcal{N}(x|0, \underline{\ 2\ }),$$

$$\text{likelihood} = \mathcal{N}(0|x, 2),$$

$$\implies \text{posterior} = \mathcal{N}(x|\underline{\ 0\ }, \underline{\ 1\ }).$$

Now apply a (square-root) EnKF to it.

Two scalar filtering problems

Consider the problem with:

$$\text{prior} = \mathcal{N}(x|0, \underline{\ 2\ }),$$
$$\text{likelihood} = \mathcal{N}(0|x, 2),$$
$$\implies \text{posterior} = \mathcal{N}(x|\underline{\ 0\ }, \underline{\ 1\ }).$$

Now apply a (square-root) EnKF to it.

Two scalar filtering problems

Consider the problem with:

$$\begin{aligned} \text{prior} &= \mathcal{N}(x|0, \underline{\;2\;}), \\ \text{likelihood} &= \mathcal{N}(0|x, 2), \\ \implies \text{posterior} &= \mathcal{N}(x|\underline{\;0\;}, \underline{\;1\;})\,. \end{aligned}$$

Now apply a (square-root) EnKF to it.

Two scalar filtering problems

Consider the problem with:

$$\text{prior} = \mathcal{N}(x|0, \underline{\;2\;}),$$
$$\text{likelihood} = \mathcal{N}(0|x, 2),$$
$$\implies \text{posterior} = \mathcal{N}(x|\underline{\;0\;}, \underline{\;1\;})\,.$$

dynamical model: $\mathcal{M}_{\mathsf{Lin}}(x) = \sqrt{2}x\,.$

Now apply a (square-root) EnKF to it.

Two scalar filtering problems

Consider the problem with:

$$\begin{aligned} \text{prior} &= \mathcal{N}(x|0, \underline{2}), \\ \text{likelihood} &= \mathcal{N}(0|x, 2), \\ \implies \text{posterior} &= \mathcal{N}(x|\underline{0}, \underline{1}). \end{aligned}$$

dynamical model: $\mathcal{M}_{\mathsf{Lin}}(x) = \sqrt{2}x$.

Now apply a (square-root) EnKF to it.

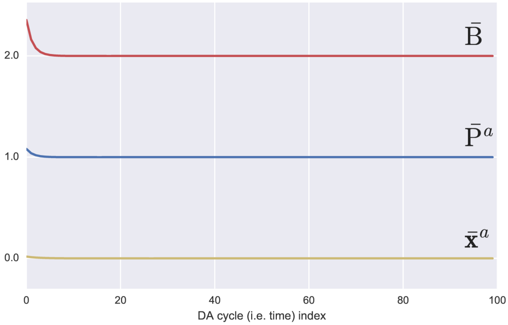
Two scalar filtering problems

Consider the problem with:

$$\text{prior} = \mathcal{N}(x|0, \underline{2}),$$
$$\text{likelihood} = \mathcal{N}(0|x, 2),$$
$$\implies \text{posterior} = \mathcal{N}(x|\underline{0}, \underline{1}).$$

dynamical model: $\mathcal{M}_{\text{Lin}}(x) = \sqrt{2}x$.

Now apply a (square-root) EnKF to it.



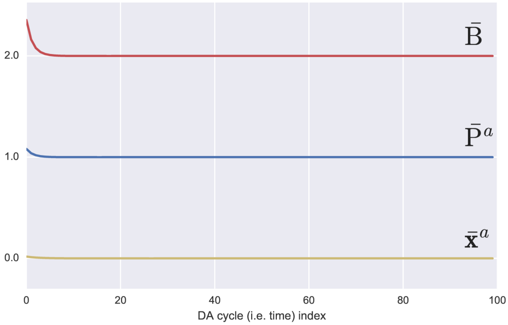
Two scalar filtering problems

Consider the problem with:

$$\text{prior} = \mathcal{N}(x|0, \underline{2}),$$
$$\text{likelihood} = \mathcal{N}(0|x, 2),$$
$$\implies \text{posterior} = \mathcal{N}(x|\underline{0}, \underline{1}).$$

dynamical model: ~~$\mathcal{M}_{\text{Lin}}(x) = \sqrt{2}x$~~ .

Now apply a (square-root) EnKF to it.



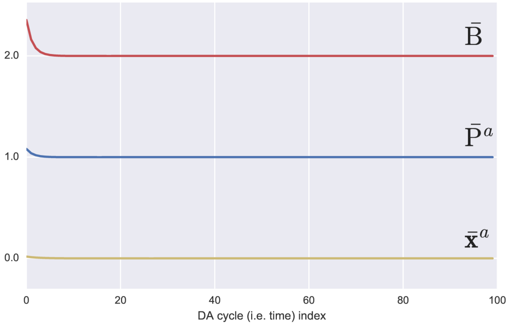
Two scalar filtering problems

Consider the problem with:

$$\text{prior} = \mathcal{N}(x|0, \underline{2}),$$
$$\text{likelihood} = \mathcal{N}(0|x, 2),$$
$$\implies \text{posterior} = \mathcal{N}(x|\underline{0}, \underline{1}).$$

$$\text{dynamical model: } \mathcal{M}_{\text{NonLin}}(x) = \dots .$$

Now apply a (square-root) EnKF to it.



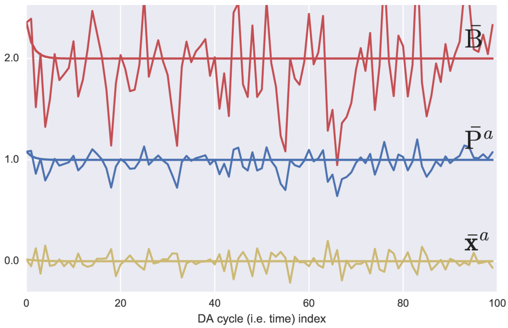
Two scalar filtering problems

Consider the problem with:

$$\text{prior} = \mathcal{N}(x|0, \underline{2}),$$
$$\text{likelihood} = \mathcal{N}(0|x, 2),$$
$$\implies \text{posterior} = \mathcal{N}(x|\underline{0}, \underline{1}).$$

$$\text{dynamical model: } \mathcal{M}_{\text{NonLin}}(x) = \dots .$$

Now apply a (square-root) EnKF to it.



Riccati eqn.

Assume linear $(\mathbf{M})$ dynamics, $\mathbf{Q} = 0$, $\mathbf{H} = \mathbf{I}$, and a deterministic EnKF.

The ensemble covariance obeys:

- the forecast step: $$\bar{\mathbf{B}}_k = \mathbf{M}^2 \bar{\mathbf{P}}_{k-1} \, ; \tag{1}$$
- the analysis step: $$\bar{\mathbf{P}}_k = (\mathbf{I} - \bar{\mathbf{K}}_k)\bar{\mathbf{B}}_k \tag{2}$$

$$\iff \quad \bar{\mathbf{P}}_k^{-1} = \bar{\mathbf{B}}_k^{-1} + \mathbf{R}^{-1} \, . \tag{3}$$

$\Longrightarrow$ the "Riccati recursion":

$$\bar{\mathbf{P}}_k^{-1} = (\mathbf{M}^2 \bar{\mathbf{P}}_{k-1})^{-1} + \mathbf{R}^{-1} \tag{4}$$

Riccati eqn.

Assume linear ($\mathbf{M}$) dynamics, $\mathbf{Q} = 0$, $\mathbf{H} = \mathbf{I}$,
and a deterministic EnKF.

The ensemble covariance obeys:

- the forecast step: $$\bar{\mathbf{B}}_k = \mathbf{M}^2 \bar{\mathbf{P}}_{k-1} \; ; \tag{1}$$
- the analysis step: $$\bar{\mathbf{P}}_k = (\mathbf{I} - \bar{\mathbf{K}}_k)\bar{\mathbf{B}}_k \tag{2}$$

$$\iff \quad \bar{\mathbf{P}}_k^{-1} = \bar{\mathbf{B}}_k^{-1} + \mathbf{R}^{-1} . \tag{3}$$

$\implies$ the "Riccati recursion":

$$\bar{\mathbf{P}}_k^{-1} = (\mathbf{M}^2 \bar{\mathbf{P}}_{k-1})^{-1} + \mathbf{R}^{-1} \tag{4}$$

Riccati eqn.

Assume linear ($\mathbf{M}$) dynamics, $\mathbf{Q} = 0$, $\mathbf{H} = \mathbf{I}$,
and a deterministic EnKF.

The ensemble covariance obeys:

- the forecast step: $$\bar{\mathbf{B}}_k = \mathbf{M}^2 \bar{\mathbf{P}}_{k-1}\,; \tag{1}$$
- the analysis step: $$\bar{\mathbf{P}}_k = (\mathbf{I} - \bar{\mathbf{K}}_k)\bar{\mathbf{B}}_k \tag{2}$$

$$\iff \quad \bar{\mathbf{P}}_k^{-1} = \bar{\mathbf{B}}_k^{-1} + \mathbf{R}^{-1}\,. \tag{3}$$

$\implies$ the "Riccati recursion":

$$\bar{\mathbf{P}}_k^{-1} = (\mathbf{M}^2 \bar{\mathbf{P}}_{k-1})^{-1} + \mathbf{R}^{-1} \tag{4}$$

Riccati eqn.

Assume linear ($\mathbf{M}$) dynamics, $\mathbf{Q} = 0$, $\mathbf{H} = \mathbf{I}$,
and a deterministic EnKF.

The ensemble covariance obeys:

- the forecast step: $$\bar{\mathbf{B}}_k = \mathbf{M}^2 \bar{\mathbf{P}}_{k-1} \,; \tag{1}$$
- the analysis step: $$\bar{\mathbf{P}}_k = (\mathbf{I} - \bar{\mathbf{K}}_k) \bar{\mathbf{B}}_k \tag{2}$$

$$\iff \quad \bar{\mathbf{P}}_k^{-1} = \bar{\mathbf{B}}_k^{-1} + \mathbf{R}^{-1} \,. \tag{3}$$

$\implies$ the "Riccati recursion":

$$\bar{\mathbf{P}}_k^{-1} = (\mathbf{M}^2 \bar{\mathbf{P}}_{k-1})^{-1} + \mathbf{R}^{-1} \tag{4}$$

Deductions from Riccati: Attenuation

Stationary solution:

$$\bar{\mathbf{P}}_\infty^{-1} = (\mathbf{M}^2 \bar{\mathbf{P}}_\infty)^{-1} + \mathbf{R}^{-1} \tag{5}$$

$$\iff \quad \bar{\mathbf{P}}_\infty = \bar{\mathbf{K}}_\infty \mathbf{R} \,, \quad \bar{\mathbf{K}}_\infty = \begin{cases} \mathbf{I} - \mathbf{M}^{-2} & \text{if } \mathbf{M} \geq 1 \,, \\ 0 & \text{otherwise.} \end{cases} \tag{6}$$

Initial conditions (ICs) don't appear
$\implies$ ICs are "forgotten".
$\implies$ Sampling error is attenuated.

Deductions from Riccati: Attenuation

Stationary solution:

$$\bar{\mathbf{P}}_\infty^{-1} = (\mathbf{M}^2 \bar{\mathbf{P}}_\infty)^{-1} + \mathbf{R}^{-1} \tag{5}$$

$$\iff \quad \bar{\mathbf{P}}_\infty = \bar{\mathbf{K}}_\infty \mathbf{R}\,, \quad \bar{\mathbf{K}}_\infty = \begin{cases} \mathbf{I} - \mathbf{M}^{-2} & \text{if } \mathbf{M} \geq 1\,, \\ 0 & \text{otherwise.} \end{cases} \tag{6}$$

Initial conditions (ICs) don't appear

$\implies$ ICs are "forgotten".

$\implies$ Sampling error is attenuated.

Deductions from Riccati: Attenuation

Stationary solution:

$$\bar{\mathbf{P}}_\infty^{-1} = (\mathbf{M}^2 \bar{\mathbf{P}}_\infty)^{-1} + \mathbf{R}^{-1} \tag{5}$$

$$\iff \quad \bar{\mathbf{P}}_\infty = \bar{\mathbf{K}}_\infty \mathbf{R}\,, \quad \bar{\mathbf{K}}_\infty = \begin{cases} \mathbf{I} - \mathbf{M}^{-2} & \text{if } \mathbf{M} \geq 1\,, \\ 0 & \text{otherwise.} \end{cases} \tag{6}$$

Initial conditions (ICs) don't appear
$\implies$ ICs are "forgotten".
$\implies$ Sampling error is attenuated.

Deductions from Riccati: Attenuation

Stationary solution:

$$\bar{\mathbf{P}}_\infty^{-1} = (\mathbf{M}^2 \bar{\mathbf{P}}_\infty)^{-1} + \mathbf{R}^{-1} \tag{5}$$

$$\iff \quad \bar{\mathbf{P}}_\infty = \bar{\mathbf{K}}_\infty \mathbf{R} \,, \quad \bar{\mathbf{K}}_\infty = \begin{cases} \mathbf{I} - \mathbf{M}^{-2} & \text{if } \mathbf{M} \geq 1 \,, \\ 0 & \text{otherwise.} \end{cases} \tag{6}$$

Initial conditions (ICs) don't appear

$\implies$ ICs are "forgotten".

$\implies$ Sampling error is attenuated.

Deductions from Riccati: Attenuation

Stationary solution:

$$\bar{\mathbf{P}}_\infty^{-1} = (\mathbf{M}^2 \bar{\mathbf{P}}_\infty)^{-1} + \mathbf{R}^{-1} \tag{5}$$

$$\iff \quad \bar{\mathbf{P}}_\infty = \bar{\mathbf{K}}_\infty \mathbf{R} \,, \quad \bar{\mathbf{K}}_\infty = \begin{cases} \mathbf{I} - \mathbf{M}^{-2} & \text{if } \mathbf{M} \geq 1 \,, \\ 0 & \text{otherwise.} \end{cases} \tag{6}$$

Initial conditions (ICs) don't appear

$\implies$ ICs are "forgotten".

$\implies$ Sampling error is attenuated.

Deductions from Riccati: Attenuation

Stationary solution:

$$\bar{\mathbf{P}}_\infty^{-1} = (\mathbf{M}^2 \bar{\mathbf{P}}_\infty)^{-1} + \mathbf{R}^{-1} \tag{5}$$

$$\iff \quad \bar{\mathbf{P}}_\infty = \bar{\mathbf{K}}_\infty \mathbf{R} \,, \quad \bar{\mathbf{K}}_\infty = \begin{cases} \mathbf{I} - \mathbf{M}^{-2} & \text{if } \mathbf{M} \geq 1 \,, \\ 0 & \text{otherwise.} \end{cases} \tag{6}$$

Initial conditions (ICs) don't appear
$\implies$ ICs are "forgotten".
$\implies$ Sampling error is attenuated.

Deductions from Riccati: Attenuation

Stationary solution:

$$\bar{\mathbf{P}}_\infty^{-1} = (\mathbf{M}^2 \bar{\mathbf{P}}_\infty)^{-1} + \mathbf{R}^{-1} \tag{5}$$

$$\iff \quad \bar{\mathbf{P}}_\infty = \bar{\mathbf{K}}_\infty \mathbf{R} \,, \quad \bar{\mathbf{K}}_\infty = \begin{cases} \mathbf{I} - \mathbf{M}^{-2} & \text{if } \mathbf{M} \geq 1 \,, \\ 0 & \text{otherwise.} \end{cases} \tag{6}$$

Initial conditions (ICs) don't appear
$\implies$ ICs are "forgotten".
$\implies$ Sampling error is attenuated.

Deductions from Riccati: Filter divergence

Perturbation analysis:

$$\text{Recall:} \qquad \bar{\mathbf{P}}_k = \underbrace{(\mathbf{I} - \bar{\mathbf{K}}_k)}\, \mathbf{M}^2 \bar{\mathbf{P}}_{k-1} \, . \tag{7}$$

$$\text{By contrast,} \qquad \delta\bar{\mathbf{P}}_k \approx (\mathbf{I} - \bar{\mathbf{K}}_k)^2 \left[\mathbf{M}^2 + \mathcal{M}\mathcal{M}''\right] \delta\bar{\mathbf{P}}_{k-1} \, , \tag{8}$$

Yielding $\delta\bar{\mathbf{P}}_k \xrightarrow[k\to\infty]{} 0$ in the linear case ($\mathcal{M}'' = 0$), as we found previously.

By contrast, no such guarantee exists when $\mathcal{M}'' \neq 0$
$\implies$ filter divergence.

Also, $\mathcal{M}''$ may grow worse with k
$\implies$ vicious circle.

Deductions from Riccati: Filter divergence

Perturbation analysis:

Recall:
$$\bar{\mathbf{P}}_k = \underbrace{(\mathbf{I} - \bar{\mathbf{K}}_k)} \mathbf{M}^2 \bar{\mathbf{P}}_{k-1} \,. \tag{7}$$

By contrast,
$$\delta\bar{\mathbf{P}}_k \approx (\mathbf{I} - \bar{\mathbf{K}}_k)^2 \left[\mathbf{M}^2 + \mathcal{M}\mathcal{M}''\right] \delta\bar{\mathbf{P}}_{k-1} \,, \tag{8}$$

Yielding $\delta\bar{\mathbf{P}}_k \xrightarrow[k\to\infty]{} 0$ in the linear case ($\mathcal{M}'' = 0$), as we found previously.

By contrast, no such guarantee exists when $\mathcal{M}'' \neq 0$
$\implies$ filter divergence.

Also, $\mathcal{M}''$ may grow worse with k
$\implies$ vicious circle.

Deductions from Riccati: Filter divergence

Perturbation analysis:

Recall: $$\bar{\mathbf{P}}_k = \underbrace{(\mathbf{I} - \bar{\mathbf{K}}_k)} \mathbf{M}^2 \bar{\mathbf{P}}_{k-1} \,. \tag{7}$$

By contrast, $$\delta\bar{\mathbf{P}}_k \approx (\mathbf{I} - \bar{\mathbf{K}}_k)^2 \left[\mathbf{M}^2 + \mathcal{M}\mathcal{M}''\right] \delta\bar{\mathbf{P}}_{k-1} \,, \tag{8}$$

Yielding $\delta\bar{\mathbf{P}}_k \xrightarrow[k\to\infty]{} 0$ in the linear case ($\mathcal{M}'' = 0$), as we found previously.

By contrast, no such guarantee exists when $\mathcal{M}'' \neq 0$
$\implies$ filter divergence.

Also, $\mathcal{M}''$ may grow worse with k
$\implies$ vicious circle.

Deductions from Riccati: Filter divergence

Perturbation analysis:

Recall: $$\bar{\mathbf{P}}_k = \underbrace{(\mathbf{I} - \bar{\mathbf{K}}_k)}_{\xrightarrow[k\to\infty]{}\mathbf{M}^{-2}} \mathbf{M}^2 \bar{\mathbf{P}}_{k-1} \, . \tag{7}$$

By contrast, $$\delta\bar{\mathbf{P}}_k \approx (\mathbf{I} - \bar{\mathbf{K}}_k)^2 \, [\mathbf{M}^2 + \mathcal{M}\mathcal{M}''] \, \delta\bar{\mathbf{P}}_{k-1} \, . \tag{8}$$

Yielding $\delta\bar{\mathbf{P}}_k \xrightarrow[k\to\infty]{} 0$ in the linear case ($\mathcal{M}'' = 0$),
as we found previously.

By contrast, no such guarantee exists when $\mathcal{M}'' \neq 0$
$\implies$ filter divergence.

Also, $\mathcal{M}''$ may grow worse with k
$\implies$ vicious circle.

Deductions from Riccati: Filter divergence

Perturbation analysis:

Recall: $$\bar{\mathbf{P}}_k = \underbrace{(\,\mathbf{I} - \bar{\mathbf{K}}_k\,)}_{\xrightarrow[k\to\infty]{}\mathbf{M}^{-2}} \; \mathbf{M}^2 \bar{\mathbf{P}}_{k-1}\,. \tag{7}$$

By contrast, $$\delta\bar{\mathbf{P}}_k \approx (\,\mathbf{I} - \bar{\mathbf{K}}_k\,)^2 \; [\mathbf{M}^2 + \mathcal{M}\mathcal{M}''] \; \delta\bar{\mathbf{P}}_{k-1}\,, \tag{8}$$

Yielding $\delta\bar{\mathbf{P}}_k \xrightarrow[k\to\infty]{} 0$ in the linear case ($\mathcal{M}'' = 0$), as we found previously.

By contrast, no such guarantee exists when $\mathcal{M}'' \neq 0$
$\Longrightarrow$ filter divergence.

Also, $\mathcal{M}''$ may grow worse with k
$\Longrightarrow$ vicious circle.

Deductions from Riccati: Filter divergence

Perturbation analysis:

Recall: $$\bar{\mathbf{P}}_k = \underbrace{(\mathbf{I} - \bar{\mathbf{K}}_k)}_{\xrightarrow[k\to\infty]{} \mathbf{M}^{-2}} \mathbf{M}^2 \bar{\mathbf{P}}_{k-1}\,. \tag{7}$$

By contrast, $$\delta\bar{\mathbf{P}}_k \approx (\mathbf{I} - \bar{\mathbf{K}}_k)^2 \, [\mathbf{M}^2 + \mathcal{M}\mathcal{M}''] \, \delta\bar{\mathbf{P}}_{k-1}\,, \tag{8}$$

Yielding $\delta\bar{\mathbf{P}}_k \xrightarrow[k\to\infty]{} 0$ in the linear case ($\mathcal{M}'' = 0$),

as we found previously.

By contrast, no such guarantee exists when $\mathcal{M}'' \neq 0$

$\Longrightarrow$ filter divergence.

Also, $\mathcal{M}''$ may grow worse with k

$\Longrightarrow$ vicious circle.

Deductions from Riccati: Filter divergence

Perturbation analysis:

Recall: $$\bar{\mathbf{P}}_k = \underbrace{(\mathbf{I} - \bar{\mathbf{K}}_k)}_{\xrightarrow[k\to\infty]{} \mathbf{M}^{-2}} \mathbf{M}^2 \bar{\mathbf{P}}_{k-1} \,. \tag{7}$$

By contrast, $$\delta\bar{\mathbf{P}}_k \approx (\mathbf{I} - \bar{\mathbf{K}}_k)^2 \, [\mathbf{M}^2 + \mathcal{M}\mathcal{M}''] \, \delta\bar{\mathbf{P}}_{k-1} \,, \tag{8}$$

Yielding $\delta\bar{\mathbf{P}}_k \xrightarrow[k\to\infty]{} 0$ in the linear case ($\mathcal{M}'' = 0$), as we found previously.

By contrast, no such guarantee exists when $\mathcal{M}'' \neq 0$
$\implies$ filter divergence.

Also, $\mathcal{M}''$ may grow worse with k
$\implies$ vicious circle.

Deductions from Riccati: Filter divergence

Perturbation analysis:

Recall: $$\bar{\mathbf{P}}_k = \underbrace{(\,\mathbf{I} - \bar{\mathbf{K}}_k\,)}_{\xrightarrow[k\to\infty]{}\mathbf{M}^{-2}}\ \mathbf{M}^2\bar{\mathbf{P}}_{k-1}\,. \tag{7}$$

By contrast, $$\delta\bar{\mathbf{P}}_k \approx (\,\mathbf{I} - \bar{\mathbf{K}}_k\,)^2\ [\mathbf{M}^2 + \mathcal{M}\mathcal{M}'']\ \delta\bar{\mathbf{P}}_{k-1}\,, \tag{8}$$

Yielding $\delta\bar{\mathbf{P}}_k \xrightarrow[k\to\infty]{} 0$ in the linear case ($\mathcal{M}'' = 0$),
as we found previously.

By contrast, no such guarantee exists when $\mathcal{M}'' \neq 0$
$\implies$ filter divergence.

Also, $\mathcal{M}''$ may grow worse with k
$\implies$ vicious circle.

Deductions from Riccati: Filter divergence

Perturbation analysis:

Recall: $$\bar{\mathbf{P}}_k = \underbrace{(\mathbf{I} - \bar{\mathbf{K}}_k)}_{\xrightarrow[k\to\infty]{} \mathbf{M}^{-2}} \mathbf{M}^2 \bar{\mathbf{P}}_{k-1} \, . \tag{7}$$

By contrast, $$\delta\bar{\mathbf{P}}_k \approx (\mathbf{I} - \bar{\mathbf{K}}_k)^2 \, [\mathbf{M}^2 + \mathcal{M}\mathcal{M}''] \, \delta\bar{\mathbf{P}}_{k-1} \, , \tag{8}$$

Yielding $\delta\bar{\mathbf{P}}_k \xrightarrow[k\to\infty]{} 0$ in the linear case ($\mathcal{M}'' = 0$), as we found previously.

By contrast, no such guarantee exists when $\mathcal{M}'' \neq 0$
$\implies$ filter divergence.

Also, $\mathcal{M}''$ may grow worse with k
$\implies$ vicious circle.

Deductions from Riccati: Filter divergence

Perturbation analysis:

Recall: $$\bar{\mathbf{P}}_k = \underbrace{(\,\mathbf{I} - \bar{\mathbf{K}}_k\,)}_{\xrightarrow[k\to\infty]{}\mathbf{M}^{-2}} \mathbf{M}^2 \bar{\mathbf{P}}_{k-1}\,. \tag{7}$$

By contrast, $$\delta\bar{\mathbf{P}}_k \approx (\,\mathbf{I} - \bar{\mathbf{K}}_k\,)^2\,[\mathbf{M}^2 + \mathcal{M}\mathcal{M}'']\,\delta\bar{\mathbf{P}}_{k-1}\,, \tag{8}$$

Yielding $\delta\bar{\mathbf{P}}_k \xrightarrow[k\to\infty]{} 0$ in the linear case ($\mathcal{M}'' = 0$), as we found previously.

By contrast, no such guarantee exists when $\mathcal{M}'' \neq 0$

$\implies$ filter divergence.

Also, $\mathcal{M}''$ may grow worse with k

$\implies$ vicious circle.

Deductions from Riccati: Filter divergence

Perturbation analysis:

Recall: $$\bar{\mathbf{P}}_k = \underbrace{(\mathbf{I} - \bar{\mathbf{K}}_k)}_{\xrightarrow[k\to\infty]{} \mathbf{M}^{-2}} \mathbf{M}^2 \bar{\mathbf{P}}_{k-1} \,. \tag{7}$$

By contrast, $$\delta\bar{\mathbf{P}}_k \approx (\mathbf{I} - \bar{\mathbf{K}}_k)^2 \, [\mathbf{M}^2 + \mathcal{M}\mathcal{M}''] \, \delta\bar{\mathbf{P}}_{k-1} \,, \tag{8}$$

Yielding $\delta\bar{\mathbf{P}}_k \xrightarrow[k\to\infty]{} 0$ in the linear case ($\mathcal{M}'' = 0$), as we found previously.

By contrast, no such guarantee exists when $\mathcal{M}'' \neq 0$
$\implies$ filter divergence.

Also, $\mathcal{M}''$ may grow worse with k
$\implies$ vicious circle.

Deductions from Riccati: Why $(N-1)$?

Riccati invariant to change $\tilde{\mathbf{P}} = \alpha\bar{\mathbf{P}}$, hence:

$$1/\tilde{\mathbf{P}}_\infty = 1/(\mathbf{M}^2\tilde{\mathbf{P}}_\infty) + 1/\mathbf{R} \tag{9}$$

Sampling error from nonlinearity – why?

- Consider the m-th "true" and "sample" moments:

$$\mu_m = \mathbb{E}[x^m], \tag{10}$$

$$\hat{\mu}_m = N^{-1} \sum_{n=1}^{N} x_n^m. \tag{11}$$

- Define: $\text{Error}_m = \hat{\mu}_m - \mu_m$.
- Define: $\mu_m^{\mathsf{f}} = \mathbb{E}[(\mathcal{M}(x))^m]$.
- Assume degree-d Taylor-exp. of $\mathcal{M}$ is accurate. Then

$$\mu_m^{\mathsf{f}} = \sum_{i=1}^{md} C_{m,i}\mu_i. \tag{12}$$

- Due to coupling of moments,

$$\text{Error}_m^{\mathsf{f}} = \sum_{i=1}^{md} C_{m,i}\text{Error}_i^{a}, \tag{13}$$

which defeats moment-matching.

Sampling error from nonlinearity – why?

- Consider the m-th "true" and "sample" moments:

$$\mu_m = \mathbb{E}[x^m] \,, \tag{10}$$

$$\hat{\mu}_m = N^{-1} \sum_{n=1}^{N} x_n^m \,. \tag{11}$$

- Define: $\text{Error}_m = \hat{\mu}_m - \mu_m$.
- Define: $\mu_m^{\mathsf{f}} = \mathbb{E}[(\mathcal{M}(x))^m]$.
- Assume degree-d Taylor-exp. of $\mathcal{M}$ is accurate. Then

$$\mu_m^{\mathsf{f}} = \sum_{i=1}^{md} C_{m,i} \mu_i \,. \tag{12}$$

- Due to coupling of moments,

$$\text{Error}_m^{\mathsf{f}} = \sum_{i=1}^{md} C_{m,i} \text{Error}_i^a \,, \tag{13}$$

which defeats moment-matching.

Sampling error from nonlinearity – why?

- Consider the m-th "true" and "sample" moments:

$$\mu_m = \mathbb{E}[x^m]\,, \tag{10}$$

$$\hat{\mu}_m = N^{-1} \sum_{n=1}^{N} x_n^m\,. \tag{11}$$

- Define: $\text{Error}_m = \hat{\mu}_m - \mu_m\,.$
- Define: $\mu^{\text{f}}_m = \mathbb{E}[(\mathcal{M}(x))^m]$.
- Assume degree-d Taylor-exp. of $\mathcal{M}$ is accurate. Then

$$\mu^{\text{f}}_m = \sum_{i=1}^{md} C_{m,i}\mu_i\,. \tag{12}$$

- Due to coupling of moments,

$$\text{Error}^{\text{f}}_m = \sum_{i=1}^{md} C_{m,i}\text{Error}_i\,, \tag{13}$$

which defeats moment-matching.

Sampling error from nonlinearity – why?

- Consider the m-th "true" and "sample" moments:

$$\mu_m = \mathbb{E}[x^m] \,, \tag{10}$$

$$\hat{\mu}_m = N^{-1} \sum_{n=1}^{N} x_n^m \,. \tag{11}$$

- Define: $\text{Error}_m = \hat{\mu}_m - \mu_m \,.$
- Define: $\mu_m^{\mathsf{f}} = \mathbb{E}\big[\big(\mathcal{M}(x)\big)^m\big]$.
- Assume degree-d Taylor-exp. of $\mathcal{M}$ is accurate. Then

$$\mu_m^{\mathsf{f}} = \sum_{i=1}^{md} C_{m,i} \mu_i \,. \tag{12}$$

- Due to coupling of moments,

$$\text{Error}_m^{\mathsf{f}} = \sum_{i=1}^{md} C_{m,i} \text{Error}_i \,, \tag{13}$$

which defeats moment-matching.

Sampling error from nonlinearity – why?

- Consider the m-th "true" and "sample" moments:

$$\mu_m = \mathbb{E}[x^m]\,, \tag{10}$$

$$\hat{\mu}_m = N^{-1} \sum_{n=1}^{N} x_n^m\,. \tag{11}$$

- Define: $\text{Error}_m = \hat{\mu}_m - \mu_m\,.$
- Define: $\mu_m^{\mathsf{f}} = \mathbb{E}\big[\big(\mathcal{M}(x)\big)^m\big].$
- Assume degree-d Taylor-exp. of $\mathcal{M}$ is accurate. Then

$$\mu_m^{\mathsf{f}} = \sum_{i=1}^{md} C_{m,i}\mu_i\,. \tag{12}$$

- Due to coupling of moments,

$$\text{Error}_m^{\mathsf{f}} = \sum_{i=1}^{md} C_{m,i}\text{Error}_i\,, \tag{13}$$

which defeats moment-matching.

Sampling error from nonlinearity – why?

- Consider the m-th "true" and "sample" moments:

$$\mu_m = \mathbb{E}[x^m]\,, \tag{10}$$

$$\hat{\mu}_m = N^{-1} \sum_{n=1}^{N} x_n^m\,. \tag{11}$$

- Define: $\mathsf{Error}_m = \hat{\mu}_m - \mu_m\,.$
- Define: $\mu_m^{\mathsf{f}} = \mathbb{E}\big[\big(\mathcal{M}(x)\big)^m\big]$.
- Assume degree-d Taylor-exp. of $\mathcal{M}$ is accurate. Then

$$\mu_m^{\mathsf{f}} = \sum_{i=1}^{md} C_{m,i}\mu_i\,. \tag{12}$$

- Hence,

$$\mathsf{Error}_m^{\mathsf{f}} = \sum_{i=1}^{md} C_{m,i}\mathsf{Error}_i\,, \tag{13}$$

  which defeats moment-matching.

Why do we prefer the Kalman gain "form"?

Not equivalent when $(N-1) < M$:

$$\bar{\mathbf{P}} = [\mathbf{I} - \bar{\mathbf{K}}\mathbf{H}]\bar{\mathbf{B}} \tag{14}$$

$$\bar{\mathbf{P}} = (\bar{\mathbf{B}}^{+} + \mathbf{H}^{\mathsf{T}}\mathbf{R}^{-1}\mathbf{H})^{-1} \tag{15}$$

Which is better?

Note that eqn. (15) follows from

$$\text{prior} \propto \exp[-\tfrac{1}{2}(x - \bar{x})^{\mathsf{T}}\,\bar{\mathbf{B}}^{+}\,(x - \bar{x})]\,, \tag{16}$$

which is "flat" in the directions outside of $\text{col}(\bar{\mathbf{B}})$.
$\Longrightarrow$ eqn. (15) yields "opposite" of the correct update.

Note: further complications in case $\bar{\mathbf{P}}$ not defined in eqn. (15).

Why do we prefer the Kalman gain "form"?

Not equivalent when $(N-1) < M$:

$$\bar{\mathbf{P}} = [\mathbf{I} - \bar{\mathbf{K}}\mathbf{H}]\bar{\mathbf{B}} \tag{14}$$

$$\bar{\mathbf{P}} = (\bar{\mathbf{B}}^{+} + \mathbf{H}^{\mathsf{T}}\mathbf{R}^{-1}\mathbf{H})^{-1} \tag{15}$$

Which is better?

Note that eqn. (15) follows from

$$\text{prior} \propto \exp[-\tfrac{1}{2}(x - \bar{x})^{\mathsf{T}}\,\bar{\mathbf{B}}^{+}\,(x - \bar{x})]\,, \tag{16}$$

which is "flat" in the directions outside of $\mathrm{col}(\bar{\mathbf{B}})$.

$\Longrightarrow$ eqn. (15) yields "opposite" of the correct update.

Note: further complications in case $\bar{\mathbf{P}}$ not defined in eqn. (15).

Why do we prefer the Kalman gain "form"?

Not equivalent when $(N-1) < M$:

$$\bar{\mathbf{P}} = [\mathbf{I} - \bar{\mathbf{K}}\mathbf{H}]\bar{\mathbf{B}} \tag{14}$$

$$\bar{\mathbf{P}} = (\bar{\mathbf{B}}^{+} + \mathbf{H}^{\mathsf{T}}\mathbf{R}^{-1}\mathbf{H})^{-1} \tag{15}$$

Which is better?

Note that eqn. (15) follows from

$$\text{prior} \propto \exp[-\tfrac{1}{2}(\boldsymbol{x} - \bar{\boldsymbol{x}})^{\mathsf{T}}\, \bar{\mathbf{B}}^{+}(\boldsymbol{x} - \bar{\boldsymbol{x}})]\,, \tag{16}$$

which is "flat" in the directions outside of $\mathrm{col}(\bar{\mathbf{B}})$.

$\Longrightarrow$ eqn. (15) yields "opposite" of the correct update.

Note: further complications in case $\bar{\mathbf{P}}$ not defined in eqn. (15).

Why do we prefer the Kalman gain "form"?

Not equivalent when $(N-1) < M$:

$$\bar{\mathbf{P}} = [\mathbf{I} - \bar{\mathbf{K}}\mathbf{H}]\bar{\mathbf{B}} \tag{14}$$

$$\bar{\mathbf{P}} = (\bar{\mathbf{B}}^{+} + \mathbf{H}^{\mathsf{T}}\mathbf{R}^{-1}\mathbf{H})^{-1} \tag{15}$$

Which is better?

Note that eqn. (15) follows from

$$\text{prior} \propto \exp[-\tfrac{1}{2}(\boldsymbol{x} - \bar{\boldsymbol{x}})^{\mathsf{T}}\, \bar{\mathbf{B}}^{+}(\boldsymbol{x} - \bar{\boldsymbol{x}})]\,, \tag{16}$$

which is "flat" in the directions outside of $\operatorname{col}(\bar{\mathbf{B}})$.

$\Longrightarrow$ eqn. (15) yields "opposite" of the correct update.

Note: further complications in case $\bar{\mathbf{P}}$ not defined in eqn. (15).

Why do we prefer the Kalman gain "form"?

Not equivalent when $(N-1) < M$:

$$\bar{\mathbf{P}} = [\mathbf{I} - \bar{\mathbf{K}}\mathbf{H}]\bar{\mathbf{B}} \tag{14}$$

$$\bar{\mathbf{P}} = (\bar{\mathbf{B}}^{+} + \mathbf{H}^{\mathsf{T}}\mathbf{R}^{-1}\mathbf{H})^{-1} \tag{15}$$

Which is better?

Note that eqn. (15) follows from

$$\text{prior} \propto \exp[-\tfrac{1}{2}(\boldsymbol{x} - \bar{\boldsymbol{x}})^{\mathsf{T}}\,\bar{\mathbf{B}}^{+}(\boldsymbol{x} - \bar{\boldsymbol{x}})]\,, \tag{16}$$

which is "flat" in the directions outside of $\mathrm{col}(\bar{\mathbf{B}})$.

$\Longrightarrow$ eqn. (15) yields "opposite" of the correct update.

Note: further complications in case $\bar{\mathbf{P}}$ not defined in eqn. (15).

Why do we prefer the Kalman gain "form"?

Not equivalent when $(N-1) < M$:

$$\bar{\mathbf{P}} = [\mathbf{I} - \bar{\mathbf{K}}\mathbf{H}]\bar{\mathbf{B}} \tag{14}$$

$$\bar{\mathbf{P}} = (\bar{\mathbf{B}}^{+} + \mathbf{H}^{\mathsf{T}}\mathbf{R}^{-1}\mathbf{H})^{-1} \tag{15}$$

Which is better?

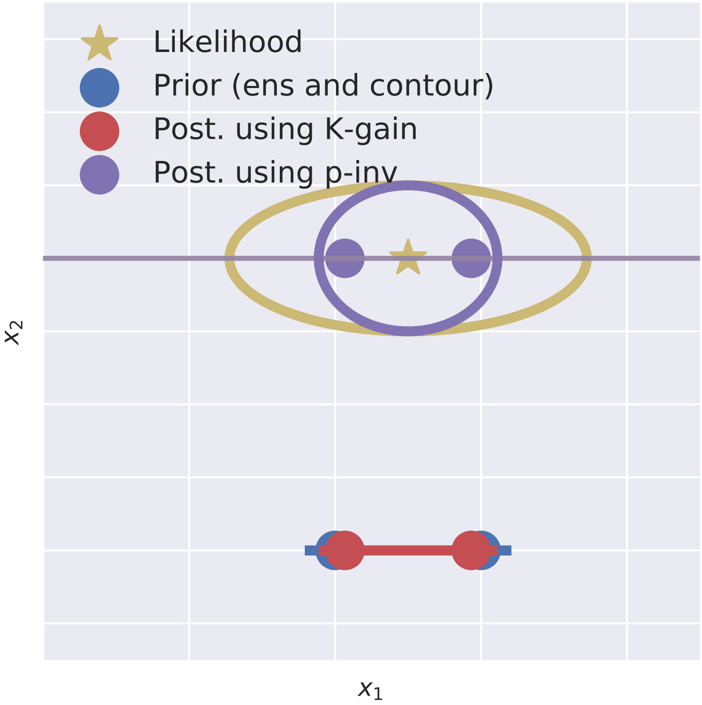
Note that eqn. (15) follows from

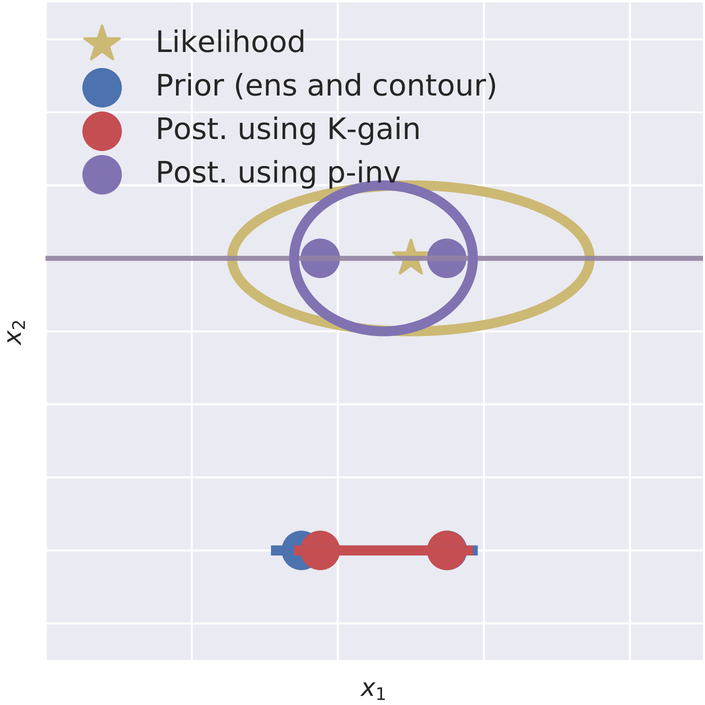
$$\text{prior} \propto \exp[-\tfrac{1}{2}(\boldsymbol{x} - \bar{\boldsymbol{x}})^{\mathsf{T}}\, \bar{\mathbf{B}}^{+}(\boldsymbol{x} - \bar{\boldsymbol{x}})]\,, \tag{16}$$

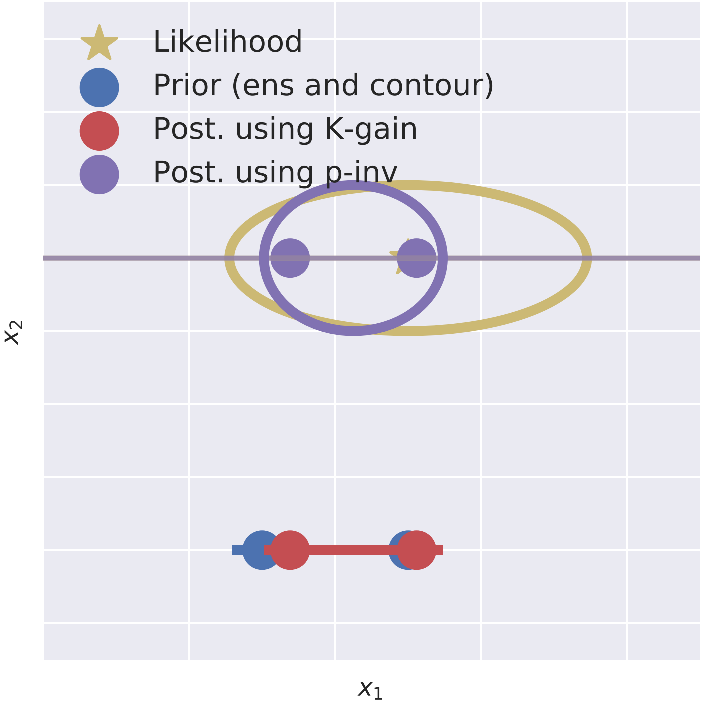
which is "flat" in the directions outside of $\operatorname{col}(\bar{\mathbf{B}})$.

$\implies$ eqn. (15) yields "opposite" of the correct update.

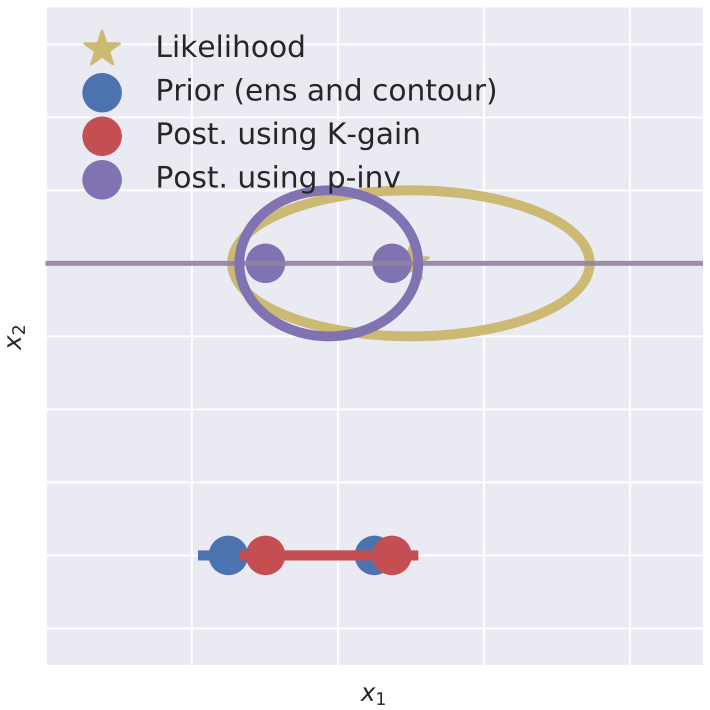
Note: further complications in case $\bar{\mathbf{P}}$ not defined in eqn. (15).


Likelihood
Prior (ens and contour)
Post. using K-gain
Post. using p-inv
x_2
x_1

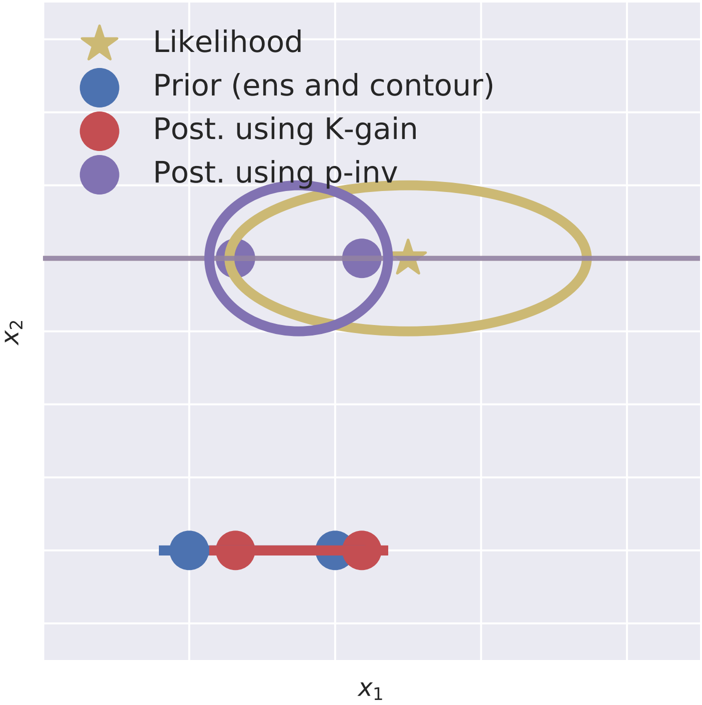

Likelihood
Prior (ens and contour)
Post. using K-gain
Post. using p-inv
x_2
x_1



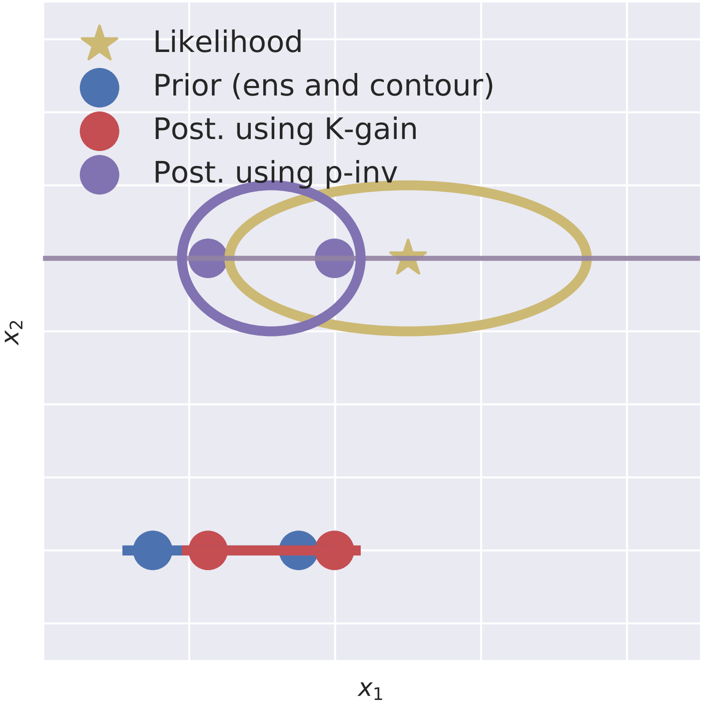
Likelihood
Prior (ens and contour)
Post. using K-gain
Post. using p-inv
x_2
x_1



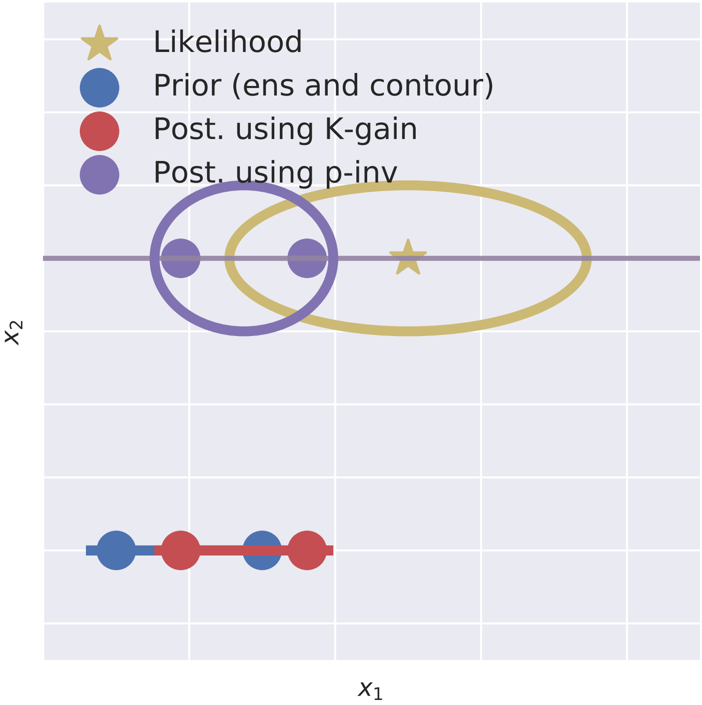
Likelihood
Prior (ens and contour)
Post. using K-gain
Post. using p-inv
x_2
x_1

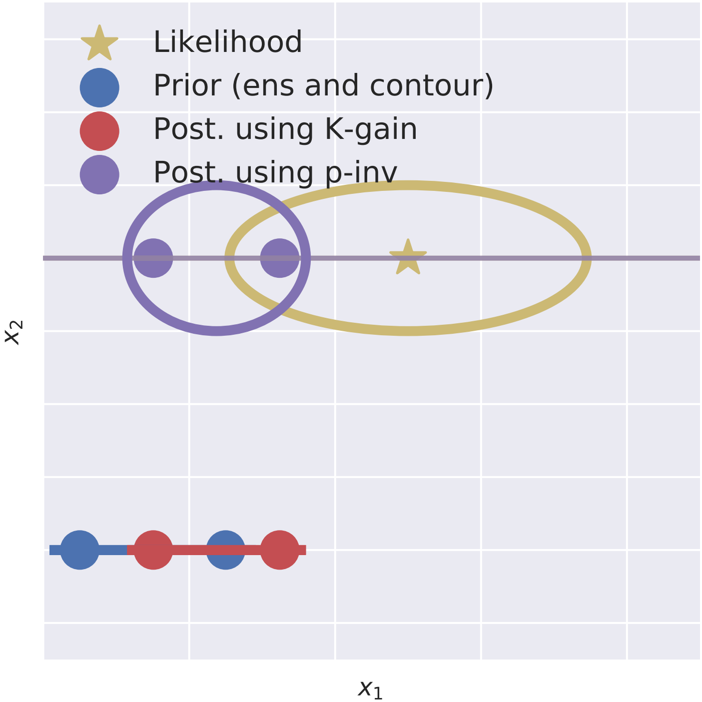


Likelihood
Prior (ens and contour)
Post. using K-gain
Post. using p-inv
x_2
x_1

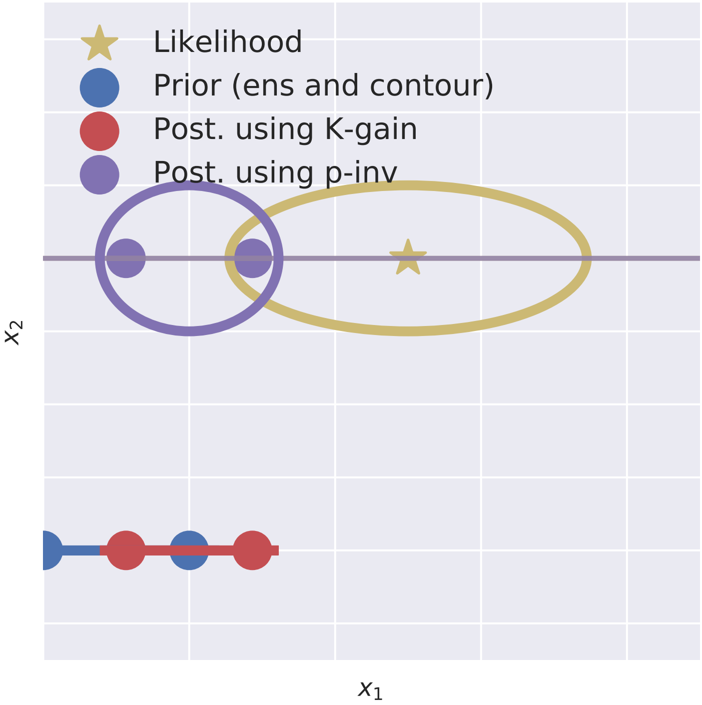


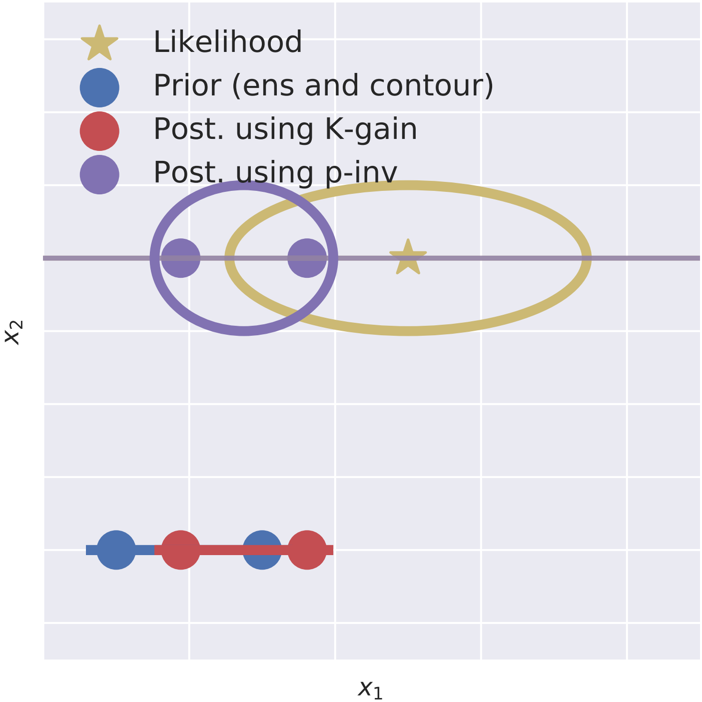
Likelihood
Prior (ens and contour)
Post. using K-gain
Post. using p-inv
x_2
x_1


Likelihood
Prior (ens and contour)
Post. using K-gain
Post. using p-inv
x_2
x_1

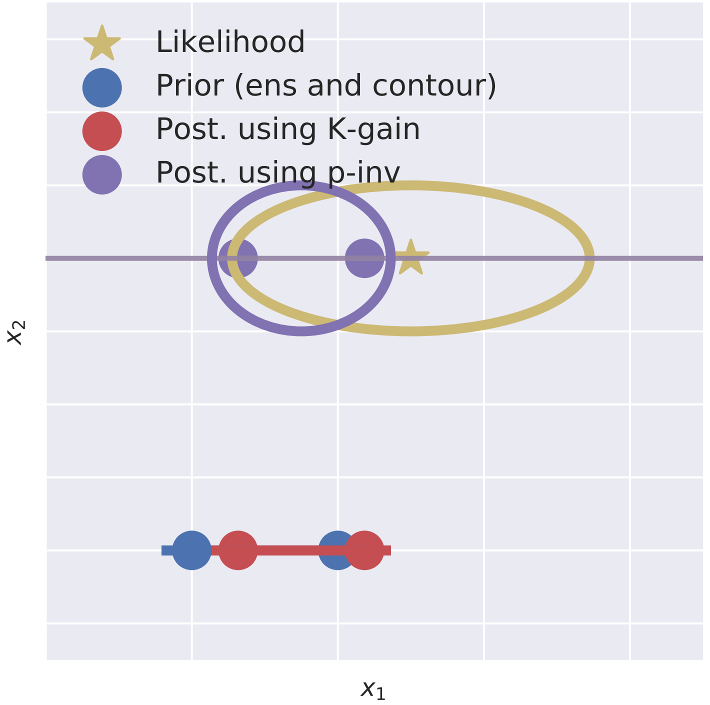


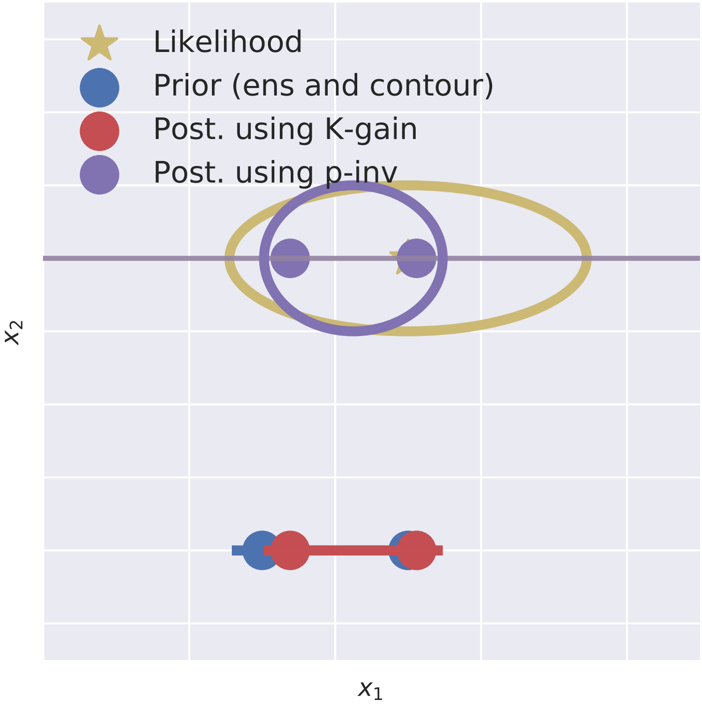
Likelihood
Prior (ens and contour)
Post. using K-gain
Post. using p-inv
x_2
x_1


Likelihood
Prior (ens and contour)
Post. using K-gain
Post. using p-inv
x_2
x_1

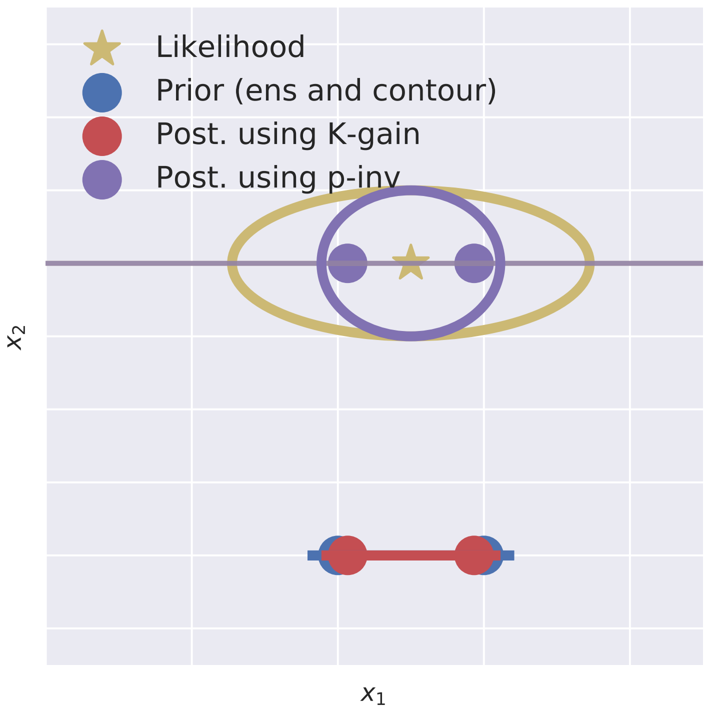


Likelihood
Prior (ens and contour)
Post. using K-gain
Post. using p-inv
x_2
x_1

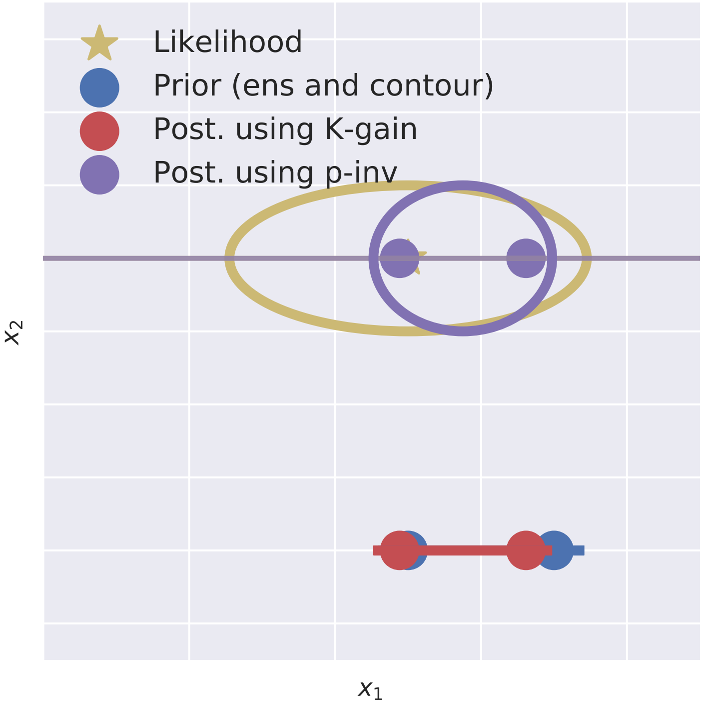

Likelihood
Prior (ens and contour)
Post. using K-gain
Post. using p-inv
x_2
x_1

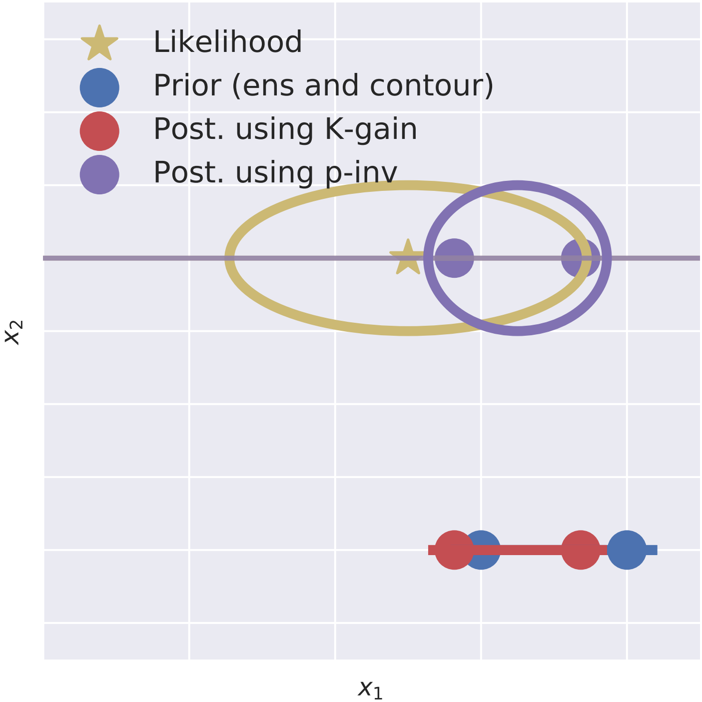


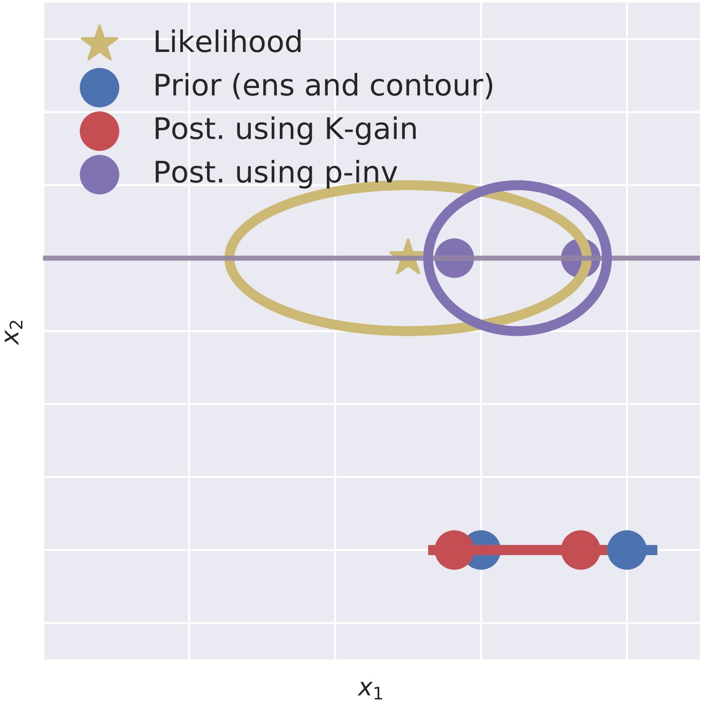
Likelihood
Prior (ens and contour)
Post. using K-gain
Post. using p-inv
x_2
x_1

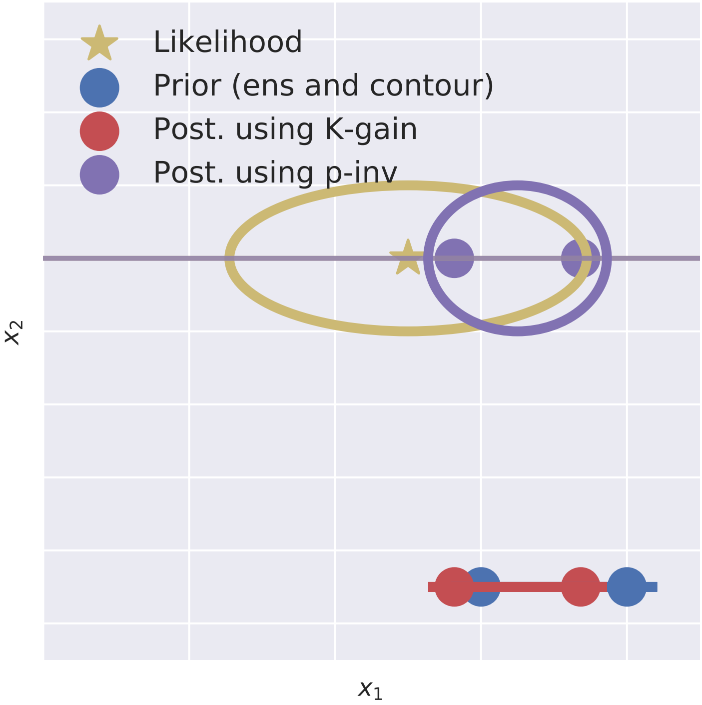


Likelihood
Prior (ens and contour)
Post. using K-gain
Post. using p-inv
x_2
x_1

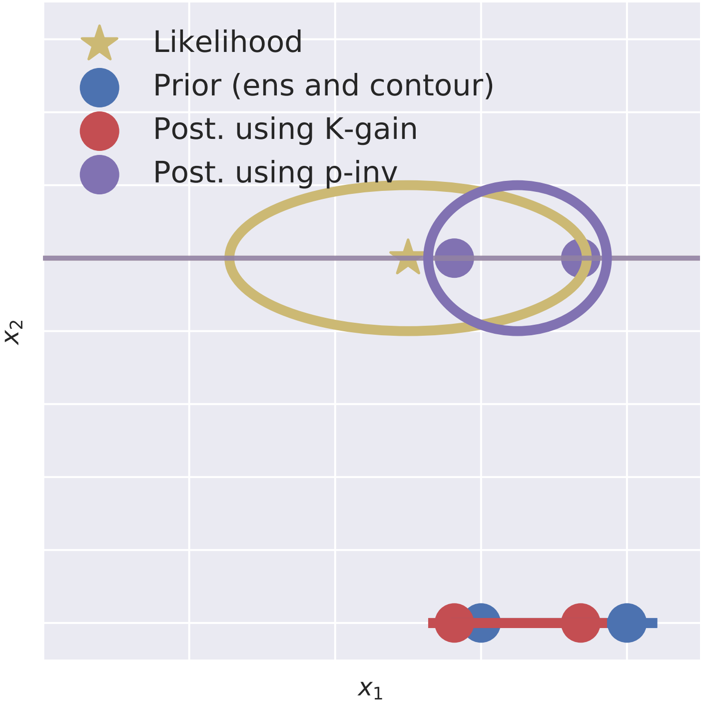

Likelihood
Prior (ens and contour)
Post. using K-gain
Post. using p-inv
x_2
x_1


Likelihood
Prior (ens and contour)
Post. using K-gain
Post. using p-inv
x_2
x_1

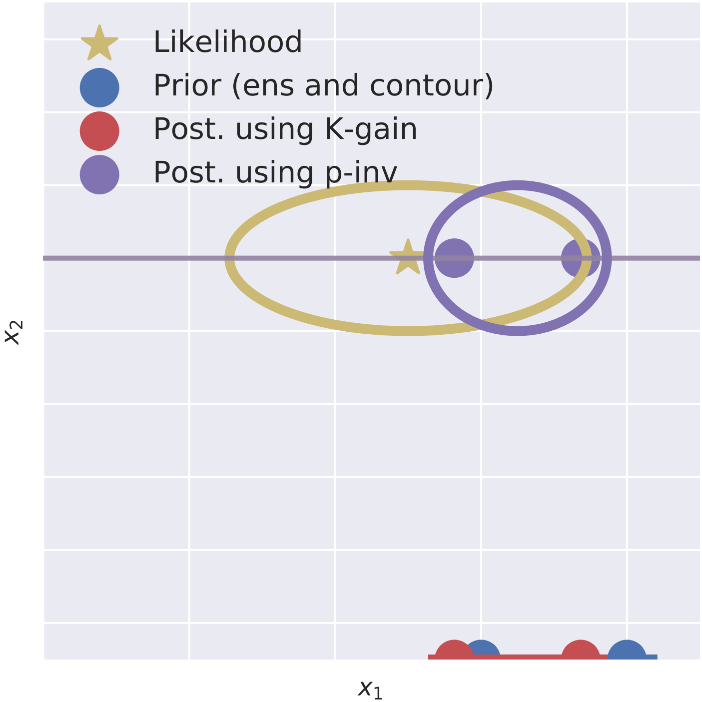

Likelihood
Prior (ens and contour)
Post. using K-gain
Post. using p-inv
x_2
x_1

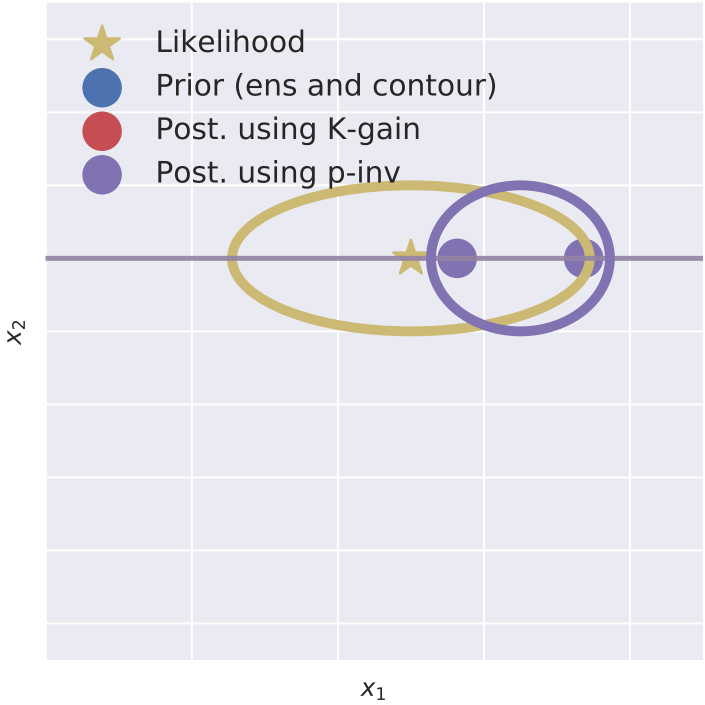


Likelihood
Prior (ens and contour)
Post. using K-gain
Post. using p-inv
x_2
x_1

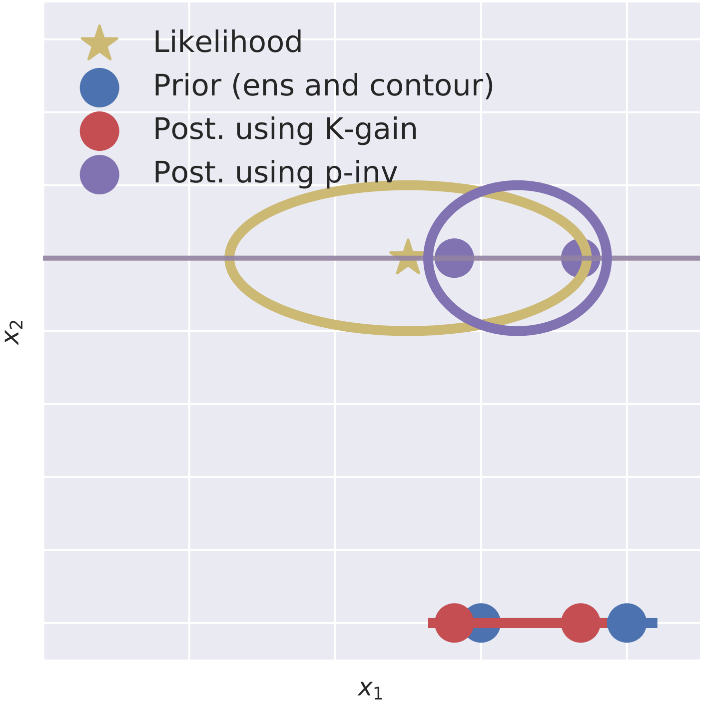


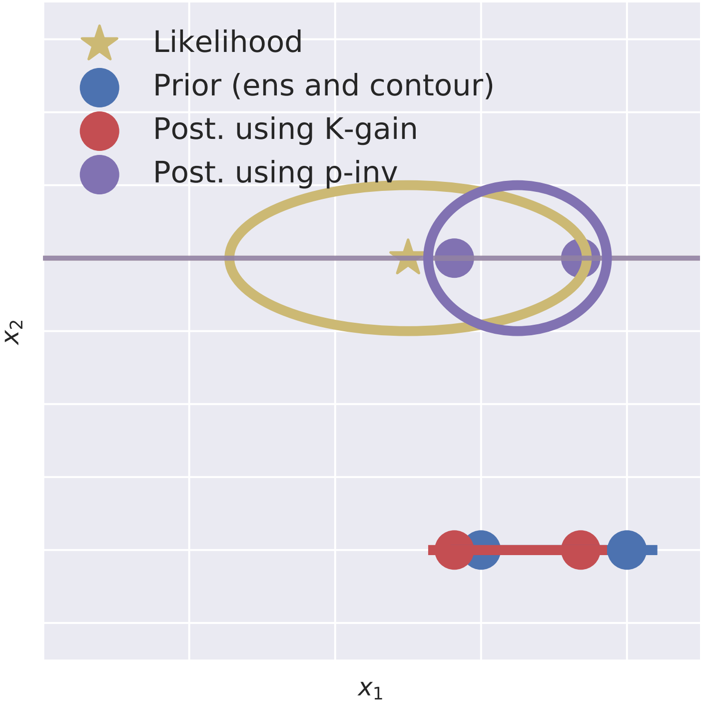
Likelihood
Prior (ens and contour)
Post. using K-gain
Post. using p-inv
x_2
x_1

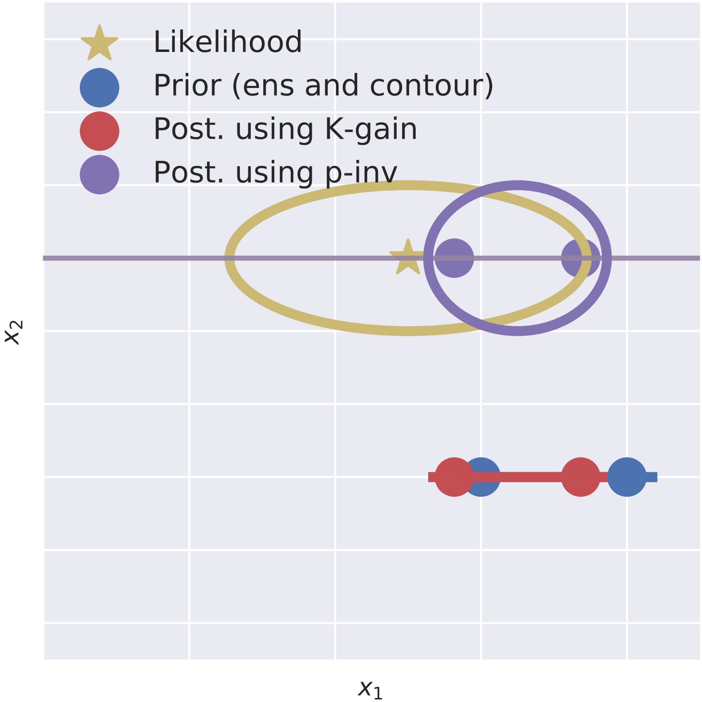

Likelihood
Prior (ens and contour)
Post. using K-gain
Post. using p-inv
x_2
x_1

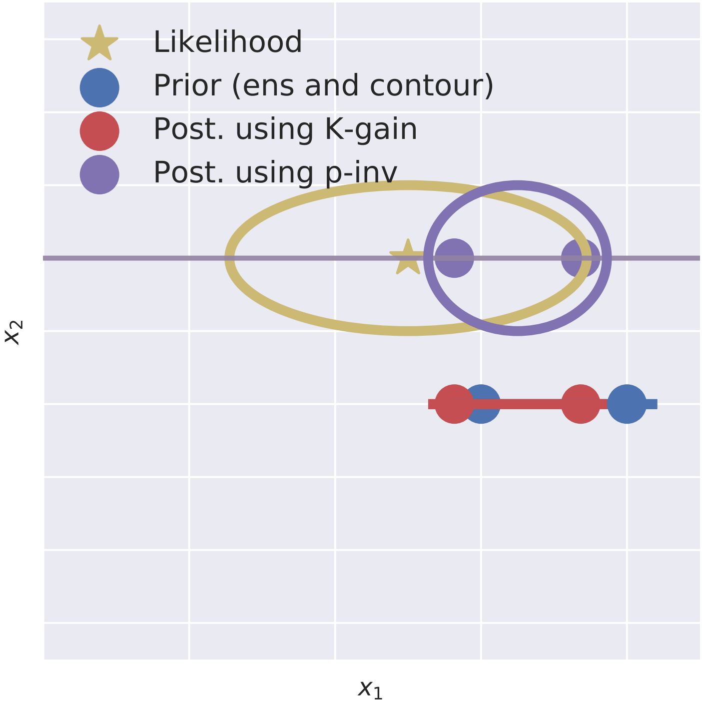


Likelihood
Prior (ens and contour)
Post. using K-gain
Post. using p-inv
x_2
x_1

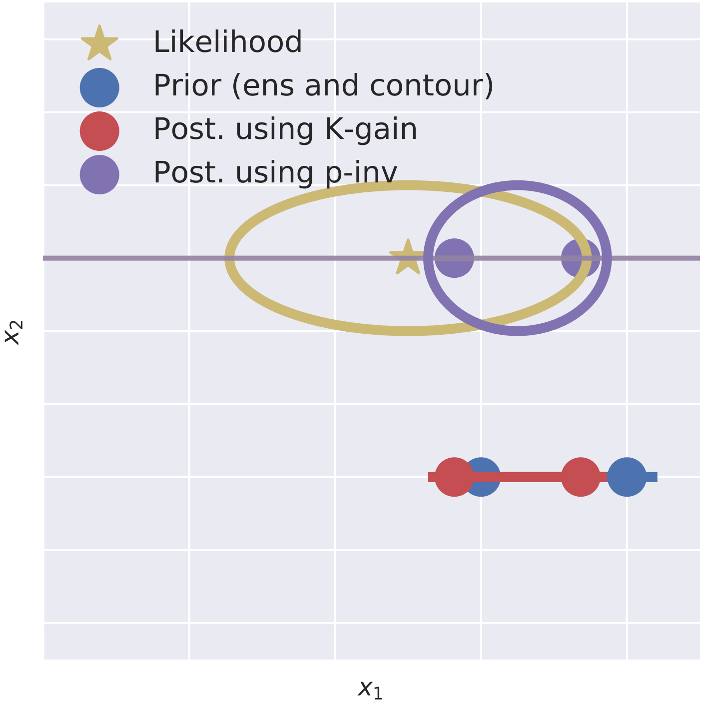

Likelihood
Prior (ens and contour)
Post. using K-gain
Post. using p-inv
x_2
x_1


Likelihood
Prior (ens and contour)
Post. using K-gain
Post. using p-inv
x_2
x_1

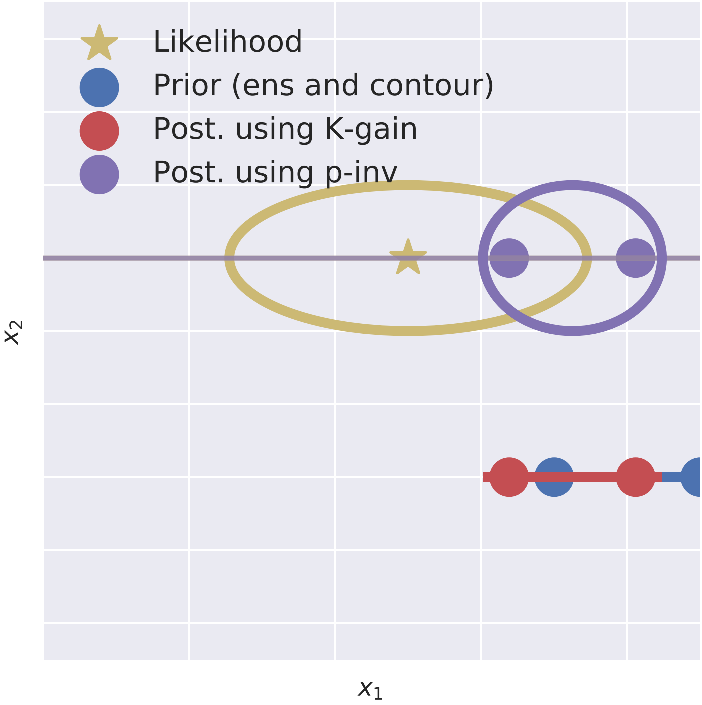

Likelihood
Prior (ens and contour)
Post. using K-gain
Post. using p-inv
x_2
x_1



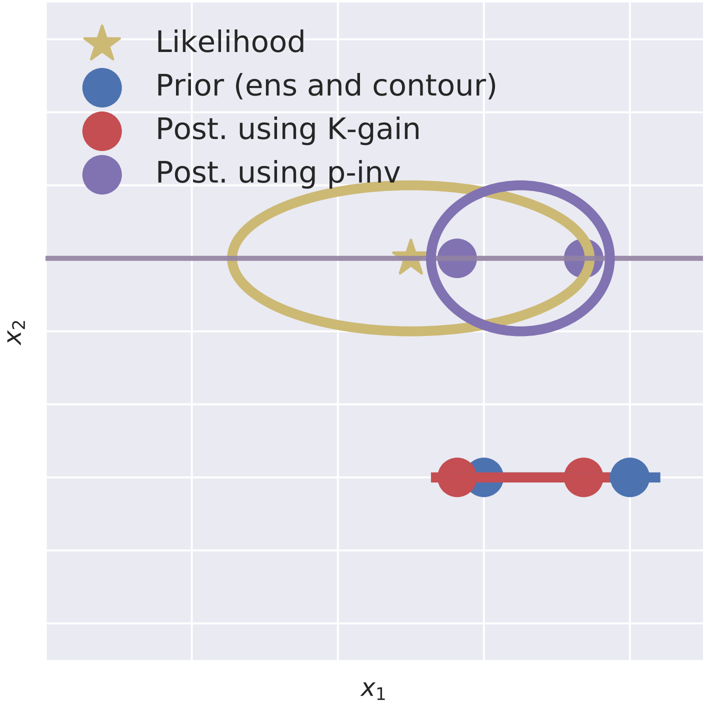
Likelihood
Prior (ens and contour)
Post. using K-gain
Post. using p-inv
x_2
x_1

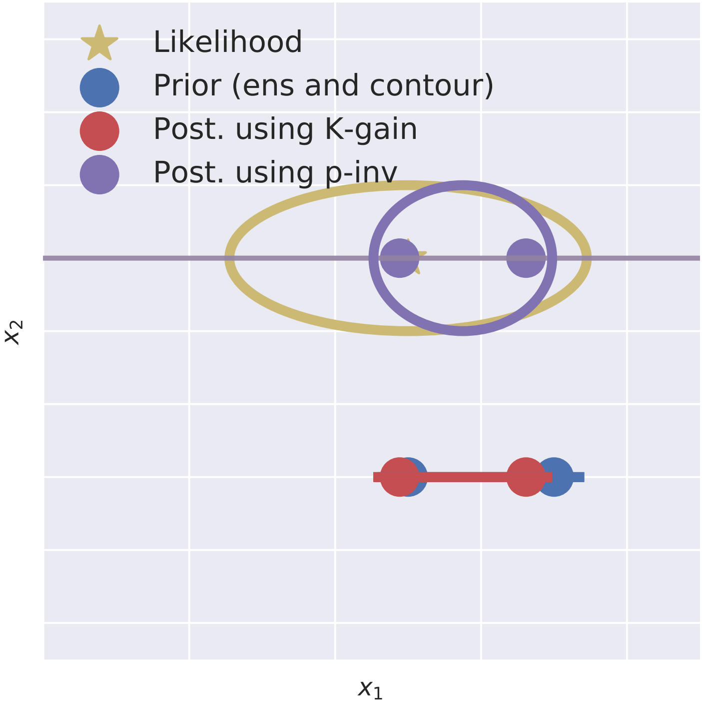


Likelihood
Prior (ens and contour)
Post. using K-gain
Post. using p-inv
x_2
x_1

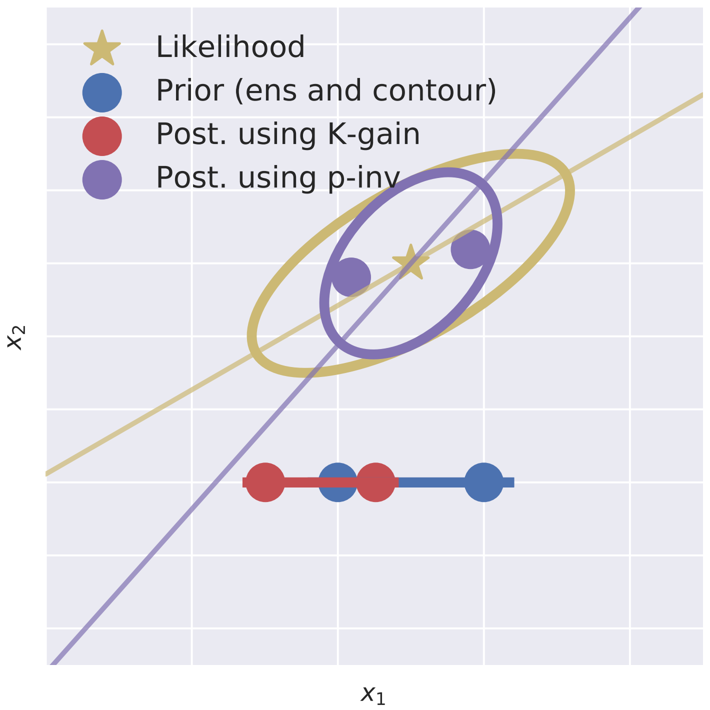


Likelihood
Prior (ens and contour)
Post. using K-gain
Post. using p-inv
x_2
x_1

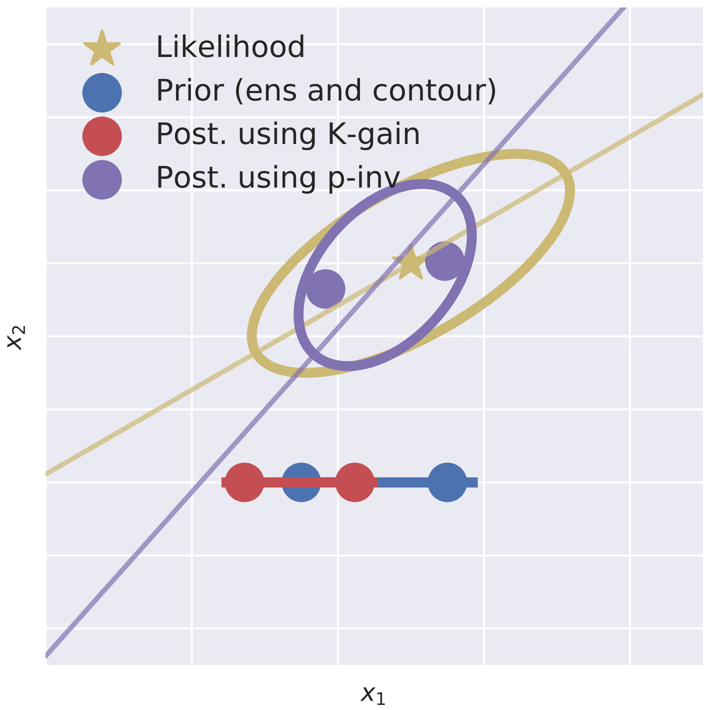

Likelihood
Prior (ens and contour)
Post. using K-gain
Post. using p-inv
x_2
x_1



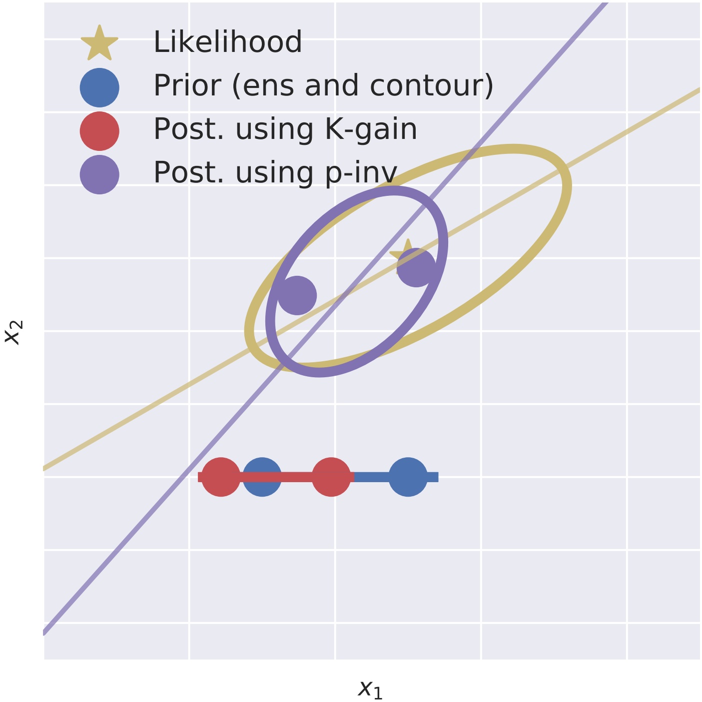
Likelihood
Prior (ens and contour)
Post. using K-gain
Post. using p-inv
x_2
x_1



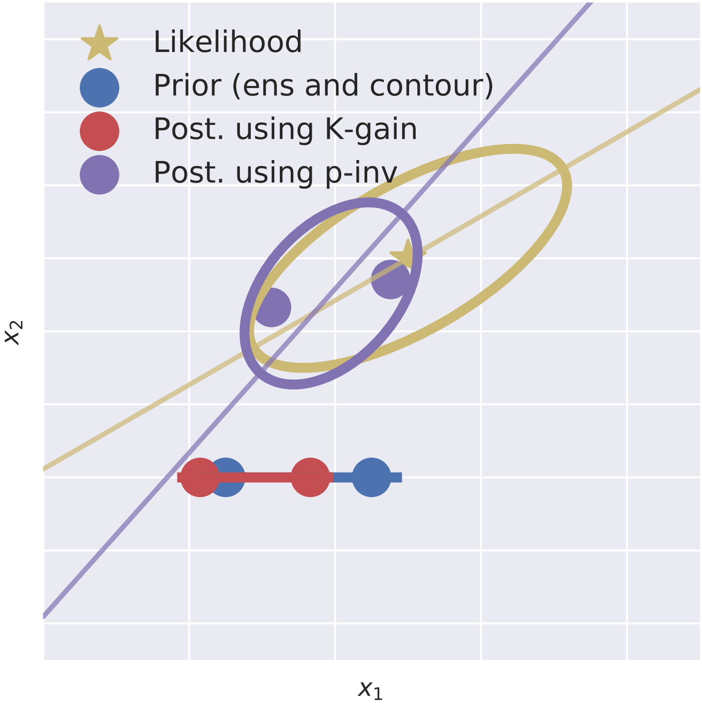
Likelihood
Prior (ens and contour)
Post. using K-gain
Post. using p-inv
x_2
x_1



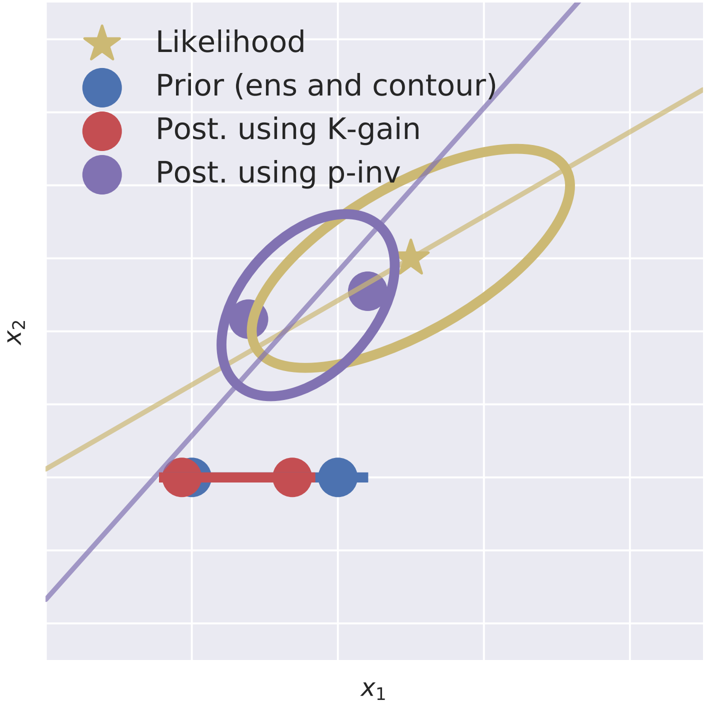
Likelihood
Prior (ens and contour)
Post. using K-gain
Post. using p-inv
x_2
x_1

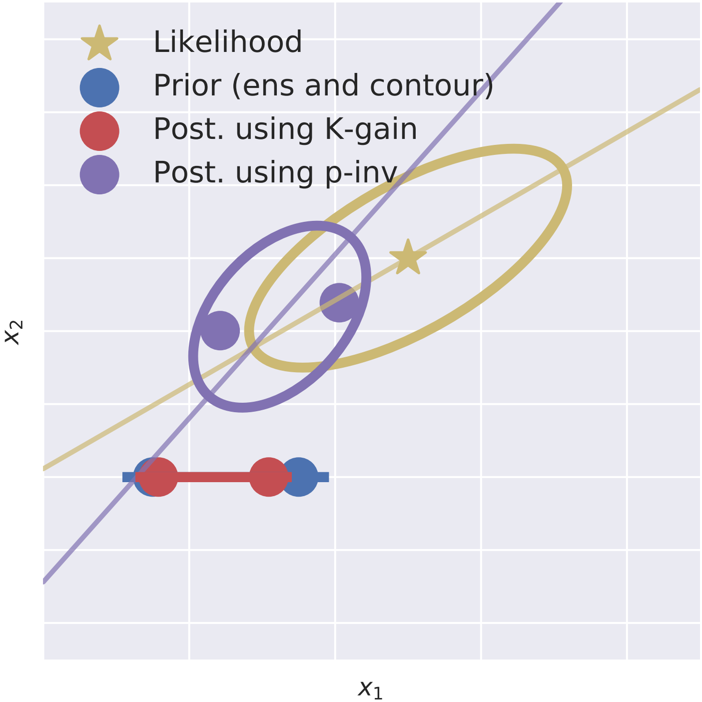


Likelihood
Prior (ens and contour)
Post. using K-gain
Post. using p-inv
x_2
x_1

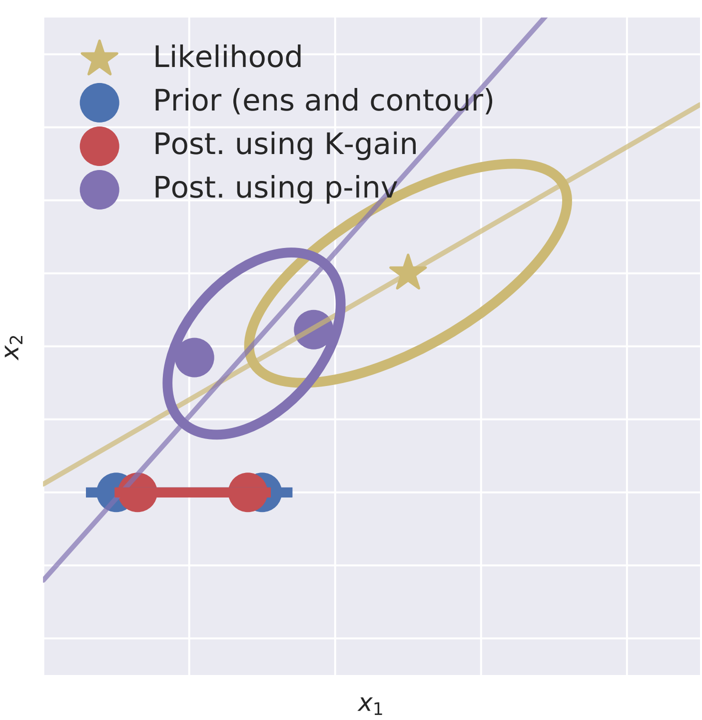


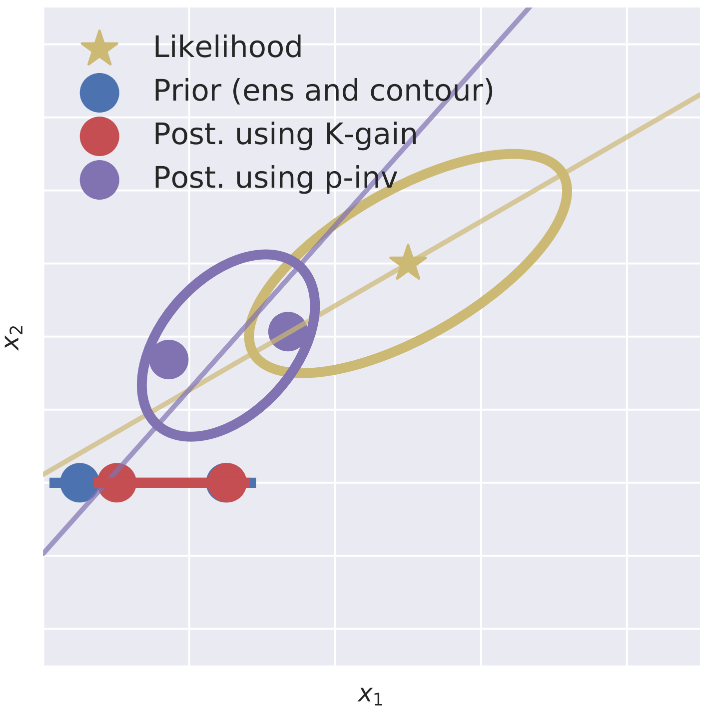
Likelihood
Prior (ens and contour)
Post. using K-gain
Post. using p-inv
x_2
x_1

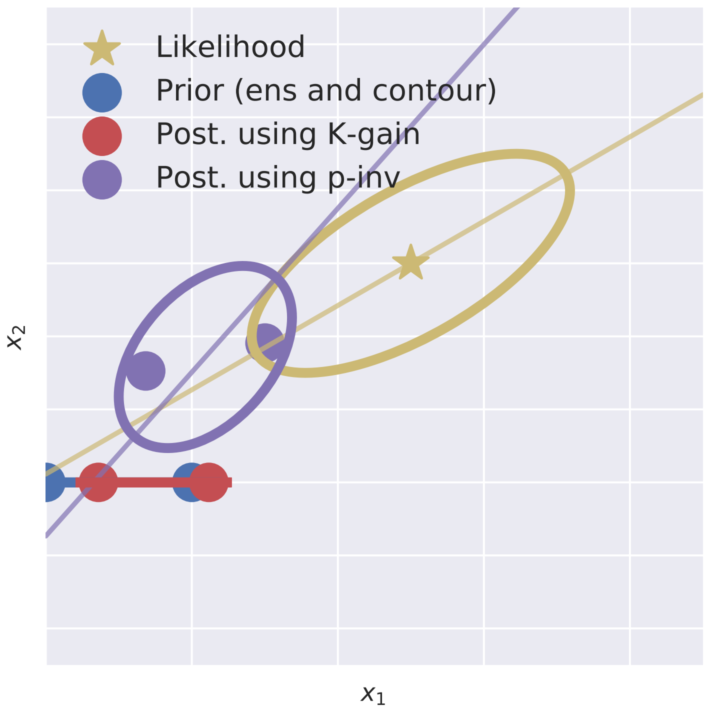

Likelihood
Prior (ens and contour)
Post. using K-gain
Post. using p-inv
x_2
x_1



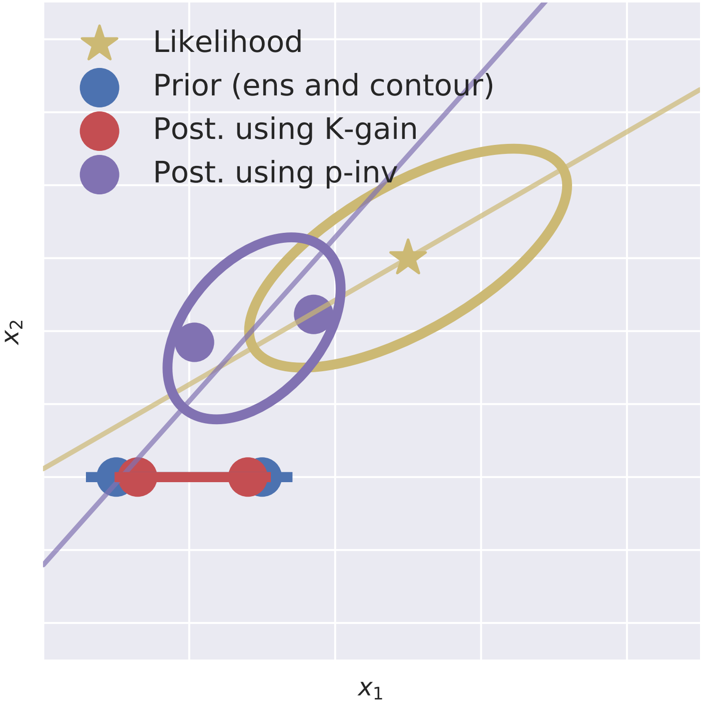
Likelihood
Prior (ens and contour)
Post. using K-gain
Post. using p-inv
x_2
x_1

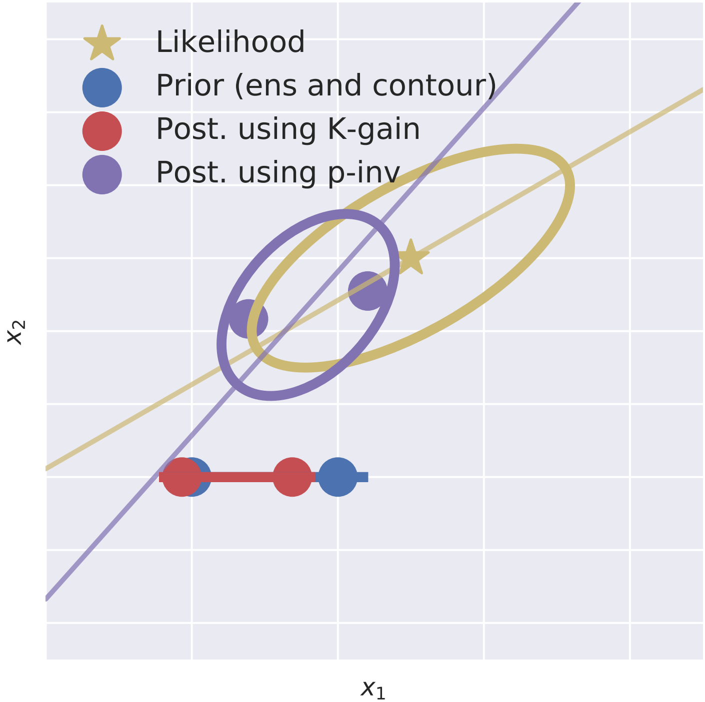

Likelihood
Prior (ens and contour)
Post. using K-gain
Post. using p-inv
x_2
x_1


Likelihood
Prior (ens and contour)
Post. using K-gain
Post. using p-inv
x_2
x_1

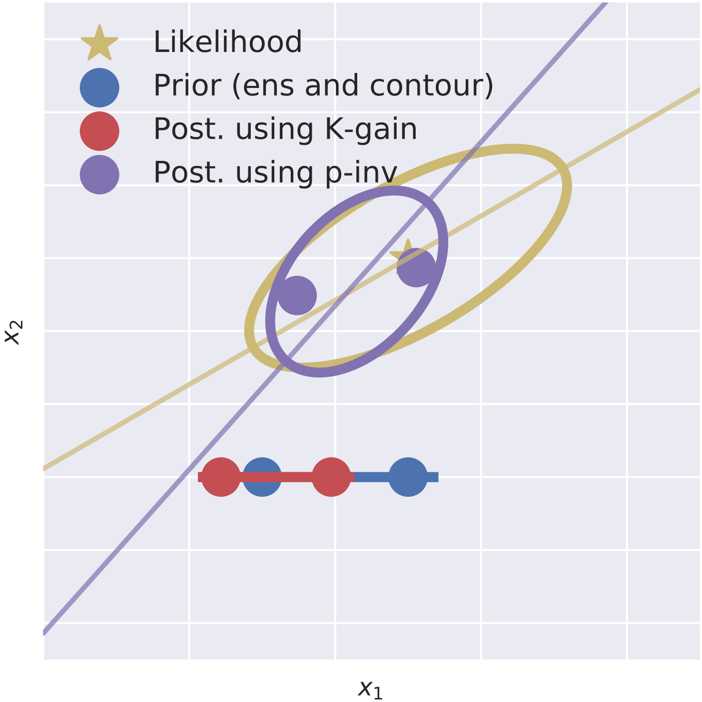


Likelihood
Prior (ens and contour)
Post. using K-gain
Post. using p-inv
x_2
x_1

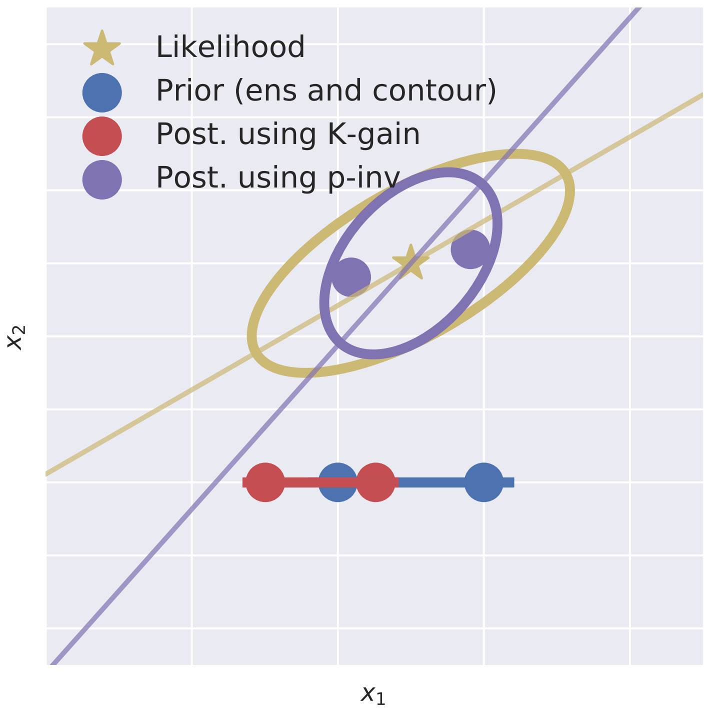

Likelihood
Prior (ens and contour)
Post. using K-gain
Post. using p-inv
x_2
x_1



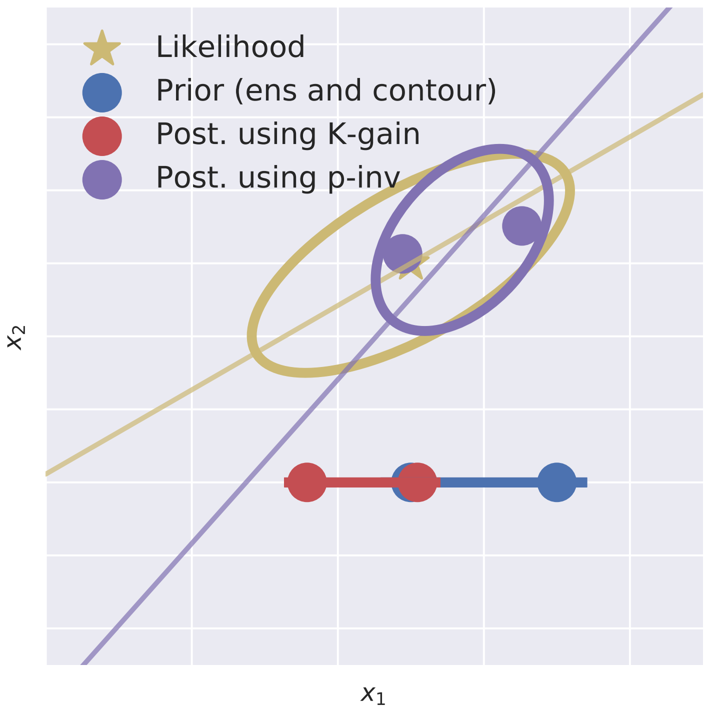
Likelihood
Prior (ens and contour)
Post. using K-gain
Post. using p-inv
x_2
x_1

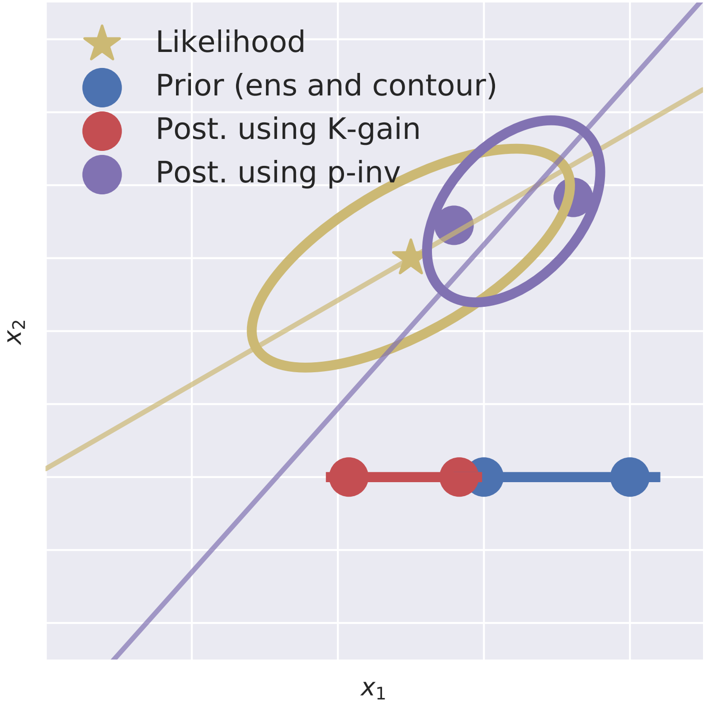


Likelihood
Prior (ens and contour)
Post. using K-gain
Post. using p-inv
x_2
x_1



Likelihood
Prior (ens and contour)
Post. using K-gain
Post. using p-inv
x_2
x_1


Likelihood
Prior (ens and contour)
Post. using K-gain
Post. using p-inv
x_2
x_1


Likelihood
Prior (ens and contour)
Post. using K-gain
Post. using p-inv
x_2
x_1

Ensemble linearizations

Recall the EnKF gain:

$$\bar{\mathbf{K}} = \underbrace{\bar{\mathbf{C}}_{\boldsymbol{xy}}}_{\frac{1}{N-1}\mathbf{X}\mathbf{Y}^\mathsf{T}} \Big(\underbrace{\bar{\mathbf{C}}_{\boldsymbol{y}}}_{\frac{1}{N-1}\mathbf{Y}\mathbf{Y}^\mathsf{T}} + \mathbf{R} \Big)^{-1} . \tag{17}$$

Question: Is there a matrix $\bar{\mathbf{H}}$ such that

$$\bar{\mathbf{K}} = \bar{\mathbf{C}}_{\boldsymbol{x}} \bar{\mathbf{H}}^\mathsf{T} \Big(\bar{\mathbf{H}} \bar{\mathbf{C}}_{\boldsymbol{x}} \bar{\mathbf{H}}^\mathsf{T} + \mathbf{R} \Big)^{-1} ? \tag{18}$$

Answer: yes (mostly):

$$\bar{\mathbf{H}} = \mathbf{Y}\mathbf{X}^{+} , \tag{19}$$

Follow up questions:

Ensemble linearizations

Recall the EnKF gain:

$$\bar{\mathbf{K}} = \underbrace{\bar{\mathbf{C}}_{\boldsymbol{xy}}}_{\frac{1}{N-1}\mathbf{X}\mathbf{Y}^\mathsf{T}} \Big(\underbrace{\bar{\mathbf{C}}_{\boldsymbol{y}}}_{\frac{1}{N-1}\mathbf{Y}\mathbf{Y}^\mathsf{T}} + \mathbf{R} \Big)^{-1} . \tag{17}$$

Question: Is there a matrix $\bar{\mathbf{H}}$ such that

$$\bar{\mathbf{K}} = \bar{\mathbf{C}}_{\boldsymbol{x}} \bar{\mathbf{H}}^\mathsf{T} \left(\bar{\mathbf{H}} \bar{\mathbf{C}}_{\boldsymbol{x}} \bar{\mathbf{H}}^\mathsf{T} + \mathbf{R} \right)^{-1} ? \tag{18}$$

Answer: yes (mostly): $\bar{\mathbf{H}} = \mathbf{Y}\mathbf{X}^{+}$. (19)

Follow up questions:

Ensemble linearizations

Recall the EnKF gain:

$$\bar{\mathbf{K}} = \underbrace{\bar{\mathbf{C}}_{\boldsymbol{xy}}}_{\frac{1}{N-1}\mathbf{X}\mathbf{Y}^\mathsf{T}} \Big(\underbrace{\bar{\mathbf{C}}_{\boldsymbol{y}}}_{\frac{1}{N-1}\mathbf{Y}\mathbf{Y}^\mathsf{T}} + \mathbf{R} \Big)^{-1} . \tag{17}$$

Question: Is there a matrix $\bar{\mathbf{H}}$ such that

$$\bar{\mathbf{K}} = \bar{\mathbf{C}}_{\boldsymbol{x}} \bar{\mathbf{H}}^\mathsf{T} \big(\bar{\mathbf{H}} \bar{\mathbf{C}}_{\boldsymbol{x}} \bar{\mathbf{H}}^\mathsf{T} + \mathbf{R} \big)^{-1} \; ? \tag{18}$$

Answer: yes (mostly): $\bar{\mathbf{H}} = \mathbf{Y}\mathbf{X}^{+} \, .$ (19)

Follow up questions:

Ensemble linearizations

Recall the EnKF gain:

$$\bar{\mathbf{K}} = \underbrace{\bar{\mathbf{C}}_{\boldsymbol{xy}}}_{\frac{1}{N-1}\mathbf{X}\mathbf{Y}^\mathsf{T}} \Big(\underbrace{\bar{\mathbf{C}}_{\boldsymbol{y}}}_{\frac{1}{N-1}\mathbf{Y}\mathbf{Y}^\mathsf{T}} + \mathbf{R} \Big)^{-1} . \tag{17}$$

Question: Is there a matrix $\bar{\mathbf{H}}$ such that

$$\bar{\mathbf{K}} = \bar{\mathbf{C}}_{\boldsymbol{x}}\bar{\mathbf{H}}^\mathsf{T} \big(\bar{\mathbf{H}}\bar{\mathbf{C}}_{\boldsymbol{x}}\bar{\mathbf{H}}^\mathsf{T} + \mathbf{R} \big)^{-1} \; ? \tag{18}$$

Answer: yes (mostly): $\bar{\mathbf{H}} = \mathbf{Y}\mathbf{X}^{+} \, .$ (19)

Follow up questions:

Ensemble linearizations

Recall the EnKF gain:

$$\bar{\mathbf{K}} = \underbrace{\bar{\mathbf{C}}_{\boldsymbol{xy}}}_{\frac{1}{N-1}\mathbf{X}\mathbf{Y}^\mathsf{T}} \Big(\underbrace{\bar{\mathbf{C}}_{\boldsymbol{y}}}_{\frac{1}{N-1}\mathbf{Y}\mathbf{Y}^\mathsf{T}} + \mathbf{R} \Big)^{-1} . \tag{17}$$

Question: Is there a matrix $\bar{\mathbf{H}}$ such that

$$\bar{\mathbf{K}} = \bar{\mathbf{C}}_{\boldsymbol{x}} \bar{\mathbf{H}}^\mathsf{T} \big(\bar{\mathbf{H}} \bar{\mathbf{C}}_{\boldsymbol{x}} \bar{\mathbf{H}}^\mathsf{T} + \mathbf{R} \big)^{-1} \; ? \tag{18}$$

Answer: yes (mostly): $$\bar{\mathbf{H}} = \mathbf{Y}\mathbf{X}^{+} . \tag{19}$$

Follow up questions:

- Why $\mathbf{Y}\mathbf{X}^{+}$? (what is it?)
- Why is this rarely discussed?
- Does it relate to the analytic derivative ($\mathcal{H}'$) ?

Ensemble linearizations

Recall the EnKF gain:

$$\bar{\mathbf{K}} = \underbrace{\bar{\mathbf{C}}_{xy}}_{\frac{1}{N-1}\mathbf{X}\mathbf{Y}^\mathsf{T}} \Big(\underbrace{\bar{\mathbf{C}}_{y}}_{\frac{1}{N-1}\mathbf{Y}\mathbf{Y}^\mathsf{T}} + \mathbf{R} \Big)^{-1} . \tag{17}$$

Question: Is there a matrix $\bar{\mathbf{H}}$ such that

$$\bar{\mathbf{K}} = \bar{\mathbf{C}}_x \bar{\mathbf{H}}^\mathsf{T} \left(\bar{\mathbf{H}} \bar{\mathbf{C}}_x \bar{\mathbf{H}}^\mathsf{T} + \mathbf{R} \right)^{-1} ? \tag{18}$$

Answer: yes (mostly): $\bar{\mathbf{H}} = \mathbf{Y}\mathbf{X}^+ \, .$ (19)

Follow up questions:

- Why $\mathbf{Y}\mathbf{X}^+$? (what is it?)
- Why is this rarely discussed?
- Does it relate to the analytic derivative ($\mathcal{H}'$) ?

Ensemble linearizations

Recall the EnKF gain:

$$\bar{\mathbf{K}} = \underbrace{\bar{\mathbf{C}}_{\boldsymbol{xy}}}_{\frac{1}{N-1}\mathbf{X}\mathbf{Y}^{\mathsf{T}}} \Big(\underbrace{\bar{\mathbf{C}}_{\boldsymbol{y}}}_{\frac{1}{N-1}\mathbf{Y}\mathbf{Y}^{\mathsf{T}}} + \mathbf{R} \Big)^{-1} . \tag{17}$$

Question: Is there a matrix $\bar{\mathbf{H}}$ such that

$$\bar{\mathbf{K}} = \bar{\mathbf{C}}_{\boldsymbol{x}} \bar{\mathbf{H}}^{\mathsf{T}} \left(\bar{\mathbf{H}} \bar{\mathbf{C}}_{\boldsymbol{x}} \bar{\mathbf{H}}^{\mathsf{T}} + \mathbf{R} \right)^{-1} ? \tag{18}$$

Answer: yes (mostly): $$\bar{\mathbf{H}} = \mathbf{Y}\mathbf{X}^{+} . \tag{19}$$

Follow up questions:

- Why $\mathbf{Y}\mathbf{X}^{+}$? (what is it?)
- Why is this rarely discussed?
- Does it relate to the analytic derivative ($\mathcal{H}'$) ?

Why $\bar{\mathbf{H}} = \mathbf{Y}\mathbf{X}^+$? (what is it?)

$\bar{\mathbf{H}}$ is:

- Linear least-squares (LLS) estimate of $\mathcal{H}$ given $\mathbf{Y}$ and $\mathbf{X}$.
- BLUE ?
- MVUE ?

EnKF also doing LLS of x given (x^f, y) and $\bar{\mathbf{H}}$.

Why $\bar{\mathbf{H}} = \mathbf{Y}\mathbf{X}^{+}$? (what is it?)

$\bar{\mathbf{H}}$ is:

- Linear least-squares (LLS) estimate of $\mathcal{H}$ given $\mathbf{Y}$ and $\mathbf{X}$.
- BLUE ?
- MVUE ?

EnKF also doing LLS of x given (x^f, y) and $\bar{\mathbf{H}}$.

Why $\bar{\mathbf{H}} = \mathbf{Y}\mathbf{X}^+$? (what is it?)

$\bar{\mathbf{H}}$ is:

- Linear least-squares (LLS) estimate of $\mathcal{H}$ given $\mathbf{Y}$ and $\mathbf{X}$.
- BLUE ?
- MVUE ?

EnKF also doing LLS of x given (x^f, y) and $\bar{\mathbf{H}}$.

Why $\bar{\mathbf{H}} = \mathbf{Y}\mathbf{X}^+$? (what is it?)

$\bar{\mathbf{H}}$ is:

- Linear least-squares (LLS) estimate of $\mathcal{H}$ given $\mathbf{Y}$ and $\mathbf{X}$.
- BLUE ?
- MVUE ?

EnKF also doing LLS of x given (x^f, y) and $\bar{\mathbf{H}}$.

Why $\bar{\mathbf{H}} = \mathbf{Y}\mathbf{X}^+$? (what is it?)

$\bar{\mathbf{H}}$ is:

- Linear least-squares (LLS) estimate of $\mathcal{H}$ given $\mathbf{Y}$ and $\mathbf{X}$.
- BLUE ?
- MVUE ?

EnKF also doing LLS of $\boldsymbol{x}$ given $(\boldsymbol{x}^f, \boldsymbol{y})$ and $\bar{\mathbf{H}}$.

Why is $\bar{\mathbf{H}}$ estimate rarely explicit?

Insert $\bar{\mathbf{H}} = \mathbf{Y}\mathbf{X}^{+}$ in $\bar{\mathbf{K}}$:

$$\bar{\mathbf{K}} = \bar{\mathbf{C}}_{\boldsymbol{x}} \bar{\mathbf{H}}^{\mathsf{T}} \big(\bar{\mathbf{H}} \bar{\mathbf{C}}_{\boldsymbol{x}} \bar{\mathbf{H}}^{\mathsf{T}} + \mathbf{R}\big)^{-1} \tag{20}$$

$$= \mathbf{X}\mathbf{Y}^{\mathsf{T}} \big(\mathbf{Y} \boldsymbol{\Pi}_{\mathbf{X}^{\mathsf{T}}} \mathbf{Y}^{\mathsf{T}} + (N{-}1)\mathbf{R}\big)^{-1}, \tag{21}$$

where $\boldsymbol{\Pi}_{\mathbf{X}^{\mathsf{T}}} = \mathbf{X}^{+}\mathbf{X}$.

Note:

- Appearance of $\boldsymbol{\Pi}_{\mathbf{X}^{\mathsf{T}}}$ can be understood from chain rule.
- Vanishing of $\boldsymbol{\Pi}_{\mathbf{X}^{\mathsf{T}}}$ if $\mathcal{H}$ is linear, or $(N{-}1) \leq M$.
- Version without $\boldsymbol{\Pi}_{\mathbf{X}^{\mathsf{T}}}$ generally preferable.

Why is $\bar{\mathbf{H}}$ estimate rarely explicit?

Insert $\bar{\mathbf{H}} = \mathbf{Y}\mathbf{X}^{+}$ in $\bar{\mathbf{K}}$:

$$\bar{\mathbf{K}} = \bar{\mathbf{C}}_{\boldsymbol{x}}\bar{\mathbf{H}}^{\mathsf{T}}\big(\bar{\mathbf{H}}\bar{\mathbf{C}}_{\boldsymbol{x}}\bar{\mathbf{H}}^{\mathsf{T}} + \mathbf{R}\big)^{-1} \tag{20}$$

$$= \mathbf{X}\mathbf{Y}^{\mathsf{T}}\big(\mathbf{Y}\boldsymbol{\Pi}_{\mathbf{X}^{\mathsf{T}}}\mathbf{Y}^{\mathsf{T}} + (N-1)\mathbf{R}\big)^{-1}, \tag{21}$$

where $\boldsymbol{\Pi}_{\mathbf{X}^{\mathsf{T}}} = \mathbf{X}^{+}\mathbf{X}$.

Note:

- Appearance of $\boldsymbol{\Pi}_{\mathbf{X}^{\mathsf{T}}}$ can be understood from chain rule.
- Vanishing of $\boldsymbol{\Pi}_{\mathbf{X}^{\mathsf{T}}}$ if $\mathcal{H}$ is linear, or $(N-1) \leq M$.
- Version without $\boldsymbol{\Pi}_{\mathbf{X}^{\mathsf{T}}}$ generally preferable.

Why is $\bar{\mathbf{H}}$ estimate rarely explicit?

Insert $\bar{\mathbf{H}} = \mathbf{Y}\mathbf{X}^{+}$ in $\bar{\mathbf{K}}$:

$$\bar{\mathbf{K}} = \bar{\mathbf{C}}_{\boldsymbol{x}}\bar{\mathbf{H}}^{\mathsf{T}}\big(\bar{\mathbf{H}}\bar{\mathbf{C}}_{\boldsymbol{x}}\bar{\mathbf{H}}^{\mathsf{T}} + \mathbf{R}\big)^{-1} \tag{20}$$

$$= \mathbf{X}\mathbf{Y}^{\mathsf{T}}\big(\mathbf{Y}\boldsymbol{\Pi}_{\mathbf{X}^{\mathsf{T}}}\mathbf{Y}^{\mathsf{T}} + (N-1)\mathbf{R}\big)^{-1}\,, \tag{21}$$

where $\boldsymbol{\Pi}_{\mathbf{X}^{\mathsf{T}}} = \mathbf{X}^{+}\mathbf{X}$.

Note:

- Appearance of $\boldsymbol{\Pi}_{\mathbf{X}^{\mathsf{T}}}$ can be understood from chain rule.
- Vanishing of $\boldsymbol{\Pi}_{\mathbf{X}^{\mathsf{T}}}$ if $\mathcal{H}$ is linear, or $(N-1) \leq M$.
- Version without $\boldsymbol{\Pi}_{\mathbf{X}^{\mathsf{T}}}$ generally preferable.

Why is $\bar{\mathbf{H}}$ estimate rarely explicit?

Insert $\bar{\mathbf{H}} = \mathbf{Y}\mathbf{X}^{+}$ in $\bar{\mathbf{K}}$:

$$\bar{\mathbf{K}} = \bar{\mathbf{C}}_{\boldsymbol{x}}\bar{\mathbf{H}}^{\mathsf{T}}(\bar{\mathbf{H}}\bar{\mathbf{C}}_{\boldsymbol{x}}\bar{\mathbf{H}}^{\mathsf{T}} + \mathbf{R})^{-1} \tag{20}$$

$$= \mathbf{X}\mathbf{Y}^{\mathsf{T}}(\mathbf{Y}\boldsymbol{\Pi}_{\mathbf{X}^{\mathsf{T}}}\mathbf{Y}^{\mathsf{T}} + (N-1)\mathbf{R})^{-1}\,, \tag{21}$$

where $\boldsymbol{\Pi}_{\mathbf{X}^{\mathsf{T}}} = \mathbf{X}^{+}\mathbf{X}$.

Note:

- Appearance of $\boldsymbol{\Pi}_{\mathbf{X}^{\mathsf{T}}}$ can be understood from chain rule.
- Vanishing of $\boldsymbol{\Pi}_{\mathbf{X}^{\mathsf{T}}}$ if $\mathcal{H}$ is linear, or $(N-1) \leq M$.
- Version without $\boldsymbol{\Pi}_{\mathbf{X}^{\mathsf{T}}}$ generally preferable.

Why is $\bar{\mathbf{H}}$ estimate rarely explicit?

Insert $\bar{\mathbf{H}} = \mathbf{Y}\mathbf{X}^+$ in $\bar{\mathbf{K}}$:

$$\bar{\mathbf{K}} = \bar{\mathbf{C}}_{\boldsymbol{x}} \bar{\mathbf{H}}^{\mathsf{T}} \big(\bar{\mathbf{H}} \bar{\mathbf{C}}_{\boldsymbol{x}} \bar{\mathbf{H}}^{\mathsf{T}} + \mathbf{R}\big)^{-1} \tag{20}$$

$$= \mathbf{X}\mathbf{Y}^{\mathsf{T}} \big(\mathbf{Y} \mathbf{\Pi}_{\mathbf{X}^{\mathsf{T}}} \mathbf{Y}^{\mathsf{T}} + (N-1)\mathbf{R}\big)^{-1}, \tag{21}$$

where $\mathbf{\Pi}_{\mathbf{X}^{\mathsf{T}}} = \mathbf{X}^+\mathbf{X}$.

Note:

- Appearance of $\mathbf{\Pi}_{\mathbf{X}^{\mathsf{T}}}$ can be understood from chain rule.
- Vanishing of $\mathbf{\Pi}_{\mathbf{X}^{\mathsf{T}}}$ if $\mathcal{H}$ is linear, or $(N-1) \leq M$.
- Version without $\mathbf{\Pi}_{\mathbf{X}^{\mathsf{T}}}$ generally preferable.

Why is $\bar{\mathbf{H}}$ estimate rarely explicit?

Insert $\bar{\mathbf{H}} = \mathbf{Y}\mathbf{X}^{+}$ in $\bar{\mathbf{K}}$:

$$\bar{\mathbf{K}} = \bar{\mathbf{C}}_{\boldsymbol{x}}\bar{\mathbf{H}}^{\mathsf{T}}\big(\bar{\mathbf{H}}\bar{\mathbf{C}}_{\boldsymbol{x}}\bar{\mathbf{H}}^{\mathsf{T}} + \mathbf{R}\big)^{-1} \tag{20}$$

$$= \mathbf{X}\mathbf{Y}^{\mathsf{T}}\big(\mathbf{Y}\boldsymbol{\Pi}_{\mathbf{X}^{\mathsf{T}}}\mathbf{Y}^{\mathsf{T}} + (N-1)\mathbf{R}\big)^{-1}, \tag{21}$$

where $\boldsymbol{\Pi}_{\mathbf{X}^{\mathsf{T}}} = \mathbf{X}^{+}\mathbf{X}$.

Note:

- Appearance of $\boldsymbol{\Pi}_{\mathbf{X}^{\mathsf{T}}}$ can be understood from chain rule.
- Vanishing of $\boldsymbol{\Pi}_{\mathbf{X}^{\mathsf{T}}}$ if $\mathcal{H}$ is linear, or $(N-1) \leq M$.
- Version without $\boldsymbol{\Pi}_{\mathbf{X}^{\mathsf{T}}}$ generally preferable.

Does $\bar{\mathbf{H}}$ relate to $\mathcal{H}'$?

Theorem:

$$
\begin{array}{ccc}
\lim_{N\to\infty} \bar{\mathbf{H}} & = & \mathbb{E}[\mathcal{H}'(x)] \\
\| & & \| \\
\lim_{N\to\infty} \mathbf{Y}\mathbf{X}^{+} & & \mathbf{C}_{yx}\mathbf{C}_{x}^{-1} \\
\| & & \\
\lim_{N\to\infty} \mathbf{Y}\mathbf{X}^{\top}(\mathbf{X}\mathbf{X}^{\top})^{-1} & & \\
\| & & \\
\lim_{N\to\infty} \bar{\mathbf{C}}_{yx}\bar{\mathbf{C}}_{x}^{-1} & & \\
\| & & \\
\mathbf{C}_{yx}\mathbf{C}_{x}^{-1} \quad \text{(a.s., by Slutsky, sub. to reg.)} & &
\end{array}
$$

I.e. $\bar{\mathbf{H}}$ is (indeed) the average derivative.

- $p(x)$ same as for ensemble (used for $\bar{\mathbf{H}}$).
- $p(x)$ must be Gaussian!

Does $\bar{\mathbf{H}}$ relate to $\mathcal{H}'$?

Theorem:

$$\lim_{N\to\infty} \bar{\mathbf{H}} = \mathbb{E}[\mathcal{H}'(\boldsymbol{x})]$$

$$\lim_{N\to\infty} \mathbf{Y}\mathbf{X}^{+} \qquad\qquad \mathbf{C}_{yx}\mathbf{C}_x^{-1}$$

$$= \lim_{N\to\infty} \mathbf{Y}\mathbf{X}^{\top}(\mathbf{X}\mathbf{X}^{\top})^{-1}$$

$$= \lim_{N\to\infty} \bar{\mathbf{C}}_{yx}\bar{\mathbf{C}}_x^{-1}$$

$$= \mathbf{C}_{yx}\mathbf{C}_x^{-1}$$ (a.s., by Slutsky, sub. to reg.)

I.e. $\bar{\mathbf{H}}$ is (indeed) the average derivative.

- $p(\boldsymbol{x})$ same as for ensemble (used for $\bar{\mathbf{H}}$).
- $p(\boldsymbol{x})$ must be Gaussian!

Does $\bar{\mathbf{H}}$ relate to $\mathcal{H}'$?

Theorem:

$$\lim_{N\to\infty} \bar{\mathbf{H}} \quad = \quad \mathbb{E}[\mathcal{H}'(\boldsymbol{x})]$$

$$\lim_{N\to\infty} \mathbf{Y}\mathbf{X}^{+} \qquad\qquad \mathbf{C}_{yx}\mathbf{C}_{x}^{-1}$$

$$= \lim_{N\to\infty} \mathbf{Y}\mathbf{X}^{\mathsf{T}}(\mathbf{X}\mathbf{X}^{\mathsf{T}})^{-1}$$

$$= \lim_{N\to\infty} \bar{\mathbf{C}}_{yx}\bar{\mathbf{C}}_{x}^{-1}$$

$$= \mathbf{C}_{yx}\mathbf{C}_{x}^{-1} \quad \text{(a.s., by Slutsky, sub. to reg.)}$$

I.e. $\bar{\mathbf{H}}$ is (indeed) the average derivative.

- $p(\boldsymbol{x})$ same as for ensemble (used for $\bar{\mathbf{H}}$).
- $p(\boldsymbol{x})$ must be Gaussian!

Does $\bar{\mathbf{H}}$ relate to $\mathcal{H}'$?

Theorem:

$$
\begin{array}{ccc}
\lim_{N\to\infty} \bar{\mathbf{H}} & = & \mathbb{E}[\mathcal{H}'(\boldsymbol{x})] \\
\parallel & & \parallel \\
\lim_{N\to\infty} \mathbf{Y}\mathbf{X}^{+} & & \mathbf{C}_{yx}\mathbf{C}_{x}^{-1} \\
\parallel & & \\
\lim_{N\to\infty} \mathbf{Y}\mathbf{X}^{\mathsf{T}}(\mathbf{X}\mathbf{X}^{\mathsf{T}})^{-1} & & \\
\parallel & & \\
\lim_{N\to\infty} \bar{\mathbf{C}}_{yx}\bar{\mathbf{C}}_{x}^{-1} & & \\
\parallel & & \\
\mathbf{C}_{yx}\mathbf{C}_{x}^{-1} \quad \text{(a.s., by Slutsky, sub. to reg.)} & &
\end{array}
$$

I.e. $\bar{\mathbf{H}}$ is (indeed) the average derivative.

- $p(\boldsymbol{x})$ same as for ensemble (used for $\bar{\mathbf{H}}$).
- $p(\boldsymbol{x})$ must be Gaussian!

Does $\bar{\mathbf{H}}$ relate to $\mathcal{H}'$?

Theorem:

$$\begin{array}{ccc}
\lim_{N\to\infty} \bar{\mathbf{H}} & = & \mathbb{E}[\mathcal{H}'(\boldsymbol{x})] \\
\parallel & & \parallel \\
\lim_{N\to\infty} \mathbf{Y}\mathbf{X}^{+} & & \mathbf{C}_{yx}\mathbf{C}_{x}^{-1} \\
\parallel & & \\
\lim_{N\to\infty} \mathbf{Y}\mathbf{X}^{\top}(\mathbf{X}\mathbf{X}^{\top})^{-1} & & \\
\parallel & & \\
\lim_{N\to\infty} \bar{\mathbf{C}}_{yx}\bar{\mathbf{C}}_{x}^{-1} & & \\
\parallel & & \\
\mathbf{C}_{yx}\mathbf{C}_{x}^{-1} \quad \text{(a.s., by Slutsky, sub. to reg.)} & &
\end{array}$$

I.e. $\bar{\mathbf{H}}$ is (indeed) the average derivative.

- $p(\boldsymbol{x})$ same as for ensemble (used for $\bar{\mathbf{H}}$).
- $p(\boldsymbol{x})$ must be Gaussian!

Does $\bar{\mathbf{H}}$ relate to $\mathcal{H}'$?

Theorem:

$$\lim_{N\to\infty} \bar{\mathbf{H}} = \mathbb{E}[\mathcal{H}'(\boldsymbol{x})]$$

$$\lim_{N\to\infty} \bar{\mathbf{H}} = \lim_{N\to\infty} \mathbf{Y}\mathbf{X}^{+} = \lim_{N\to\infty} \mathbf{Y}\mathbf{X}^{\top}(\mathbf{X}\mathbf{X}^{\top})^{-1} = \lim_{N\to\infty} \bar{\mathbf{C}}_{yx}\bar{\mathbf{C}}_{x}^{-1} = \mathbf{C}_{yx}\mathbf{C}_{x}^{-1} \quad \text{(a.s., by Slutsky, sub. to reg.)}$$

$$\mathbb{E}[\mathcal{H}'(\boldsymbol{x})] = \mathbf{C}_{yx}\mathbf{C}_{x}^{-1}$$

I.e. $\bar{\mathbf{H}}$ is (indeed) the average derivative.

- $p(\boldsymbol{x})$ same as for ensemble (used for $\bar{\mathbf{H}}$).
- $p(\boldsymbol{x})$ must be Gaussian!

Does $\bar{\mathbf{H}}$ relate to $\mathcal{H}'$?

Theorem:

$$
\begin{array}{ccc}
\lim_{N\to\infty} \bar{\mathbf{H}} & = & \mathbb{E}[\mathcal{H}'(\boldsymbol{x})] \\
\| & & \| \\
\lim_{N\to\infty} \mathbf{Y}\mathbf{X}^{+} & & \mathbf{C}_{yx}\mathbf{C}_x^{-1} \\
\| & & \\
\lim_{N\to\infty} \mathbf{Y}\mathbf{X}^{\mathsf{T}}(\mathbf{X}\mathbf{X}^{\mathsf{T}})^{-1} & & \text{(by Stein/IBP)} \\
\| & & \\
\lim_{N\to\infty} \bar{\mathbf{C}}_{yx}\bar{\mathbf{C}}_x^{-1} & & \\
\| & & \\
\mathbf{C}_{yx}\mathbf{C}_x^{-1} \quad \text{(a.s., by Slutsky, sub. to reg.)} & &
\end{array}
$$

I.e. $\bar{\mathbf{H}}$ is (indeed) the average derivative.

- $p(\boldsymbol{x})$ same as for ensemble (used for $\bar{\mathbf{H}}$).
- $p(\boldsymbol{x})$ must be Gaussian!

Does $\bar{\mathbf{H}}$ relate to $\mathcal{H}'$?

Theorem:

$$
\begin{array}{ccc}
\lim\limits_{N\to\infty} \bar{\mathbf{H}} & = & \mathbb{E}[\mathcal{H}'(\boldsymbol{x})] \\
\parallel & & \parallel \\
\lim\limits_{N\to\infty} \mathbf{Y}\mathbf{X}^{+} & & \mathbf{C}_{\boldsymbol{y}\boldsymbol{x}}\mathbf{C}_{\boldsymbol{x}}^{-1} \\
\parallel & & \\
\lim\limits_{N\to\infty} \mathbf{Y}\mathbf{X}^{\mathsf{T}}(\mathbf{X}\mathbf{X}^{\mathsf{T}})^{-1} & & \text{(by Stein/IBP)} \\
\parallel & & \\
\lim\limits_{N\to\infty} \bar{\mathbf{C}}_{\boldsymbol{y}\boldsymbol{x}}\bar{\mathbf{C}}_{\boldsymbol{x}}^{-1} & & \\
\parallel & & \\
\mathbf{C}_{\boldsymbol{y}\boldsymbol{x}}\mathbf{C}_{\boldsymbol{x}}^{-1} \quad \text{(a.s., by Slutsky, sub. to reg.)} & &
\end{array}
$$

I.e. $\bar{\mathbf{H}}$ is (indeed) the average derivative.

- $p(\boldsymbol{x})$ same as for ensemble (used for $\bar{\mathbf{H}}$).
- $p(\boldsymbol{x})$ must be Gaussian!

Does $\bar{\mathbf{H}}$ relate to $\mathcal{H}'$?

Theorem:

$$
\begin{array}{lcl}
\lim_{N\to\infty} \bar{\mathbf{H}} & = & \mathbb{E}[\mathcal{H}'(\boldsymbol{x})] \\
\shortparallel & & \shortparallel \\
\lim_{N\to\infty} \mathbf{Y}\mathbf{X}^{+} & & \mathbf{C}_{\boldsymbol{yx}}\mathbf{C}_{\boldsymbol{x}}^{-1} \\
\shortparallel & & \\
\lim_{N\to\infty} \mathbf{Y}\mathbf{X}^{\mathsf{T}}(\mathbf{X}\mathbf{X}^{\mathsf{T}})^{-1} & & \text{(by Stein/IBP)} \\
\shortparallel & & \\
\lim_{N\to\infty} \bar{\mathbf{C}}_{\boldsymbol{yx}}\bar{\mathbf{C}}_{\boldsymbol{x}}^{-1} & & \\
\shortparallel & & \\
\mathbf{C}_{\boldsymbol{yx}}\mathbf{C}_{\boldsymbol{x}}^{-1} \quad \text{(a.s., by Slutsky, sub. to reg.)} & &
\end{array}
$$

I.e. $\bar{\mathbf{H}}$ is (indeed) the average derivative.

- $p(\boldsymbol{x})$ same as for ensemble (used for $\bar{\mathbf{H}}$).
- $p(\boldsymbol{x})$ must be Gaussian!

Does $\bar{\mathbf{H}}$ relate to $\mathcal{H}'$?

Theorem:

$$\lim_{N\to\infty} \bar{\mathbf{H}} \quad = \quad \mathbb{E}[\mathcal{H}'(\boldsymbol{x})]$$

$$\lim_{N\to\infty} \bar{\mathbf{H}} = \lim_{N\to\infty} \mathbf{Y}\mathbf{X}^{+} = \lim_{N\to\infty} \mathbf{Y}\mathbf{X}^{\mathsf{T}}(\mathbf{X}\mathbf{X}^{\mathsf{T}})^{-1} = \lim_{N\to\infty} \bar{\mathbf{C}}_{\boldsymbol{yx}}\bar{\mathbf{C}}_{\boldsymbol{x}}^{-1} = \mathbf{C}_{\boldsymbol{yx}}\mathbf{C}_{\boldsymbol{x}}^{-1}$$

(a.s., by Slutsky, sub. to reg.)

$$\mathbb{E}[\mathcal{H}'(\boldsymbol{x})] = \mathbf{C}_{\boldsymbol{yx}}\mathbf{C}_{\boldsymbol{x}}^{-1}$$

(by Stein/IBP)

I.e. $\bar{\mathbf{H}}$ is (indeed) the average derivative.

- $p(\boldsymbol{x})$ same as for ensemble (used for $\bar{\mathbf{H}}$).
- $p(\boldsymbol{x})$ must be Gaussian!