DS-GA 3001.008 Modelling time series data

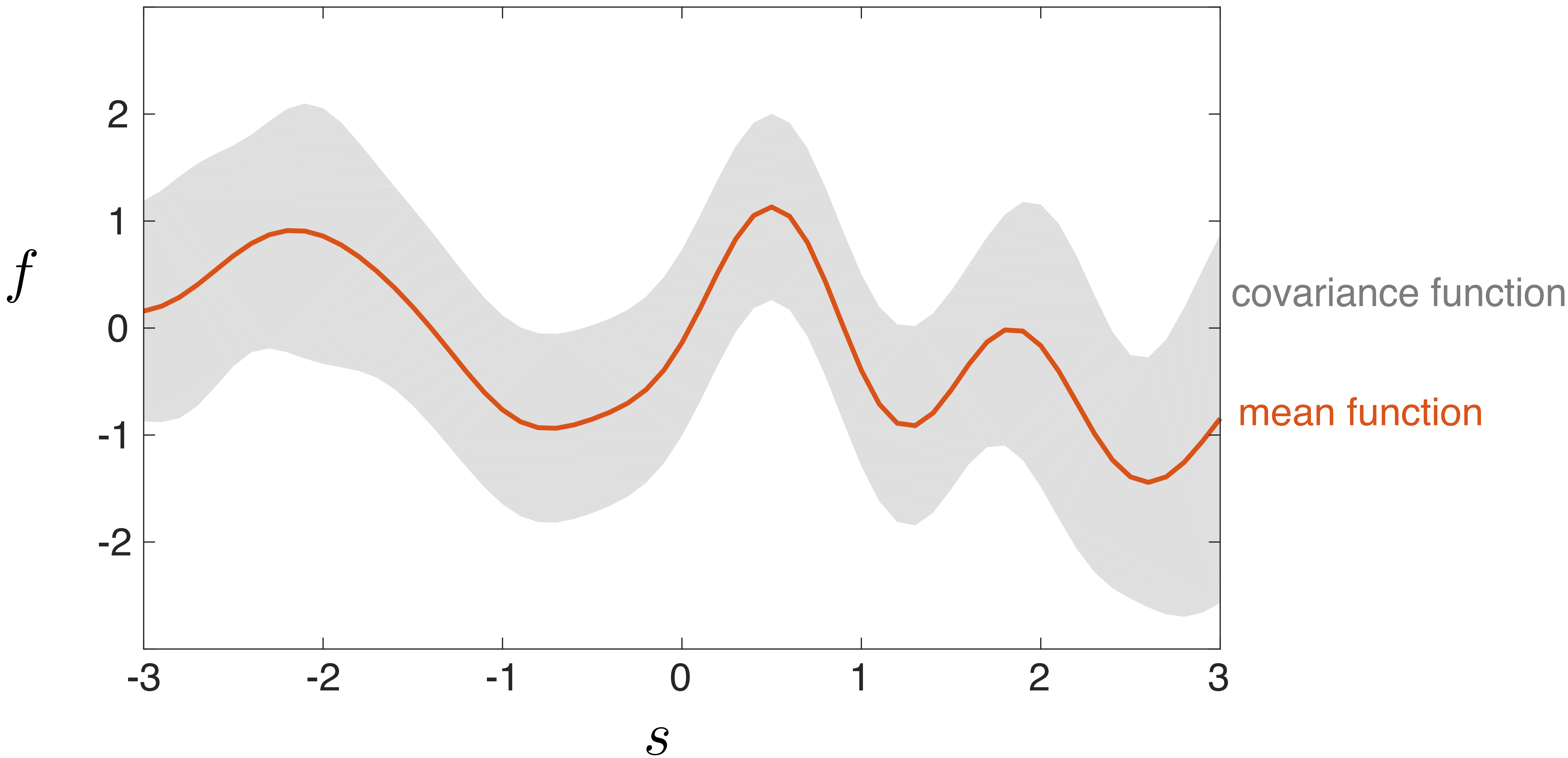
L9. An intro to Gaussian Processes

Instructor: Cristina Savin

NYU, CNS & CDS

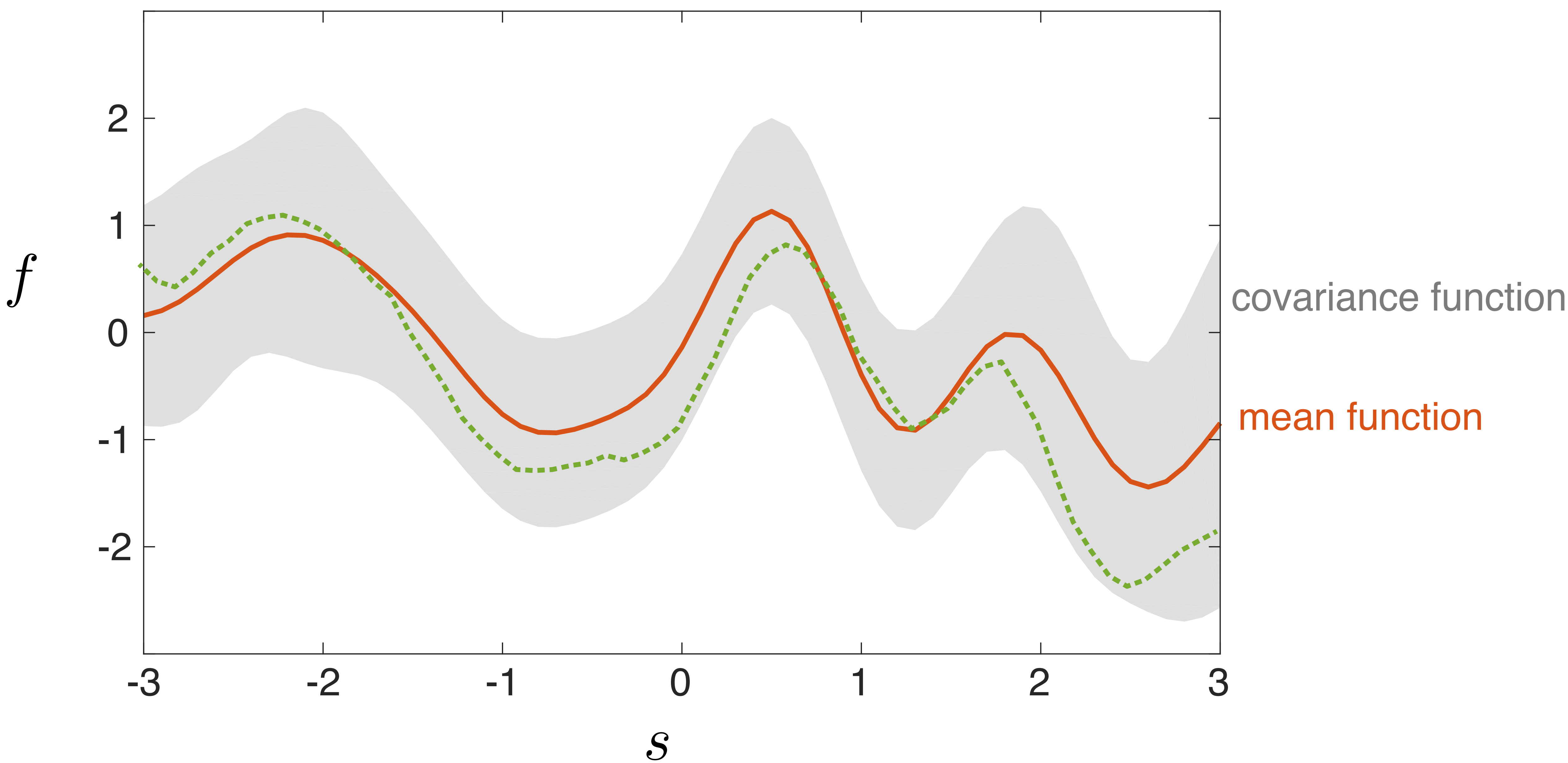
Probabilistic inference with functions

$$f \sim \mathrm{GP}\left(f; \mu(s), k(s_1, s_2)\right)$$



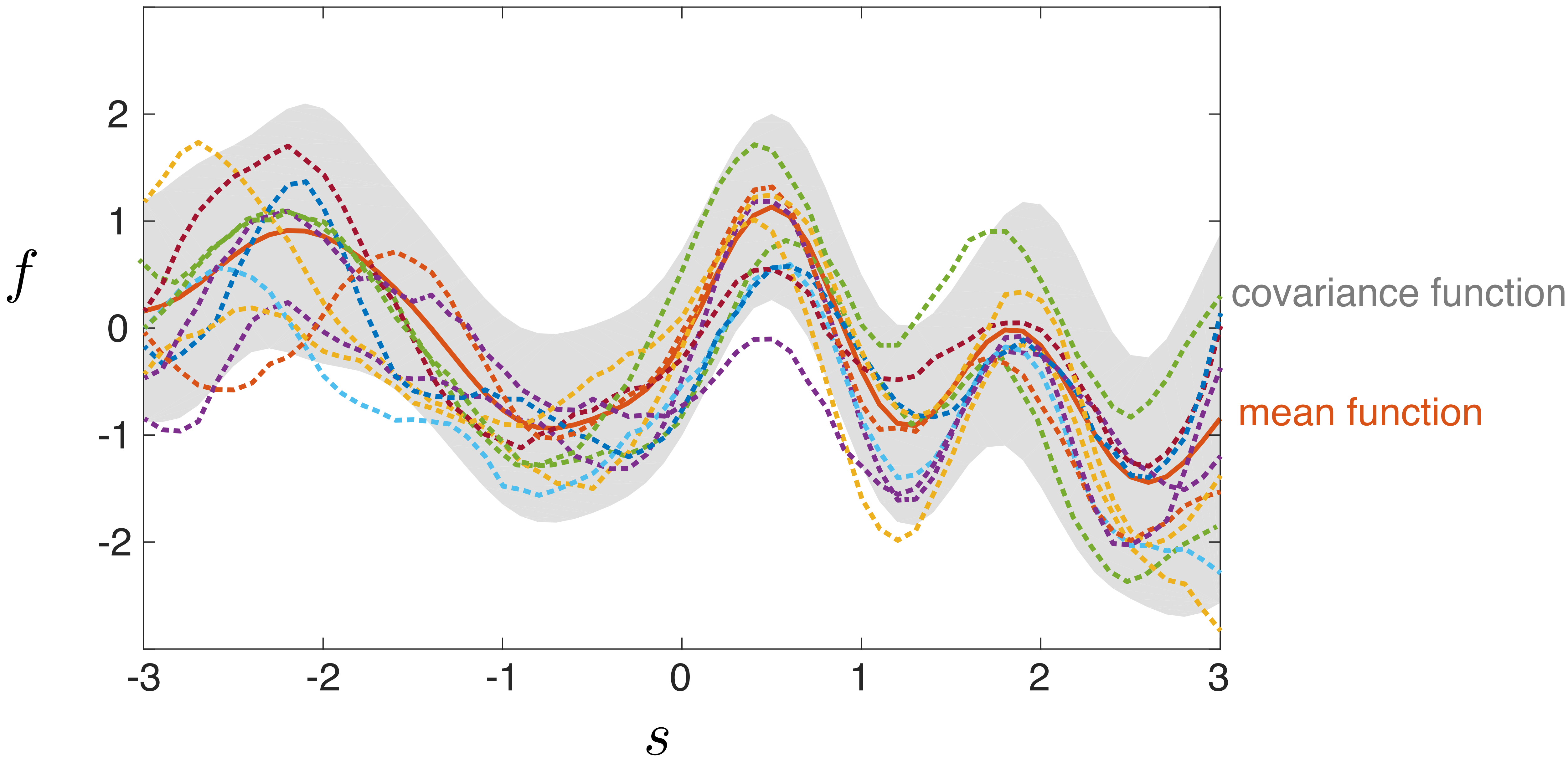
Probabilistic inference with functions

$$f \sim \mathrm{GP}\left(f; \mu(s), k(s_1, s_2)\right)$$



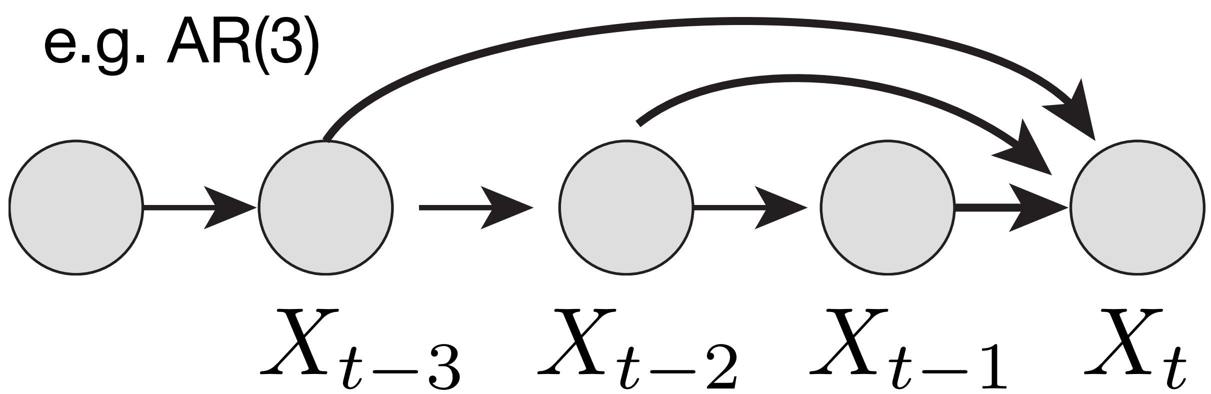
Probabilistic inference with functions

$$f \sim \mathrm{GP}\left(f; \mu(s), k(s_1, s_2)\right)$$



Big picture

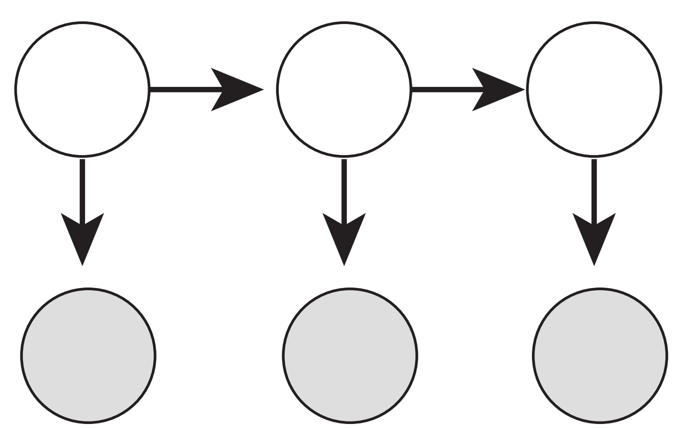
ARIMA: directly model data statistics



simple dependencies

linear prediction

Latent state models: introduce latent variables



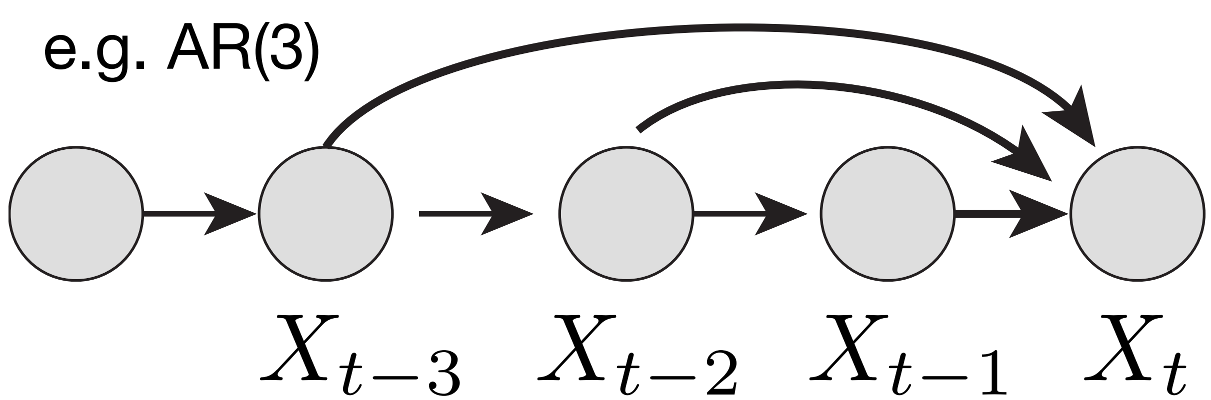
more complex dependencies

Markov dynamics in latent space

more complicated prediction

Big picture

ARIMA: directly model data statistics

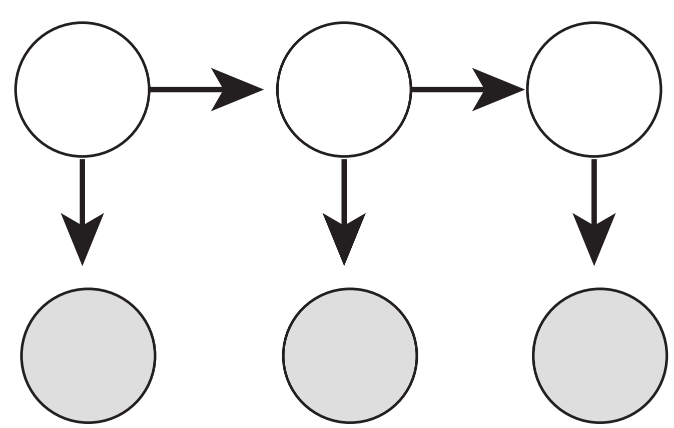


simple dependencies

linear prediction

GP: complex functional dependencies

Latent state models: introduce latent variables



more complex dependencies

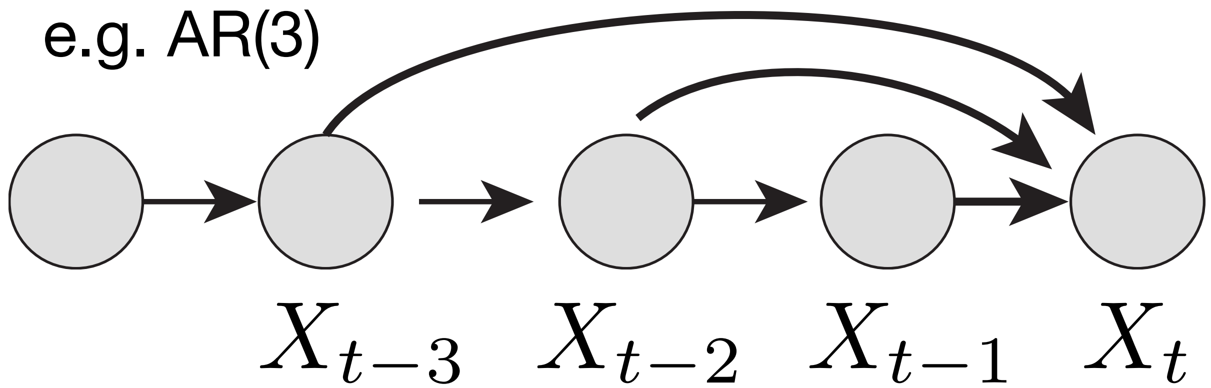
Markov dynamics in latent space

more complicated prediction

Big picture

ARIMA: directly model data statistics

e.g. AR(3)

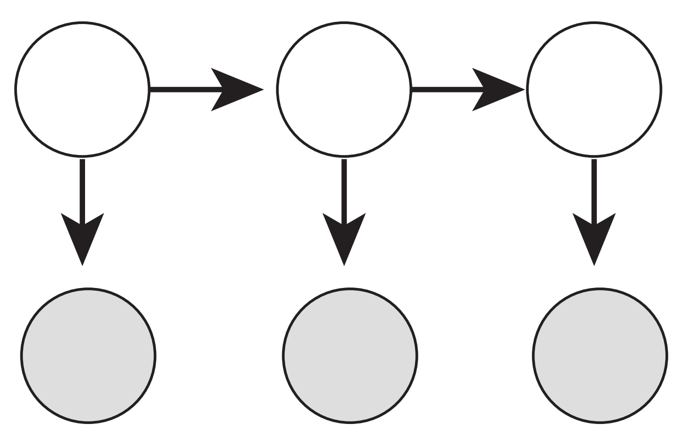


simple dependencies

linear prediction

GP: complex functional dependencies

Latent state models: introduce latent variables



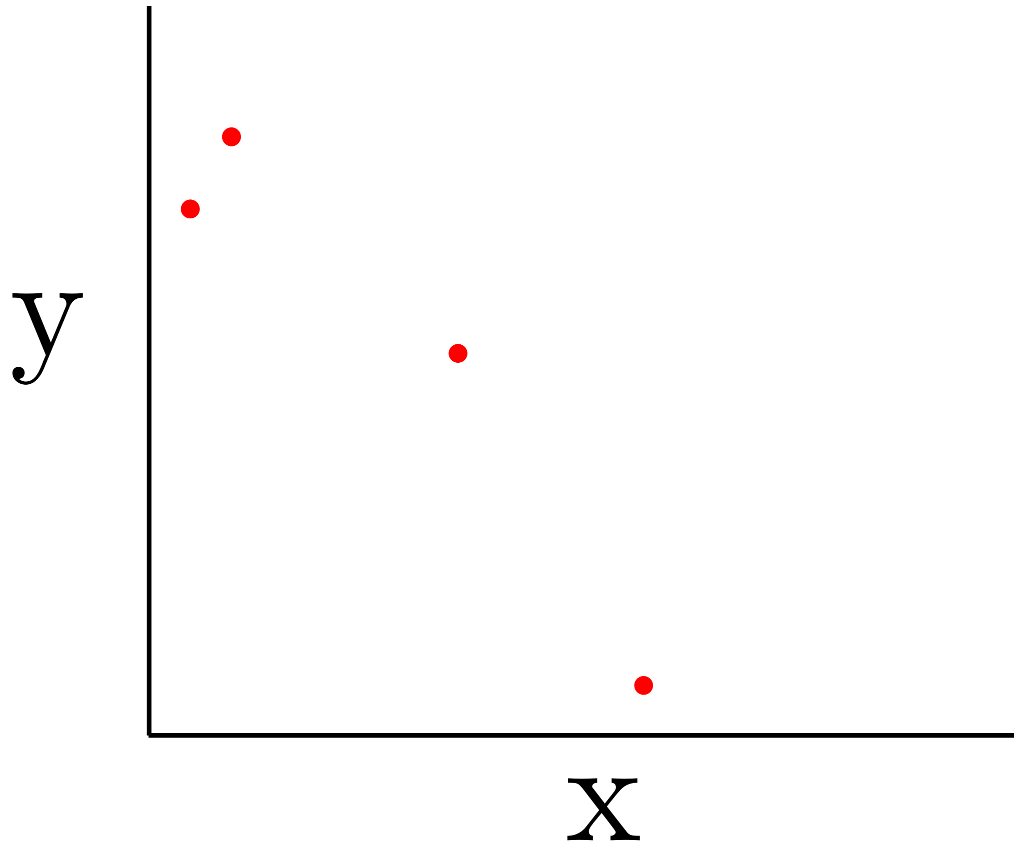
more complex dependencies

Markov dynamics in latent space

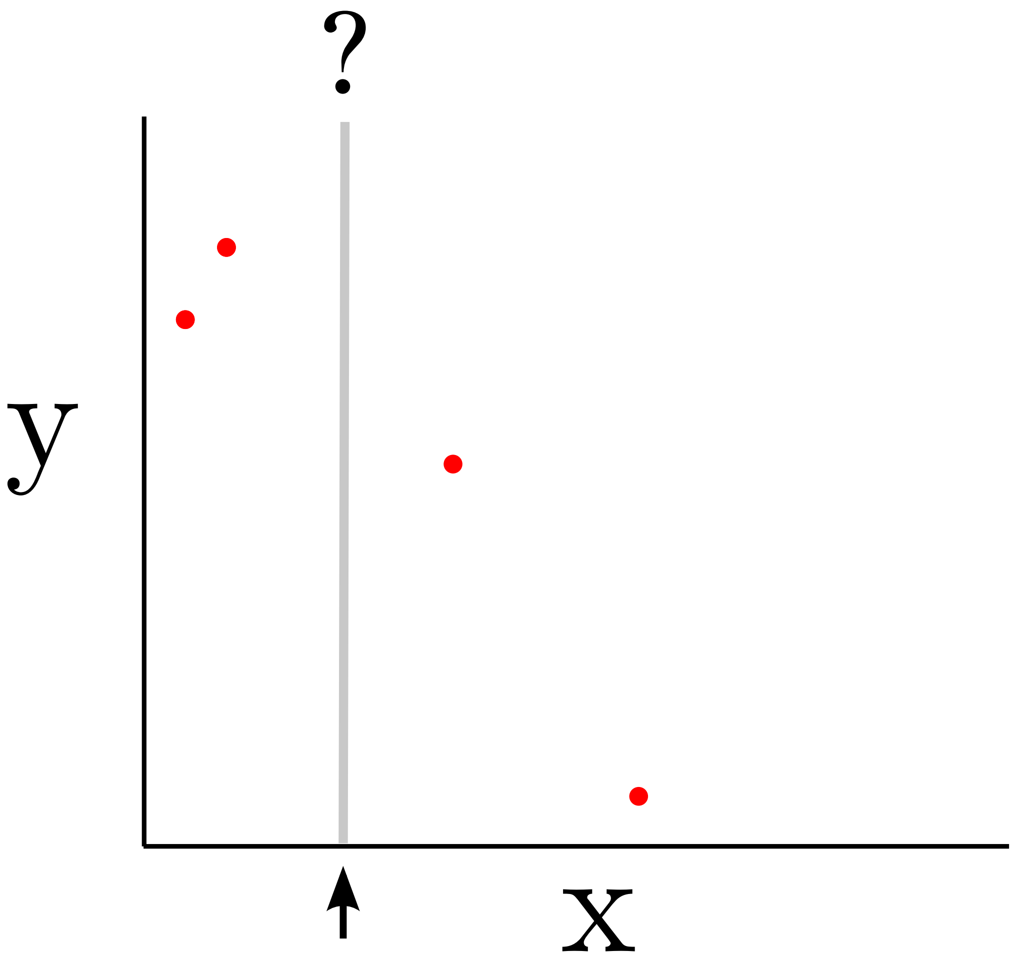
more complicated prediction

GP: complex latent dynamics

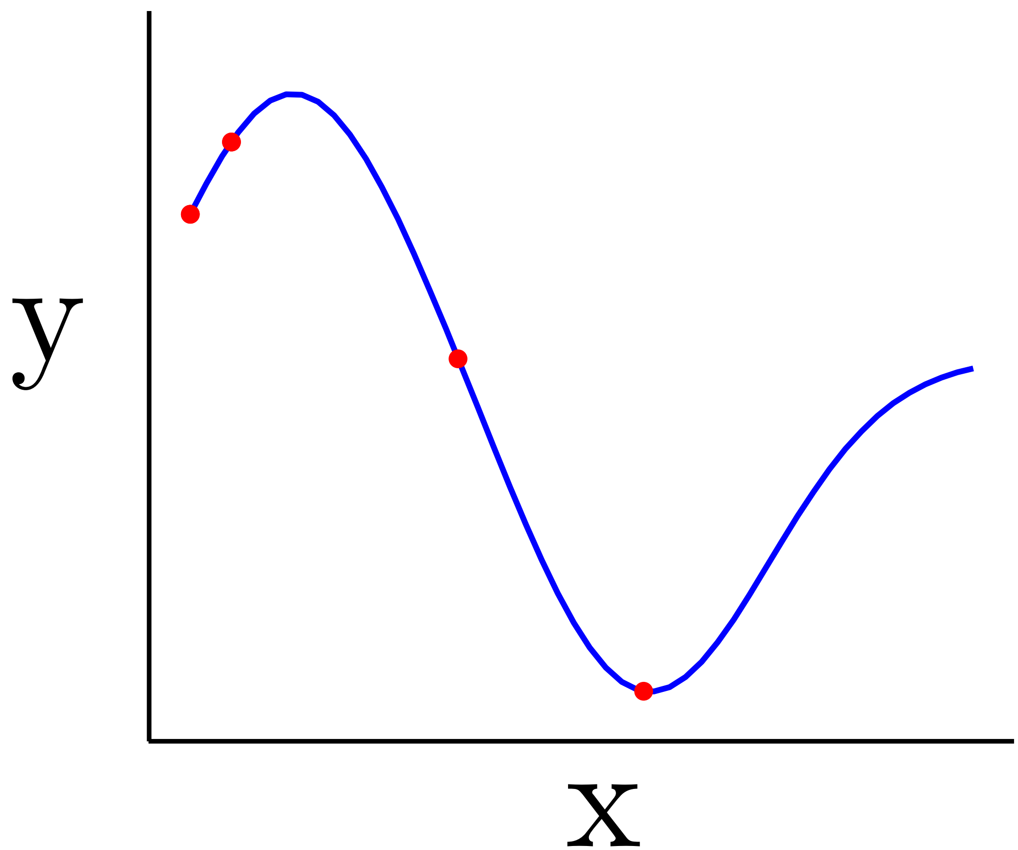
Motivation: non-linear regression



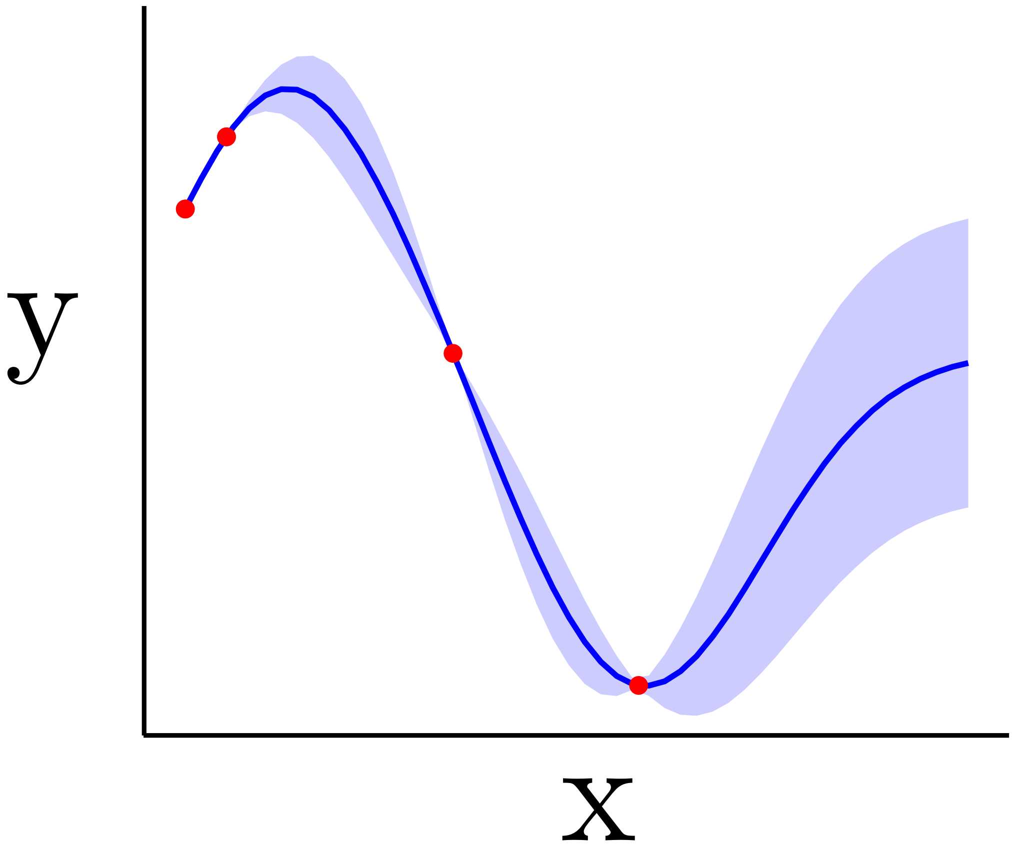
Motivation: non-linear regression



Motivation: non-linear regression

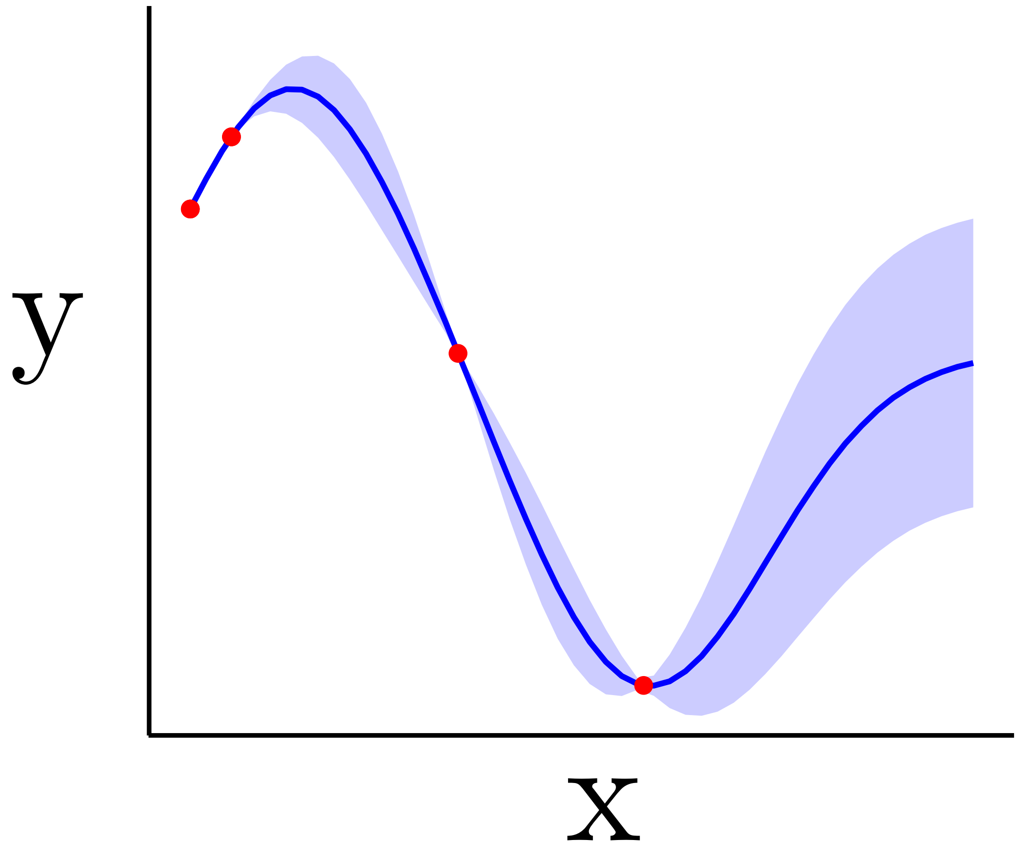


Motivation: non-linear regression



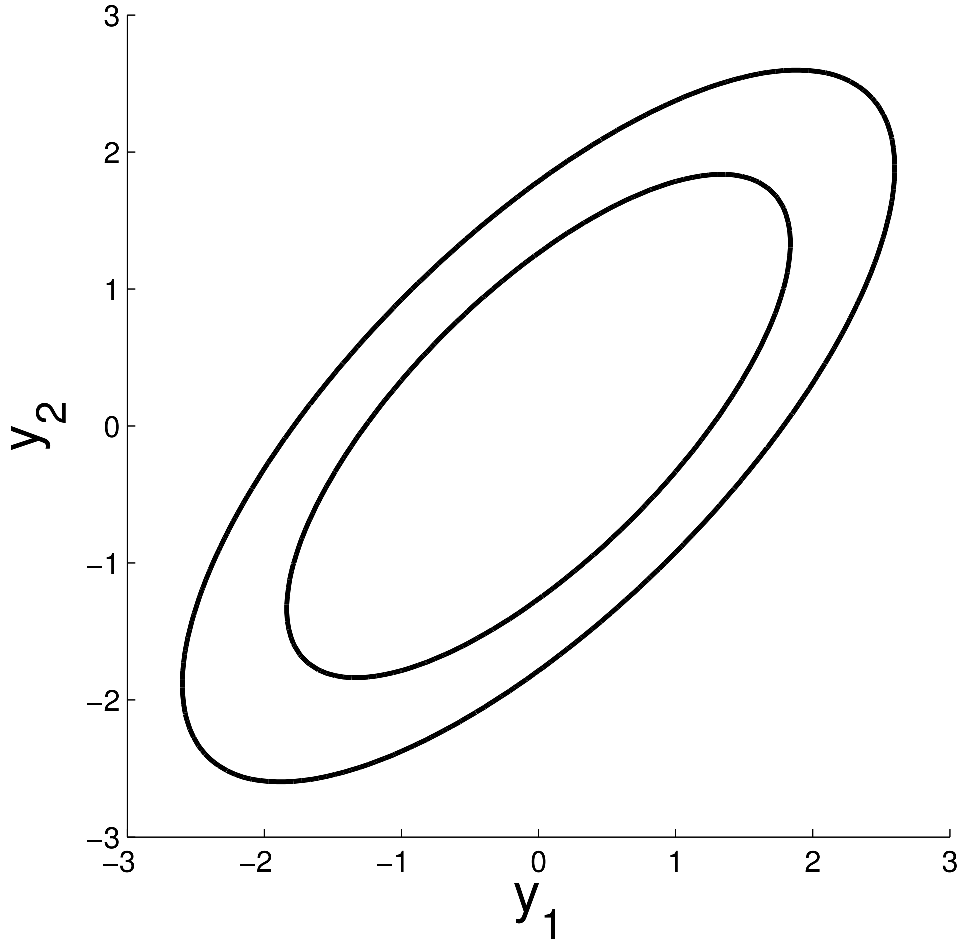
Motivation: non-linear regression

Can we do this with a plain old Gaussian?



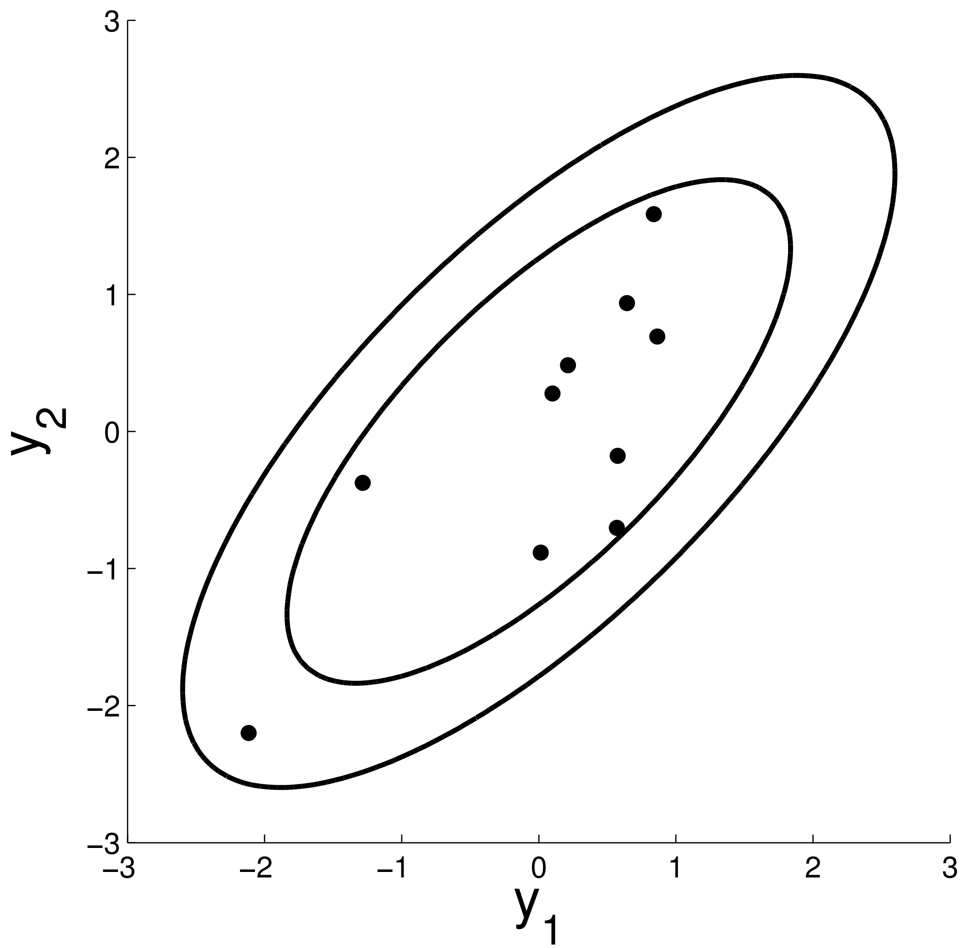
Gaussian distribution

$$p(\mathbf{y}|\Sigma) \propto \exp\left(-\tfrac{1}{2}\mathbf{y}^\mathsf{T}\Sigma^{-1}\mathbf{y}\right) \qquad \Sigma = \begin{bmatrix} 1 & .7 \\ .7 & 1 \end{bmatrix}$$



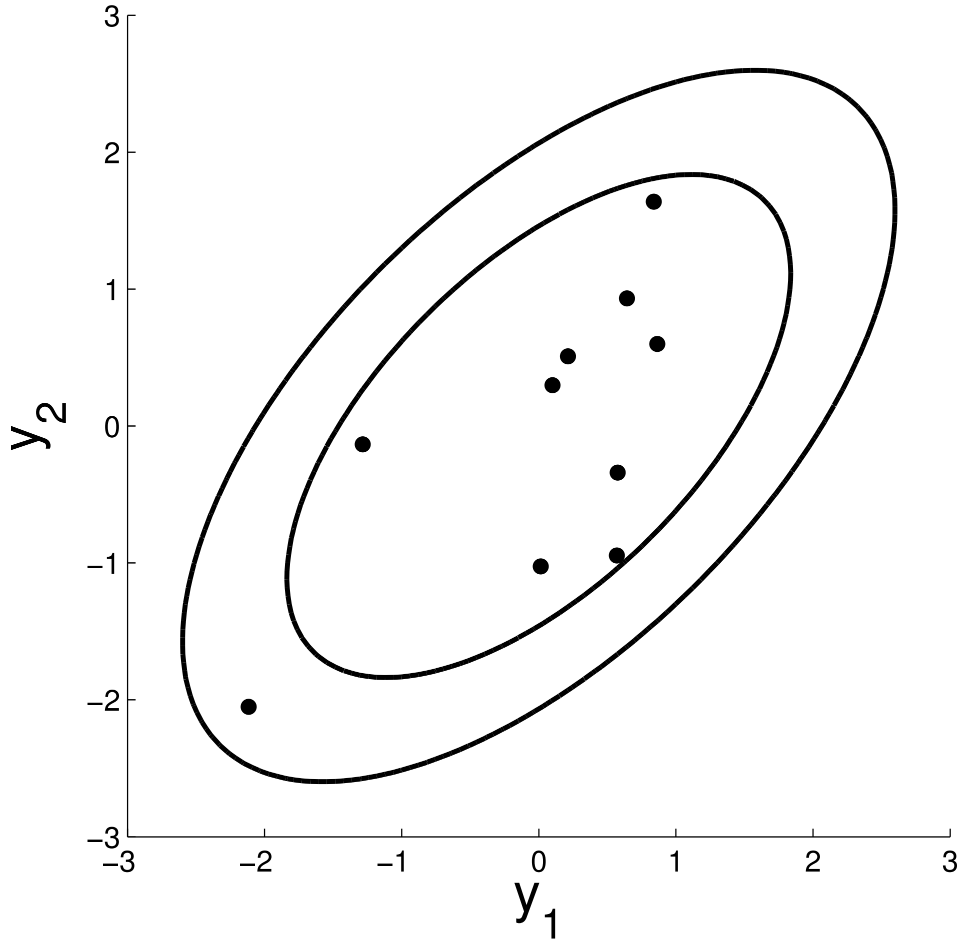
Gaussian distribution

$$p(\mathbf{y}|\Sigma) \propto \exp\left(-\tfrac{1}{2}\mathbf{y}^\mathsf{T}\Sigma^{-1}\mathbf{y}\right) \qquad \Sigma = \begin{bmatrix} 1 & .7 \\ .7 & 1 \end{bmatrix}$$



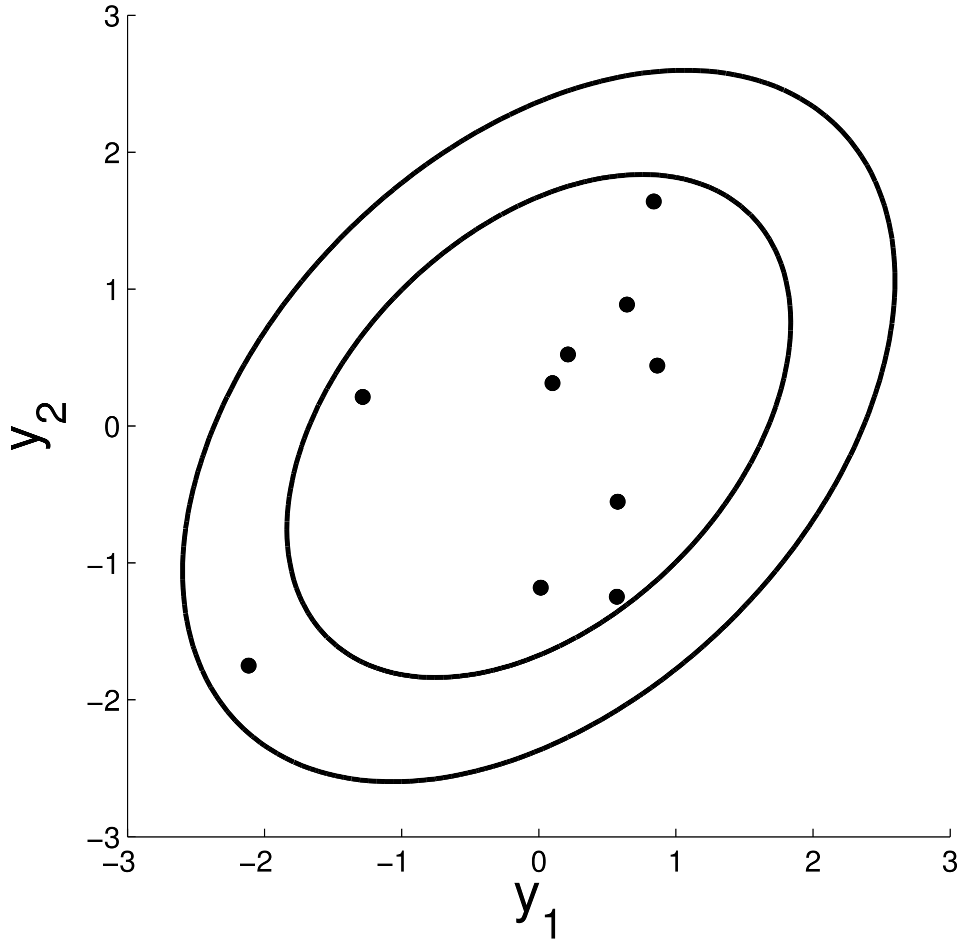
Gaussian distribution

$$p(\mathbf{y}|\Sigma) \propto \exp\left(-\tfrac{1}{2}\mathbf{y}^\intercal\Sigma^{-1}\mathbf{y}\right) \qquad \Sigma = \begin{bmatrix} 1 & .6 \\ .6 & 1 \end{bmatrix}$$



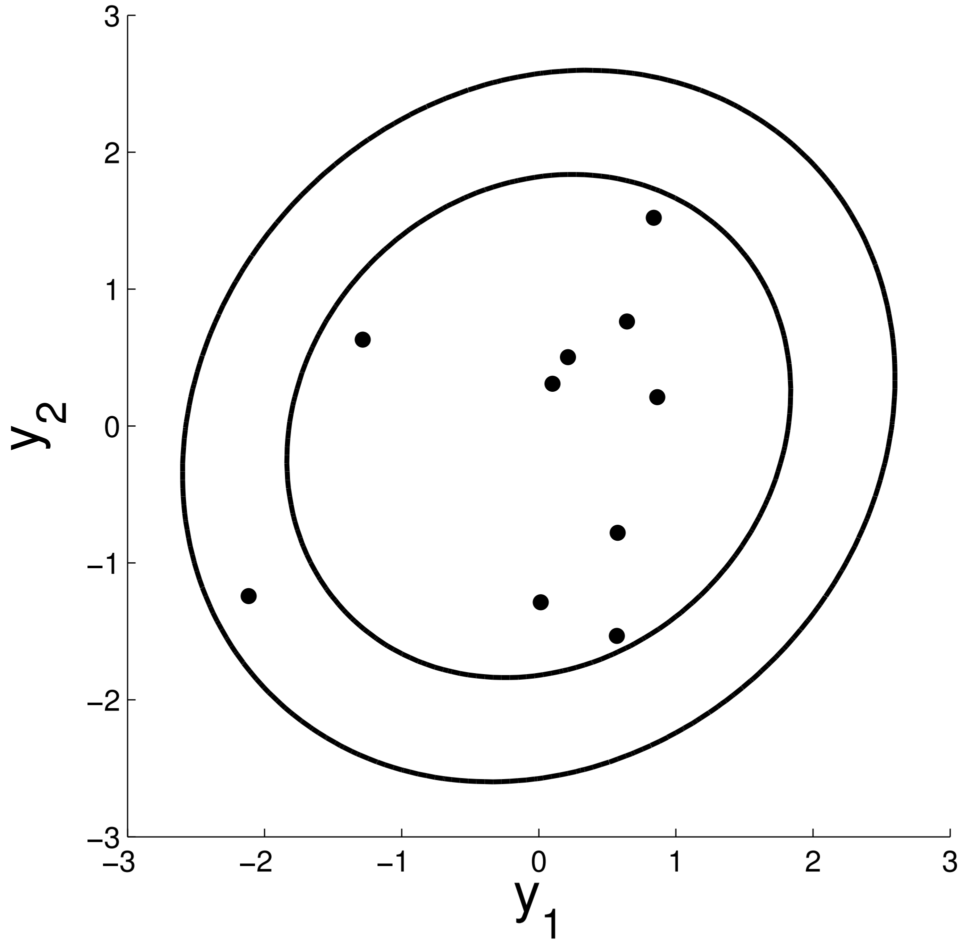
Gaussian distribution

$$p(\mathbf{y}|\Sigma) \propto \exp\left(-\tfrac{1}{2}\mathbf{y}^{\intercal}\Sigma^{-1}\mathbf{y}\right) \qquad \Sigma = \begin{bmatrix} 1 & .4 \\ .4 & 1 \end{bmatrix}$$



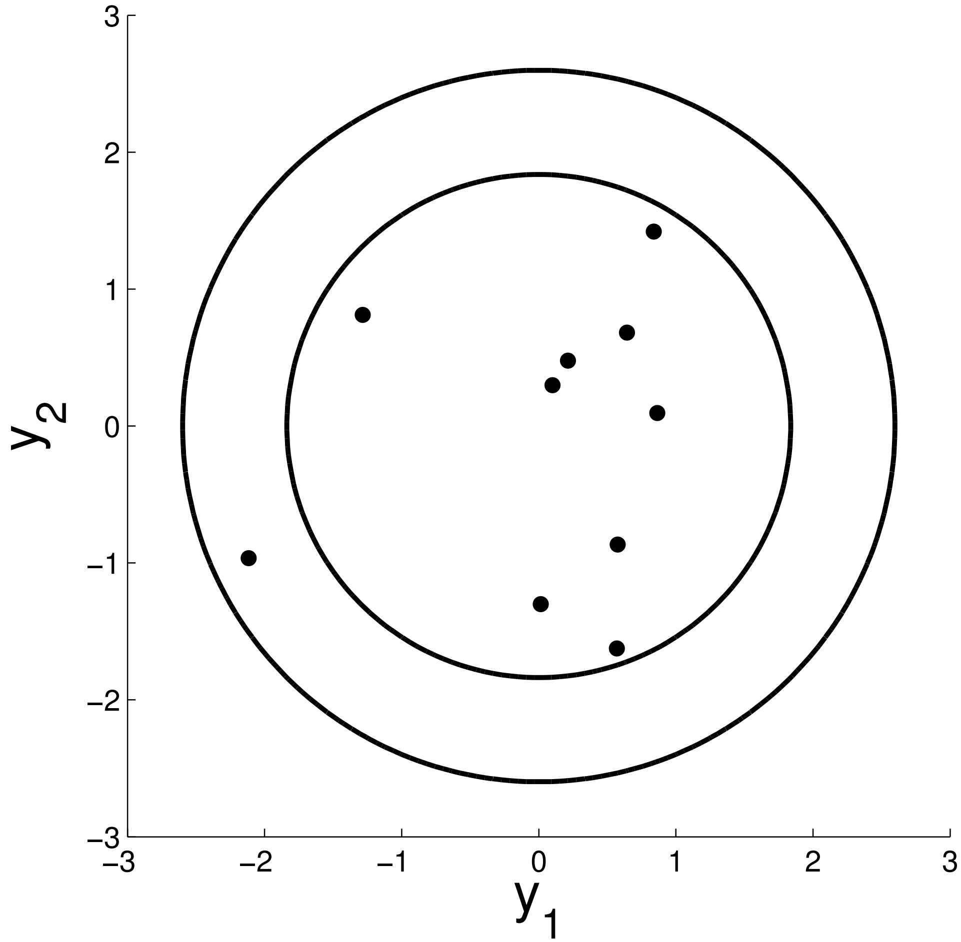
Gaussian distribution

$$p(\mathbf{y}|\Sigma) \propto \exp\left(-\tfrac{1}{2}\mathbf{y}^{\mathsf{T}}\Sigma^{-1}\mathbf{y}\right) \qquad \Sigma = \begin{bmatrix} 1 & .1 \\ .1 & 1 \end{bmatrix}$$



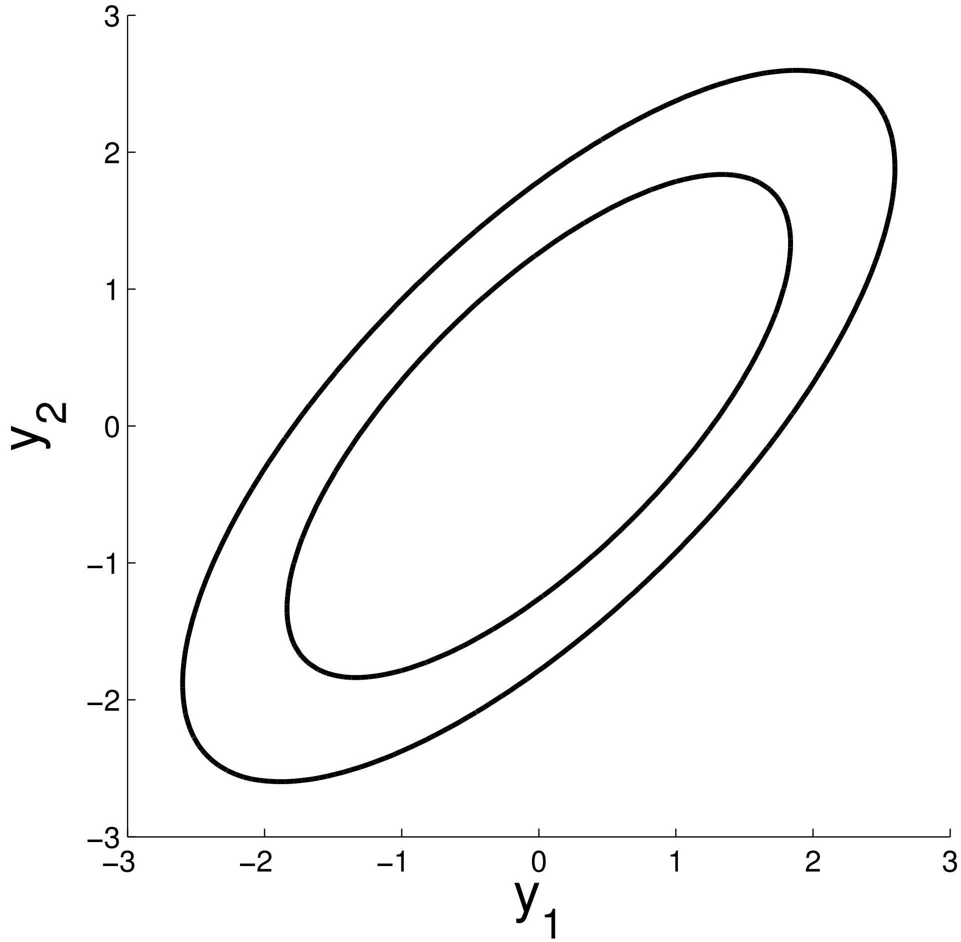
Gaussian distribution

$$p(\mathbf{y}|\Sigma) \propto \exp\left(-\frac{1}{2}\mathbf{y}^\top \Sigma^{-1} \mathbf{y}\right) \qquad \Sigma = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$



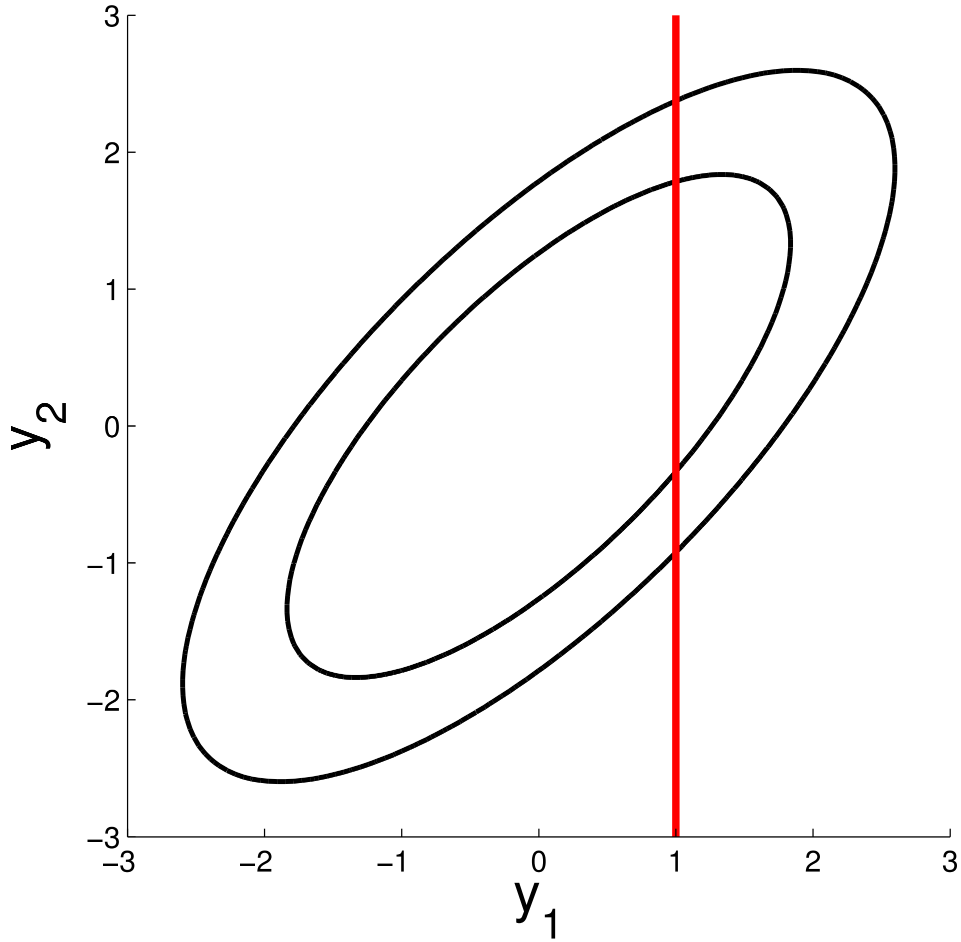
Gaussian distribution

$$p(\mathbf{y}|\Sigma) \propto \exp\left(-\tfrac{1}{2}\mathbf{y}^{\mathsf{T}}\Sigma^{-1}\mathbf{y}\right) \qquad \Sigma = \begin{bmatrix} 1 & .7 \\ .7 & 1 \end{bmatrix}$$



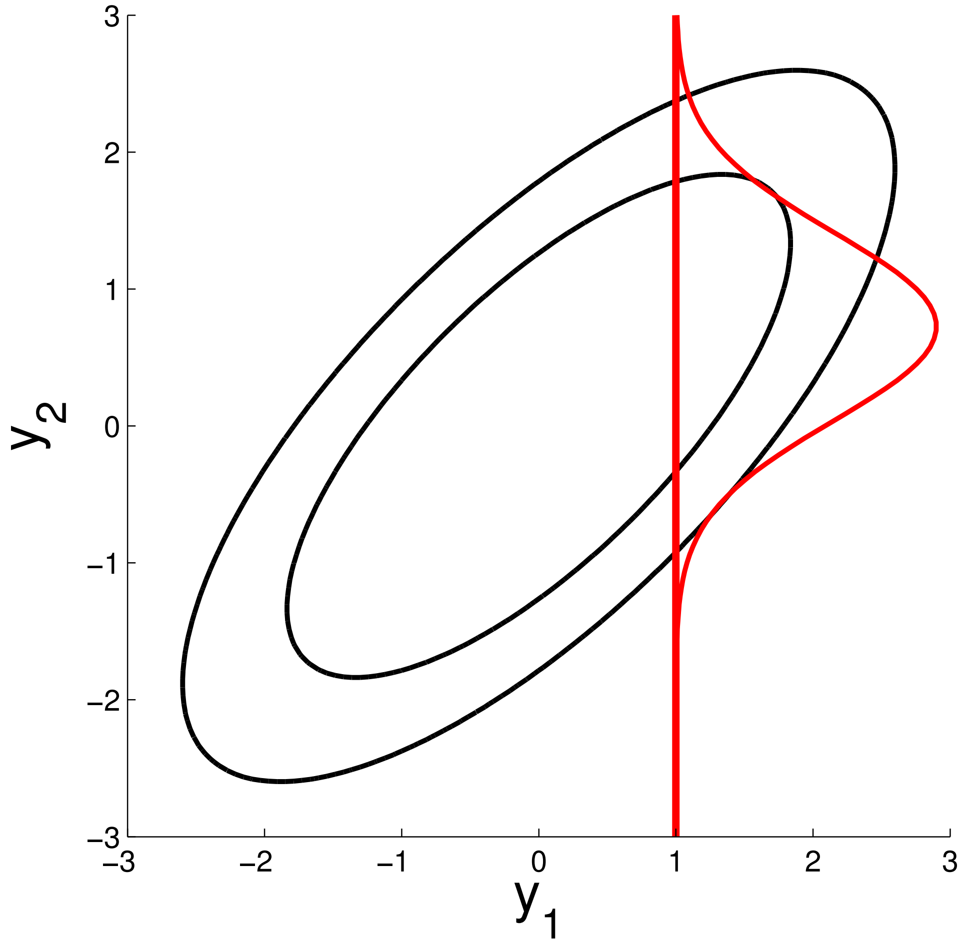
Gaussian distribution

$$p(\mathbf{y}|\Sigma) \propto \exp\left(-\frac{1}{2}\mathbf{y}^{\mathsf{T}}\Sigma^{-1}\mathbf{y}\right) \qquad \Sigma = \begin{bmatrix} 1 & .7 \\ .7 & 1 \end{bmatrix}$$



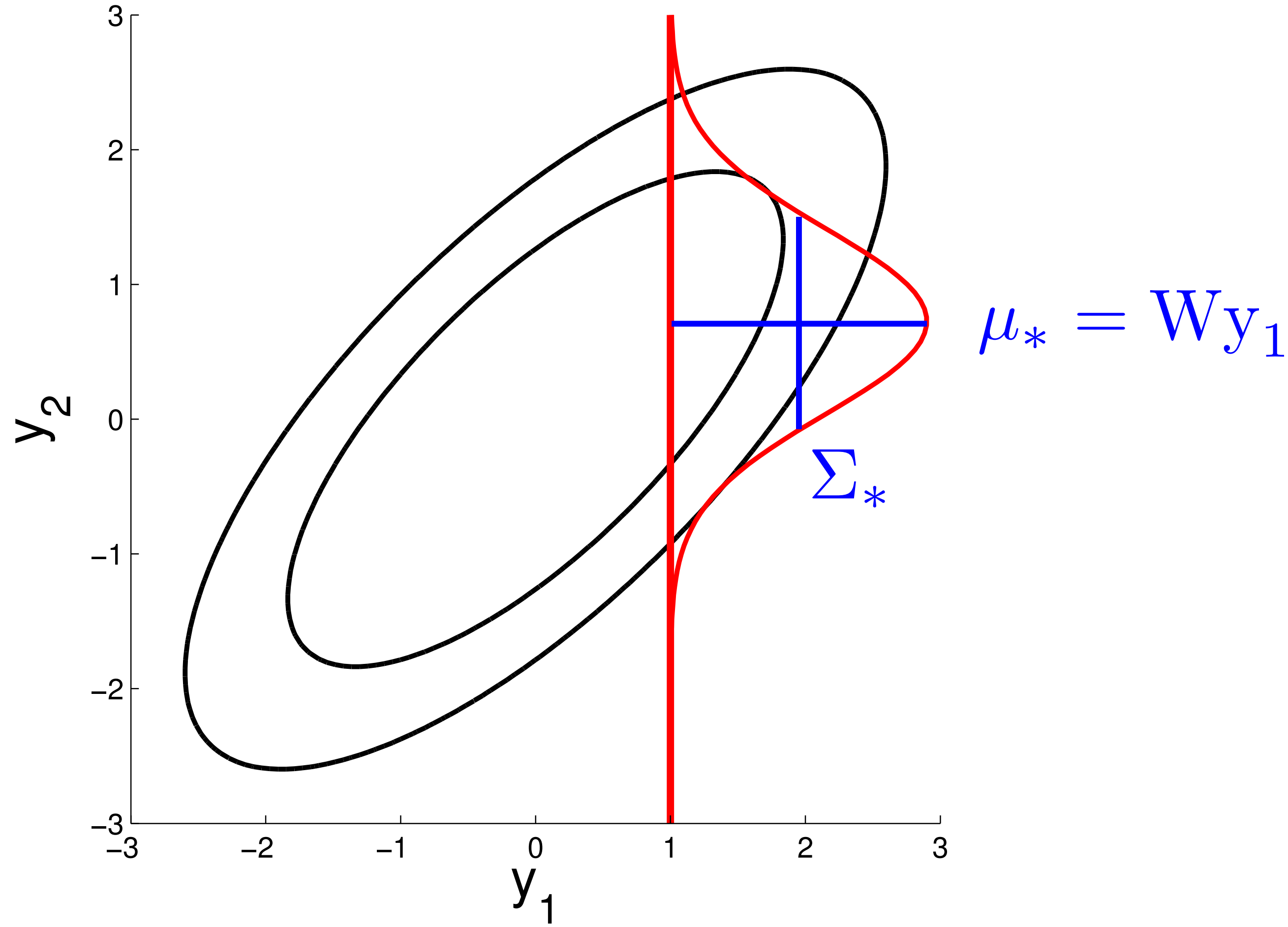
Gaussian distribution

$$p(\mathsf{y}_2|\mathsf{y}_1, \Sigma) \propto \exp\left(-\frac{1}{2}(\mathsf{y}_2 - \mu_*)\Sigma_*{}^{-1}(\mathsf{y}_2 - \mu_*)\right)$$



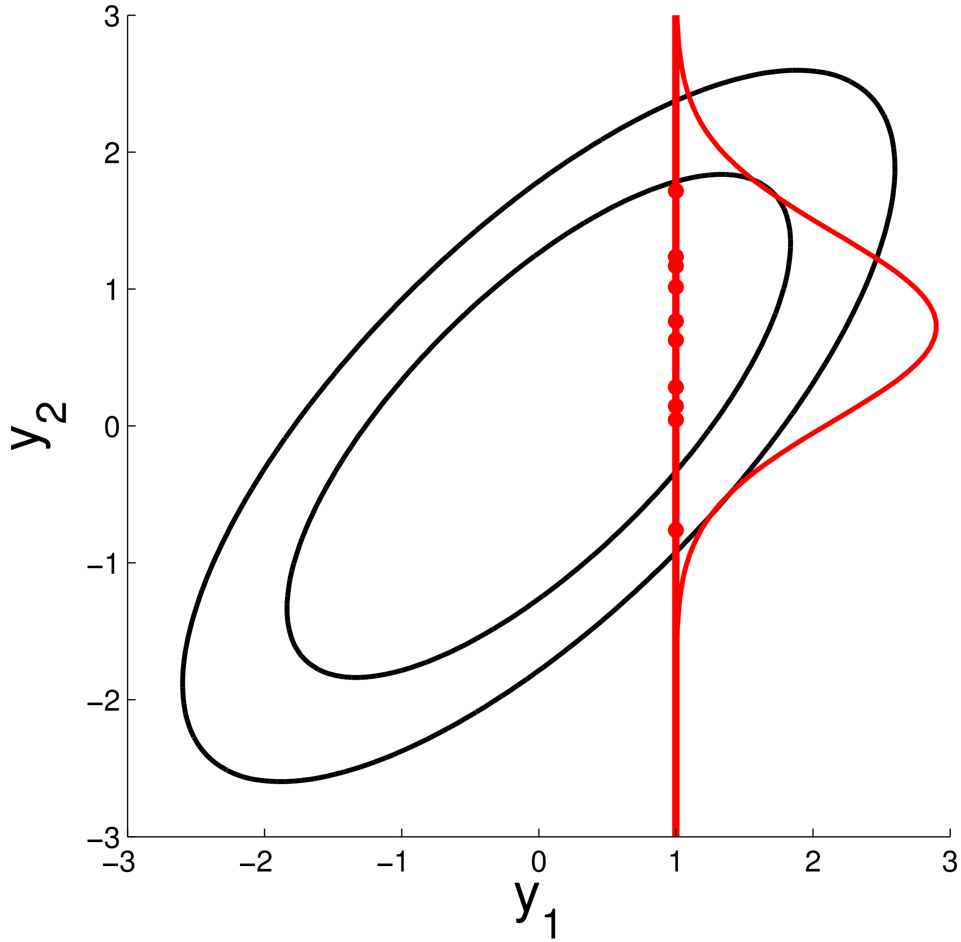
Gaussian distribution

$$p(\mathsf{y}_2|\mathsf{y}_1, \Sigma) \propto \exp\left(-\frac{1}{2}(\mathsf{y}_2 - \mu_*)\Sigma_*{}^{-1}(\mathsf{y}_2 - \mu_*)\right)$$



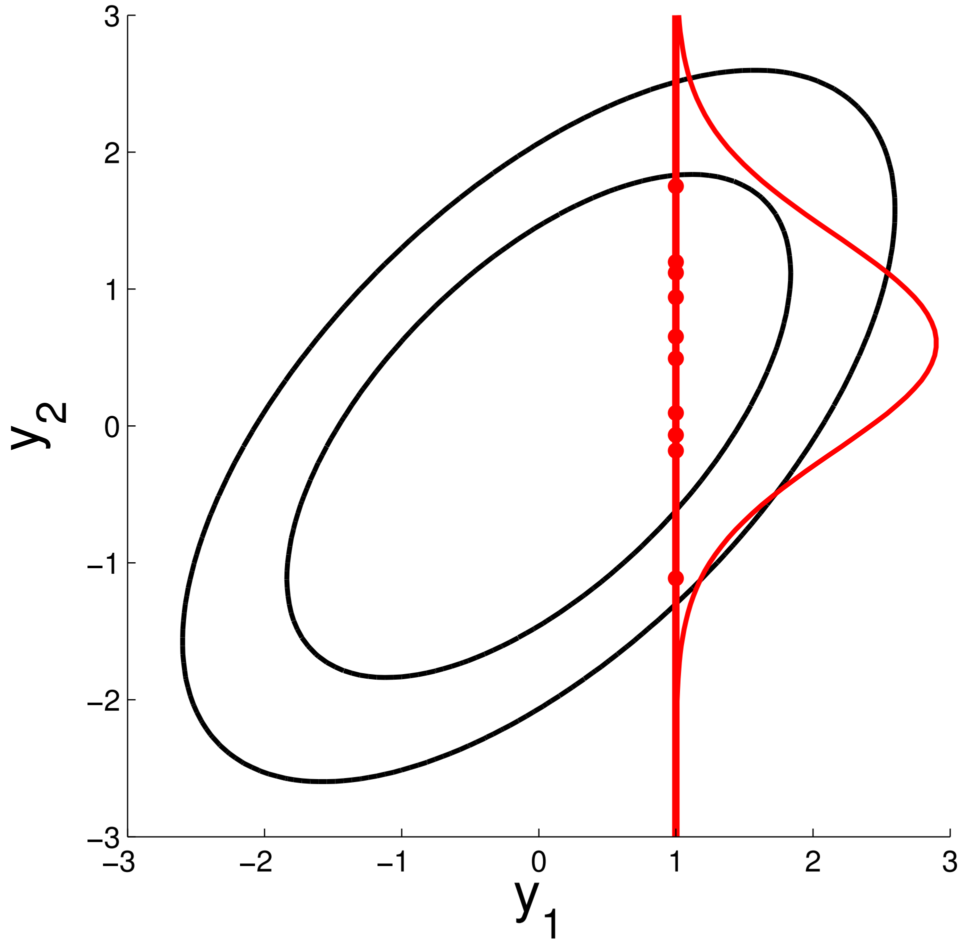
Gaussian distribution

$$p(\mathsf{y}_2|\mathsf{y}_1, \Sigma) \propto \exp\left(-\frac{1}{2}(\mathsf{y}_2 - \mu_*)\Sigma_*{}^{-1}(\mathsf{y}_2 - \mu_*)\right)$$



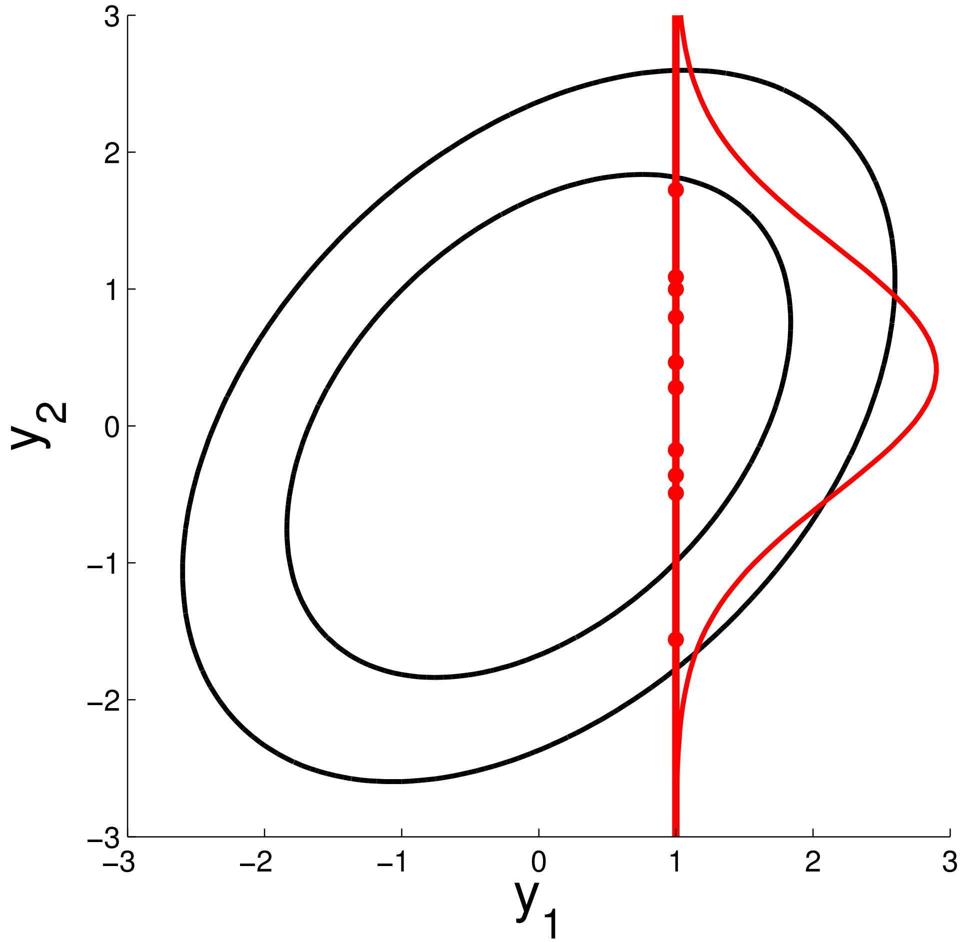
Gaussian distribution

$$p(\mathsf{y}_2|\mathsf{y}_1,\Sigma) \propto \exp\left(-\frac{1}{2}(\mathsf{y}_2 - \mu_*)\Sigma_*{}^{-1}(\mathsf{y}_2 - \mu_*)\right)$$



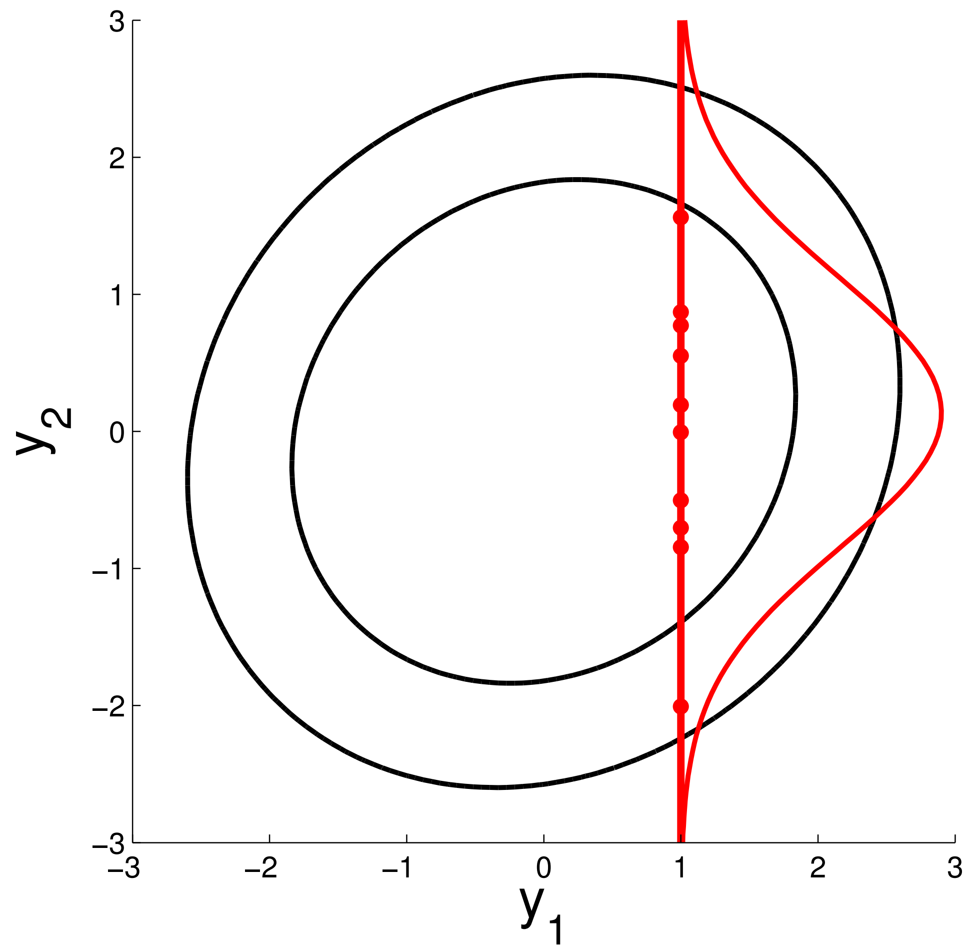
Gaussian distribution

$$p(\mathsf{y}_2|\mathsf{y}_1,\Sigma) \propto \exp\left(-\frac{1}{2}(\mathsf{y}_2 - \mu_*)\Sigma_*{}^{-1}(\mathsf{y}_2 - \mu_*)\right)$$



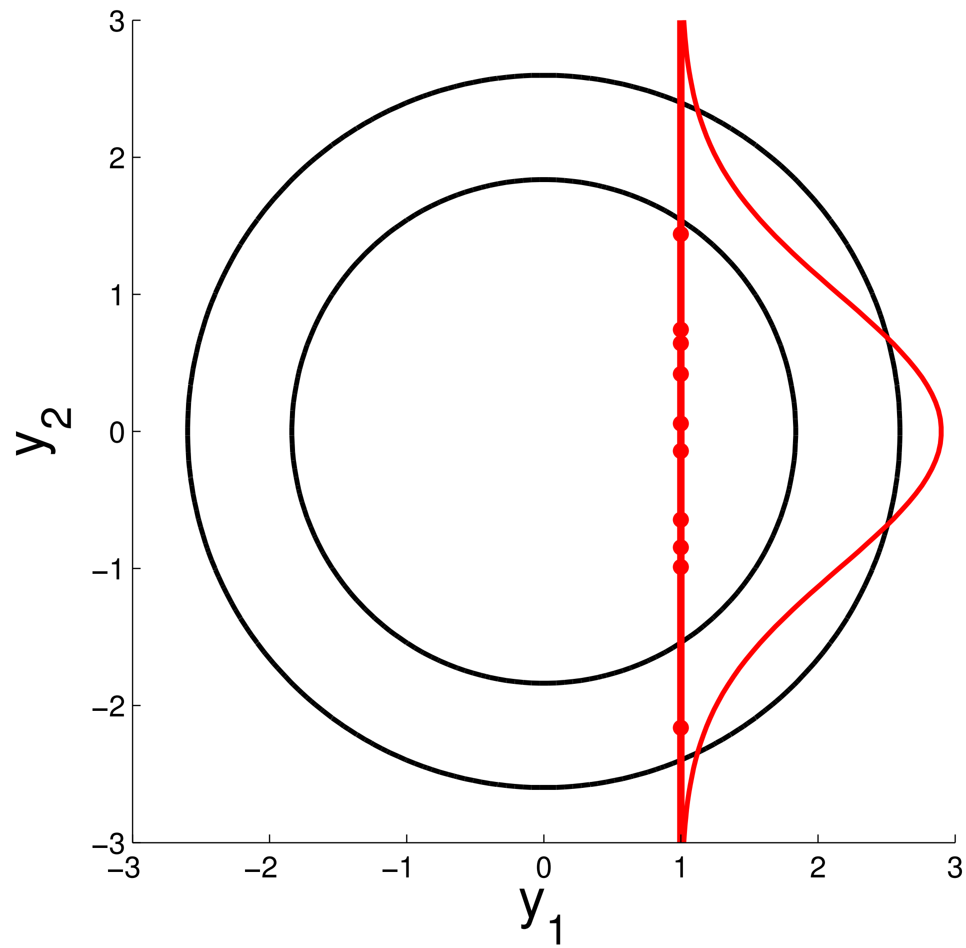
Gaussian distribution

$$p(\mathsf{y}_2|\mathsf{y}_1,\Sigma) \propto \exp\left(-\frac{1}{2}(\mathsf{y}_2 - \mu_*)\Sigma_*{}^{-1}(\mathsf{y}_2 - \mu_*)\right)$$

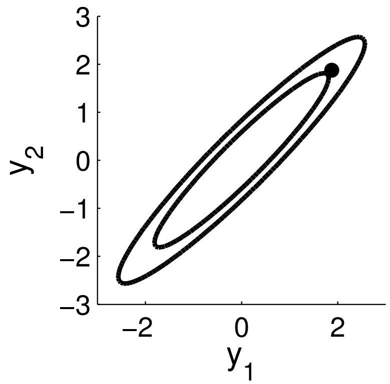


Gaussian distribution

$$p(\mathsf{y}_2|\mathsf{y}_1,\Sigma) \propto \exp\left(-\frac{1}{2}(\mathsf{y}_2-\mu_*)\Sigma_*{}^{-1}(\mathsf{y}_2-\mu_*)\right)$$

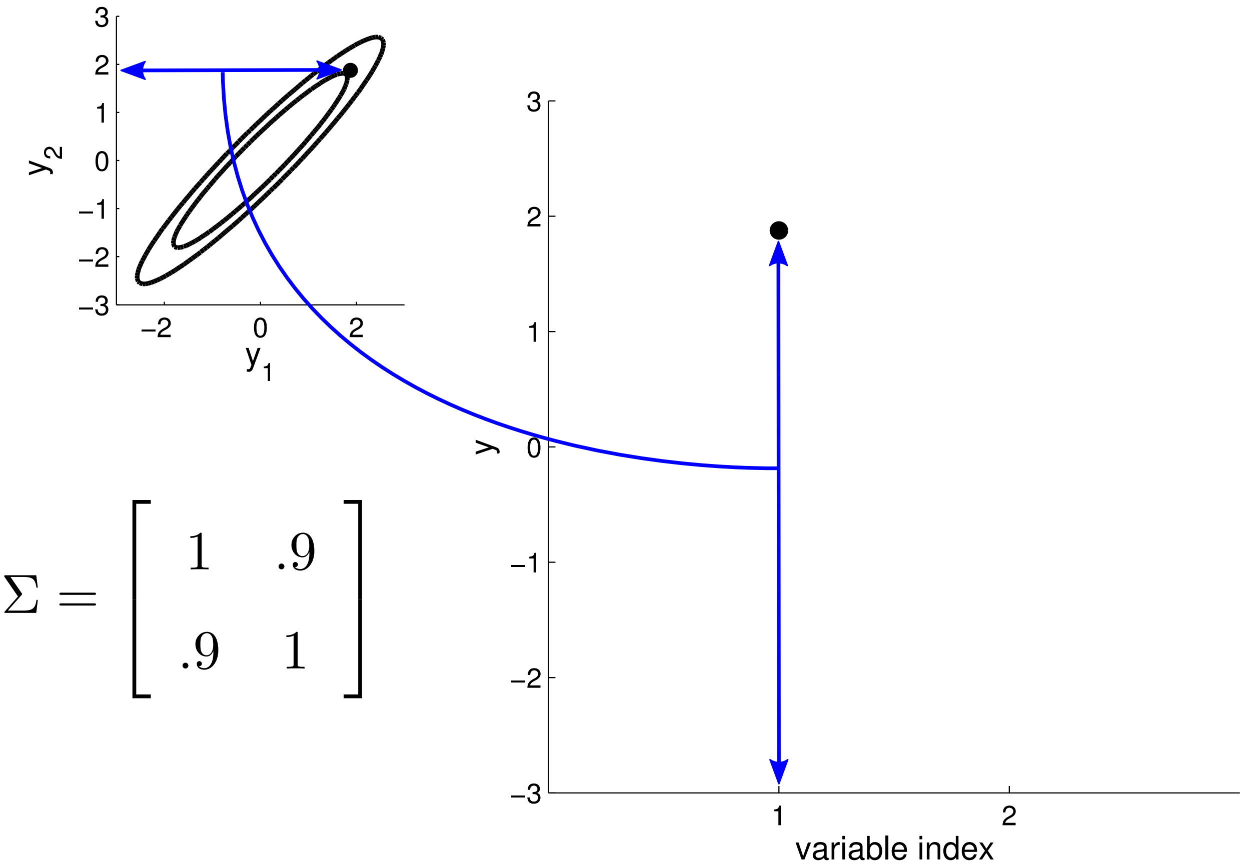


New visualisation



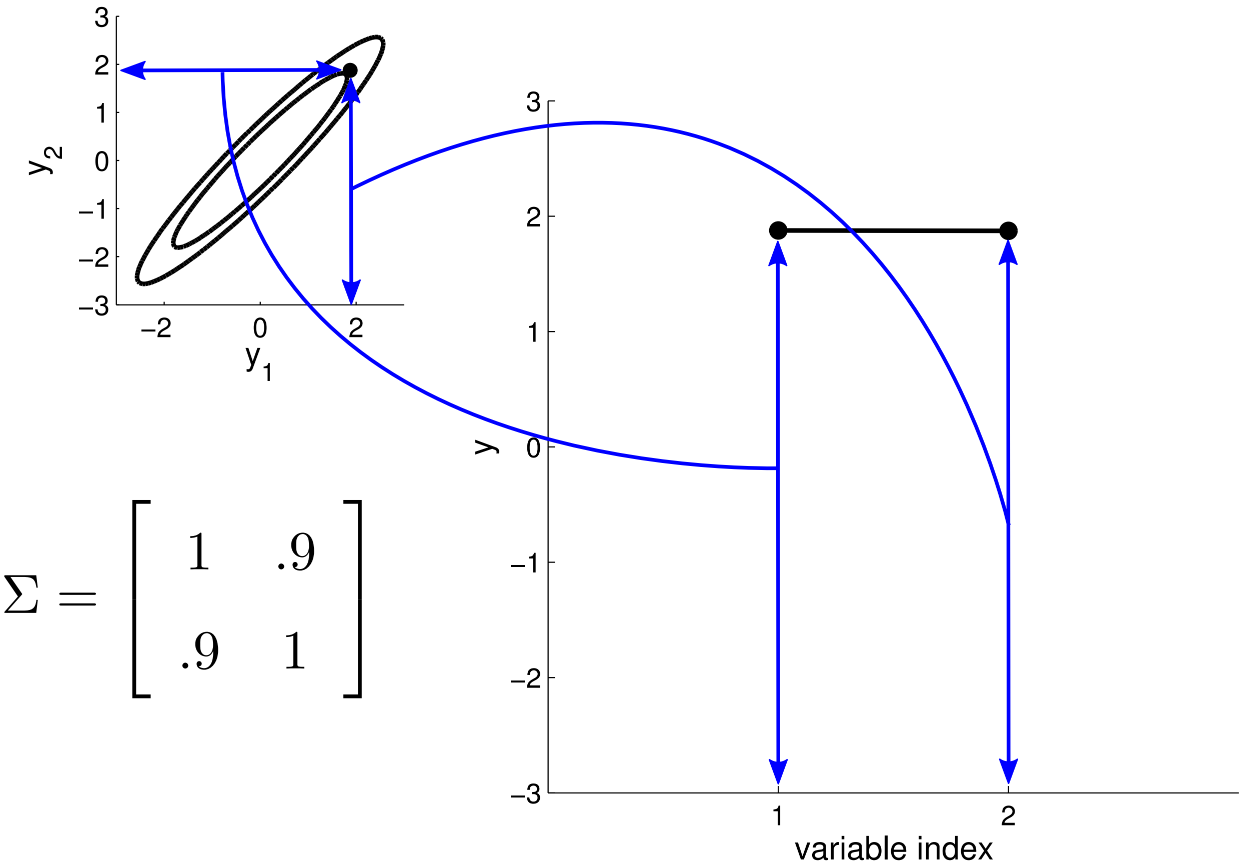
$$\Sigma = \begin{bmatrix} 1 & .9 \\ .9 & 1 \end{bmatrix}$$

New visualisation



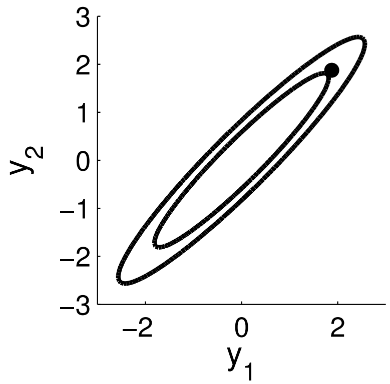
$$\Sigma = \begin{bmatrix} 1 & .9 \\ .9 & 1 \end{bmatrix}$$

New visualisation

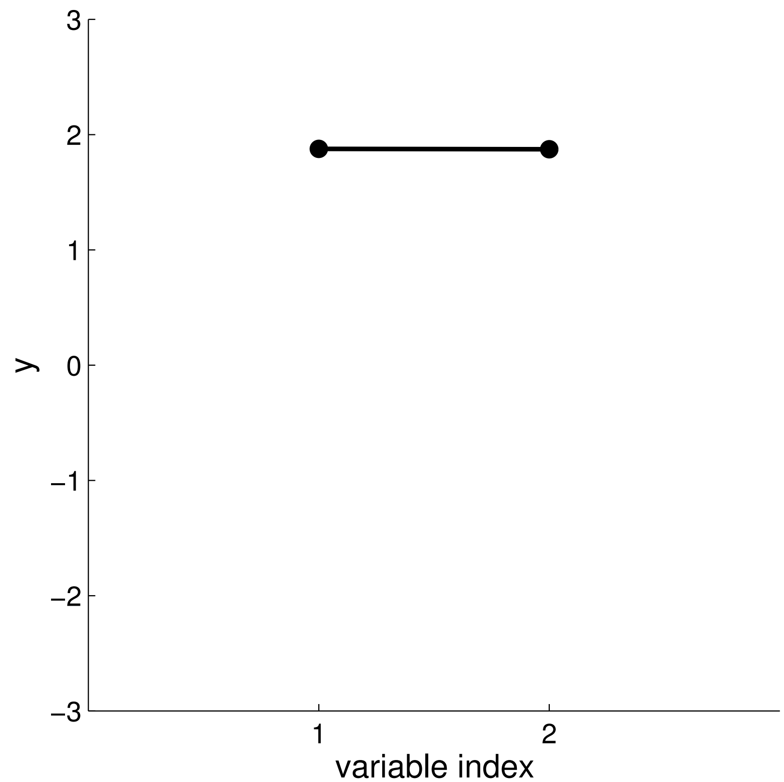


$$\Sigma = \begin{bmatrix} 1 & .9 \\ .9 & 1 \end{bmatrix}$$

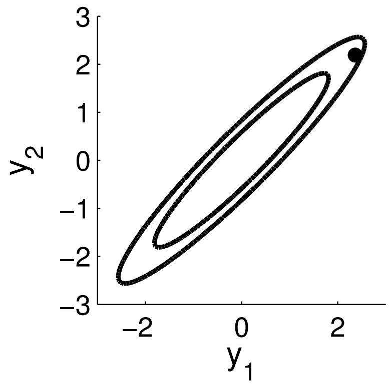
New visualisation

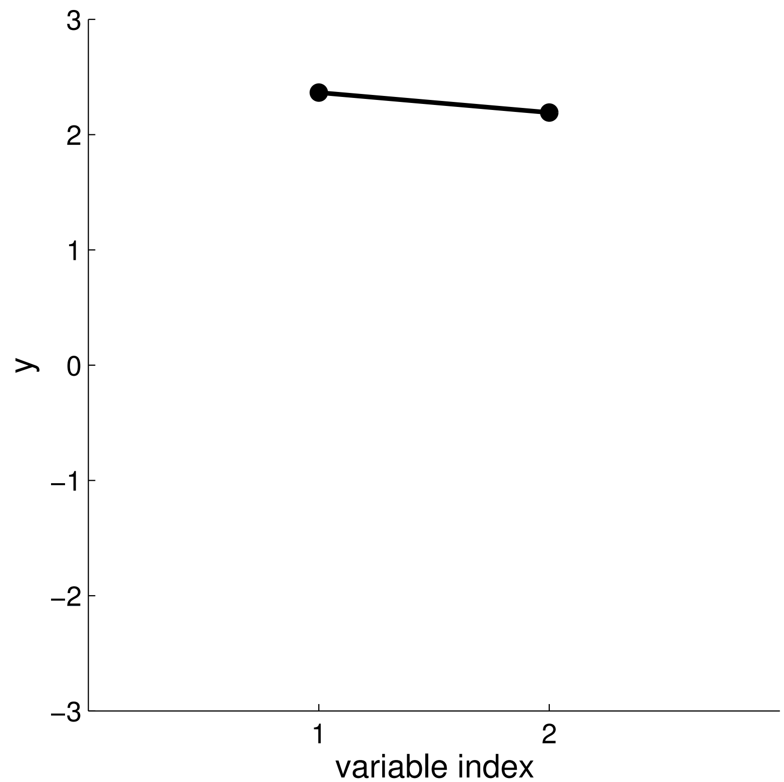


$$\Sigma = \begin{bmatrix} 1 & .9 \\ .9 & 1 \end{bmatrix}$$



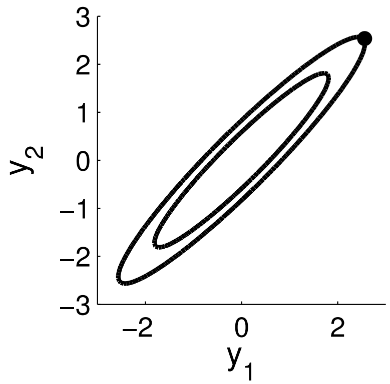
New visualisation



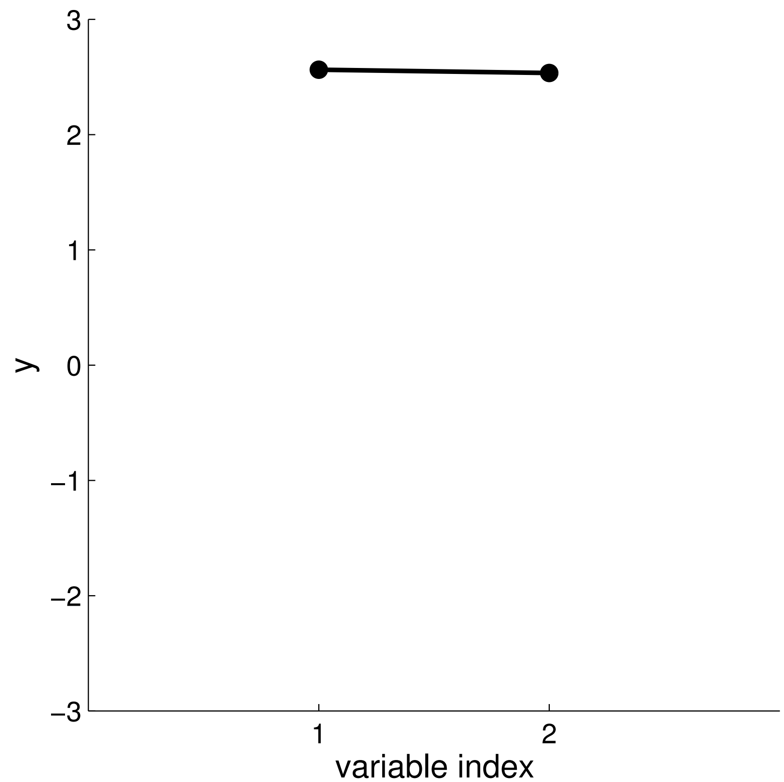


$$\Sigma = \begin{bmatrix} 1 & .9 \\ .9 & 1 \end{bmatrix}$$

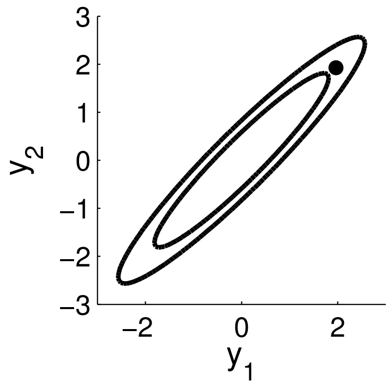
New visualisation



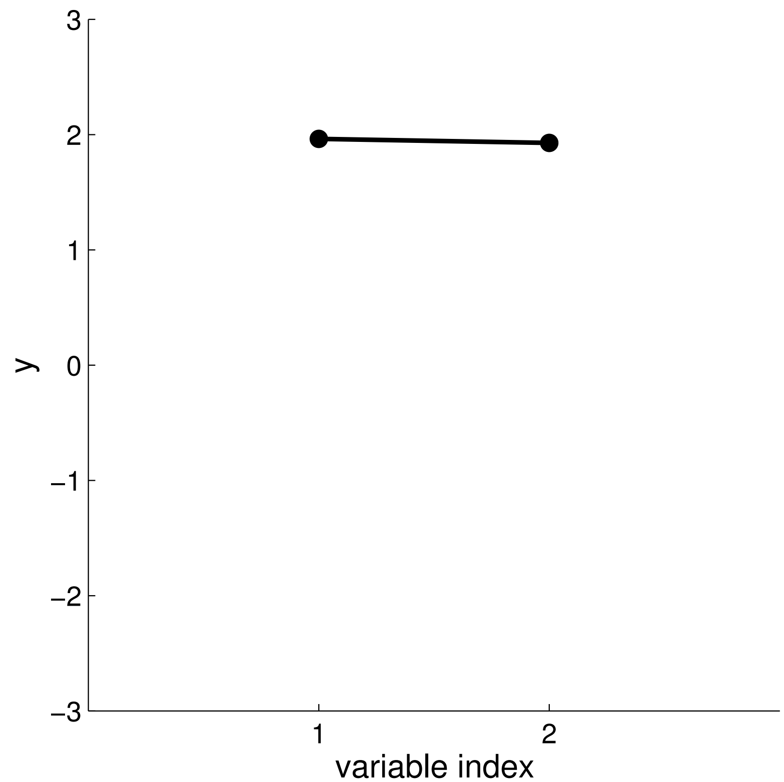
$$\Sigma = \begin{bmatrix} 1 & .9 \\ .9 & 1 \end{bmatrix}$$



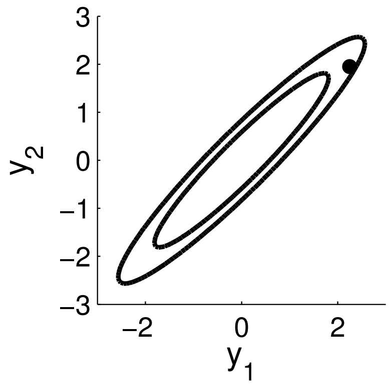
New visualisation



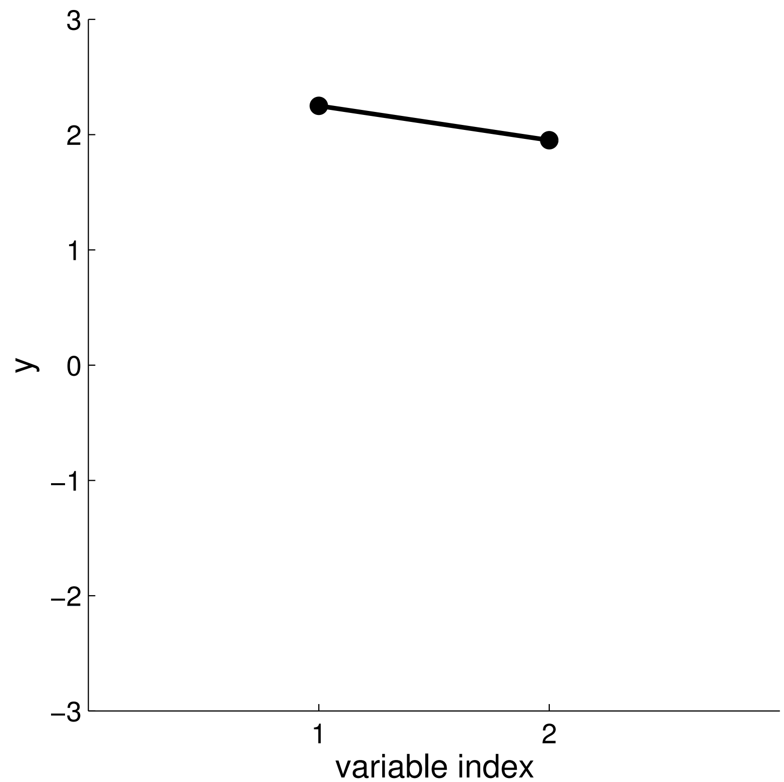
$$\Sigma = \begin{bmatrix} 1 & .9 \\ .9 & 1 \end{bmatrix}$$



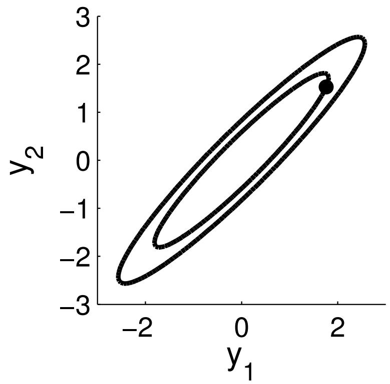
New visualisation



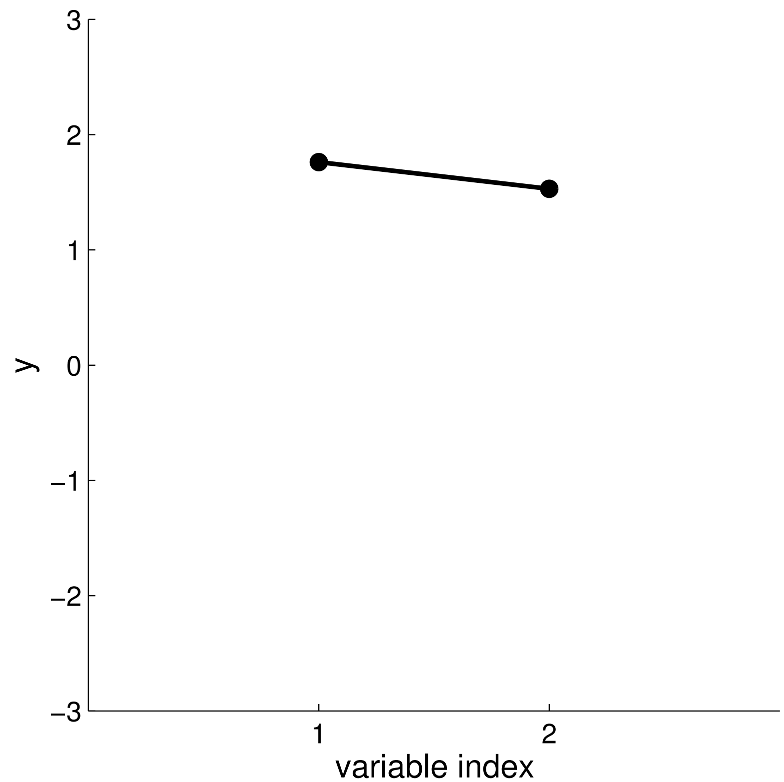
$$\Sigma = \begin{bmatrix} 1 & .9 \\ .9 & 1 \end{bmatrix}$$



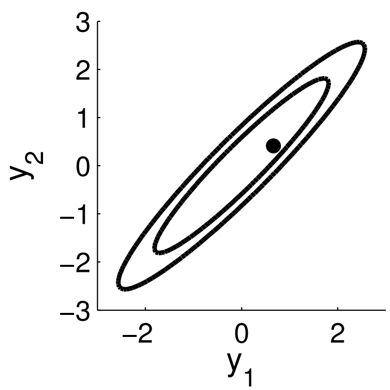
New visualisation



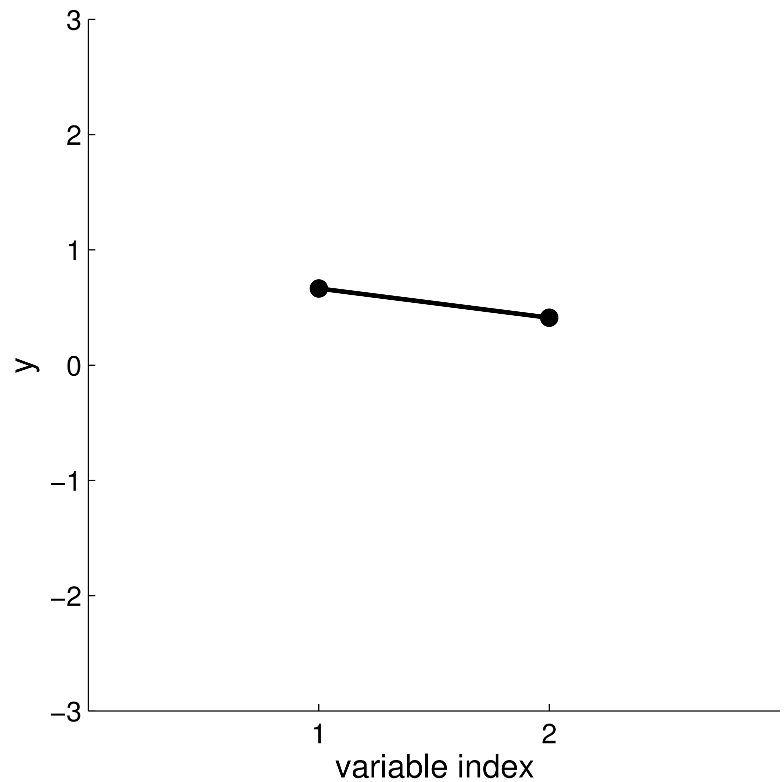
$$\Sigma = \begin{bmatrix} 1 & .9 \\ .9 & 1 \end{bmatrix}$$



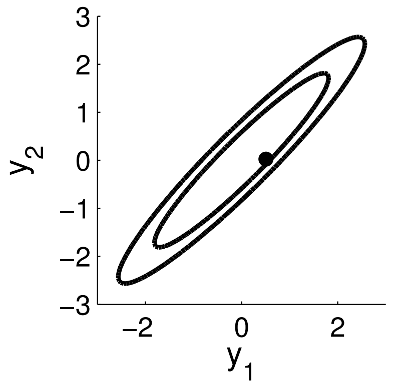
New visualisation



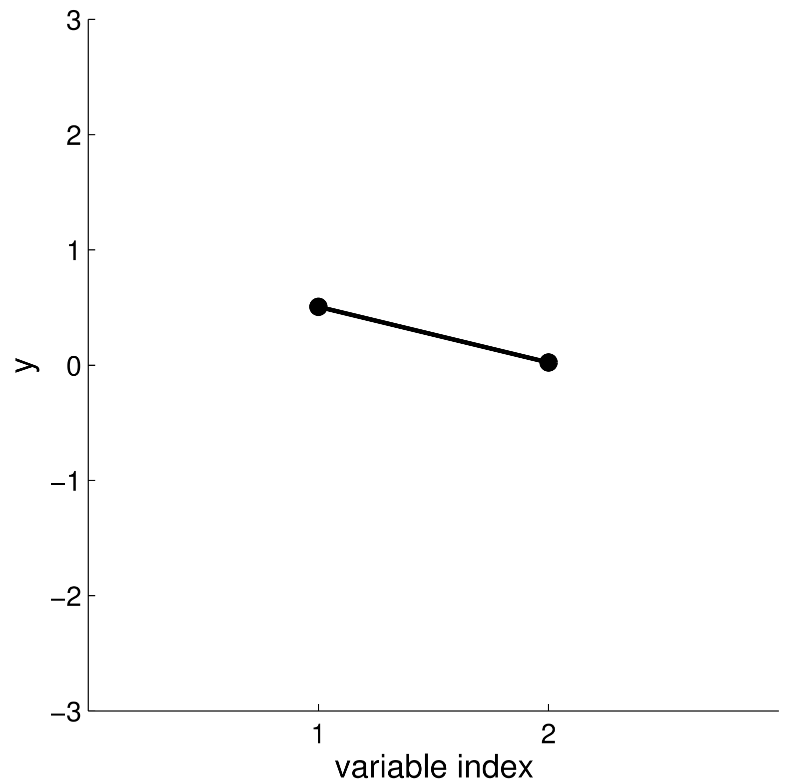
$$\Sigma = \begin{bmatrix} 1 & .9 \\ .9 & 1 \end{bmatrix}$$



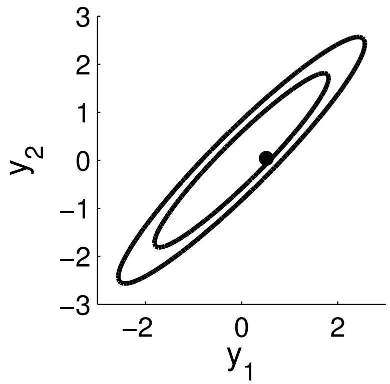
New visualisation



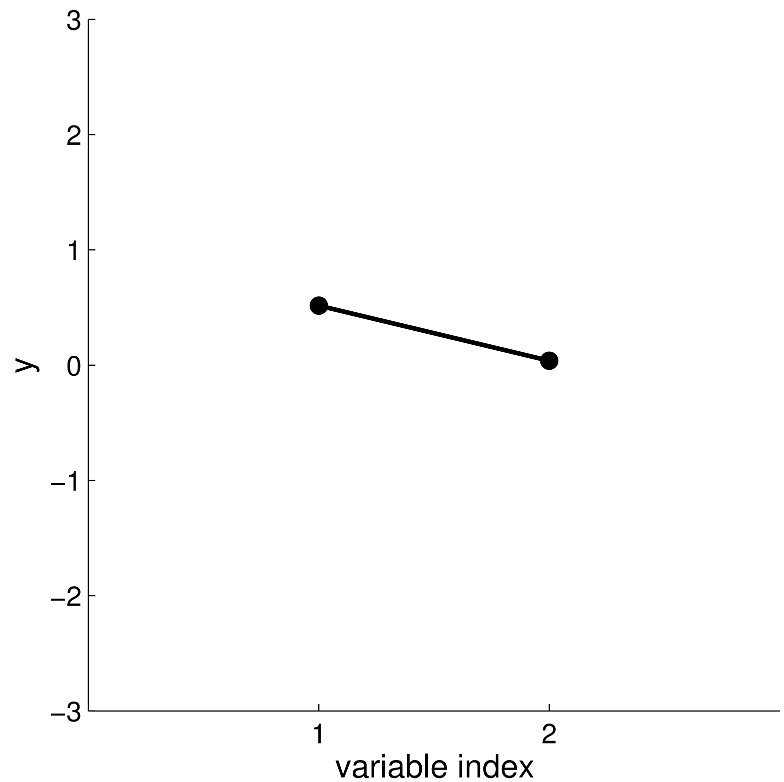
$$\Sigma = \begin{bmatrix} 1 & .9 \\ .9 & 1 \end{bmatrix}$$



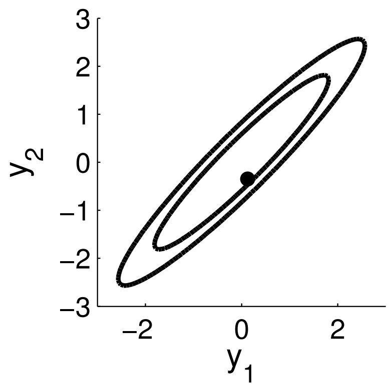
New visualisation



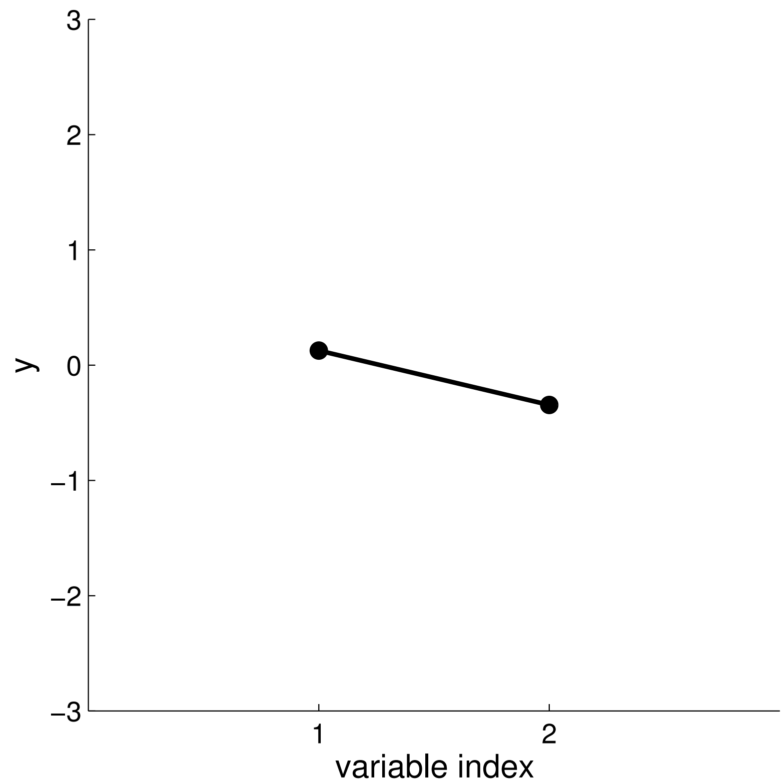
$$\Sigma = \begin{bmatrix} 1 & .9 \\ .9 & 1 \end{bmatrix}$$



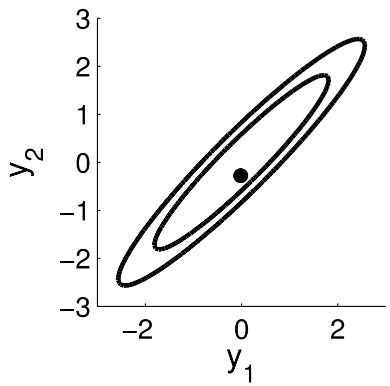
New visualisation



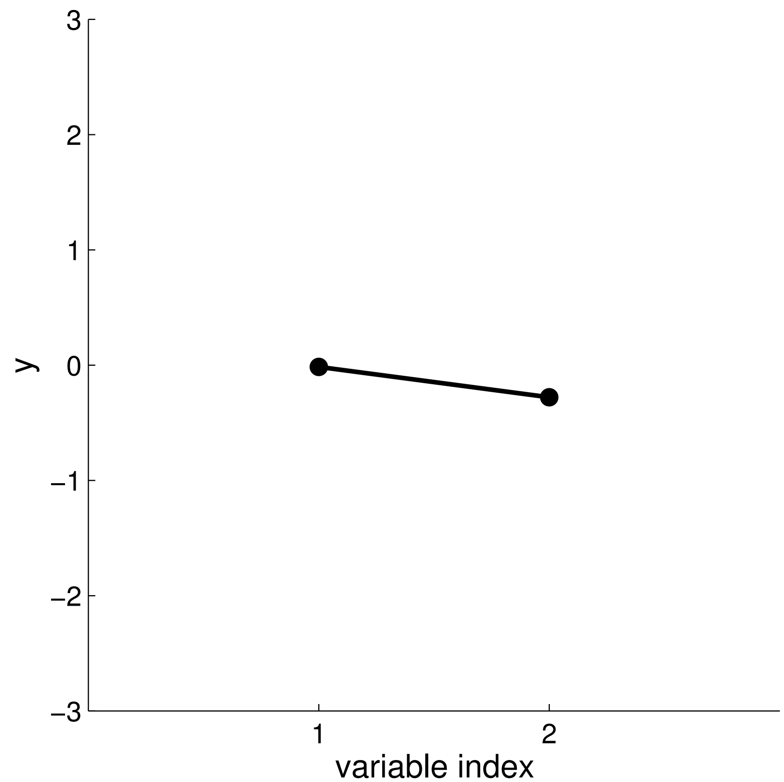
$$\Sigma = \begin{bmatrix} 1 & .9 \\ .9 & 1 \end{bmatrix}$$



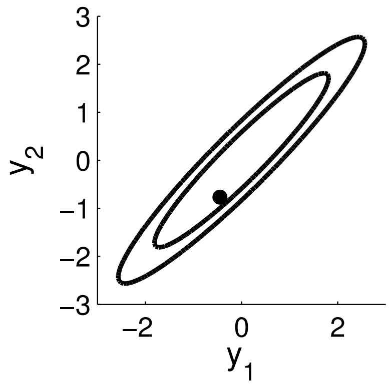
New visualisation



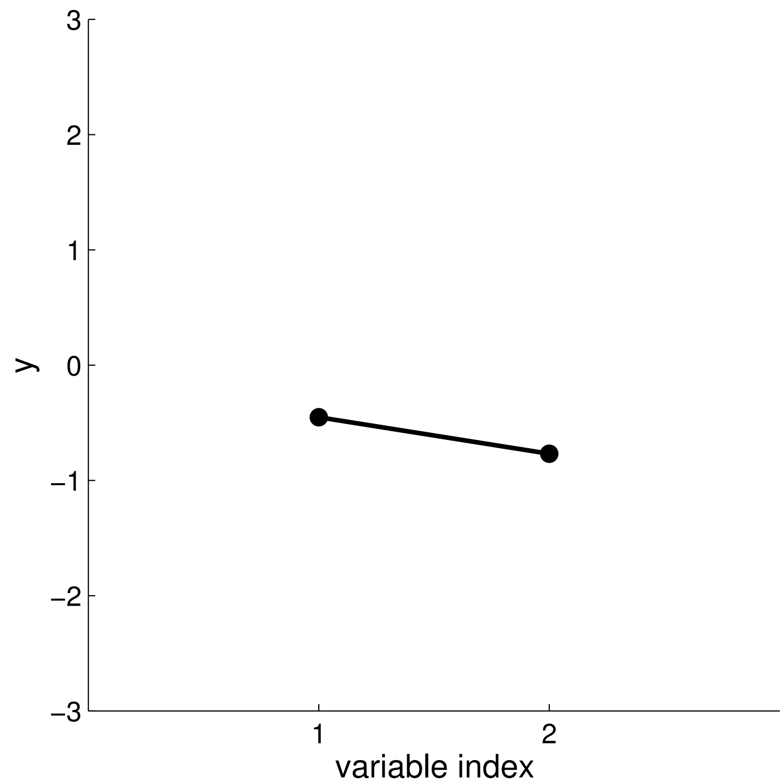
$$\Sigma = \begin{bmatrix} 1 & .9 \\ .9 & 1 \end{bmatrix}$$



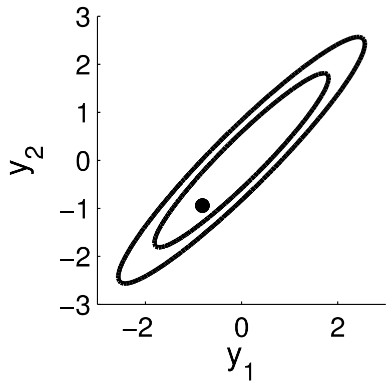
New visualisation



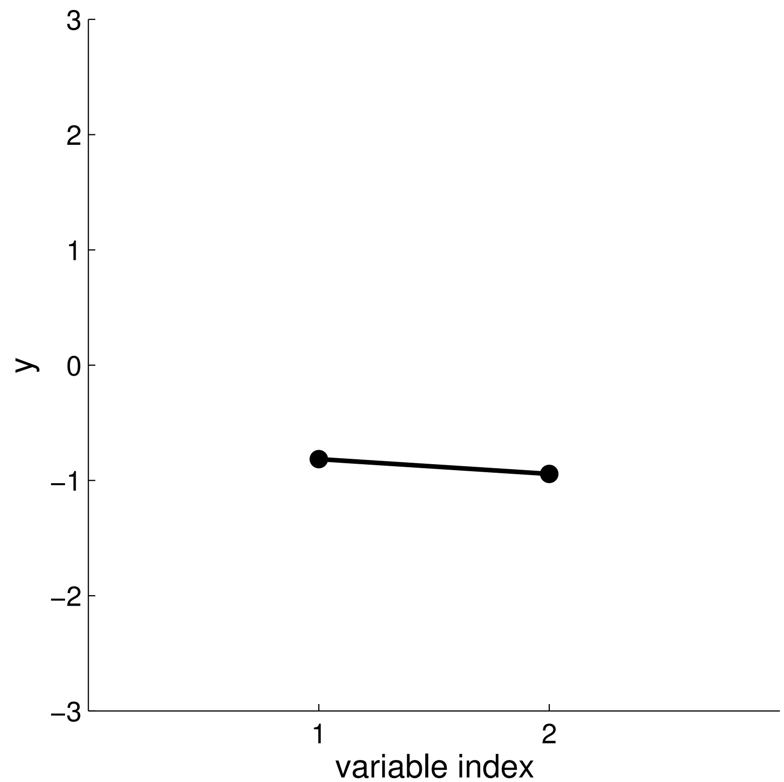
$$\Sigma = \begin{bmatrix} 1 & .9 \\ .9 & 1 \end{bmatrix}$$



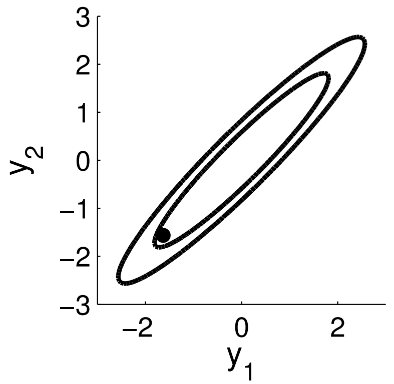
New visualisation



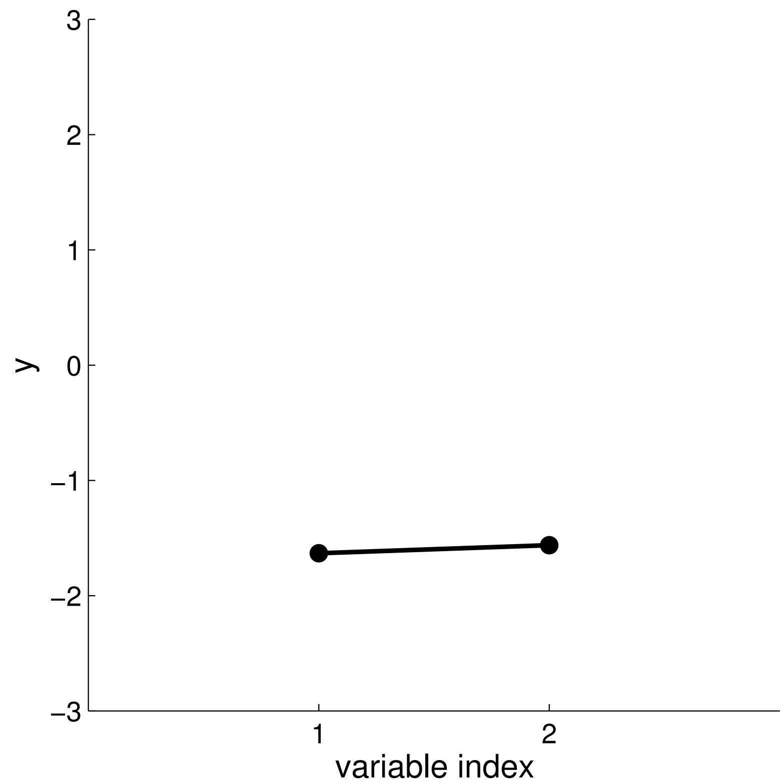
$$\Sigma = \begin{bmatrix} 1 & .9 \\ .9 & 1 \end{bmatrix}$$



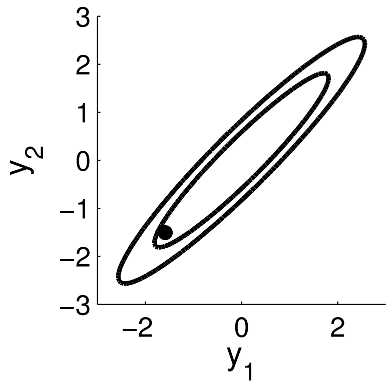
New visualisation



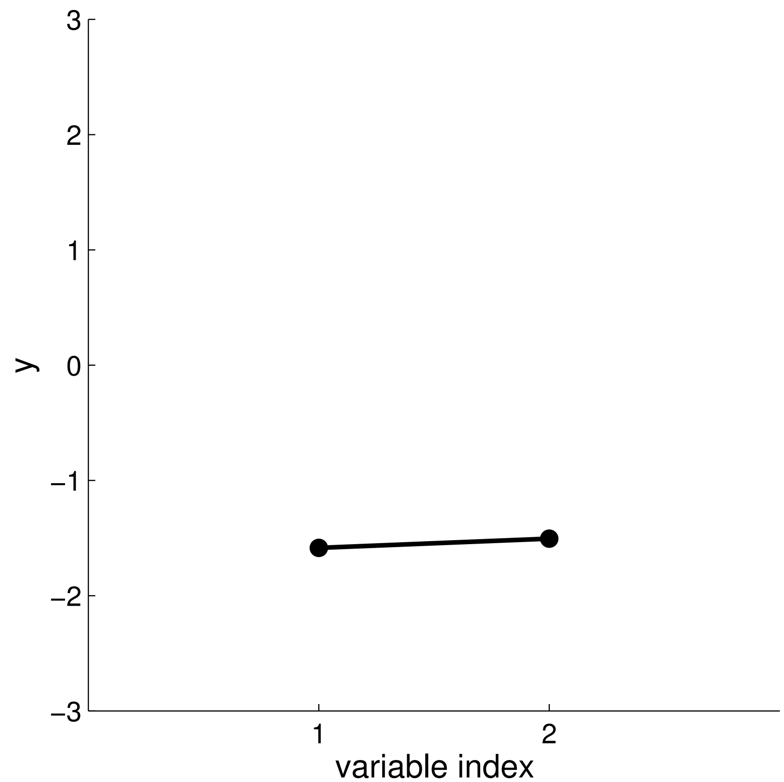
$$\Sigma = \begin{bmatrix} 1 & .9 \\ .9 & 1 \end{bmatrix}$$



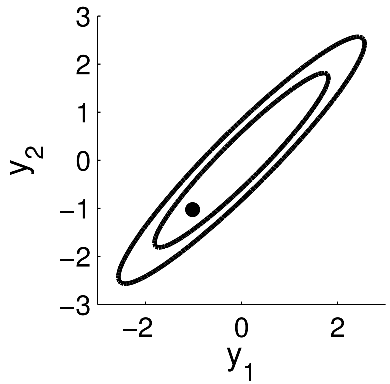
New visualisation



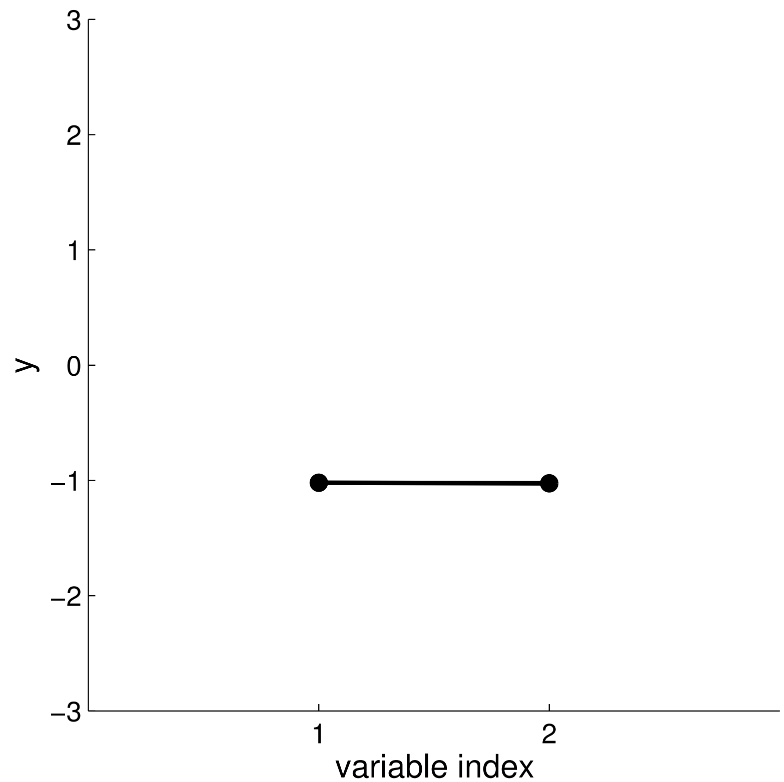
$$\Sigma = \begin{bmatrix} 1 & .9 \\ .9 & 1 \end{bmatrix}$$



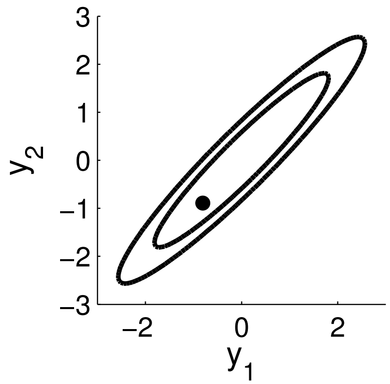
New visualisation



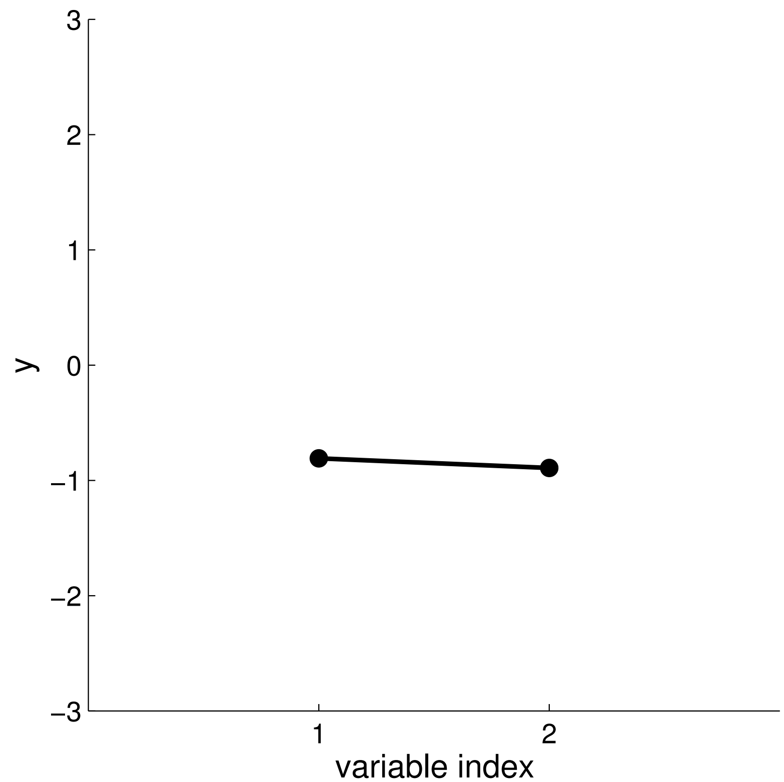
$$\Sigma = \begin{bmatrix} 1 & .9 \\ .9 & 1 \end{bmatrix}$$



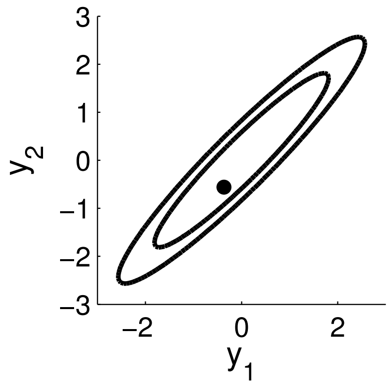
New visualisation



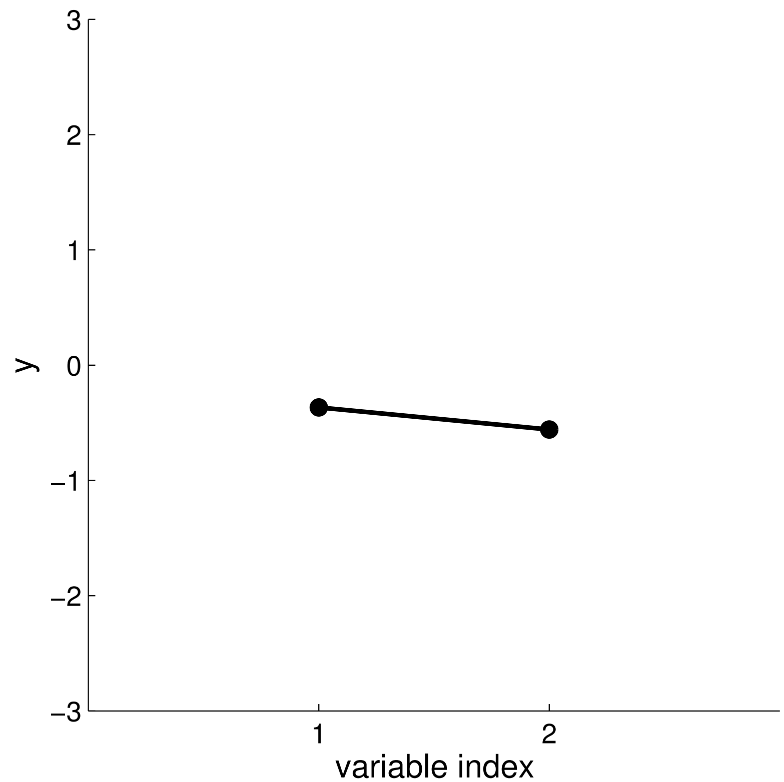
$$\Sigma = \begin{bmatrix} 1 & .9 \\ .9 & 1 \end{bmatrix}$$



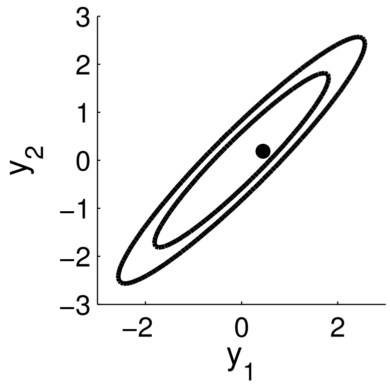
New visualisation



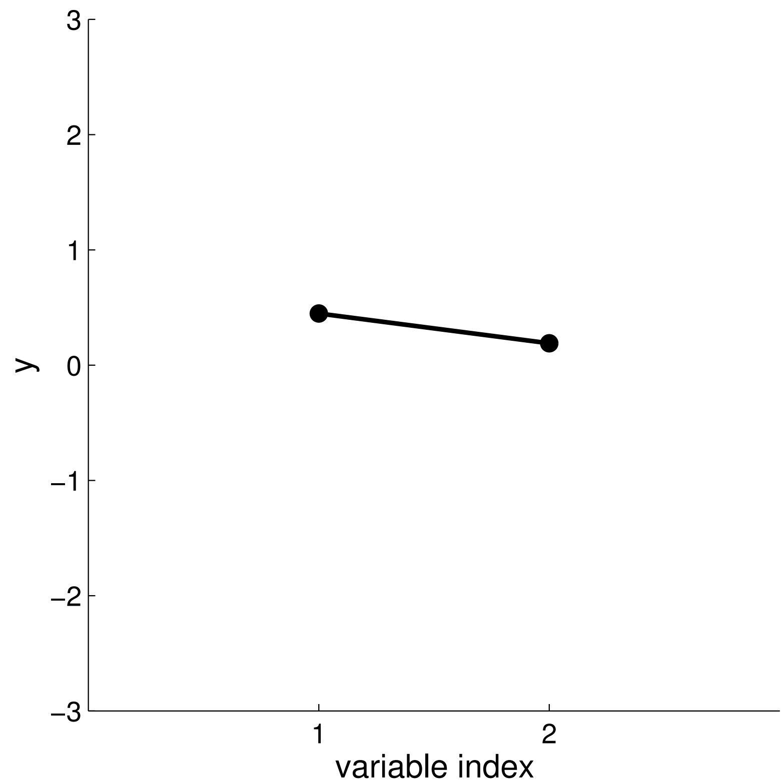
$$\Sigma = \begin{bmatrix} 1 & .9 \\ .9 & 1 \end{bmatrix}$$



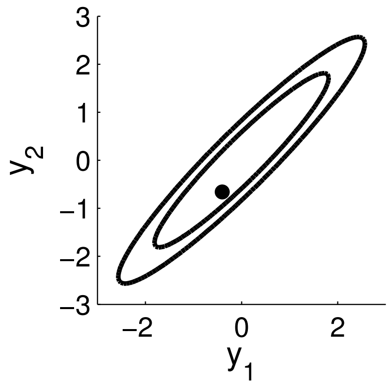
New visualisation



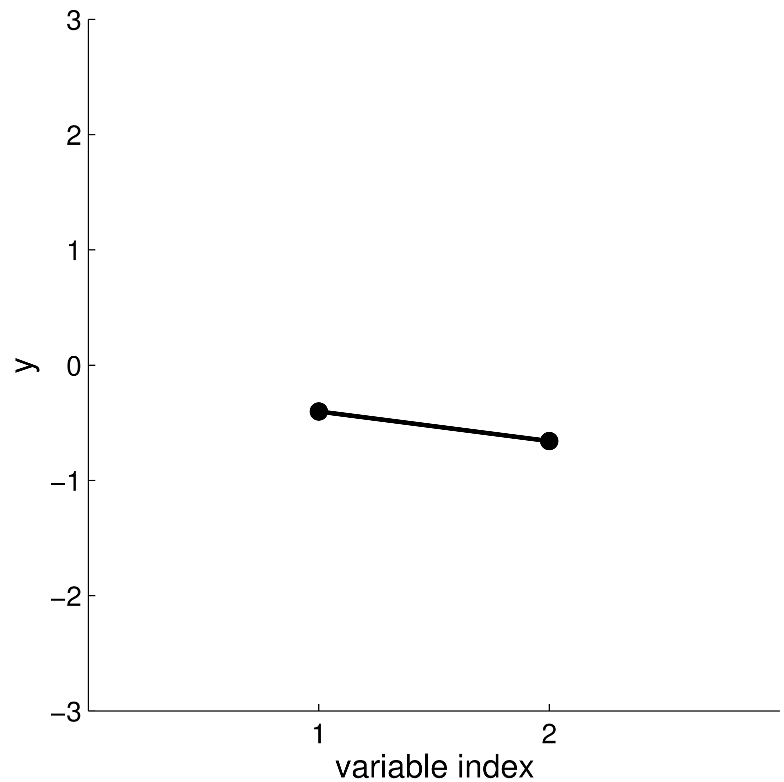
$$\Sigma = \begin{bmatrix} 1 & .9 \\ .9 & 1 \end{bmatrix}$$



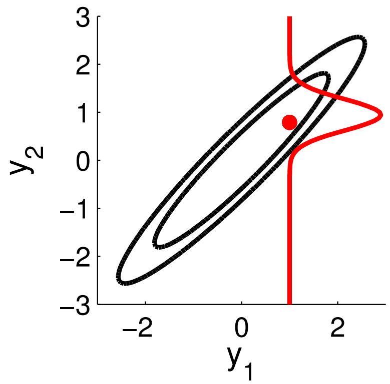
New visualisation



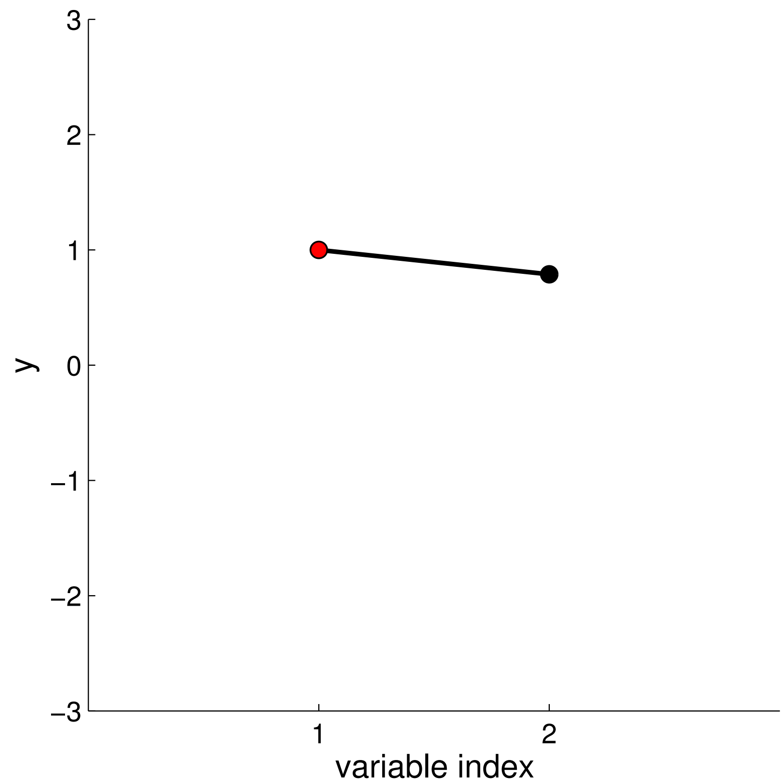
$$\Sigma = \begin{bmatrix} 1 & .9 \\ .9 & 1 \end{bmatrix}$$



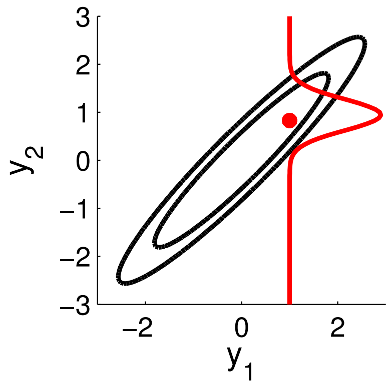
New visualisation

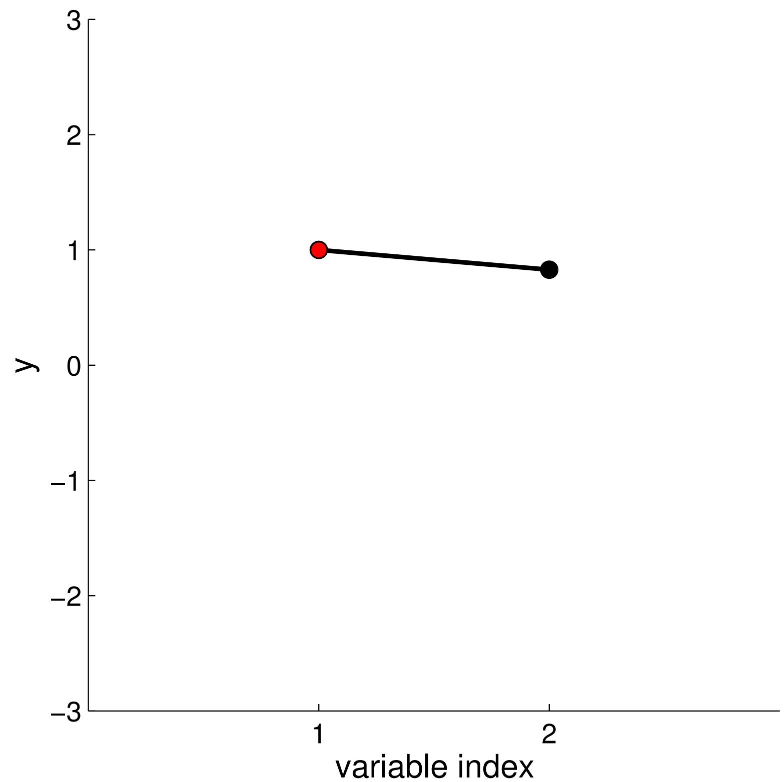


$$\Sigma = \begin{bmatrix} 1 & .9 \\ .9 & 1 \end{bmatrix}$$



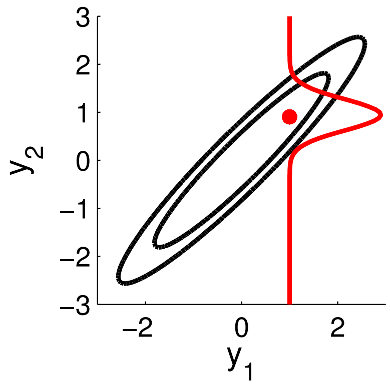
New visualisation



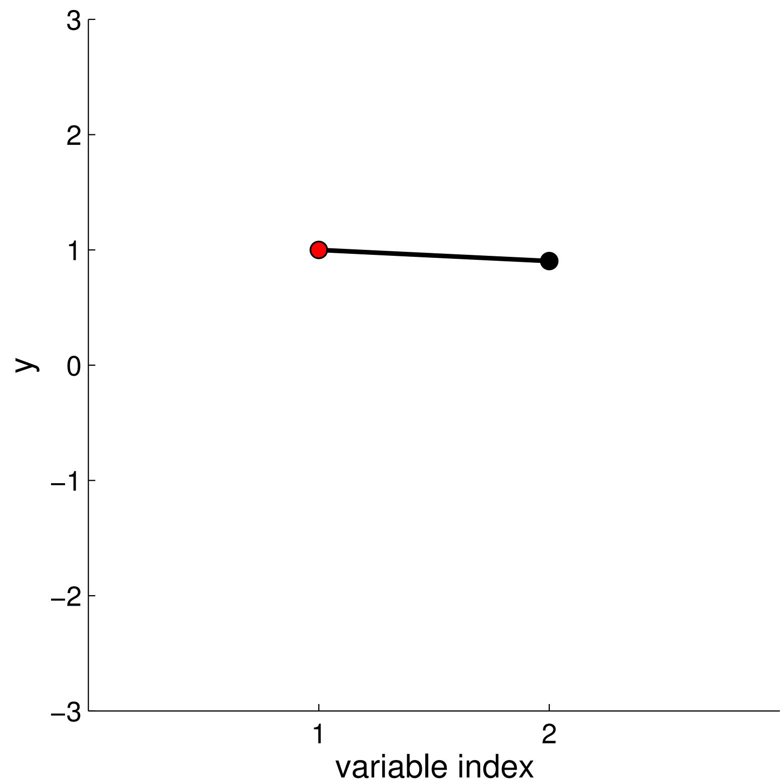


$$\Sigma = \begin{bmatrix} 1 & .9 \\ .9 & 1 \end{bmatrix}$$

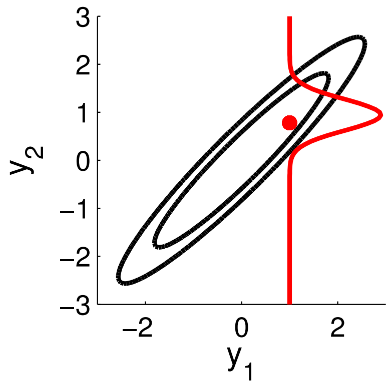
New visualisation



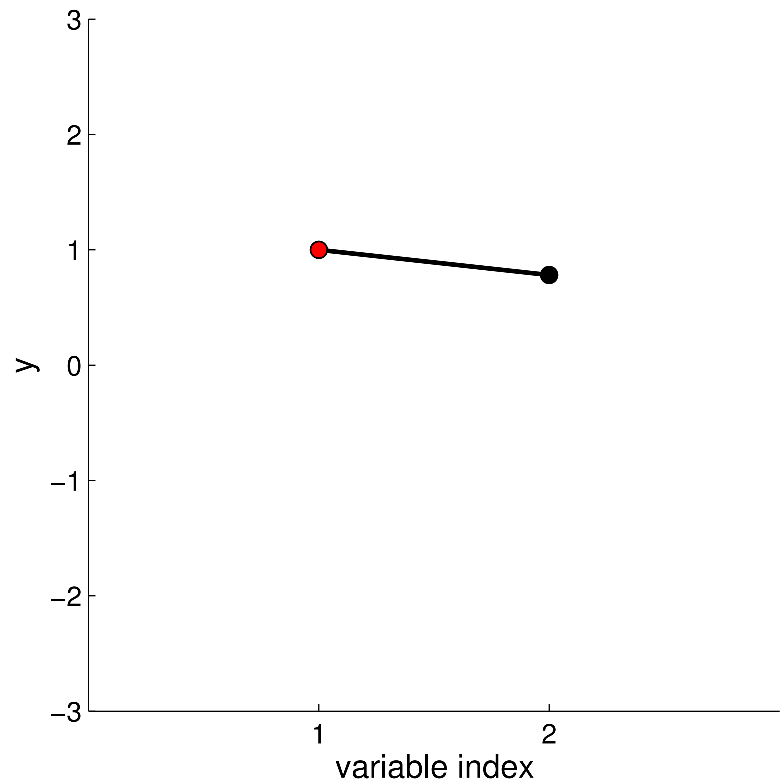
$$\Sigma = \begin{bmatrix} 1 & .9 \\ .9 & 1 \end{bmatrix}$$



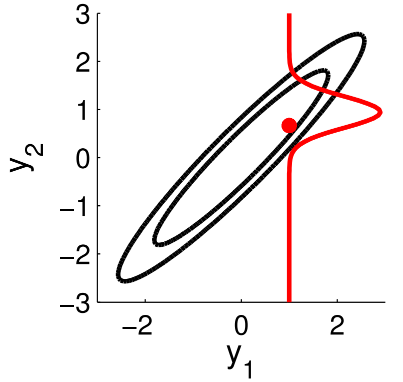
New visualisation



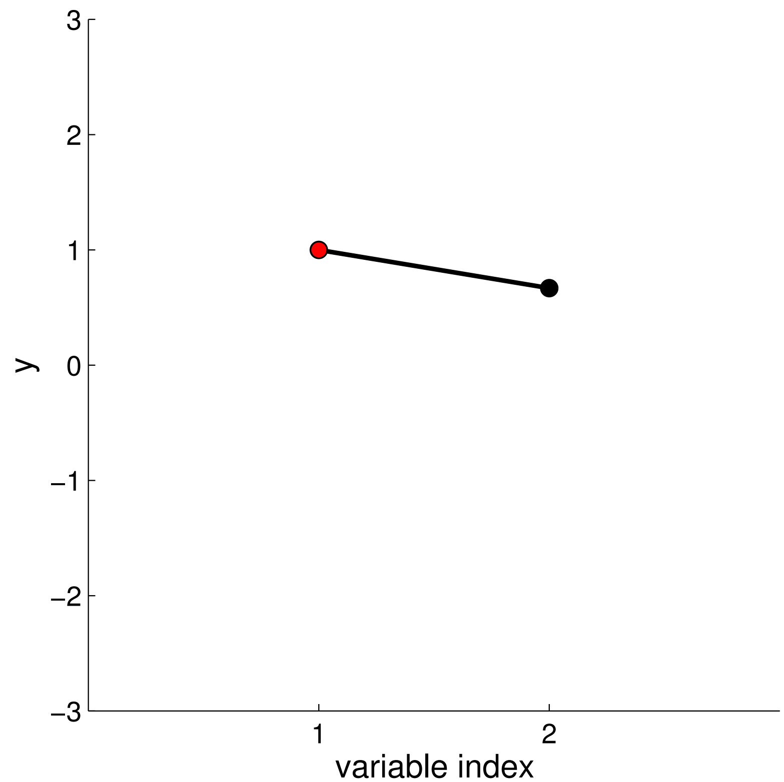
$$\Sigma = \begin{bmatrix} 1 & .9 \\ .9 & 1 \end{bmatrix}$$



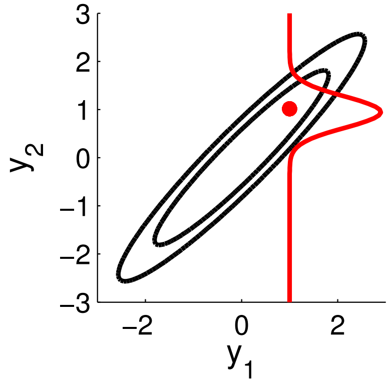
New visualisation



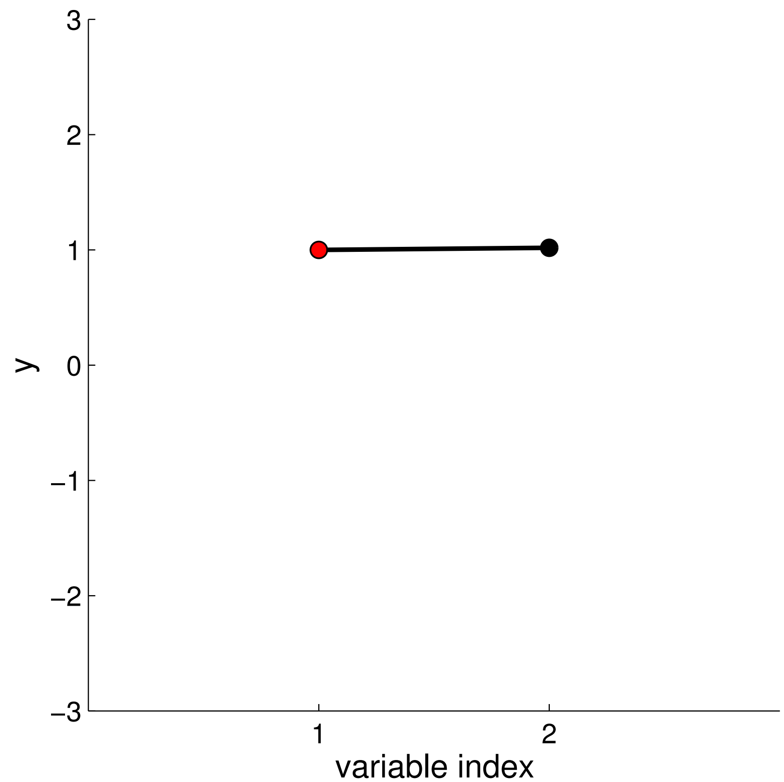
$$\Sigma = \begin{bmatrix} 1 & .9 \\ .9 & 1 \end{bmatrix}$$



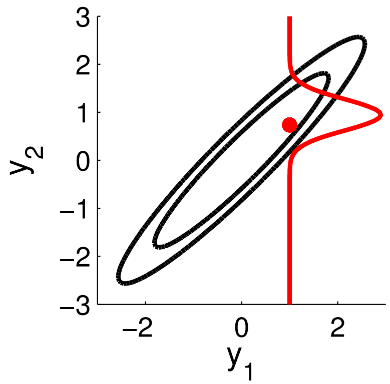
New visualisation



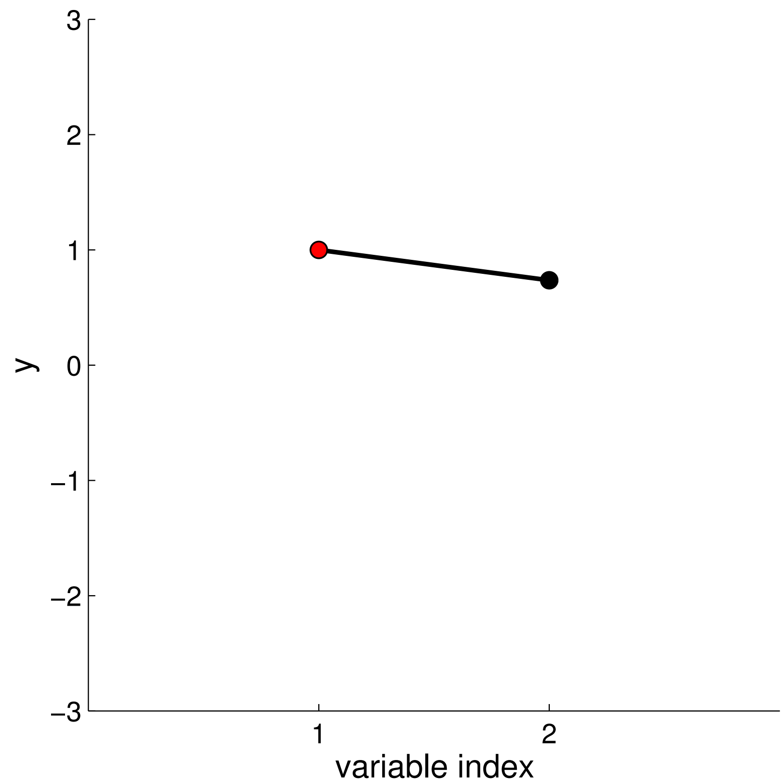
$$\Sigma = \begin{bmatrix} 1 & .9 \\ .9 & 1 \end{bmatrix}$$



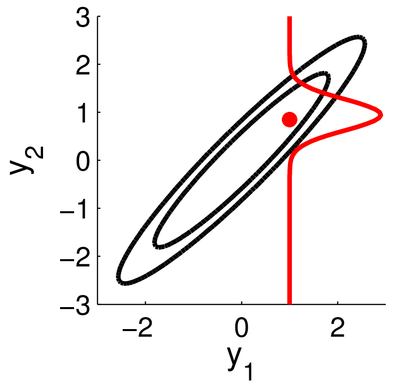
New visualisation



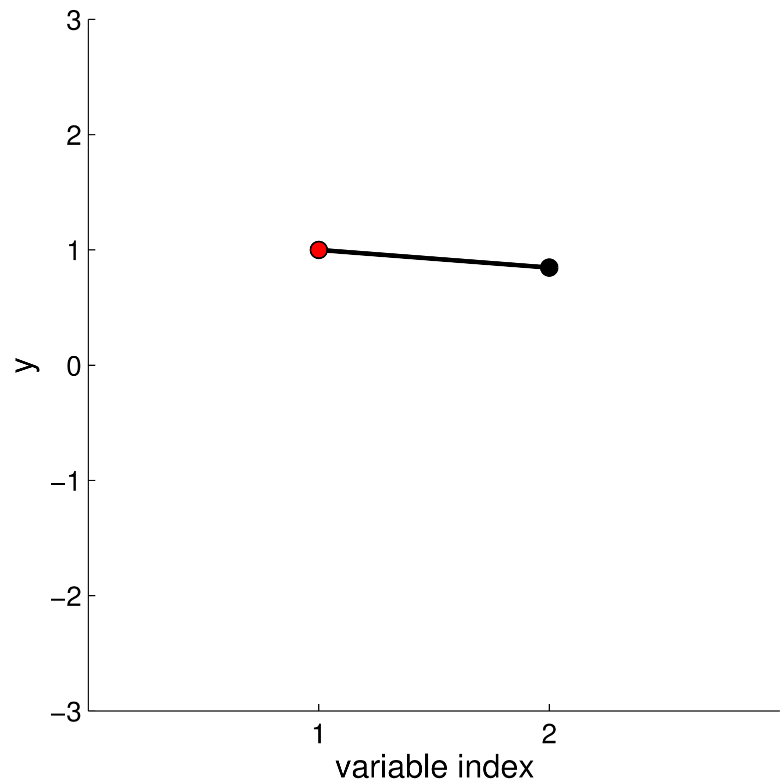
$$\Sigma = \begin{bmatrix} 1 & .9 \\ .9 & 1 \end{bmatrix}$$



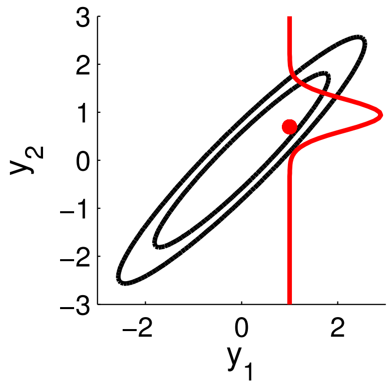
New visualisation



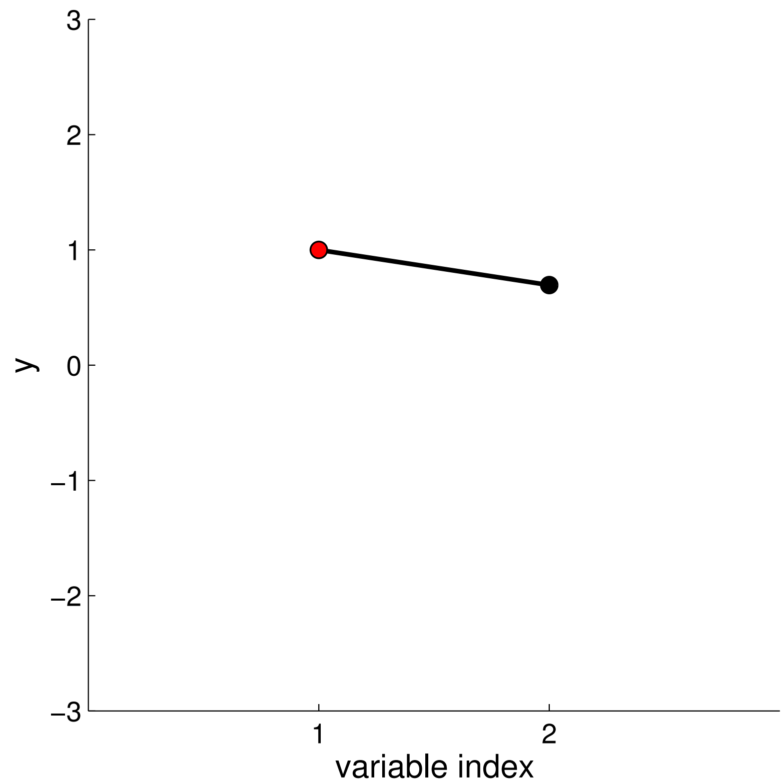
$$\Sigma = \begin{bmatrix} 1 & .9 \\ .9 & 1 \end{bmatrix}$$



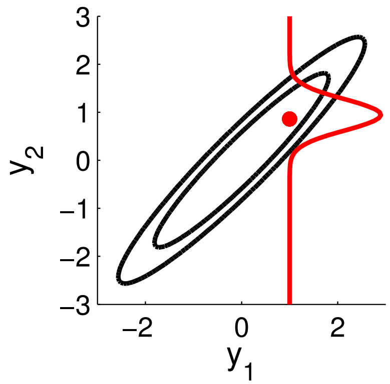
New visualisation



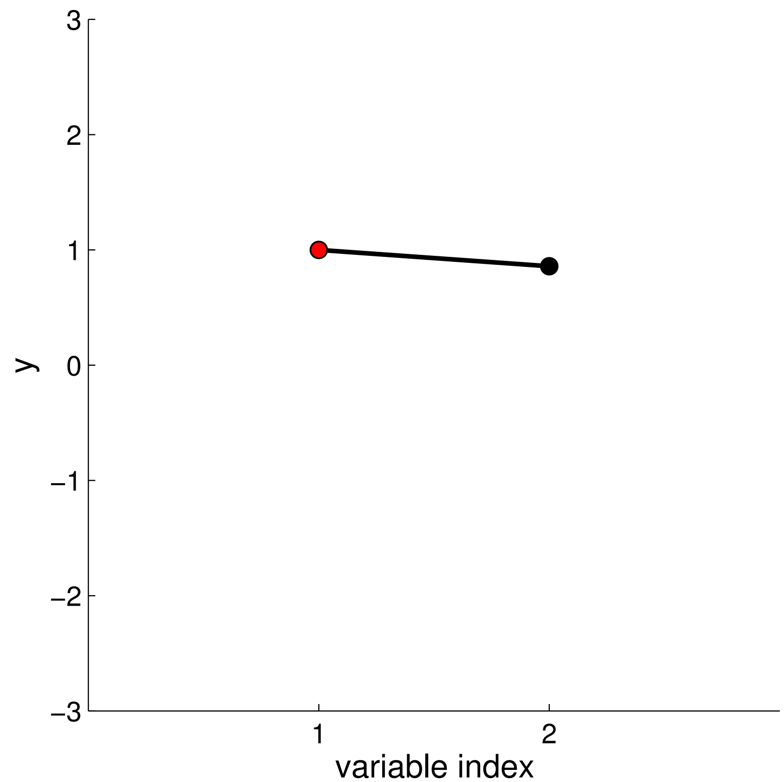
$$\Sigma = \begin{bmatrix} 1 & .9 \\ .9 & 1 \end{bmatrix}$$



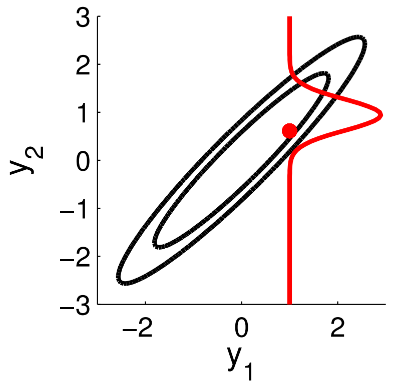
New visualisation



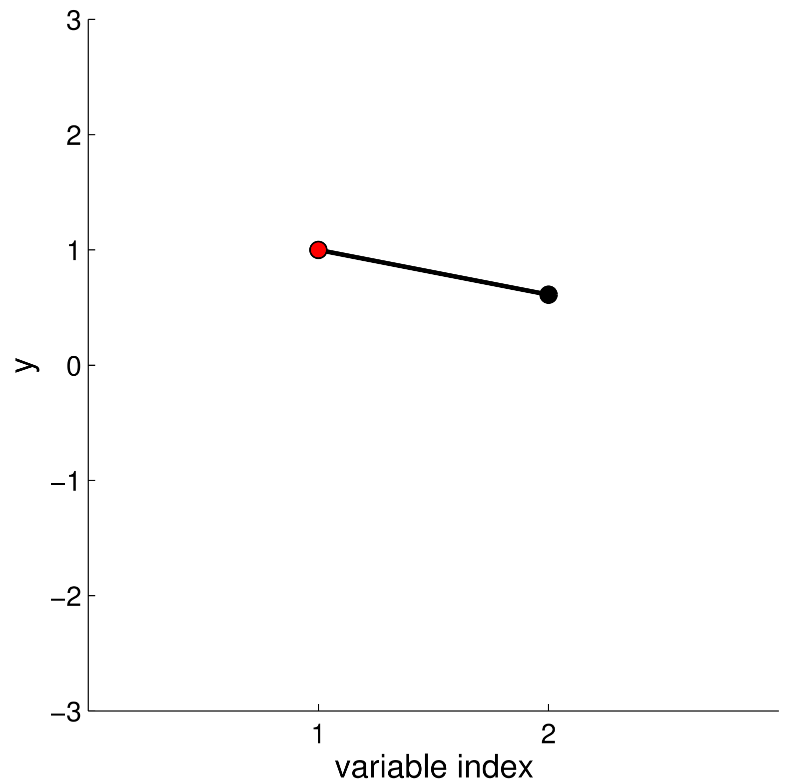
$$\Sigma = \begin{bmatrix} 1 & .9 \\ .9 & 1 \end{bmatrix}$$



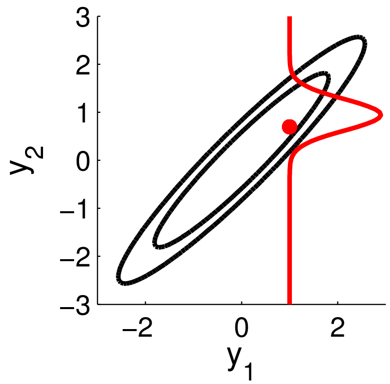
New visualisation



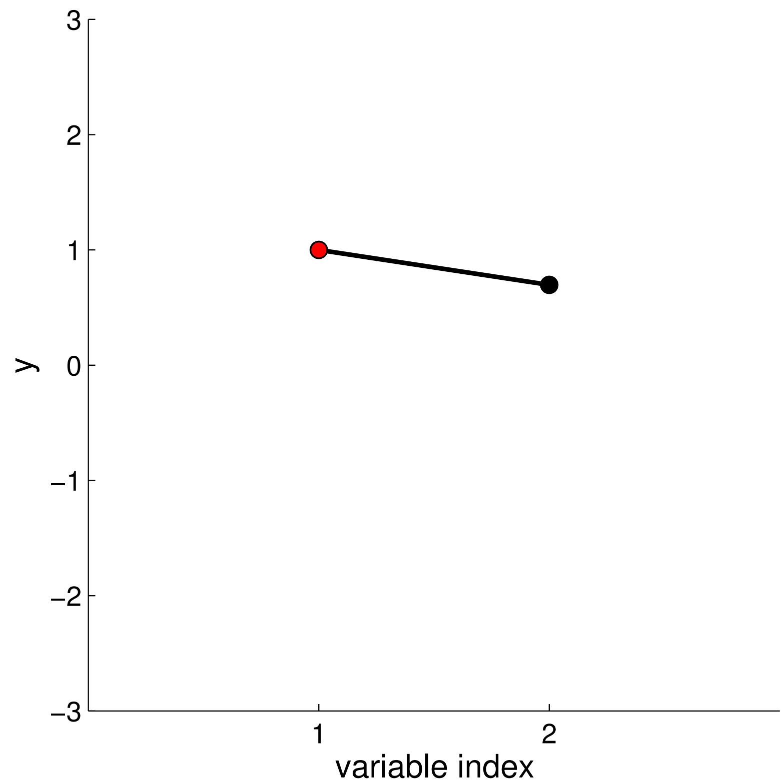
$$\Sigma = \begin{bmatrix} 1 & .9 \\ .9 & 1 \end{bmatrix}$$



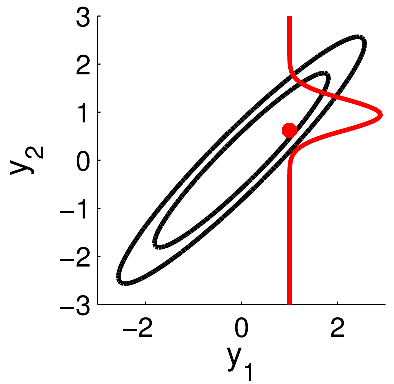
New visualisation



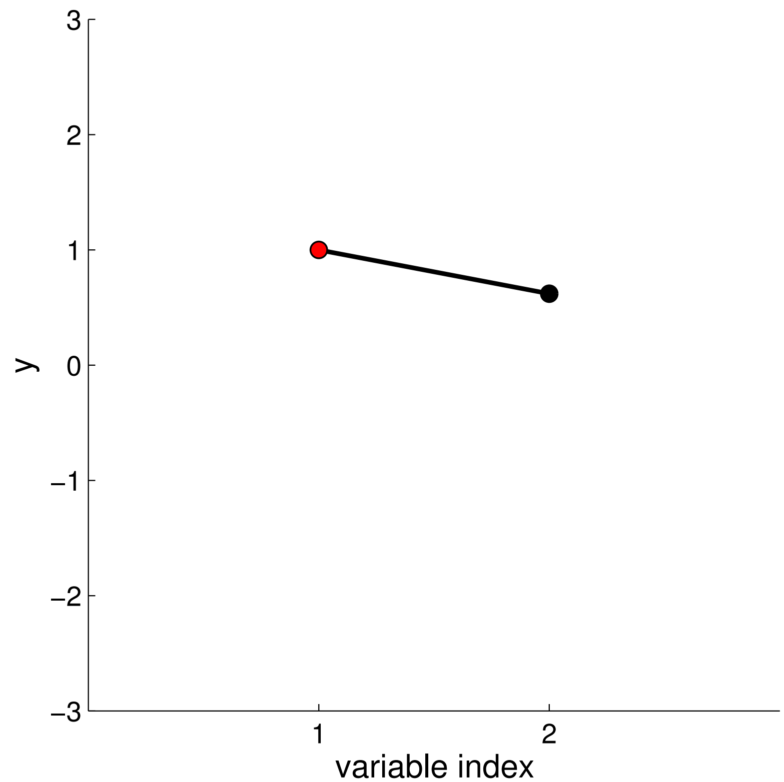
$$\Sigma = \begin{bmatrix} 1 & .9 \\ .9 & 1 \end{bmatrix}$$



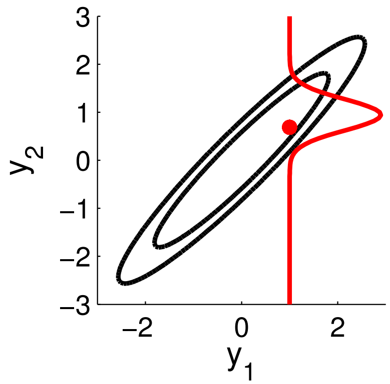
New visualisation



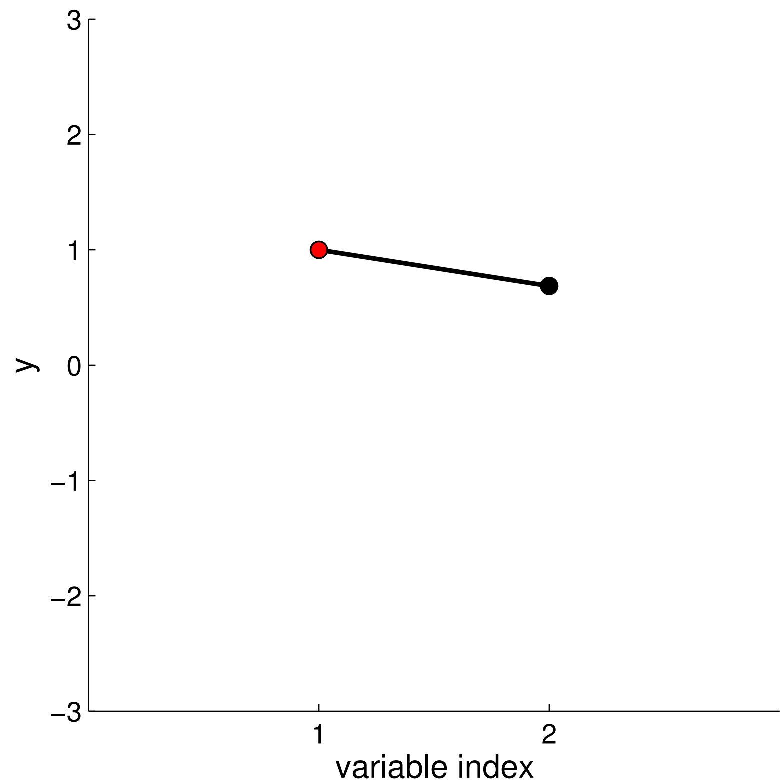
$$\Sigma = \begin{bmatrix} 1 & .9 \\ .9 & 1 \end{bmatrix}$$



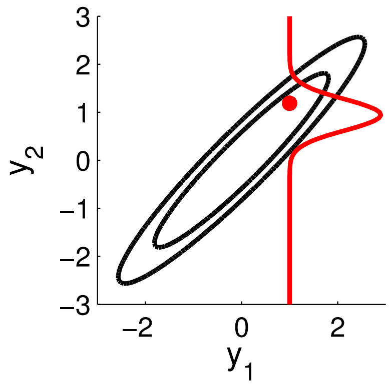
New visualisation



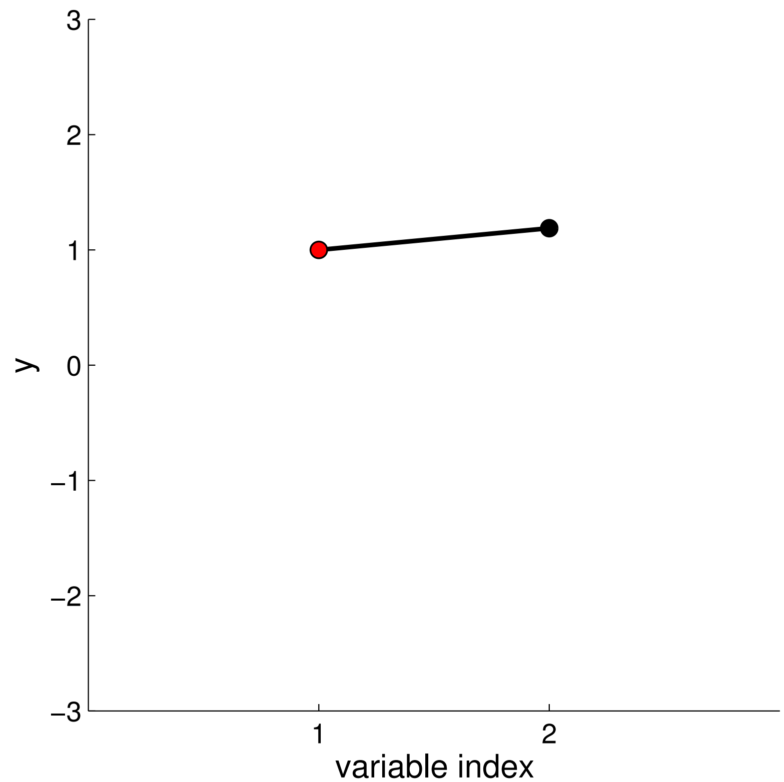
$$\Sigma = \begin{bmatrix} 1 & .9 \\ .9 & 1 \end{bmatrix}$$



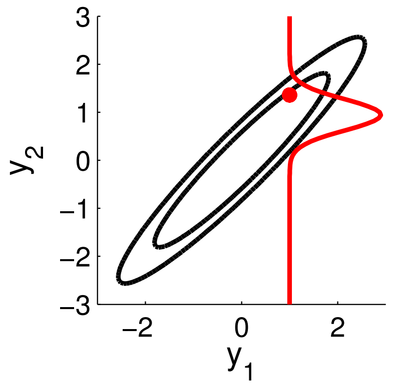
New visualisation



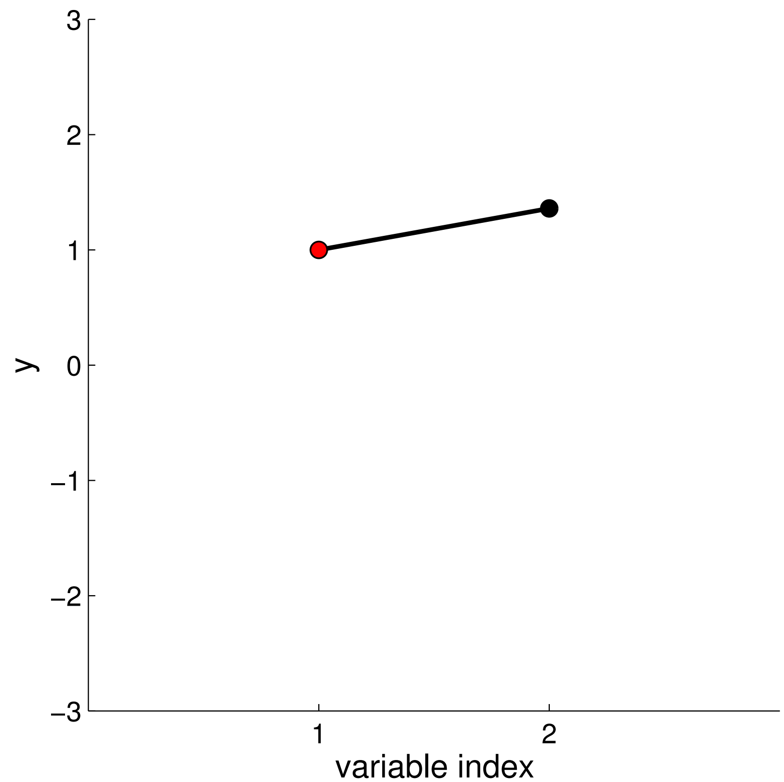
$$\Sigma = \begin{bmatrix} 1 & .9 \\ .9 & 1 \end{bmatrix}$$



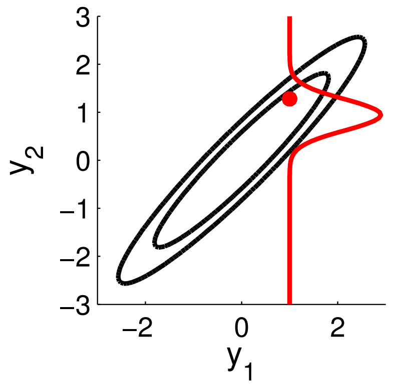
New visualisation



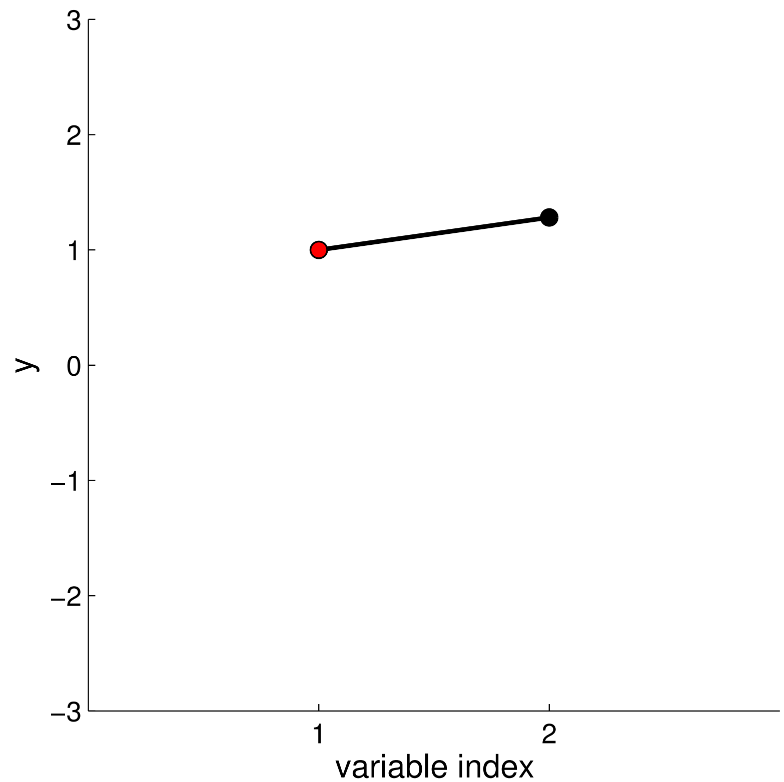
$$\Sigma = \begin{bmatrix} 1 & .9 \\ .9 & 1 \end{bmatrix}$$



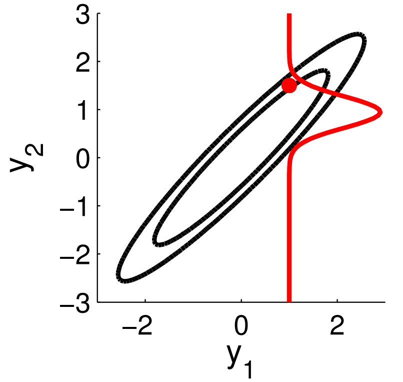
New visualisation



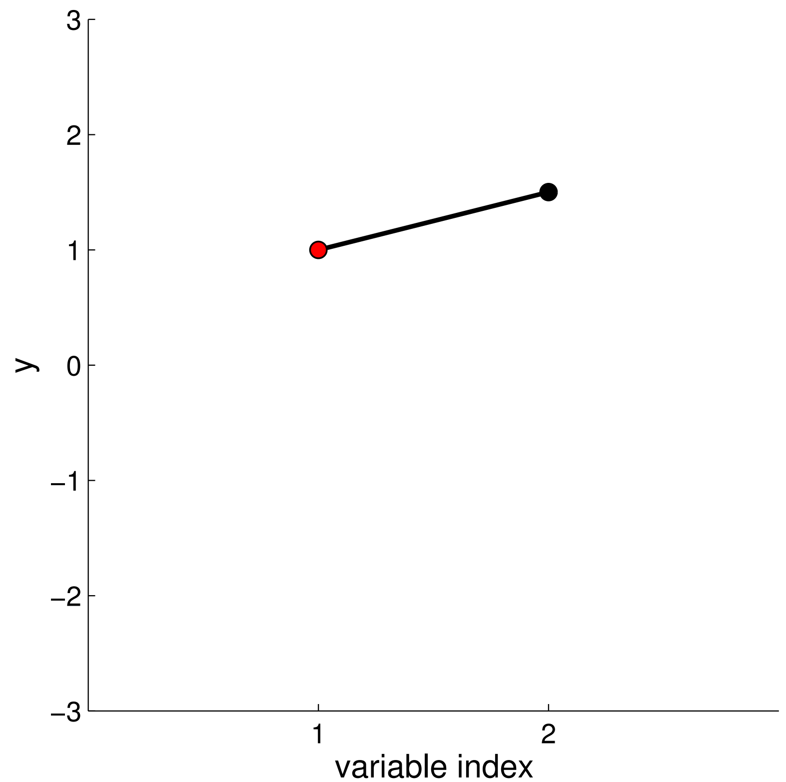
$$\Sigma = \begin{bmatrix} 1 & .9 \\ .9 & 1 \end{bmatrix}$$



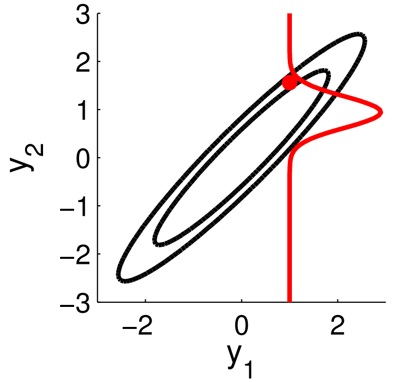
New visualisation



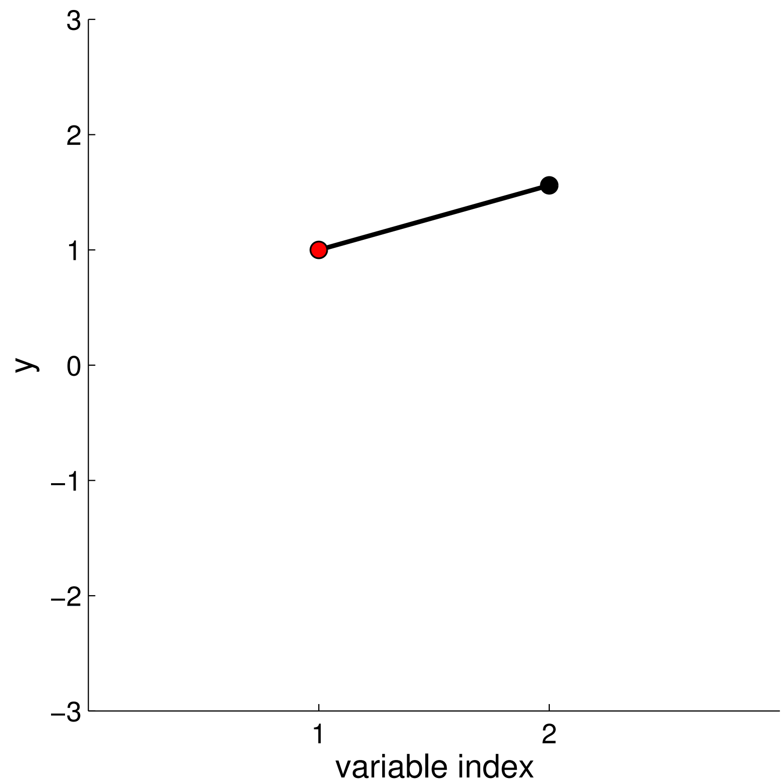
$$\Sigma = \begin{bmatrix} 1 & .9 \\ .9 & 1 \end{bmatrix}$$



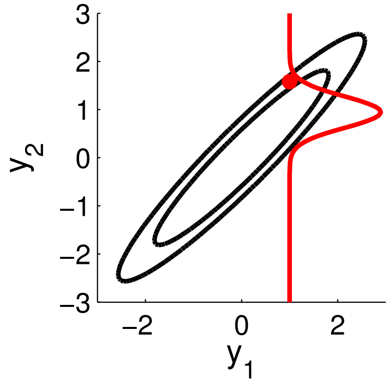
New visualisation

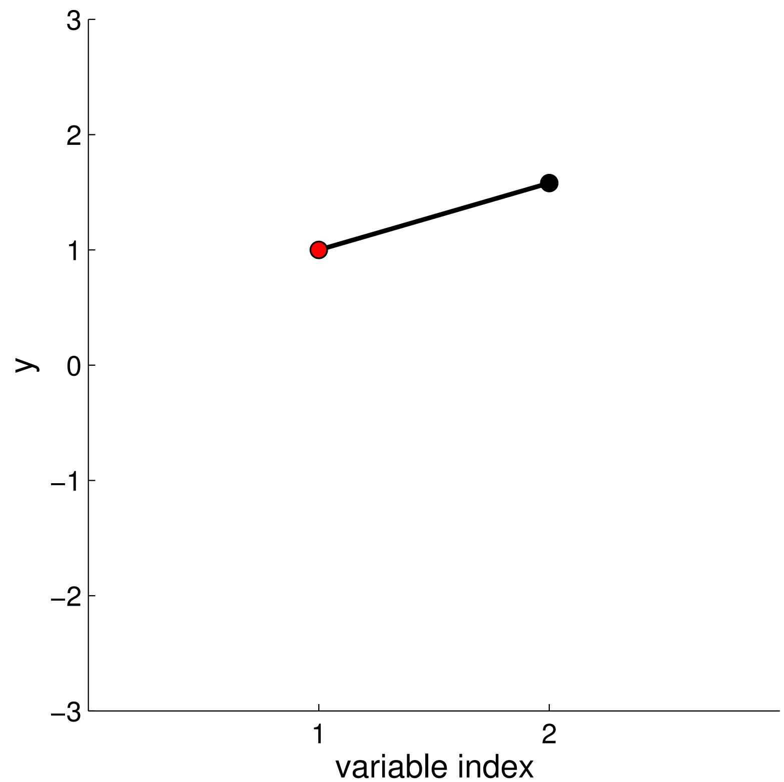


$$\Sigma = \begin{bmatrix} 1 & .9 \\ .9 & 1 \end{bmatrix}$$



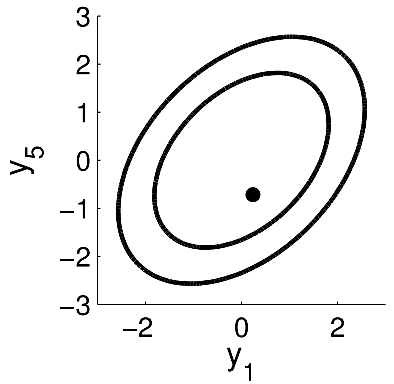
New visualisation

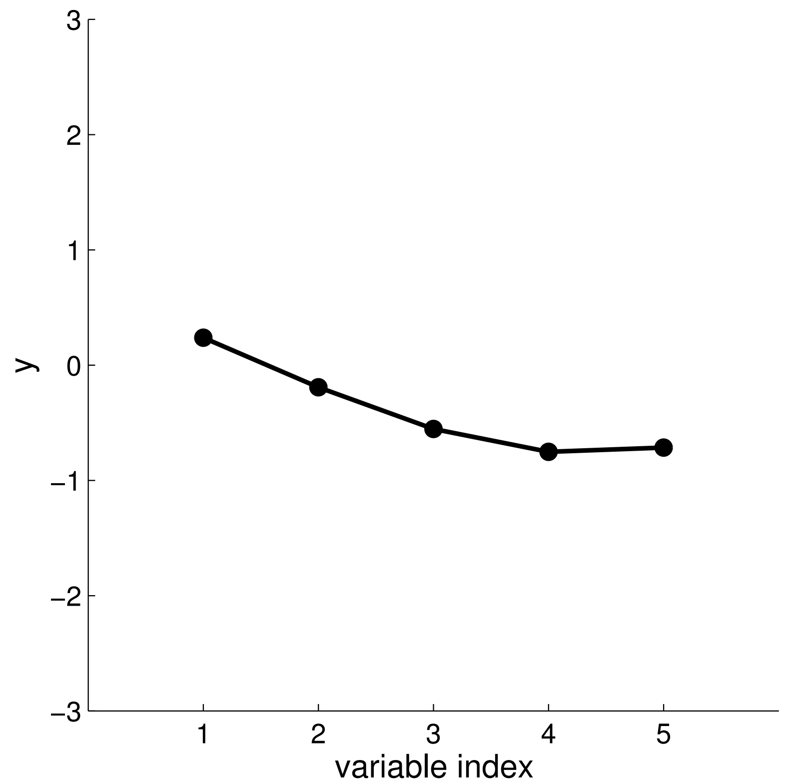




$$\Sigma = \begin{bmatrix} 1 & .9 \\ .9 & 1 \end{bmatrix}$$

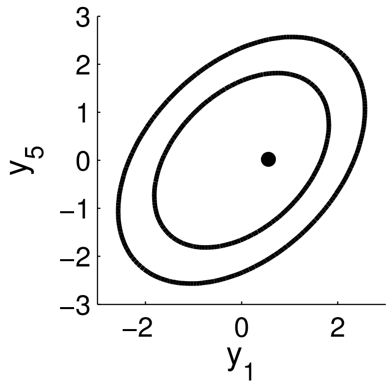
New visualisation

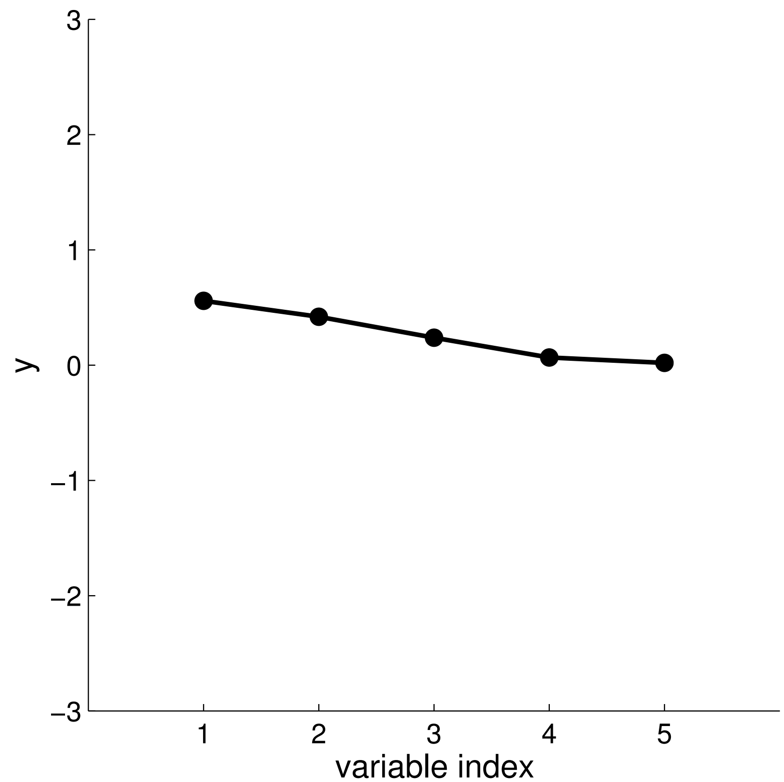




$$\Sigma = \begin{bmatrix} 1 & .9 & .8 & .6 & .4 \\ .9 & 1 & .9 & .8 & .6 \\ .8 & .9 & 1 & .9 & .8 \\ .6 & .8 & .9 & 1 & .9 \\ .4 & .6 & .8 & .9 & 1 \end{bmatrix}$$

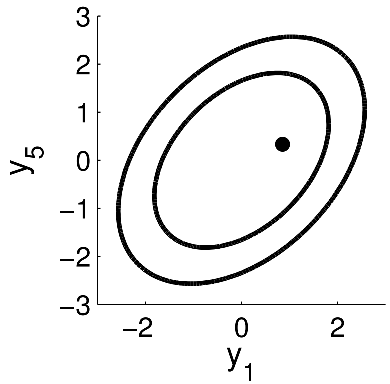
New visualisation

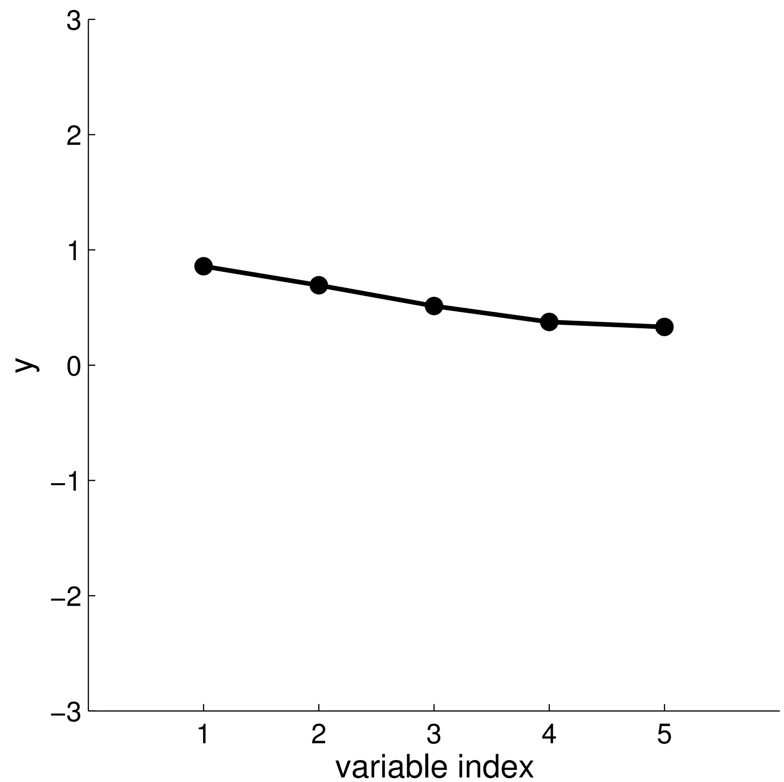




$$\Sigma = \begin{bmatrix} 1 & .9 & .8 & .6 & .4 \\ .9 & 1 & .9 & .8 & .6 \\ .8 & .9 & 1 & .9 & .8 \\ .6 & .8 & .9 & 1 & .9 \\ .4 & .6 & .8 & .9 & 1 \end{bmatrix}$$

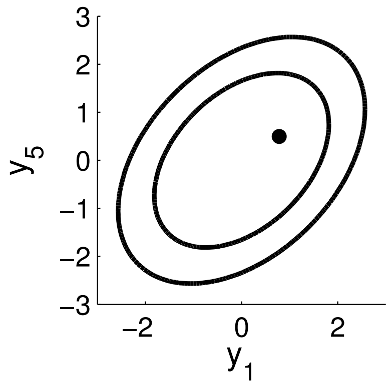
New visualisation

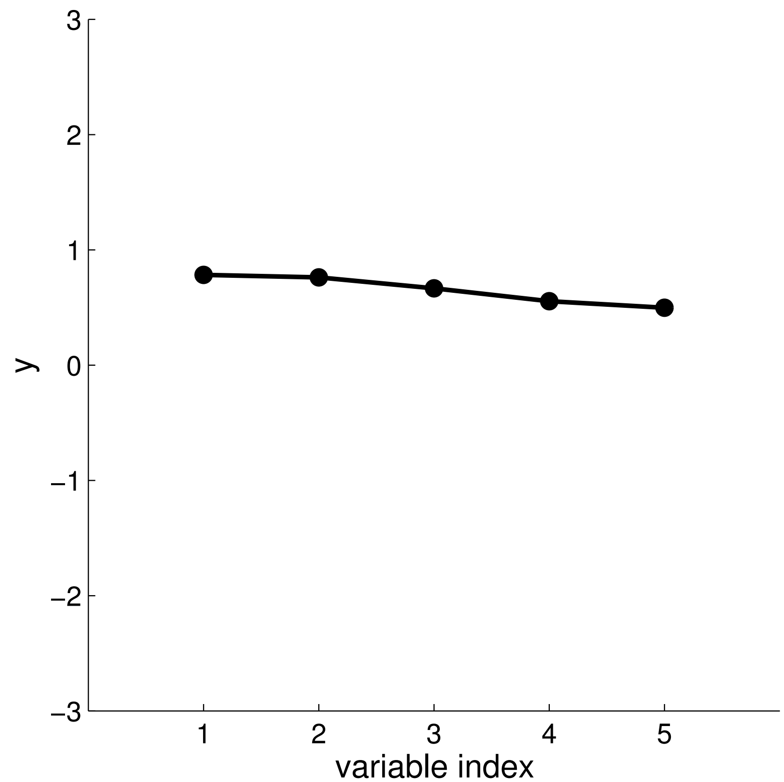




$$\Sigma = \begin{bmatrix} 1 & .9 & .8 & .6 & .4 \\ .9 & 1 & .9 & .8 & .6 \\ .8 & .9 & 1 & .9 & .8 \\ .6 & .8 & .9 & 1 & .9 \\ .4 & .6 & .8 & .9 & 1 \end{bmatrix}$$

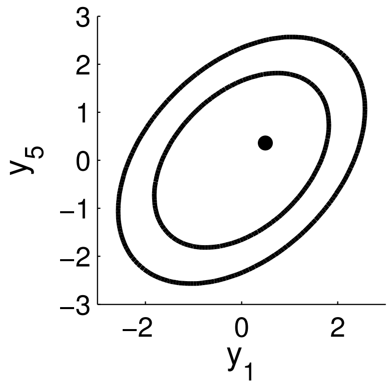
New visualisation

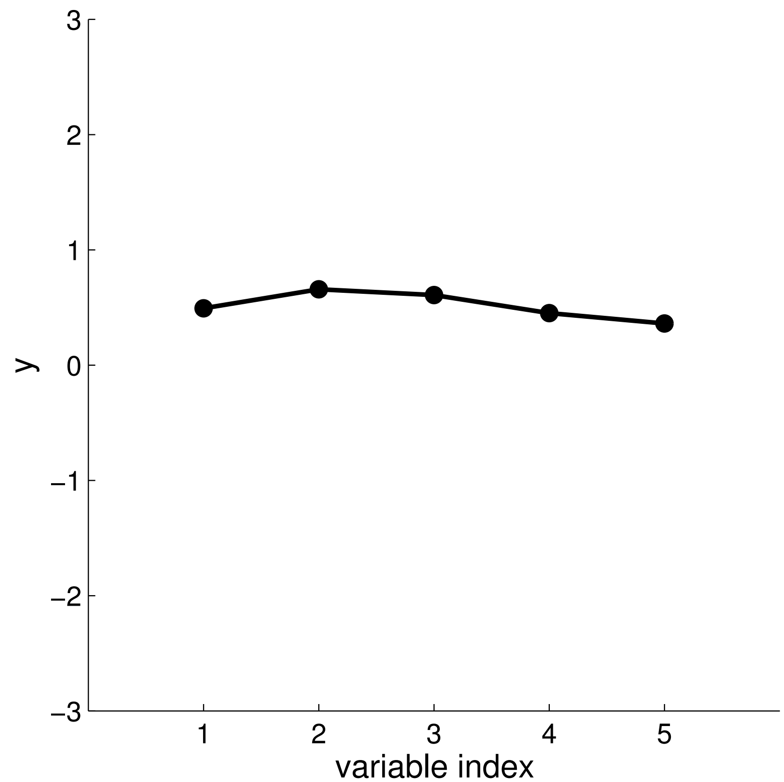




$$\Sigma = \begin{bmatrix} 1 & .9 & .8 & .6 & .4 \\ .9 & 1 & .9 & .8 & .6 \\ .8 & .9 & 1 & .9 & .8 \\ .6 & .8 & .9 & 1 & .9 \\ .4 & .6 & .8 & .9 & 1 \end{bmatrix}$$

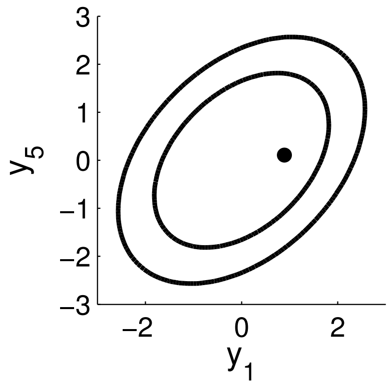
New visualisation



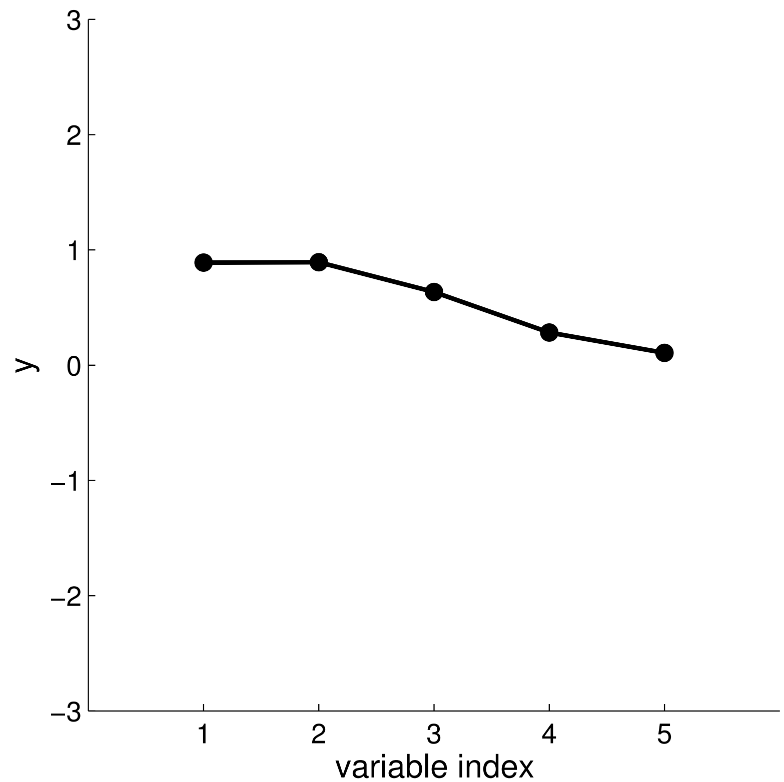


$$\Sigma = \begin{bmatrix} 1 & .9 & .8 & .6 & .4 \\ .9 & 1 & .9 & .8 & .6 \\ .8 & .9 & 1 & .9 & .8 \\ .6 & .8 & .9 & 1 & .9 \\ .4 & .6 & .8 & .9 & 1 \end{bmatrix}$$

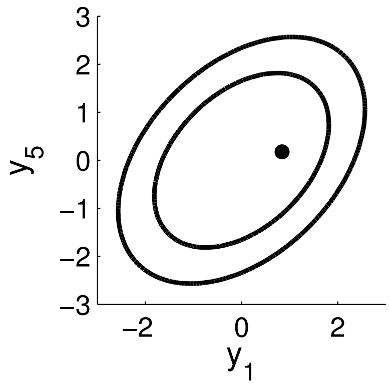
New visualisation

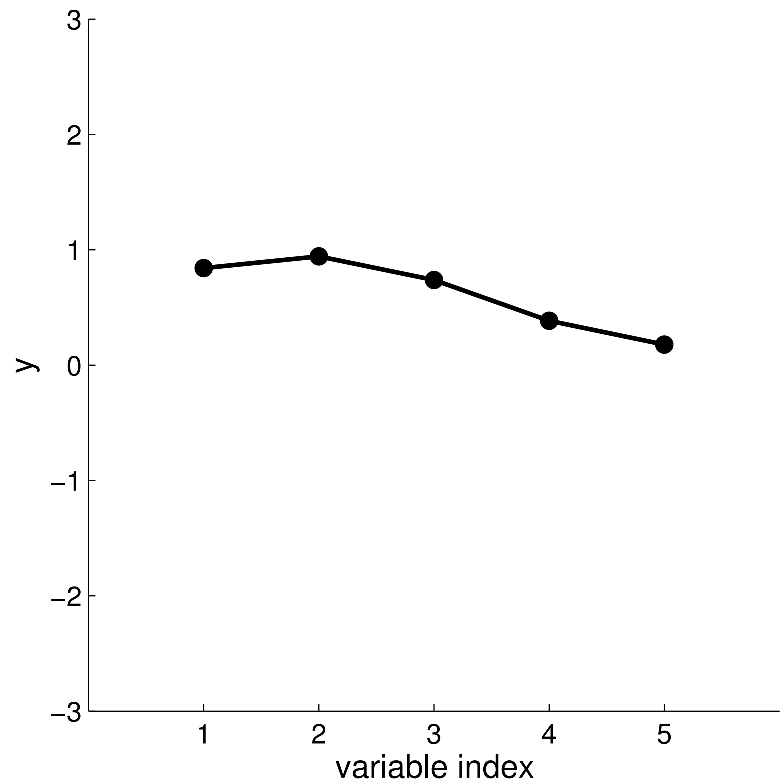


$$\Sigma = \begin{bmatrix} 1 & .9 & .8 & .6 & .4 \\ .9 & 1 & .9 & .8 & .6 \\ .8 & .9 & 1 & .9 & .8 \\ .6 & .8 & .9 & 1 & .9 \\ .4 & .6 & .8 & .9 & 1 \end{bmatrix}$$



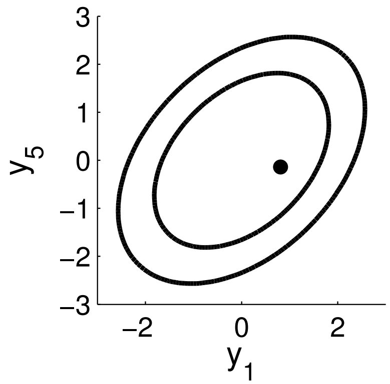
New visualisation

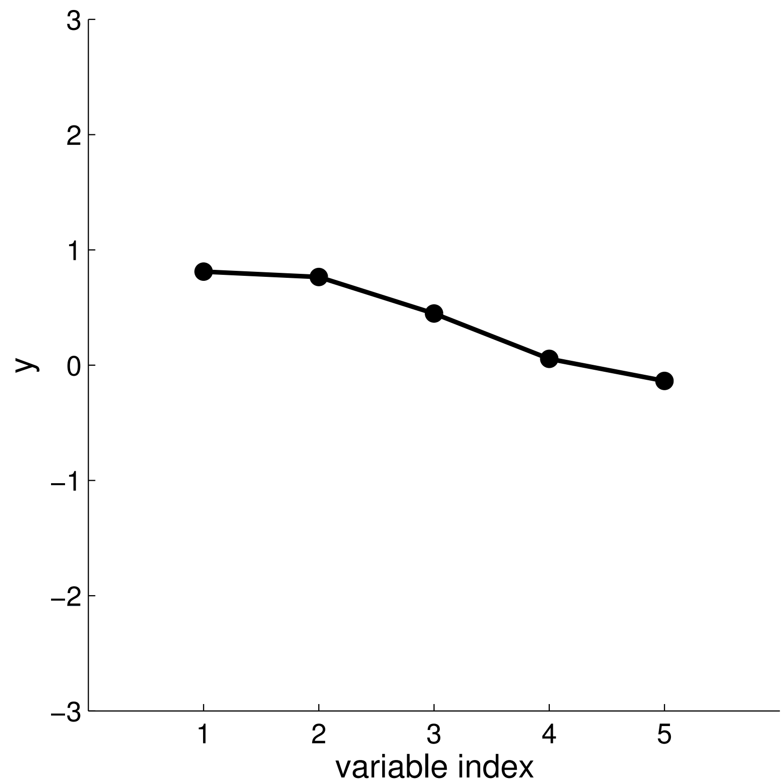




$$\Sigma = \begin{bmatrix} 1 & .9 & .8 & .6 & .4 \\ .9 & 1 & .9 & .8 & .6 \\ .8 & .9 & 1 & .9 & .8 \\ .6 & .8 & .9 & 1 & .9 \\ .4 & .6 & .8 & .9 & 1 \end{bmatrix}$$

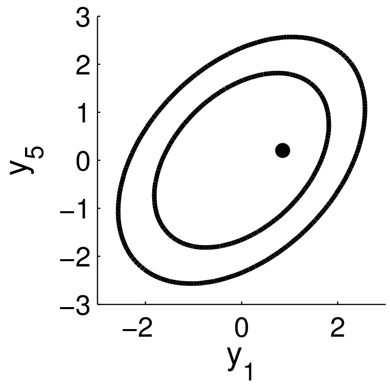
New visualisation

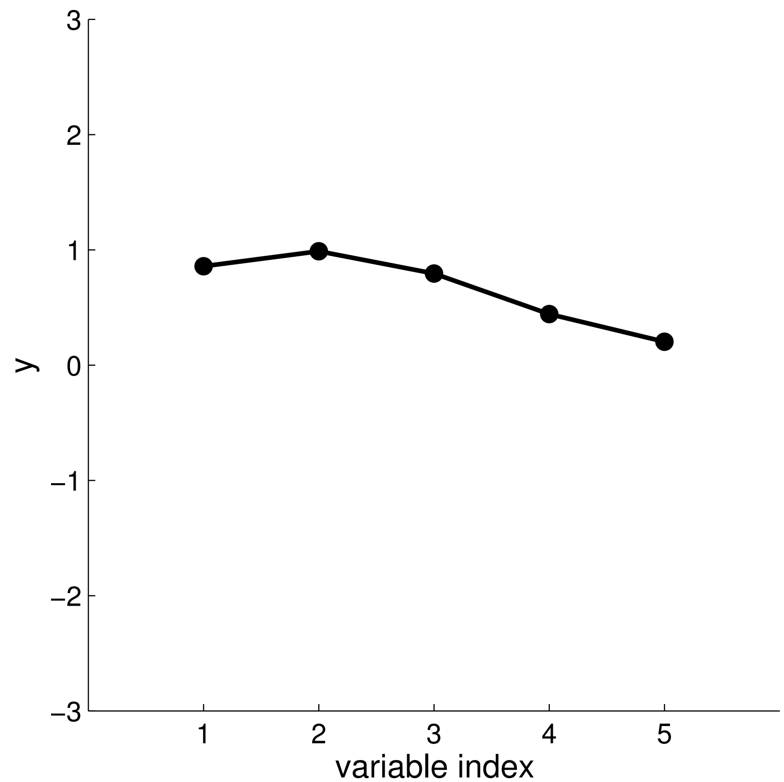




$$\Sigma = \begin{bmatrix} 1 & .9 & .8 & .6 & .4 \\ .9 & 1 & .9 & .8 & .6 \\ .8 & .9 & 1 & .9 & .8 \\ .6 & .8 & .9 & 1 & .9 \\ .4 & .6 & .8 & .9 & 1 \end{bmatrix}$$

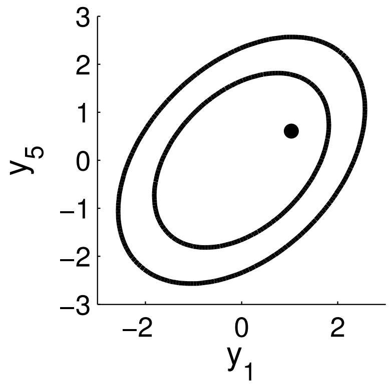
New visualisation

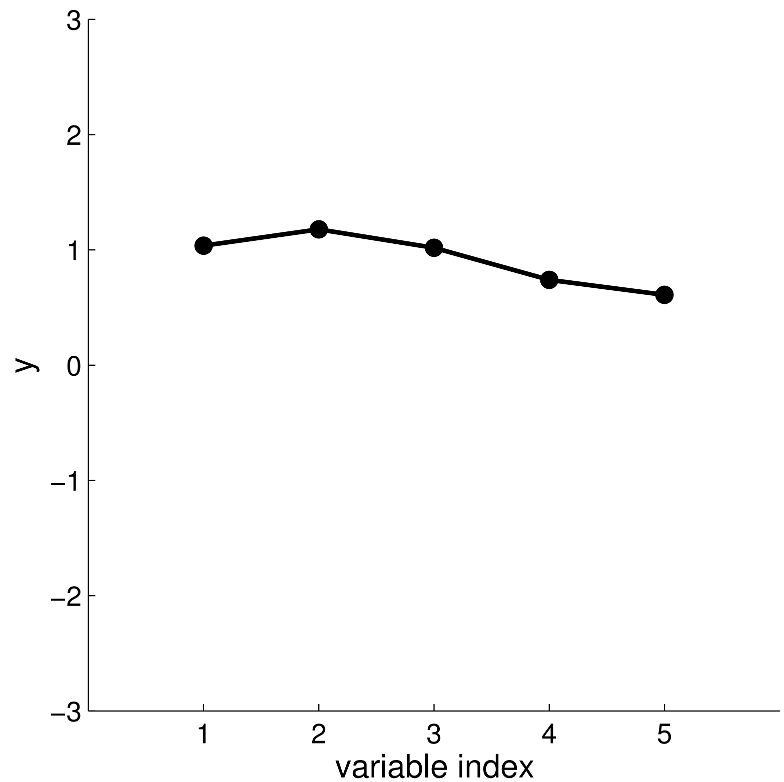




$$\Sigma = \begin{bmatrix} 1 & .9 & .8 & .6 & .4 \\ .9 & 1 & .9 & .8 & .6 \\ .8 & .9 & 1 & .9 & .8 \\ .6 & .8 & .9 & 1 & .9 \\ .4 & .6 & .8 & .9 & 1 \end{bmatrix}$$

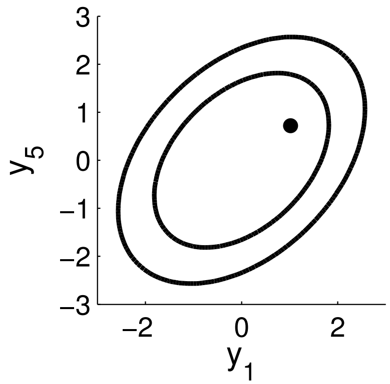
New visualisation

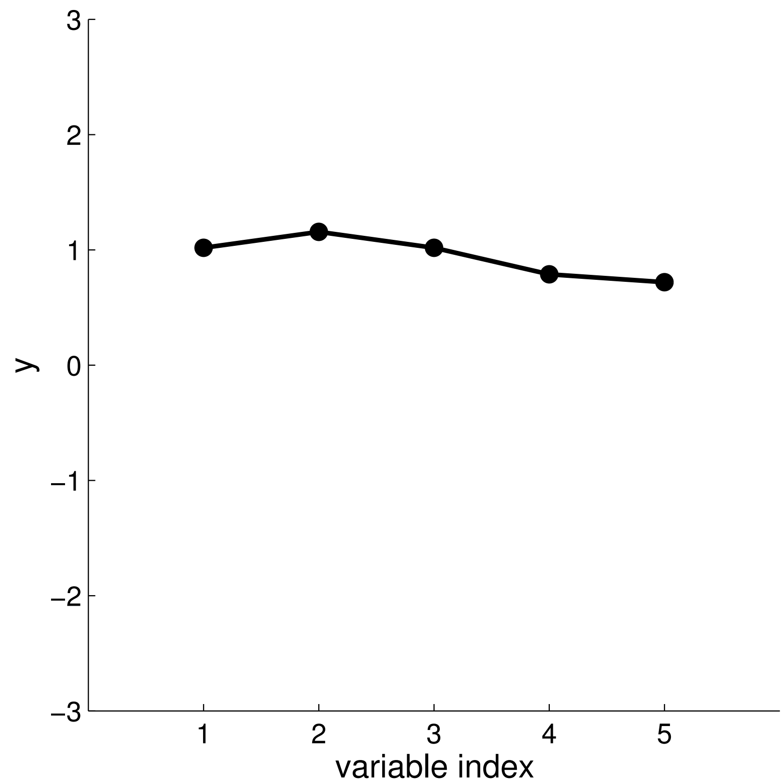




$$\Sigma = \begin{bmatrix} 1 & .9 & .8 & .6 & .4 \\ .9 & 1 & .9 & .8 & .6 \\ .8 & .9 & 1 & .9 & .8 \\ .6 & .8 & .9 & 1 & .9 \\ .4 & .6 & .8 & .9 & 1 \end{bmatrix}$$

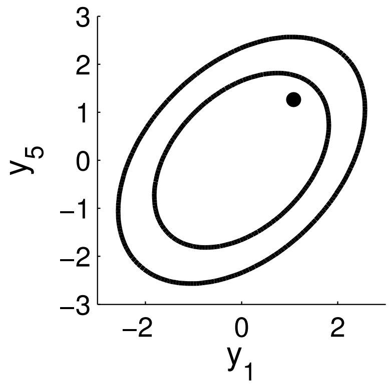
New visualisation

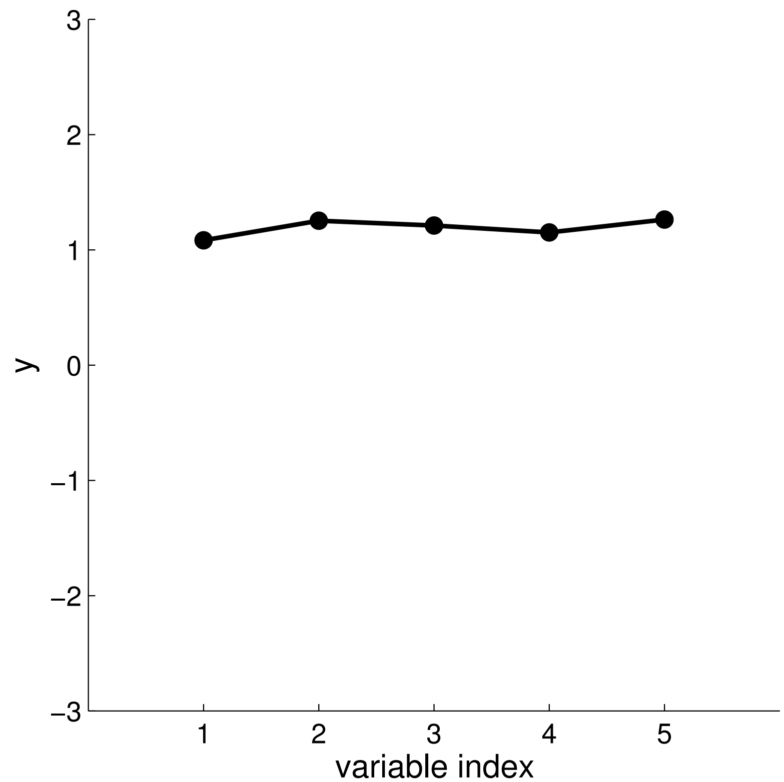




$$\Sigma = \begin{bmatrix} 1 & .9 & .8 & .6 & .4 \\ .9 & 1 & .9 & .8 & .6 \\ .8 & .9 & 1 & .9 & .8 \\ .6 & .8 & .9 & 1 & .9 \\ .4 & .6 & .8 & .9 & 1 \end{bmatrix}$$

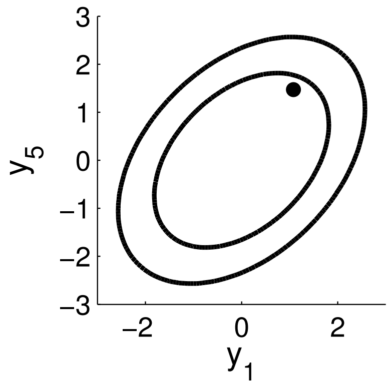
New visualisation



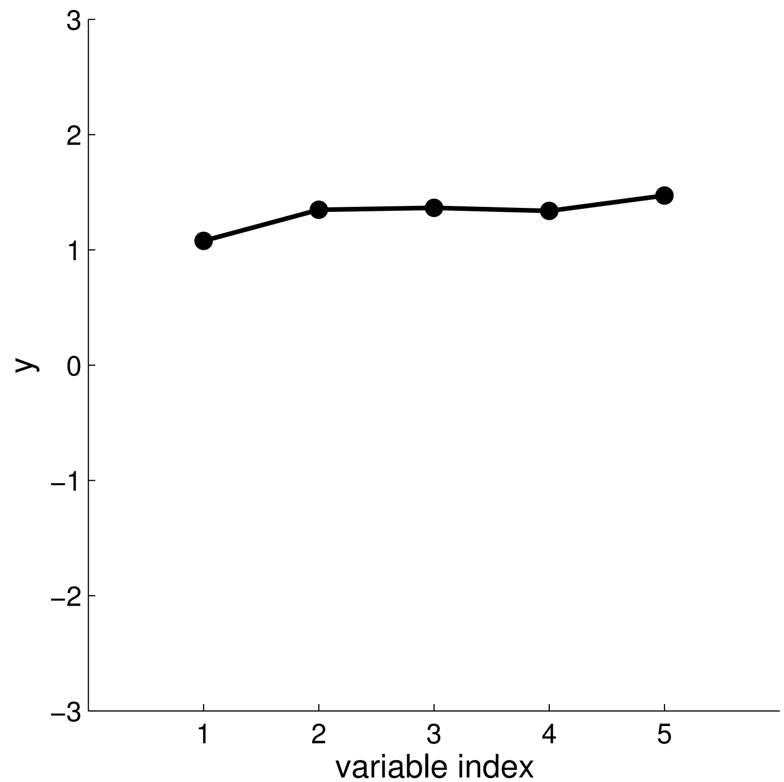


$$\Sigma = \begin{bmatrix} 1 & .9 & .8 & .6 & .4 \\ .9 & 1 & .9 & .8 & .6 \\ .8 & .9 & 1 & .9 & .8 \\ .6 & .8 & .9 & 1 & .9 \\ .4 & .6 & .8 & .9 & 1 \end{bmatrix}$$

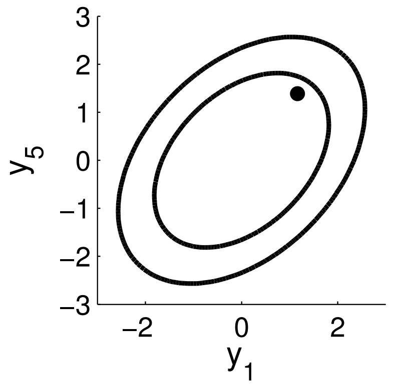
New visualisation

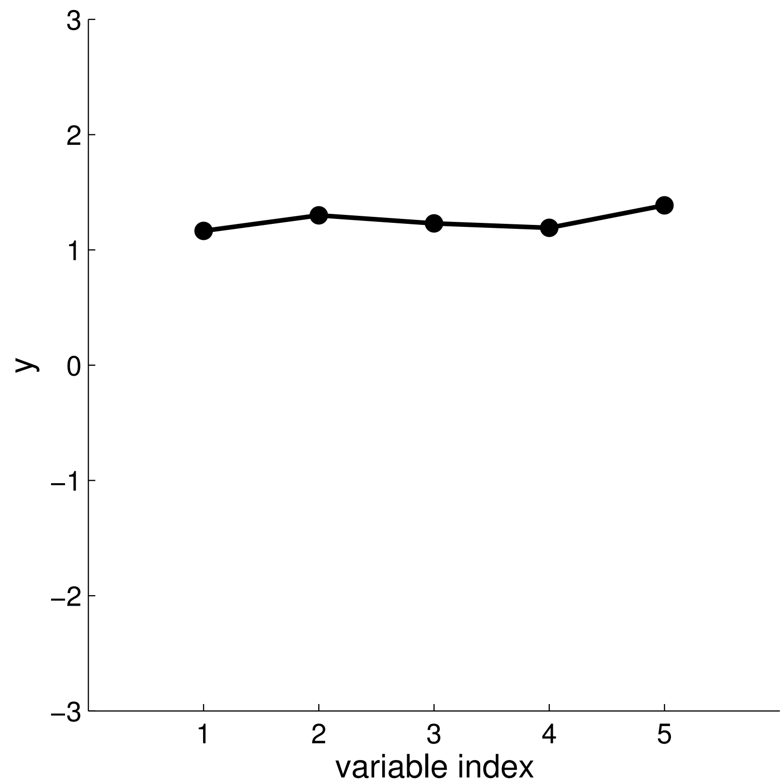


$$\Sigma = \begin{bmatrix} 1 & .9 & .8 & .6 & .4 \\ .9 & 1 & .9 & .8 & .6 \\ .8 & .9 & 1 & .9 & .8 \\ .6 & .8 & .9 & 1 & .9 \\ .4 & .6 & .8 & .9 & 1 \end{bmatrix}$$



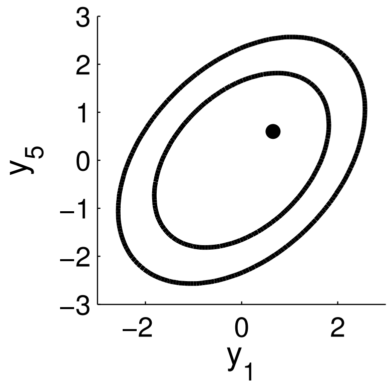
New visualisation

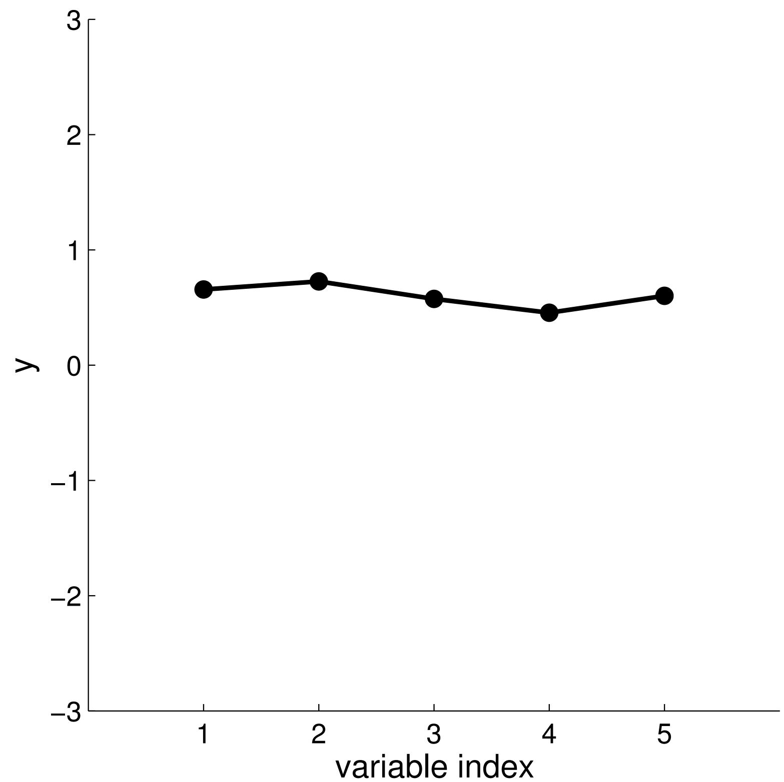




$$\Sigma = \begin{bmatrix} 1 & .9 & .8 & .6 & .4 \\ .9 & 1 & .9 & .8 & .6 \\ .8 & .9 & 1 & .9 & .8 \\ .6 & .8 & .9 & 1 & .9 \\ .4 & .6 & .8 & .9 & 1 \end{bmatrix}$$

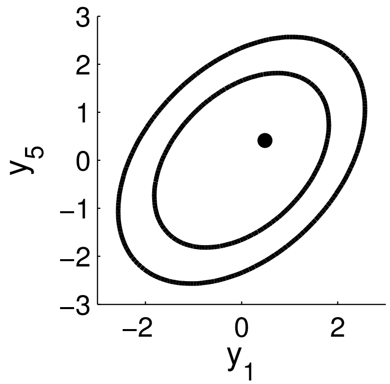
New visualisation

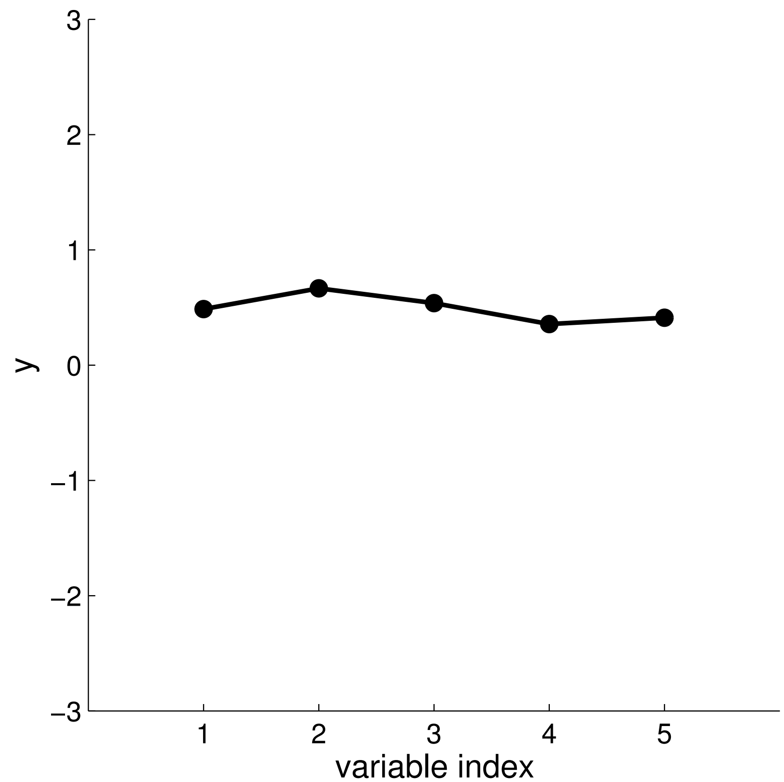




$$\Sigma = \begin{bmatrix} 1 & .9 & .8 & .6 & .4 \\ .9 & 1 & .9 & .8 & .6 \\ .8 & .9 & 1 & .9 & .8 \\ .6 & .8 & .9 & 1 & .9 \\ .4 & .6 & .8 & .9 & 1 \end{bmatrix}$$

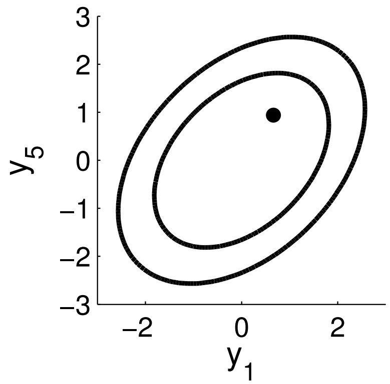
New visualisation

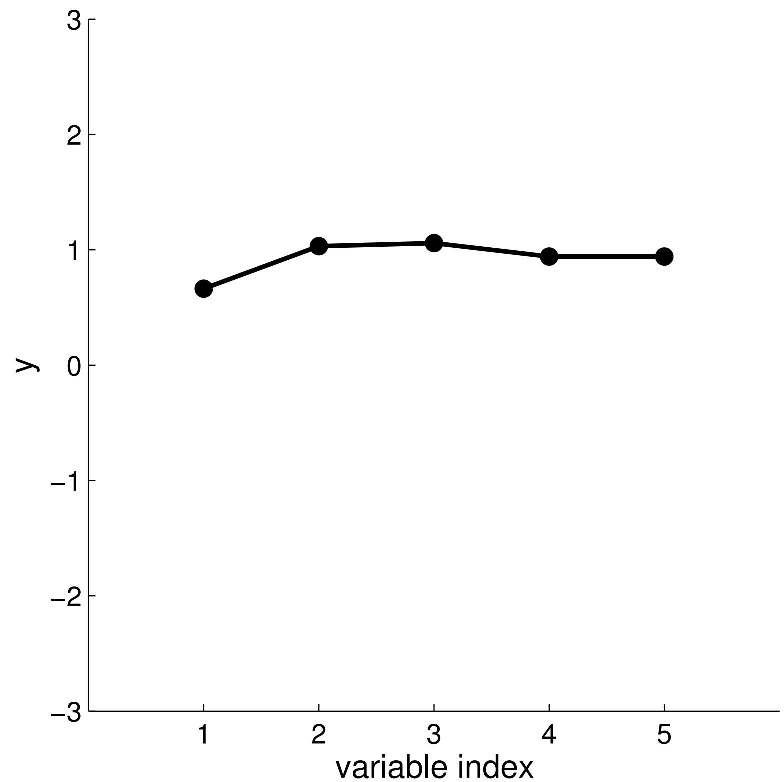




$$\Sigma = \begin{bmatrix} 1 & .9 & .8 & .6 & .4 \\ .9 & 1 & .9 & .8 & .6 \\ .8 & .9 & 1 & .9 & .8 \\ .6 & .8 & .9 & 1 & .9 \\ .4 & .6 & .8 & .9 & 1 \end{bmatrix}$$

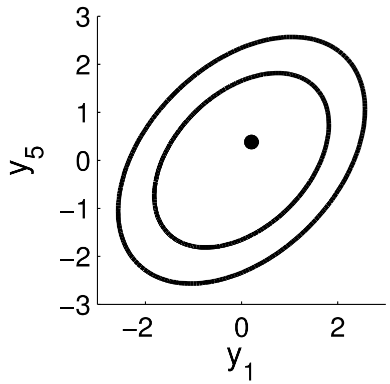
New visualisation

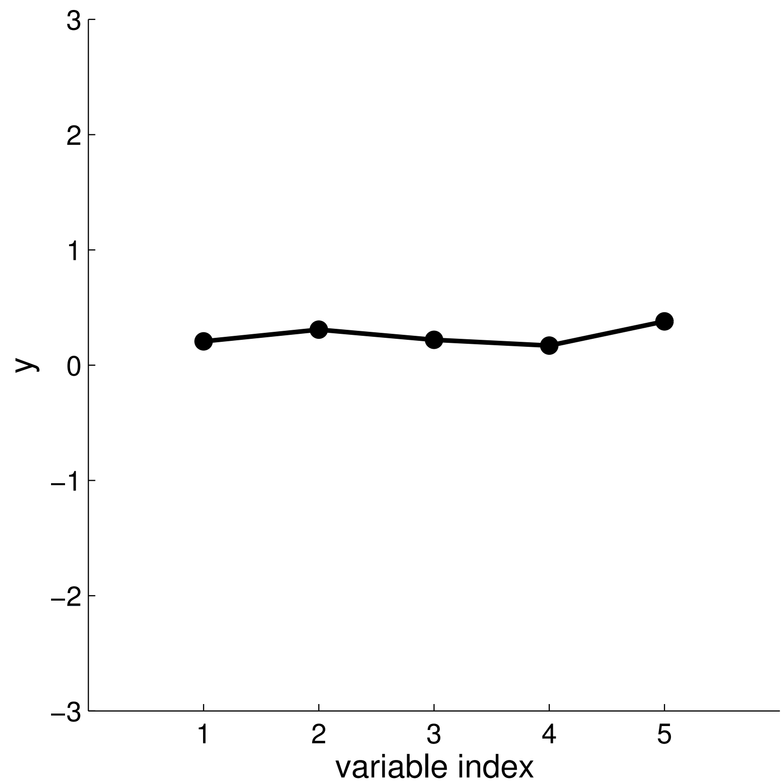




$$\Sigma = \begin{bmatrix} 1 & .9 & .8 & .6 & .4 \\ .9 & 1 & .9 & .8 & .6 \\ .8 & .9 & 1 & .9 & .8 \\ .6 & .8 & .9 & 1 & .9 \\ .4 & .6 & .8 & .9 & 1 \end{bmatrix}$$

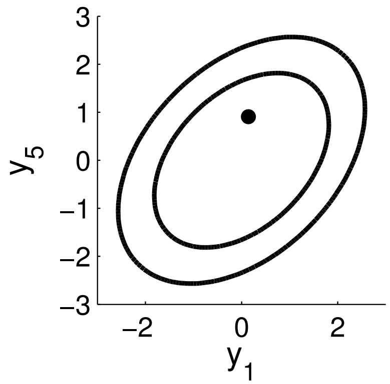
New visualisation

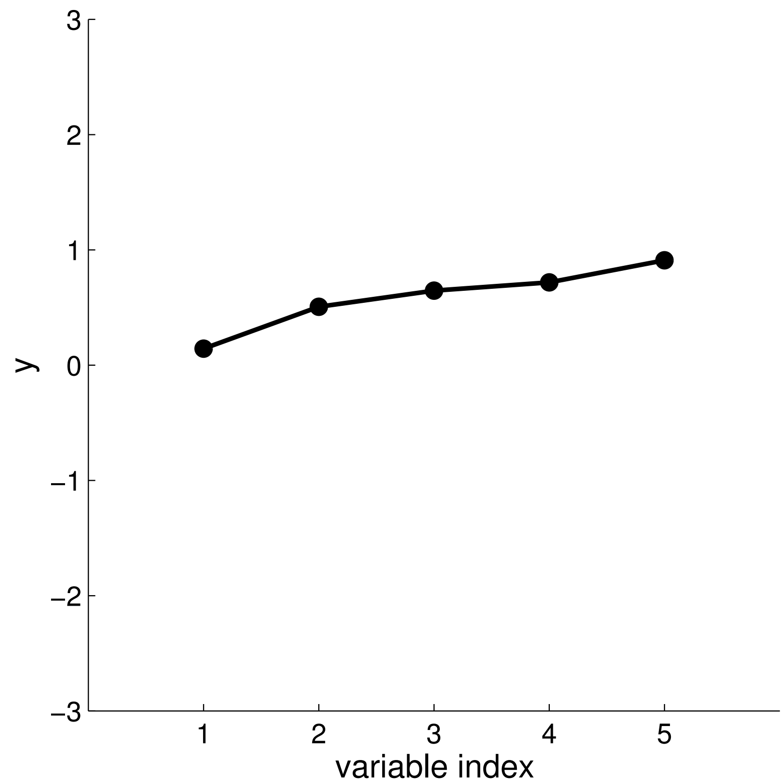




$$\Sigma = \begin{bmatrix} 1 & .9 & .8 & .6 & .4 \\ .9 & 1 & .9 & .8 & .6 \\ .8 & .9 & 1 & .9 & .8 \\ .6 & .8 & .9 & 1 & .9 \\ .4 & .6 & .8 & .9 & 1 \end{bmatrix}$$

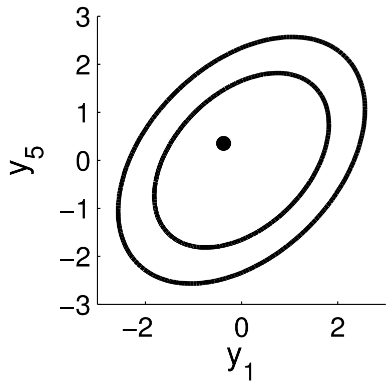
New visualisation

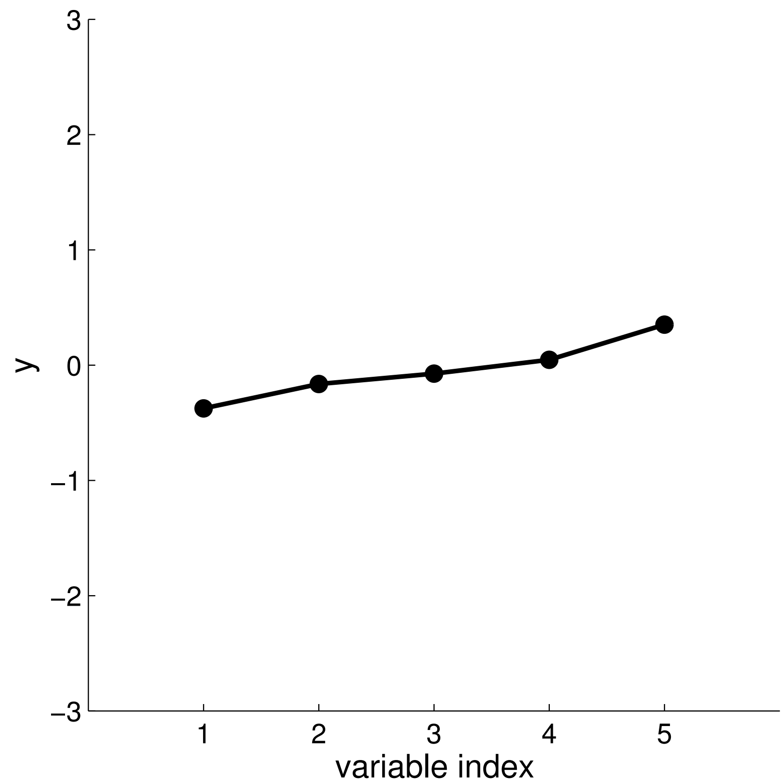




$$\Sigma = \begin{bmatrix} 1 & .9 & .8 & .6 & .4 \\ .9 & 1 & .9 & .8 & .6 \\ .8 & .9 & 1 & .9 & .8 \\ .6 & .8 & .9 & 1 & .9 \\ .4 & .6 & .8 & .9 & 1 \end{bmatrix}$$

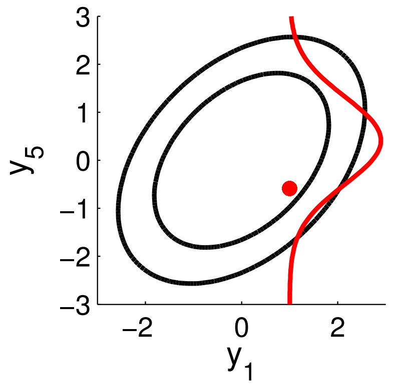
New visualisation

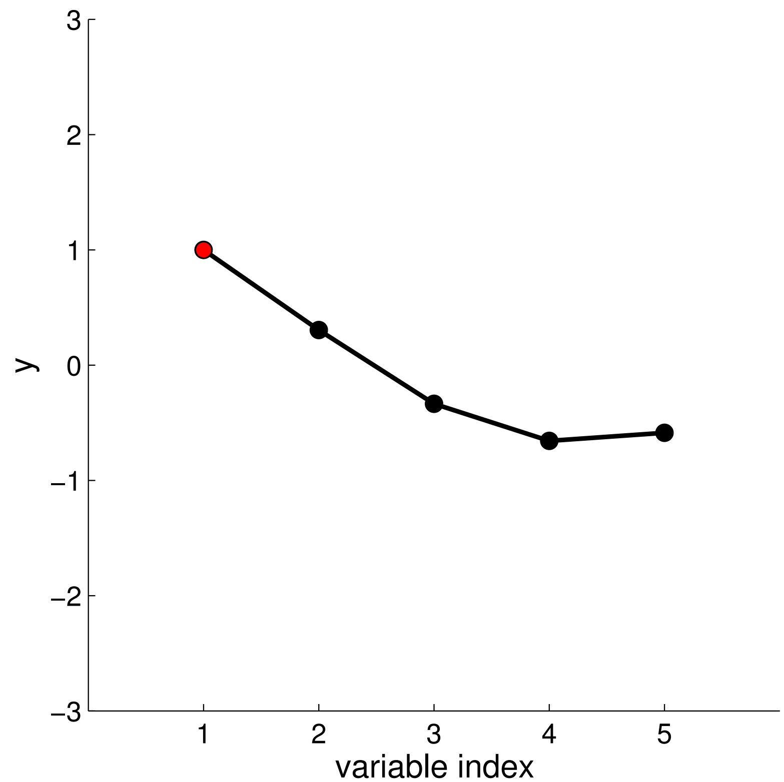




$$\Sigma = \begin{bmatrix} 1 & .9 & .8 & .6 & .4 \\ .9 & 1 & .9 & .8 & .6 \\ .8 & .9 & 1 & .9 & .8 \\ .6 & .8 & .9 & 1 & .9 \\ .4 & .6 & .8 & .9 & 1 \end{bmatrix}$$

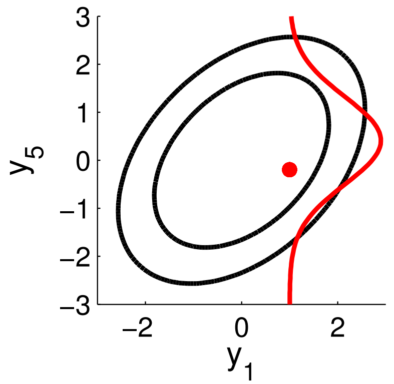
New visualisation

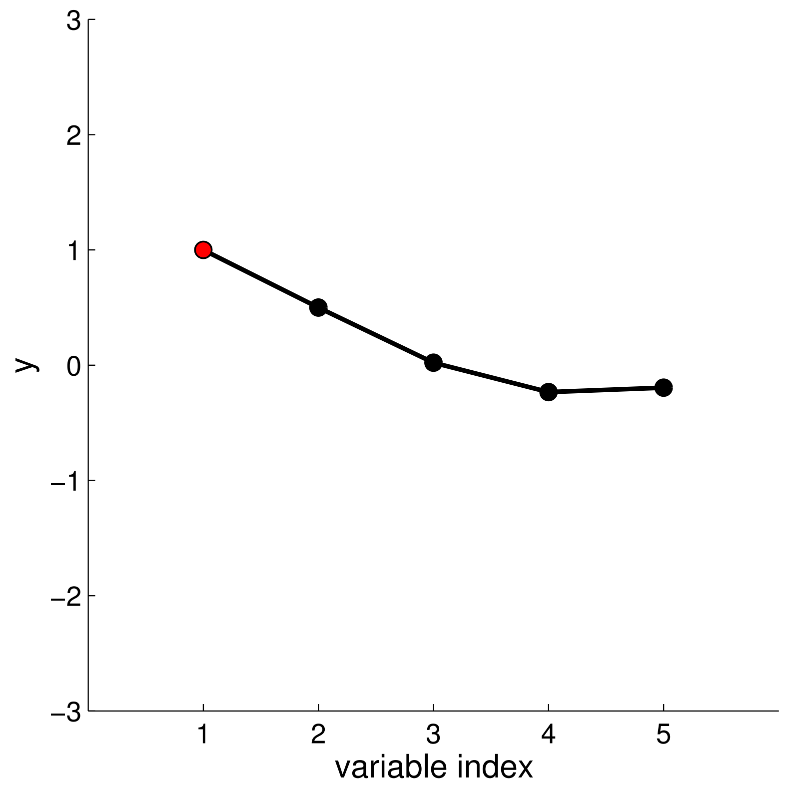




$$\Sigma = \begin{bmatrix} 1 & .9 & .8 & .6 & .4 \\ .9 & 1 & .9 & .8 & .6 \\ .8 & .9 & 1 & .9 & .8 \\ .6 & .8 & .9 & 1 & .9 \\ .4 & .6 & .8 & .9 & 1 \end{bmatrix}$$

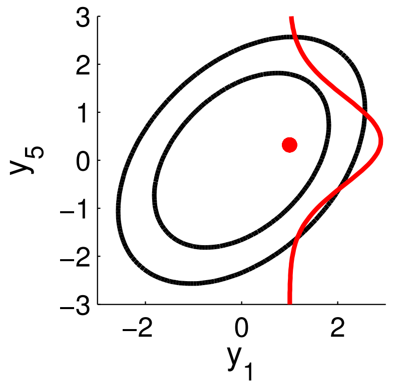
New visualisation

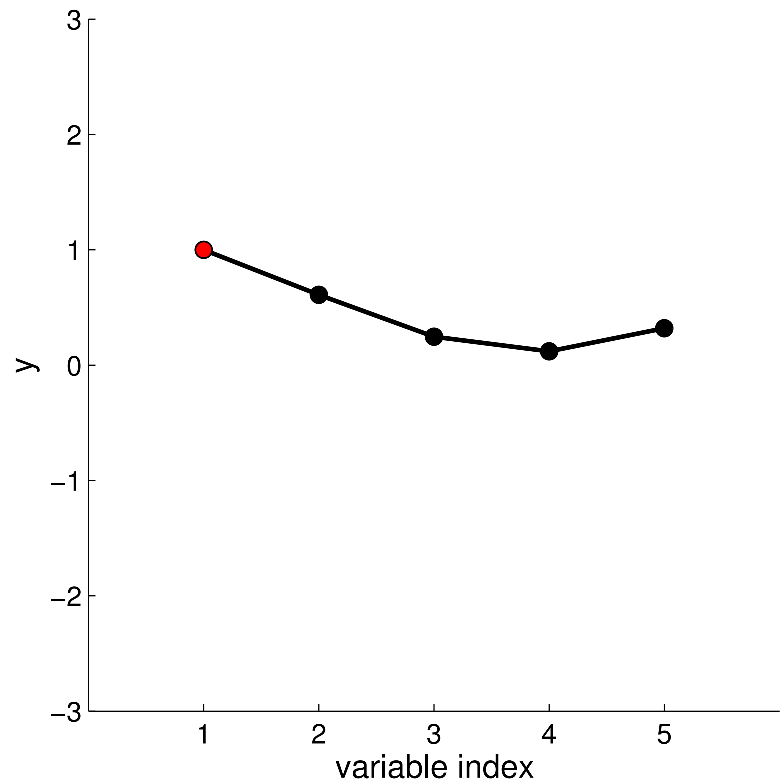




$$\Sigma = \begin{bmatrix} 1 & .9 & .8 & .6 & .4 \\ .9 & 1 & .9 & .8 & .6 \\ .8 & .9 & 1 & .9 & .8 \\ .6 & .8 & .9 & 1 & .9 \\ .4 & .6 & .8 & .9 & 1 \end{bmatrix}$$

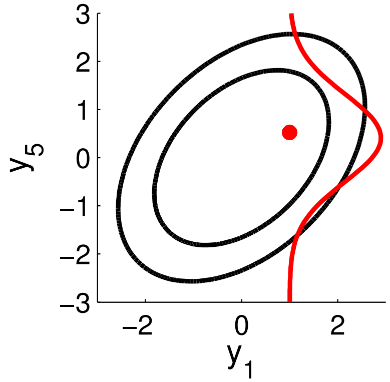
New visualisation

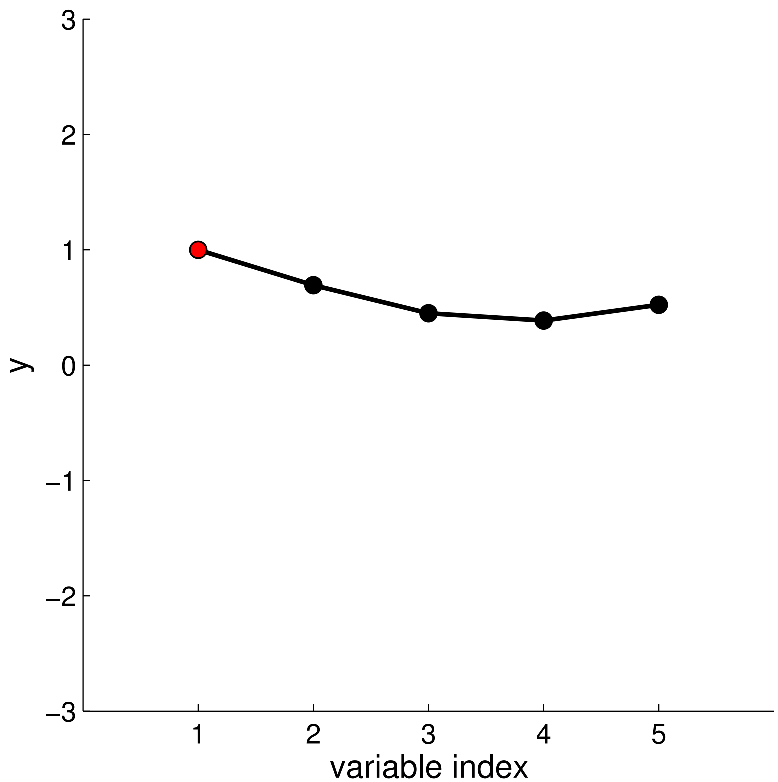




$$\Sigma = \begin{bmatrix} 1 & .9 & .8 & .6 & .4 \\ .9 & 1 & .9 & .8 & .6 \\ .8 & .9 & 1 & .9 & .8 \\ .6 & .8 & .9 & 1 & .9 \\ .4 & .6 & .8 & .9 & 1 \end{bmatrix}$$

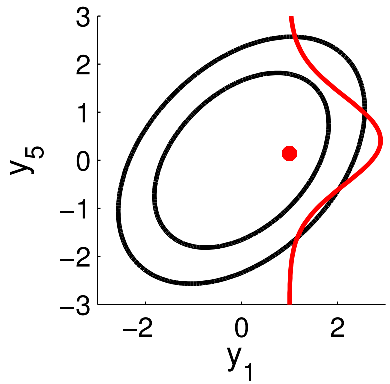
New visualisation

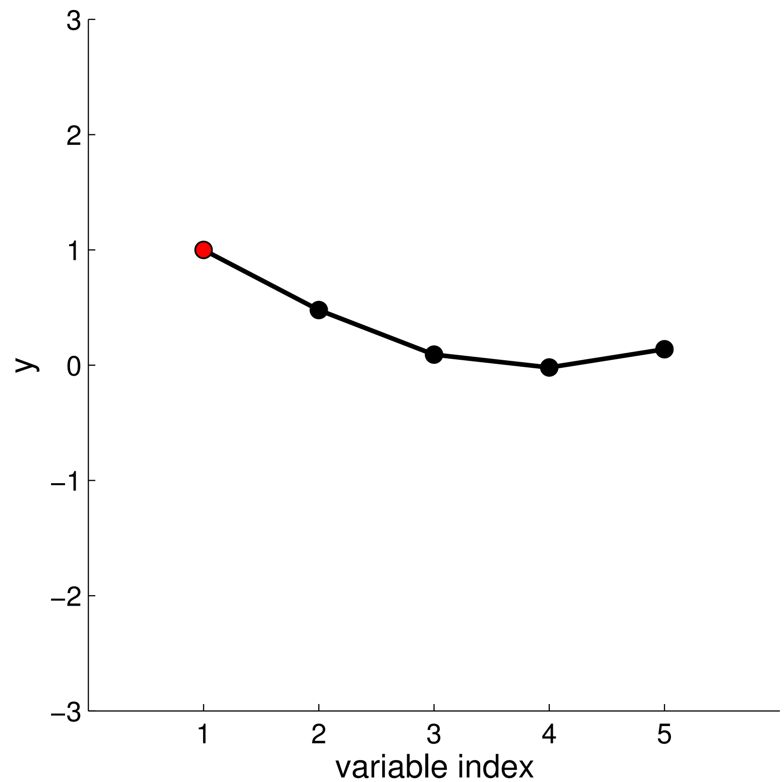




$$\Sigma = \begin{bmatrix} 1 & .9 & .8 & .6 & .4 \\ .9 & 1 & .9 & .8 & .6 \\ .8 & .9 & 1 & .9 & .8 \\ .6 & .8 & .9 & 1 & .9 \\ .4 & .6 & .8 & .9 & 1 \end{bmatrix}$$

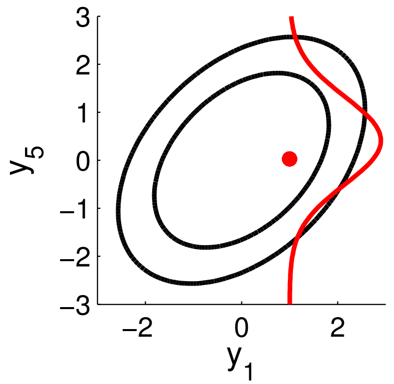
New visualisation

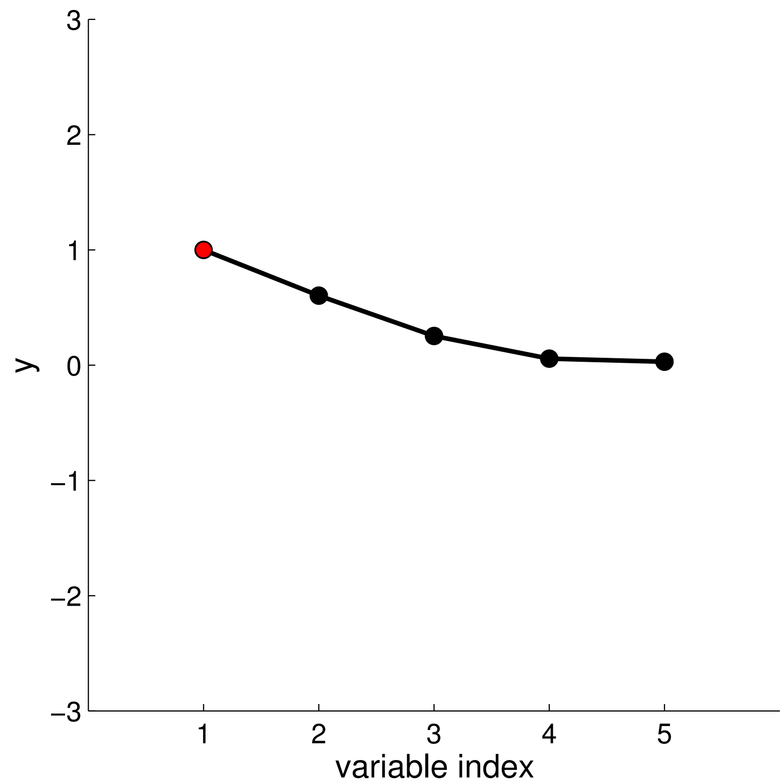




$$\Sigma = \begin{bmatrix} 1 & .9 & .8 & .6 & .4 \\ .9 & 1 & .9 & .8 & .6 \\ .8 & .9 & 1 & .9 & .8 \\ .6 & .8 & .9 & 1 & .9 \\ .4 & .6 & .8 & .9 & 1 \end{bmatrix}$$

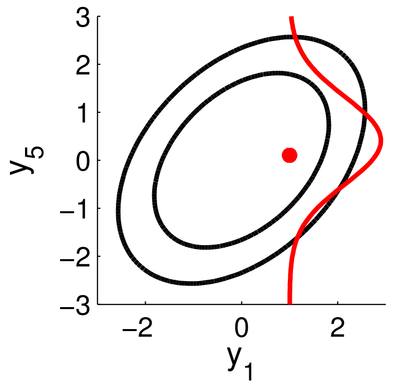
New visualisation

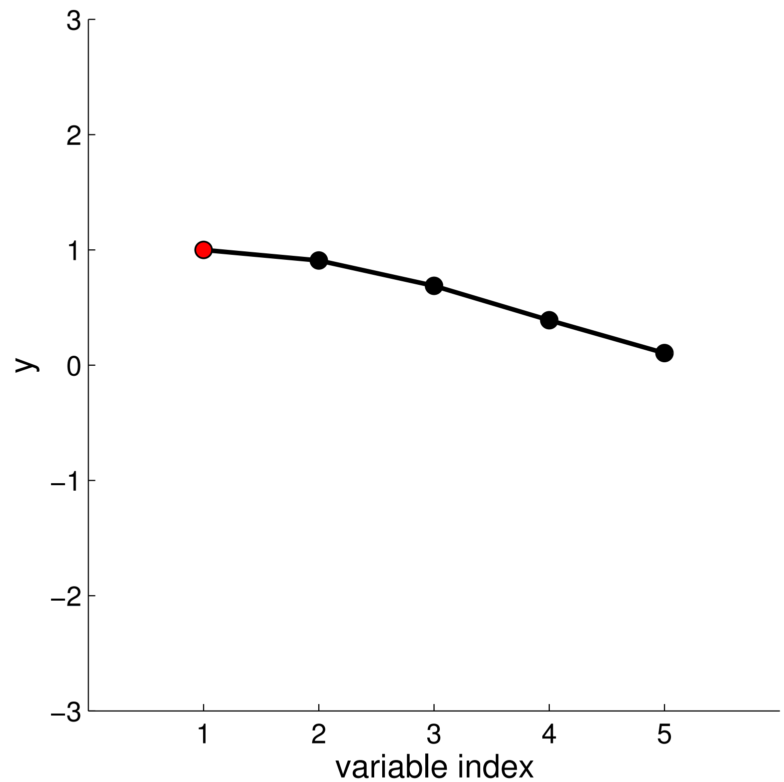




$$\Sigma = \begin{bmatrix} 1 & .9 & .8 & .6 & .4 \\ .9 & 1 & .9 & .8 & .6 \\ .8 & .9 & 1 & .9 & .8 \\ .6 & .8 & .9 & 1 & .9 \\ .4 & .6 & .8 & .9 & 1 \end{bmatrix}$$

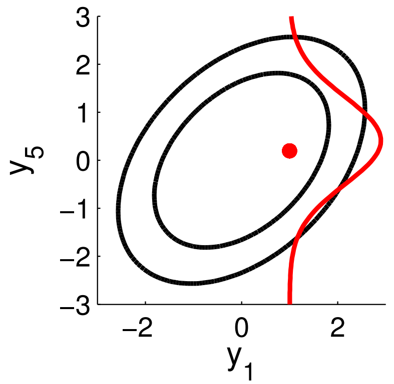
New visualisation

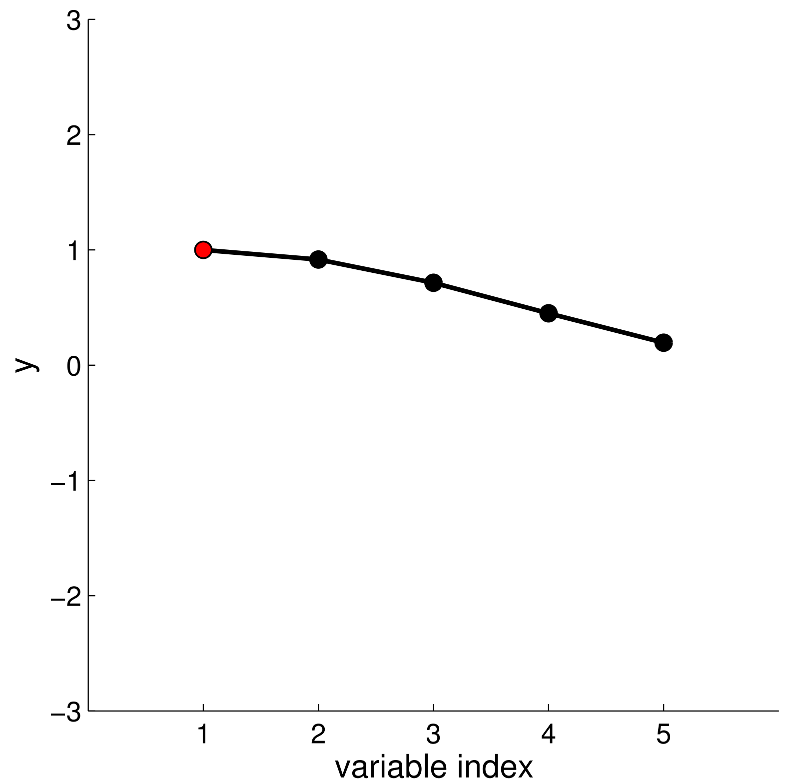




$$\Sigma = \begin{bmatrix} 1 & .9 & .8 & .6 & .4 \\ .9 & 1 & .9 & .8 & .6 \\ .8 & .9 & 1 & .9 & .8 \\ .6 & .8 & .9 & 1 & .9 \\ .4 & .6 & .8 & .9 & 1 \end{bmatrix}$$

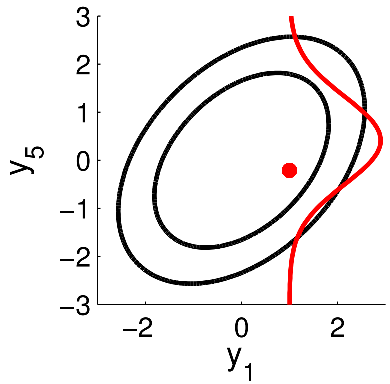
New visualisation

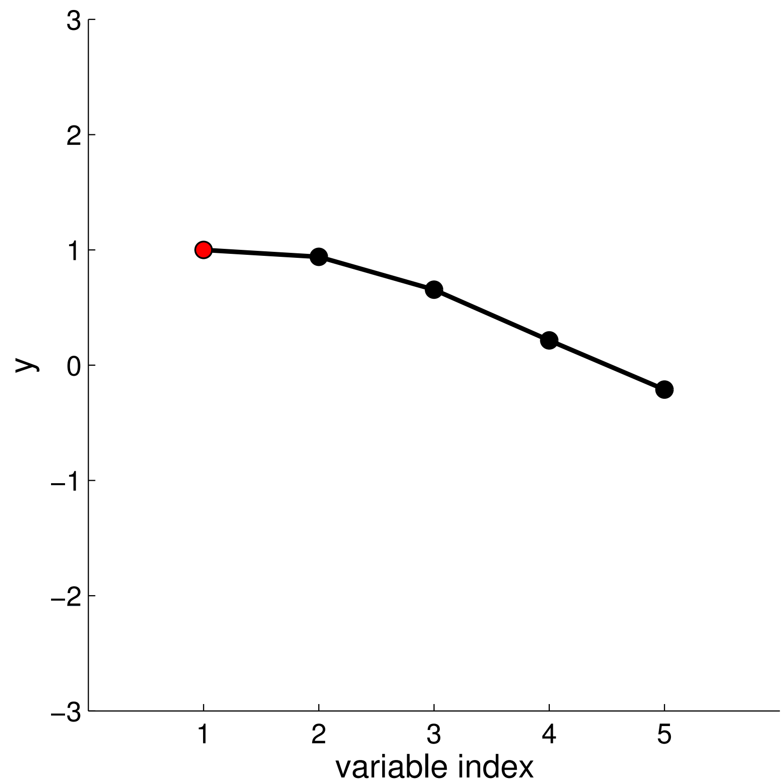




$$\Sigma = \begin{bmatrix} 1 & .9 & .8 & .6 & .4 \\ .9 & 1 & .9 & .8 & .6 \\ .8 & .9 & 1 & .9 & .8 \\ .6 & .8 & .9 & 1 & .9 \\ .4 & .6 & .8 & .9 & 1 \end{bmatrix}$$

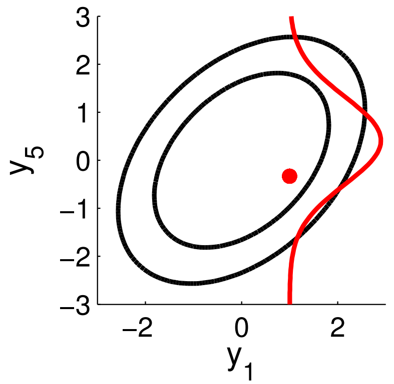
New visualisation

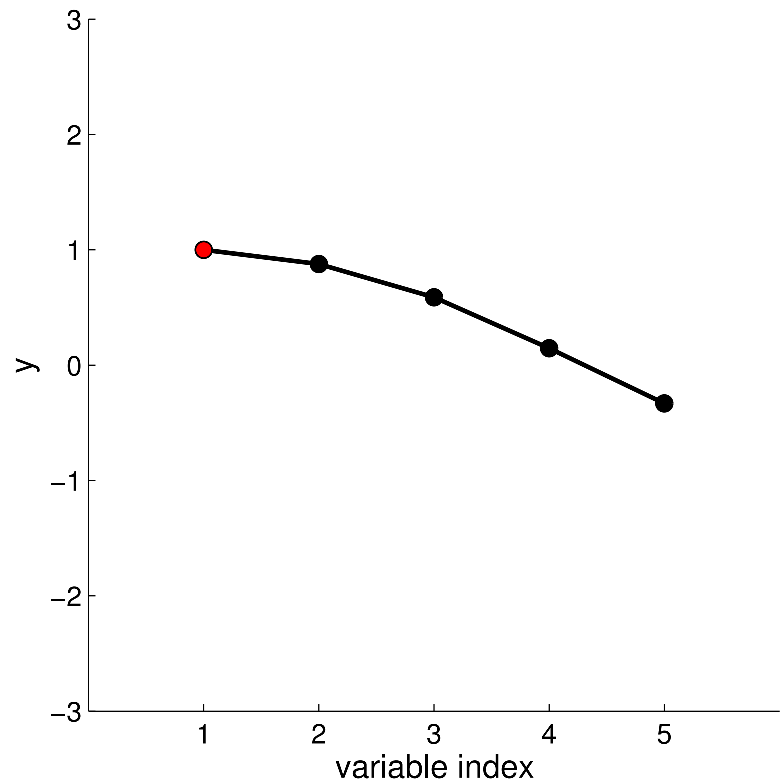




$$\Sigma = \begin{bmatrix} 1 & .9 & .8 & .6 & .4 \\ .9 & 1 & .9 & .8 & .6 \\ .8 & .9 & 1 & .9 & .8 \\ .6 & .8 & .9 & 1 & .9 \\ .4 & .6 & .8 & .9 & 1 \end{bmatrix}$$

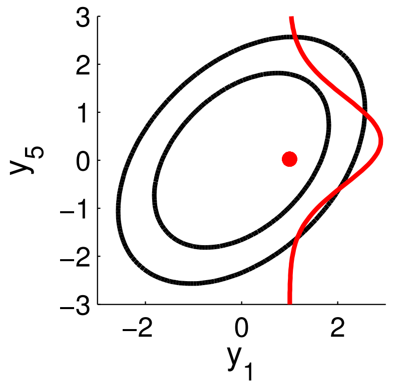
New visualisation

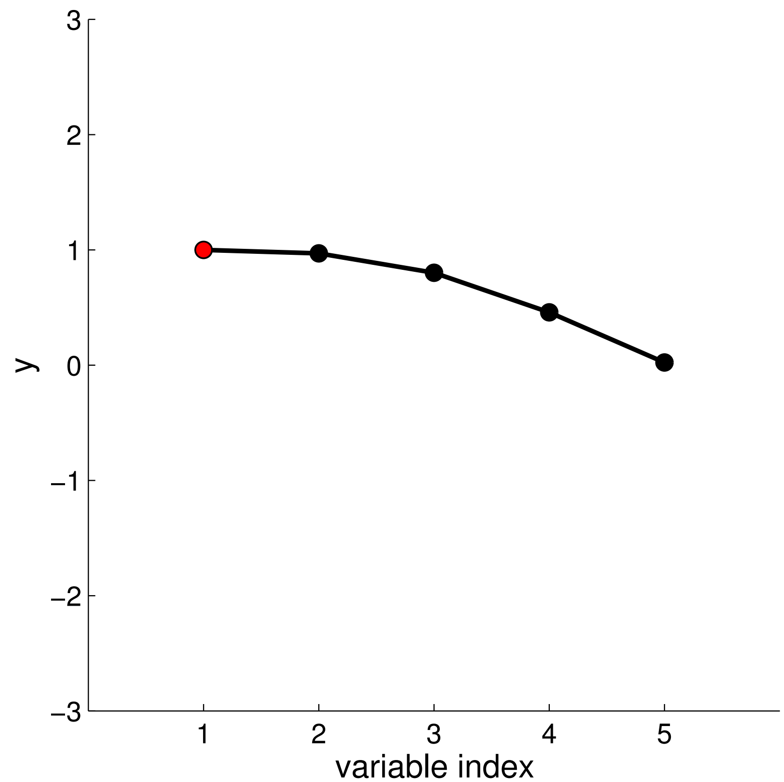




$$\Sigma = \begin{bmatrix} 1 & .9 & .8 & .6 & .4 \\ .9 & 1 & .9 & .8 & .6 \\ .8 & .9 & 1 & .9 & .8 \\ .6 & .8 & .9 & 1 & .9 \\ .4 & .6 & .8 & .9 & 1 \end{bmatrix}$$

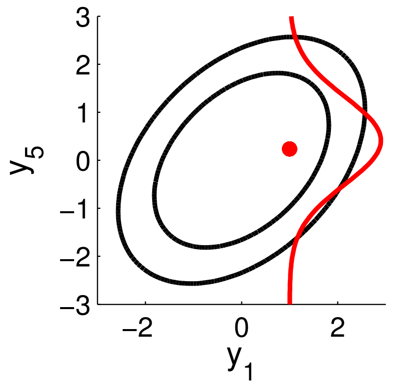
New visualisation

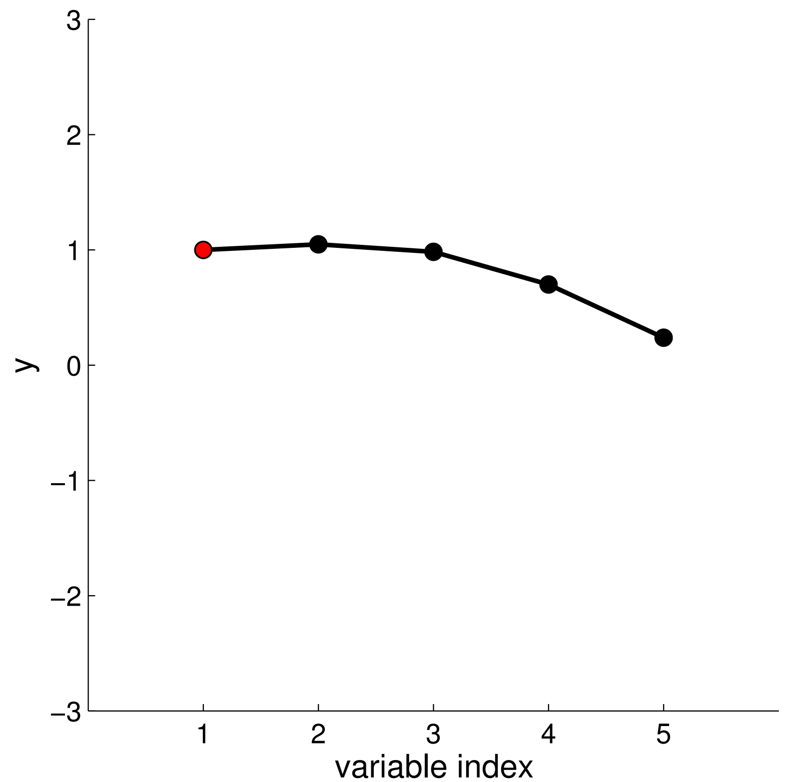




$$\Sigma = \begin{bmatrix} 1 & .9 & .8 & .6 & .4 \\ .9 & 1 & .9 & .8 & .6 \\ .8 & .9 & 1 & .9 & .8 \\ .6 & .8 & .9 & 1 & .9 \\ .4 & .6 & .8 & .9 & 1 \end{bmatrix}$$

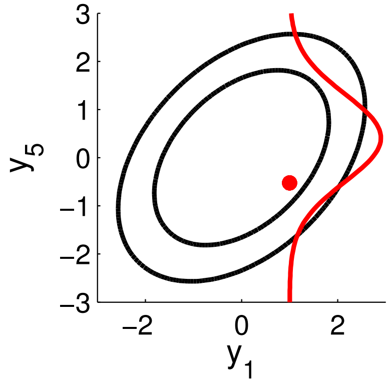
New visualisation

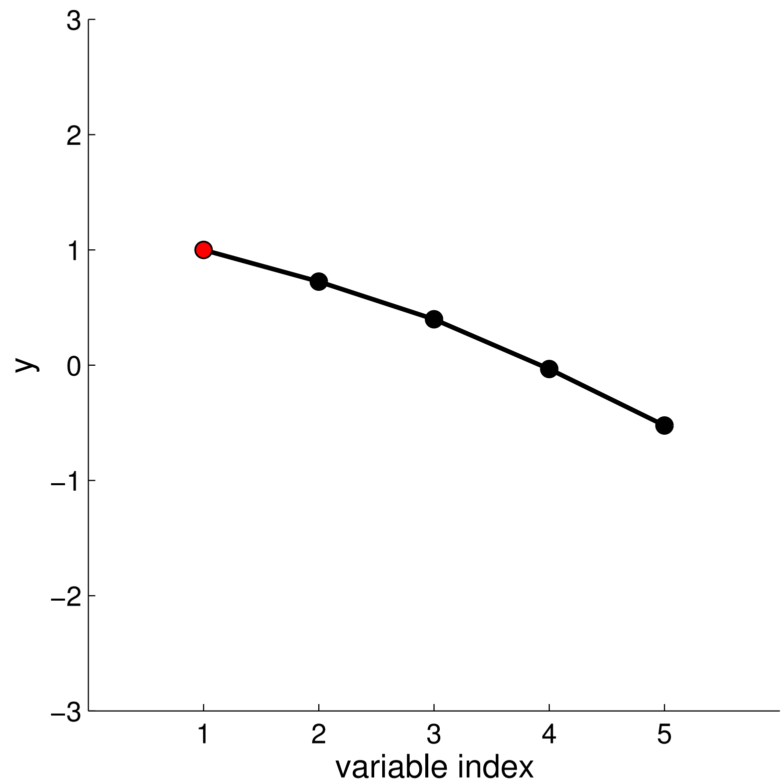




$$\Sigma = \begin{bmatrix} 1 & .9 & .8 & .6 & .4 \\ .9 & 1 & .9 & .8 & .6 \\ .8 & .9 & 1 & .9 & .8 \\ .6 & .8 & .9 & 1 & .9 \\ .4 & .6 & .8 & .9 & 1 \end{bmatrix}$$

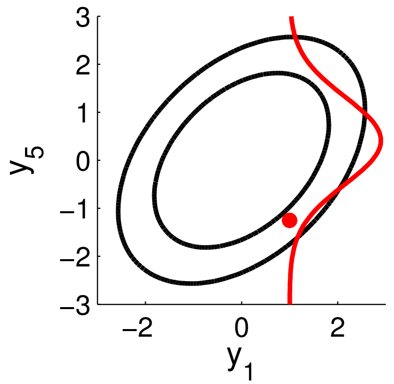
New visualisation

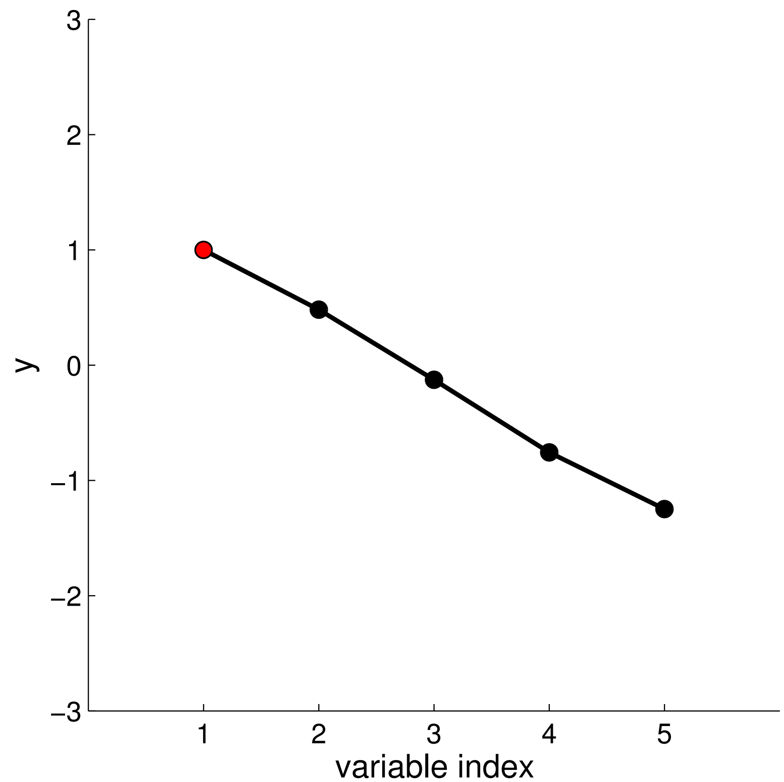




$$\Sigma = \begin{bmatrix} 1 & .9 & .8 & .6 & .4 \\ .9 & 1 & .9 & .8 & .6 \\ .8 & .9 & 1 & .9 & .8 \\ .6 & .8 & .9 & 1 & .9 \\ .4 & .6 & .8 & .9 & 1 \end{bmatrix}$$

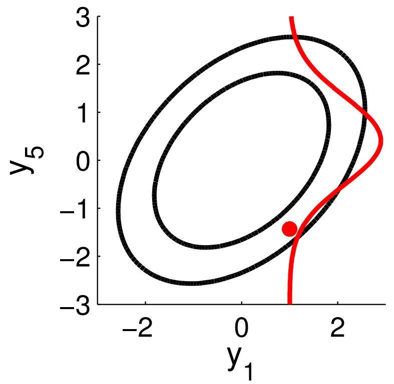
New visualisation

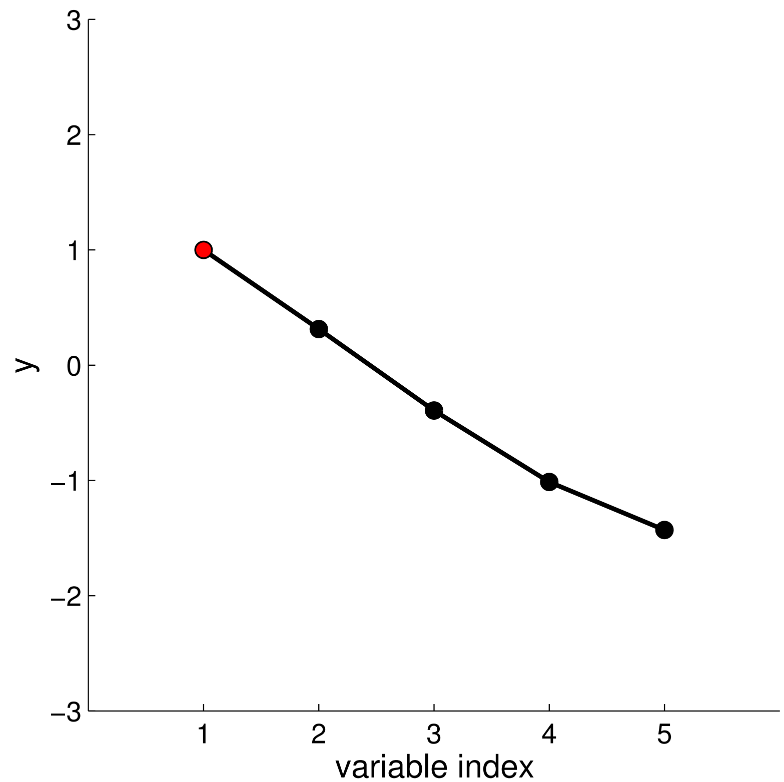




$$\Sigma = \begin{bmatrix} 1 & .9 & .8 & .6 & .4 \\ .9 & 1 & .9 & .8 & .6 \\ .8 & .9 & 1 & .9 & .8 \\ .6 & .8 & .9 & 1 & .9 \\ .4 & .6 & .8 & .9 & 1 \end{bmatrix}$$

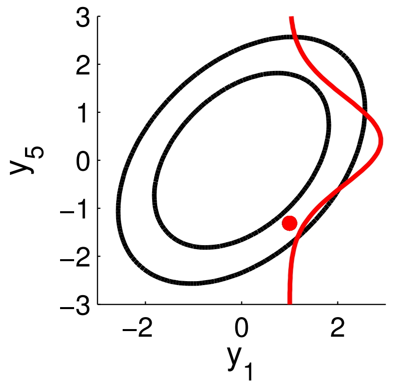
New visualisation

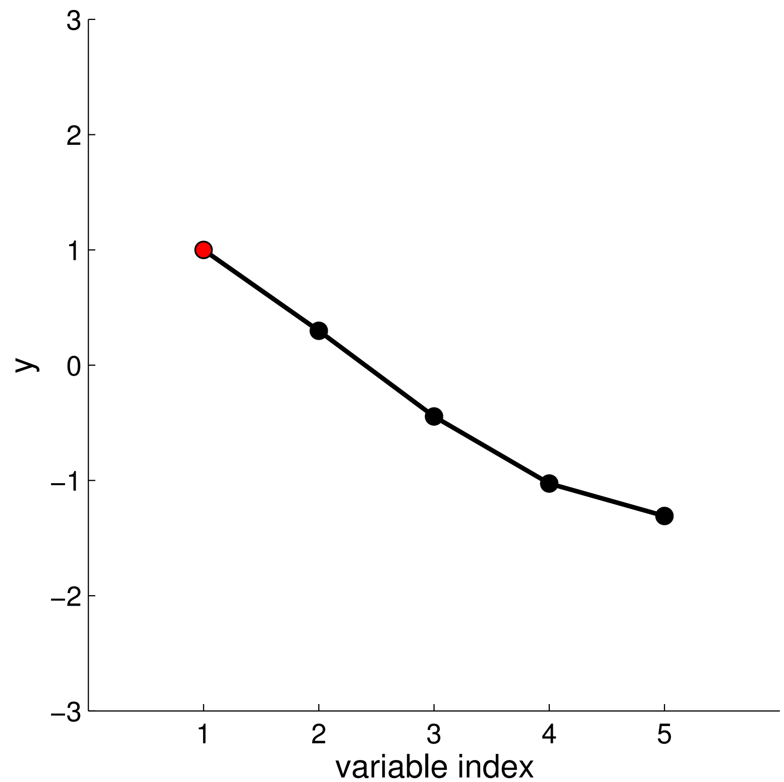




$$\Sigma = \begin{bmatrix} 1 & .9 & .8 & .6 & .4 \\ .9 & 1 & .9 & .8 & .6 \\ .8 & .9 & 1 & .9 & .8 \\ .6 & .8 & .9 & 1 & .9 \\ .4 & .6 & .8 & .9 & 1 \end{bmatrix}$$

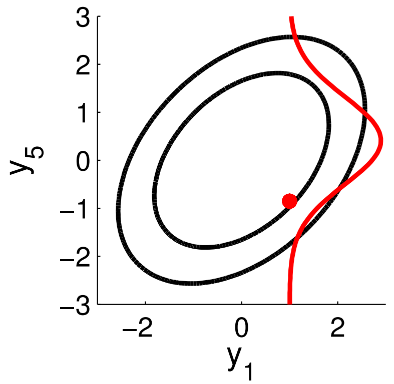
New visualisation

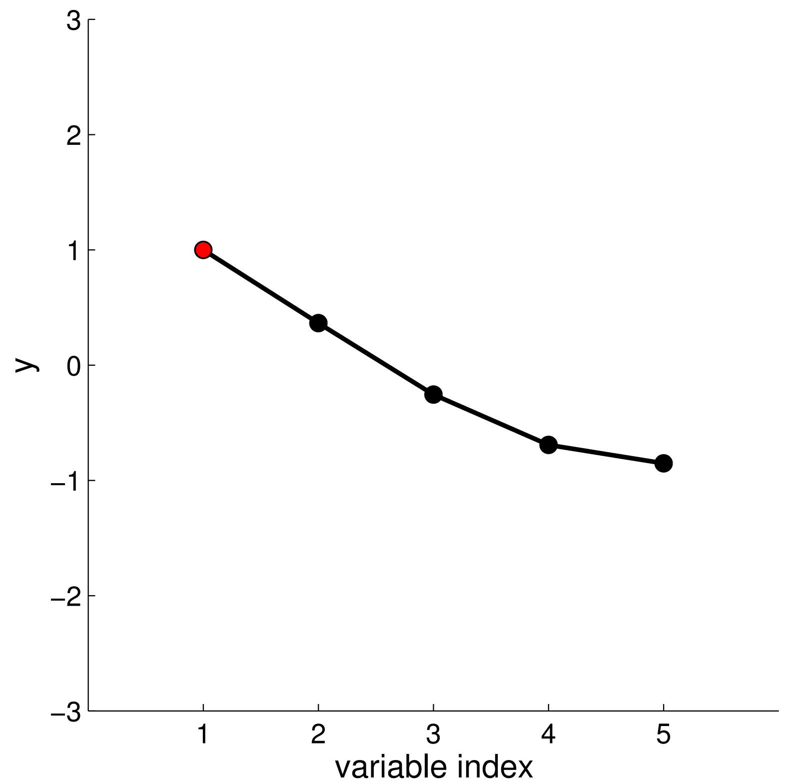




$$\Sigma = \begin{bmatrix} 1 & .9 & .8 & .6 & .4 \\ .9 & 1 & .9 & .8 & .6 \\ .8 & .9 & 1 & .9 & .8 \\ .6 & .8 & .9 & 1 & .9 \\ .4 & .6 & .8 & .9 & 1 \end{bmatrix}$$

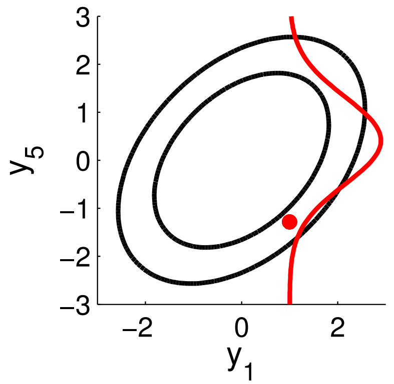
New visualisation

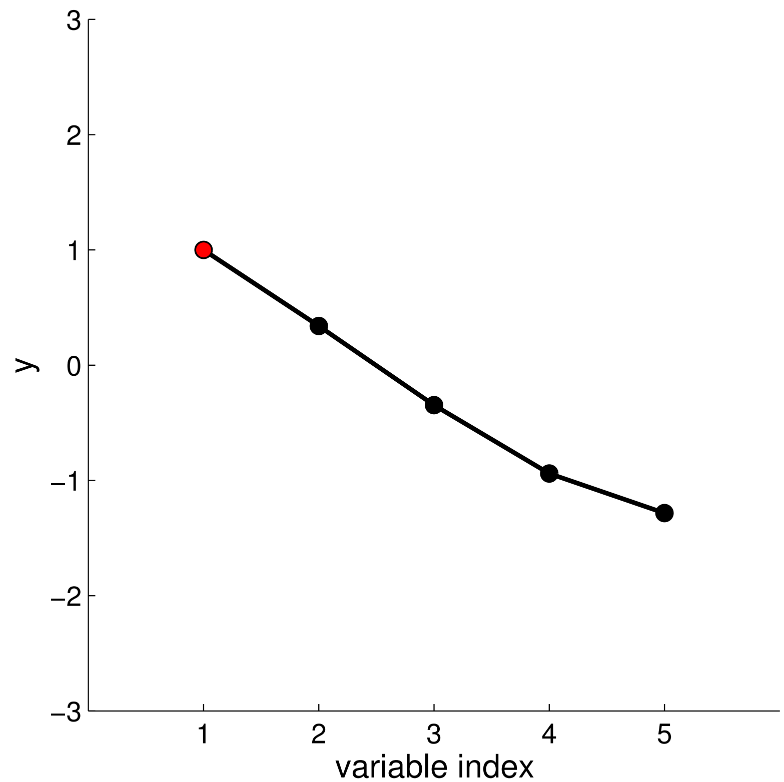




$$\Sigma = \begin{bmatrix} 1 & .9 & .8 & .6 & .4 \\ .9 & 1 & .9 & .8 & .6 \\ .8 & .9 & 1 & .9 & .8 \\ .6 & .8 & .9 & 1 & .9 \\ .4 & .6 & .8 & .9 & 1 \end{bmatrix}$$

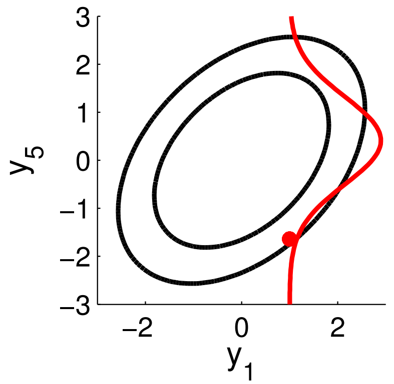
New visualisation

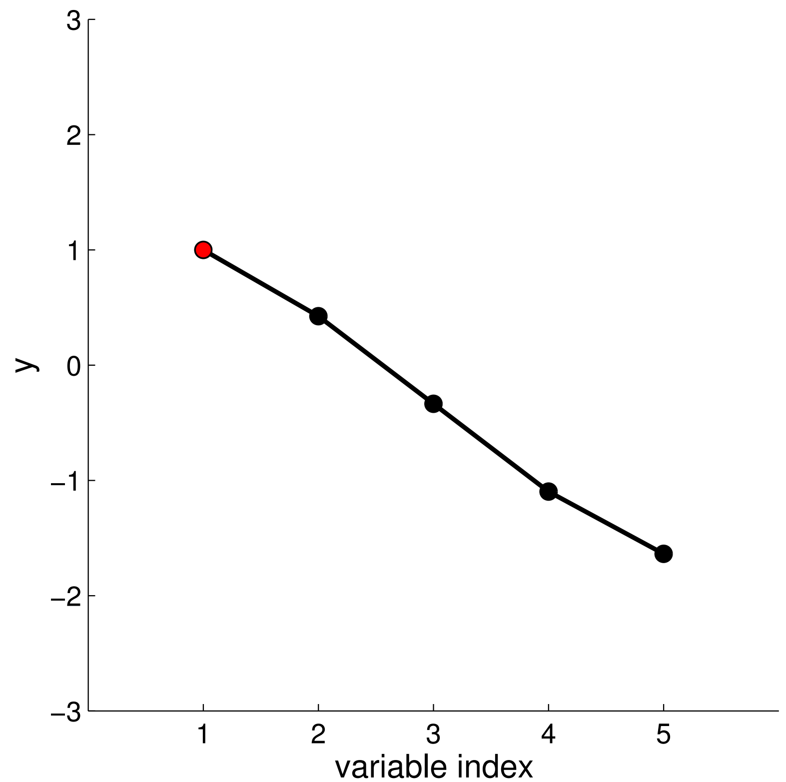




$$\Sigma = \begin{bmatrix} 1 & .9 & .8 & .6 & .4 \\ .9 & 1 & .9 & .8 & .6 \\ .8 & .9 & 1 & .9 & .8 \\ .6 & .8 & .9 & 1 & .9 \\ .4 & .6 & .8 & .9 & 1 \end{bmatrix}$$

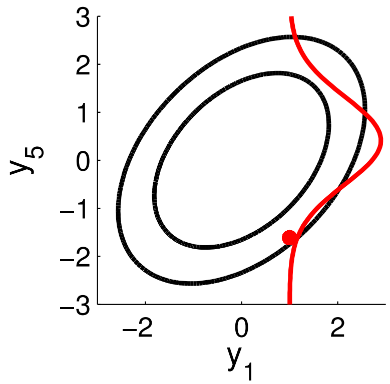
New visualisation

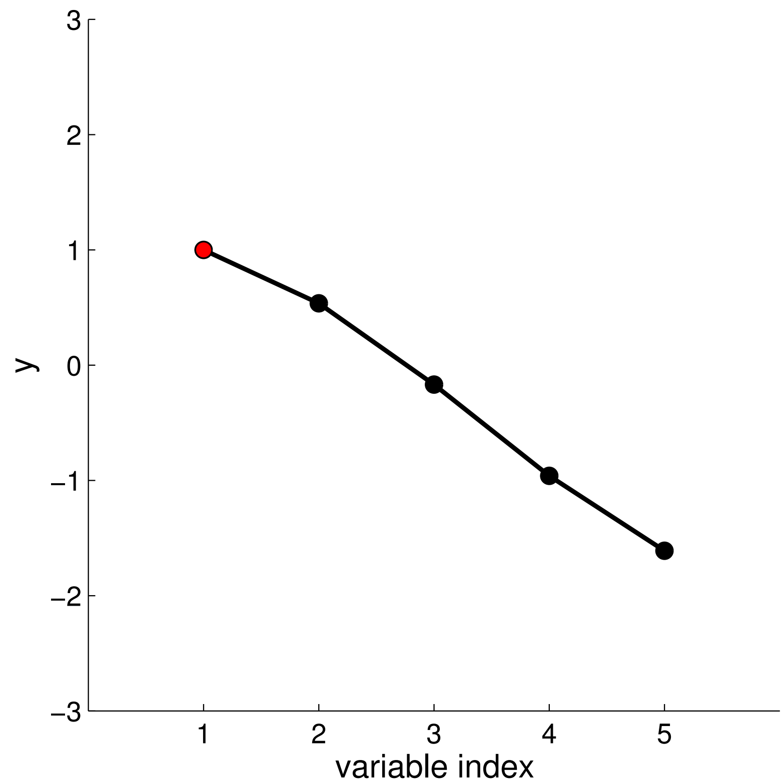




$$\Sigma = \begin{bmatrix} 1 & .9 & .8 & .6 & .4 \\ .9 & 1 & .9 & .8 & .6 \\ .8 & .9 & 1 & .9 & .8 \\ .6 & .8 & .9 & 1 & .9 \\ .4 & .6 & .8 & .9 & 1 \end{bmatrix}$$

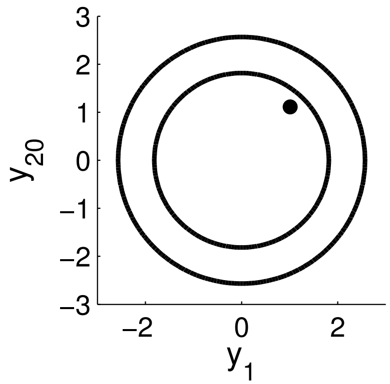
New visualisation

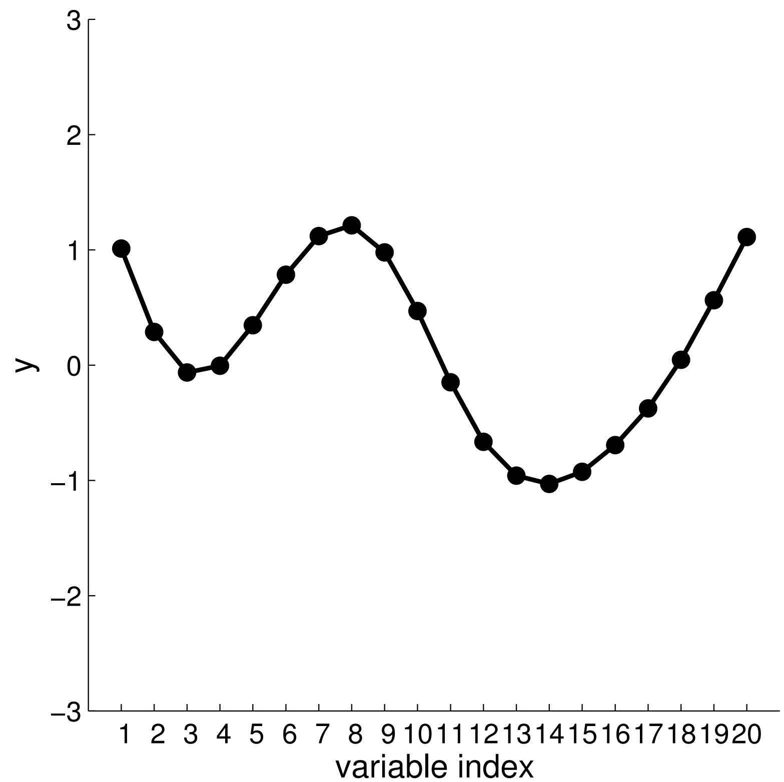




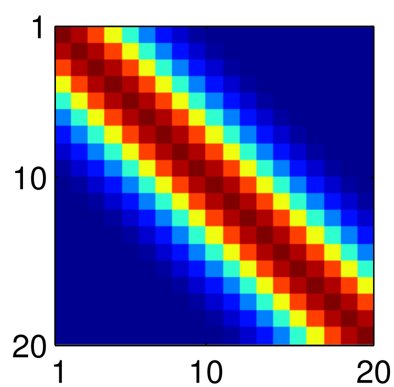
$$\Sigma = \begin{bmatrix} 1 & .9 & .8 & .6 & .4 \\ .9 & 1 & .9 & .8 & .6 \\ .8 & .9 & 1 & .9 & .8 \\ .6 & .8 & .9 & 1 & .9 \\ .4 & .6 & .8 & .9 & 1 \end{bmatrix}$$

New visualisation

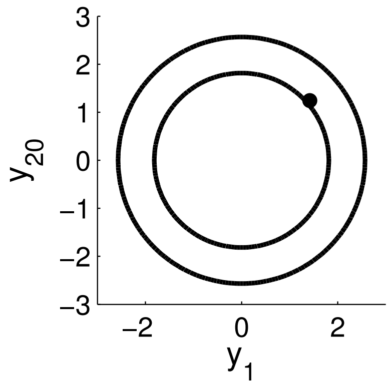


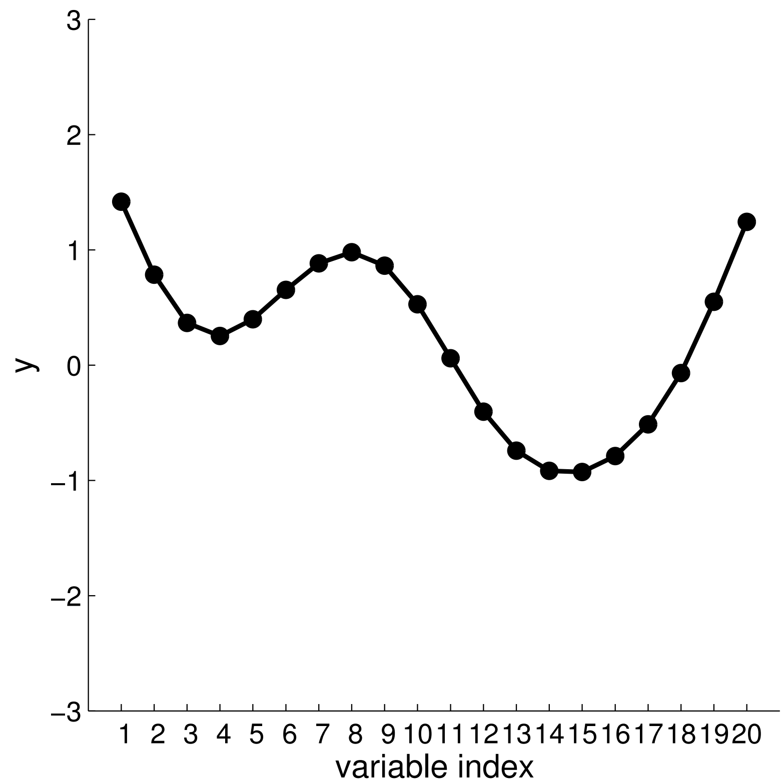


$\Sigma =$

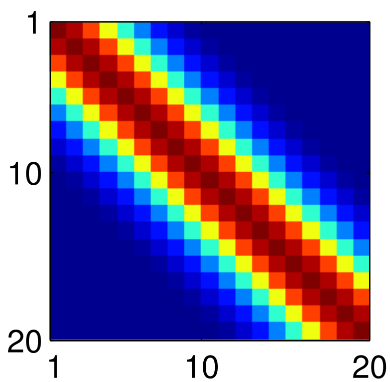


New visualisation

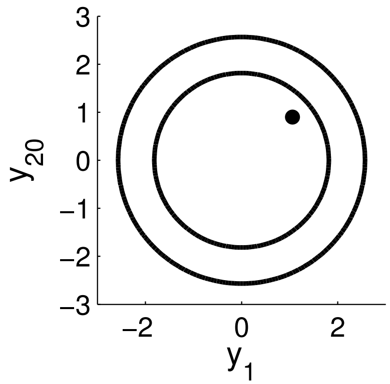


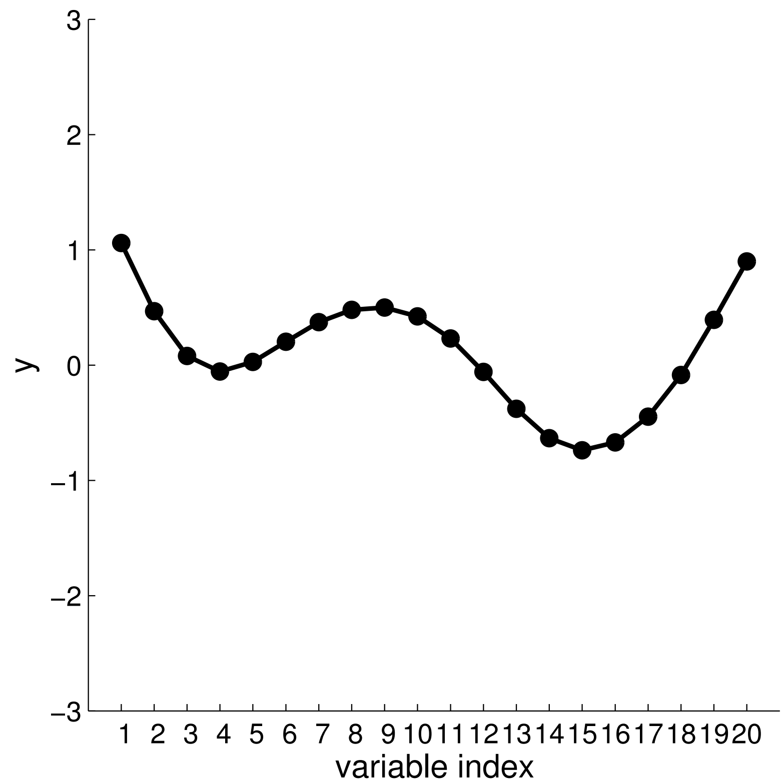


$\Sigma =$

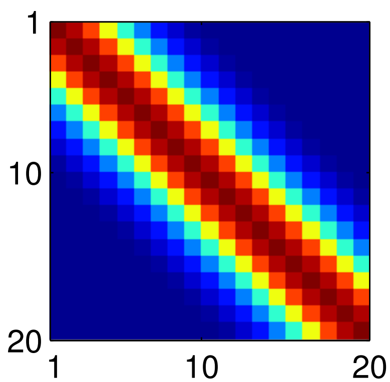


New visualisation

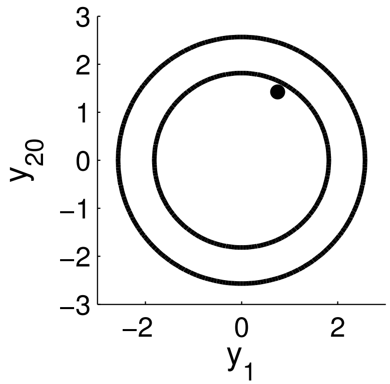


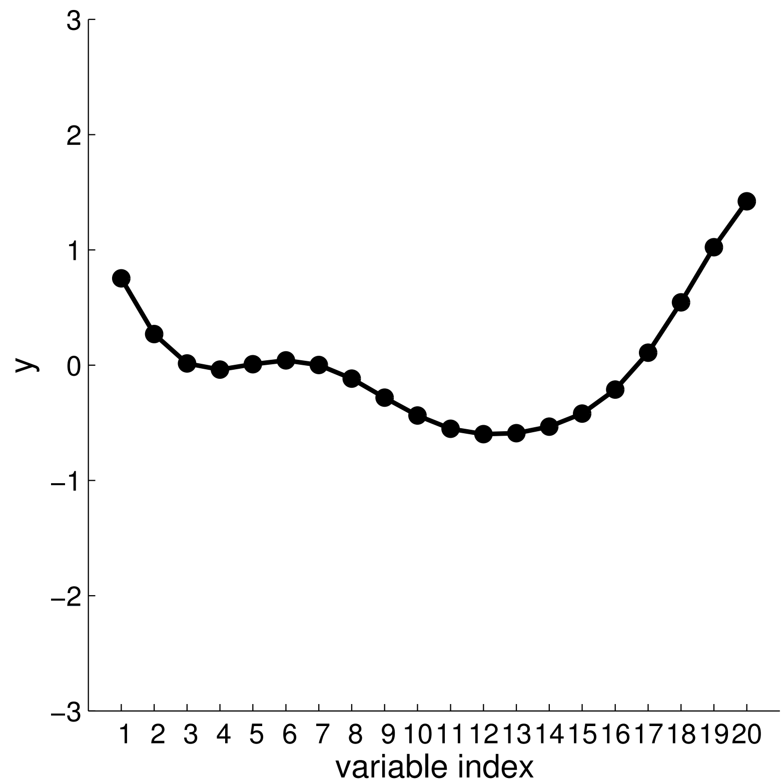


$\Sigma =$

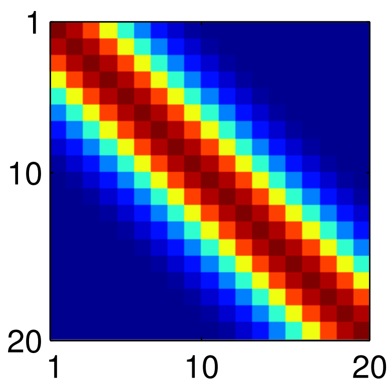


New visualisation

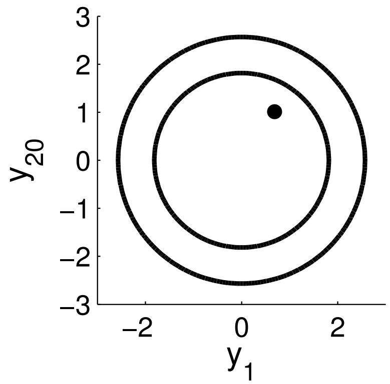


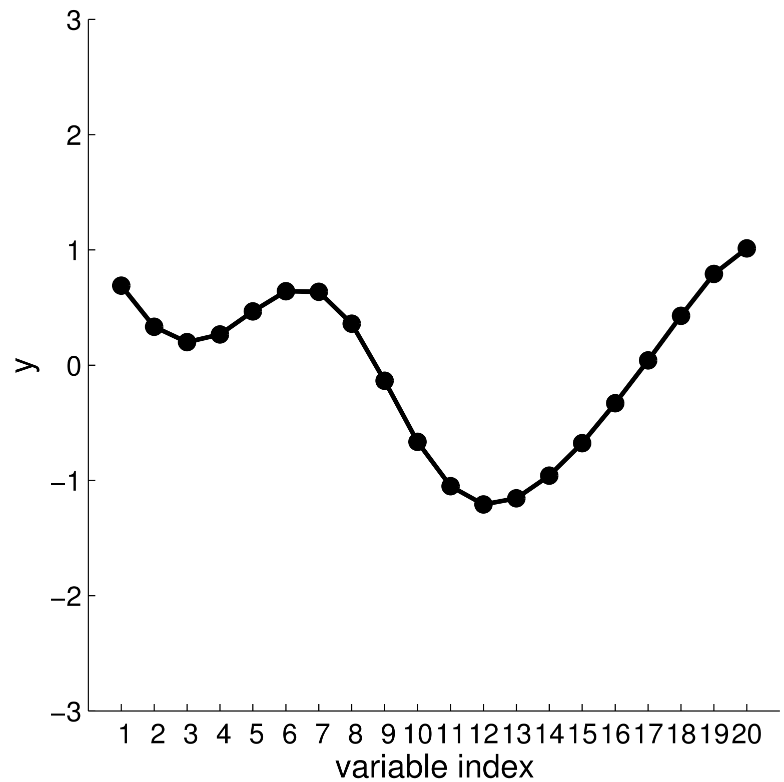


$\Sigma =$

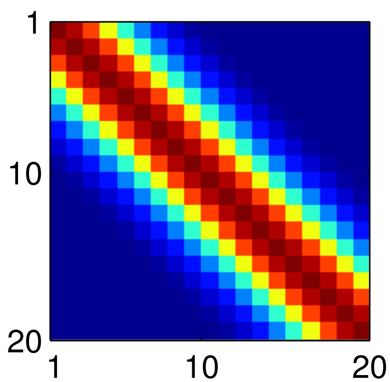


New visualisation

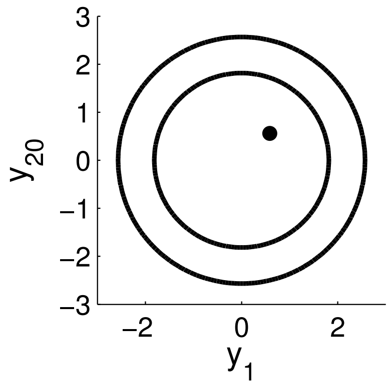


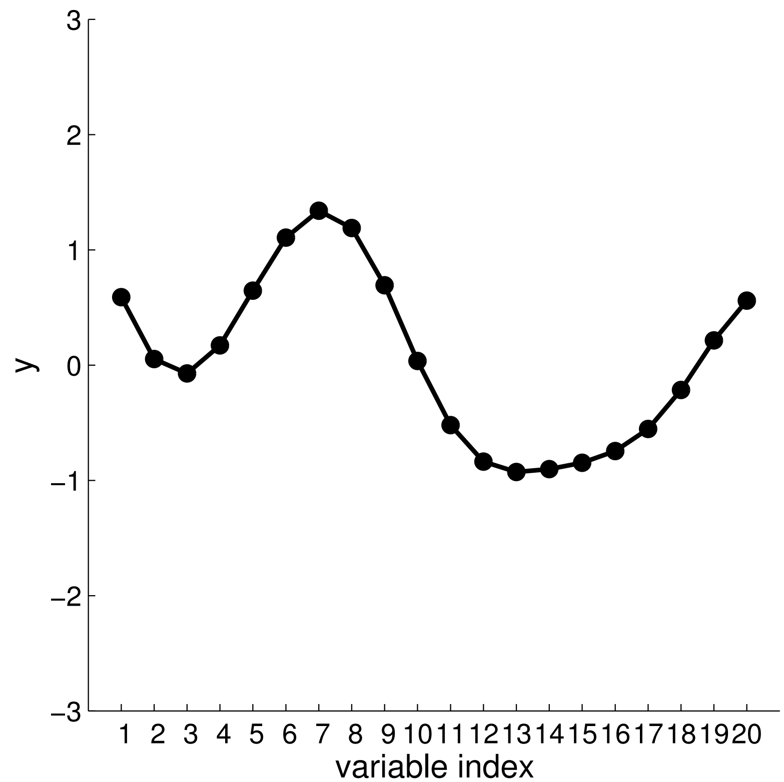


$$\Sigma =$$

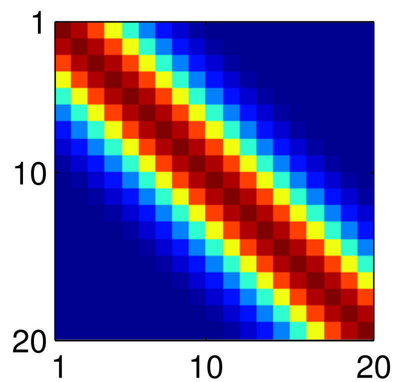


New visualisation

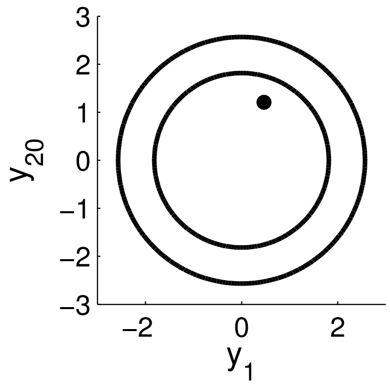


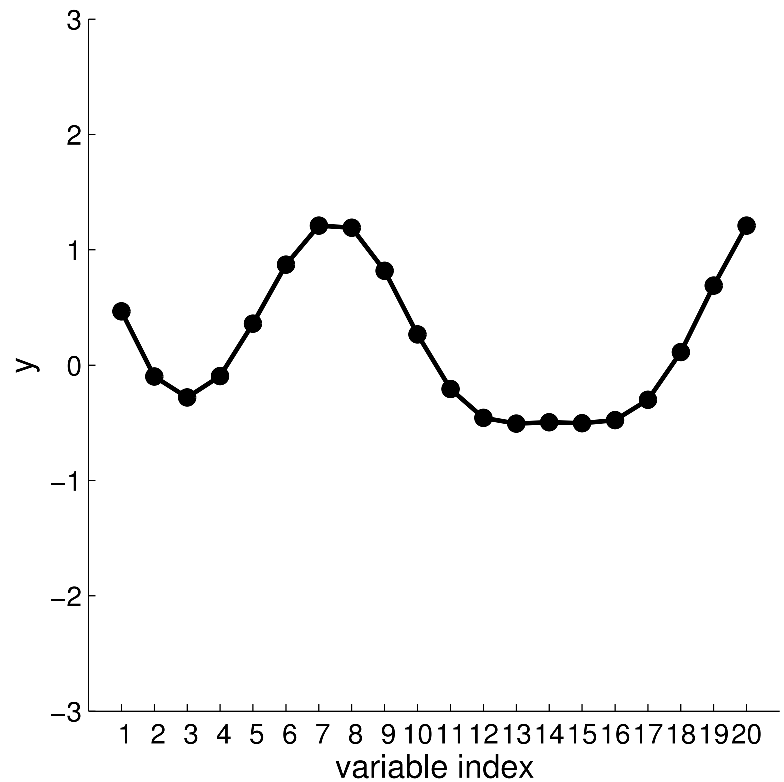


$\Sigma =$

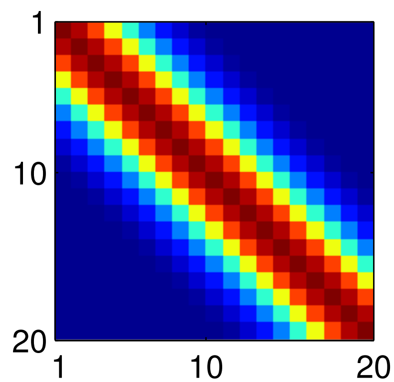


New visualisation

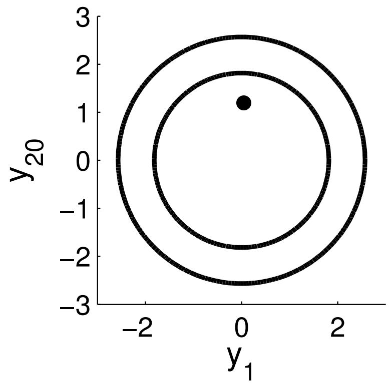


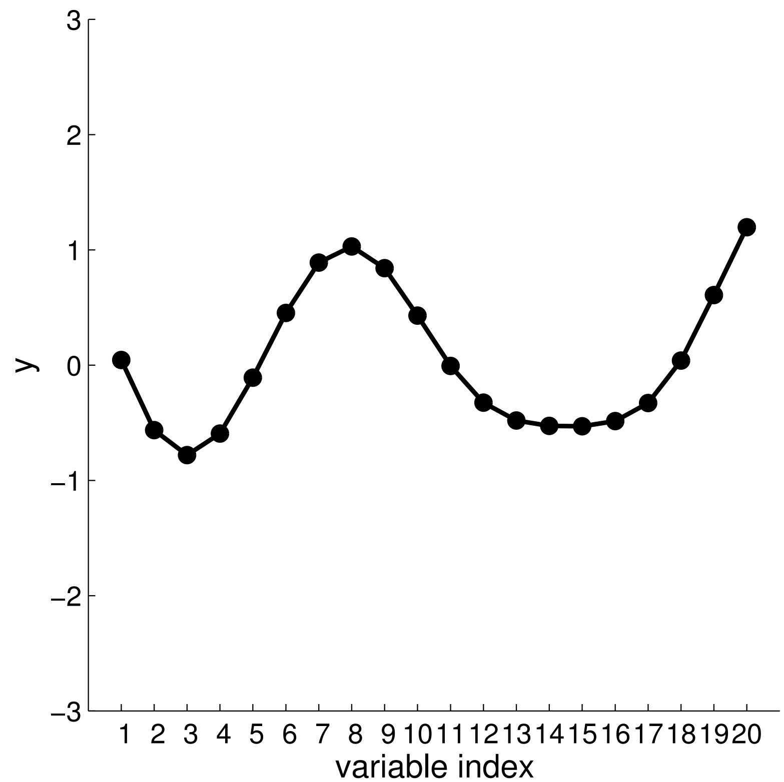


$\Sigma =$

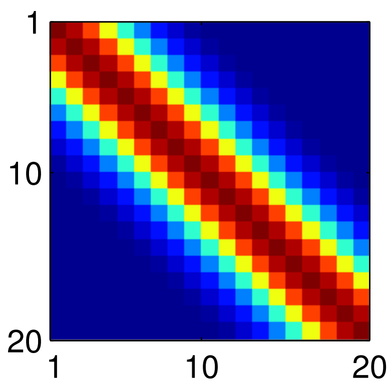


New visualisation

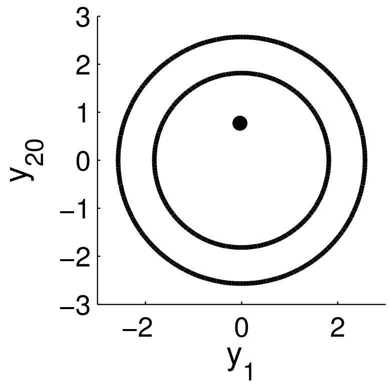


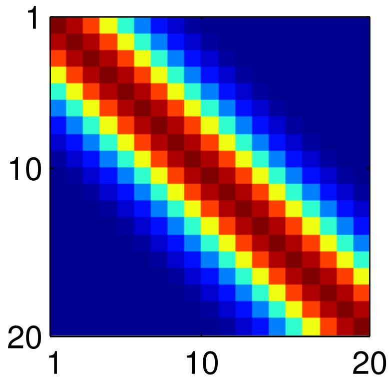


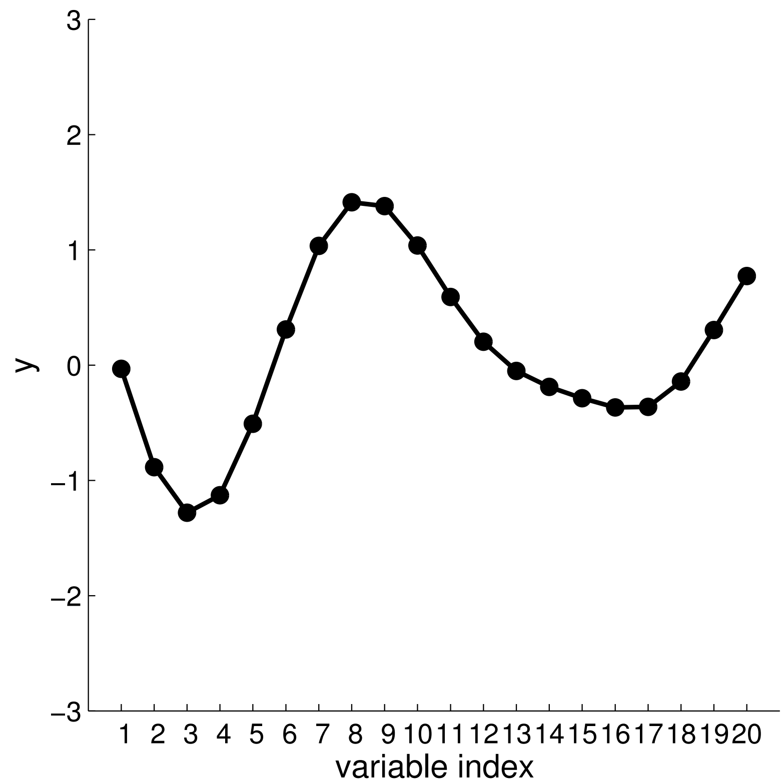
$$\Sigma =$$



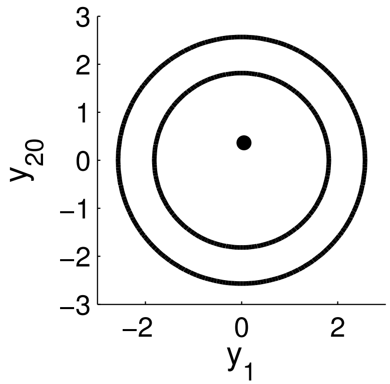
New visualisation

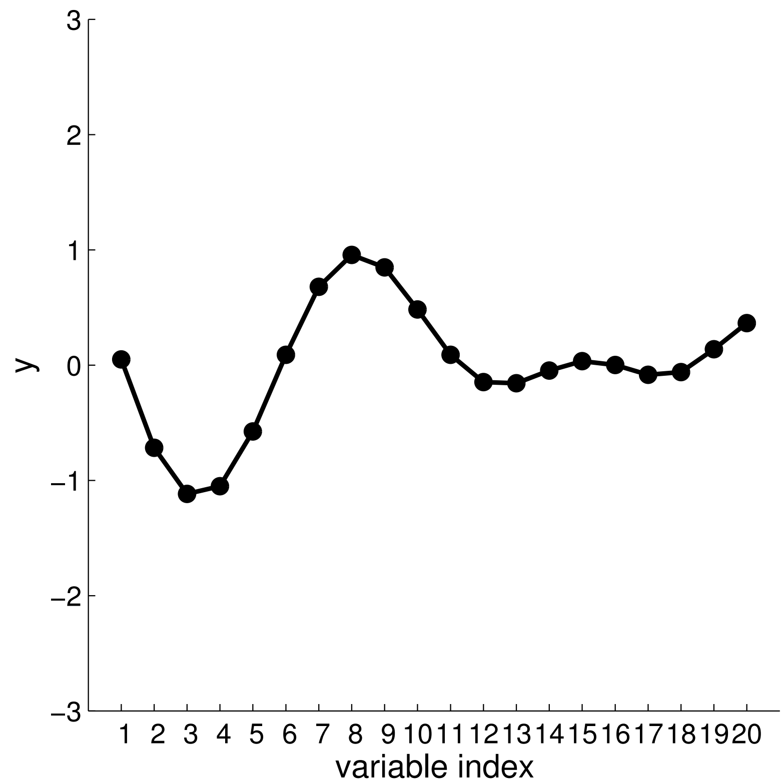


$\Sigma =$ 

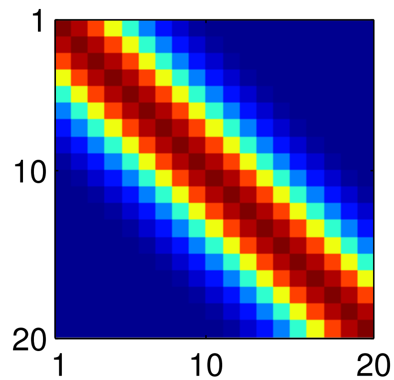


New visualisation

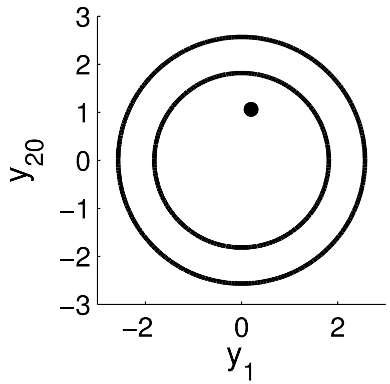


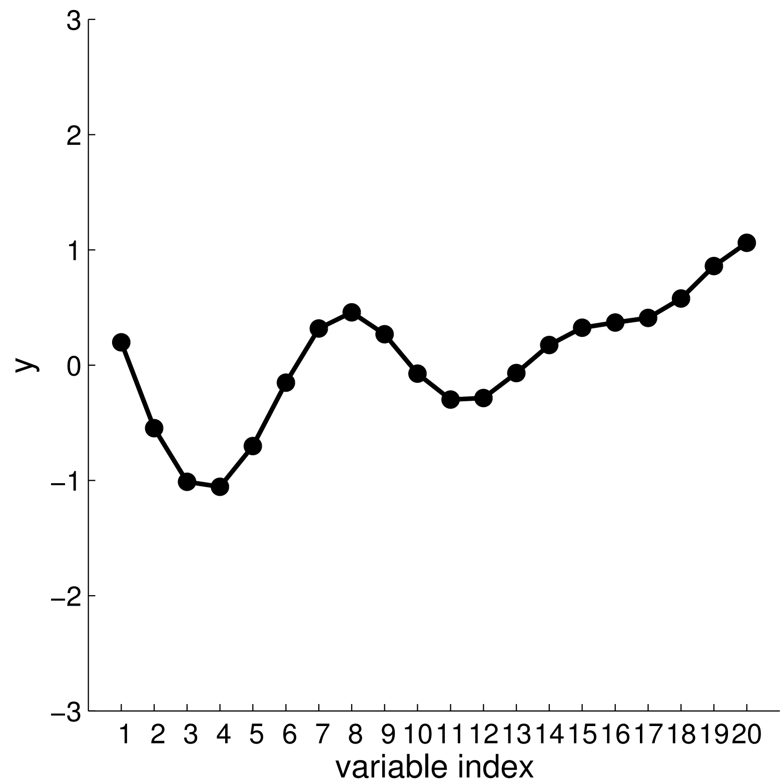


$\Sigma =$

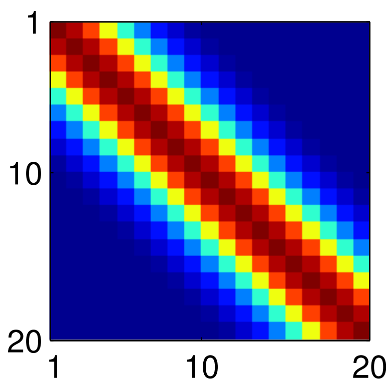


New visualisation

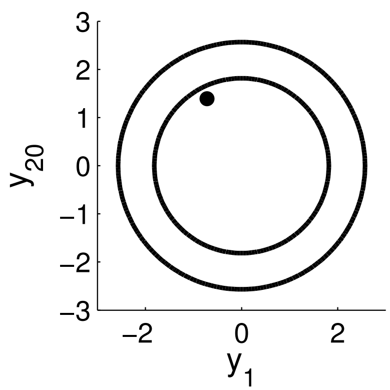


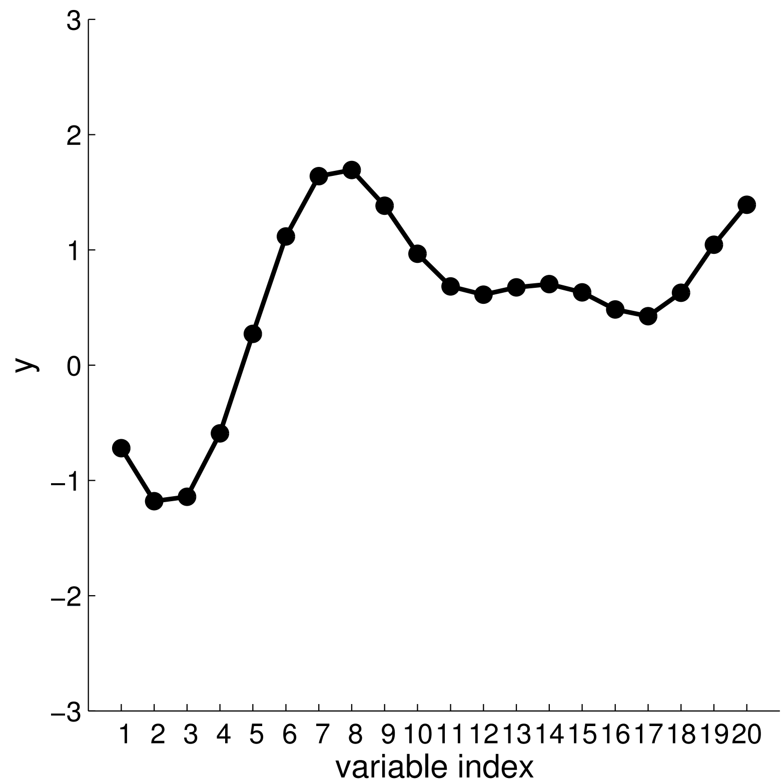


$\Sigma =$

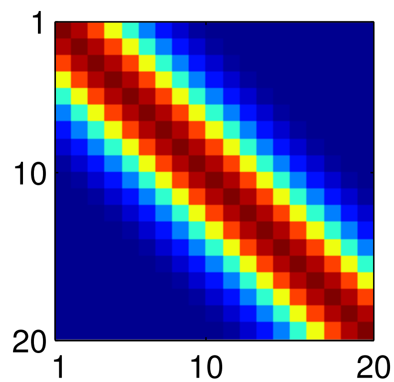


New visualisation

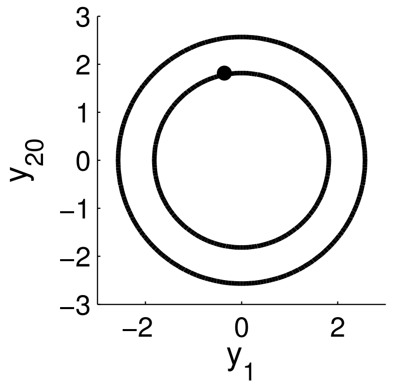


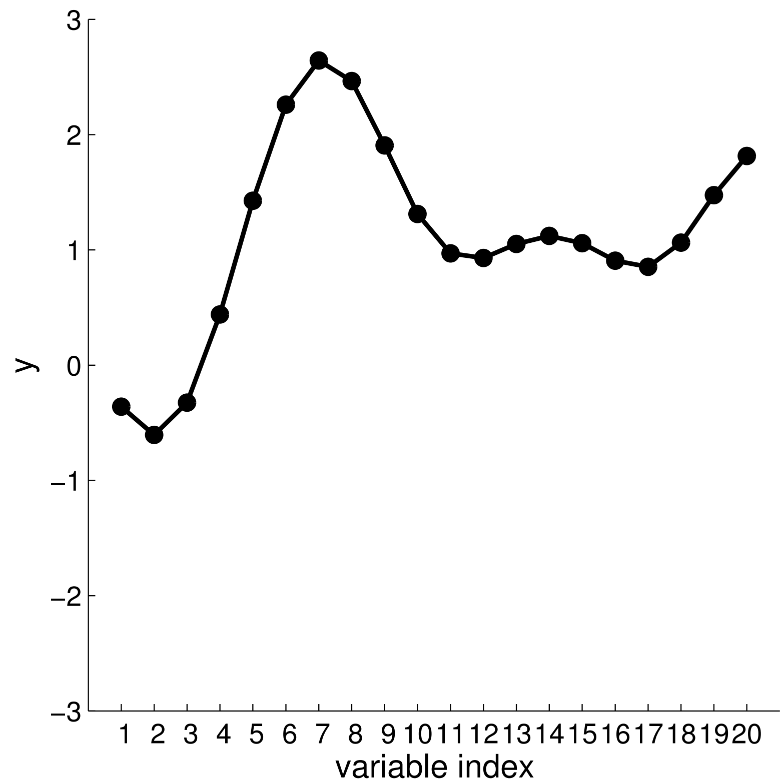


$\Sigma =$

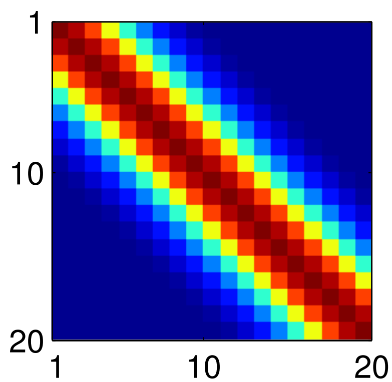


New visualisation

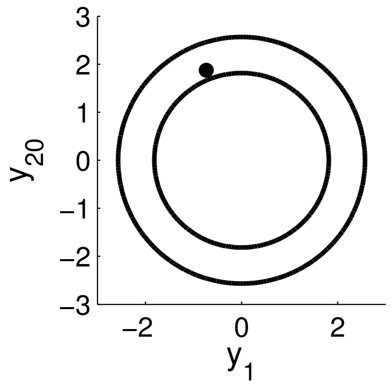




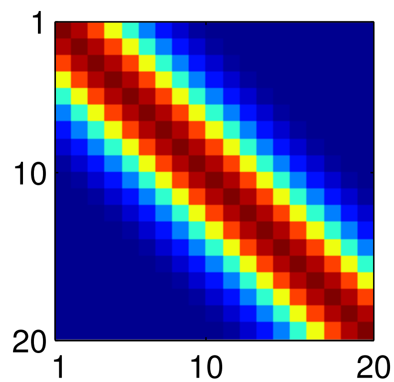
$$\Sigma =$$



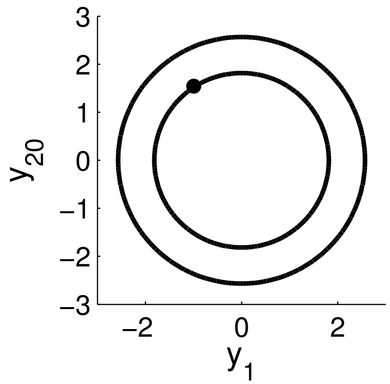
New visualisation

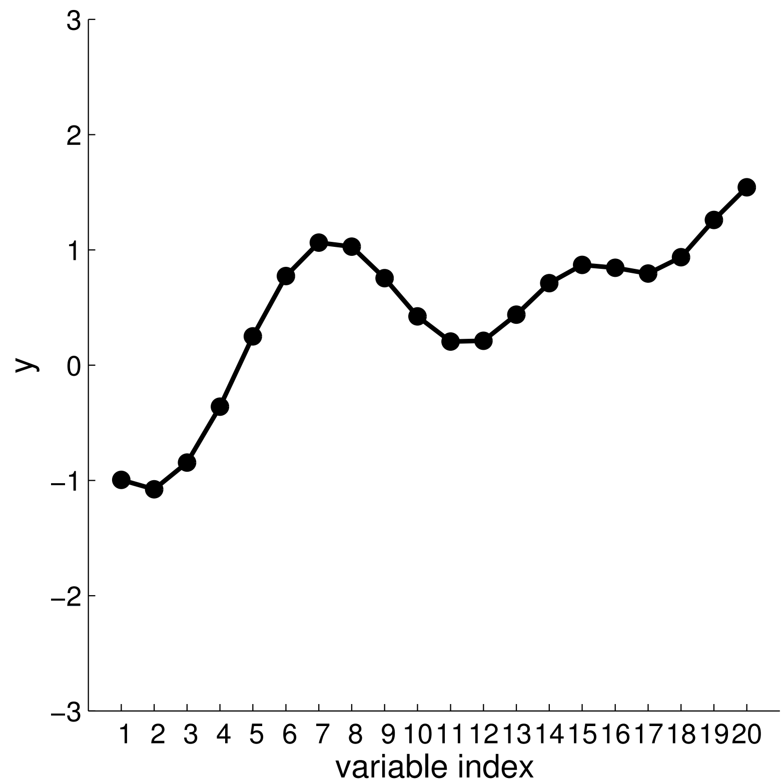


$\Sigma =$

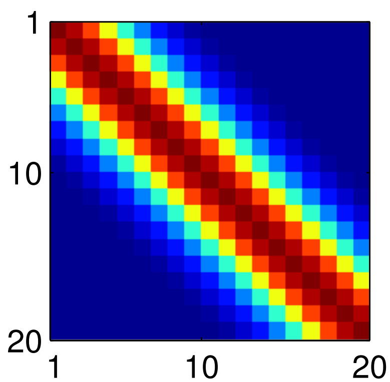


New visualisation

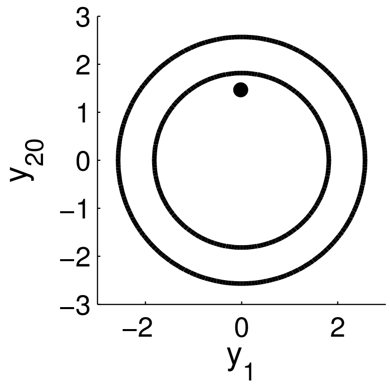


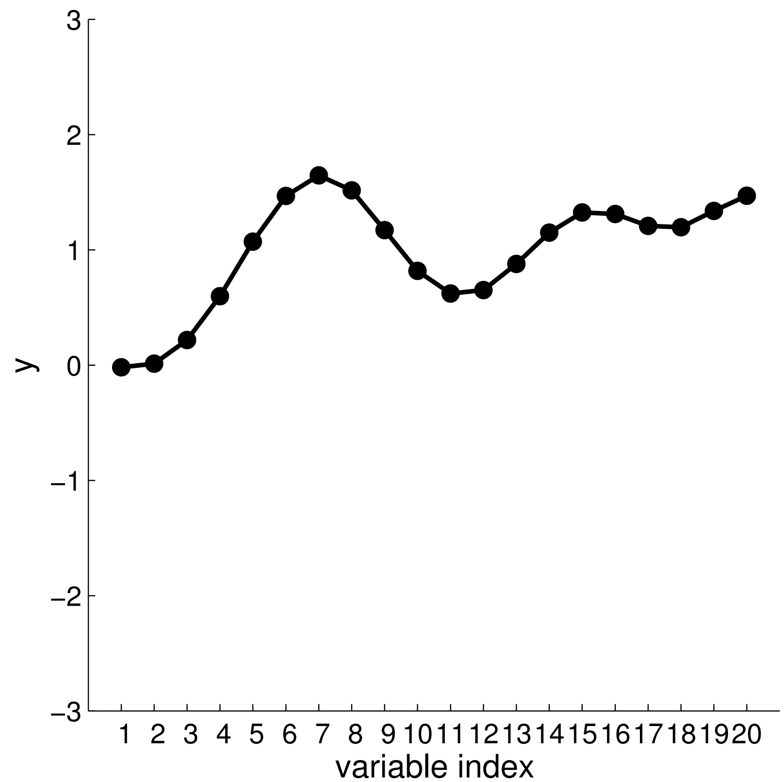


$\Sigma =$

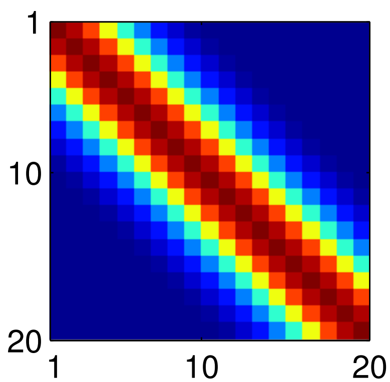


New visualisation

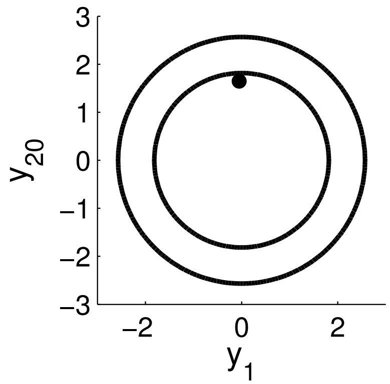


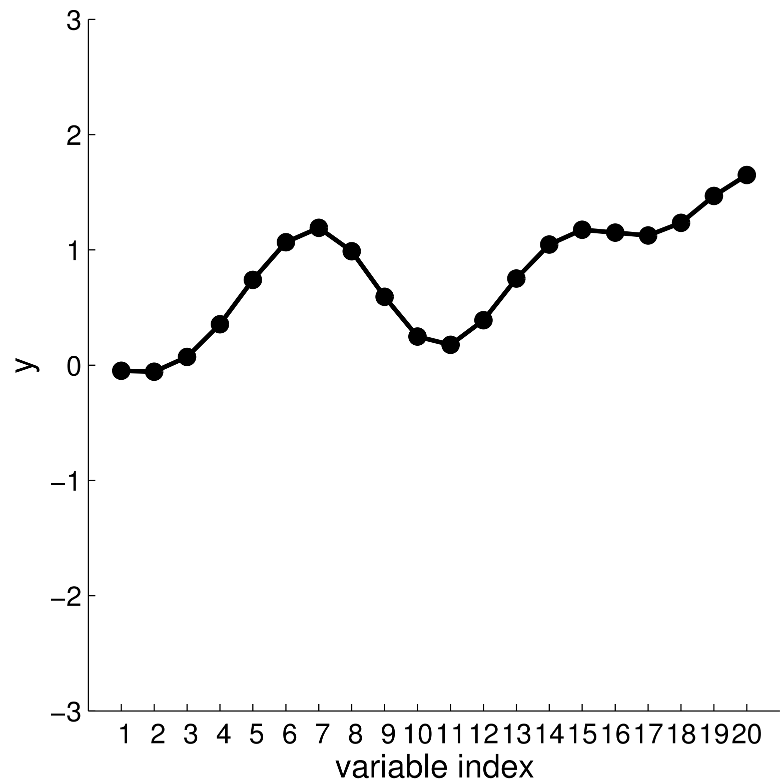


$\Sigma =$

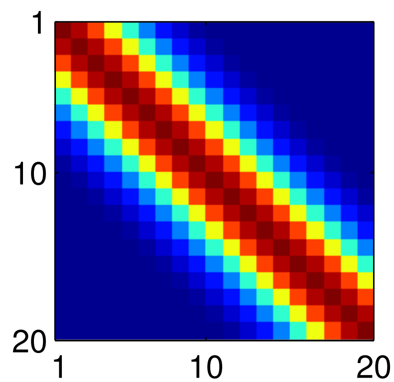


New visualisation

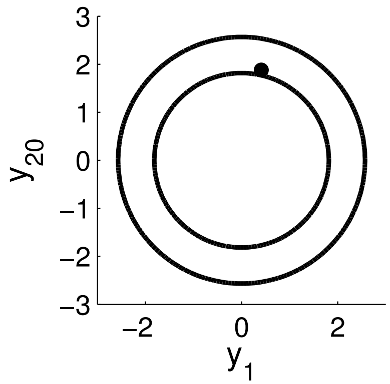


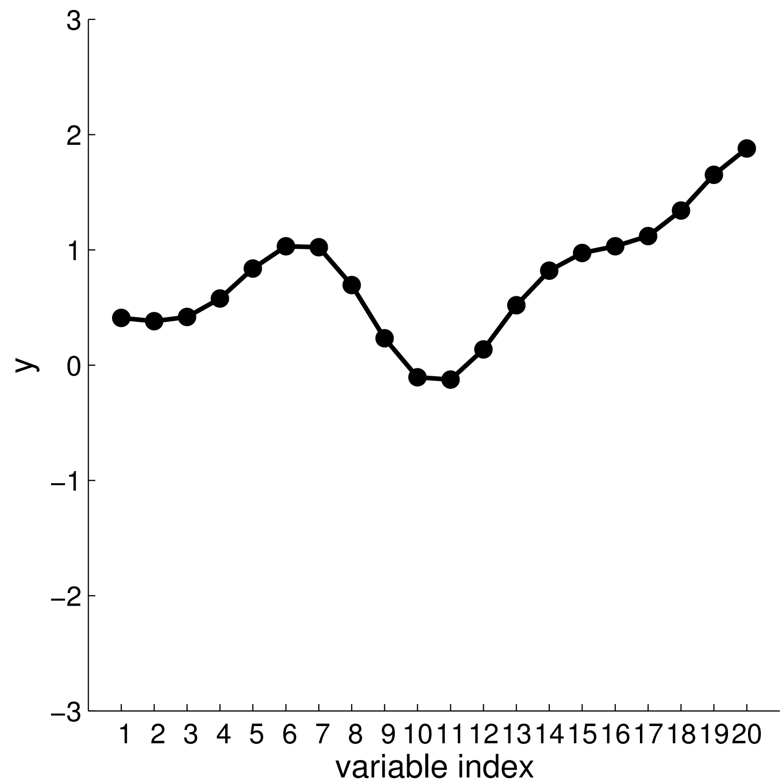


$$\Sigma =$$

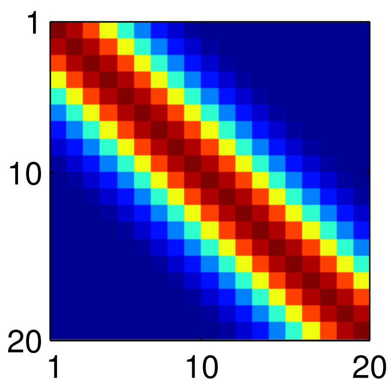


New visualisation

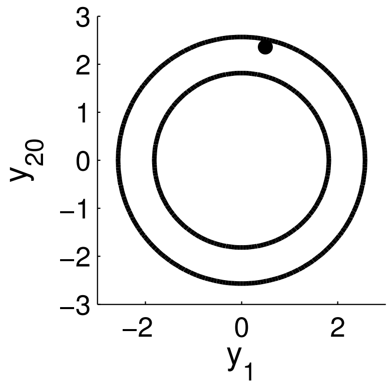


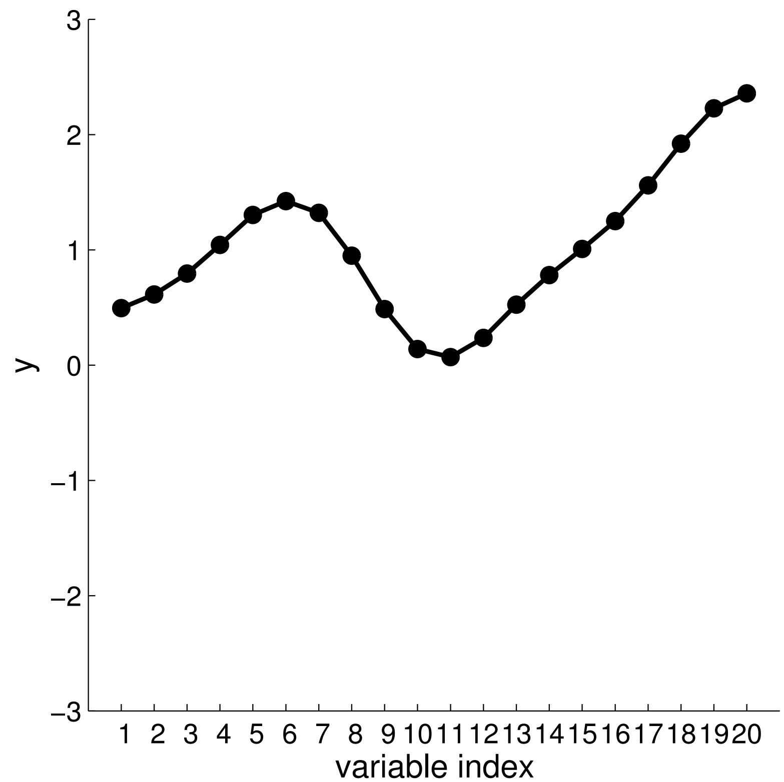


$\Sigma =$

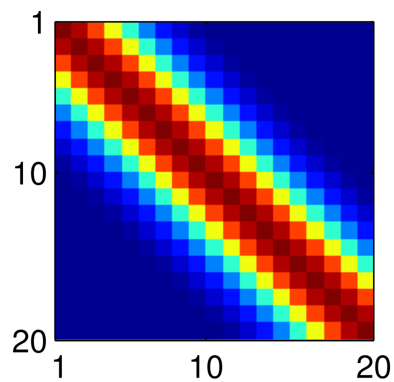


New visualisation

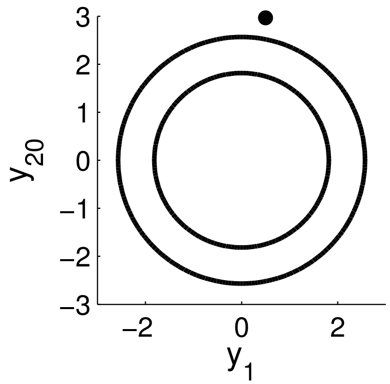


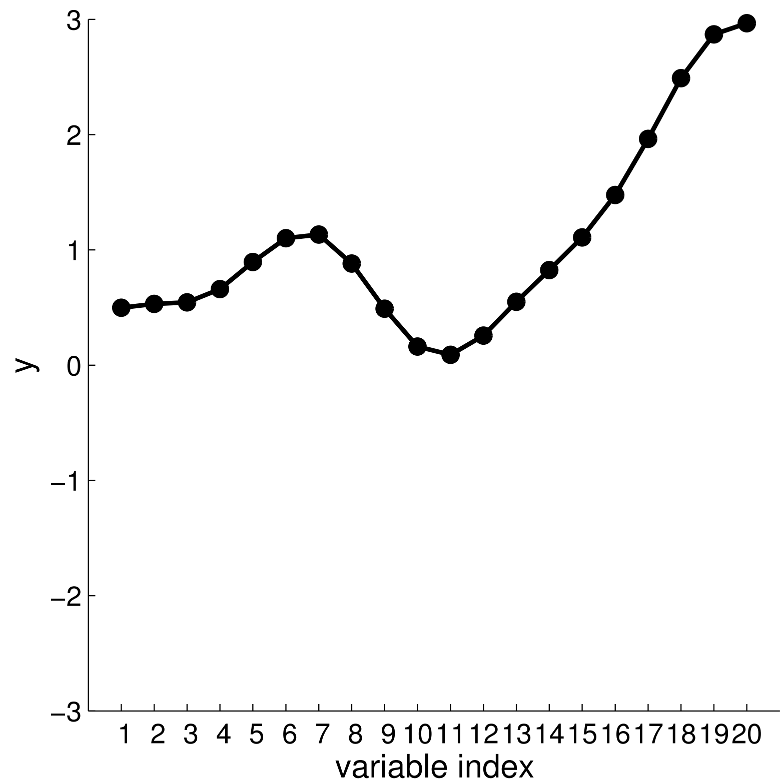


$\Sigma =$

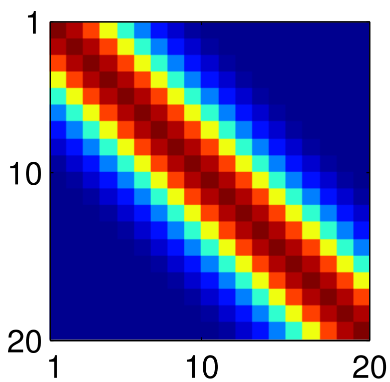


New visualisation

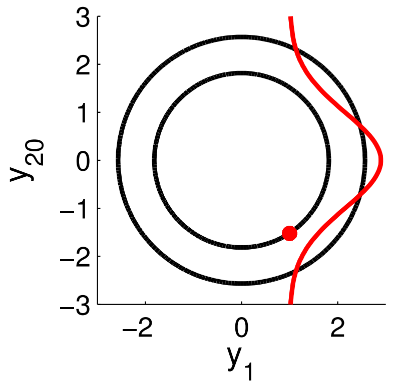


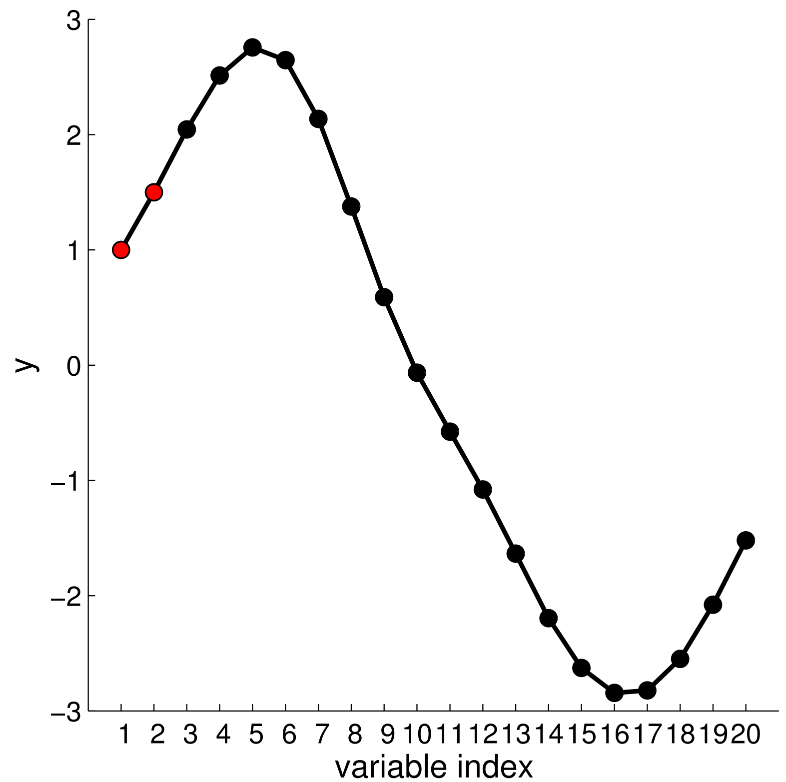


$\Sigma =$

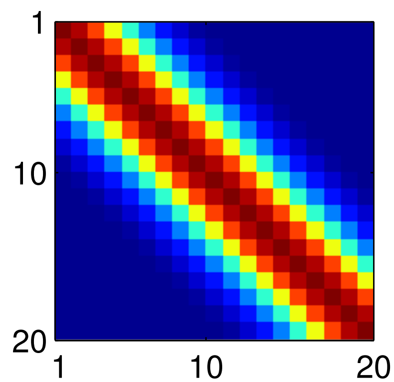


New visualisation

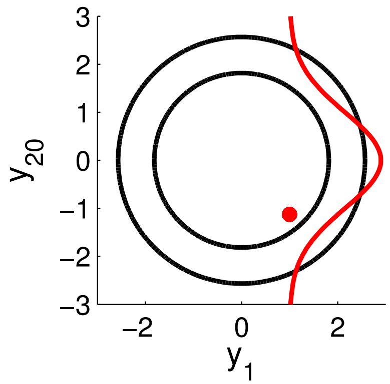


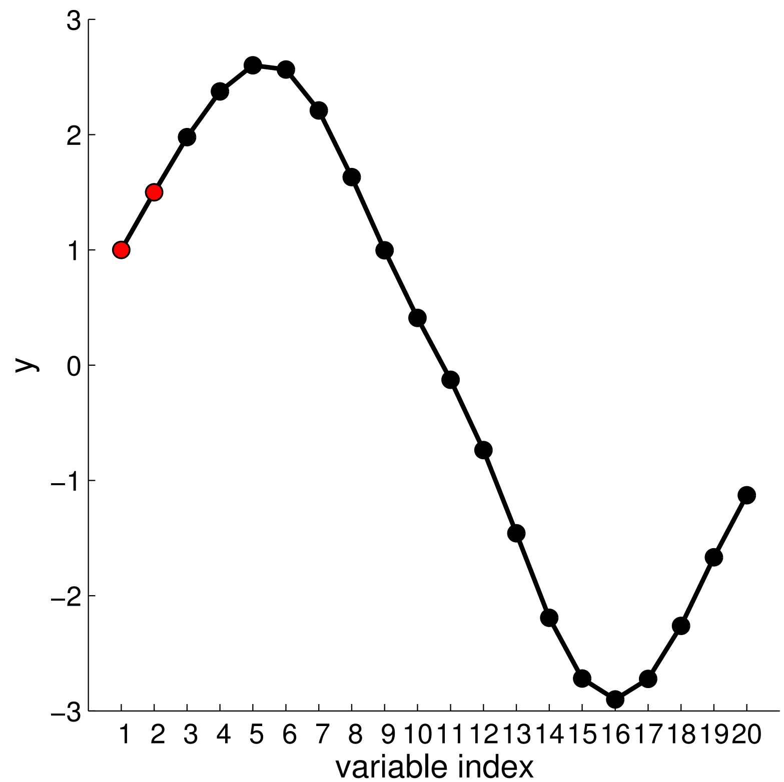


$\Sigma =$

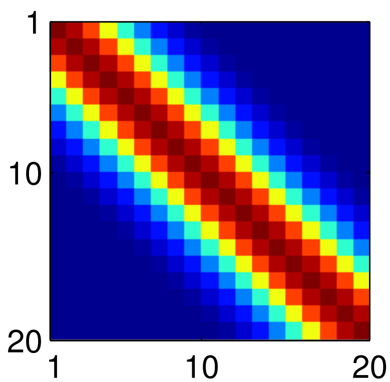


New visualisation

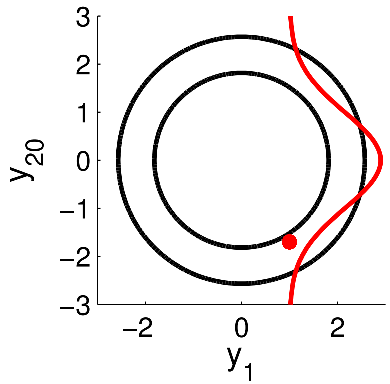


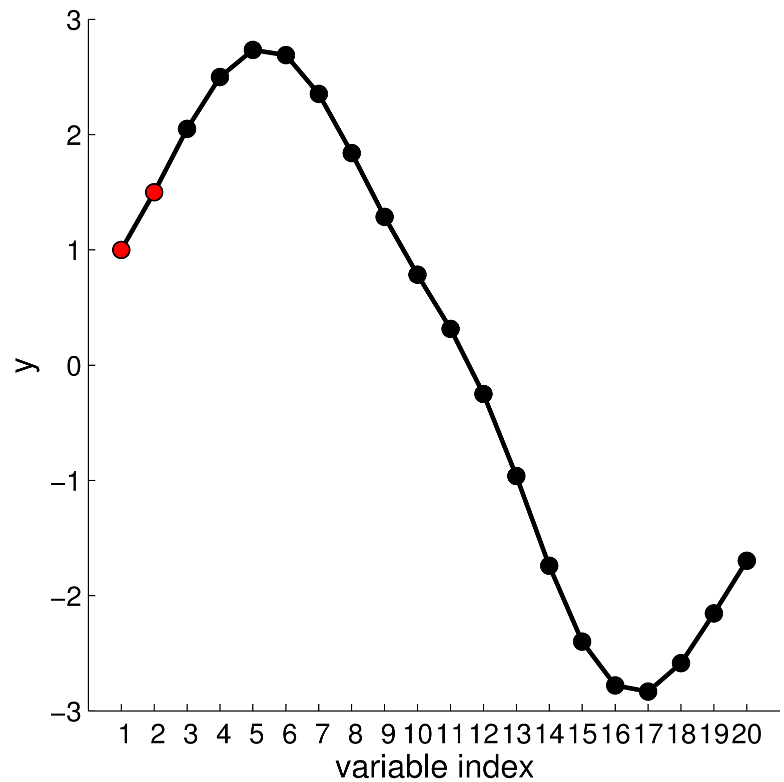


$$\Sigma =$$

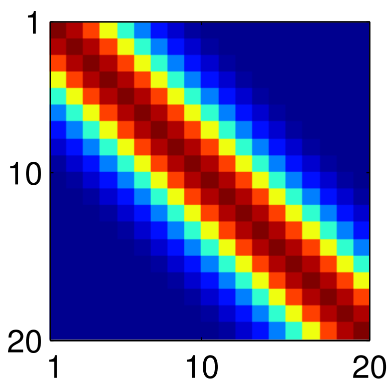


New visualisation

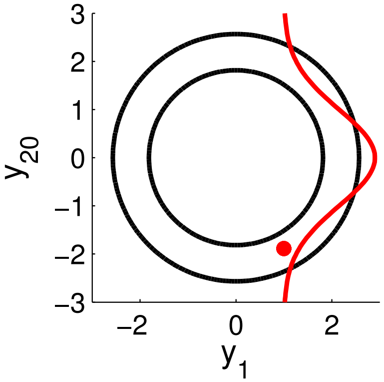




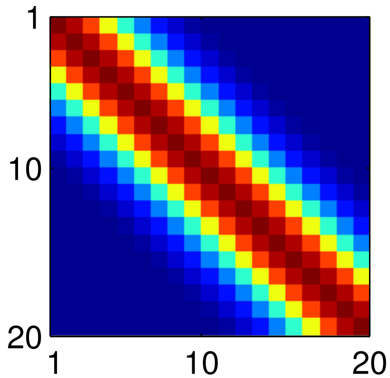
$$\Sigma =$$

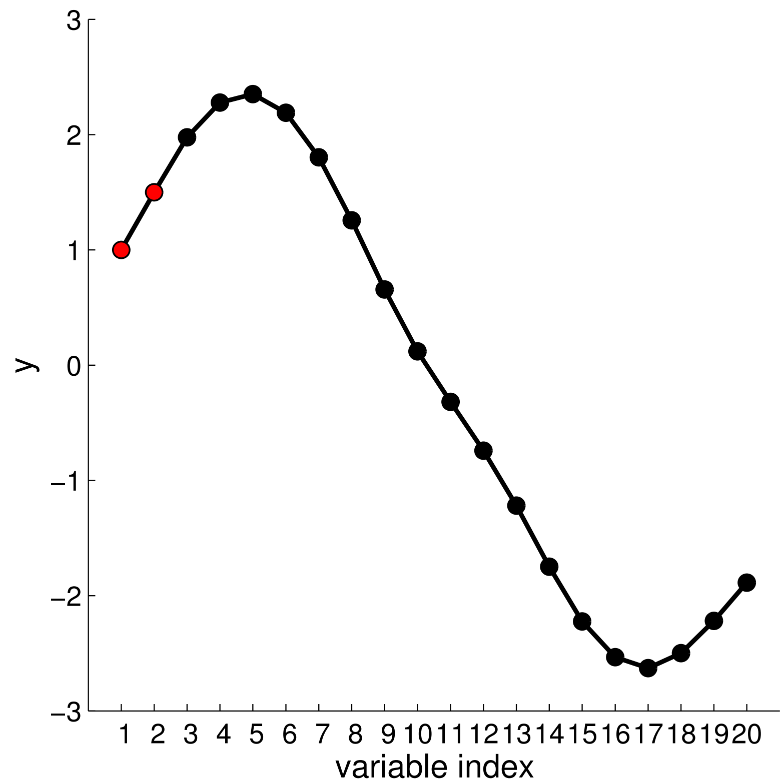


New visualisation

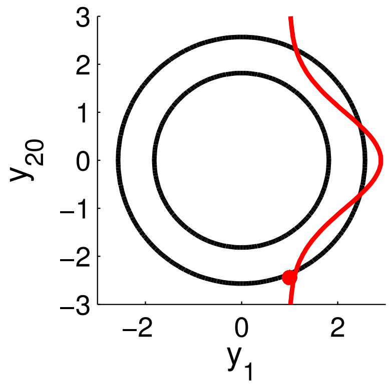


$\Sigma =$

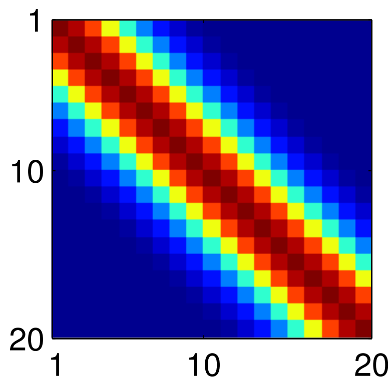


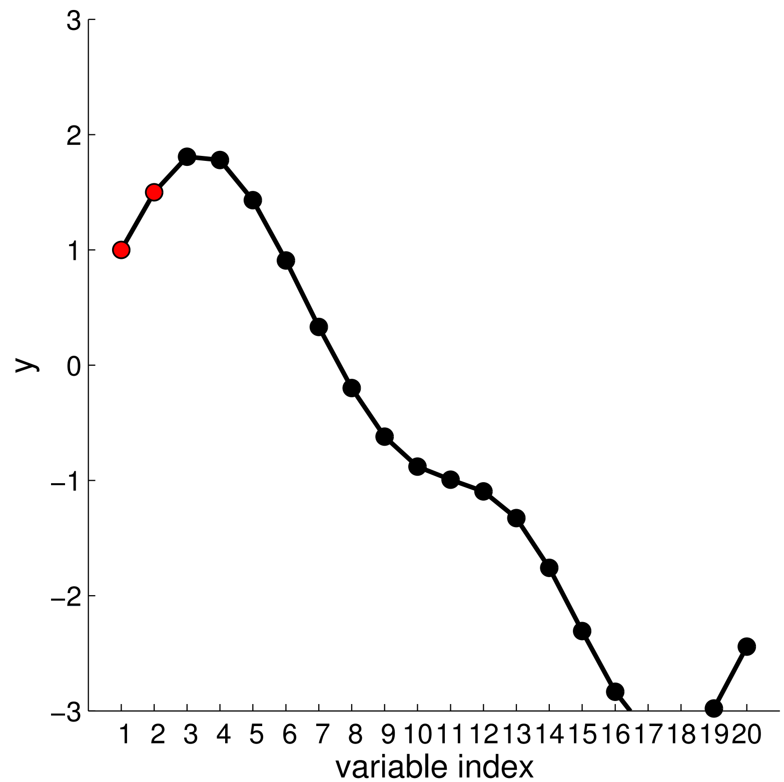


New visualisation

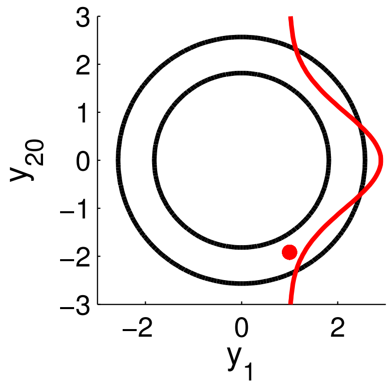


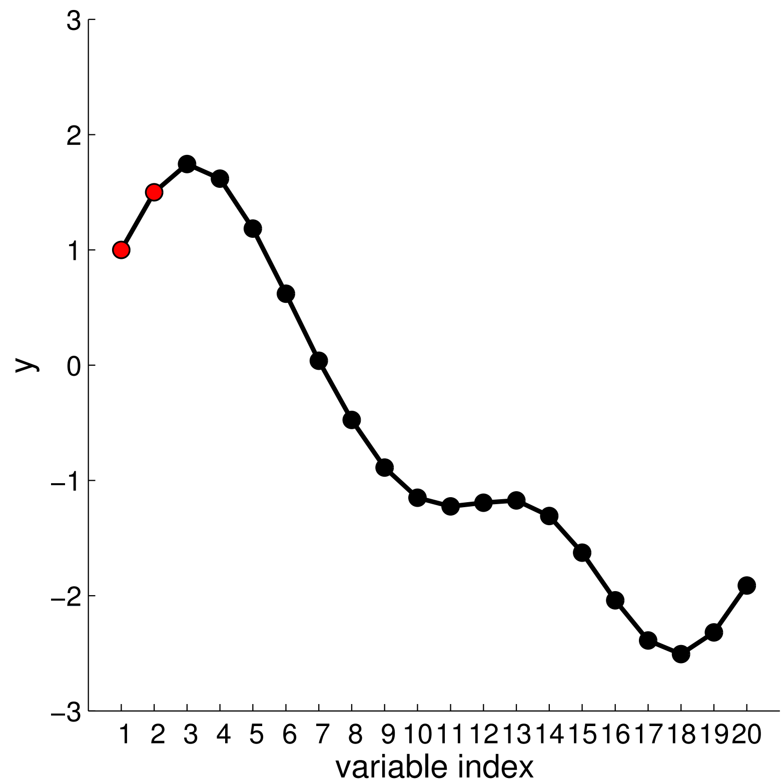
$\Sigma =$



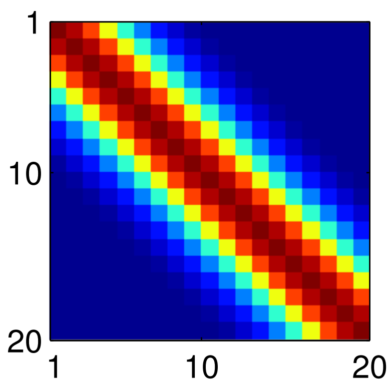


New visualisation

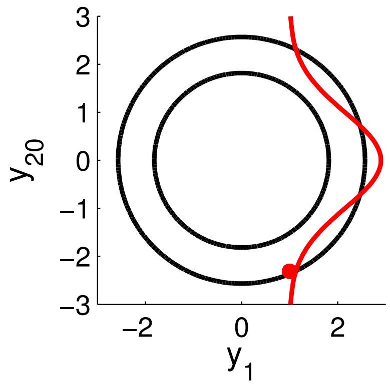




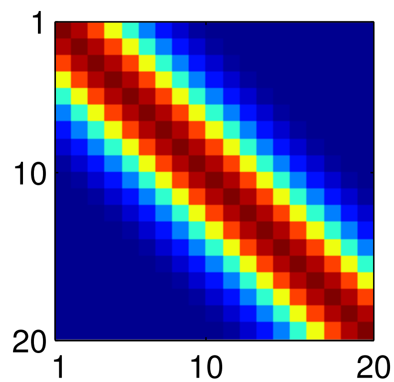
$\Sigma =$

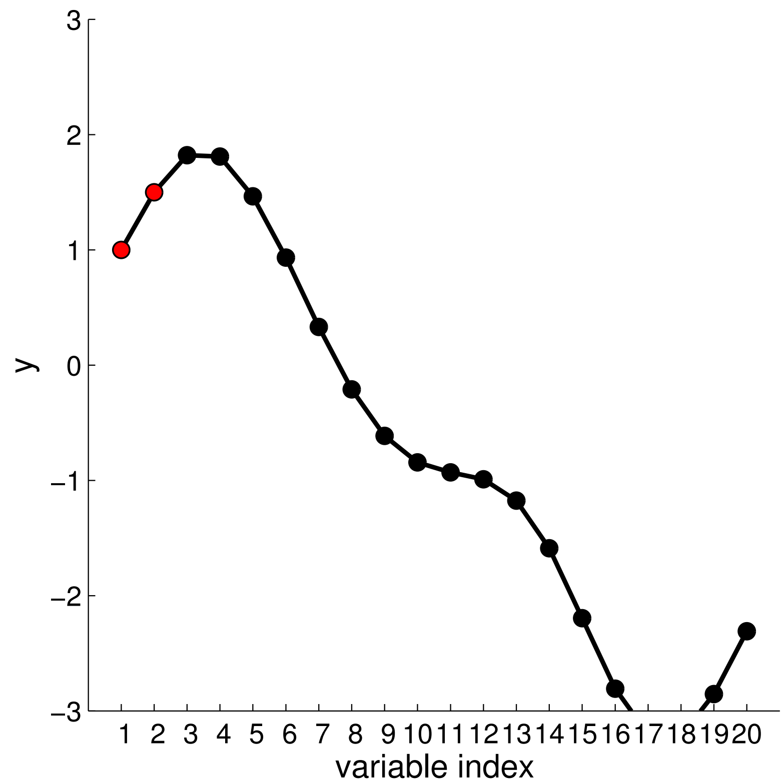


New visualisation

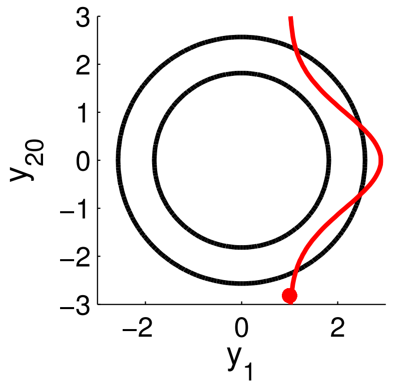


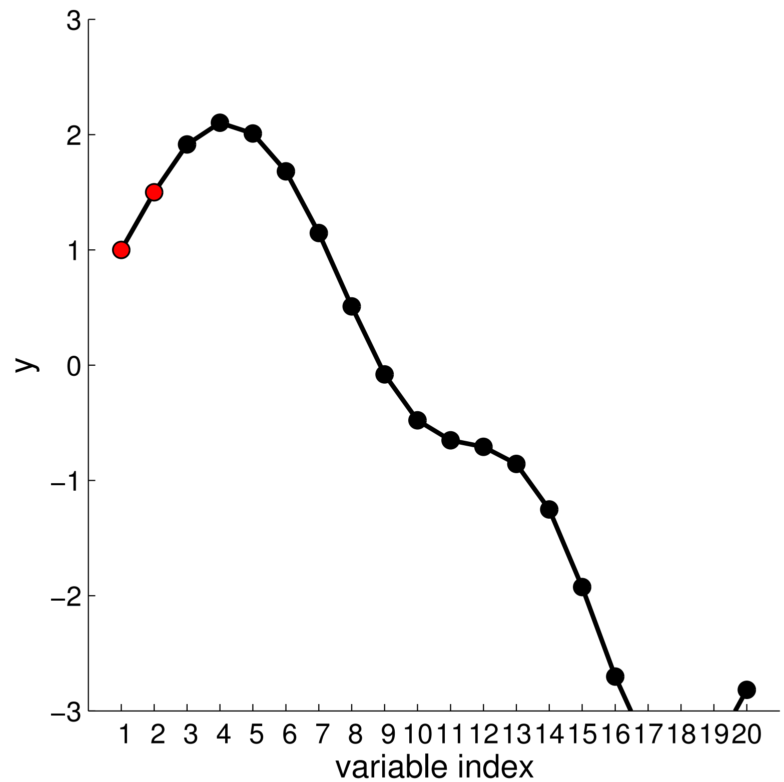
$$\Sigma =$$



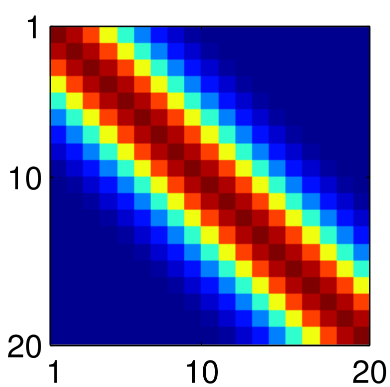


New visualisation

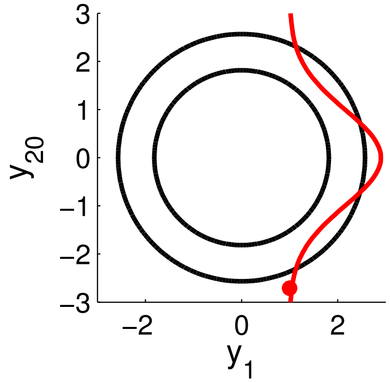


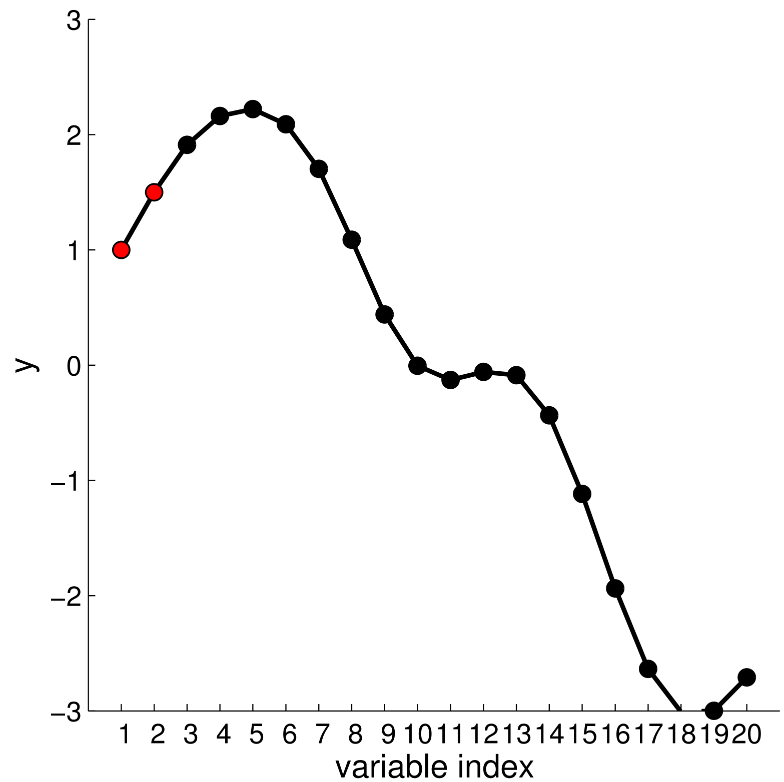


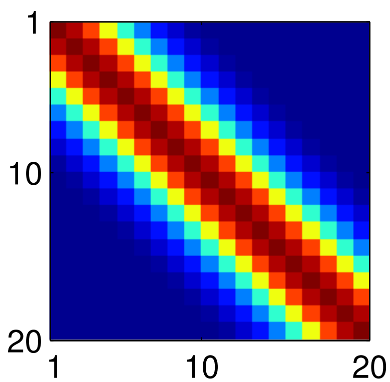
$$\Sigma =$$



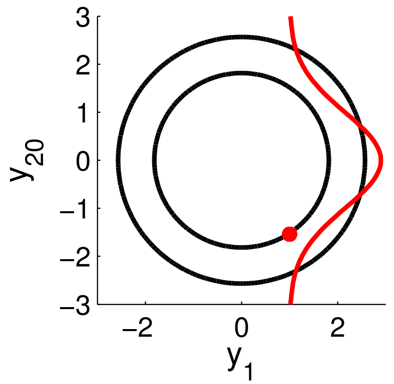
New visualisation

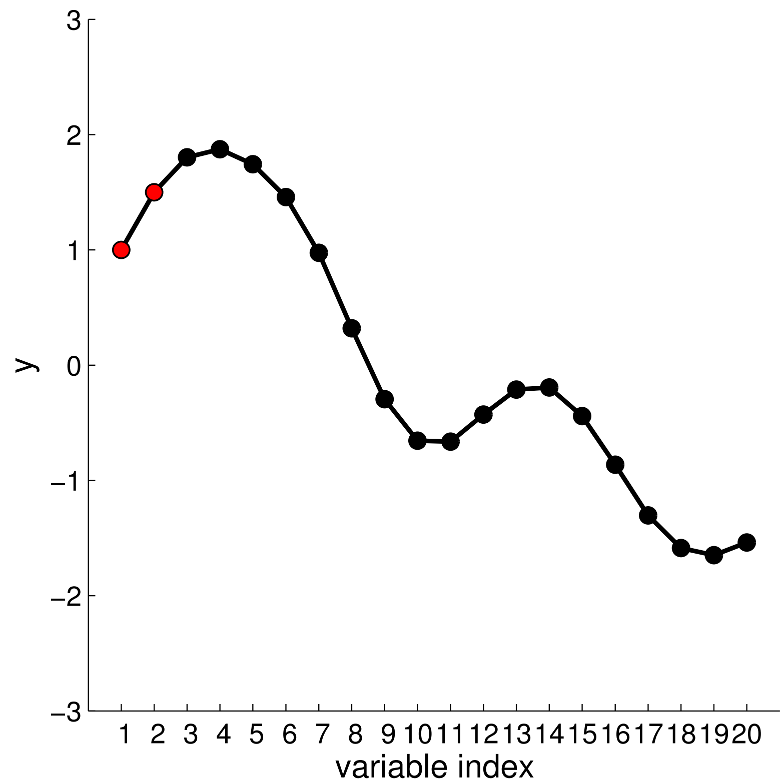




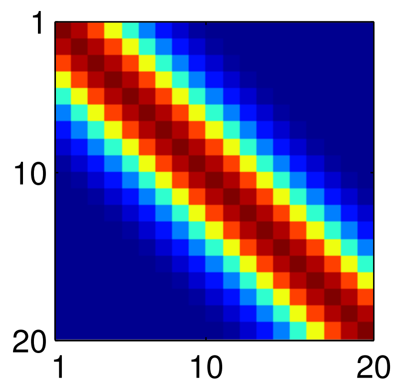
$\Sigma =$ 

New visualisation

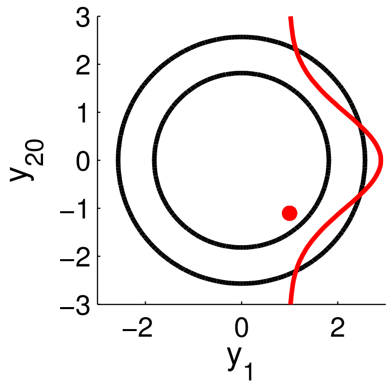


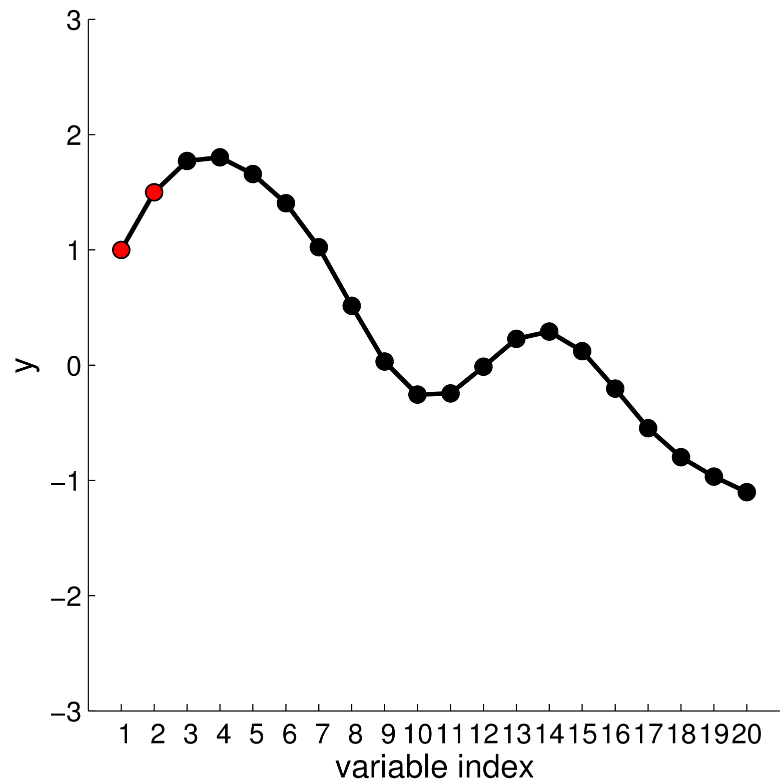


$\Sigma =$

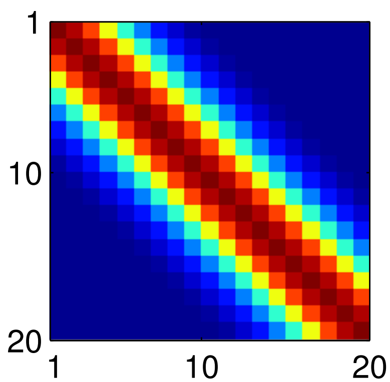


New visualisation

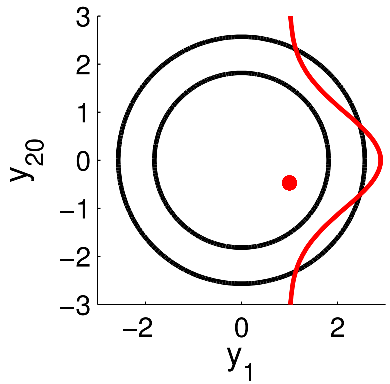


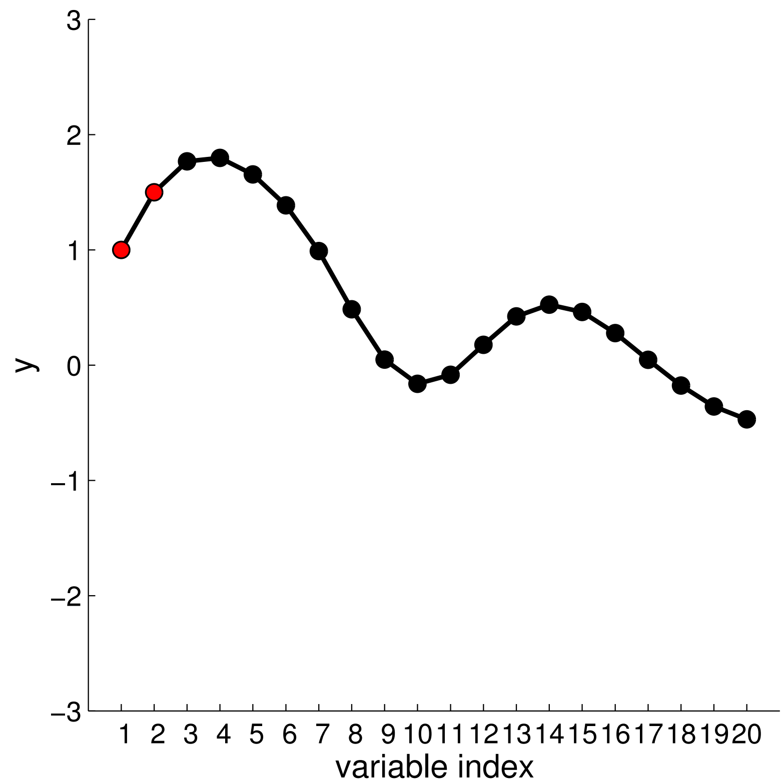


$\Sigma =$

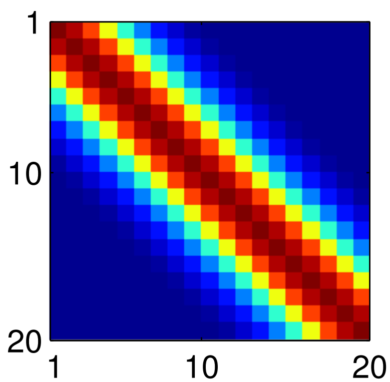


New visualisation

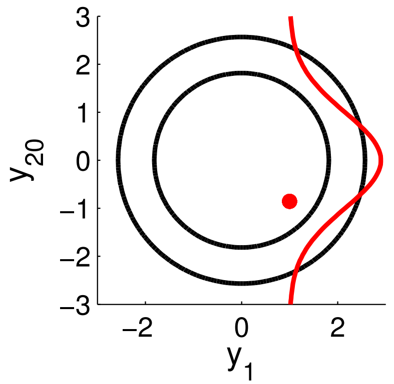


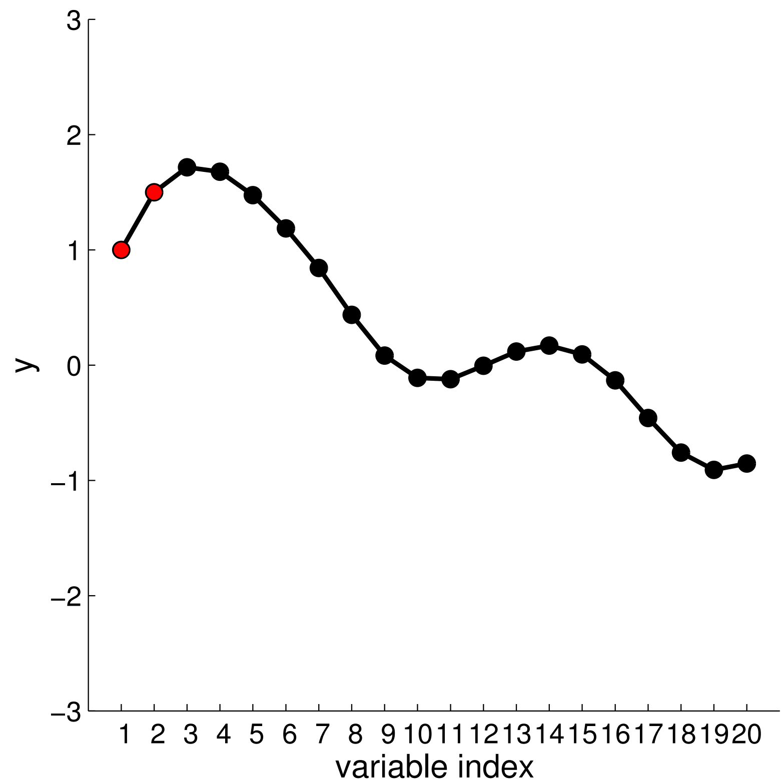


$\Sigma =$

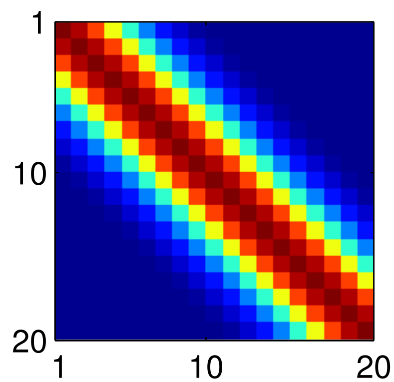


New visualisation

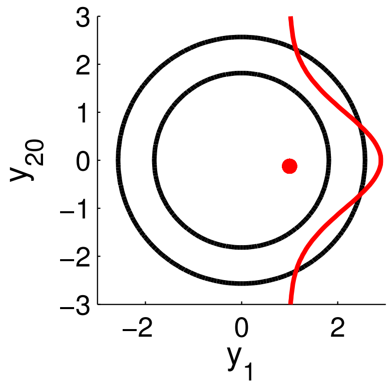


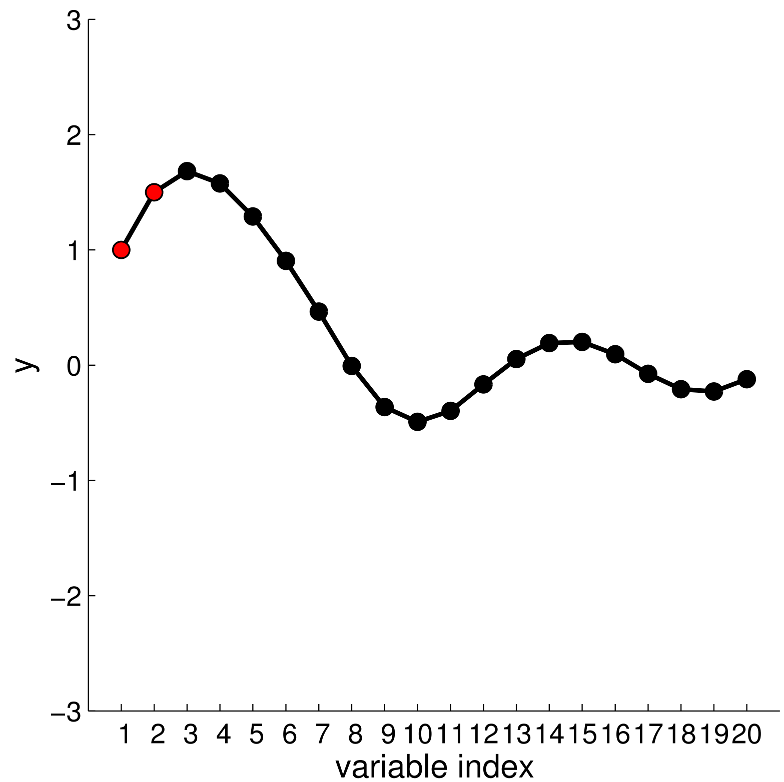


$\Sigma =$

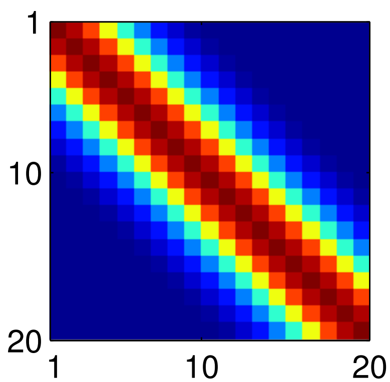


New visualisation

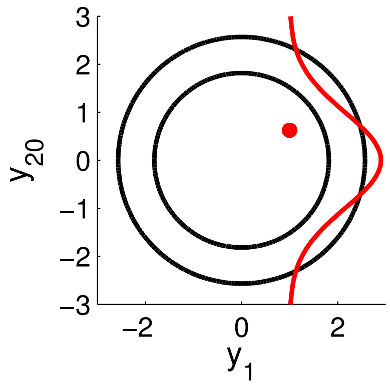


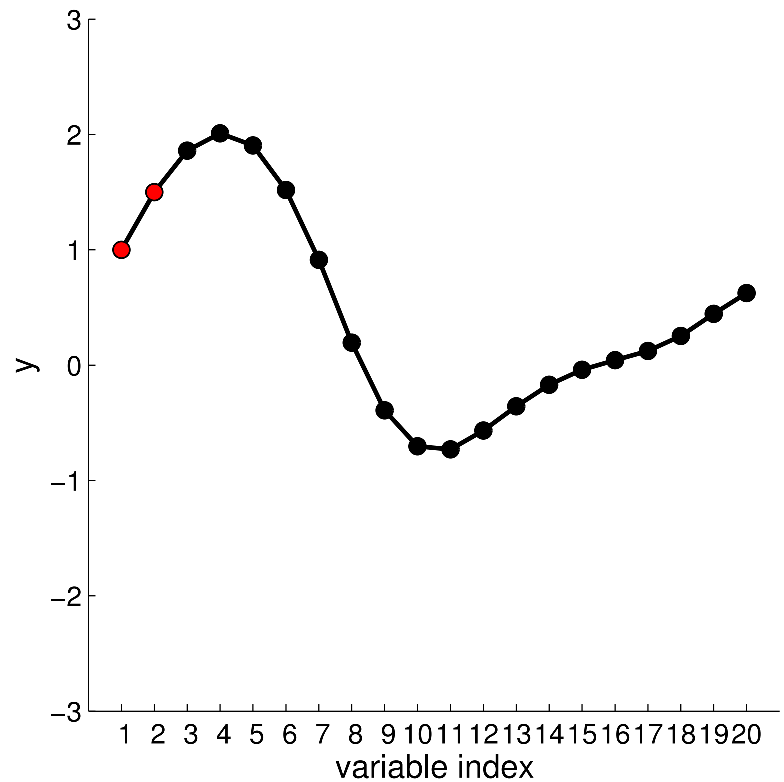


$$\Sigma =$$

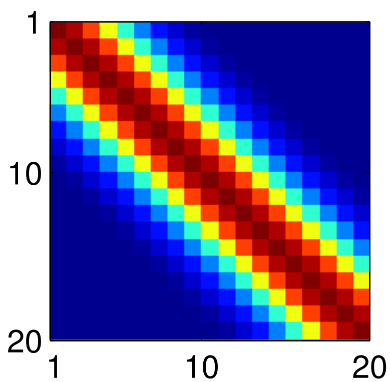


New visualisation

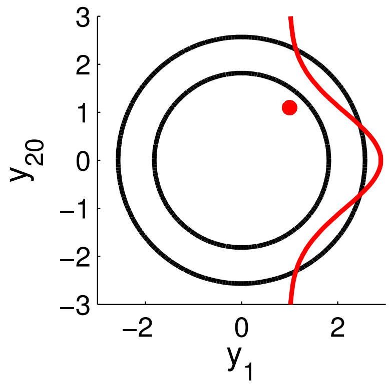


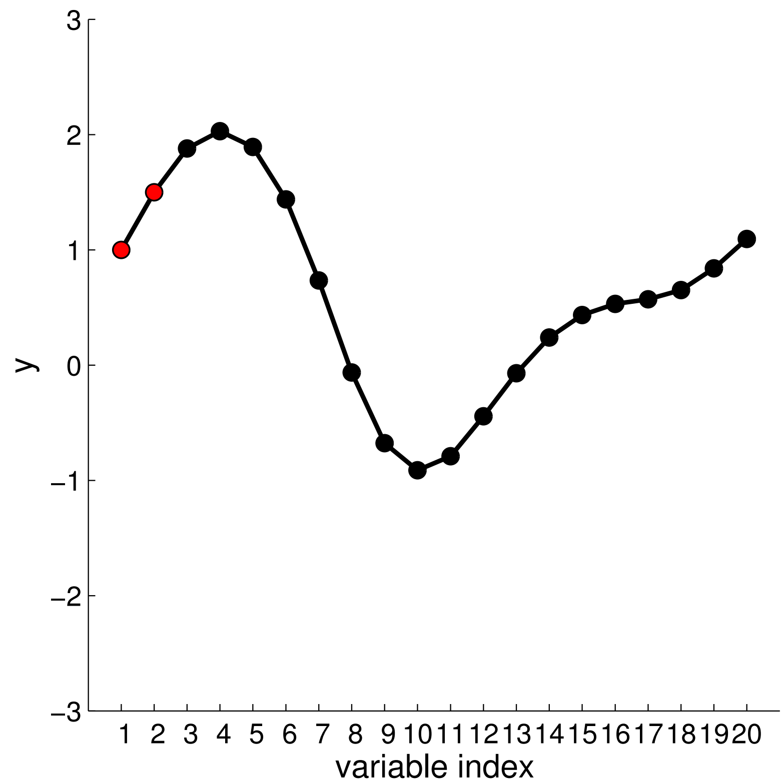


$$\Sigma =$$

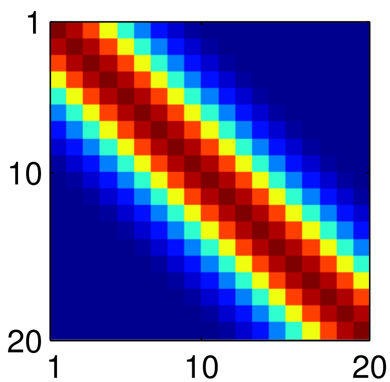


New visualisation

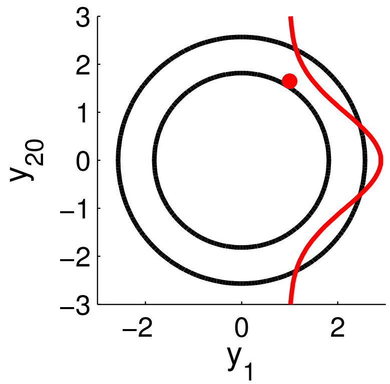


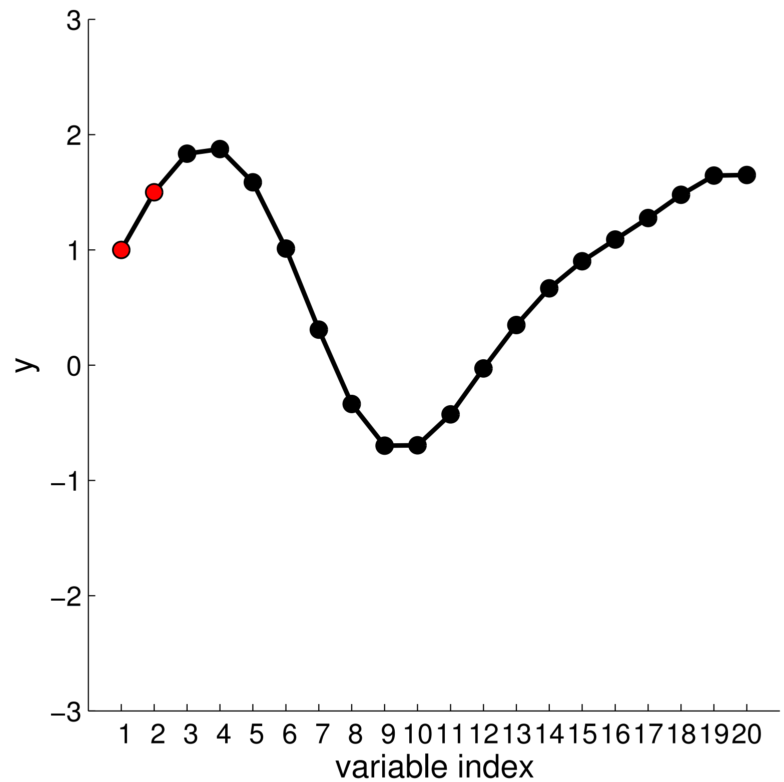


$\Sigma =$

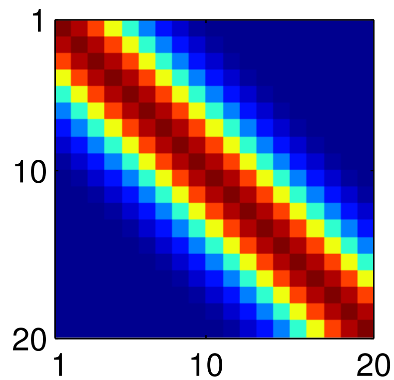


New visualisation

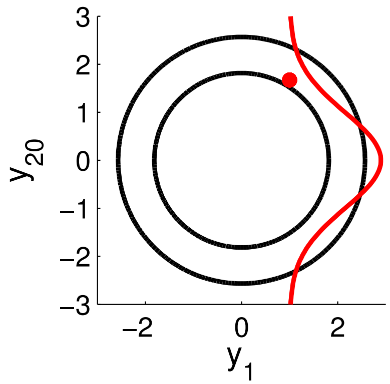


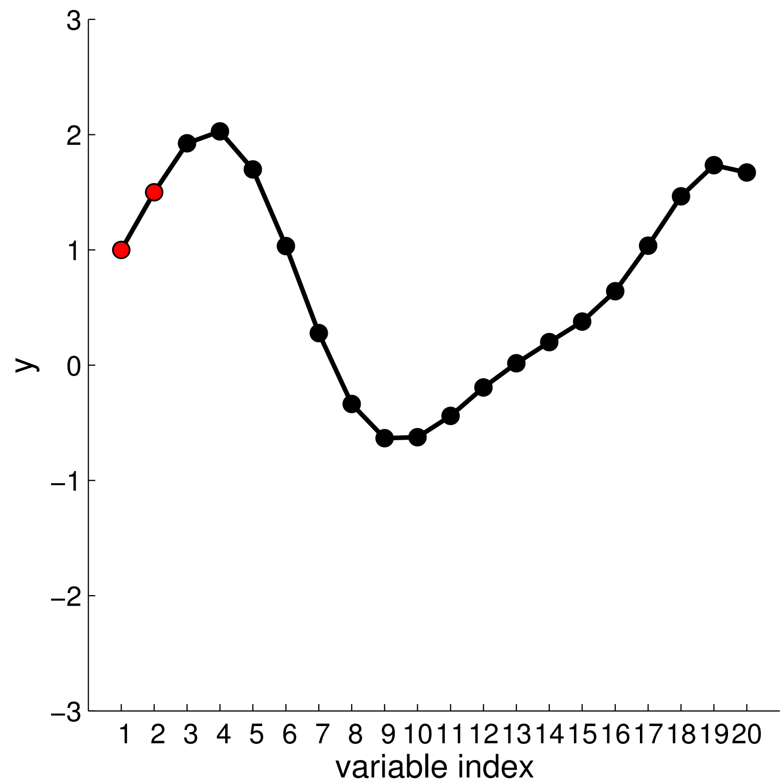


$\Sigma =$

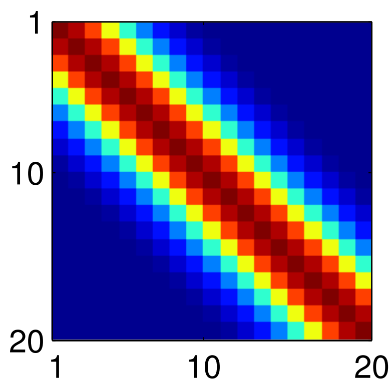


New visualisation

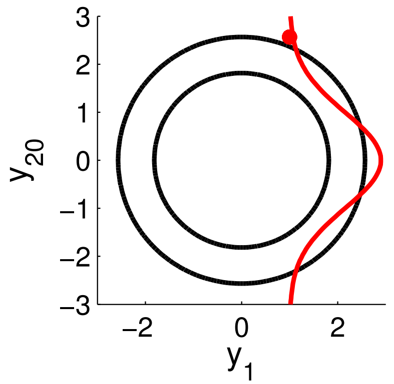




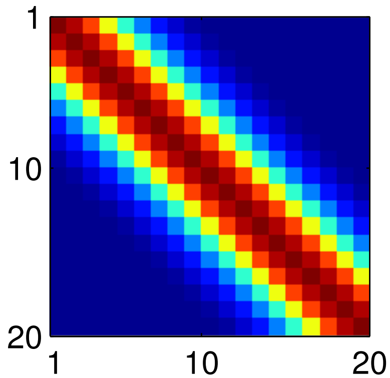
$\Sigma =$

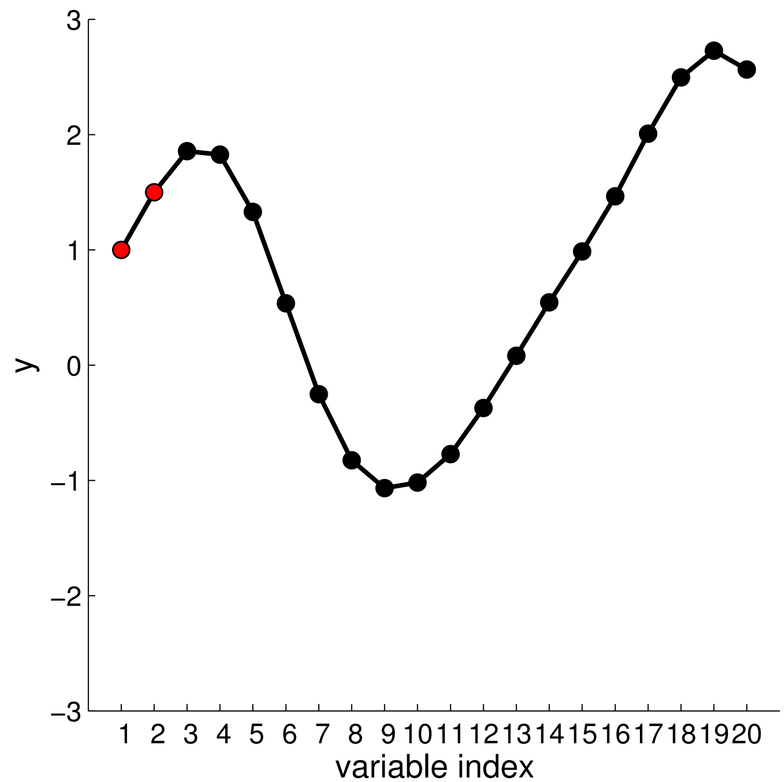


New visualisation

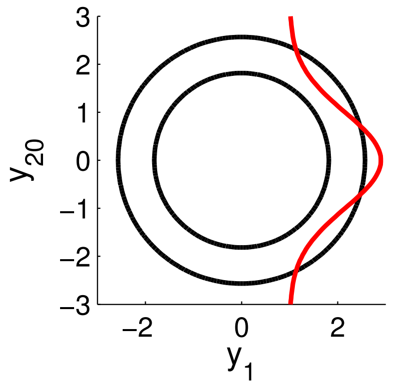


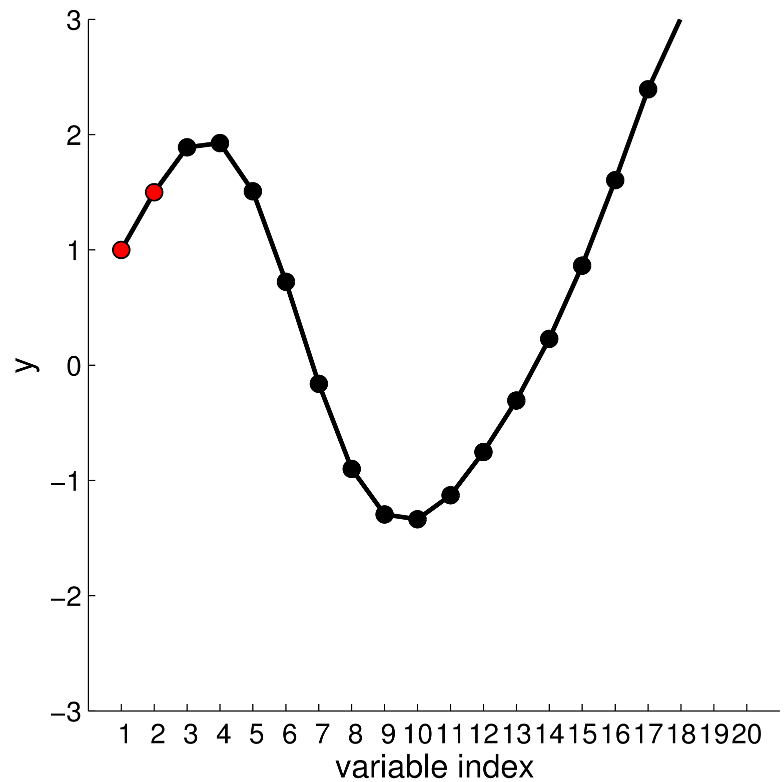
$\Sigma =$



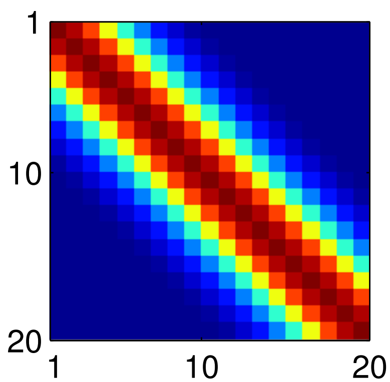


New visualisation



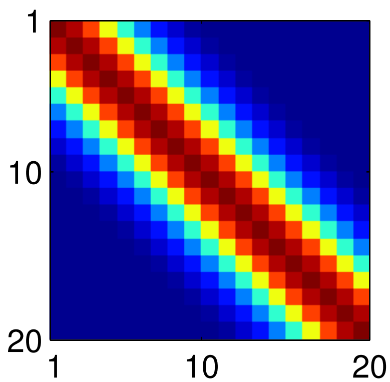


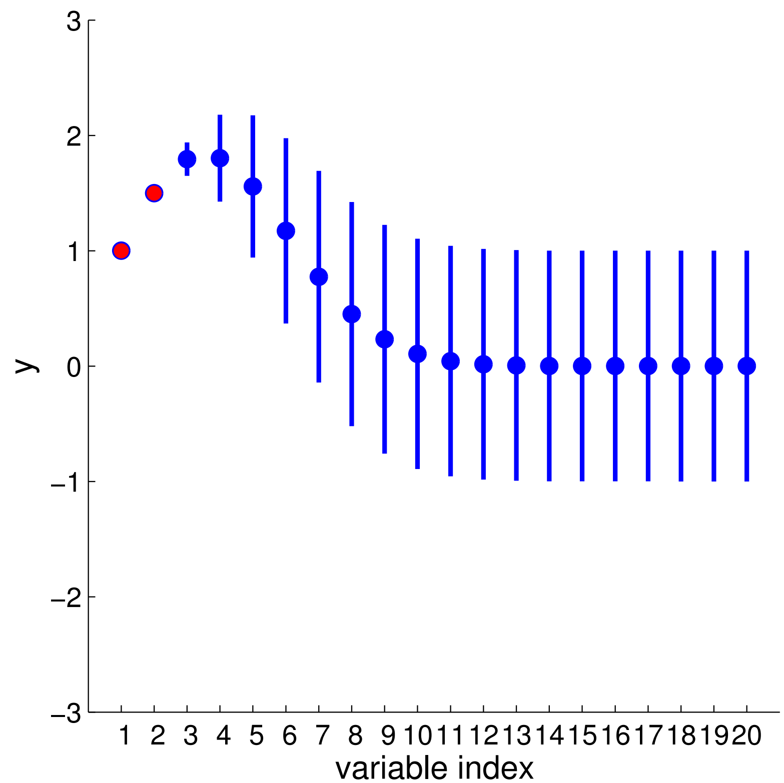
$\Sigma =$



Regression using Gaussians

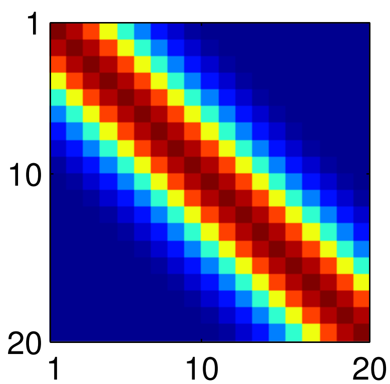
$$\Sigma =$$

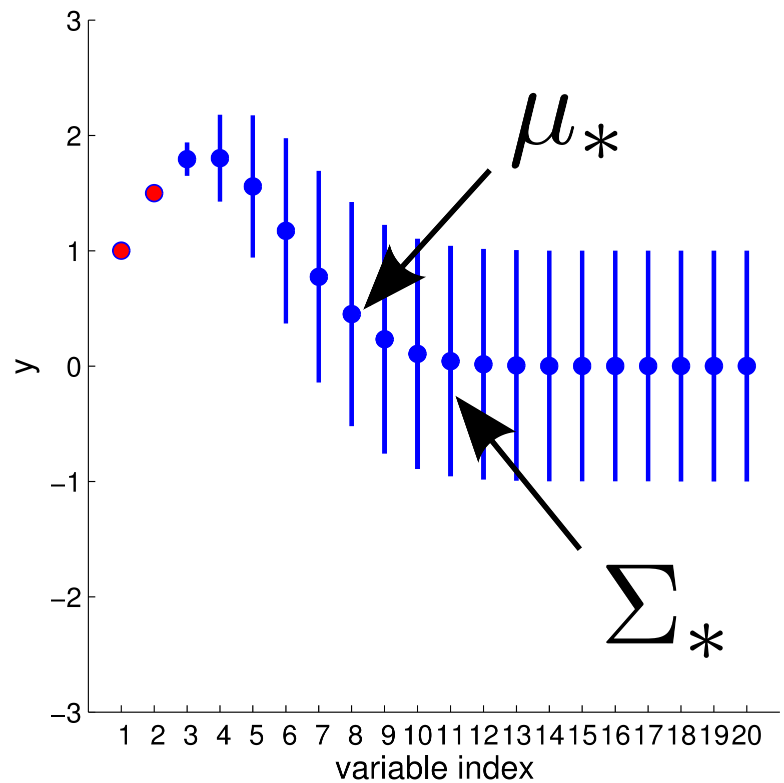




Regression using Gaussians

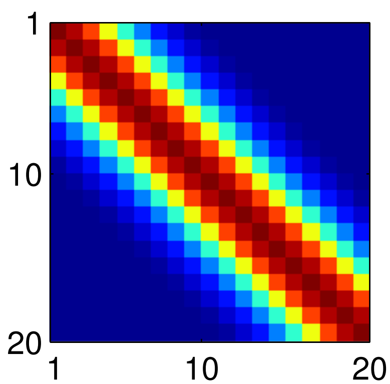
$\Sigma =$

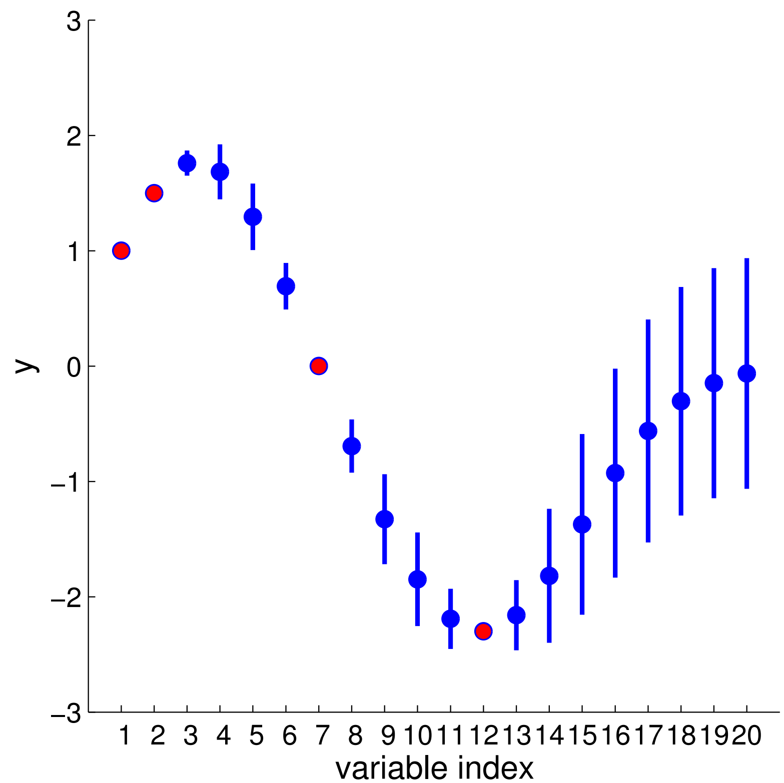




Regression using Gaussians

$$\Sigma =$$



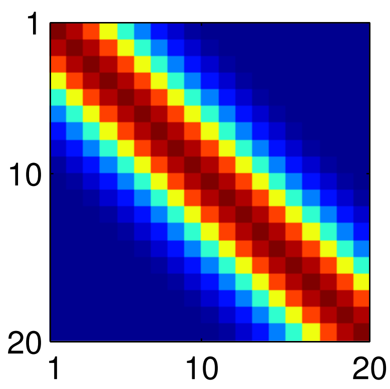


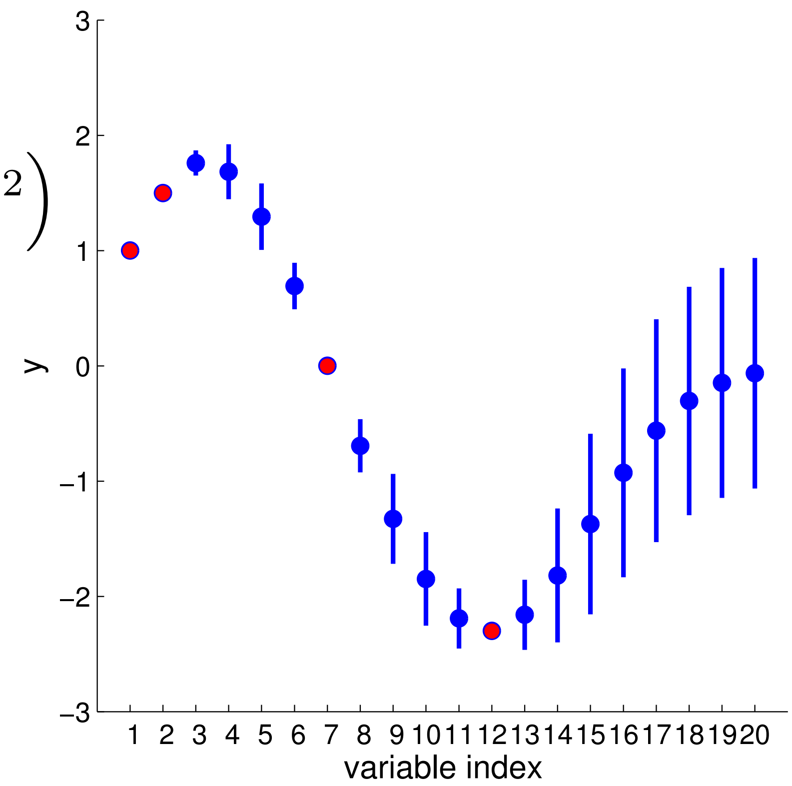
Regression using Gaussians

$$\Sigma(\mathsf{x}_1, \mathsf{x}_2) = \mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) + \mathsf{I}\sigma_{\mathsf{y}}^2$$

$$\mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) = \sigma^2 \exp\left(-\frac{1}{2\mathsf{l}^2}(\mathsf{x}_1 - \mathsf{x}_2)^2\right)$$

$\Sigma =$

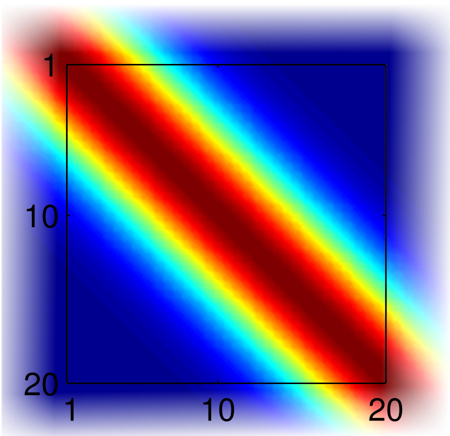


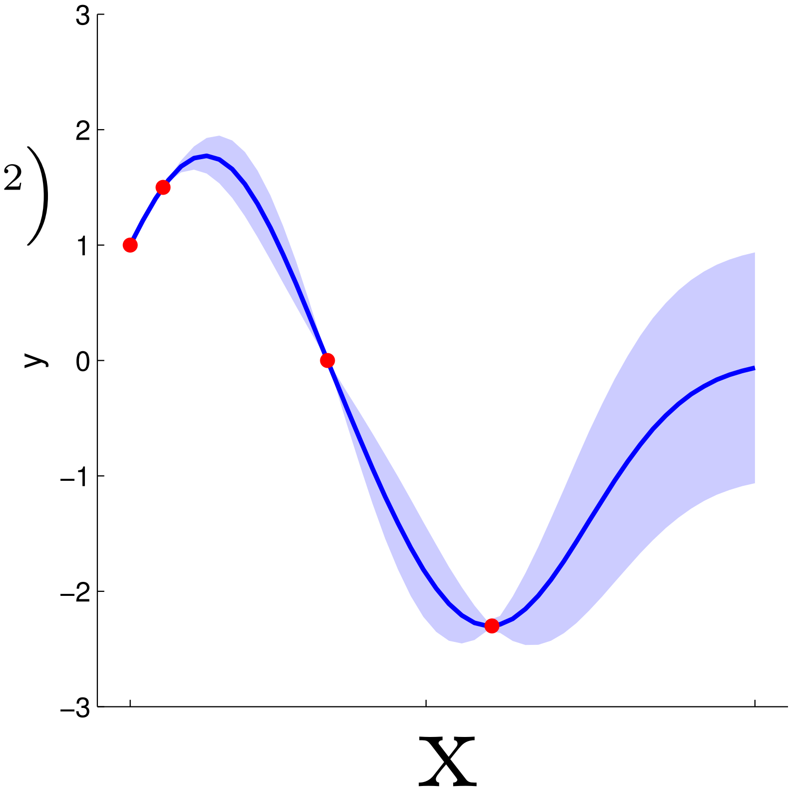


Regression: probabilistic inference in function space

$$\Sigma(\mathsf{x}_1, \mathsf{x}_2) = \mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) + \mathsf{I}\sigma_{\mathsf{y}}^2$$

$$\mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) = \sigma^2 \exp\left(-\frac{1}{2\mathsf{l}^2}(\mathsf{x}_1 - \mathsf{x}_2)^2\right)$$

$\Sigma =$ 



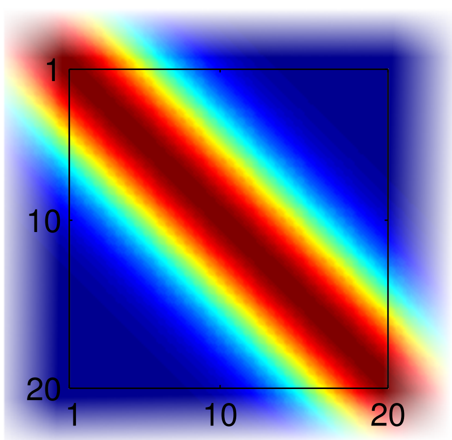
Regression: probabilistic inference in function space

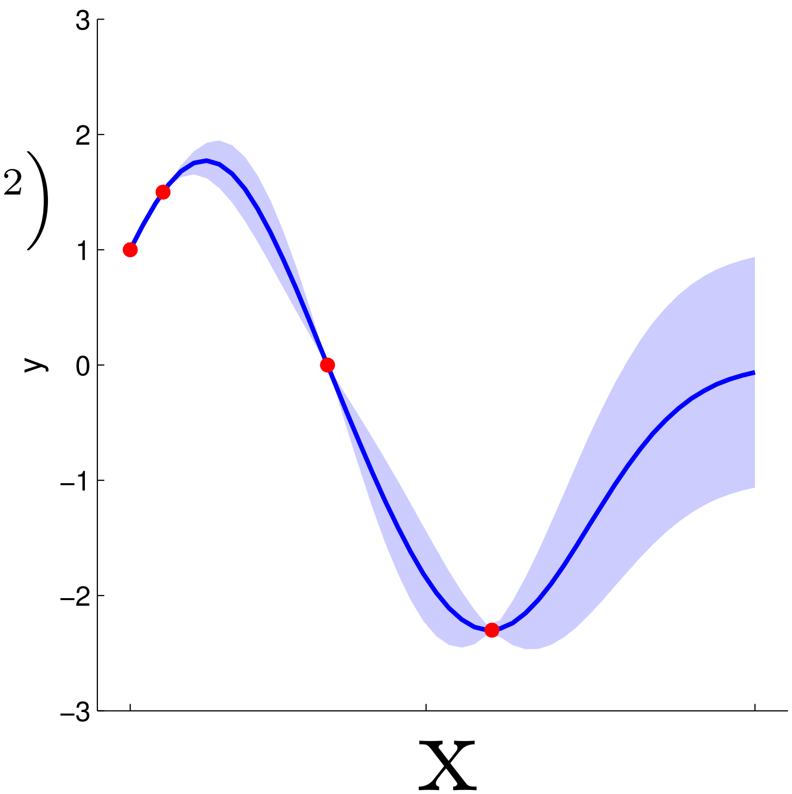
Non-parametric (∞-parametric)

$$p(\mathsf{y}|\theta) = \mathcal{N}(0, \Sigma)$$

$$\Sigma(\mathsf{x}_1, \mathsf{x}_2) = \mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) + \mathsf{I}\sigma_{\mathsf{y}}^2$$

$$\mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) = \sigma^2 \exp\left(-\frac{1}{2\mathsf{l}^2}(\mathsf{x}_1 - \mathsf{x}_2)^2\right)$$

$\Sigma =$ 



Parametric model

$$\mathsf{y}(\mathsf{x}) = f(\mathsf{x}; \theta) + \sigma_{\mathsf{y}}\epsilon$$

$$\epsilon \sim \mathcal{N}(0, 1)$$

Regression: probabilistic inference in function space

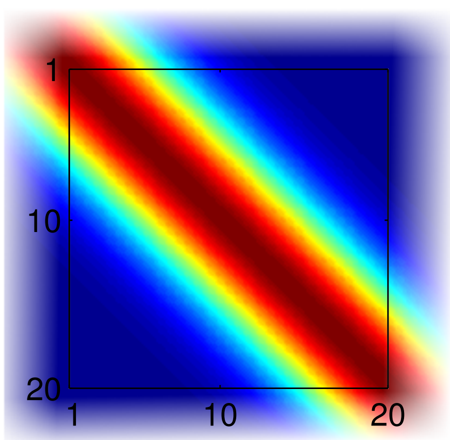
Non-parametric (∞-parametric)

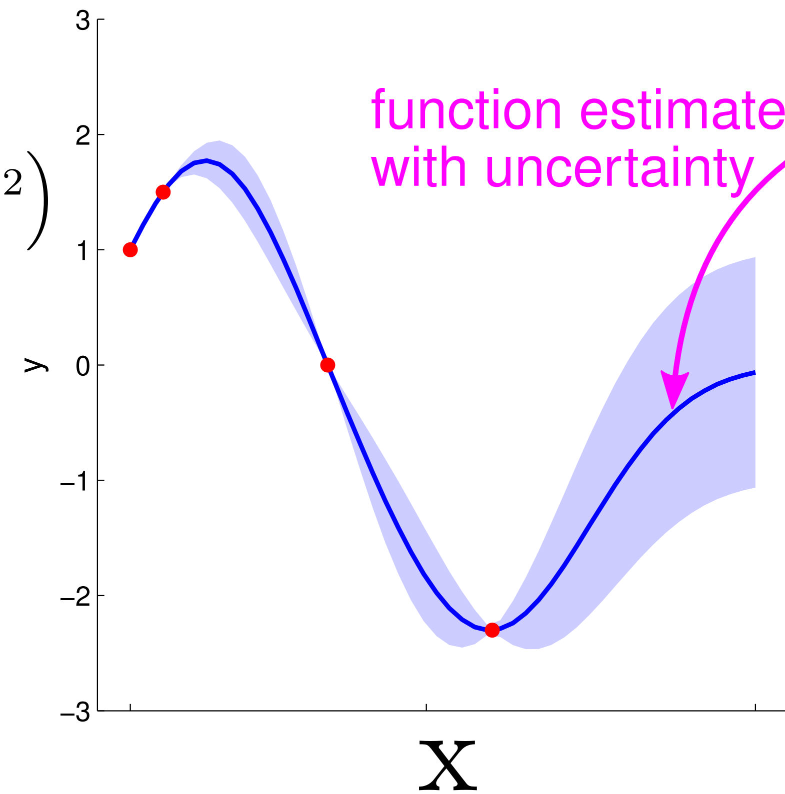
$$p(\mathsf{y}|\theta) = \mathcal{N}(0, \Sigma)$$

$$\Sigma(\mathsf{x}_1, \mathsf{x}_2) = \mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) + \mathsf{I}\sigma_{\mathsf{y}}^2$$

$$\mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) = \sigma^2 \exp\left(-\frac{1}{2\mathsf{l}^2}(\mathsf{x}_1 - \mathsf{x}_2)^2\right)$$

$\Sigma =$





Parametric model

$$\mathsf{y}(\mathsf{x}) = f(\mathsf{x}; \theta) + \sigma_{\mathsf{y}}\epsilon$$

$$\epsilon \sim \mathcal{N}(0, 1)$$

Regression: probabilistic inference in function space

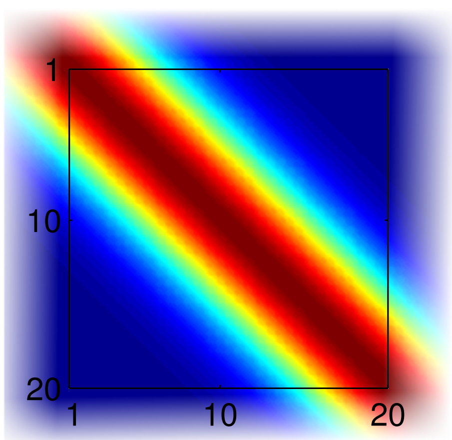
Non-parametric (∞-parametric)

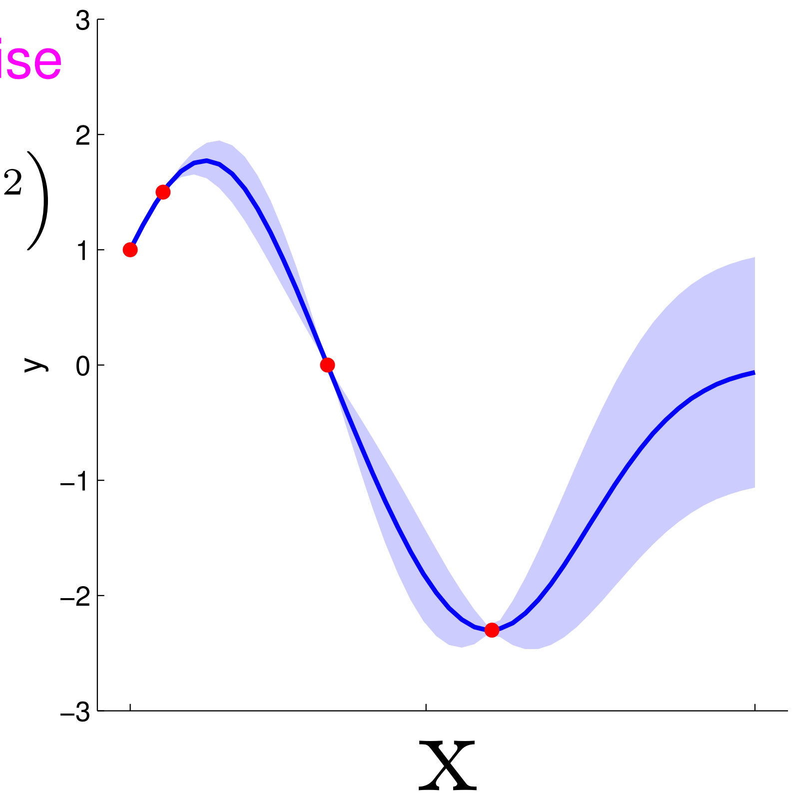
$p(\mathsf{y}|\theta) = \mathcal{N}(0, \Sigma)$

$\Sigma(\mathsf{x}_1, \mathsf{x}_2) = \mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) + \mathsf{I}\sigma_{\mathsf{y}}^2$ ←noise

$\mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) = \sigma^2 \exp\left(-\frac{1}{2\mathsf{l}^2}(\mathsf{x}_1 - \mathsf{x}_2)^2\right)$

$\Sigma =$





Parametric model

$\mathsf{y}(\mathsf{x}) = f(\mathsf{x}; \theta) + \sigma_{\mathsf{y}}\epsilon$

$\epsilon \sim \mathcal{N}(0, 1)$

Regression: probabilistic inference in function space

Non-parametric (∞-parametric)

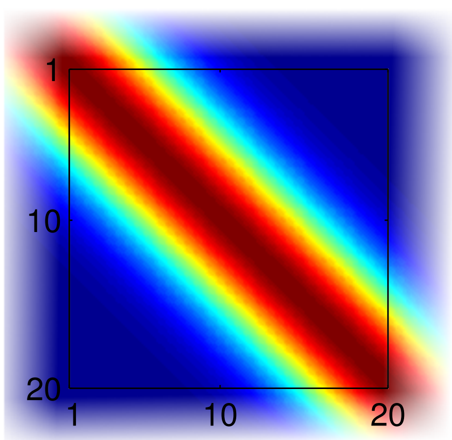
$p(\mathsf{y}|\theta) = \mathcal{N}(0, \Sigma)$

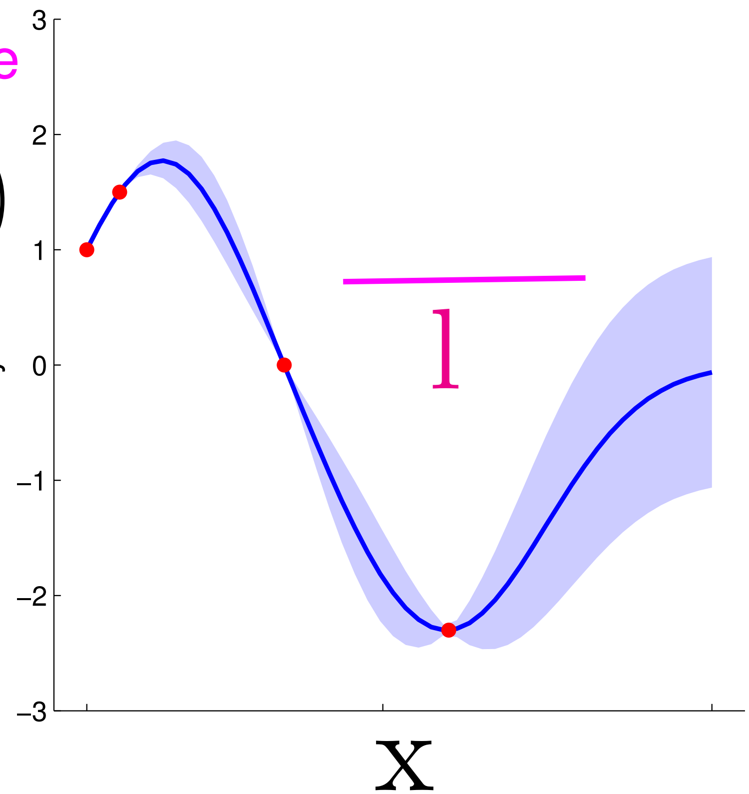
$\Sigma(\mathsf{x}_1, \mathsf{x}_2) = \mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) + \mathsf{I}\sigma_{\mathsf{y}}^2$ ← noise

$\mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) = \sigma^2 \exp\left(-\frac{1}{2\mathsf{l}^2}(\mathsf{x}_1 - \mathsf{x}_2)^2\right)$

↑ horizontal-scale

$\Sigma =$





Parametric model

$\mathsf{y}(\mathsf{x}) = f(\mathsf{x}; \theta) + \sigma_{\mathsf{y}}\epsilon$

$\epsilon \sim \mathcal{N}(0, 1)$

Regression: probabilistic inference in function space

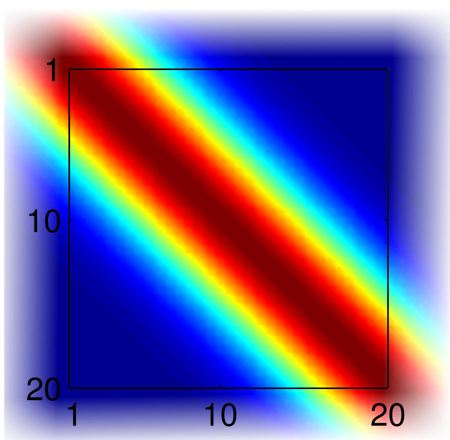
Non-parametric (∞-parametric)

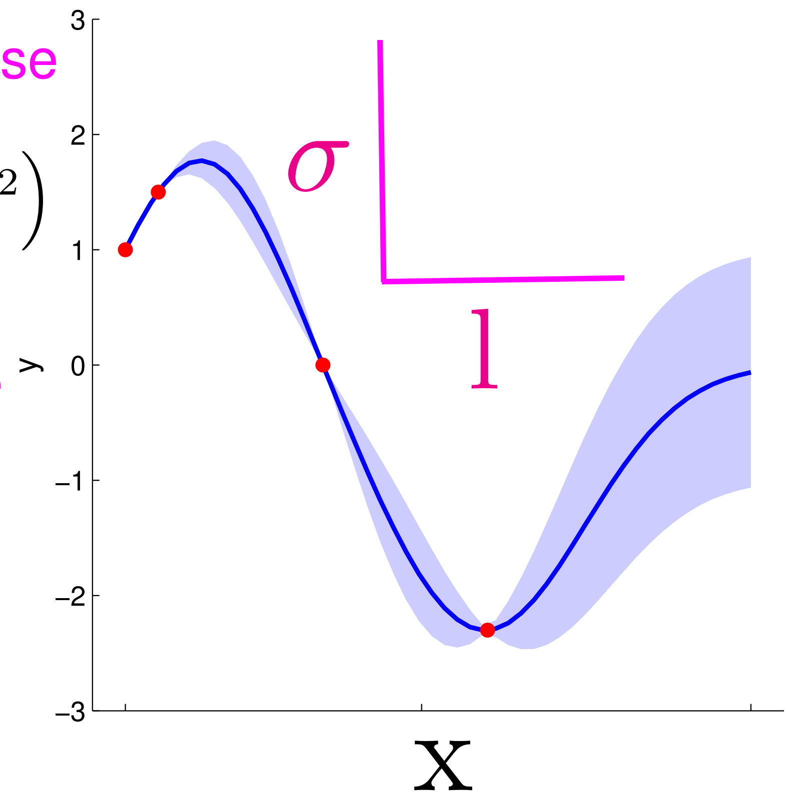
$$p(\mathsf{y}|\theta) = \mathcal{N}(0, \Sigma)$$

$$\Sigma(\mathsf{x}_1, \mathsf{x}_2) = \mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) + \mathsf{I}\sigma_{\mathsf{y}}^2$$ ← noise

$$\mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) = \sigma^2 \exp\left(-\frac{1}{2\mathsf{l}^2}(\mathsf{x}_1 - \mathsf{x}_2)^2\right)$$

vertical-scale horizontal-scale

$\Sigma =$ 



Parametric model

$$\mathsf{y}(\mathsf{x}) = f(\mathsf{x}; \theta) + \sigma_{\mathsf{y}}\epsilon$$

$$\epsilon \sim \mathcal{N}(0, 1)$$

Mathematical Foundations: Definition

Gaussian process = generalization of multivariate Gaussian distribution to infinitely many variables.

Definition: a Gaussian process is a collection of random variables, any finite number of which have (consistent) Gaussian distributions.

A Gaussian distribution is fully specified by a mean vector, $\boldsymbol{\mu}$, and covariance matrix Σ:

$$\mathbf{f} = (\mathsf{f}_1, \ldots, \mathsf{f}_n) \sim \mathcal{N}(\mu, \Sigma), \quad \text{indices} \quad i = 1, \ldots, n$$

A Gaussian process is fully specified by a mean function $m(\mathbf{x})$ and covariance function $\mathsf{K}(\mathbf{x}, \mathbf{x}')$:

$$f(\mathbf{x}) \sim \mathcal{GP}\left(m(\mathbf{x}), \mathsf{K}(\mathbf{x}, \mathbf{x}')\right), \quad \text{indices} \quad \mathbf{x}$$

Mathematical Foundations: Marginalisation

Q1. A GP is "like" a Gaussian distribution with an infinitely long mean vector and an "infinite by infinite" covariance matrix, so how do we represent it on a computer?

Mathematical Foundations: Marginalisation

Q1. A GP is "like" a Gaussian distribution with an infinitely long mean vector and an "infinite by infinite" covariance matrix, so how do we represent it on a computer?

We are saved by the marginalisation property:

$$p(\mathbf{y}_1) = \int p(\mathbf{y}_1, \mathbf{y}_2) \mathrm{d}\mathbf{y}_2$$

$$p(\mathbf{y}_1, \mathbf{y}_2) = \mathcal{N}\left(\begin{bmatrix} \mathbf{a} \\ \mathbf{b} \end{bmatrix}, \begin{bmatrix} \mathrm{A} & \mathrm{B} \\ \mathrm{B}^\mathsf{T} & \mathrm{C} \end{bmatrix} \right) \implies p(\mathbf{y}_1) = \mathcal{N}(\mathbf{a}, \mathrm{A})$$

$\implies$ Only need to represent finite dimensional projections of GPs on computer.

nonparametric model: complexity increases with size of the data

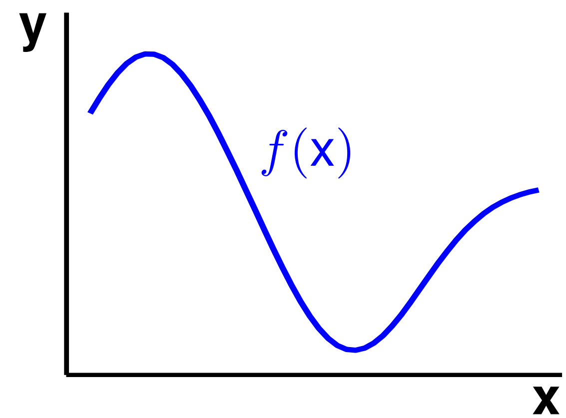
blackboard: revisiting ARIMA

Mathematical Foundations: Regression

Q3. What's the formal justification for how we were using GPs for regression?

Generative model (like non-linear regression)

$\mathsf{y}(\mathsf{x}) = f(\mathsf{x}) + \epsilon\sigma_{\mathsf{y}}$



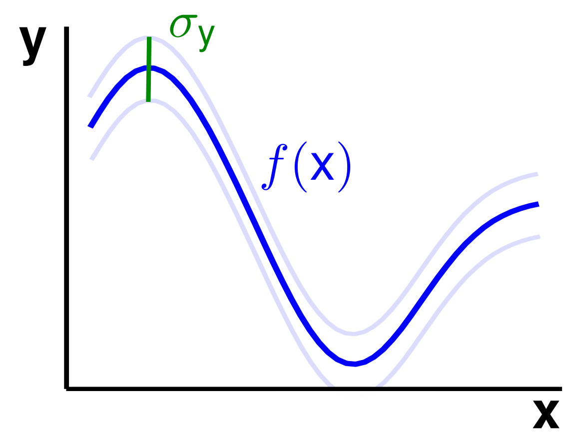
Mathematical Foundations: Regression

Q3. What's the formal justification for how we were using GPs for regression?

Generative model (like non-linear regression)

$$\mathsf{y}(\mathsf{x}) = f(\mathsf{x}) + \epsilon \sigma_{\mathsf{y}}$$

$$p(\epsilon) = \mathcal{N}(0, 1)$$



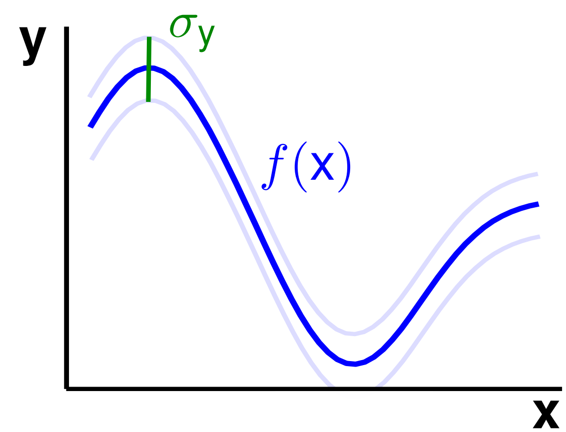
Mathematical Foundations: Regression

Q3. What's the formal justification for how we were using GPs for regression?

Generative model (like non-linear regression)

$$\mathsf{y}(\mathsf{x}) = f(\mathsf{x}) + \epsilon\sigma_{\mathsf{y}}$$

$$p(\epsilon) = \mathcal{N}(0, 1)$$



place GP prior over the non-linear function

$$p(f(\mathsf{x})|\theta) = \mathcal{GP}(0, \mathsf{K}(\mathsf{x}, \mathsf{x}'))$$

$$\mathsf{K}(\mathsf{x}, \mathsf{x}') = \sigma^2 \exp\left(-\frac{1}{2l^2}(\mathsf{x} - \mathsf{x}')^2\right)$$ (smoothly wiggling functions expected)

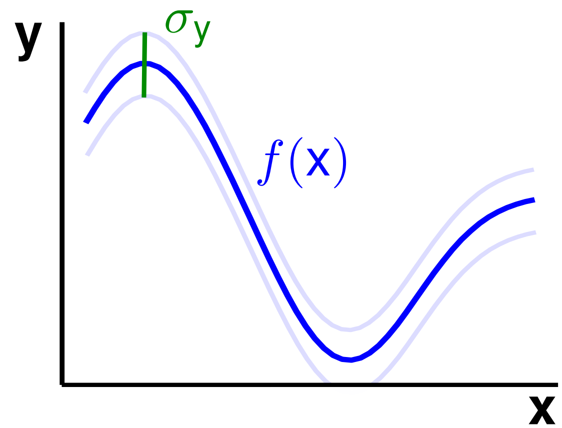
Mathematical Foundations: Regression

Q3. What's the formal justification for how we were using GPs for regression?

Generative model (like non-linear regression)

$$\mathsf{y}(\mathsf{x}) = f(\mathsf{x}) + \epsilon\sigma_{\mathsf{y}}$$

$$p(\epsilon) = \mathcal{N}(0, 1)$$



place GP prior over the non-linear function

$$p(f(\mathsf{x})|\theta) = \mathcal{GP}(0, \mathsf{K}(\mathsf{x}, \mathsf{x}'))$$

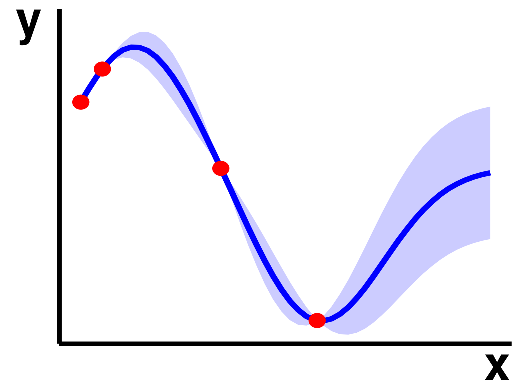
$$\mathsf{K}(\mathsf{x}, \mathsf{x}') = \sigma^2 \exp\left(-\frac{1}{2l^2}(\mathsf{x} - \mathsf{x}')^2\right)$$ (smoothly wiggling functions expected)

since the sum of two Gaussians is a Gaussian, the model induces a GP over $\mathsf{y}(\mathsf{x})$

$$p(\mathsf{y}(\mathsf{x})|\theta) = \mathcal{GP}(0, \mathsf{K}(\mathsf{x}, \mathsf{x}') + \mathrm{I}\sigma_{\mathsf{y}}^2)$$

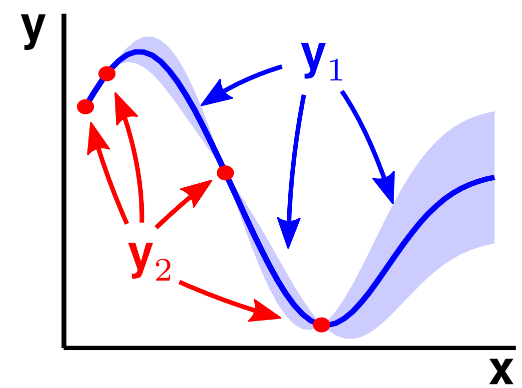
Mathematical foundations: Prediction

Q4. How do we make predictions?



Mathematical foundations: Prediction

Q4. How do we make predictions?



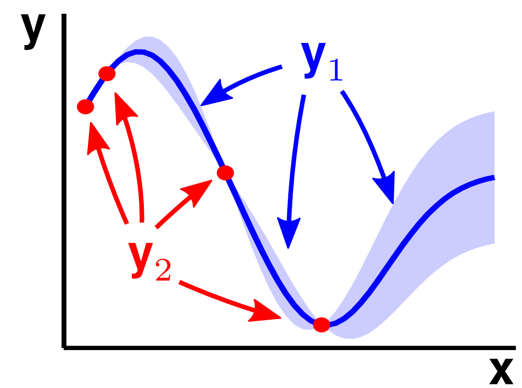
$$p(\mathbf{y}_1|\mathbf{y}_2) = \frac{p(\mathbf{y}_1, \mathbf{y}_2)}{p(\mathbf{y}_2)}$$

Mathematical foundations: Prediction

Q4. How do we make predictions?

$$p(\mathbf{y}_1, \mathbf{y}_2) = \mathcal{N}\left(\begin{bmatrix} \mathbf{a} \\ \mathbf{b} \end{bmatrix}, \begin{bmatrix} \mathrm{A} & \mathrm{B} \\ \mathrm{B}^\mathsf{T} & \mathrm{C} \end{bmatrix} \right)$$

$$p(\mathbf{y}_1 | \mathbf{y}_2) = \frac{p(\mathbf{y}_1, \mathbf{y}_2)}{p(\mathbf{y}_2)}$$



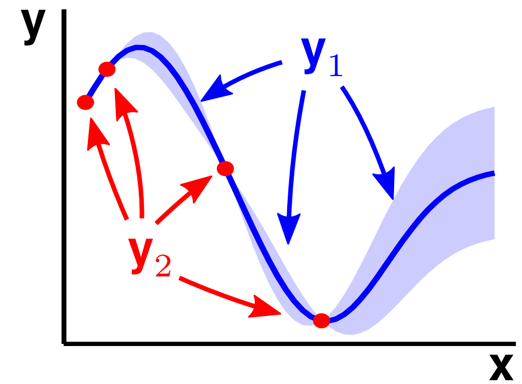
Mathematical foundations: Prediction

Q4. How do we make predictions?

$$p(\mathbf{y}_1, \mathbf{y}_2) = \mathcal{N}\left(\begin{bmatrix} \mathbf{a} \\ \mathbf{b} \end{bmatrix}, \begin{bmatrix} \mathrm{A} & \mathrm{B} \\ \mathrm{B}^\mathrm{T} & \mathrm{C} \end{bmatrix}\right)$$

$$p(\mathbf{y}_1 | \mathbf{y}_2) = \frac{p(\mathbf{y}_1, \mathbf{y}_2)}{p(\mathbf{y}_2)}$$

$$p(\mathbf{y}_2) = \mathcal{N}(\mathbf{b}, \mathrm{C})$$



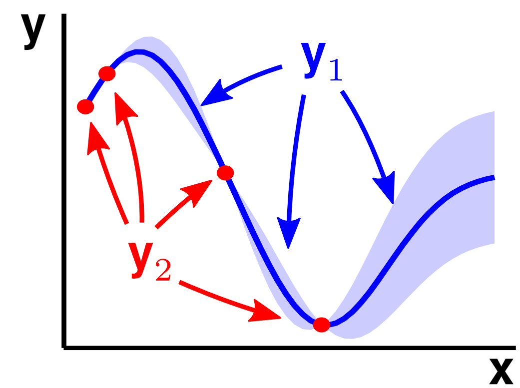
Mathematical foundations: Prediction

Q4. How do we make predictions?

$$p(\mathbf{y}_1, \mathbf{y}_2) = \mathcal{N}\left(\begin{bmatrix} \mathbf{a} \\ \mathbf{b} \end{bmatrix}, \begin{bmatrix} \mathsf{A} & \mathsf{B} \\ \mathsf{B}^\mathsf{T} & \mathsf{C} \end{bmatrix}\right)$$

$$p(\mathbf{y}_1|\mathbf{y}_2) = \frac{p(\mathbf{y}_1, \mathbf{y}_2)}{p(\mathbf{y}_2)}$$

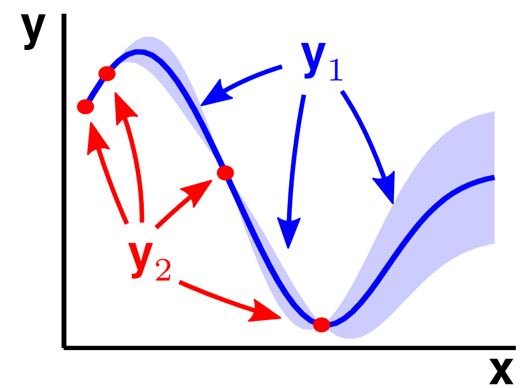
$$p(\mathbf{y}_2) = \mathcal{N}(\mathbf{b}, \mathsf{C})$$



$$\Longrightarrow \quad p(\mathbf{y}_1|\mathbf{y}_2) = \mathcal{N}(\mathbf{a} + \mathsf{B}\mathsf{C}^{-1}(\mathbf{y}_2 - \mathbf{b}), \mathsf{A} - \mathsf{B}\mathsf{C}^{-1}\mathsf{B}^\mathsf{T})$$

Mathematical foundations: Prediction

Q4. How do we make predictions?



$$p(\mathbf{y}_1, \mathbf{y}_2) = \mathcal{N}\left(\begin{bmatrix} \mathbf{a} \\ \mathbf{b} \end{bmatrix}, \begin{bmatrix} \mathrm{A} & \mathrm{B} \\ \mathrm{B}^\mathsf{T} & \mathrm{C} \end{bmatrix}\right)$$

$$p(\mathbf{y}_1 | \mathbf{y}_2) = \frac{p(\mathbf{y}_1, \mathbf{y}_2)}{p(\mathbf{y}_2)}$$

$$p(\mathbf{y}_2) = \mathcal{N}(\mathbf{b}, \mathrm{C})$$

$$\Longrightarrow \quad p(\mathbf{y}_1 | \mathbf{y}_2) = \mathcal{N}(\mathbf{a} + \mathrm{B}\mathrm{C}^{-1}(\mathbf{y}_2 - \mathbf{b}), \mathrm{A} - \mathrm{B}\mathrm{C}^{-1}\mathrm{B}^\mathsf{T})$$

predictive mean

$$\mu_{\mathbf{y}_1|\mathbf{y}_2} = \mathbf{a} + \mathrm{B}\mathrm{C}^{-1}(\mathbf{y}_2 - \mathbf{b})$$
$$= \mathrm{B}\mathrm{C}^{-1}\mathbf{y}_2$$
$$= \mathrm{W}\mathbf{y}_2$$

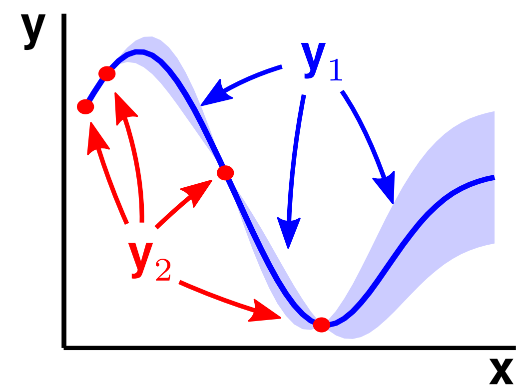
linear in the data

Mathematical foundations: Prediction

Q4. How do we make predictions?

$$p(\mathbf{y}_1, \mathbf{y}_2) = \mathcal{N}\left(\begin{bmatrix} \mathbf{a} \\ \mathbf{b} \end{bmatrix}, \begin{bmatrix} \mathsf{A} & \mathsf{B} \\ \mathsf{B}^\mathsf{T} & \mathsf{C} \end{bmatrix}\right)$$

$$p(\mathbf{y}_1|\mathbf{y}_2) = \frac{p(\mathbf{y}_1, \mathbf{y}_2)}{p(\mathbf{y}_2)} \qquad p(\mathbf{y}_2) = \mathcal{N}(\mathbf{b}, \mathsf{C})$$



$$\Longrightarrow \quad p(\mathbf{y}_1|\mathbf{y}_2) = \mathcal{N}(\mathbf{a} + \mathsf{B}\mathsf{C}^{-1}(\mathbf{y}_2 - \mathbf{b}), \mathsf{A} - \mathsf{B}\mathsf{C}^{-1}\mathsf{B}^\mathsf{T})$$

predictive mean

$$\mu_{\mathbf{y}_1|\mathbf{y}_2} = \mathbf{a} + \mathsf{B}\mathsf{C}^{-1}(\mathbf{y}_2 - \mathbf{b})$$
$$= \mathsf{B}\mathsf{C}^{-1}\mathbf{y}_2$$
$$= \mathsf{W}\mathbf{y}_2$$

linear in the data

predictive covariance

$$\Sigma_{\mathbf{y}_1|\mathbf{y}_2} = \mathsf{A} - \mathsf{B}\mathsf{C}^{-1}\mathsf{B}^\mathsf{T}$$

predictive uncertainty = prior uncertainty − reduction in uncertainty

predictions more confident than prior

What effect do the hyper-parameters have?

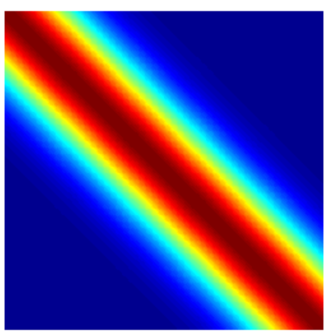
Non-parametric (∞-parametric)

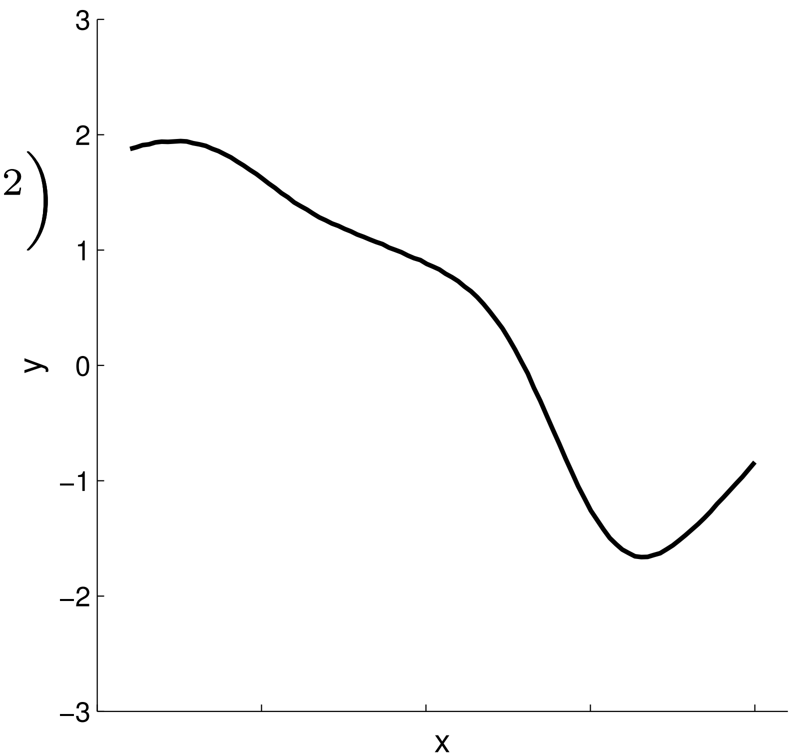
$$p(\mathsf{y}|\theta) = \mathcal{N}(0, \Sigma)$$

$$\Sigma(\mathsf{x}_1, \mathsf{x}_2) = \mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) + \mathsf{I}\sigma_{\mathsf{y}}^2$$

$$\mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) = \sigma^2 \exp\left(-\frac{1}{2\mathsf{l}^2}(\mathsf{x}_1 - \mathsf{x}_2)^2\right)$$

$$\Sigma =$$





Parametric model

$$\mathsf{y}(\mathsf{x}) = f(\mathsf{x};\theta) + \sigma_{\mathsf{y}}\epsilon$$

$$\epsilon \sim \mathcal{N}(0, 1)$$

What effect do the hyper-parameters have?

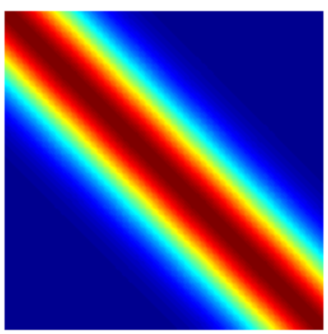
Non-parametric (∞-parametric)

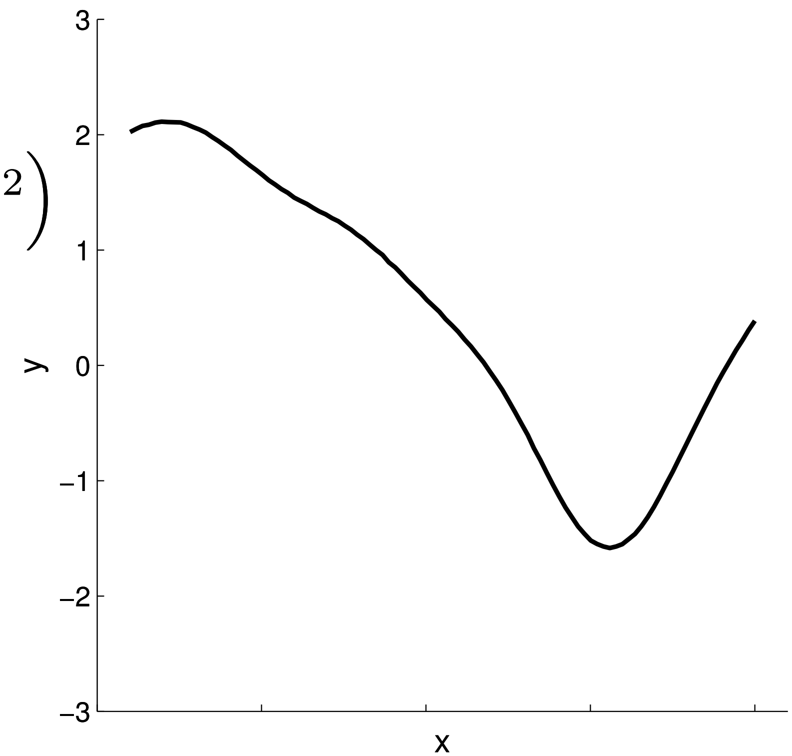
$$p(\mathsf{y}|\theta) = \mathcal{N}(0, \Sigma)$$

$$\Sigma(\mathsf{x}_1, \mathsf{x}_2) = \mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) + \mathsf{I}\sigma_\mathsf{y}^2$$

$$\mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) = \sigma^2 \exp\left(-\frac{1}{2\mathsf{l}^2}(\mathsf{x}_1 - \mathsf{x}_2)^2\right)$$

$\Sigma =$





Parametric model

$$\mathsf{y}(\mathsf{x}) = f(\mathsf{x}; \theta) + \sigma_\mathsf{y}\epsilon$$

$$\epsilon \sim \mathcal{N}(0, 1)$$

What effect do the hyper-parameters have?

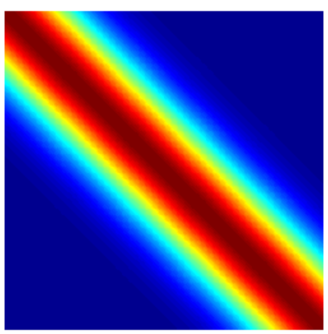
Non-parametric (∞-parametric)

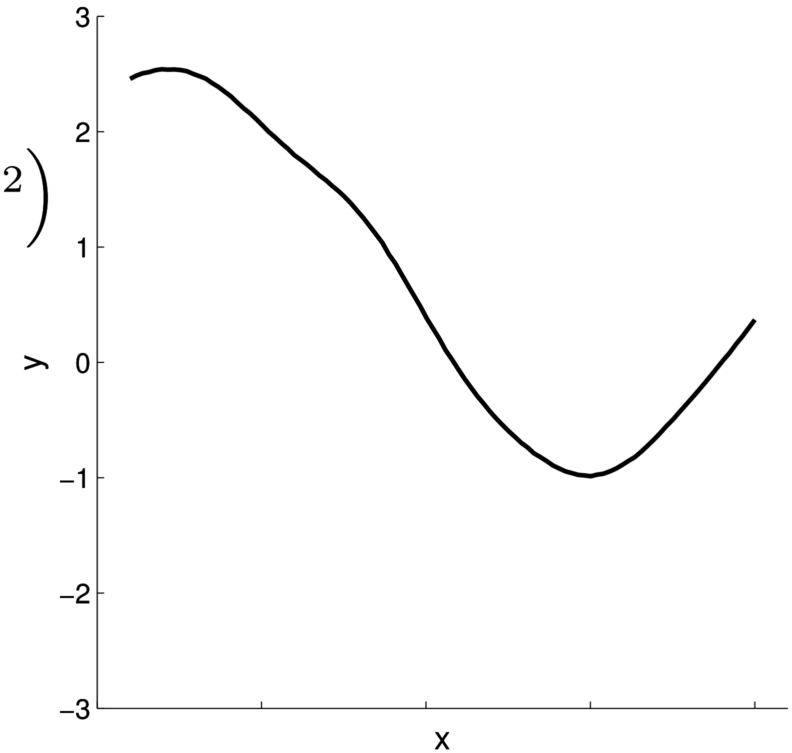
$$p(\mathsf{y}|\theta) = \mathcal{N}(0, \Sigma)$$

$$\Sigma(\mathsf{x}_1, \mathsf{x}_2) = \mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) + \mathsf{I}\sigma_{\mathsf{y}}^2$$

$$\mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) = \sigma^2 \exp\left(-\frac{1}{2\mathsf{l}^2}(\mathsf{x}_1 - \mathsf{x}_2)^2\right)$$

$$\Sigma =$$





Parametric model

$$\mathsf{y}(\mathsf{x}) = f(\mathsf{x}; \theta) + \sigma_{\mathsf{y}}\epsilon$$

$$\epsilon \sim \mathcal{N}(0, 1)$$

What effect do the hyper-parameters have?

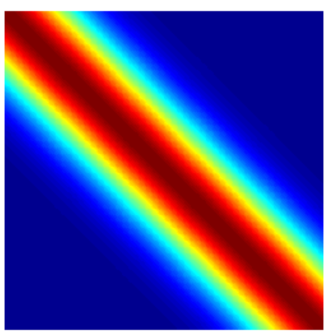
Non-parametric (∞-parametric)

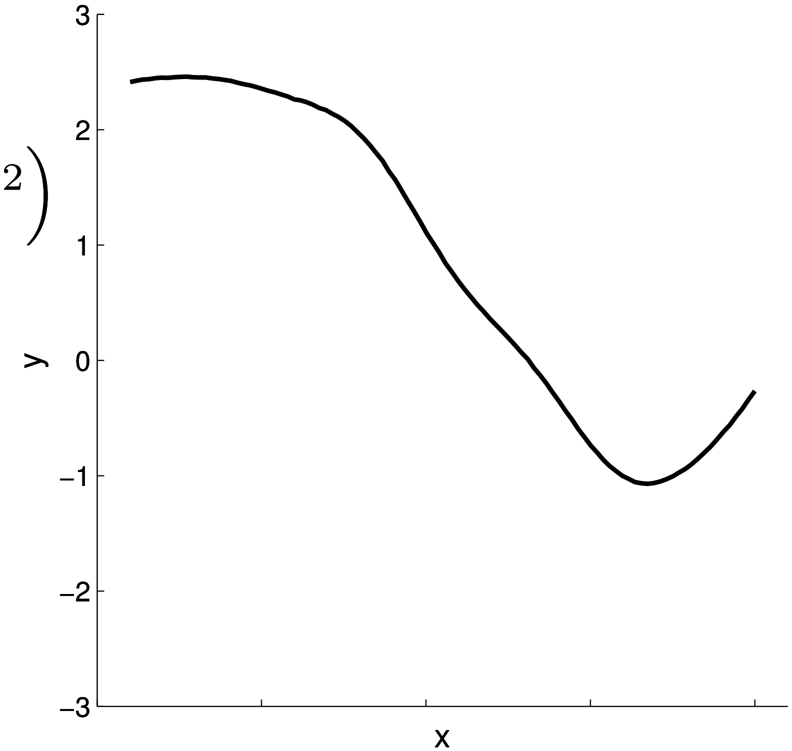
$$p(\mathsf{y}|\theta) = \mathcal{N}(0, \Sigma)$$

$$\Sigma(\mathsf{x}_1, \mathsf{x}_2) = \mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) + \mathsf{I}\sigma_{\mathsf{y}}^2$$

$$\mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) = \sigma^2 \exp\left(-\frac{1}{2\mathsf{l}^2}(\mathsf{x}_1 - \mathsf{x}_2)^2\right)$$

$$\Sigma =$$





Parametric model

$$\mathsf{y}(\mathsf{x}) = f(\mathsf{x};\theta) + \sigma_{\mathsf{y}}\epsilon$$

$$\epsilon \sim \mathcal{N}(0, 1)$$

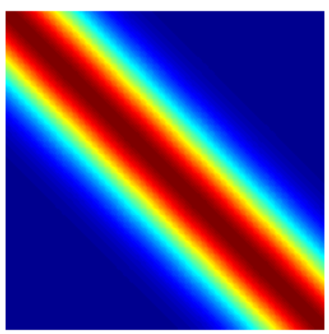
What effect do the hyper-parameters have?

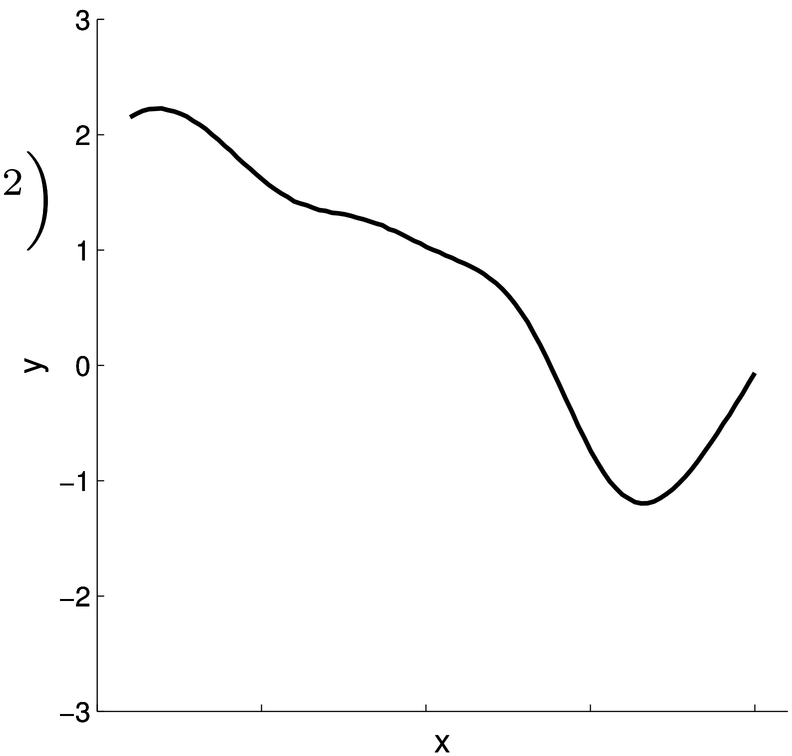
Non-parametric (∞-parametric)

$$p(\mathsf{y}|\theta) = \mathcal{N}(0, \Sigma)$$

$$\Sigma(\mathsf{x}_1, \mathsf{x}_2) = \mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) + \mathsf{I}\sigma_{\mathsf{y}}^2$$

$$\mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) = \sigma^2 \exp\left(-\frac{1}{2\mathsf{l}^2}(\mathsf{x}_1 - \mathsf{x}_2)^2\right)$$

$\Sigma =$ 



Parametric model

$$\mathsf{y}(\mathsf{x}) = f(\mathsf{x}; \theta) + \sigma_{\mathsf{y}}\epsilon$$

$$\epsilon \sim \mathcal{N}(0, 1)$$

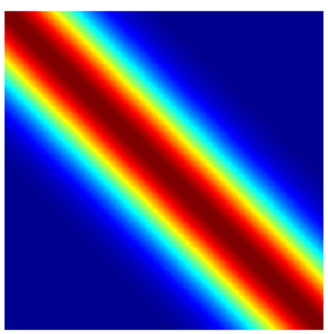
What effect do the hyper-parameters have?

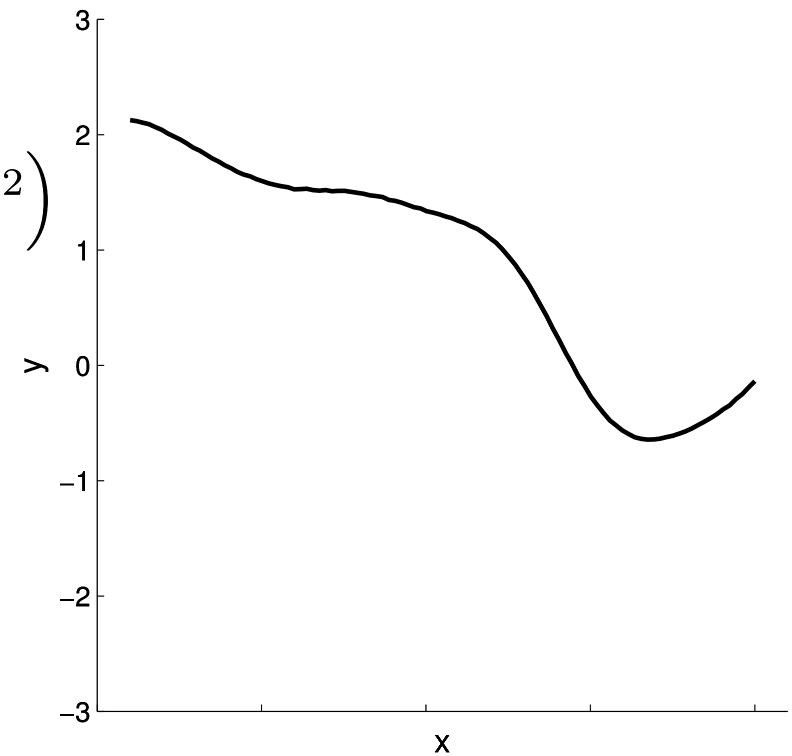
Non-parametric (∞-parametric)

$$p(\mathsf{y}|\theta) = \mathcal{N}(0, \Sigma)$$

$$\Sigma(\mathsf{x}_1, \mathsf{x}_2) = \mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) + \mathsf{I}\sigma_{\mathsf{y}}^2$$

$$\mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) = \sigma^2 \exp\left(-\frac{1}{2\mathsf{l}^2}(\mathsf{x}_1 - \mathsf{x}_2)^2\right)$$

$\Sigma =$ 



Parametric model

$$\mathsf{y}(\mathsf{x}) = f(\mathsf{x}; \theta) + \sigma_{\mathsf{y}}\epsilon$$

$$\epsilon \sim \mathcal{N}(0, 1)$$

What effect do the hyper-parameters have?

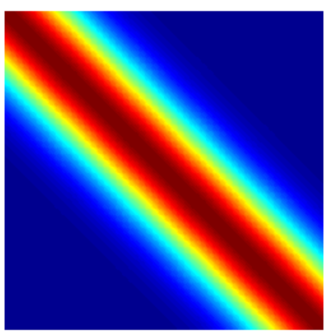
Non-parametric (∞-parametric)

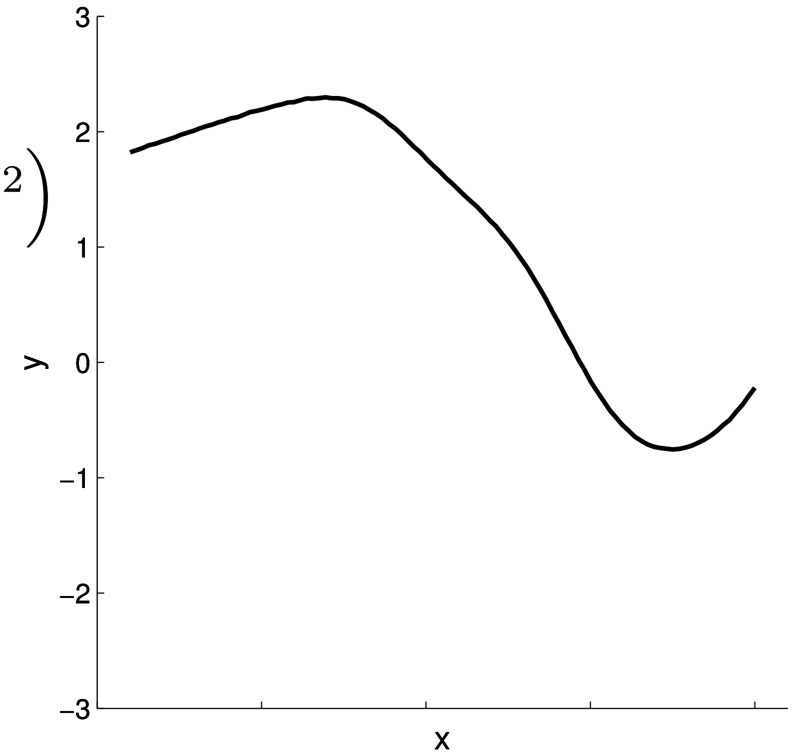
$$p(\mathsf{y}|\theta) = \mathcal{N}(0, \Sigma)$$

$$\Sigma(\mathsf{x}_1, \mathsf{x}_2) = \mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) + \mathsf{I}\sigma_\mathsf{y}^2$$

$$\mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) = \sigma^2 \exp\left(-\frac{1}{2\mathsf{l}^2}(\mathsf{x}_1 - \mathsf{x}_2)^2\right)$$

$$\Sigma =$$





Parametric model

$$\mathsf{y}(\mathsf{x}) = f(\mathsf{x}; \theta) + \sigma_\mathsf{y}\epsilon$$

$$\epsilon \sim \mathcal{N}(0, 1)$$

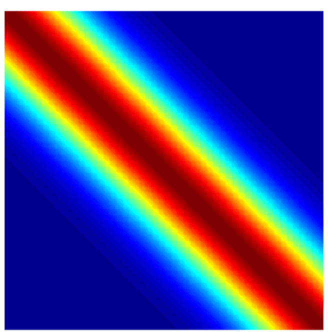
What effect do the hyper-parameters have?

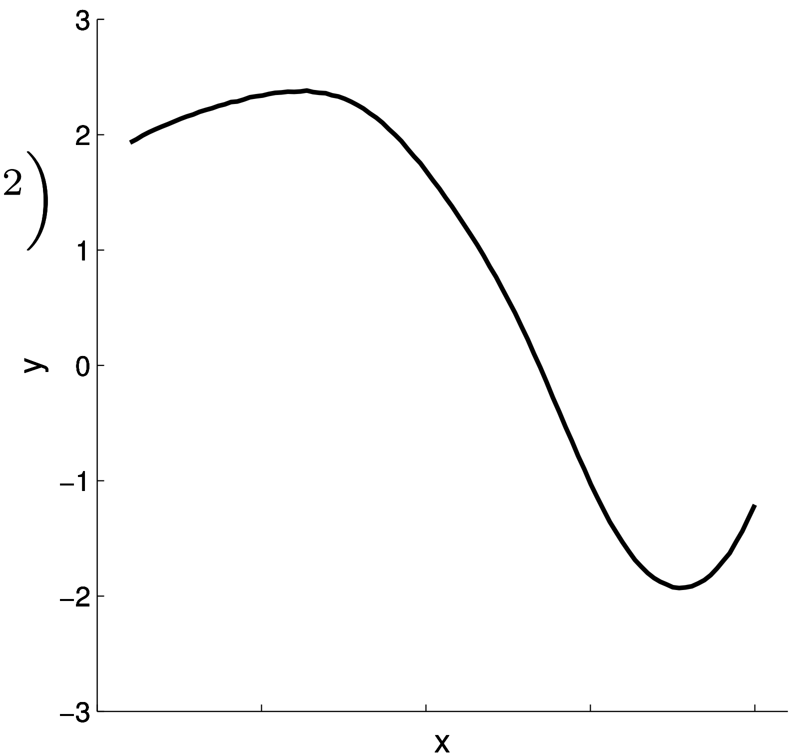
Non-parametric (∞-parametric)

$$p(\mathsf{y}|\theta) = \mathcal{N}(0, \Sigma)$$

$$\Sigma(\mathsf{x}_1, \mathsf{x}_2) = \mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) + \mathsf{I}\sigma_{\mathsf{y}}^2$$

$$\mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) = \sigma^2 \exp\left(-\frac{1}{2\mathsf{l}^2}(\mathsf{x}_1 - \mathsf{x}_2)^2\right)$$

$\Sigma =$ 



Parametric model

$$\mathsf{y}(\mathsf{x}) = f(\mathsf{x}; \theta) + \sigma_{\mathsf{y}}\epsilon$$

$$\epsilon \sim \mathcal{N}(0, 1)$$

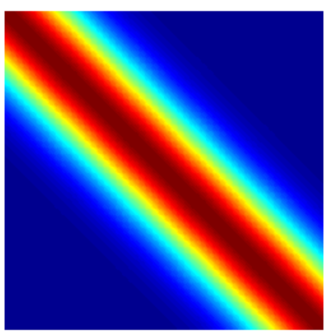
What effect do the hyper-parameters have?

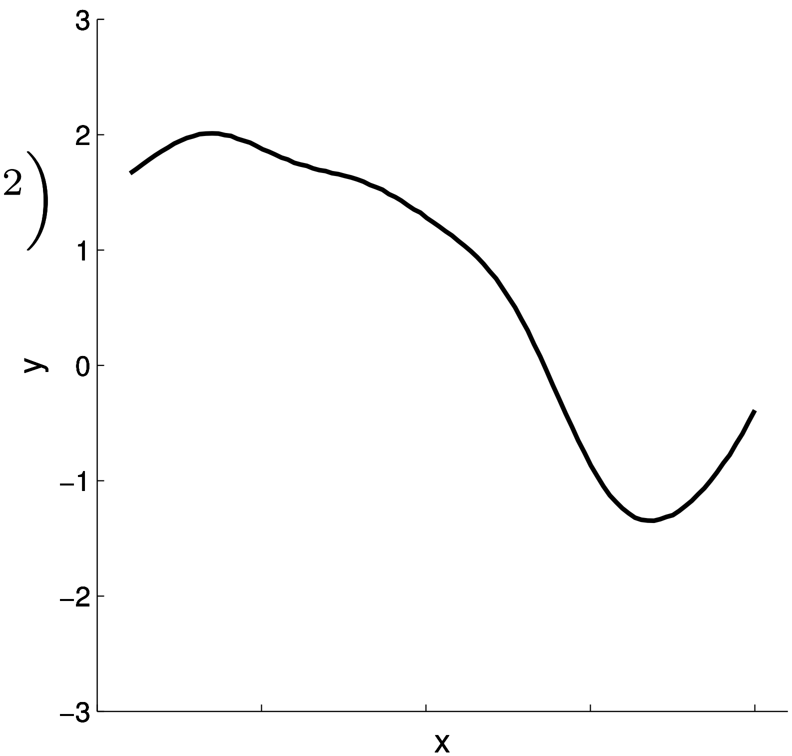
Non-parametric (∞-parametric)

$$p(\mathsf{y}|\theta) = \mathcal{N}(0, \Sigma)$$

$$\Sigma(\mathsf{x}_1, \mathsf{x}_2) = \mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) + \mathsf{I}\sigma_{\mathsf{y}}^2$$

$$\mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) = \sigma^2 \exp\left(-\frac{1}{2\mathsf{l}^2}(\mathsf{x}_1 - \mathsf{x}_2)^2\right)$$

$$\Sigma =$$ 



Parametric model

$$\mathsf{y}(\mathsf{x}) = f(\mathsf{x}; \theta) + \sigma_{\mathsf{y}}\epsilon$$

$$\epsilon \sim \mathcal{N}(0, 1)$$

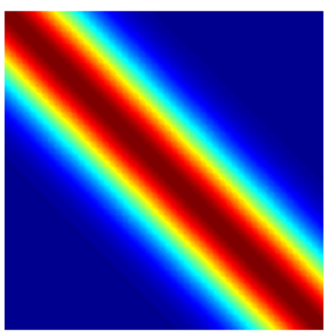
What effect do the hyper-parameters have?

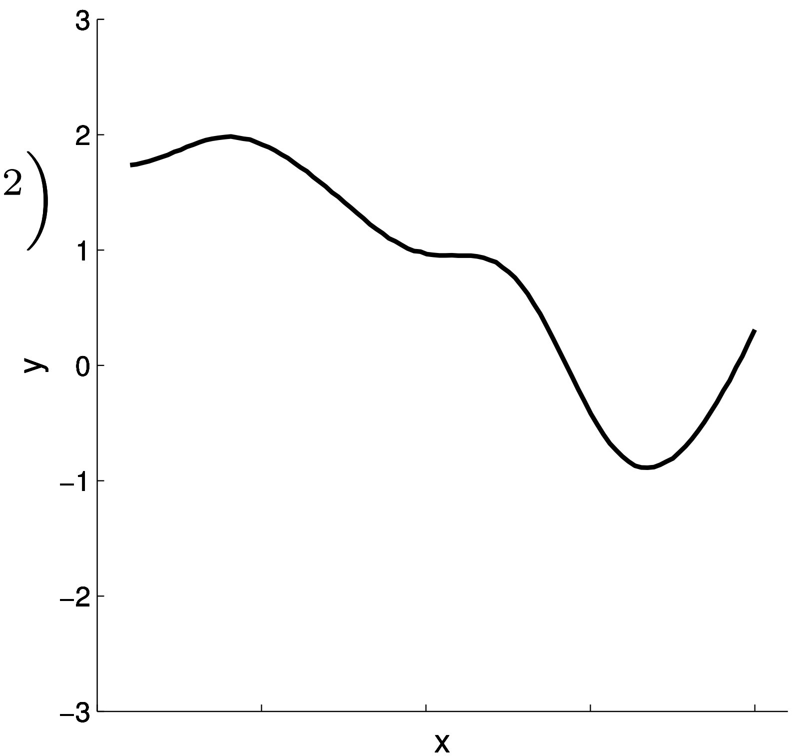
Non-parametric (∞-parametric)

$$p(\mathsf{y}|\theta) = \mathcal{N}(0, \Sigma)$$

$$\Sigma(\mathsf{x}_1, \mathsf{x}_2) = \mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) + \mathsf{I}\sigma_{\mathsf{y}}^2$$

$$\mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) = \sigma^2 \exp\left(-\frac{1}{2\mathsf{l}^2}(\mathsf{x}_1 - \mathsf{x}_2)^2\right)$$

$\Sigma =$ 



Parametric model

$$\mathsf{y}(\mathsf{x}) = f(\mathsf{x};\theta) + \sigma_{\mathsf{y}}\epsilon$$

$$\epsilon \sim \mathcal{N}(0, 1)$$

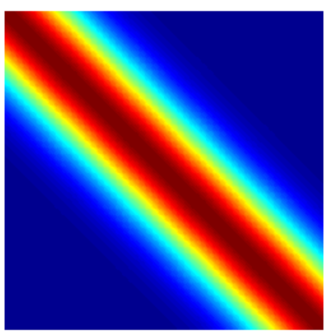
What effect do the hyper-parameters have?

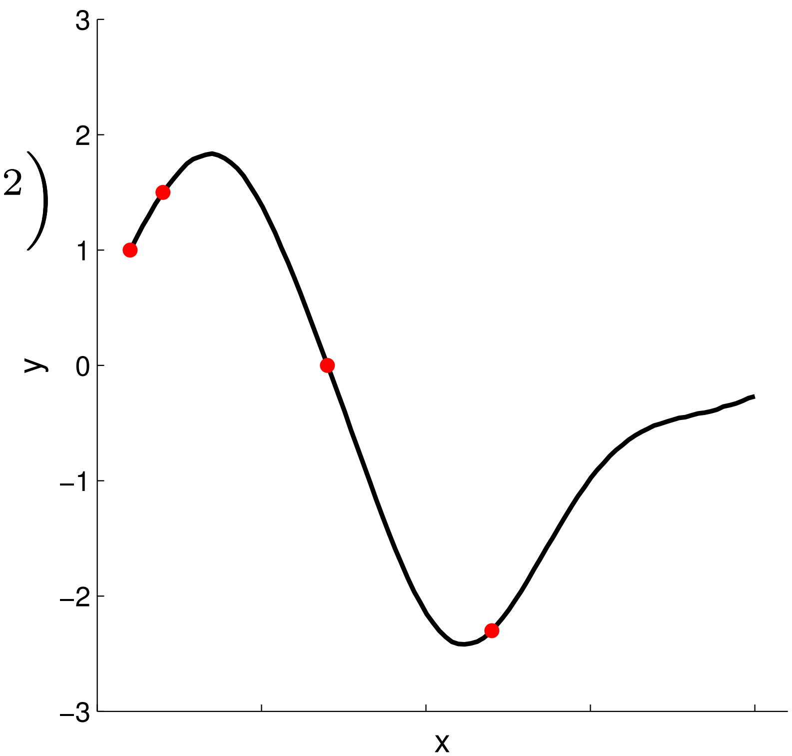
Non-parametric (∞-parametric)

$$p(\mathsf{y}|\theta) = \mathcal{N}(0, \Sigma)$$

$$\Sigma(\mathsf{x}_1, \mathsf{x}_2) = \mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) + \mathsf{I}\sigma_\mathsf{y}^2$$

$$\mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) = \sigma^2 \exp\left(-\frac{1}{2\mathsf{l}^2}(\mathsf{x}_1 - \mathsf{x}_2)^2\right)$$

$\Sigma =$ 



Parametric model

$$\mathsf{y}(\mathsf{x}) = f(\mathsf{x}; \theta) + \sigma_\mathsf{y}\epsilon$$

$$\epsilon \sim \mathcal{N}(0, 1)$$

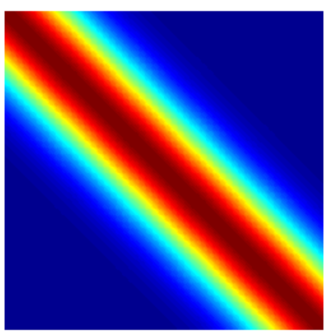
What effect do the hyper-parameters have?

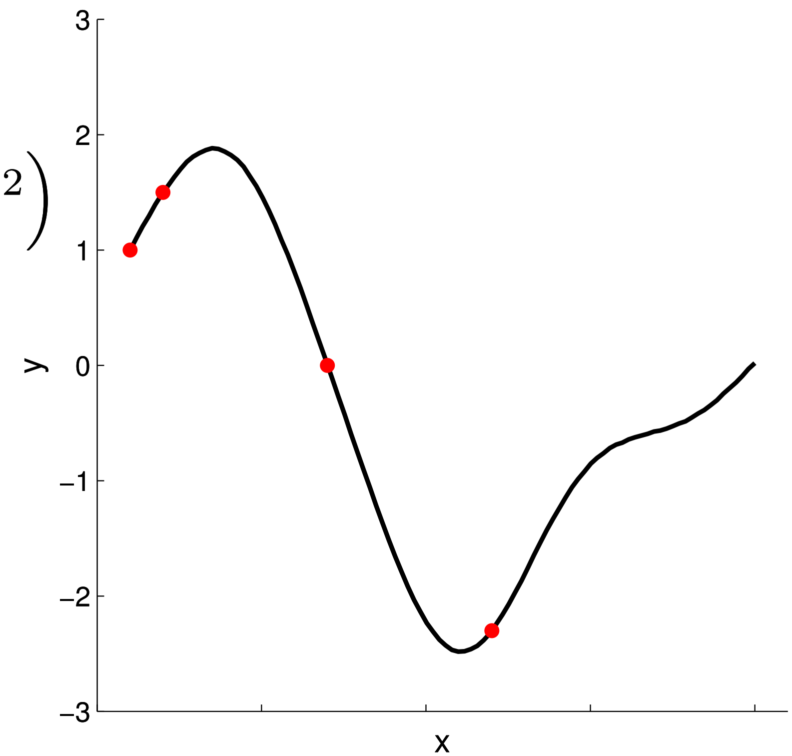
Non-parametric (∞-parametric)

$$p(\mathsf{y}|\theta) = \mathcal{N}(0, \Sigma)$$

$$\Sigma(\mathsf{x}_1, \mathsf{x}_2) = \mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) + \mathsf{I}\sigma_\mathsf{y}^2$$

$$\mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) = \sigma^2 \exp\left(-\frac{1}{2\mathsf{l}^2}(\mathsf{x}_1 - \mathsf{x}_2)^2\right)$$

$\Sigma =$ 



Parametric model

$$\mathsf{y}(\mathsf{x}) = f(\mathsf{x}; \theta) + \sigma_\mathsf{y}\epsilon$$

$$\epsilon \sim \mathcal{N}(0, 1)$$

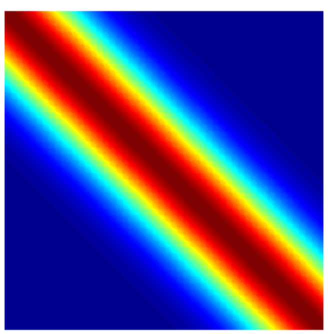
What effect do the hyper-parameters have?

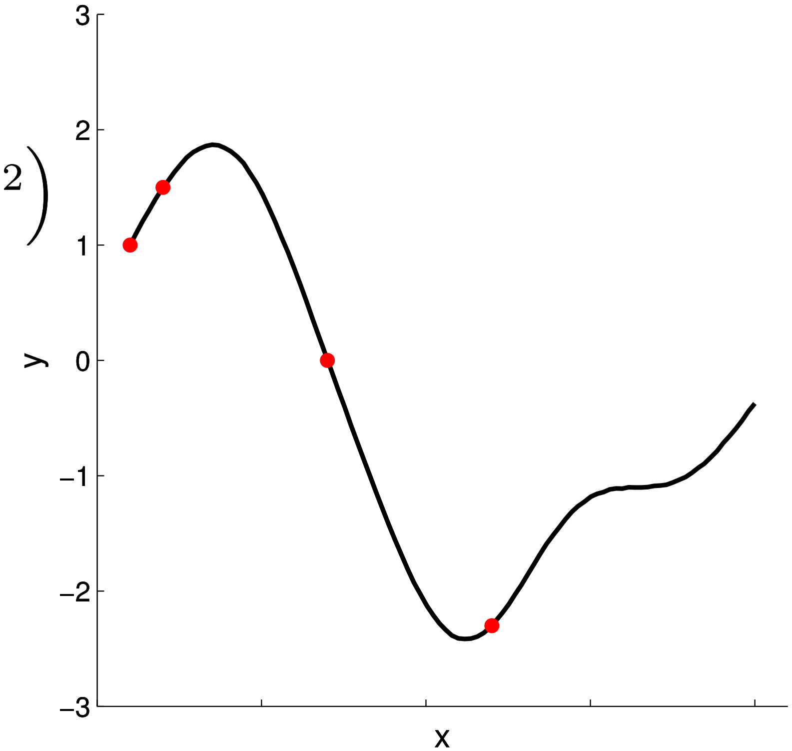
Non-parametric (∞-parametric)

$$p(\mathsf{y}|\theta) = \mathcal{N}(0, \Sigma)$$

$$\Sigma(\mathsf{x}_1, \mathsf{x}_2) = \mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) + \mathsf{I}\sigma_{\mathsf{y}}^2$$

$$\mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) = \sigma^2 \exp\left(-\frac{1}{2\mathsf{l}^2}(\mathsf{x}_1 - \mathsf{x}_2)^2\right)$$

$\Sigma =$ 



Parametric model

$$\mathsf{y}(\mathsf{x}) = f(\mathsf{x}; \theta) + \sigma_{\mathsf{y}}\epsilon$$

$$\epsilon \sim \mathcal{N}(0, 1)$$

What effect do the hyper-parameters have?

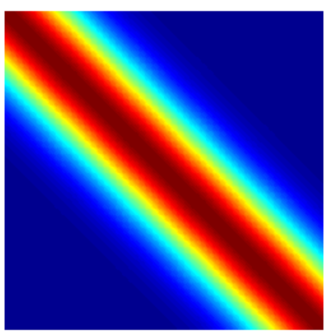
Non-parametric (∞-parametric)

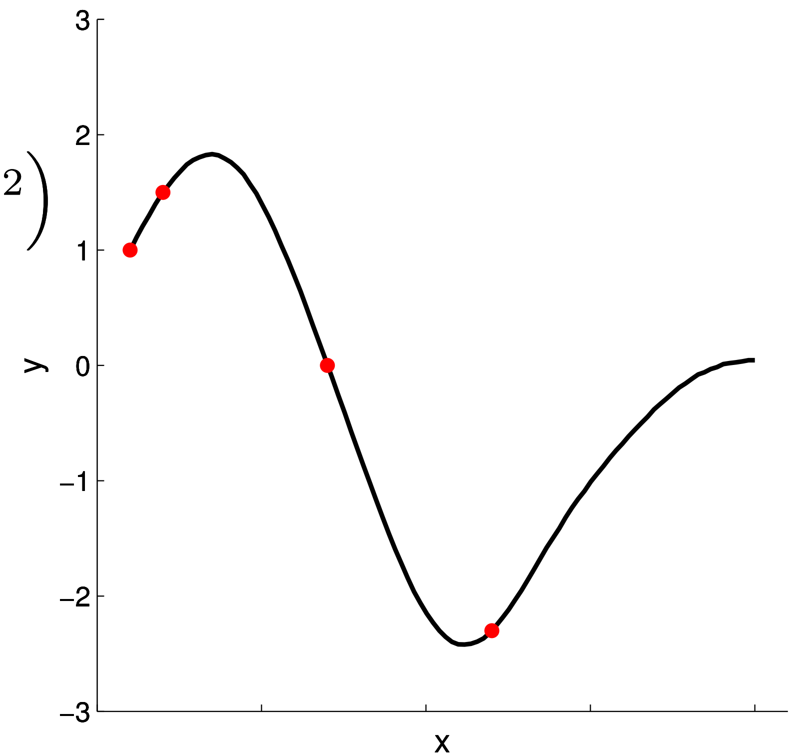
$$p(\mathsf{y}|\theta) = \mathcal{N}(0, \Sigma)$$

$$\Sigma(\mathsf{x}_1, \mathsf{x}_2) = \mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) + \mathsf{I}\sigma_\mathsf{y}^2$$

$$\mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) = \sigma^2 \exp\left(-\frac{1}{2\mathsf{l}^2}(\mathsf{x}_1 - \mathsf{x}_2)^2\right)$$

$$\Sigma =$$





Parametric model

$$\mathsf{y}(\mathsf{x}) = f(\mathsf{x}; \theta) + \sigma_\mathsf{y}\epsilon$$

$$\epsilon \sim \mathcal{N}(0, 1)$$

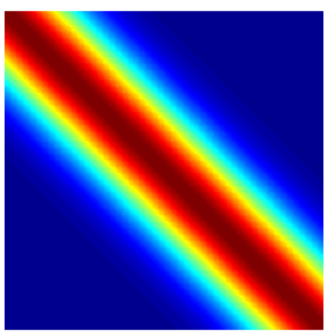
What effect do the hyper-parameters have?

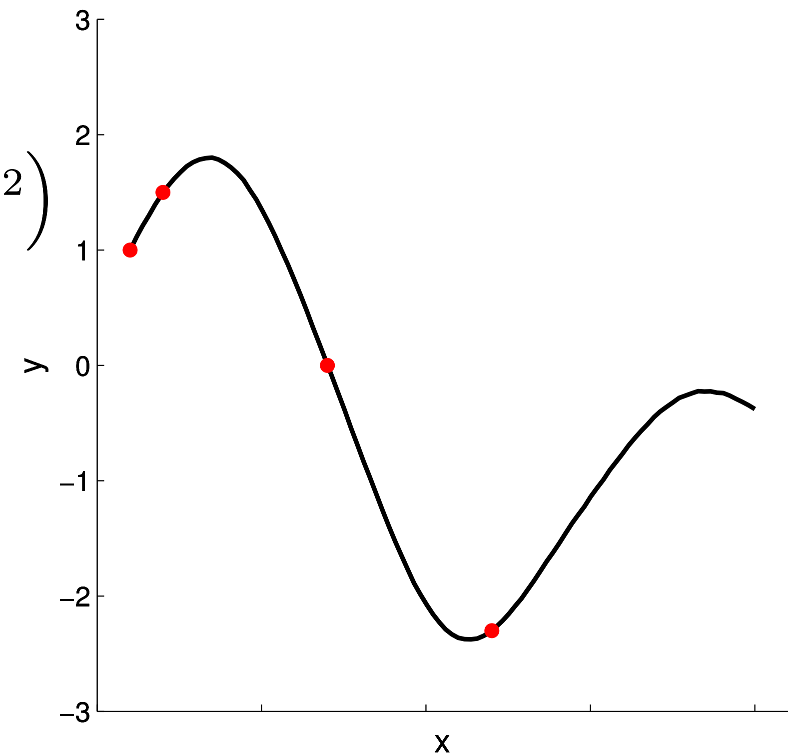
Non-parametric (∞-parametric)

$$p(\mathsf{y}|\theta) = \mathcal{N}(0, \Sigma)$$

$$\Sigma(\mathsf{x}_1, \mathsf{x}_2) = \mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) + \mathsf{I}\sigma_{\mathsf{y}}^2$$

$$\mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) = \sigma^2 \exp\left(-\frac{1}{2\mathsf{l}^2}(\mathsf{x}_1 - \mathsf{x}_2)^2\right)$$

$\Sigma =$ 



Parametric model

$$\mathsf{y}(\mathsf{x}) = f(\mathsf{x}; \theta) + \sigma_{\mathsf{y}}\epsilon$$

$$\epsilon \sim \mathcal{N}(0, 1)$$

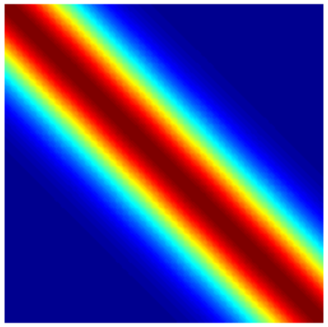
What effect do the hyper-parameters have?

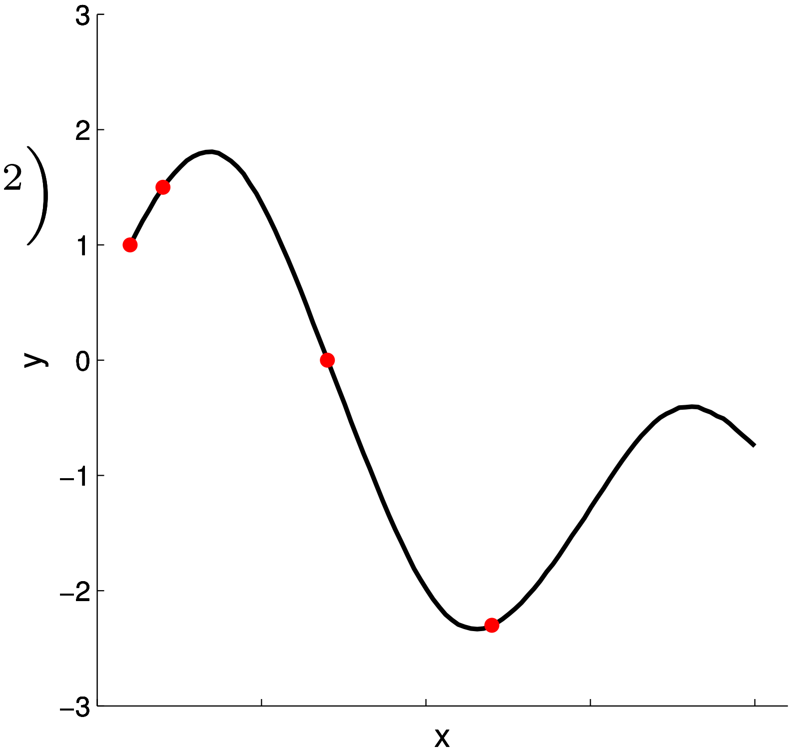
Non-parametric (∞-parametric)

$$p(\mathsf{y}|\theta) = \mathcal{N}(0, \Sigma)$$

$$\Sigma(\mathsf{x}_1, \mathsf{x}_2) = \mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) + \mathsf{I}\sigma_{\mathsf{y}}^2$$

$$\mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) = \sigma^2 \exp\left(-\frac{1}{2\mathsf{l}^2}(\mathsf{x}_1 - \mathsf{x}_2)^2\right)$$

$\Sigma =$ 



Parametric model

$$\mathsf{y}(\mathsf{x}) = f(\mathsf{x}; \theta) + \sigma_{\mathsf{y}}\epsilon$$

$$\epsilon \sim \mathcal{N}(0, 1)$$

What effect do the hyper-parameters have?

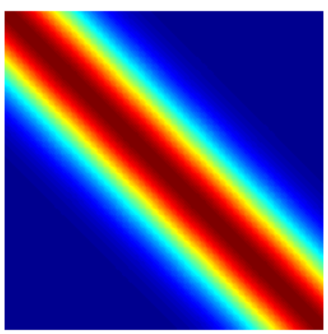
Non-parametric (∞-parametric)

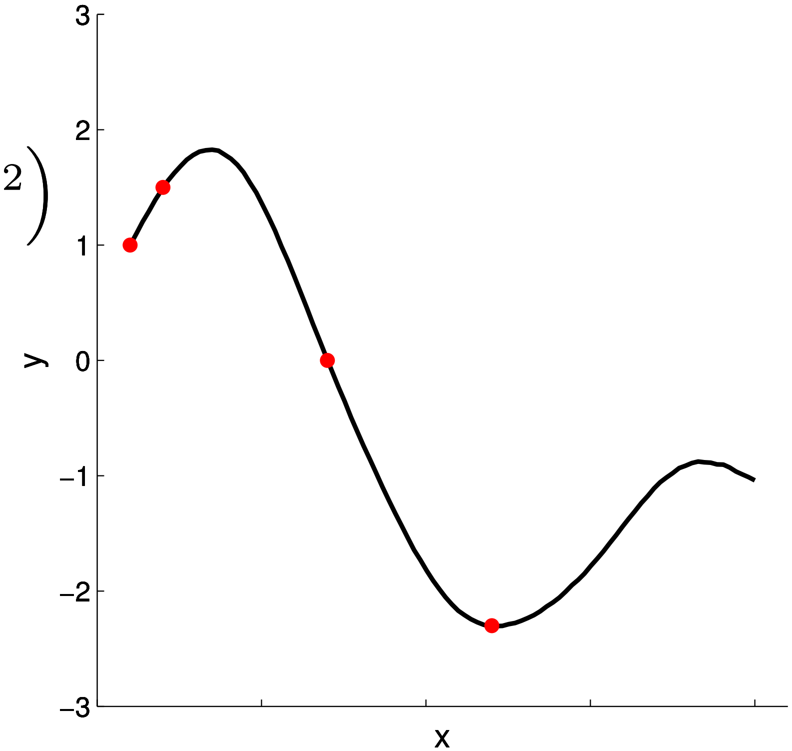
$$p(\mathsf{y}|\theta) = \mathcal{N}(0, \Sigma)$$

$$\Sigma(\mathsf{x}_1, \mathsf{x}_2) = \mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) + \mathsf{I}\sigma_\mathsf{y}^2$$

$$\mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) = \sigma^2 \exp\left(-\frac{1}{2\mathsf{l}^2}(\mathsf{x}_1 - \mathsf{x}_2)^2\right)$$

$$\Sigma =$$





Parametric model

$$\mathsf{y}(\mathsf{x}) = f(\mathsf{x}; \theta) + \sigma_\mathsf{y}\epsilon$$

$$\epsilon \sim \mathcal{N}(0, 1)$$

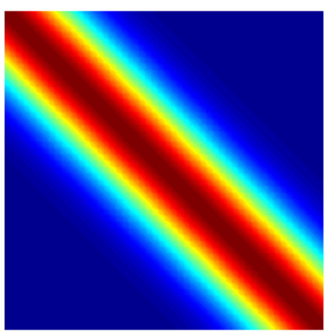
What effect do the hyper-parameters have?

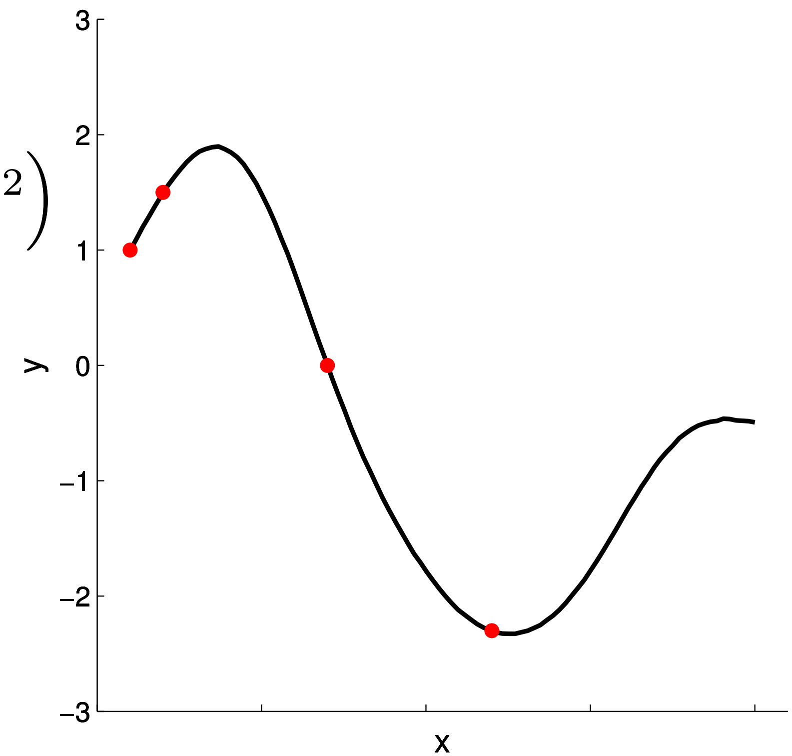
Non-parametric (∞-parametric)

$$p(\mathsf{y}|\theta) = \mathcal{N}(0, \Sigma)$$

$$\Sigma(\mathsf{x}_1, \mathsf{x}_2) = \mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) + \mathsf{I}\sigma_\mathsf{y}^2$$

$$\mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) = \sigma^2 \exp\left(-\frac{1}{2\mathsf{l}^2}(\mathsf{x}_1 - \mathsf{x}_2)^2\right)$$

$\Sigma =$ 



Parametric model

$$\mathsf{y}(\mathsf{x}) = f(\mathsf{x}; \theta) + \sigma_\mathsf{y}\epsilon$$

$$\epsilon \sim \mathcal{N}(0, 1)$$

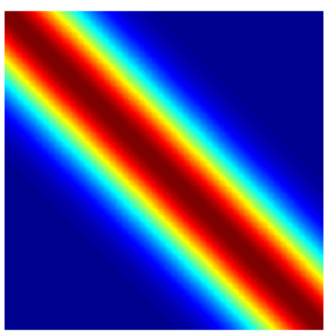
What effect do the hyper-parameters have?

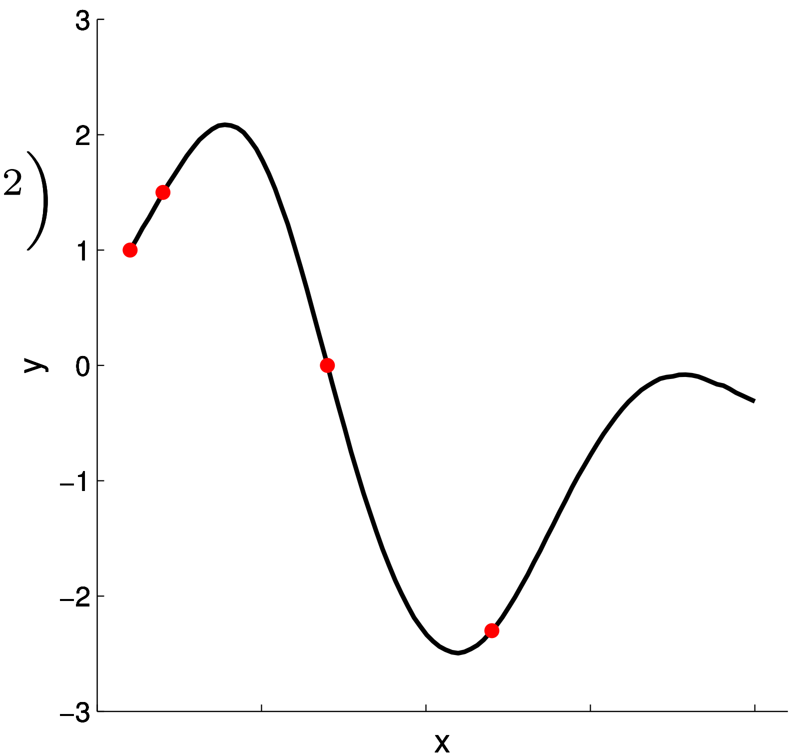
Non-parametric (∞-parametric)

$$p(\mathsf{y}|\theta) = \mathcal{N}(0, \Sigma)$$

$$\Sigma(\mathsf{x}_1, \mathsf{x}_2) = \mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) + \mathsf{I}\sigma_{\mathsf{y}}^2$$

$$\mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) = \sigma^2 \exp\left(-\frac{1}{2\mathsf{l}^2}(\mathsf{x}_1 - \mathsf{x}_2)^2\right)$$

$\Sigma =$ 



Parametric model

$$\mathsf{y}(\mathsf{x}) = f(\mathsf{x}; \theta) + \sigma_{\mathsf{y}}\epsilon$$

$$\epsilon \sim \mathcal{N}(0, 1)$$

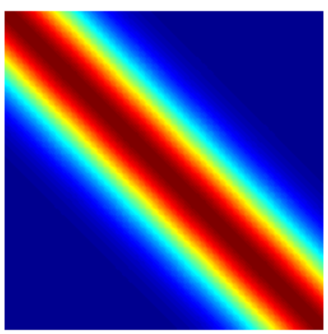
What effect do the hyper-parameters have?

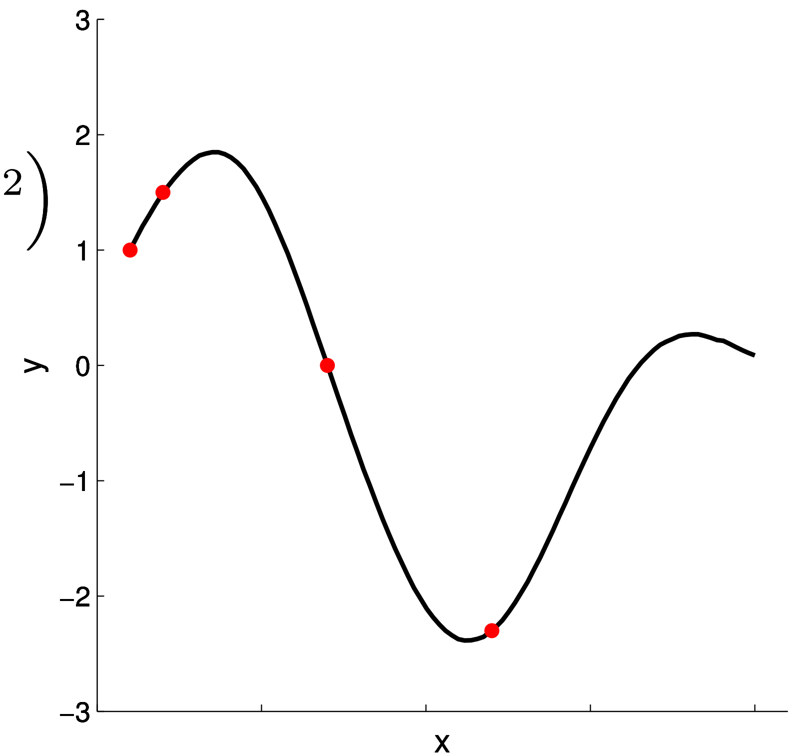
Non-parametric (∞-parametric)

$$p(\mathsf{y}|\theta) = \mathcal{N}(0, \Sigma)$$

$$\Sigma(\mathsf{x}_1, \mathsf{x}_2) = \mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) + \mathsf{I}\sigma_{\mathsf{y}}^2$$

$$\mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) = \sigma^2 \exp\left(-\frac{1}{2\mathsf{l}^2}(\mathsf{x}_1 - \mathsf{x}_2)^2\right)$$

$\Sigma =$ 



Parametric model

$$\mathsf{y}(\mathsf{x}) = f(\mathsf{x};\theta) + \sigma_{\mathsf{y}}\epsilon$$

$$\epsilon \sim \mathcal{N}(0, 1)$$

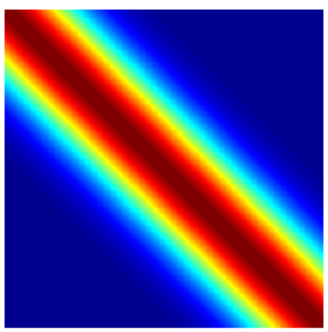
What effect do the hyper-parameters have?

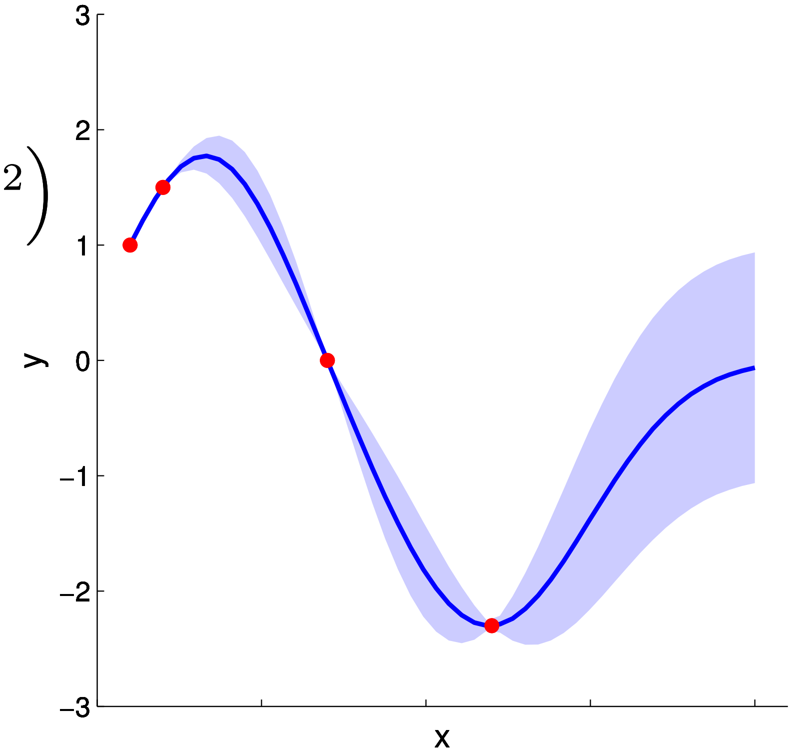
Non-parametric (∞-parametric)

$$p(\mathsf{y}|\theta) = \mathcal{N}(0, \Sigma)$$

$$\Sigma(\mathsf{x}_1, \mathsf{x}_2) = \mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) + \mathsf{I}\sigma_{\mathsf{y}}^2$$

$$\mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) = \sigma^2 \exp\left(-\frac{1}{2\mathsf{l}^2}(\mathsf{x}_1 - \mathsf{x}_2)^2\right)$$

$\Sigma =$ 



Parametric model

$$\mathsf{y}(\mathsf{x}) = f(\mathsf{x};\theta) + \sigma_{\mathsf{y}}\epsilon$$

$$\epsilon \sim \mathcal{N}(0, 1)$$

What effect do the hyper-parameters have?

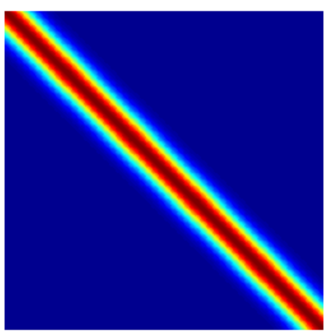
short horizontal length-scale

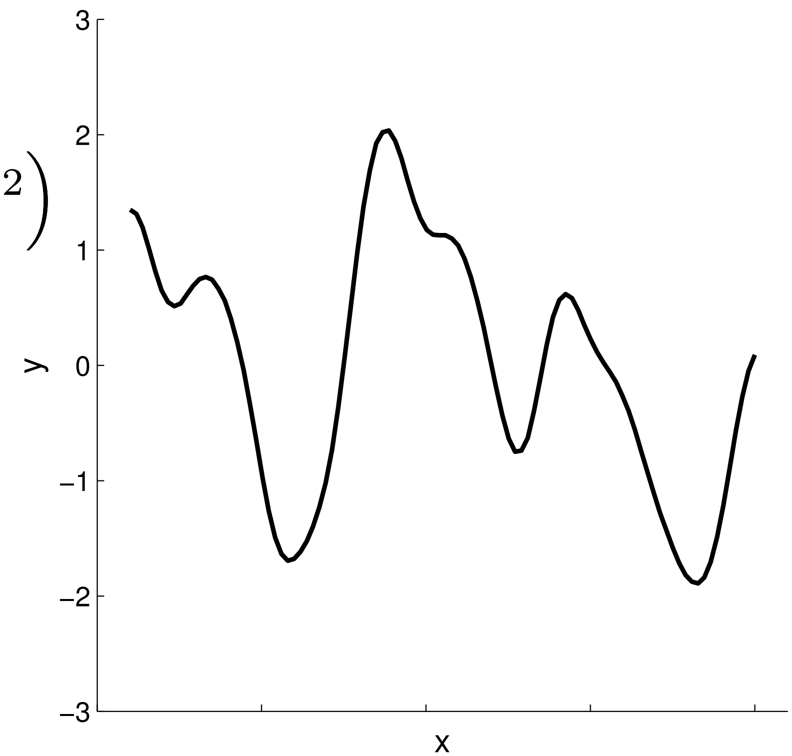
Non-parametric (∞-parametric)

$$p(\mathsf{y}|\theta) = \mathcal{N}(0, \Sigma)$$

$$\Sigma(\mathsf{x}_1, \mathsf{x}_2) = \mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) + \mathsf{I}\sigma_{\mathsf{y}}^2$$

$$\mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) = \sigma^2 \exp\left(-\frac{1}{2\mathsf{l}^2}(\mathsf{x}_1 - \mathsf{x}_2)^2\right)$$

$\Sigma =$ 



Parametric model

$$\mathsf{y}(\mathsf{x}) = f(\mathsf{x}; \theta) + \sigma_{\mathsf{y}}\epsilon$$

$$\epsilon \sim \mathcal{N}(0, 1)$$

What effect do the hyper-parameters have?

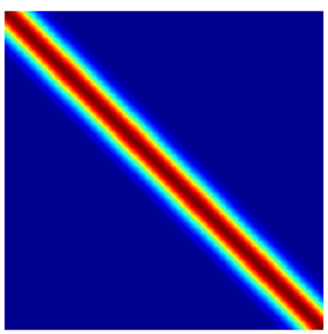
short horizontal length-scale

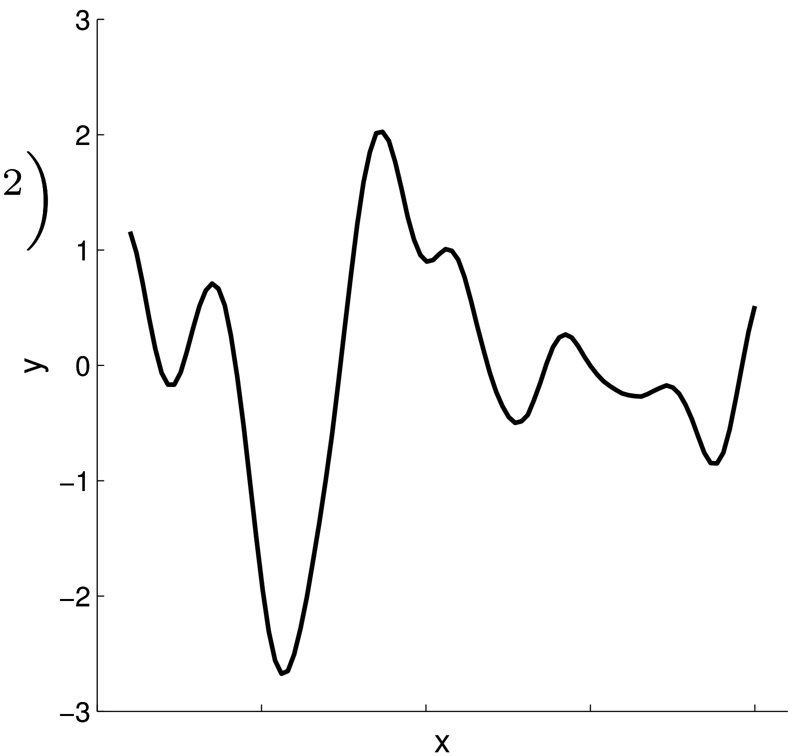
Non-parametric (∞-parametric)

$$p(\mathsf{y}|\theta) = \mathcal{N}(0, \Sigma)$$

$$\Sigma(\mathsf{x}_1, \mathsf{x}_2) = \mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) + \mathsf{I}\sigma_{\mathsf{y}}^2$$

$$\mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) = \sigma^2 \exp\left(-\frac{1}{2\mathsf{l}^2}(\mathsf{x}_1 - \mathsf{x}_2)^2\right)$$

$\Sigma =$ 



Parametric model

$$\mathsf{y}(\mathsf{x}) = f(\mathsf{x}; \theta) + \sigma_{\mathsf{y}}\epsilon$$

$$\epsilon \sim \mathcal{N}(0, 1)$$

What effect do the hyper-parameters have?

short horizontal length-scale

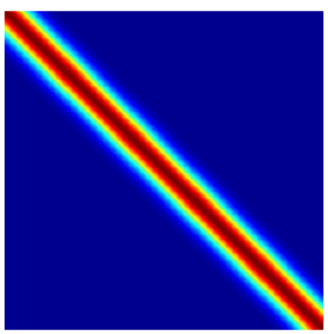
Non-parametric (∞-parametric)

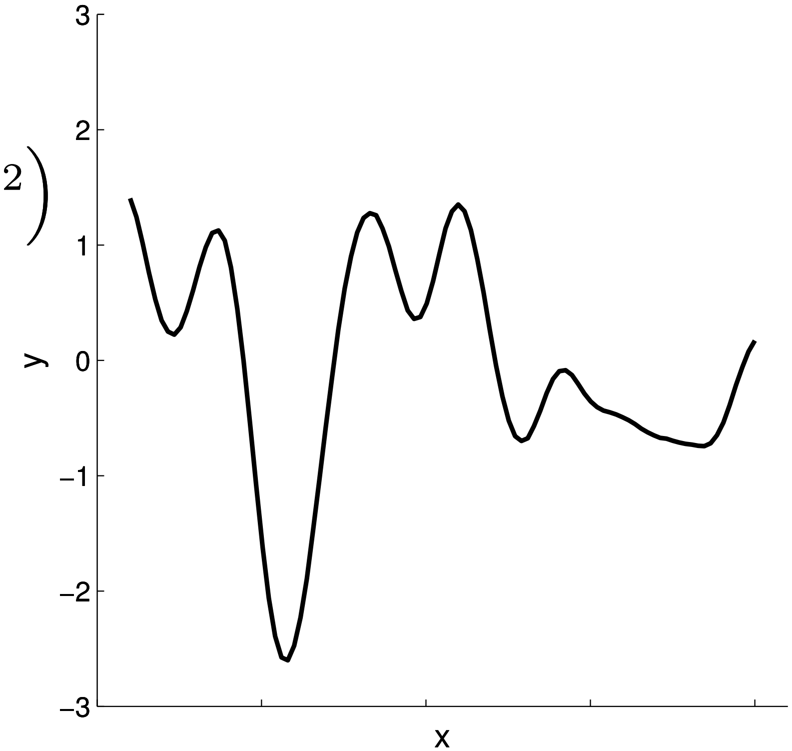
$$p(\mathsf{y}|\theta) = \mathcal{N}(0, \Sigma)$$

$$\Sigma(\mathsf{x}_1, \mathsf{x}_2) = \mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) + \mathsf{I}\sigma_{\mathsf{y}}^2$$

$$\mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) = \sigma^2 \exp\left(-\frac{1}{2\mathsf{l}^2}(\mathsf{x}_1 - \mathsf{x}_2)^2\right)$$

$$\Sigma =$$





Parametric model

$$\mathsf{y}(\mathsf{x}) = f(\mathsf{x};\theta) + \sigma_{\mathsf{y}}\epsilon$$

$$\epsilon \sim \mathcal{N}(0, 1)$$

What effect do the hyper-parameters have?

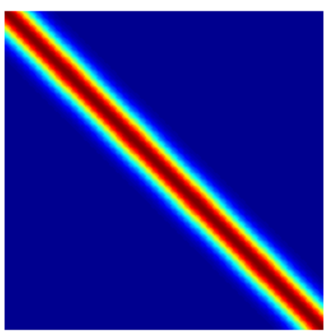
short horizontal length-scale

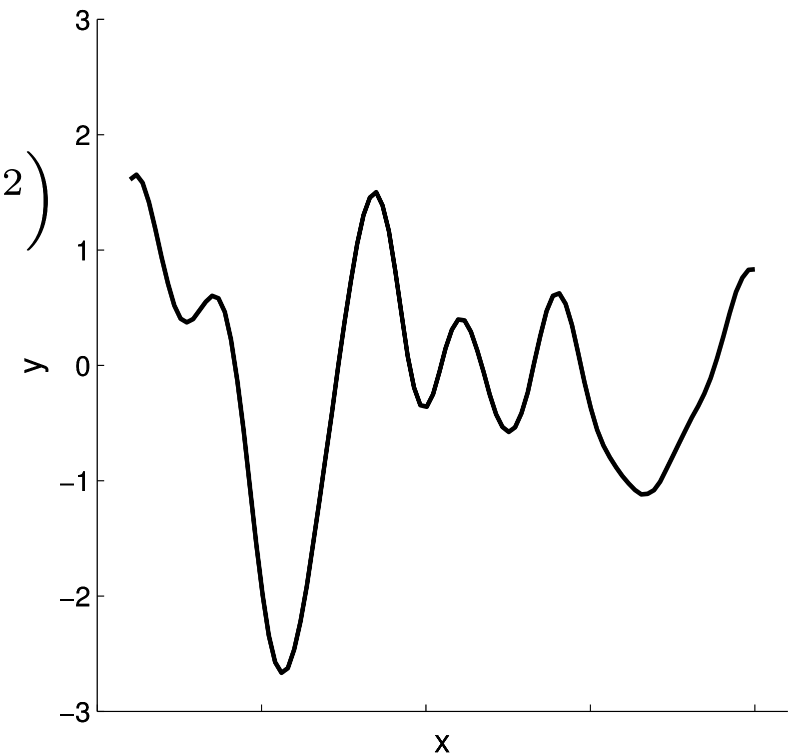
Non-parametric (∞-parametric)

$$p(\mathsf{y}|\theta) = \mathcal{N}(0, \Sigma)$$

$$\Sigma(\mathsf{x}_1, \mathsf{x}_2) = \mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) + \mathsf{I}\sigma_{\mathsf{y}}^2$$

$$\mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) = \sigma^2 \exp\left(-\frac{1}{2\mathsf{l}^2}(\mathsf{x}_1 - \mathsf{x}_2)^2\right)$$

$\Sigma =$ 



Parametric model

$$\mathsf{y}(\mathsf{x}) = f(\mathsf{x}; \theta) + \sigma_{\mathsf{y}}\epsilon$$

$$\epsilon \sim \mathcal{N}(0, 1)$$

What effect do the hyper-parameters have?

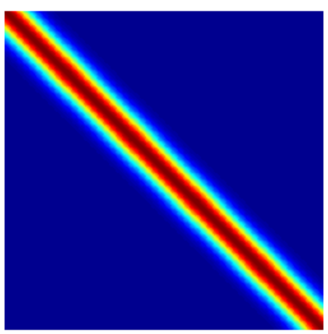
short horizontal length-scale

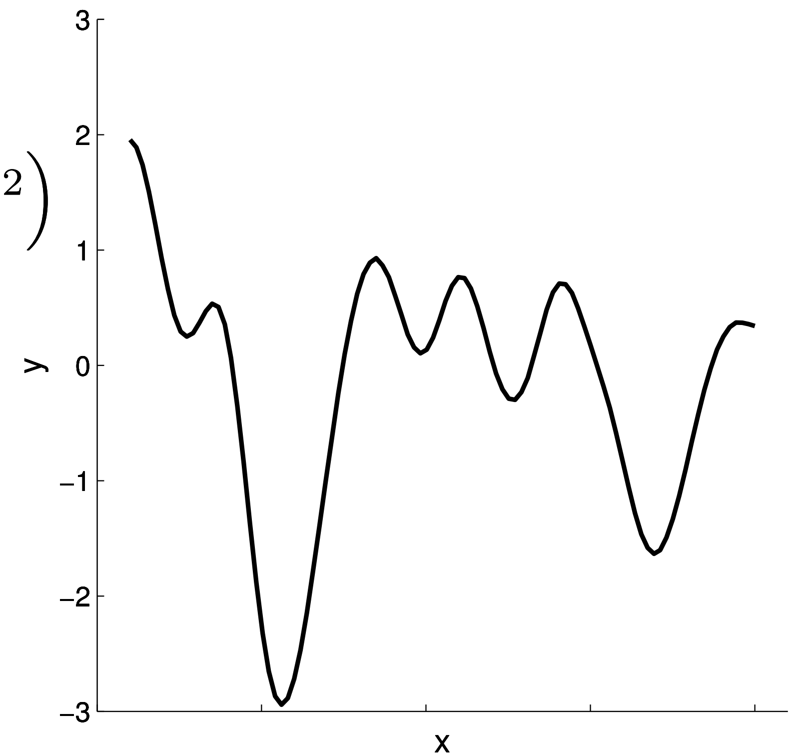
Non-parametric (∞-parametric)

$$p(\mathsf{y}|\theta) = \mathcal{N}(0, \Sigma)$$

$$\Sigma(\mathsf{x}_1, \mathsf{x}_2) = \mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) + \mathsf{I}\sigma_\mathsf{y}^2$$

$$\mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) = \sigma^2 \exp\left(-\frac{1}{2\mathsf{l}^2}(\mathsf{x}_1 - \mathsf{x}_2)^2\right)$$

$\Sigma =$ 



Parametric model

$$\mathsf{y}(\mathsf{x}) = f(\mathsf{x}; \theta) + \sigma_\mathsf{y}\epsilon$$

$$\epsilon \sim \mathcal{N}(0, 1)$$

What effect do the hyper-parameters have?

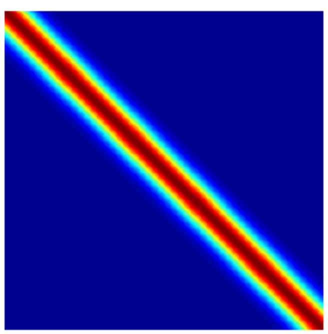
short horizontal length-scale

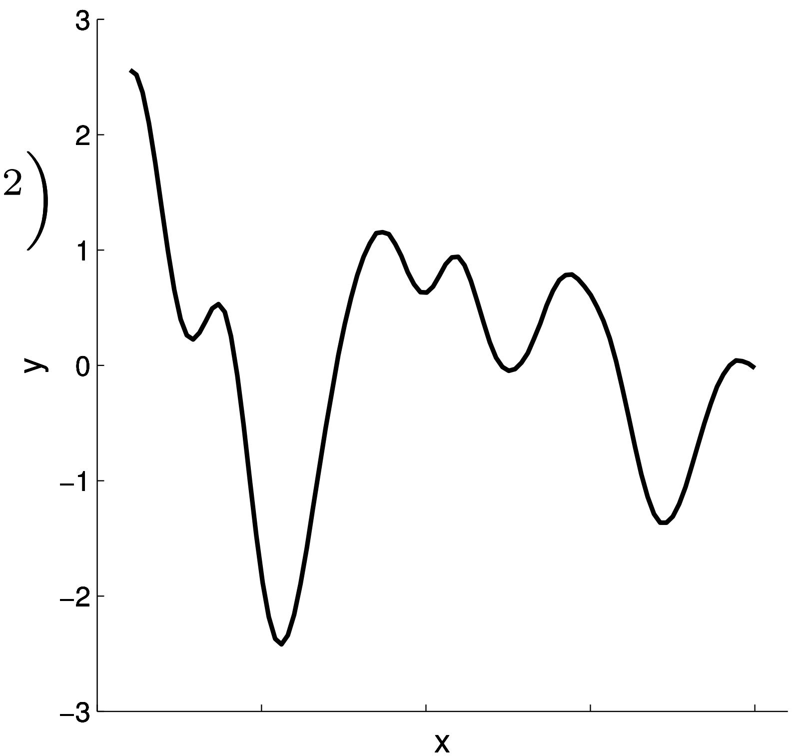
Non-parametric (∞-parametric)

$$p(\mathsf{y}|\theta) = \mathcal{N}(0, \Sigma)$$

$$\Sigma(\mathsf{x}_1, \mathsf{x}_2) = \mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) + \mathsf{I}\sigma_{\mathsf{y}}^2$$

$$\mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) = \sigma^2 \exp\left(-\frac{1}{2\mathsf{l}^2}(\mathsf{x}_1 - \mathsf{x}_2)^2\right)$$

$\Sigma =$ 



Parametric model

$$\mathsf{y}(\mathsf{x}) = f(\mathsf{x}; \theta) + \sigma_{\mathsf{y}}\epsilon$$

$$\epsilon \sim \mathcal{N}(0, 1)$$

What effect do the hyper-parameters have?

short horizontal length-scale

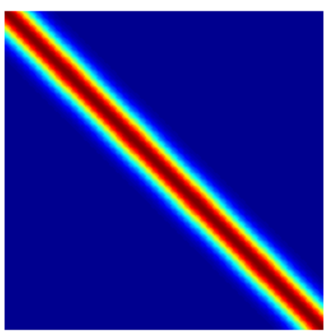
Non-parametric (∞-parametric)

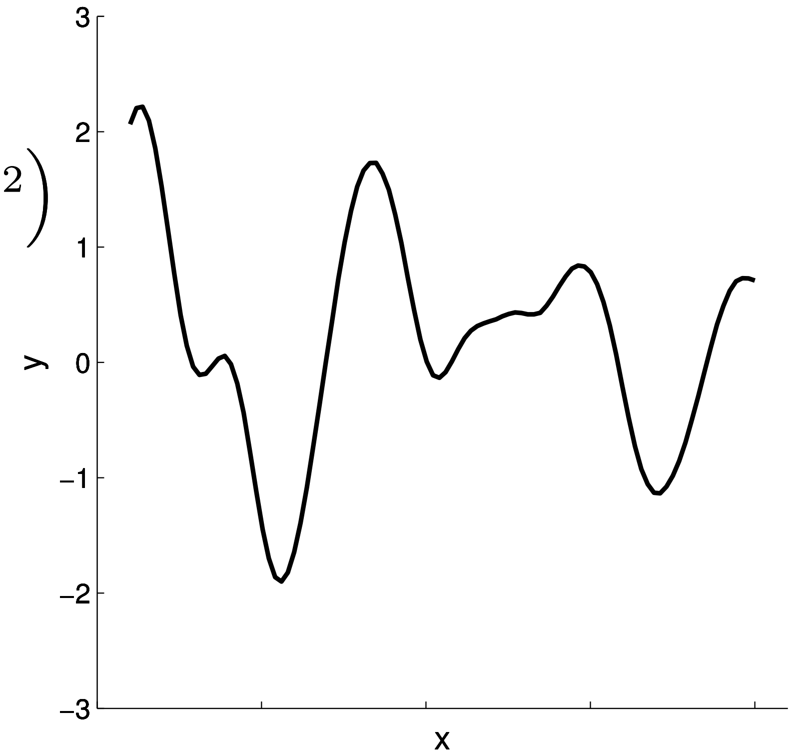
$$p(\mathsf{y}|\theta) = \mathcal{N}(0, \Sigma)$$

$$\Sigma(\mathsf{x}_1, \mathsf{x}_2) = \mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) + \mathsf{I}\sigma_{\mathsf{y}}^2$$

$$\mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) = \sigma^2 \exp\left(-\frac{1}{2\mathsf{l}^2}(\mathsf{x}_1 - \mathsf{x}_2)^2\right)$$

$$\Sigma =$$





Parametric model

$$\mathsf{y}(\mathsf{x}) = f(\mathsf{x};\theta) + \sigma_{\mathsf{y}}\epsilon$$

$$\epsilon \sim \mathcal{N}(0, 1)$$

What effect do the hyper-parameters have?

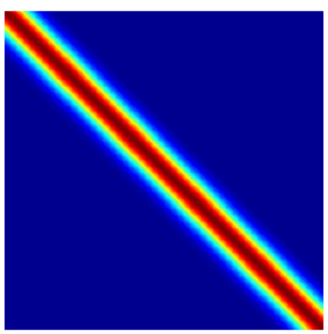
short horizontal length-scale

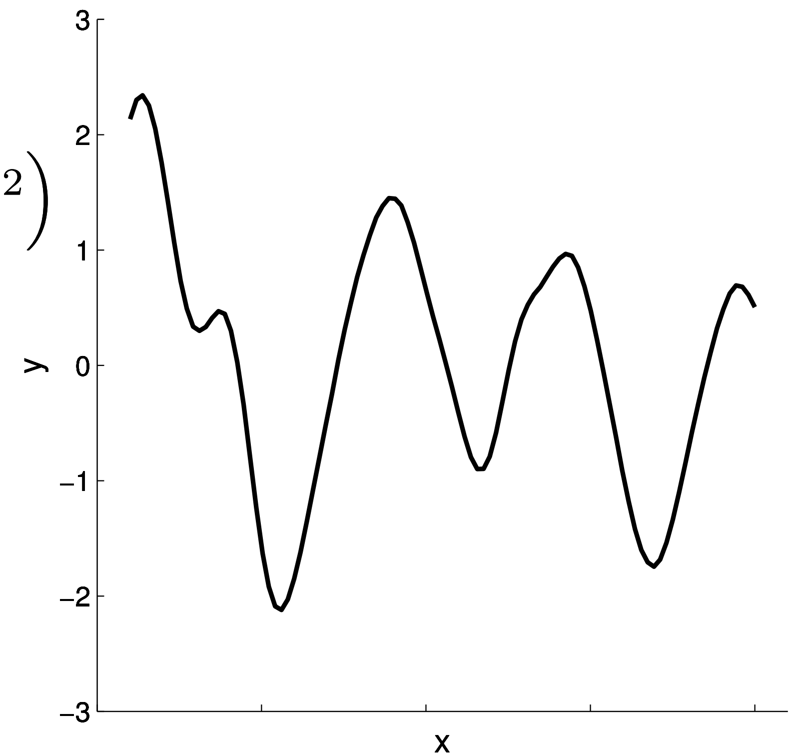
Non-parametric (∞-parametric)

$$p(\mathsf{y}|\theta) = \mathcal{N}(0, \Sigma)$$

$$\Sigma(\mathsf{x}_1, \mathsf{x}_2) = \mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) + \mathsf{I}\sigma_{\mathsf{y}}^2$$

$$\mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) = \sigma^2 \exp\left(-\frac{1}{2\mathsf{l}^2}(\mathsf{x}_1 - \mathsf{x}_2)^2\right)$$

$\Sigma =$ 



Parametric model

$$\mathsf{y}(\mathsf{x}) = f(\mathsf{x}; \theta) + \sigma_{\mathsf{y}}\epsilon$$

$$\epsilon \sim \mathcal{N}(0, 1)$$

What effect do the hyper-parameters have?

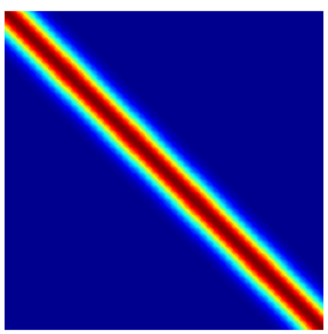
short horizontal length-scale

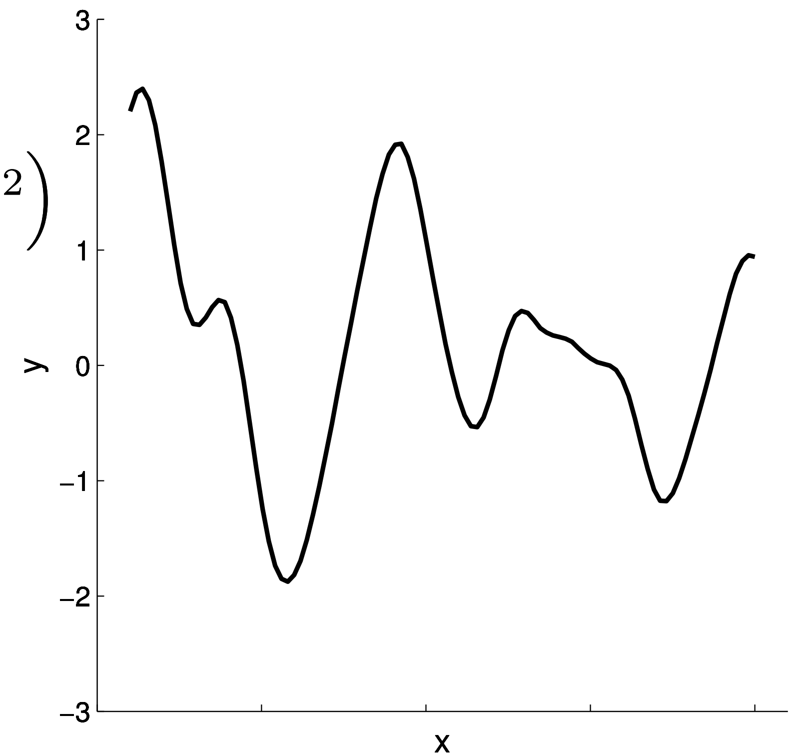
Non-parametric (∞-parametric)

$$p(\mathsf{y}|\theta) = \mathcal{N}(0, \Sigma)$$

$$\Sigma(\mathsf{x}_1, \mathsf{x}_2) = \mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) + \mathsf{I}\sigma_{\mathsf{y}}^2$$

$$\mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) = \sigma^2 \exp\left(-\frac{1}{2\mathsf{l}^2}(\mathsf{x}_1 - \mathsf{x}_2)^2\right)$$

$\Sigma =$ 



Parametric model

$$\mathsf{y}(\mathsf{x}) = f(\mathsf{x}; \theta) + \sigma_{\mathsf{y}}\epsilon$$

$$\epsilon \sim \mathcal{N}(0, 1)$$

What effect do the hyper-parameters have?

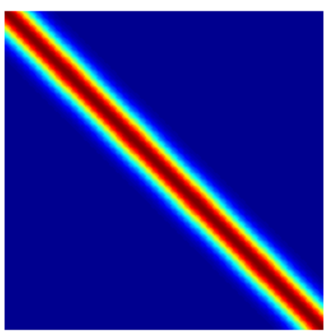
short horizontal length-scale

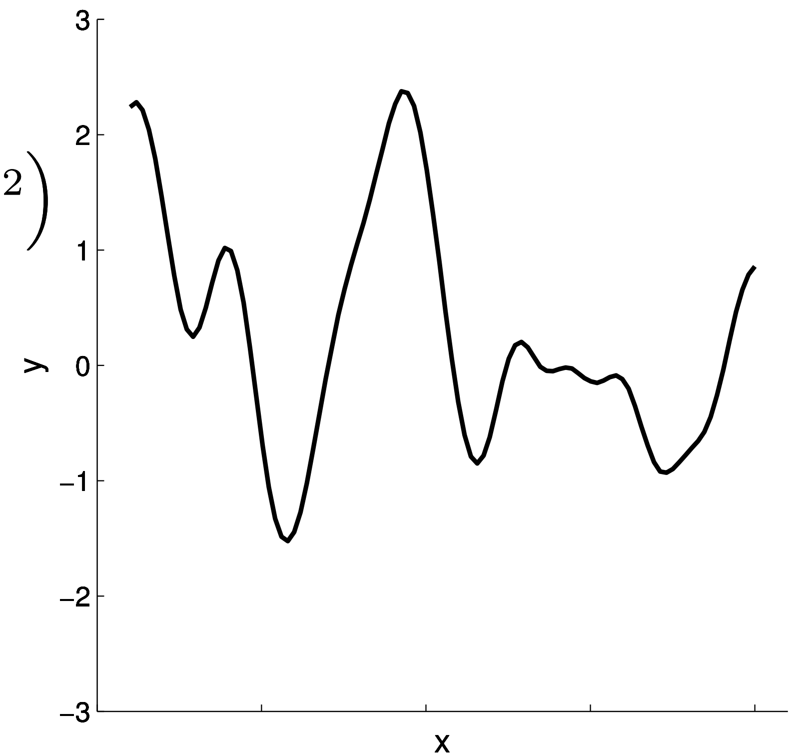
Non-parametric (∞-parametric)

$$p(\mathsf{y}|\theta) = \mathcal{N}(0, \Sigma)$$

$$\Sigma(\mathsf{x}_1, \mathsf{x}_2) = \mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) + \mathsf{I}\sigma_{\mathsf{y}}^2$$

$$\mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) = \sigma^2 \exp\left(-\frac{1}{2\mathsf{l}^2}(\mathsf{x}_1 - \mathsf{x}_2)^2\right)$$

$\Sigma =$ 



Parametric model

$$\mathsf{y}(\mathsf{x}) = f(\mathsf{x}; \theta) + \sigma_{\mathsf{y}}\epsilon$$

$$\epsilon \sim \mathcal{N}(0, 1)$$

What effect do the hyper-parameters have?

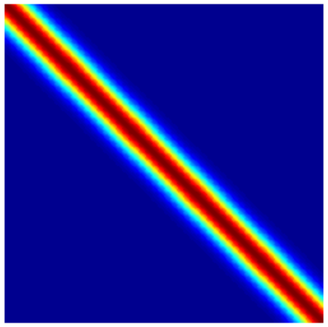
short horizontal length-scale

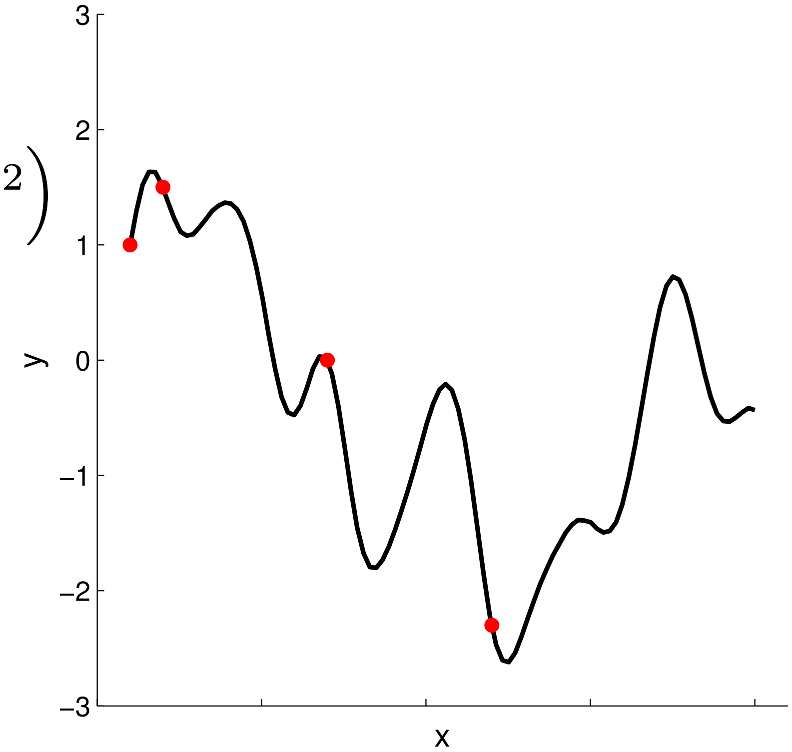
Non-parametric (∞-parametric)

$$p(\mathsf{y}|\theta) = \mathcal{N}(0, \Sigma)$$

$$\Sigma(\mathsf{x}_1, \mathsf{x}_2) = \mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) + \mathsf{I}\sigma_\mathsf{y}^2$$

$$\mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) = \sigma^2 \exp\left(-\frac{1}{2\mathsf{l}^2}(\mathsf{x}_1 - \mathsf{x}_2)^2\right)$$

$\Sigma =$ 



Parametric model

$$\mathsf{y}(\mathsf{x}) = f(\mathsf{x};\theta) + \sigma_\mathsf{y}\epsilon$$

$$\epsilon \sim \mathcal{N}(0, 1)$$

What effect do the hyper-parameters have?

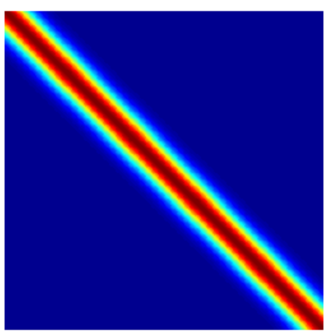
short horizontal length-scale

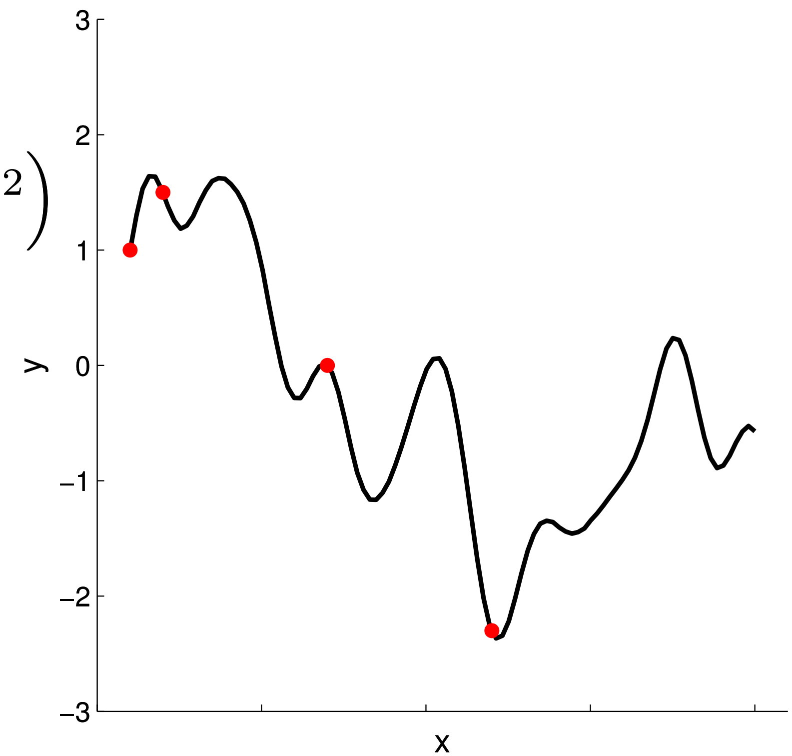
Non-parametric (∞-parametric)

$$p(\mathsf{y}|\theta) = \mathcal{N}(0, \Sigma)$$

$$\Sigma(\mathsf{x}_1, \mathsf{x}_2) = \mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) + \mathsf{I}\sigma_{\mathsf{y}}^2$$

$$\mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) = \sigma^2 \exp\left(-\frac{1}{2\mathsf{l}^2}(\mathsf{x}_1 - \mathsf{x}_2)^2\right)$$

$\Sigma =$ 



Parametric model

$$\mathsf{y}(\mathsf{x}) = f(\mathsf{x};\theta) + \sigma_{\mathsf{y}}\epsilon$$

$$\epsilon \sim \mathcal{N}(0, 1)$$

What effect do the hyper-parameters have?

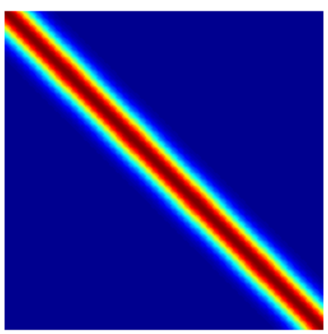
short horizontal length-scale

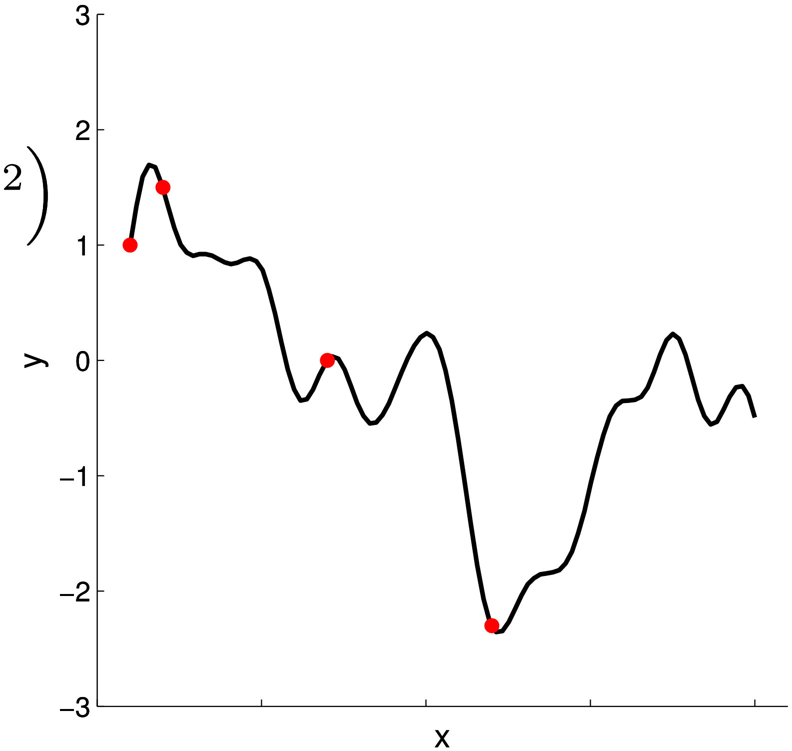
Non-parametric (∞-parametric)

$p(\mathsf{y}|\theta) = \mathcal{N}(0, \Sigma)$

$\Sigma(\mathsf{x}_1, \mathsf{x}_2) = \mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) + \mathsf{I}\sigma_{\mathsf{y}}^2$

$\mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) = \sigma^2 \exp\left(-\frac{1}{2\mathsf{l}^2}(\mathsf{x}_1 - \mathsf{x}_2)^2\right)$

$\Sigma =$ 



Parametric model

$\mathsf{y}(\mathsf{x}) = f(\mathsf{x}; \theta) + \sigma_{\mathsf{y}}\epsilon$

$\epsilon \sim \mathcal{N}(0, 1)$

What effect do the hyper-parameters have?

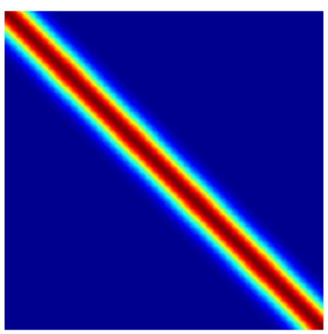
short horizontal length-scale

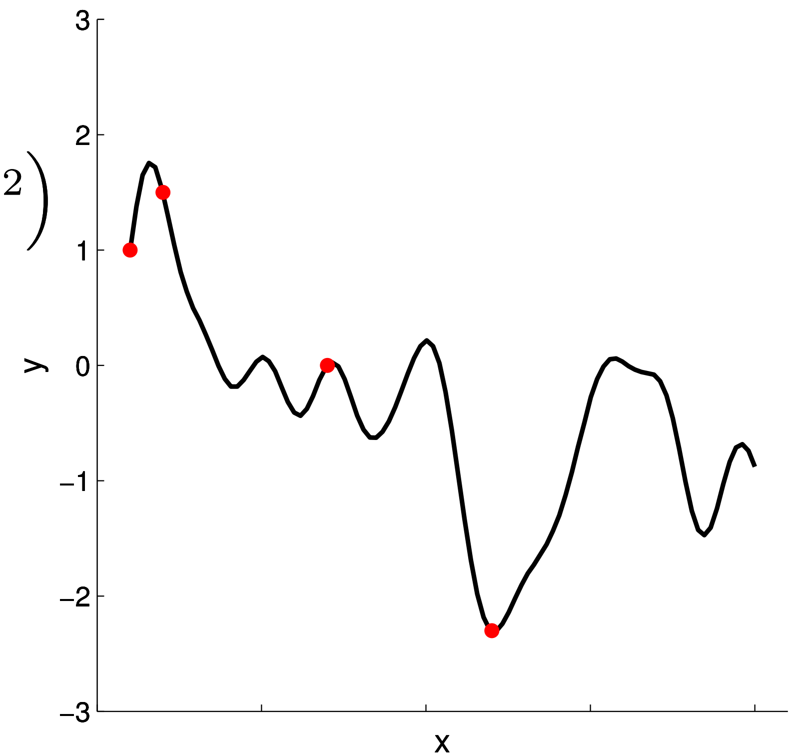
Non-parametric (∞-parametric)

$$p(\mathsf{y}|\theta) = \mathcal{N}(0, \Sigma)$$

$$\Sigma(\mathsf{x}_1, \mathsf{x}_2) = \mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) + \mathsf{I}\sigma_{\mathsf{y}}^2$$

$$\mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) = \sigma^2 \exp\left(-\frac{1}{2\mathsf{l}^2}(\mathsf{x}_1 - \mathsf{x}_2)^2\right)$$

$\Sigma =$ 



Parametric model

$$\mathsf{y}(\mathsf{x}) = f(\mathsf{x}; \theta) + \sigma_{\mathsf{y}}\epsilon$$

$$\epsilon \sim \mathcal{N}(0, 1)$$

What effect do the hyper-parameters have?

short horizontal length-scale

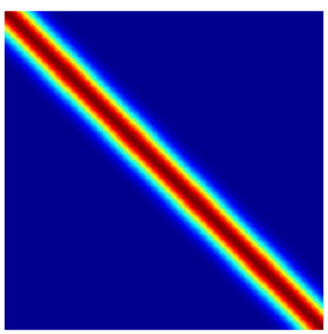
Non-parametric (∞-parametric)

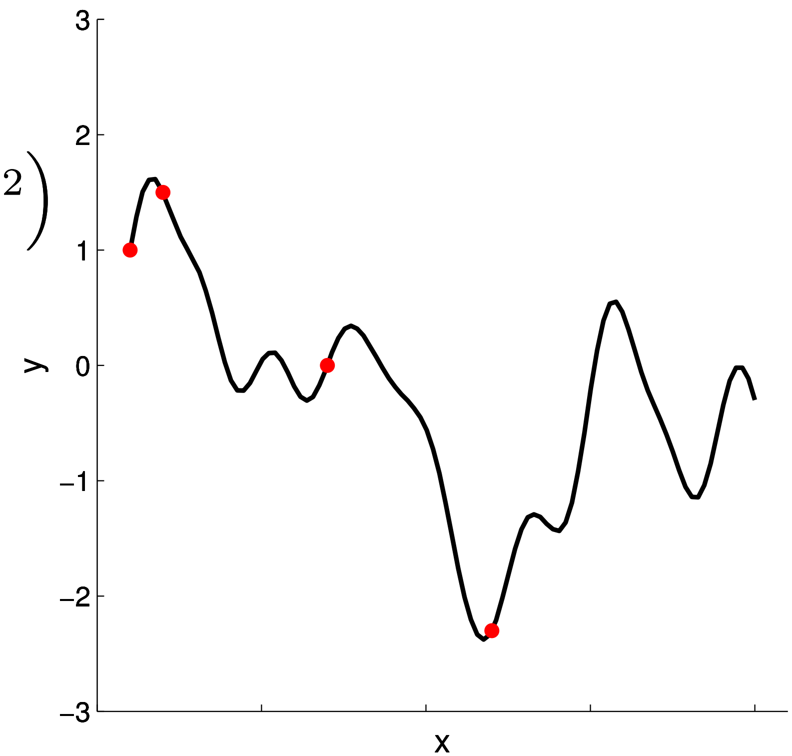
$$p(\mathsf{y}|\theta) = \mathcal{N}(0, \Sigma)$$

$$\Sigma(\mathsf{x}_1, \mathsf{x}_2) = \mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) + \mathsf{I}\sigma_\mathsf{y}^2$$

$$\mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) = \sigma^2 \exp\left(-\frac{1}{2\mathsf{l}^2}(\mathsf{x}_1 - \mathsf{x}_2)^2\right)$$

$$\Sigma =$$





Parametric model

$$\mathsf{y}(\mathsf{x}) = f(\mathsf{x}; \theta) + \sigma_\mathsf{y}\epsilon$$

$$\epsilon \sim \mathcal{N}(0, 1)$$

What effect do the hyper-parameters have?

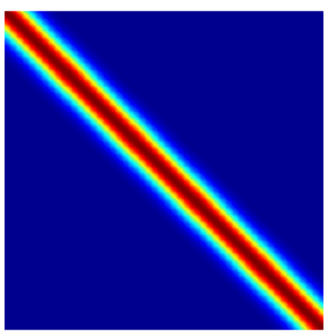
short horizontal length-scale

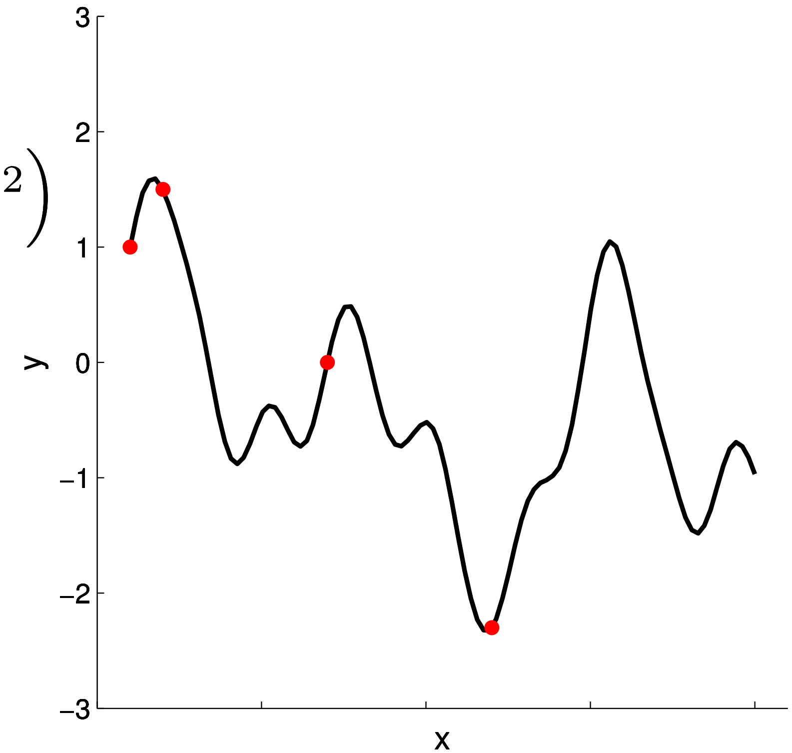
Non-parametric (∞-parametric)

$$p(\mathsf{y}|\theta) = \mathcal{N}(0, \Sigma)$$

$$\Sigma(\mathsf{x}_1, \mathsf{x}_2) = \mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) + \mathsf{I}\sigma_{\mathsf{y}}^2$$

$$\mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) = \sigma^2 \exp\left(-\frac{1}{2\mathsf{l}^2}(\mathsf{x}_1 - \mathsf{x}_2)^2\right)$$

$\Sigma =$ 



Parametric model

$$\mathsf{y}(\mathsf{x}) = f(\mathsf{x}; \theta) + \sigma_{\mathsf{y}}\epsilon$$

$$\epsilon \sim \mathcal{N}(0, 1)$$

What effect do the hyper-parameters have?

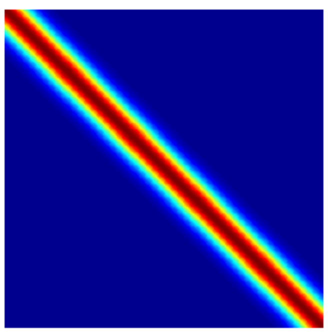
short horizontal length-scale

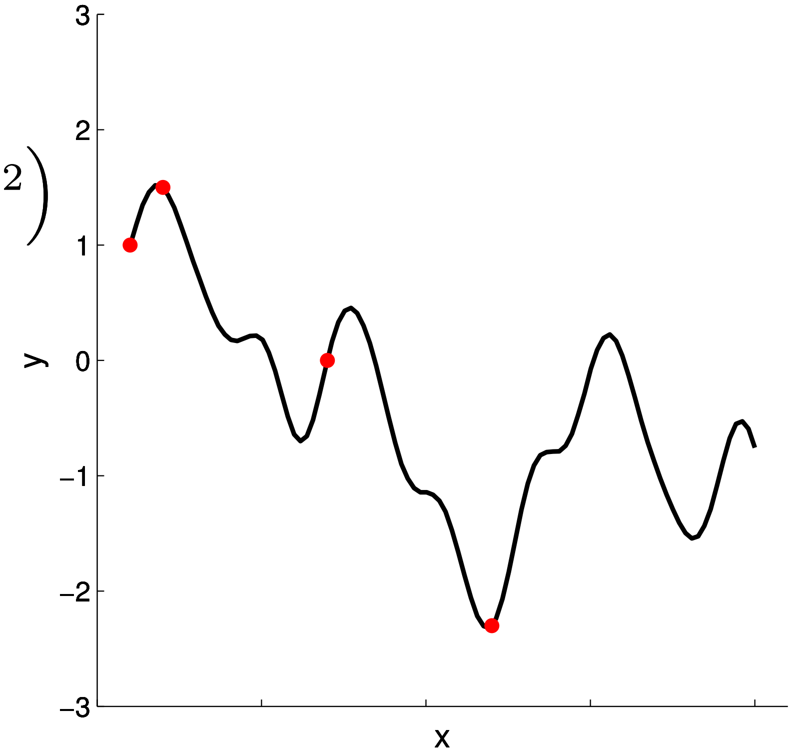
Non-parametric (∞-parametric)

$$p(\mathsf{y}|\theta) = \mathcal{N}(0, \Sigma)$$

$$\Sigma(\mathsf{x}_1, \mathsf{x}_2) = \mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) + \mathsf{I}\sigma_\mathsf{y}^2$$

$$\mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) = \sigma^2 \exp\left(-\frac{1}{2\mathsf{l}^2}(\mathsf{x}_1 - \mathsf{x}_2)^2\right)$$

$\Sigma =$ 



Parametric model

$$\mathsf{y}(\mathsf{x}) = f(\mathsf{x}; \theta) + \sigma_\mathsf{y}\epsilon$$

$$\epsilon \sim \mathcal{N}(0, 1)$$

What effect do the hyper-parameters have?

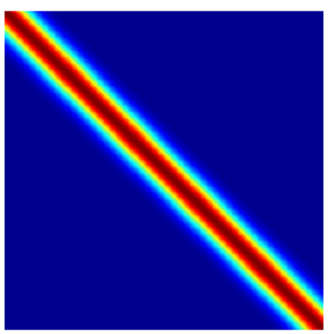
short horizontal length-scale

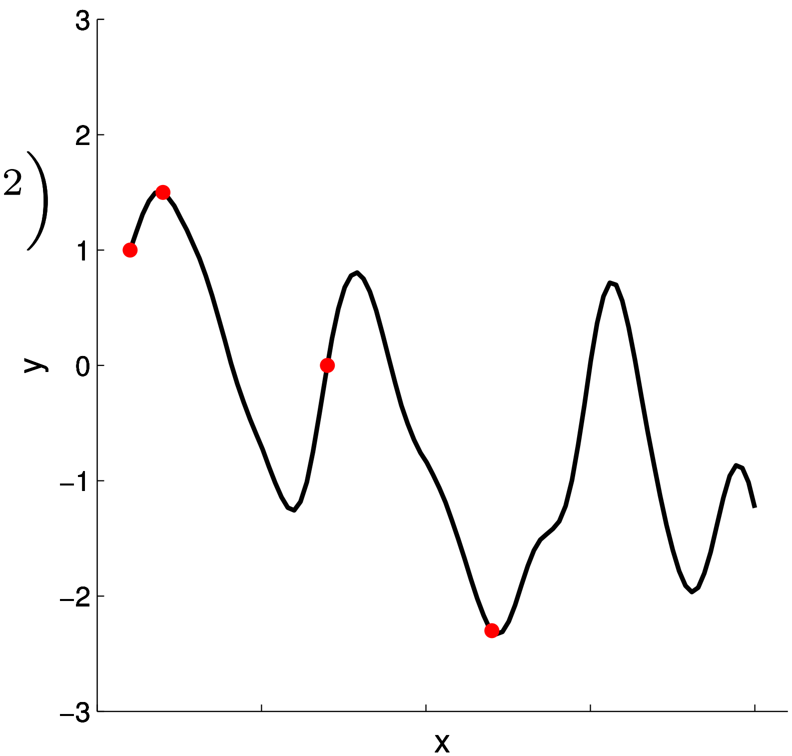
Non-parametric (∞-parametric)

$$p(\mathsf{y}|\theta) = \mathcal{N}(0, \Sigma)$$

$$\Sigma(\mathsf{x}_1, \mathsf{x}_2) = \mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) + \mathsf{I}\sigma_{\mathsf{y}}^2$$

$$\mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) = \sigma^2 \exp\left(-\frac{1}{2\mathsf{l}^2}(\mathsf{x}_1 - \mathsf{x}_2)^2\right)$$

$\Sigma =$ 



Parametric model

$$\mathsf{y}(\mathsf{x}) = f(\mathsf{x}; \theta) + \sigma_{\mathsf{y}}\epsilon$$

$$\epsilon \sim \mathcal{N}(0, 1)$$

What effect do the hyper-parameters have?

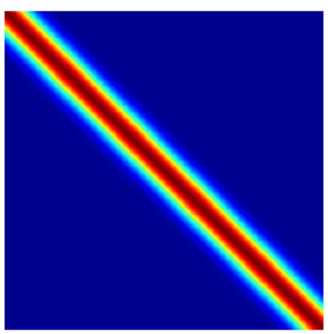
short horizontal length-scale

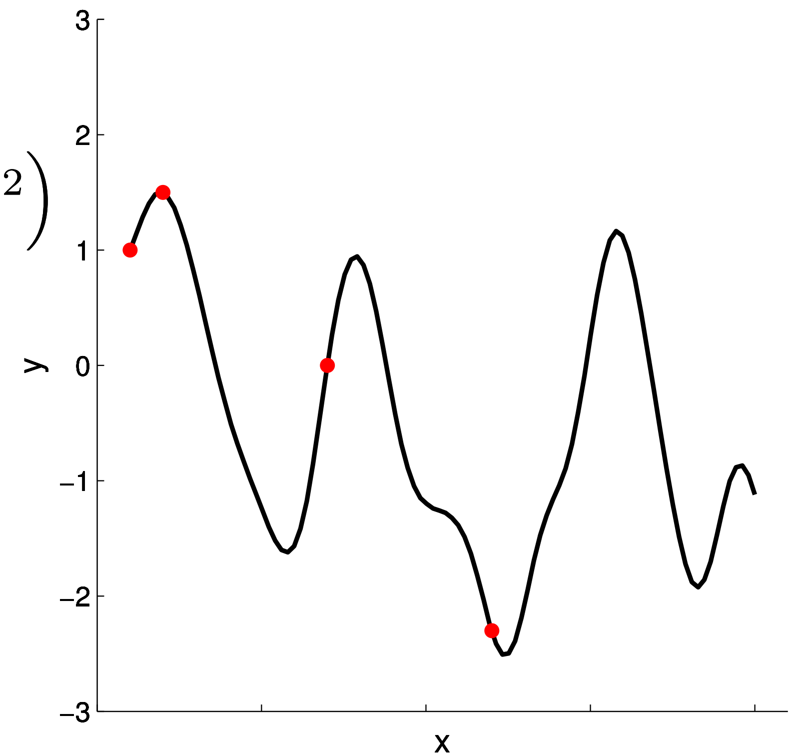
Non-parametric (∞-parametric)

$$p(\mathsf{y}|\theta) = \mathcal{N}(0, \Sigma)$$

$$\Sigma(\mathsf{x}_1, \mathsf{x}_2) = \mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) + \mathsf{I}\sigma_\mathsf{y}^2$$

$$\mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) = \sigma^2 \exp\left(-\frac{1}{2\mathsf{l}^2}(\mathsf{x}_1 - \mathsf{x}_2)^2\right)$$

$\Sigma =$ 



Parametric model

$$\mathsf{y}(\mathsf{x}) = f(\mathsf{x}; \theta) + \sigma_\mathsf{y}\epsilon$$

$$\epsilon \sim \mathcal{N}(0, 1)$$

What effect do the hyper-parameters have?

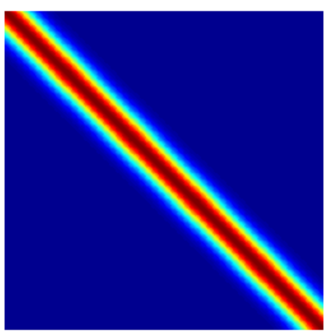
short horizontal length-scale

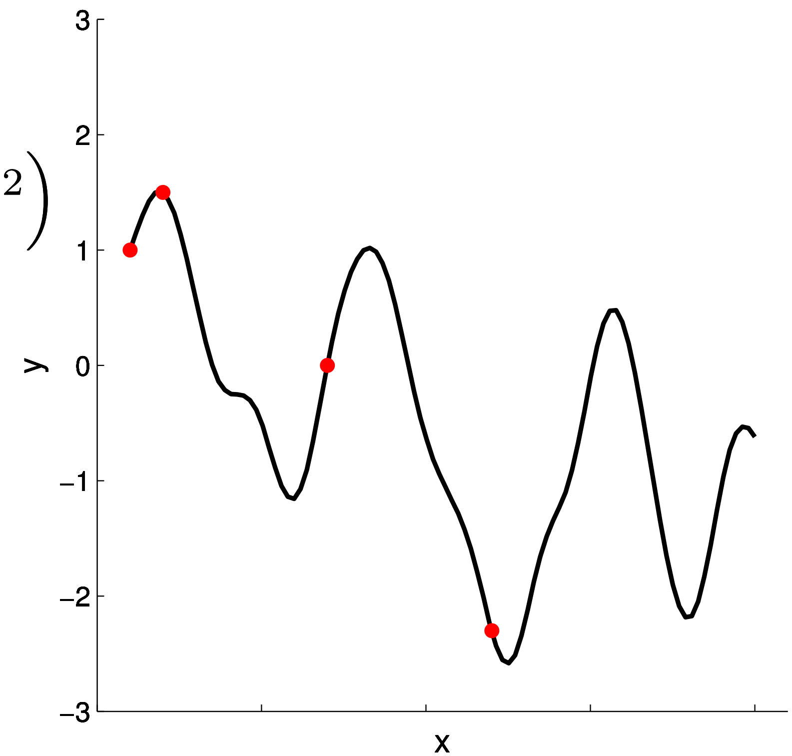
Non-parametric (∞-parametric)

$p(\mathsf{y}|\theta) = \mathcal{N}(0, \Sigma)$

$\Sigma(\mathsf{x}_1, \mathsf{x}_2) = \mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) + \mathsf{I}\sigma_\mathsf{y}^2$

$\mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) = \sigma^2 \exp\left(-\frac{1}{2\mathsf{l}^2}(\mathsf{x}_1 - \mathsf{x}_2)^2\right)$

$\Sigma =$ 



Parametric model

$\mathsf{y}(\mathsf{x}) = f(\mathsf{x}; \theta) + \sigma_\mathsf{y}\epsilon$

$\epsilon \sim \mathcal{N}(0, 1)$

What effect do the hyper-parameters have?

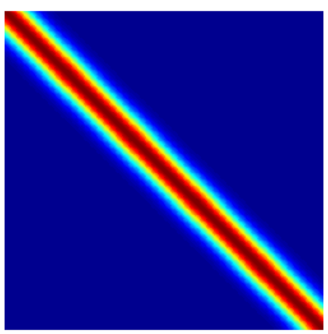
short horizontal length-scale

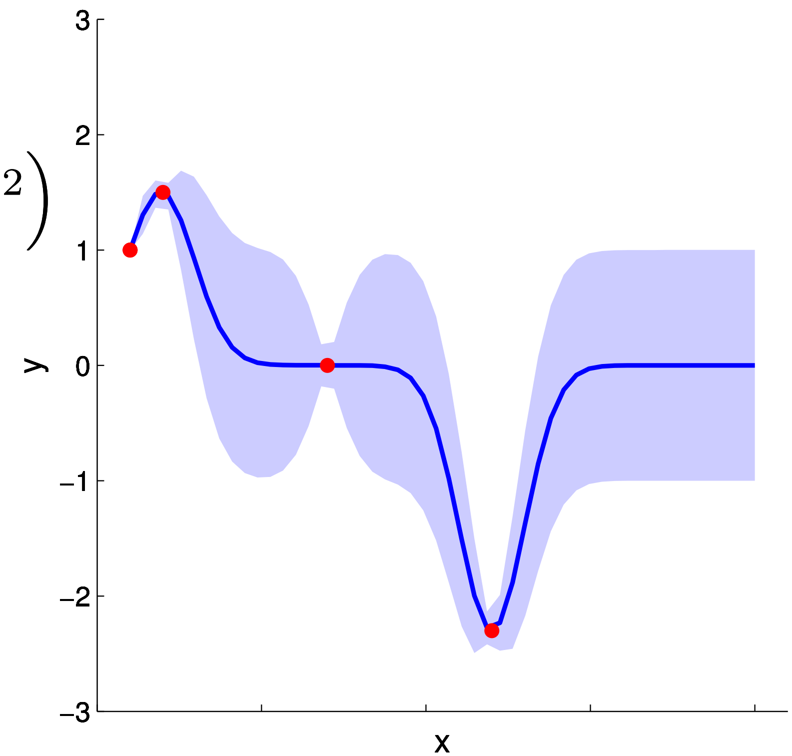
Non-parametric (∞-parametric)

$$p(\mathsf{y}|\theta) = \mathcal{N}(0, \Sigma)$$

$$\Sigma(\mathsf{x}_1, \mathsf{x}_2) = \mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) + \mathsf{I}\sigma_\mathsf{y}^2$$

$$\mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) = \sigma^2 \exp\left(-\frac{1}{2\mathsf{l}^2}(\mathsf{x}_1 - \mathsf{x}_2)^2\right)$$

$\Sigma =$ 



Parametric model

$$\mathsf{y}(\mathsf{x}) = f(\mathsf{x}; \theta) + \sigma_\mathsf{y}\epsilon$$

$$\epsilon \sim \mathcal{N}(0, 1)$$

What effect do the hyper-parameters have?

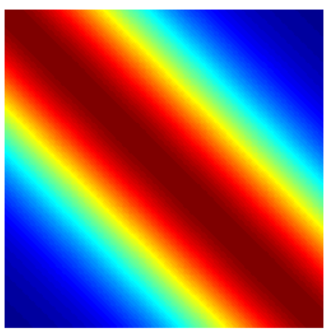
long horizontal length-scale

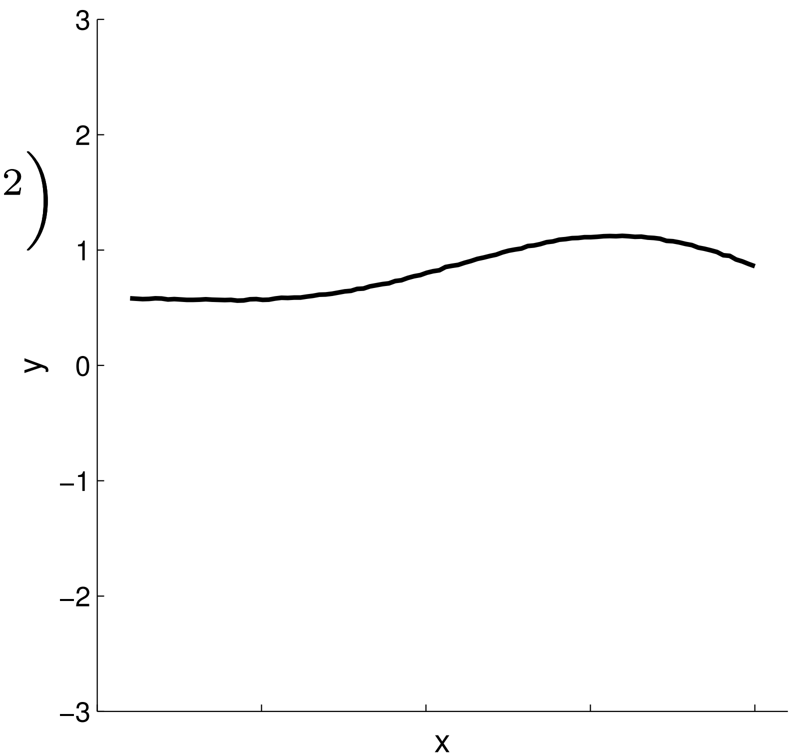
Non-parametric (∞-parametric)

$$p(\mathsf{y}|\theta) = \mathcal{N}(0, \Sigma)$$

$$\Sigma(\mathsf{x}_1, \mathsf{x}_2) = \mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) + \mathsf{I}\sigma_\mathsf{y}^2$$

$$\mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) = \sigma^2 \exp\left(-\frac{1}{2\mathsf{l}^2}(\mathsf{x}_1 - \mathsf{x}_2)^2\right)$$

$\Sigma =$ 



Parametric model

$$\mathsf{y}(\mathsf{x}) = f(\mathsf{x}; \theta) + \sigma_\mathsf{y}\epsilon$$

$$\epsilon \sim \mathcal{N}(0, 1)$$

What effect do the hyper-parameters have?

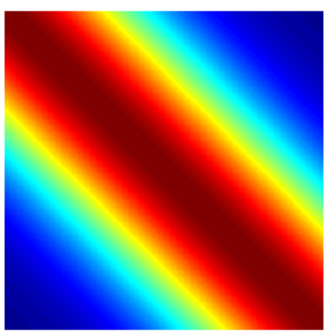
long horizontal length-scale

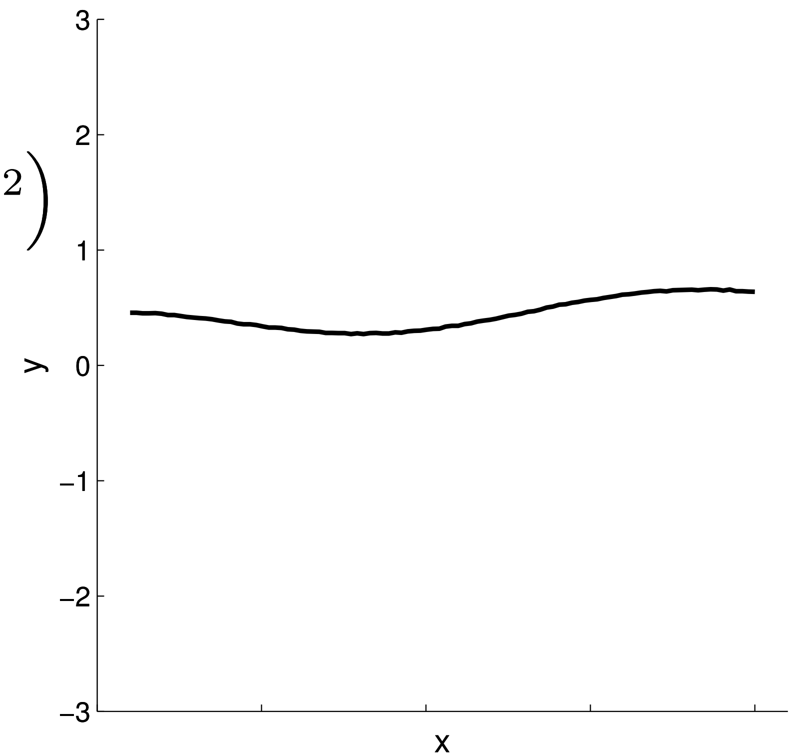
Non-parametric (∞-parametric)

$$p(\mathsf{y}|\theta) = \mathcal{N}(0, \Sigma)$$

$$\Sigma(\mathsf{x}_1, \mathsf{x}_2) = \mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) + \mathsf{I}\sigma_{\mathsf{y}}^2$$

$$\mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) = \sigma^2 \exp\left(-\frac{1}{2\mathsf{l}^2}(\mathsf{x}_1 - \mathsf{x}_2)^2\right)$$

$\Sigma =$ 



Parametric model

$$\mathsf{y}(\mathsf{x}) = f(\mathsf{x}; \theta) + \sigma_{\mathsf{y}}\epsilon$$

$$\epsilon \sim \mathcal{N}(0, 1)$$

What effect do the hyper-parameters have?

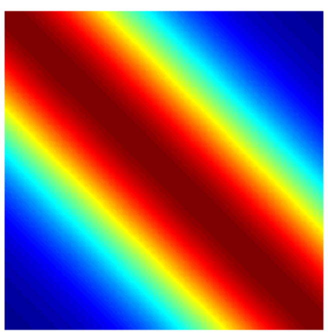
long horizontal length-scale

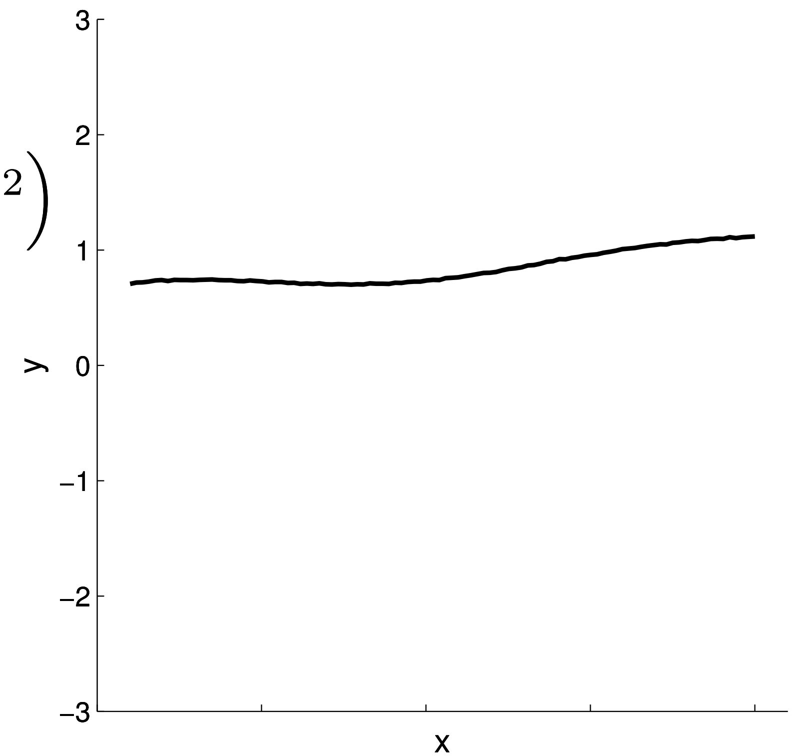
Non-parametric (∞-parametric)

$$p(\mathsf{y}|\theta) = \mathcal{N}(0, \Sigma)$$

$$\Sigma(\mathsf{x}_1, \mathsf{x}_2) = \mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) + \mathsf{I}\sigma_{\mathsf{y}}^2$$

$$\mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) = \sigma^2 \exp\left(-\frac{1}{2\mathsf{l}^2}(\mathsf{x}_1 - \mathsf{x}_2)^2\right)$$

$\Sigma =$ 



Parametric model

$$\mathsf{y}(\mathsf{x}) = f(\mathsf{x}; \theta) + \sigma_{\mathsf{y}}\epsilon$$

$$\epsilon \sim \mathcal{N}(0, 1)$$

What effect do the hyper-parameters have?

long horizontal length-scale

Non-parametric (∞-parametric)

Parametric model

$$p(\mathsf{y}|\theta) = \mathcal{N}(0, \Sigma)$$

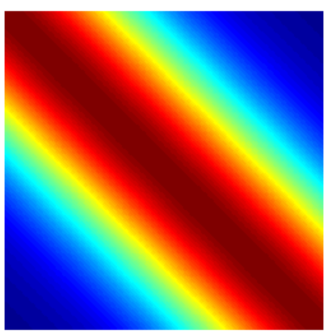
$$\Sigma(\mathsf{x}_1, \mathsf{x}_2) = \mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) + \mathsf{I}\sigma_{\mathsf{y}}^2$$

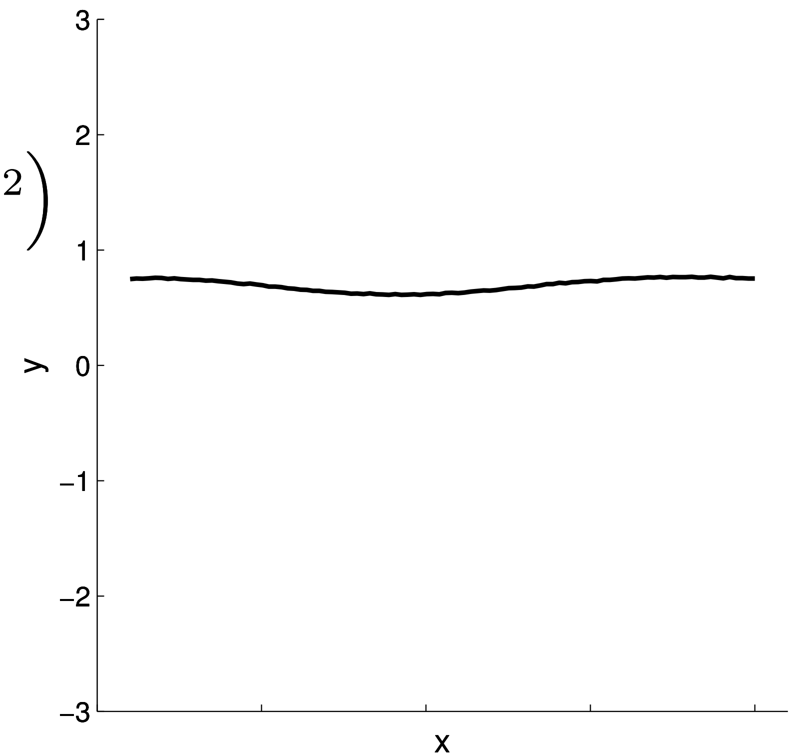
$$\mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) = \sigma^2 \exp\left(-\frac{1}{2\mathsf{l}^2}(\mathsf{x}_1 - \mathsf{x}_2)^2\right)$$

$$\mathsf{y}(\mathsf{x}) = f(\mathsf{x}; \theta) + \sigma_{\mathsf{y}}\epsilon$$

$$\epsilon \sim \mathcal{N}(0, 1)$$

$$\Sigma =$$





What effect do the hyper-parameters have?

long horizontal length-scale

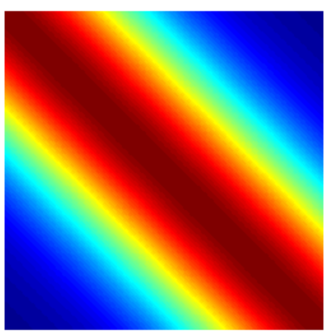
Non-parametric (∞-parametric)

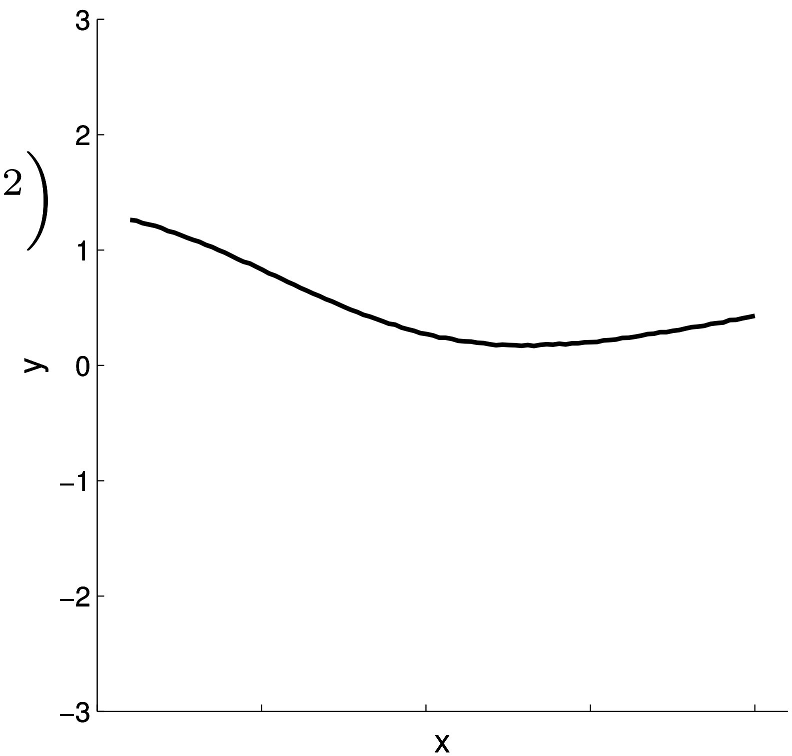
$$p(\mathsf{y}|\theta) = \mathcal{N}(0, \Sigma)$$

$$\Sigma(\mathsf{x}_1, \mathsf{x}_2) = \mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) + \mathsf{I}\sigma_{\mathsf{y}}^2$$

$$\mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) = \sigma^2 \exp\left(-\frac{1}{2\mathsf{l}^2}(\mathsf{x}_1 - \mathsf{x}_2)^2\right)$$

$\Sigma =$





Parametric model

$$\mathsf{y}(\mathsf{x}) = f(\mathsf{x}; \theta) + \sigma_{\mathsf{y}}\epsilon$$

$$\epsilon \sim \mathcal{N}(0, 1)$$

What effect do the hyper-parameters have?

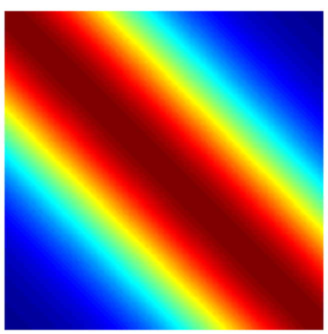
long horizontal length-scale

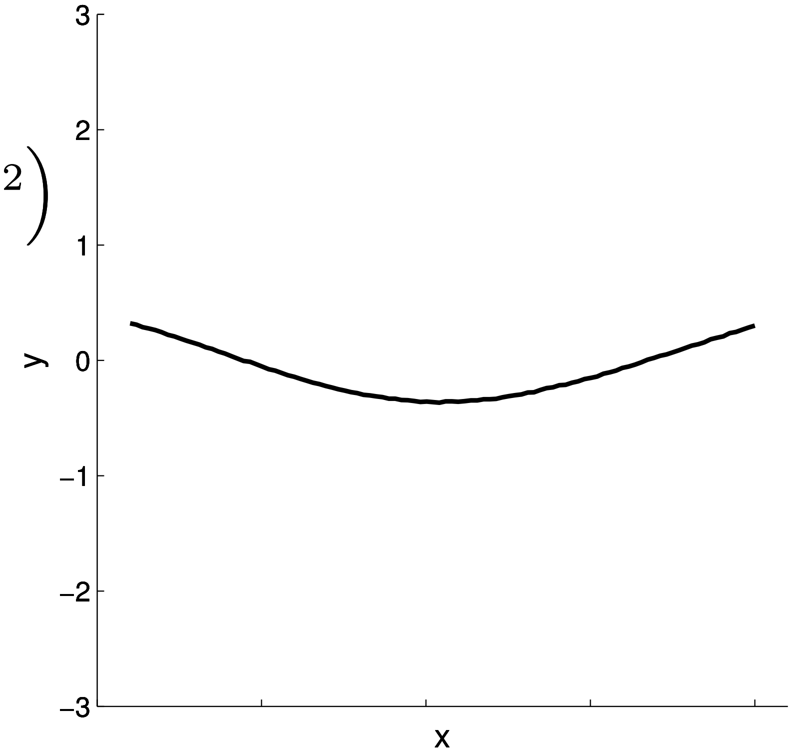
Non-parametric (∞-parametric)

$$p(\mathsf{y}|\theta) = \mathcal{N}(0, \Sigma)$$

$$\Sigma(\mathsf{x}_1, \mathsf{x}_2) = \mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) + \mathsf{I}\sigma_{\mathsf{y}}^2$$

$$\mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) = \sigma^2 \exp\left(-\frac{1}{2\mathsf{l}^2}(\mathsf{x}_1 - \mathsf{x}_2)^2\right)$$

$\Sigma =$ 



Parametric model

$$\mathsf{y}(\mathsf{x}) = f(\mathsf{x};\theta) + \sigma_{\mathsf{y}}\epsilon$$

$$\epsilon \sim \mathcal{N}(0, 1)$$

What effect do the hyper-parameters have?

long horizontal length-scale

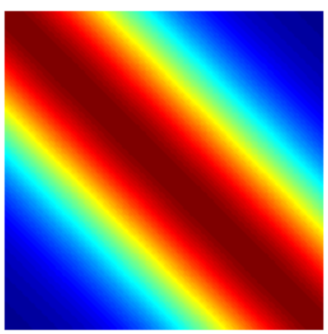
Non-parametric (∞-parametric)

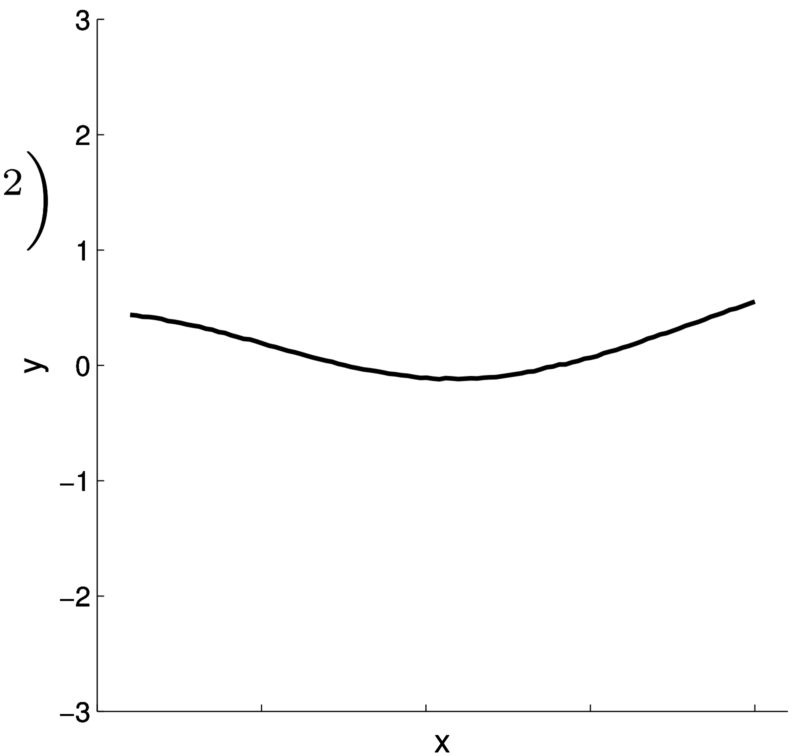
$p(\mathsf{y}|\theta) = \mathcal{N}(0, \Sigma)$

$\Sigma(\mathsf{x}_1, \mathsf{x}_2) = \mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) + \mathsf{I}\sigma_{\mathsf{y}}^2$

$\mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) = \sigma^2 \exp\left(-\frac{1}{2\mathsf{l}^2}(\mathsf{x}_1 - \mathsf{x}_2)^2\right)$

$\Sigma =$





Parametric model

$\mathsf{y}(\mathsf{x}) = f(\mathsf{x}; \theta) + \sigma_{\mathsf{y}}\epsilon$

$\epsilon \sim \mathcal{N}(0, 1)$

What effect do the hyper-parameters have?

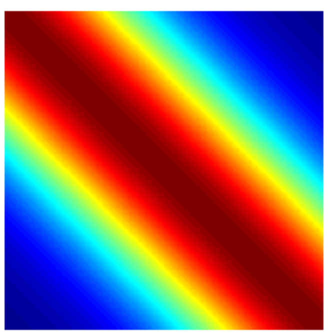
long horizontal length-scale

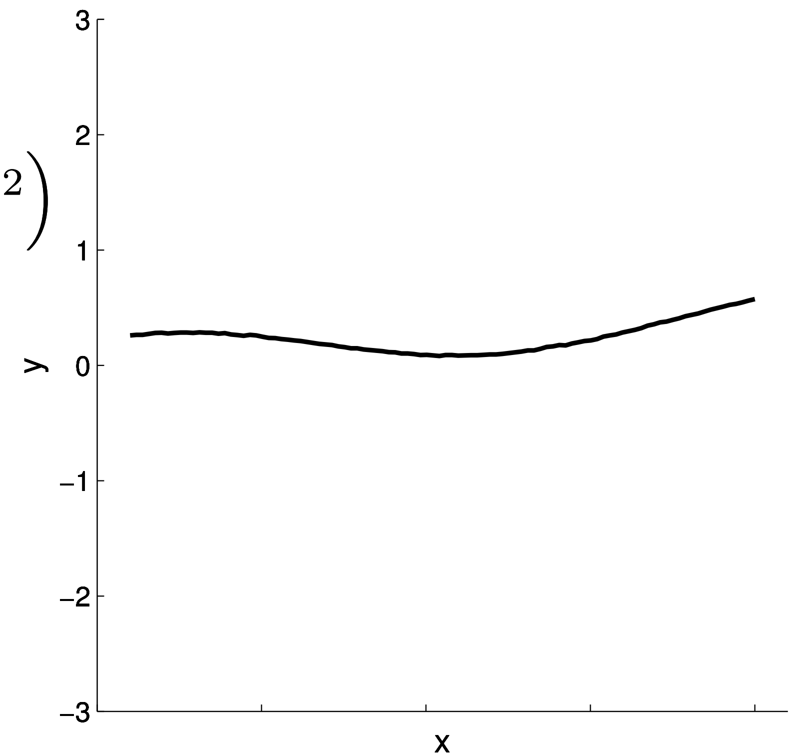
Non-parametric (∞-parametric)

$$p(\mathsf{y}|\theta) = \mathcal{N}(0, \Sigma)$$

$$\Sigma(\mathsf{x}_1, \mathsf{x}_2) = \mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) + \mathsf{I}\sigma_{\mathsf{y}}^2$$

$$\mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) = \sigma^2 \exp\left(-\frac{1}{2\mathsf{l}^2}(\mathsf{x}_1 - \mathsf{x}_2)^2\right)$$

$\Sigma =$ 



Parametric model

$$\mathsf{y}(\mathsf{x}) = f(\mathsf{x}; \theta) + \sigma_{\mathsf{y}}\epsilon$$

$$\epsilon \sim \mathcal{N}(0, 1)$$

What effect do the hyper-parameters have?

long horizontal length-scale

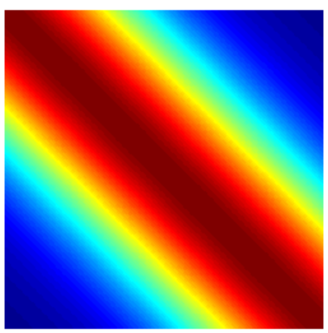
Non-parametric (∞-parametric)

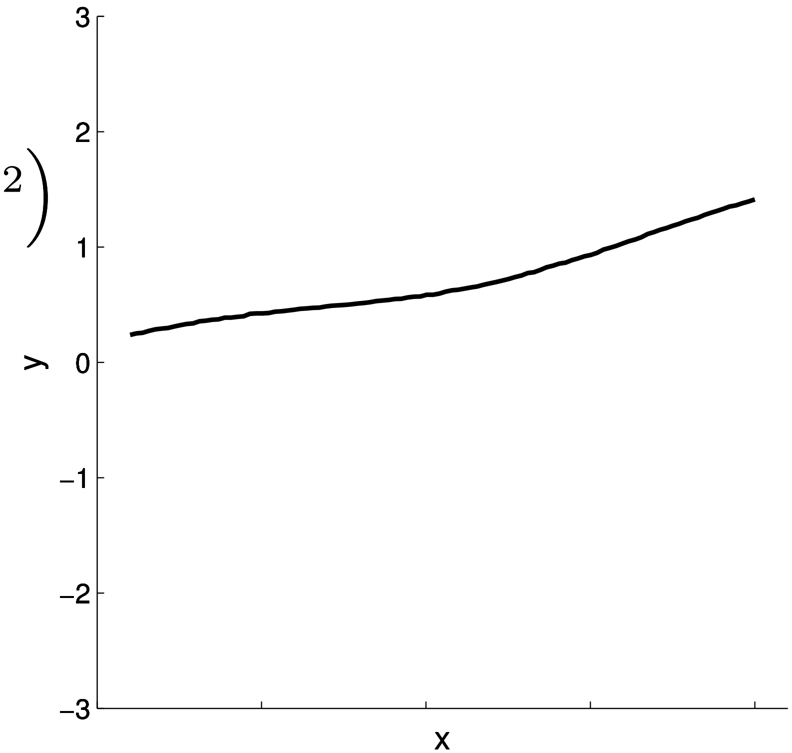
$$p(\mathsf{y}|\theta) = \mathcal{N}(0, \Sigma)$$

$$\Sigma(\mathsf{x}_1, \mathsf{x}_2) = \mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) + \mathsf{I}\sigma_{\mathsf{y}}^2$$

$$\mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) = \sigma^2 \exp\left(-\frac{1}{2\mathsf{l}^2}(\mathsf{x}_1 - \mathsf{x}_2)^2\right)$$

$\Sigma =$





Parametric model

$$\mathsf{y}(\mathsf{x}) = f(\mathsf{x}; \theta) + \sigma_{\mathsf{y}}\epsilon$$

$$\epsilon \sim \mathcal{N}(0, 1)$$

What effect do the hyper-parameters have?

long horizontal length-scale

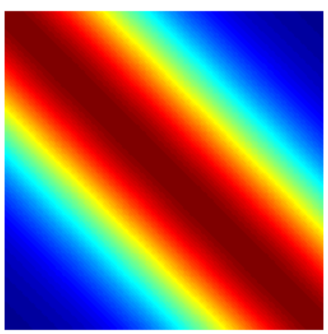
Non-parametric (∞-parametric)

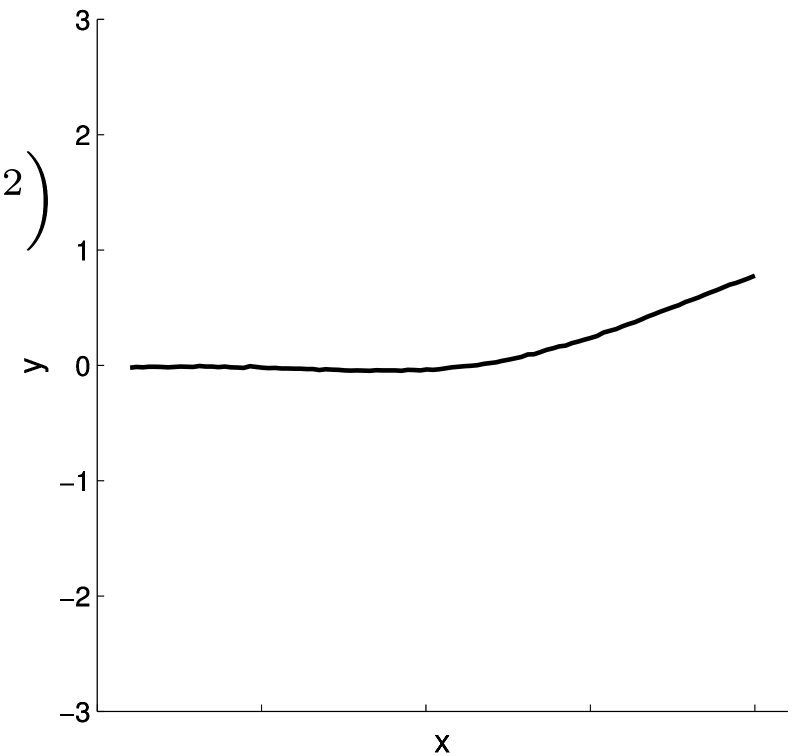
$$p(\mathsf{y}|\theta) = \mathcal{N}(0, \Sigma)$$

$$\Sigma(\mathsf{x}_1, \mathsf{x}_2) = \mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) + \mathsf{I}\sigma_{\mathsf{y}}^2$$

$$\mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) = \sigma^2 \exp\left(-\frac{1}{2\mathsf{l}^2}(\mathsf{x}_1 - \mathsf{x}_2)^2\right)$$

$\Sigma =$





Parametric model

$$\mathsf{y}(\mathsf{x}) = f(\mathsf{x}; \theta) + \sigma_{\mathsf{y}}\epsilon$$

$$\epsilon \sim \mathcal{N}(0, 1)$$

What effect do the hyper-parameters have?

long horizontal length-scale

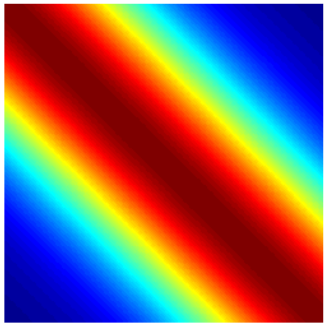
Non-parametric (∞-parametric)

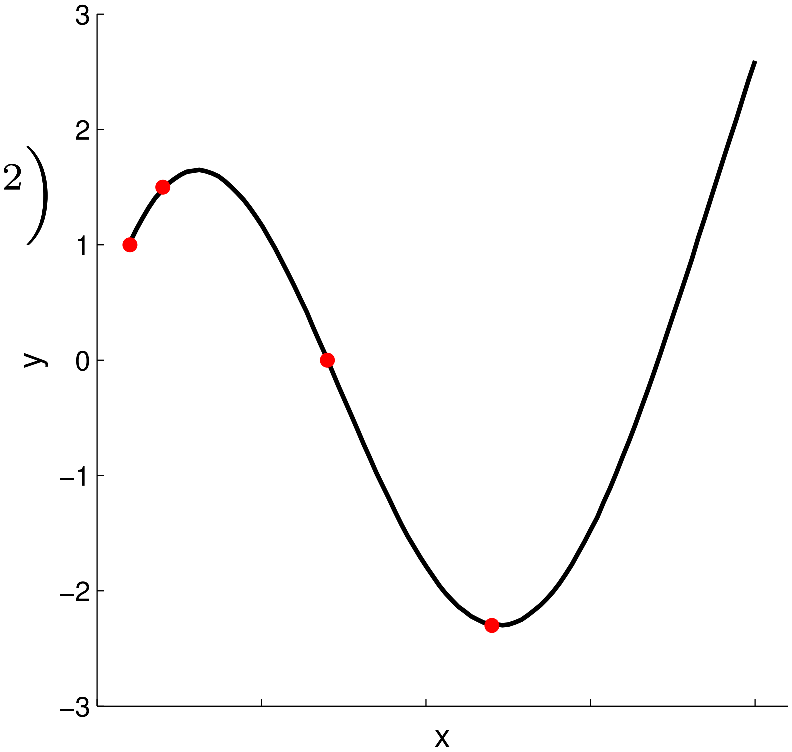
$$p(\mathsf{y}|\theta) = \mathcal{N}(0, \Sigma)$$

$$\Sigma(\mathsf{x}_1, \mathsf{x}_2) = \mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) + \mathsf{I}\sigma_\mathsf{y}^2$$

$$\mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) = \sigma^2 \exp\left(-\frac{1}{2\mathsf{l}^2}(\mathsf{x}_1 - \mathsf{x}_2)^2\right)$$

$$\Sigma =$$





Parametric model

$$\mathsf{y}(\mathsf{x}) = f(\mathsf{x}; \theta) + \sigma_\mathsf{y}\epsilon$$

$$\epsilon \sim \mathcal{N}(0, 1)$$

What effect do the hyper-parameters have?

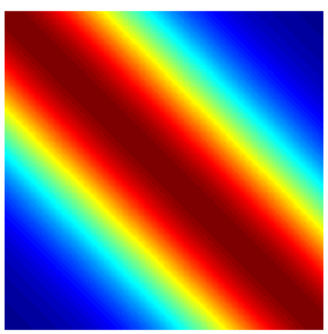
long horizontal length-scale

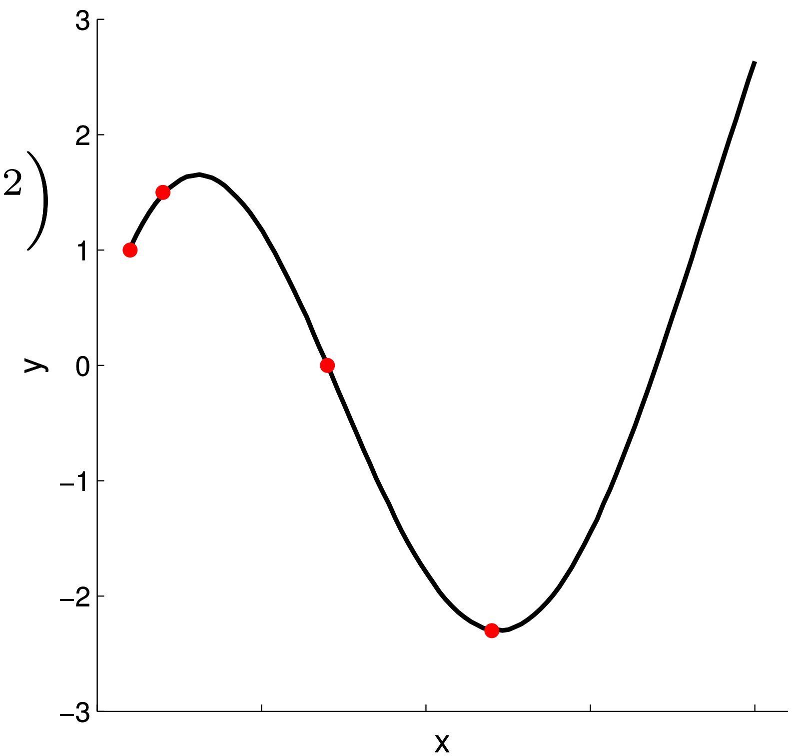
Non-parametric (∞-parametric)

$$p(\mathsf{y}|\theta) = \mathcal{N}(0, \Sigma)$$

$$\Sigma(\mathsf{x}_1, \mathsf{x}_2) = \mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) + \mathsf{I}\sigma_{\mathsf{y}}^2$$

$$\mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) = \sigma^2 \exp\left(-\frac{1}{2\mathsf{l}^2}(\mathsf{x}_1 - \mathsf{x}_2)^2\right)$$

$\Sigma =$ 



Parametric model

$$\mathsf{y}(\mathsf{x}) = f(\mathsf{x}; \theta) + \sigma_{\mathsf{y}}\epsilon$$

$$\epsilon \sim \mathcal{N}(0, 1)$$

What effect do the hyper-parameters have?

long horizontal length-scale

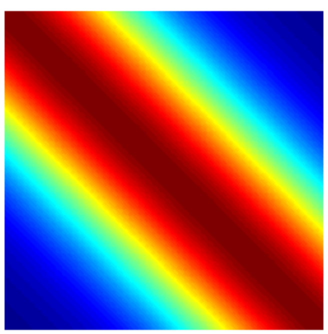
Non-parametric (∞-parametric)

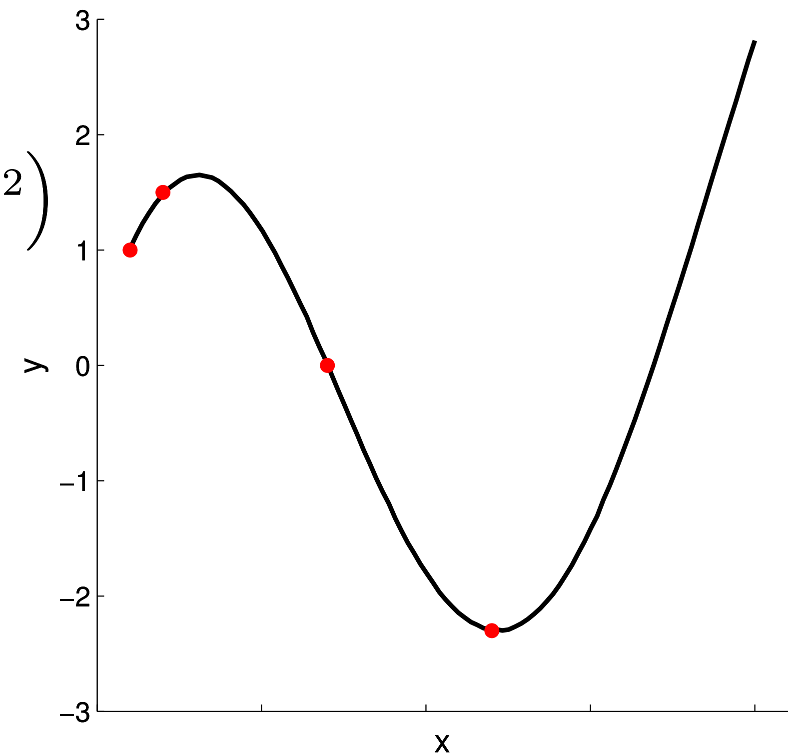
$$p(\mathsf{y}|\theta) = \mathcal{N}(0, \Sigma)$$

$$\Sigma(\mathsf{x}_1, \mathsf{x}_2) = \mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) + \mathsf{I}\sigma_\mathsf{y}^2$$

$$\mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) = \sigma^2 \exp\left(-\frac{1}{2\mathsf{l}^2}(\mathsf{x}_1 - \mathsf{x}_2)^2\right)$$

$$\Sigma =$$





Parametric model

$$\mathsf{y}(\mathsf{x}) = f(\mathsf{x};\theta) + \sigma_\mathsf{y}\epsilon$$

$$\epsilon \sim \mathcal{N}(0, 1)$$

What effect do the hyper-parameters have?

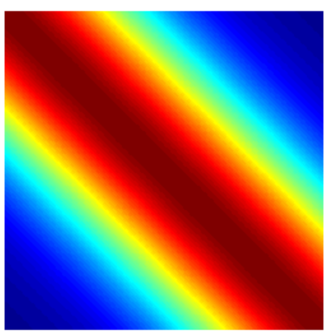
long horizontal length-scale

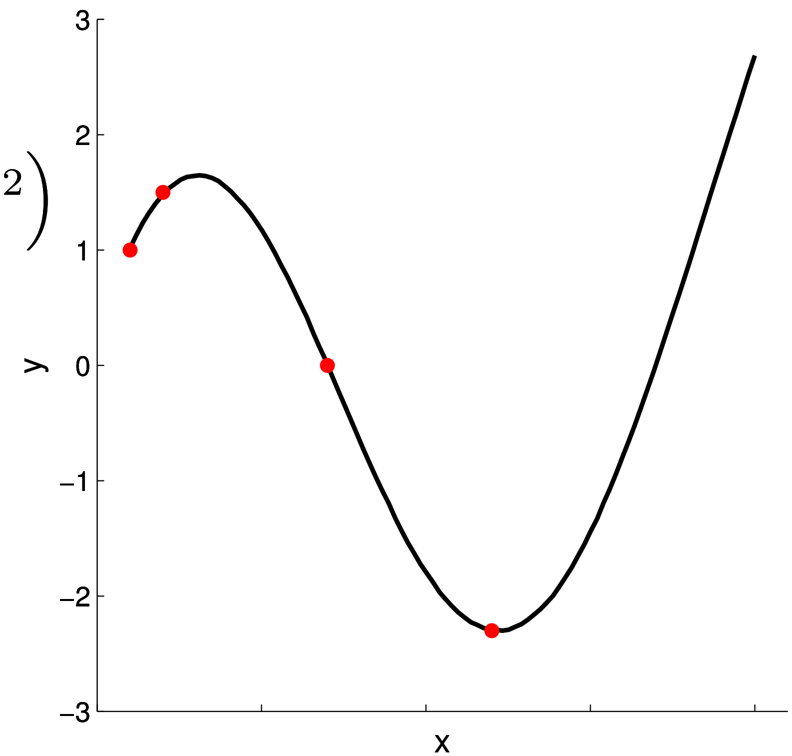
Non-parametric (∞-parametric)

$$p(\mathsf{y}|\theta) = \mathcal{N}(0, \Sigma)$$

$$\Sigma(\mathsf{x}_1, \mathsf{x}_2) = \mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) + \mathsf{I}\sigma_{\mathsf{y}}^2$$

$$\mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) = \sigma^2 \exp\left(-\frac{1}{2\mathsf{l}^2}(\mathsf{x}_1 - \mathsf{x}_2)^2\right)$$

$\Sigma =$ 



Parametric model

$$\mathsf{y}(\mathsf{x}) = f(\mathsf{x}; \theta) + \sigma_{\mathsf{y}}\epsilon$$

$$\epsilon \sim \mathcal{N}(0, 1)$$

What effect do the hyper-parameters have?

long horizontal length-scale

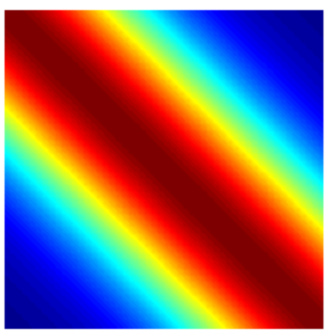
Non-parametric (∞-parametric)

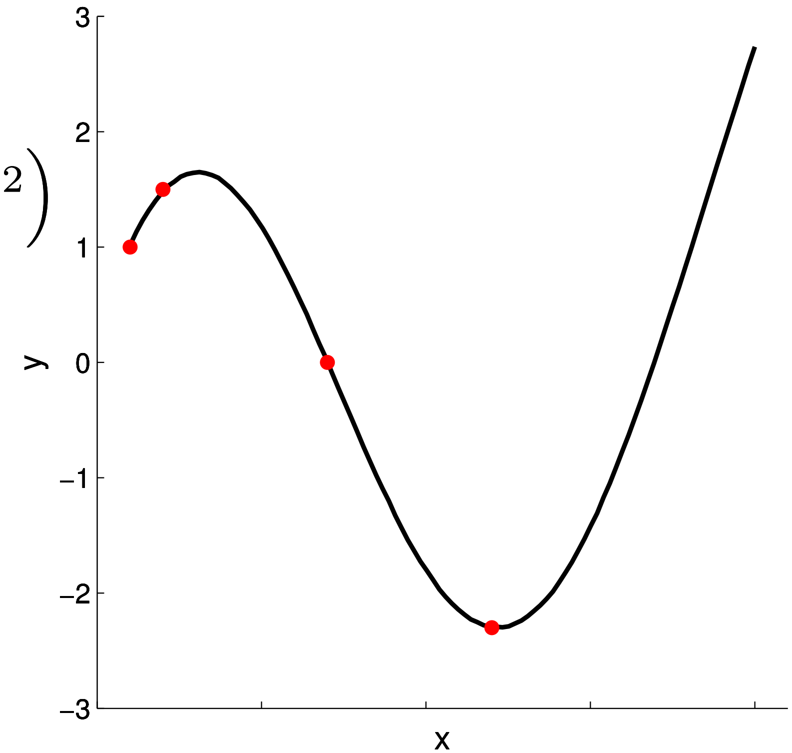
$$p(\mathsf{y}|\theta) = \mathcal{N}(0, \Sigma)$$

$$\Sigma(\mathsf{x}_1, \mathsf{x}_2) = \mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) + \mathsf{I}\sigma_\mathsf{y}^2$$

$$\mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) = \sigma^2 \exp\left(-\frac{1}{2\mathsf{l}^2}(\mathsf{x}_1 - \mathsf{x}_2)^2\right)$$

$$\Sigma =$$





Parametric model

$$\mathsf{y}(\mathsf{x}) = f(\mathsf{x}; \theta) + \sigma_\mathsf{y}\epsilon$$

$$\epsilon \sim \mathcal{N}(0, 1)$$

What effect do the hyper-parameters have?

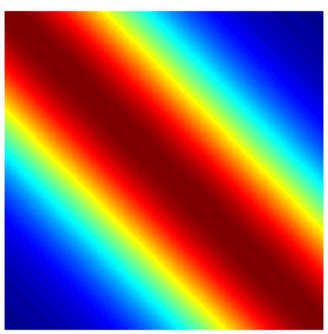
long horizontal length-scale

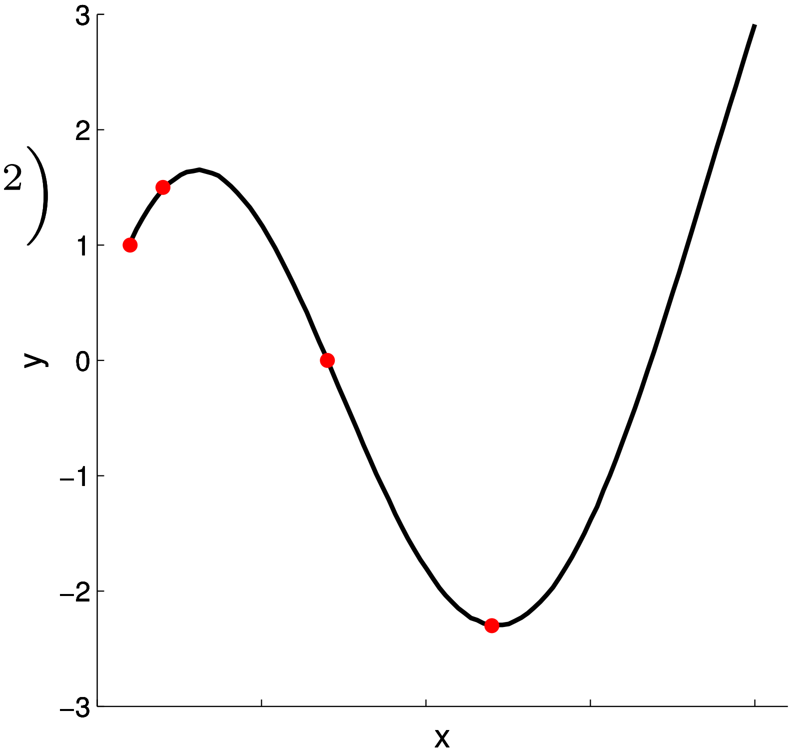
Non-parametric (∞-parametric)

$$p(\mathsf{y}|\theta) = \mathcal{N}(0, \Sigma)$$

$$\Sigma(\mathsf{x}_1, \mathsf{x}_2) = \mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) + \mathsf{I}\sigma_{\mathsf{y}}^2$$

$$\mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) = \sigma^2 \exp\left(-\frac{1}{2\mathsf{l}^2}(\mathsf{x}_1 - \mathsf{x}_2)^2\right)$$

$\Sigma =$ 



Parametric model

$$\mathsf{y}(\mathsf{x}) = f(\mathsf{x}; \theta) + \sigma_{\mathsf{y}}\epsilon$$

$$\epsilon \sim \mathcal{N}(0, 1)$$

What effect do the hyper-parameters have?

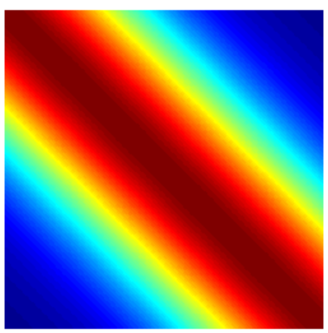
long horizontal length-scale

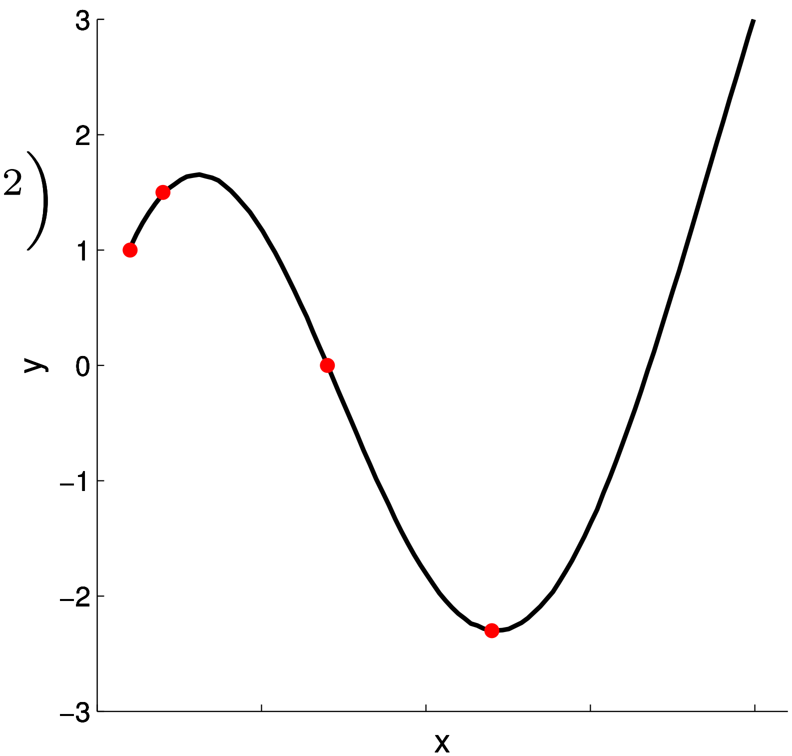
Non-parametric (∞-parametric)

$$p(\mathsf{y}|\theta) = \mathcal{N}(0, \Sigma)$$

$$\Sigma(\mathsf{x}_1, \mathsf{x}_2) = \mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) + \mathsf{I}\sigma_{\mathsf{y}}^2$$

$$\mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) = \sigma^2 \exp\left(-\frac{1}{2\mathsf{l}^2}(\mathsf{x}_1 - \mathsf{x}_2)^2\right)$$

$\Sigma =$ 



Parametric model

$$\mathsf{y}(\mathsf{x}) = f(\mathsf{x};\theta) + \sigma_{\mathsf{y}}\epsilon$$

$$\epsilon \sim \mathcal{N}(0, 1)$$

What effect do the hyper-parameters have?

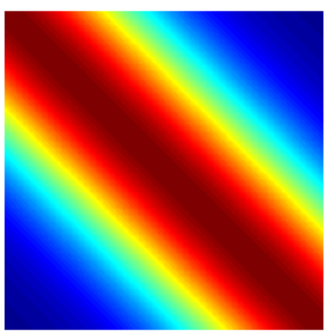
long horizontal length-scale

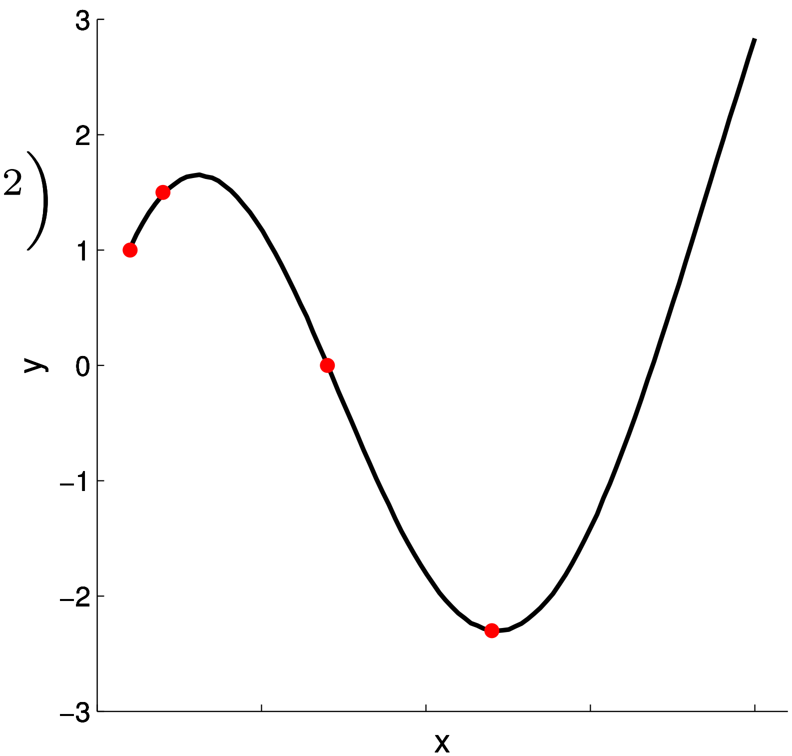
Non-parametric (∞-parametric)

$$p(\mathsf{y}|\theta) = \mathcal{N}(0, \Sigma)$$

$$\Sigma(\mathsf{x}_1, \mathsf{x}_2) = \mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) + \mathsf{I}\sigma_{\mathsf{y}}^2$$

$$\mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) = \sigma^2 \exp\left(-\frac{1}{2\mathsf{l}^2}(\mathsf{x}_1 - \mathsf{x}_2)^2\right)$$

$\Sigma =$ 



Parametric model

$$\mathsf{y}(\mathsf{x}) = f(\mathsf{x}; \theta) + \sigma_{\mathsf{y}}\epsilon$$

$$\epsilon \sim \mathcal{N}(0, 1)$$

What effect do the hyper-parameters have?

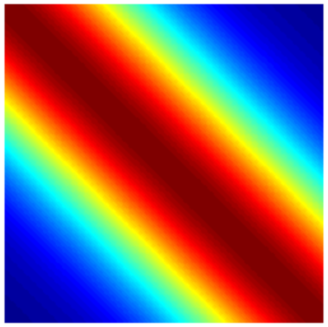
long horizontal length-scale

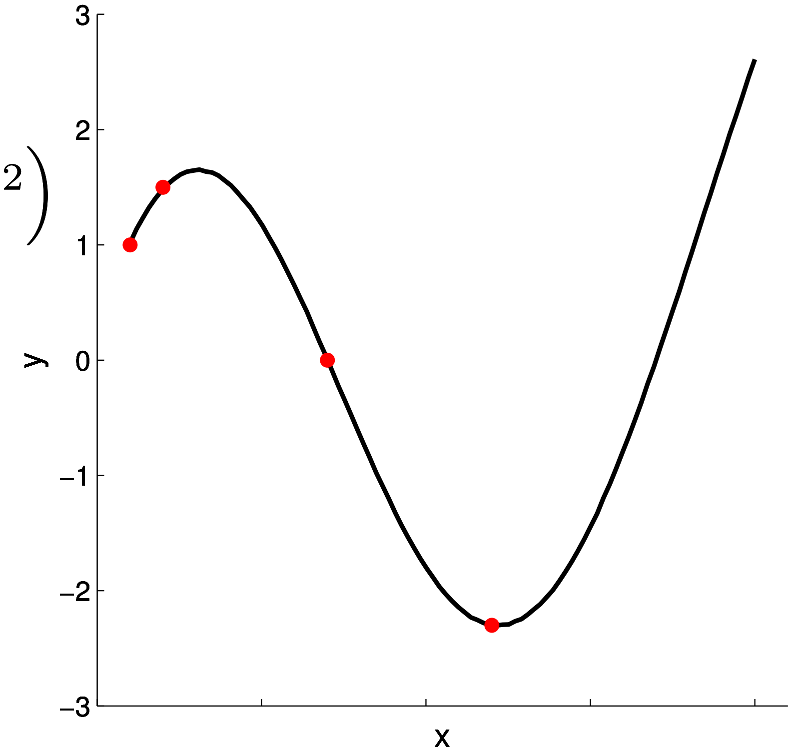
Non-parametric (∞-parametric)

$$p(\mathsf{y}|\theta) = \mathcal{N}(0, \Sigma)$$

$$\Sigma(\mathsf{x}_1, \mathsf{x}_2) = \mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) + \mathsf{I}\sigma_{\mathsf{y}}^2$$

$$\mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) = \sigma^2 \exp\left(-\frac{1}{2\mathsf{l}^2}(\mathsf{x}_1 - \mathsf{x}_2)^2\right)$$

$\Sigma =$ 



Parametric model

$$\mathsf{y}(\mathsf{x}) = f(\mathsf{x};\theta) + \sigma_{\mathsf{y}}\epsilon$$

$$\epsilon \sim \mathcal{N}(0, 1)$$

What effect do the hyper-parameters have?

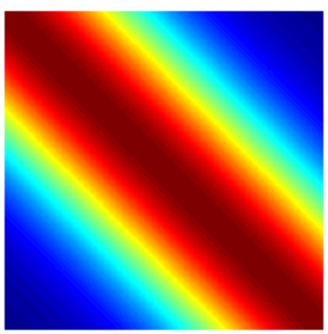
long horizontal length-scale

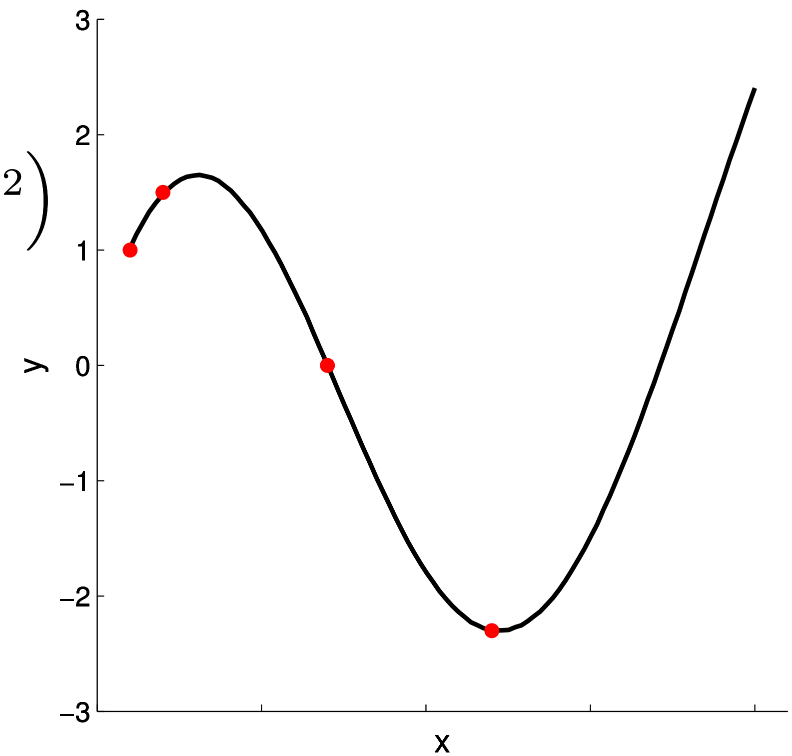
Non-parametric (∞-parametric)

$$p(\mathsf{y}|\theta) = \mathcal{N}(0, \Sigma)$$

$$\Sigma(\mathsf{x}_1, \mathsf{x}_2) = \mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) + \mathsf{I}\sigma_\mathsf{y}^2$$

$$\mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) = \sigma^2 \exp\left(-\frac{1}{2\mathsf{l}^2}(\mathsf{x}_1 - \mathsf{x}_2)^2\right)$$

$\Sigma =$ 



Parametric model

$$\mathsf{y}(\mathsf{x}) = f(\mathsf{x}; \theta) + \sigma_\mathsf{y}\epsilon$$

$$\epsilon \sim \mathcal{N}(0, 1)$$

What effect do the hyper-parameters have?

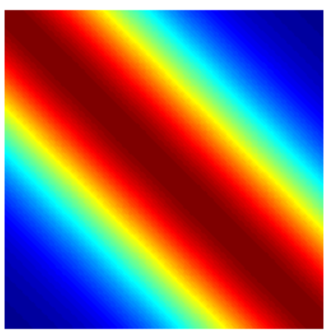
long horizontal length-scale

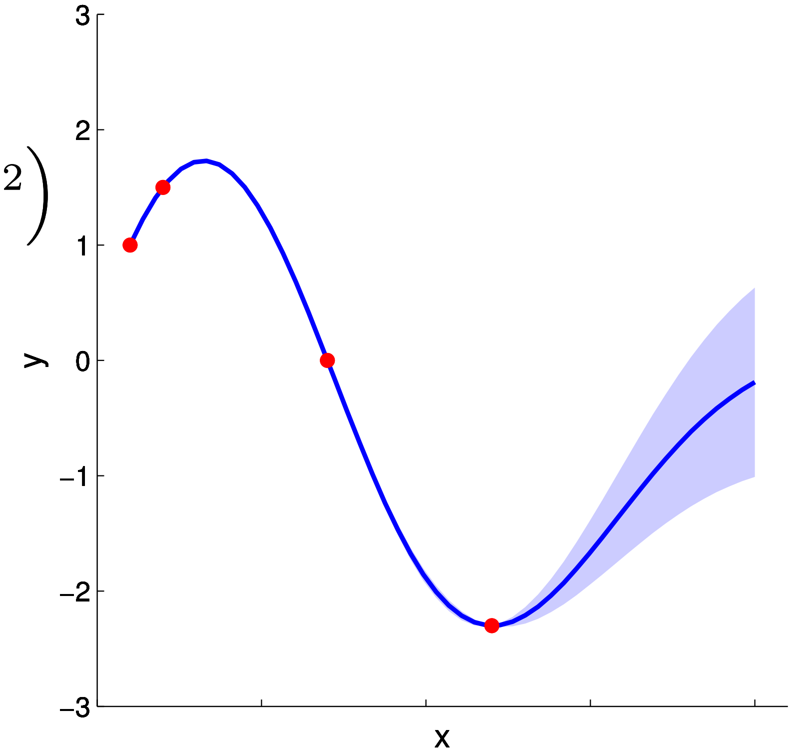
Non-parametric (∞-parametric)

$$p(\mathsf{y}|\theta) = \mathcal{N}(0, \Sigma)$$

$$\Sigma(\mathsf{x}_1, \mathsf{x}_2) = \mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) + \mathsf{I}\sigma_{\mathsf{y}}^2$$

$$\mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) = \sigma^2 \exp\left(-\frac{1}{2\mathsf{l}^2}(\mathsf{x}_1 - \mathsf{x}_2)^2\right)$$

$\Sigma =$ 



Parametric model

$$\mathsf{y}(\mathsf{x}) = f(\mathsf{x}; \theta) + \sigma_{\mathsf{y}}\epsilon$$

$$\epsilon \sim \mathcal{N}(0, 1)$$

What effect do the hyper-parameters have?

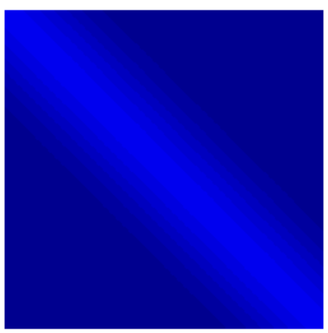
small vertical scale

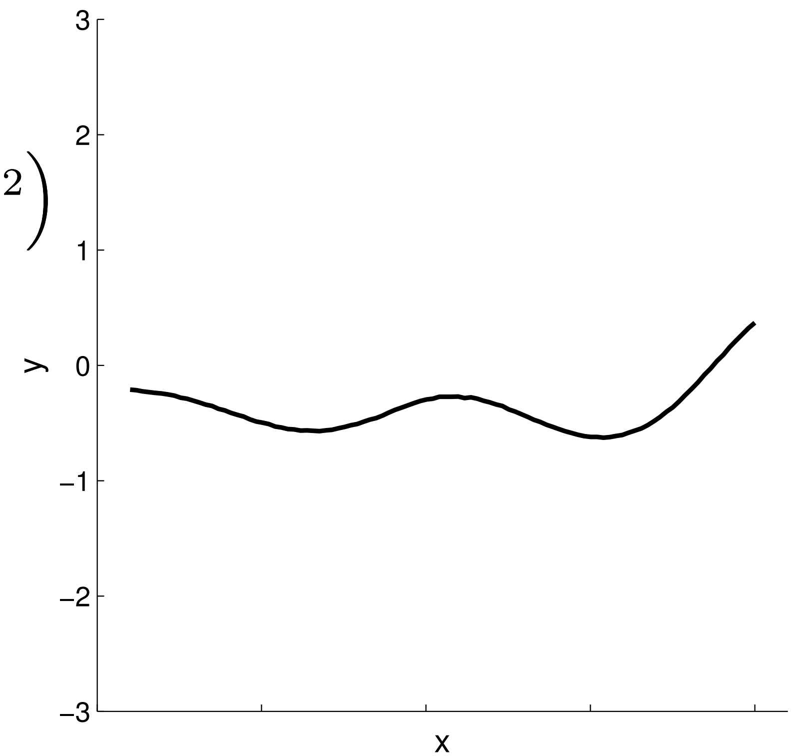
Non-parametric (∞-parametric)

$$p(\mathrm{y}|\theta) = \mathcal{N}(0, \Sigma)$$

$$\Sigma(\mathrm{x}_1, \mathrm{x}_2) = \mathrm{K}(\mathrm{x}_1, \mathrm{x}_2) + \mathrm{I}\sigma_\mathrm{y}^2$$

$$\mathrm{K}(\mathrm{x}_1, \mathrm{x}_2) = \sigma^2 \exp\left(-\frac{1}{2\mathrm{l}^2}(\mathrm{x}_1 - \mathrm{x}_2)^2\right)$$

$\Sigma =$ 



Parametric model

$$\mathrm{y}(\mathrm{x}) = f(\mathrm{x}; \theta) + \sigma_\mathrm{y}\epsilon$$

$$\epsilon \sim \mathcal{N}(0, 1)$$

What effect do the hyper-parameters have?

small vertical scale

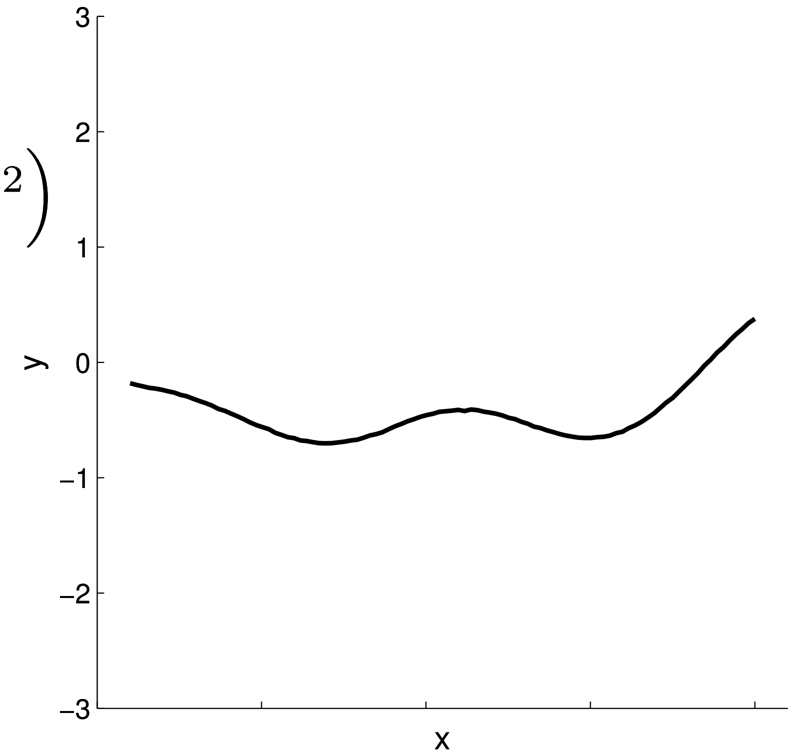
Non-parametric (∞-parametric)

$$p(\mathsf{y}|\theta) = \mathcal{N}(0, \Sigma)$$

$$\Sigma(\mathsf{x}_1, \mathsf{x}_2) = \mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) + \mathsf{I}\sigma_\mathsf{y}^2$$

$$\mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) = \sigma^2 \exp\left(-\frac{1}{2\mathsf{l}^2}(\mathsf{x}_1 - \mathsf{x}_2)^2\right)$$

$\Sigma =$ 



Parametric model

$$\mathsf{y}(\mathsf{x}) = f(\mathsf{x}; \theta) + \sigma_\mathsf{y}\epsilon$$

$$\epsilon \sim \mathcal{N}(0, 1)$$

What effect do the hyper-parameters have?

small vertical scale

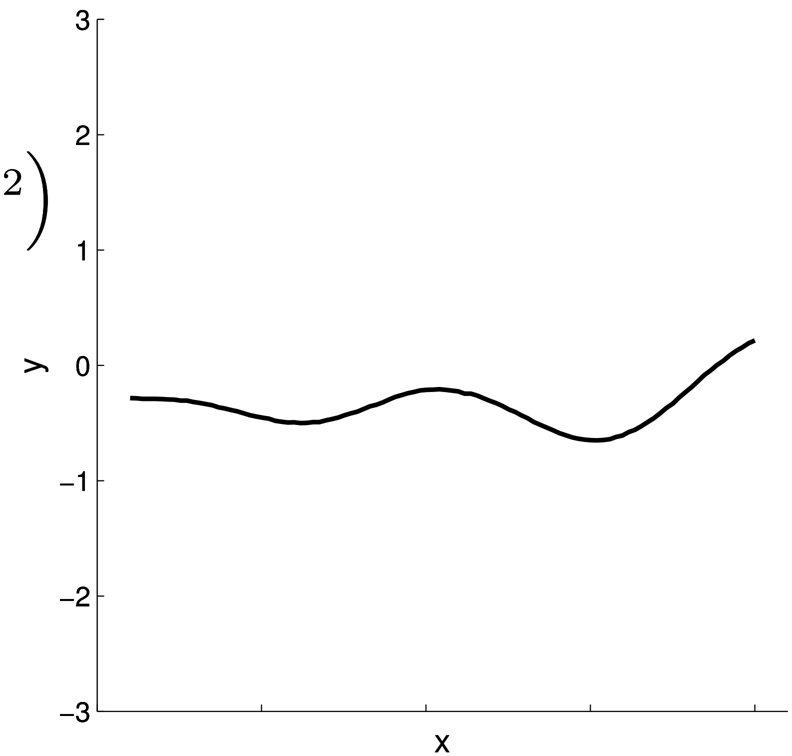
Non-parametric (∞-parametric)

$$p(\mathsf{y}|\theta) = \mathcal{N}(0, \Sigma)$$

$$\Sigma(\mathsf{x}_1, \mathsf{x}_2) = \mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) + \mathsf{I}\sigma_\mathsf{y}^2$$

$$\mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) = \sigma^2 \exp\left(-\frac{1}{2\mathsf{l}^2}(\mathsf{x}_1 - \mathsf{x}_2)^2\right)$$

$$\Sigma =$$



Parametric model

$$\mathsf{y}(\mathsf{x}) = f(\mathsf{x}; \theta) + \sigma_\mathsf{y}\epsilon$$

$$\epsilon \sim \mathcal{N}(0, 1)$$

What effect do the hyper-parameters have?

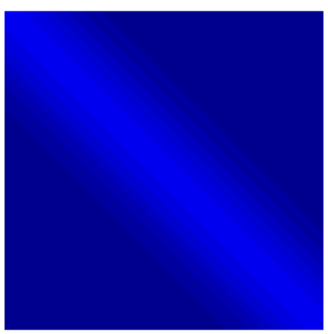
small vertical scale

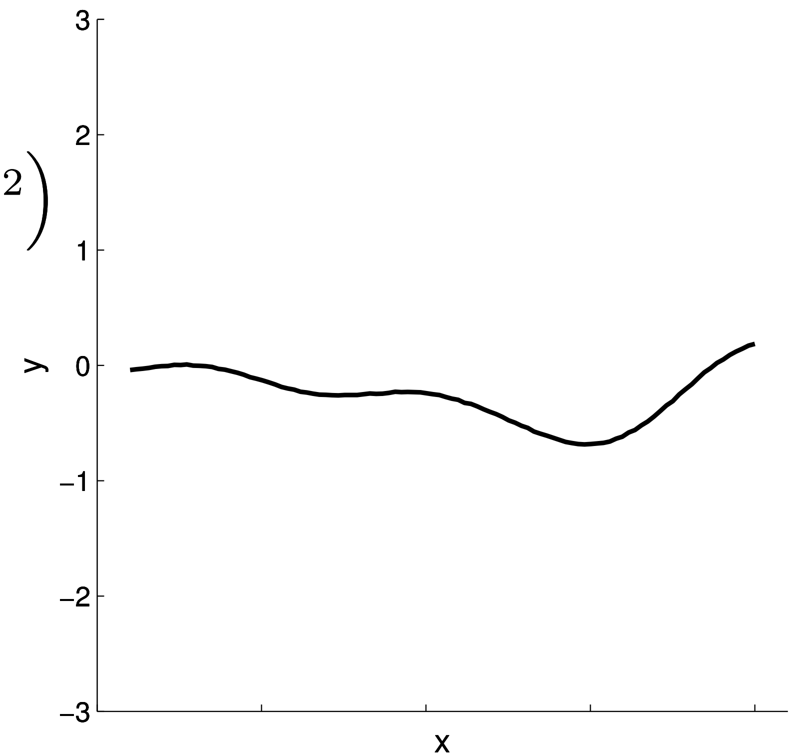
Non-parametric (∞-parametric)

$$p(\mathsf{y}|\theta) = \mathcal{N}(0, \Sigma)$$

$$\Sigma(\mathsf{x}_1, \mathsf{x}_2) = \mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) + \mathsf{I}\sigma_{\mathsf{y}}^2$$

$$\mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) = \sigma^2 \exp\left(-\frac{1}{2\mathsf{l}^2}(\mathsf{x}_1 - \mathsf{x}_2)^2\right)$$

$\Sigma =$ 



Parametric model

$$\mathsf{y}(\mathsf{x}) = f(\mathsf{x}; \theta) + \sigma_{\mathsf{y}}\epsilon$$

$$\epsilon \sim \mathcal{N}(0, 1)$$

What effect do the hyper-parameters have?

small vertical scale

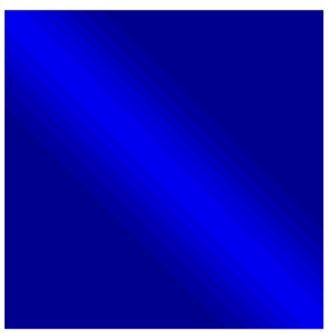
Non-parametric (∞-parametric)

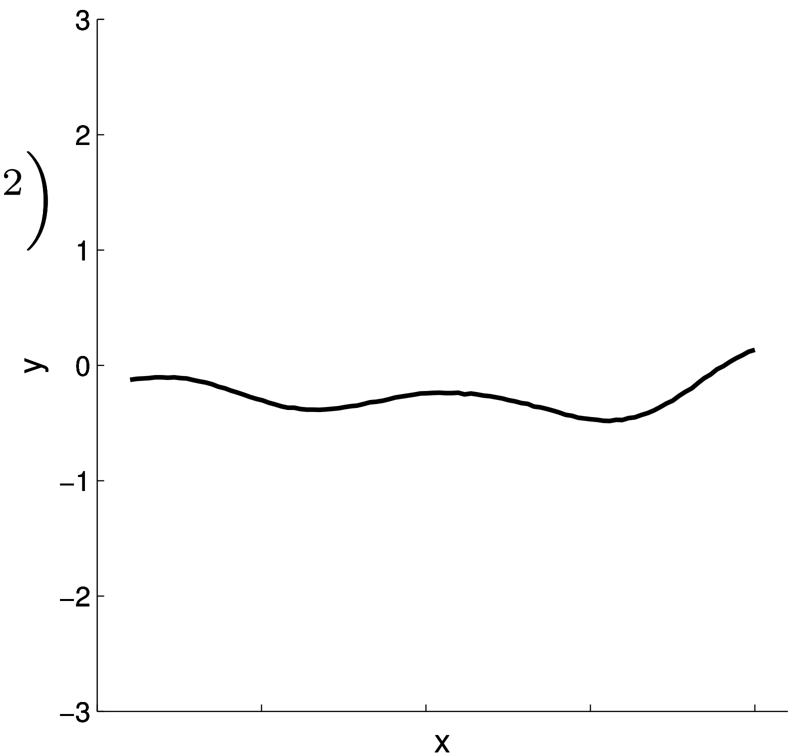
$$p(\mathsf{y}|\theta) = \mathcal{N}(0, \Sigma)$$

$$\Sigma(\mathsf{x}_1, \mathsf{x}_2) = \mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) + \mathsf{I}\sigma_{\mathsf{y}}^2$$

$$\mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) = \sigma^2 \exp\left(-\frac{1}{2\mathsf{l}^2}(\mathsf{x}_1 - \mathsf{x}_2)^2\right)$$

$$\Sigma =$$





Parametric model

$$\mathsf{y}(\mathsf{x}) = f(\mathsf{x};\theta) + \sigma_{\mathsf{y}}\epsilon$$

$$\epsilon \sim \mathcal{N}(0, 1)$$

What effect do the hyper-parameters have?

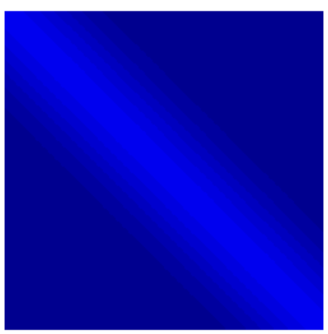
small vertical scale

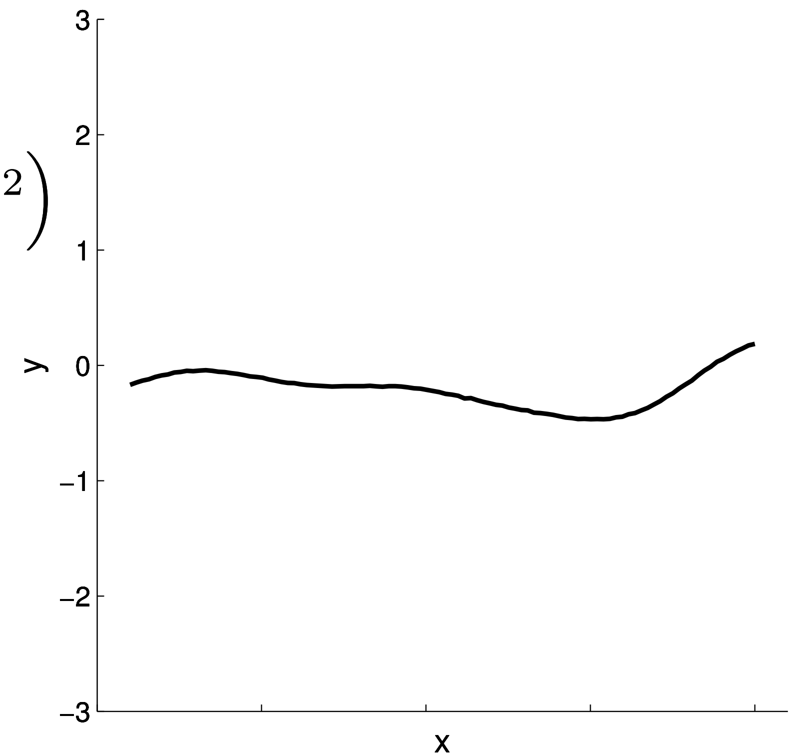
Non-parametric (∞-parametric)

$$p(\mathsf{y}|\theta) = \mathcal{N}(0, \Sigma)$$

$$\Sigma(\mathsf{x}_1, \mathsf{x}_2) = \mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) + \mathsf{I}\sigma_{\mathsf{y}}^2$$

$$\mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) = \sigma^2 \exp\left(-\frac{1}{2\mathsf{l}^2}(\mathsf{x}_1 - \mathsf{x}_2)^2\right)$$

$\Sigma =$ 



Parametric model

$$\mathsf{y}(\mathsf{x}) = f(\mathsf{x}; \theta) + \sigma_{\mathsf{y}}\epsilon$$

$$\epsilon \sim \mathcal{N}(0, 1)$$

What effect do the hyper-parameters have?

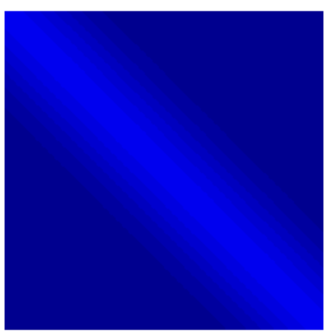
small vertical scale

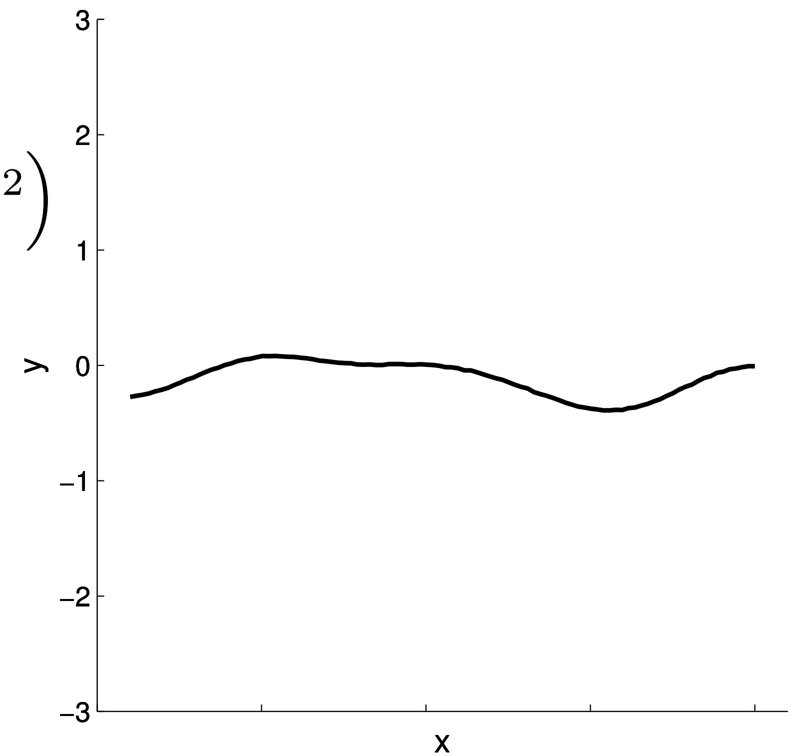
Non-parametric (∞-parametric)

$$p(\mathsf{y}|\theta) = \mathcal{N}(0, \Sigma)$$

$$\Sigma(\mathsf{x}_1, \mathsf{x}_2) = \mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) + \mathsf{I}\sigma_{\mathsf{y}}^2$$

$$\mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) = \sigma^2 \exp\left(-\frac{1}{2\mathsf{l}^2}(\mathsf{x}_1 - \mathsf{x}_2)^2\right)$$

$\Sigma =$ 



Parametric model

$$\mathsf{y}(\mathsf{x}) = f(\mathsf{x}; \theta) + \sigma_{\mathsf{y}}\epsilon$$

$$\epsilon \sim \mathcal{N}(0, 1)$$

What effect do the hyper-parameters have?

small vertical scale

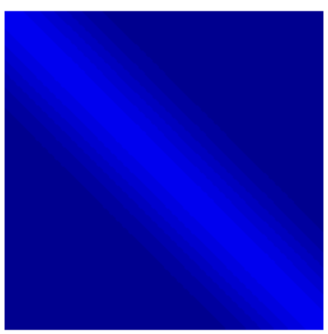
Non-parametric (∞-parametric)

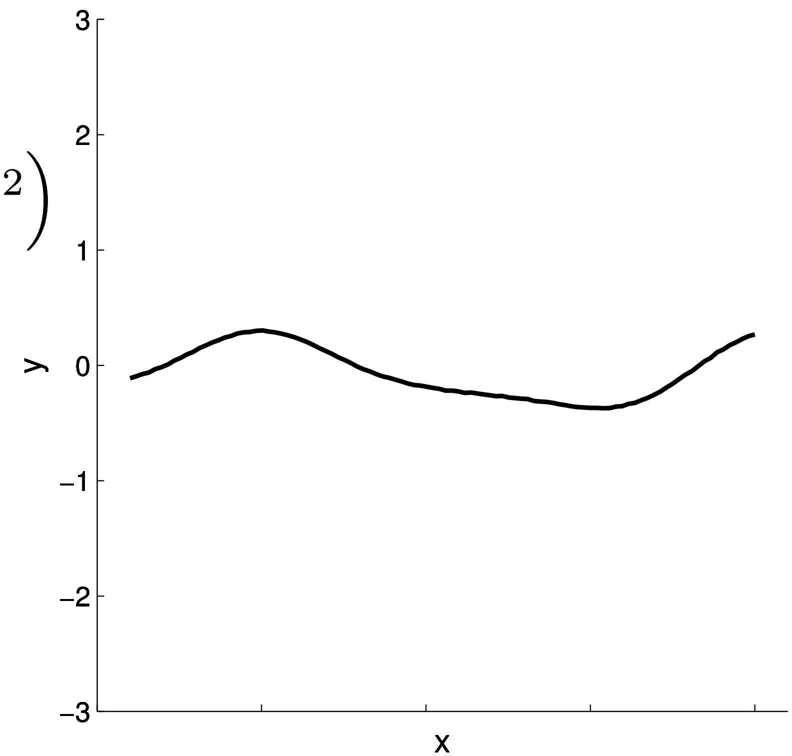
$$p(\mathsf{y}|\theta) = \mathcal{N}(0, \Sigma)$$

$$\Sigma(\mathsf{x}_1, \mathsf{x}_2) = \mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) + \mathsf{I}\sigma_{\mathsf{y}}^2$$

$$\mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) = \sigma^2 \exp\left(-\frac{1}{2\mathsf{l}^2}(\mathsf{x}_1 - \mathsf{x}_2)^2\right)$$

$$\Sigma =$$





Parametric model

$$\mathsf{y}(\mathsf{x}) = f(\mathsf{x}; \theta) + \sigma_{\mathsf{y}}\epsilon$$

$$\epsilon \sim \mathcal{N}(0, 1)$$

What effect do the hyper-parameters have?

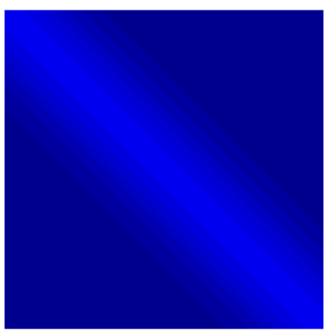
small vertical scale

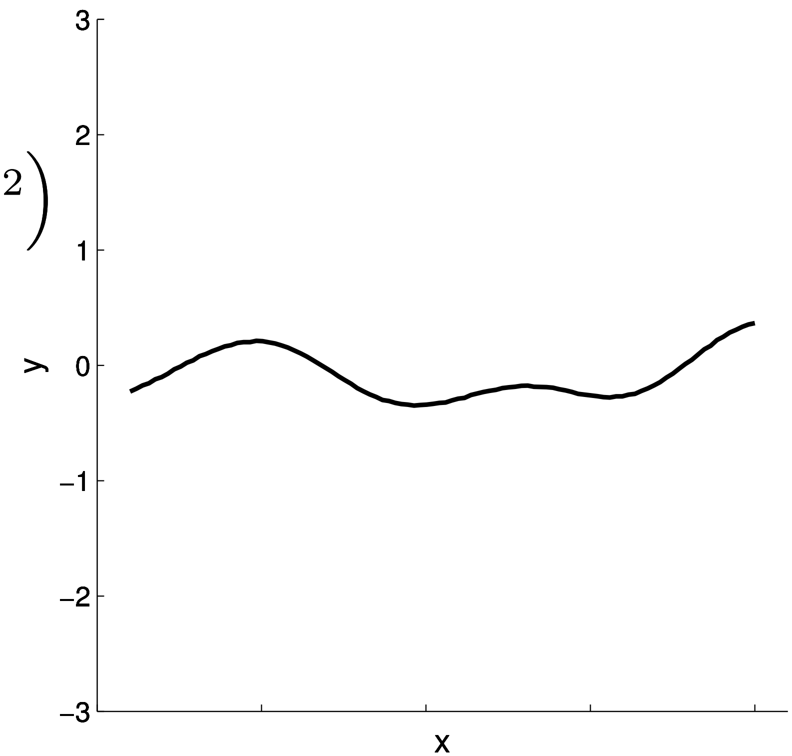
Non-parametric (∞-parametric)

$$p(\mathsf{y}|\theta) = \mathcal{N}(0, \Sigma)$$

$$\Sigma(\mathsf{x}_1, \mathsf{x}_2) = \mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) + \mathsf{I}\sigma_{\mathsf{y}}^2$$

$$\mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) = \sigma^2 \exp\left(-\frac{1}{2\mathsf{l}^2}(\mathsf{x}_1 - \mathsf{x}_2)^2\right)$$

$\Sigma =$ 



Parametric model

$$\mathsf{y}(\mathsf{x}) = f(\mathsf{x}; \theta) + \sigma_{\mathsf{y}}\epsilon$$

$$\epsilon \sim \mathcal{N}(0, 1)$$

What effect do the hyper-parameters have?

small vertical scale

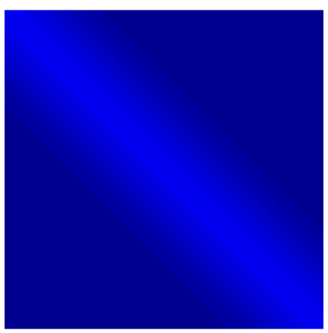
Non-parametric (∞-parametric)

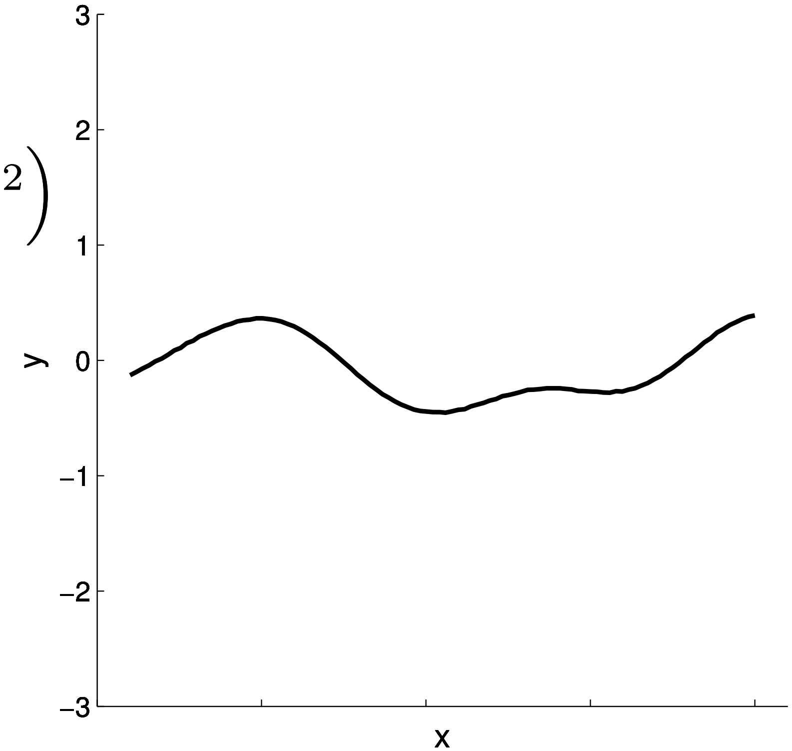
$$p(\mathsf{y}|\theta) = \mathcal{N}(0, \Sigma)$$

$$\Sigma(\mathsf{x}_1, \mathsf{x}_2) = \mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) + \mathsf{I}\sigma_{\mathsf{y}}^2$$

$$\mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) = \sigma^2 \exp\left(-\frac{1}{2\mathsf{l}^2}(\mathsf{x}_1 - \mathsf{x}_2)^2\right)$$

$$\Sigma =$$





Parametric model

$$\mathsf{y}(\mathsf{x}) = f(\mathsf{x}; \theta) + \sigma_{\mathsf{y}}\epsilon$$

$$\epsilon \sim \mathcal{N}(0, 1)$$

What effect do the hyper-parameters have?

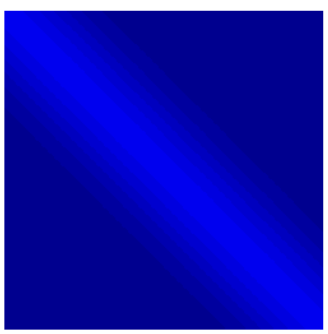
small vertical scale

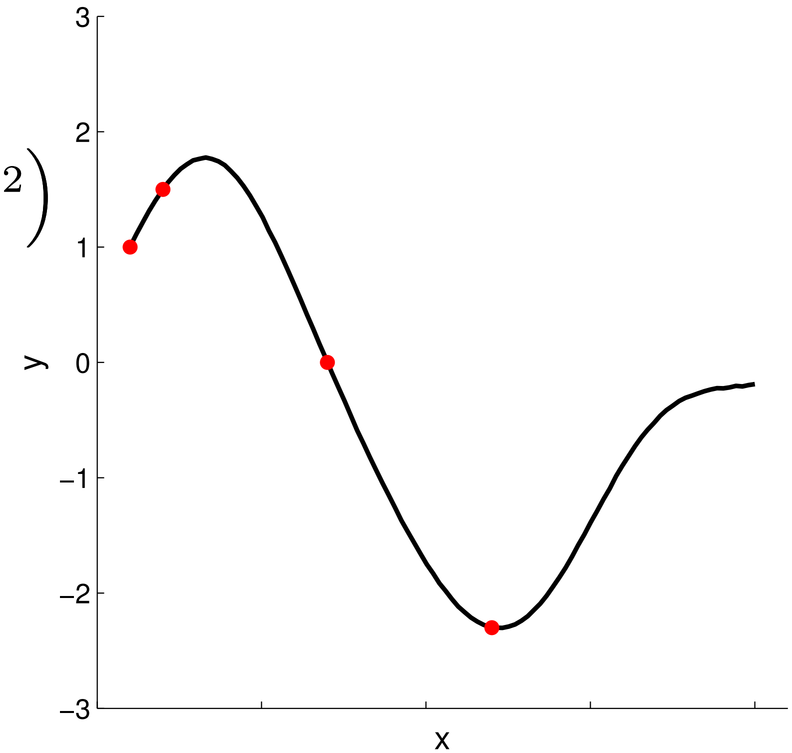
Non-parametric (∞-parametric)

$$p(\mathsf{y}|\theta) = \mathcal{N}(0, \Sigma)$$

$$\Sigma(\mathsf{x}_1, \mathsf{x}_2) = \mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) + \mathsf{I}\sigma_\mathsf{y}^2$$

$$\mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) = \sigma^2 \exp\left(-\frac{1}{2\mathsf{l}^2}(\mathsf{x}_1 - \mathsf{x}_2)^2\right)$$

$\Sigma =$ 



Parametric model

$$\mathsf{y}(\mathsf{x}) = f(\mathsf{x};\theta) + \sigma_\mathsf{y}\epsilon$$

$$\epsilon \sim \mathcal{N}(0, 1)$$

What effect do the hyper-parameters have?

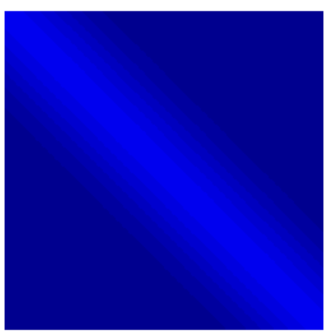
small vertical scale

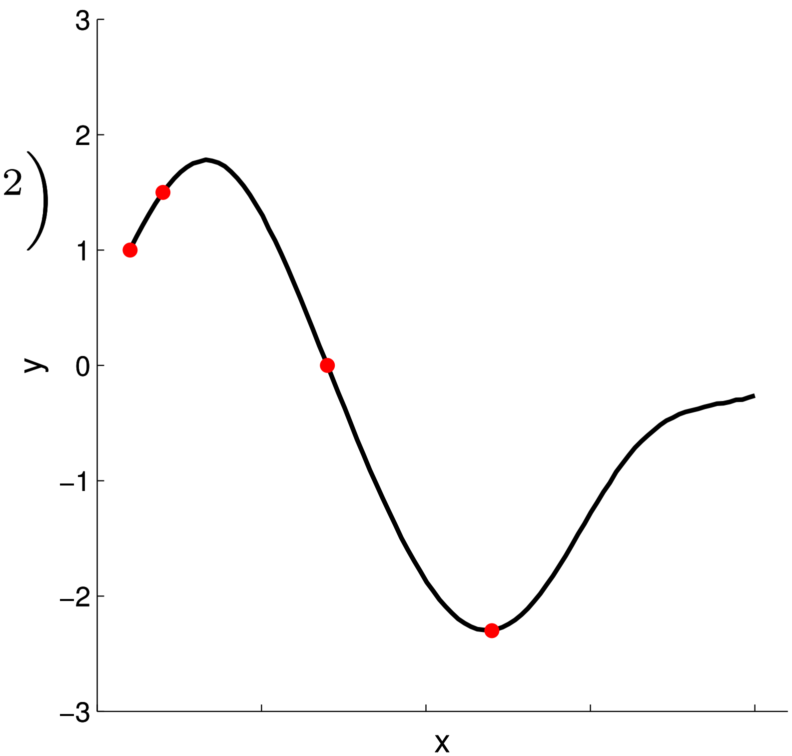
Non-parametric (∞-parametric)

$$p(\mathsf{y}|\theta) = \mathcal{N}(0, \Sigma)$$

$$\Sigma(\mathsf{x}_1, \mathsf{x}_2) = \mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) + \mathsf{I}\sigma_{\mathsf{y}}^2$$

$$\mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) = \sigma^2 \exp\left(-\frac{1}{2\mathsf{l}^2}(\mathsf{x}_1 - \mathsf{x}_2)^2\right)$$

$\Sigma =$ 



Parametric model

$$\mathsf{y}(\mathsf{x}) = f(\mathsf{x}; \theta) + \sigma_{\mathsf{y}}\epsilon$$

$$\epsilon \sim \mathcal{N}(0, 1)$$

What effect do the hyper-parameters have?

small vertical scale

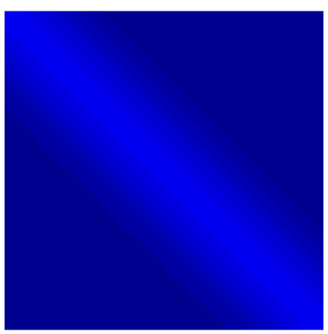
Non-parametric (∞-parametric)

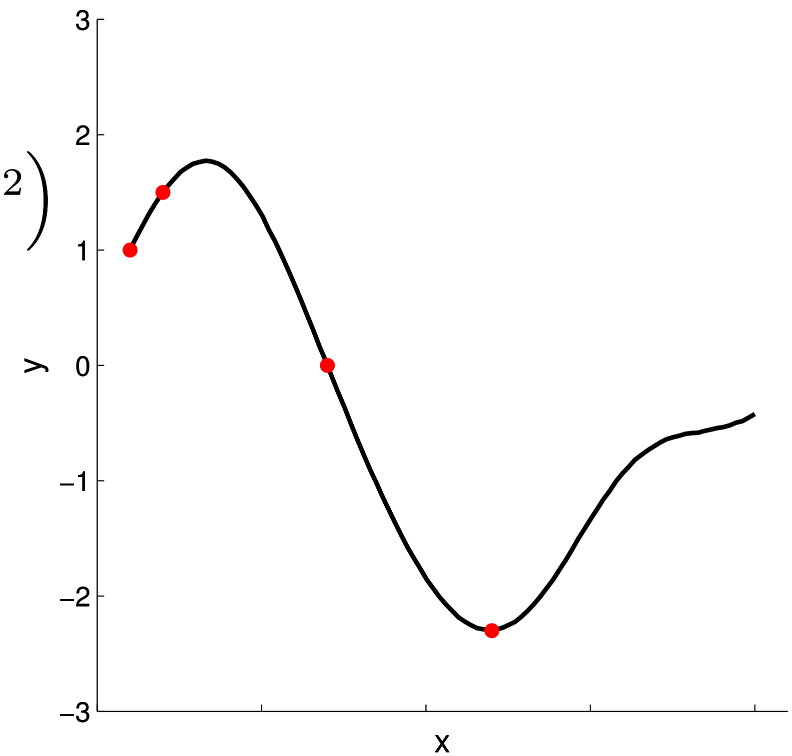
$$p(\mathsf{y}|\theta) = \mathcal{N}(0, \Sigma)$$

$$\Sigma(\mathsf{x}_1, \mathsf{x}_2) = \mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) + \mathsf{I}\sigma_\mathsf{y}^2$$

$$\mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) = \sigma^2 \exp\left(-\frac{1}{2\mathsf{l}^2}(\mathsf{x}_1 - \mathsf{x}_2)^2\right)$$

$$\Sigma =$$





Parametric model

$$\mathsf{y}(\mathsf{x}) = f(\mathsf{x}; \theta) + \sigma_\mathsf{y}\epsilon$$

$$\epsilon \sim \mathcal{N}(0, 1)$$

What effect do the hyper-parameters have?

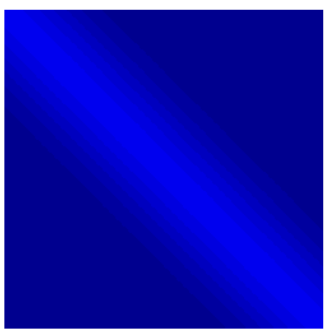
small vertical scale

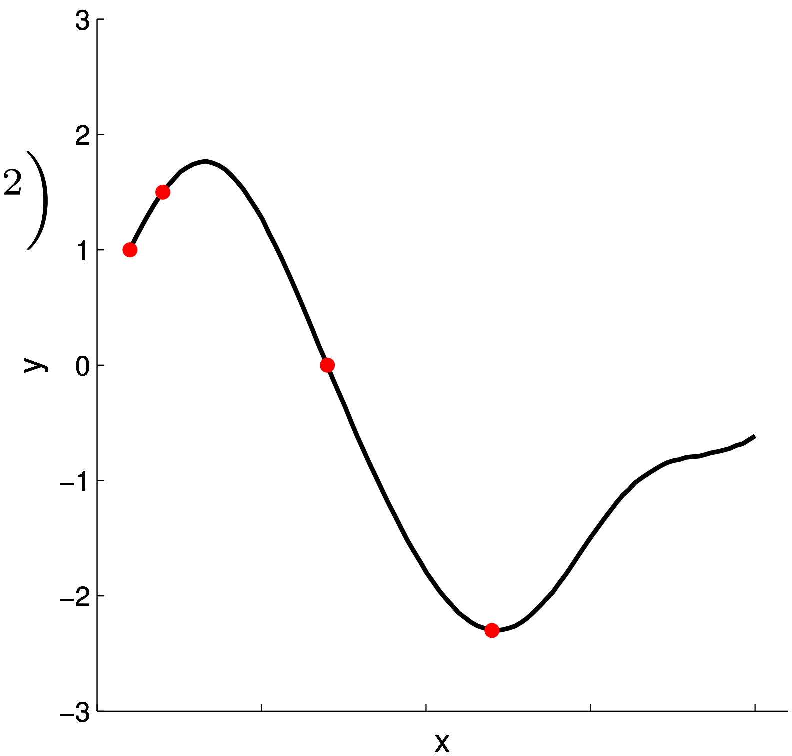
Non-parametric (∞-parametric)

$$p(\mathsf{y}|\theta) = \mathcal{N}(0, \Sigma)$$

$$\Sigma(\mathsf{x}_1, \mathsf{x}_2) = \mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) + \mathsf{I}\sigma_\mathsf{y}^2$$

$$\mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) = \sigma^2 \exp\left(-\frac{1}{2\mathsf{l}^2}(\mathsf{x}_1 - \mathsf{x}_2)^2\right)$$

$\Sigma =$ 



Parametric model

$$\mathsf{y}(\mathsf{x}) = f(\mathsf{x}; \theta) + \sigma_\mathsf{y}\epsilon$$

$$\epsilon \sim \mathcal{N}(0, 1)$$

What effect do the hyper-parameters have?

small vertical scale

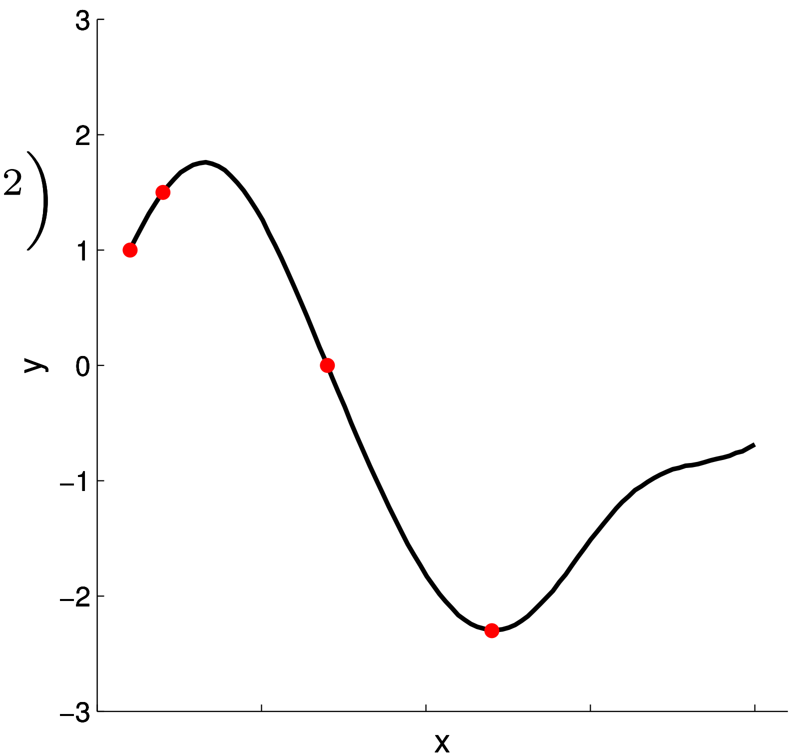
Non-parametric (∞-parametric)

$$p(\mathsf{y}|\theta) = \mathcal{N}(0, \Sigma)$$

$$\Sigma(\mathsf{x}_1, \mathsf{x}_2) = \mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) + \mathsf{I}\sigma_{\mathsf{y}}^2$$

$$\mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) = \sigma^2 \exp\left(-\frac{1}{2\mathsf{l}^2}(\mathsf{x}_1 - \mathsf{x}_2)^2\right)$$

$$\Sigma =$$



Parametric model

$$\mathsf{y}(\mathsf{x}) = f(\mathsf{x}; \theta) + \sigma_{\mathsf{y}}\epsilon$$

$$\epsilon \sim \mathcal{N}(0, 1)$$

What effect do the hyper-parameters have?

small vertical scale

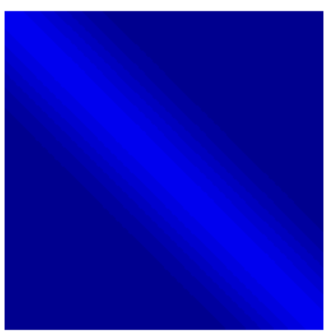
Non-parametric (∞-parametric)

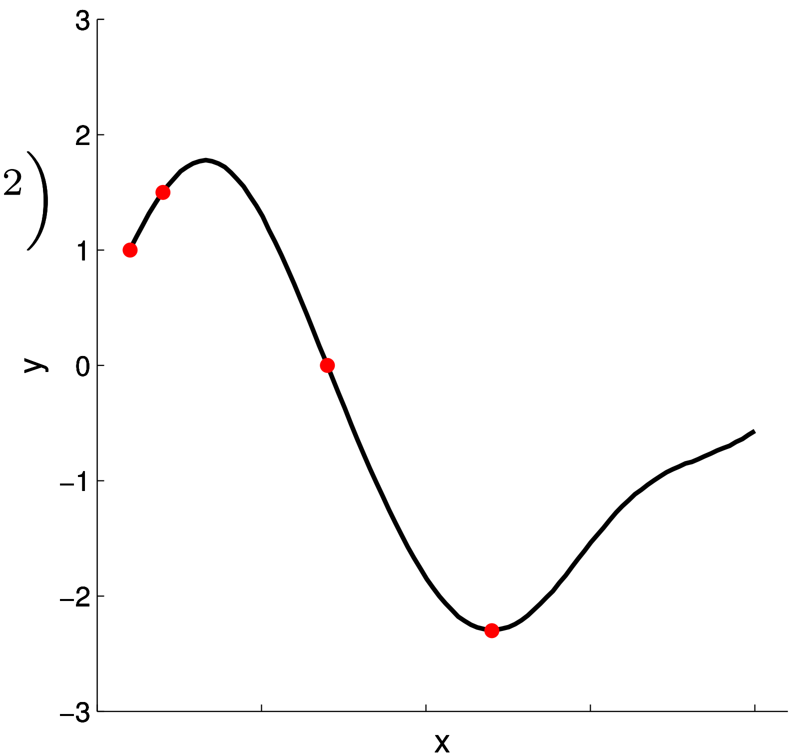
$$p(\mathsf{y}|\theta) = \mathcal{N}(0, \Sigma)$$

$$\Sigma(\mathsf{x}_1, \mathsf{x}_2) = \mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) + \mathsf{I}\sigma_{\mathsf{y}}^2$$

$$\mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) = \sigma^2 \exp\left(-\frac{1}{2\mathsf{l}^2}(\mathsf{x}_1 - \mathsf{x}_2)^2\right)$$

$$\Sigma =$$





Parametric model

$$\mathsf{y}(\mathsf{x}) = f(\mathsf{x};\theta) + \sigma_{\mathsf{y}}\epsilon$$

$$\epsilon \sim \mathcal{N}(0, 1)$$

What effect do the hyper-parameters have?

small vertical scale

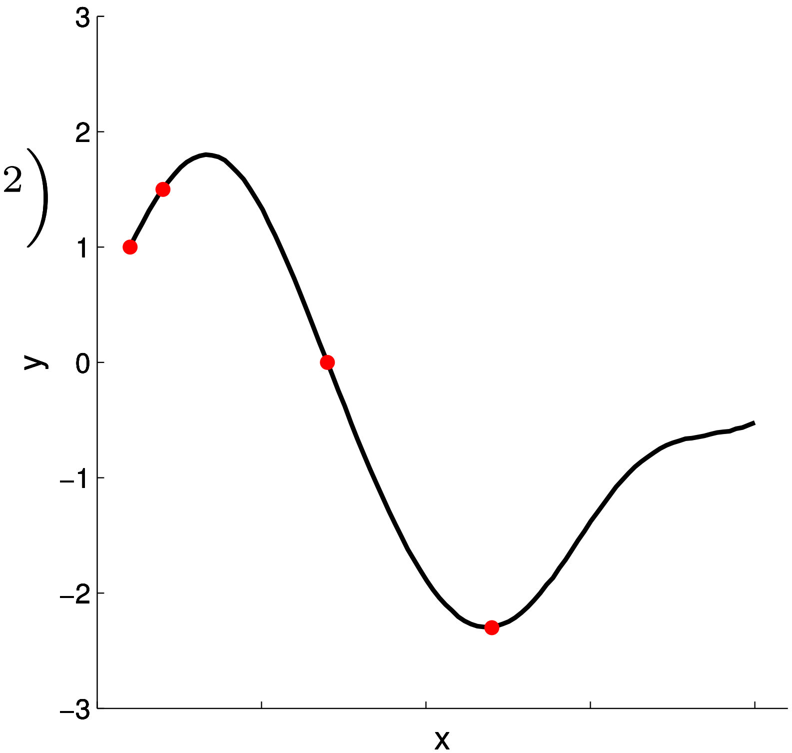
Non-parametric (∞-parametric)

$$p(\mathsf{y}|\theta) = \mathcal{N}(0, \Sigma)$$

$$\Sigma(\mathsf{x}_1, \mathsf{x}_2) = \mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) + \mathsf{I}\sigma_\mathsf{y}^2$$

$$\mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) = \sigma^2 \exp\left(-\frac{1}{2\mathsf{l}^2}(\mathsf{x}_1 - \mathsf{x}_2)^2\right)$$

$\Sigma =$ 



Parametric model

$$\mathsf{y}(\mathsf{x}) = f(\mathsf{x}; \theta) + \sigma_\mathsf{y}\epsilon$$

$$\epsilon \sim \mathcal{N}(0, 1)$$

What effect do the hyper-parameters have?

small vertical scale

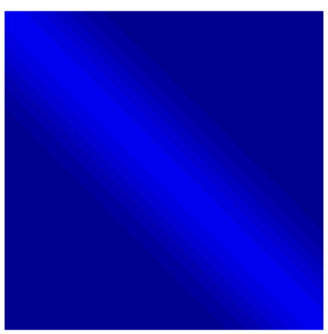
Non-parametric (∞-parametric)

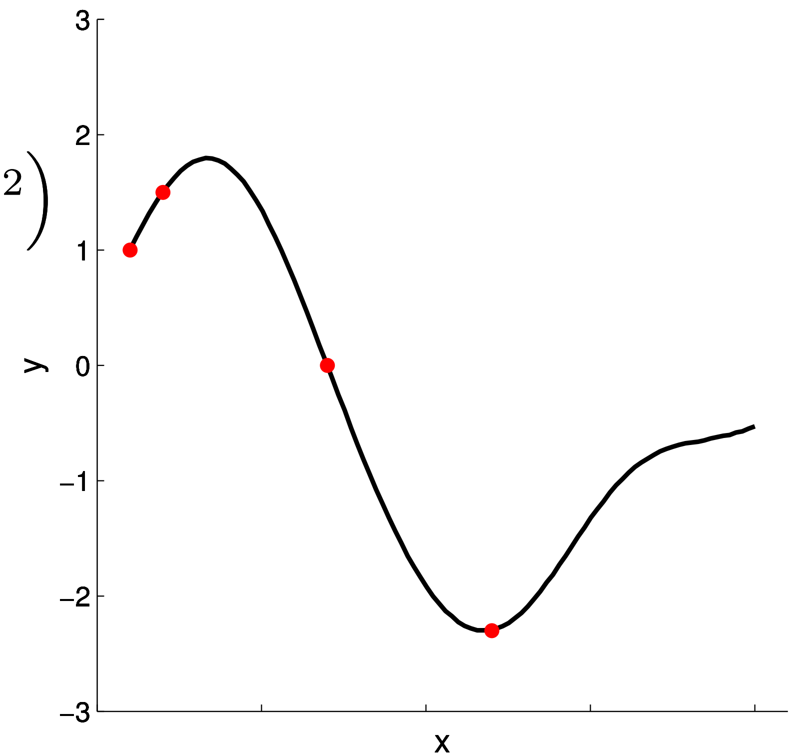
$$p(\mathsf{y}|\theta) = \mathcal{N}(0, \Sigma)$$

$$\Sigma(\mathsf{x}_1, \mathsf{x}_2) = \mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) + \mathsf{I}\sigma_{\mathsf{y}}^2$$

$$\mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) = \sigma^2 \exp\left(-\frac{1}{2\mathsf{l}^2}(\mathsf{x}_1 - \mathsf{x}_2)^2\right)$$

$$\Sigma =$$





Parametric model

$$\mathsf{y}(\mathsf{x}) = f(\mathsf{x}; \theta) + \sigma_{\mathsf{y}}\epsilon$$

$$\epsilon \sim \mathcal{N}(0, 1)$$

What effect do the hyper-parameters have?

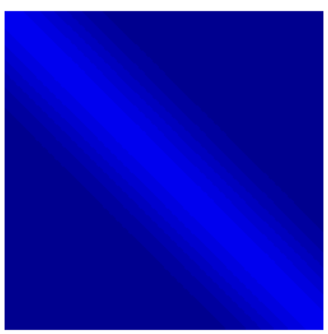
small vertical scale

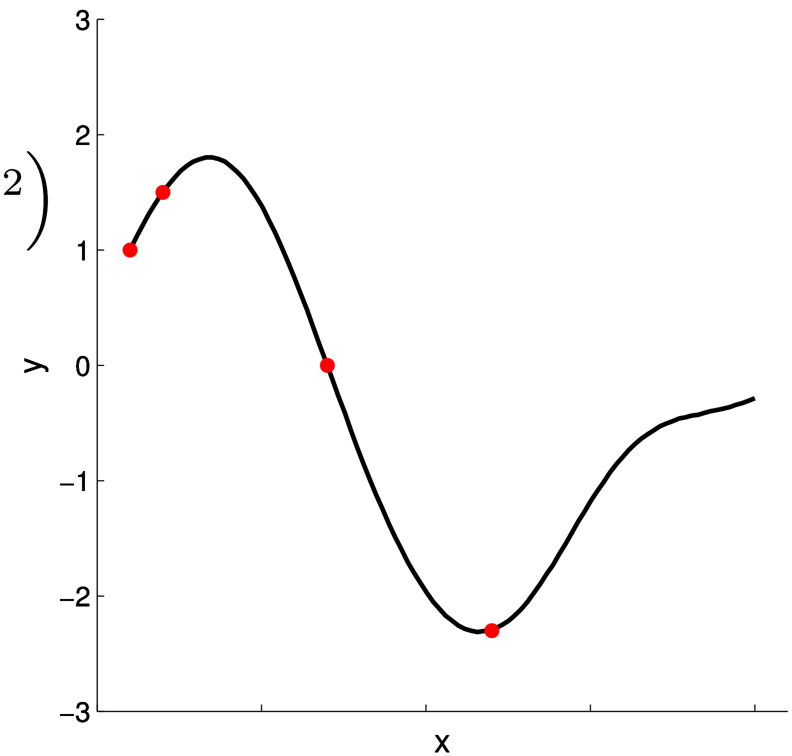
Non-parametric (∞-parametric)

$$p(\mathsf{y}|\theta) = \mathcal{N}(0, \Sigma)$$

$$\Sigma(\mathsf{x}_1, \mathsf{x}_2) = \mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) + \mathsf{I}\sigma_{\mathsf{y}}^2$$

$$\mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) = \sigma^2 \exp\left(-\frac{1}{2\mathsf{l}^2}(\mathsf{x}_1 - \mathsf{x}_2)^2\right)$$

$\Sigma =$ 



Parametric model

$$\mathsf{y}(\mathsf{x}) = f(\mathsf{x}; \theta) + \sigma_{\mathsf{y}}\epsilon$$

$$\epsilon \sim \mathcal{N}(0, 1)$$

What effect do the hyper-parameters have?

small vertical scale

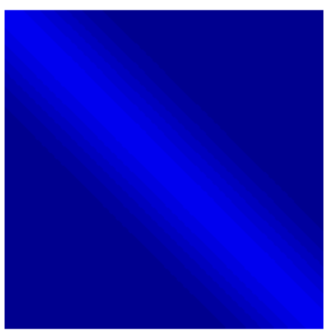
Non-parametric (∞-parametric)

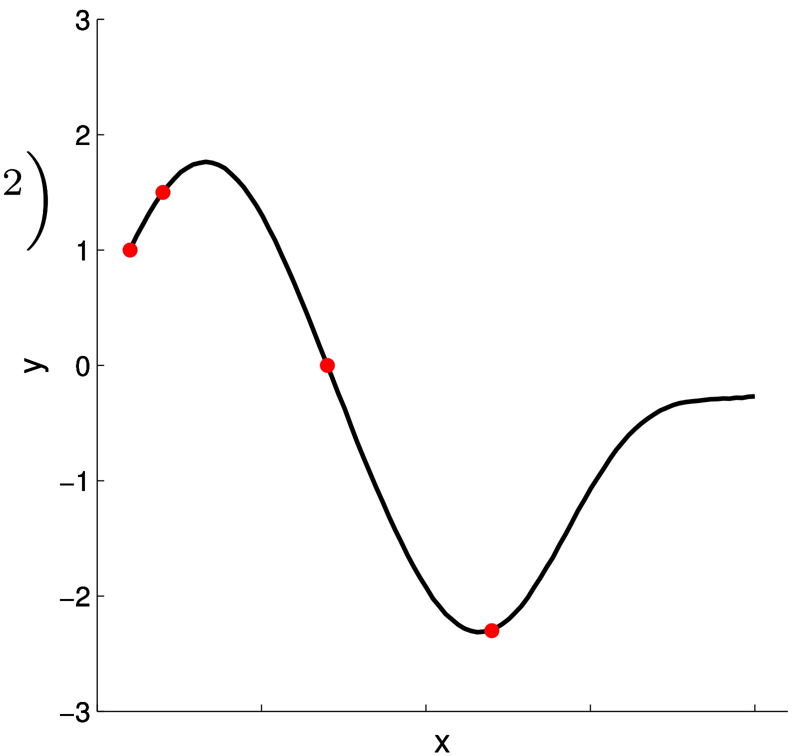
$$p(\mathsf{y}|\theta) = \mathcal{N}(0, \Sigma)$$

$$\Sigma(\mathsf{x}_1, \mathsf{x}_2) = \mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) + \mathsf{I}\sigma_{\mathsf{y}}^2$$

$$\mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) = \sigma^2 \exp\left(-\frac{1}{2\mathsf{l}^2}(\mathsf{x}_1 - \mathsf{x}_2)^2\right)$$

$$\Sigma =$$





Parametric model

$$\mathsf{y}(\mathsf{x}) = f(\mathsf{x}; \theta) + \sigma_{\mathsf{y}}\epsilon$$

$$\epsilon \sim \mathcal{N}(0, 1)$$

What effect do the hyper-parameters have?

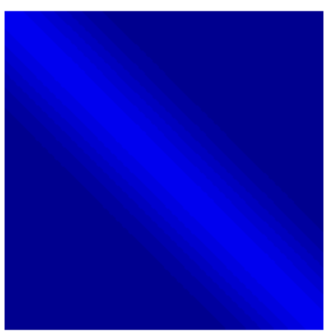
small vertical scale

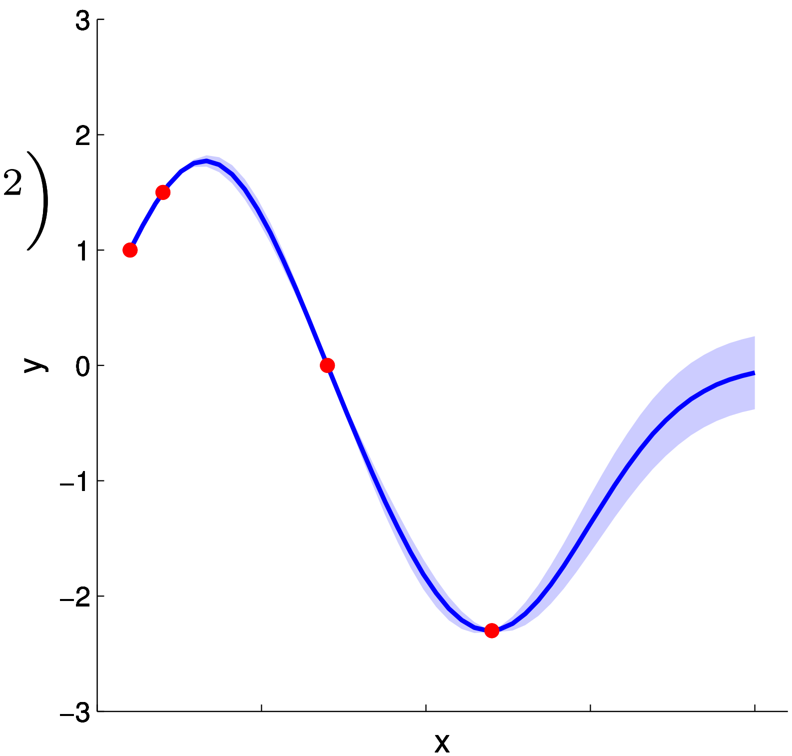
Non-parametric (∞-parametric)

$$p(\mathsf{y}|\theta) = \mathcal{N}(0, \Sigma)$$

$$\Sigma(\mathsf{x}_1, \mathsf{x}_2) = \mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) + \mathsf{I}\sigma_{\mathsf{y}}^2$$

$$\mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) = \sigma^2 \exp\left(-\frac{1}{2\mathsf{l}^2}(\mathsf{x}_1 - \mathsf{x}_2)^2\right)$$

$\Sigma =$ 



Parametric model

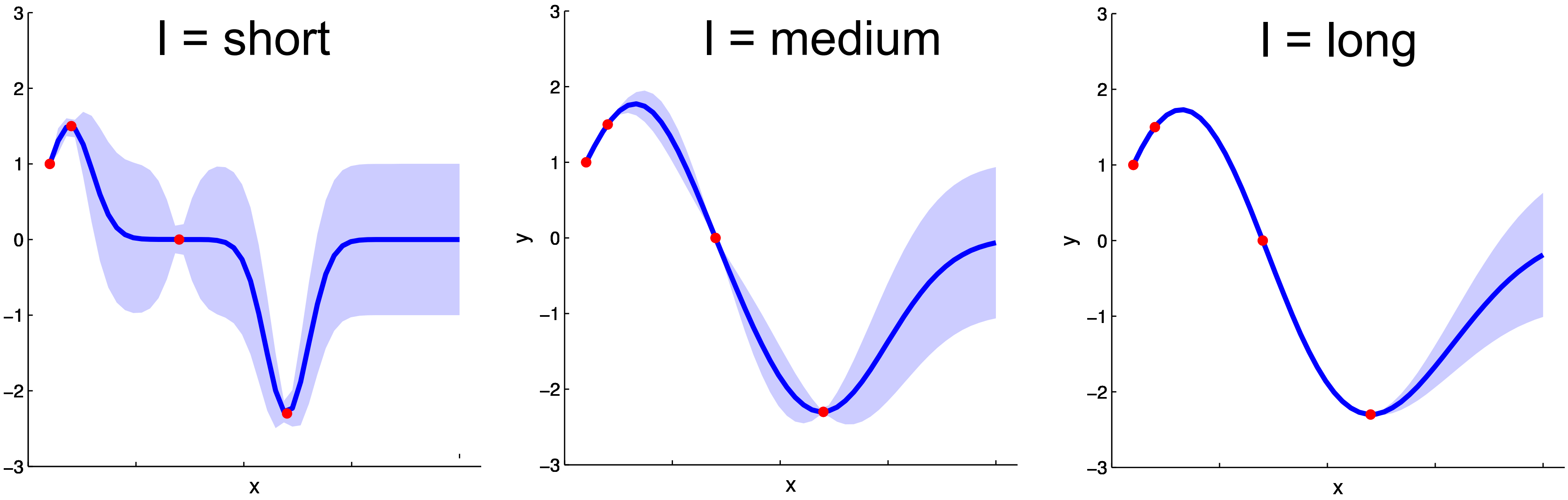
$$\mathsf{y}(\mathsf{x}) = f(\mathsf{x}; \theta) + \sigma_{\mathsf{y}}\epsilon$$

$$\epsilon \sim \mathcal{N}(0, 1)$$

What effect do the hyper-parameters have?

$$\mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) = \sigma^2 \exp\left(-\frac{1}{2\mathsf{l}^2}(\mathsf{x}_1 - \mathsf{x}_2)^2\right)$$

- Hyper-parameters have a strong effect
  - l controls the horizontal length-scale
  - σ^2 controls the vertical scale of the data
- $\implies$ we need automatic ways of learning the hyper-parameters from data



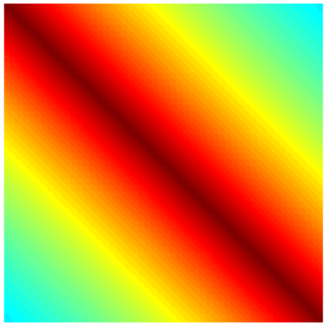
What effect does the form of the covariance function have?

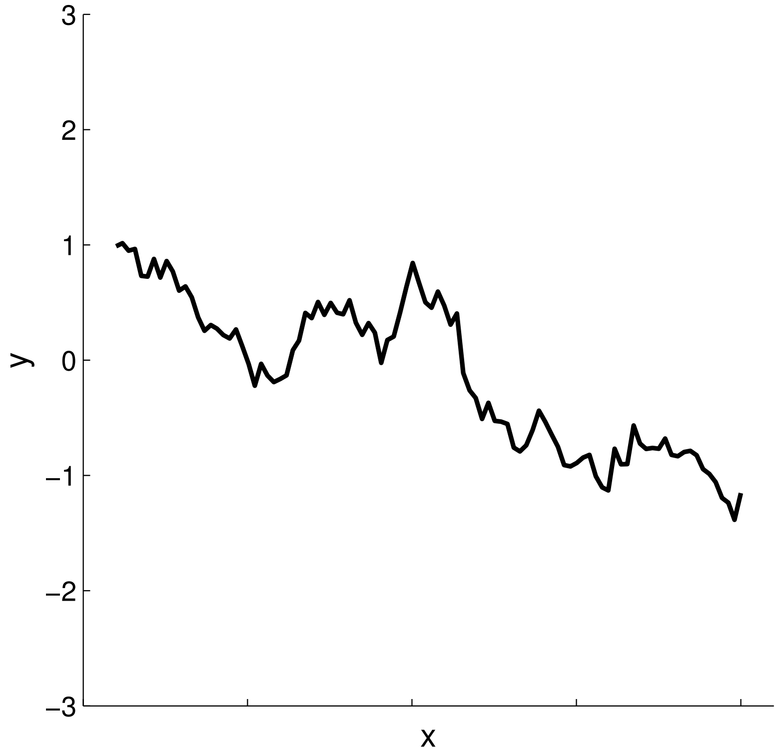
$$K(x_1, x_2) = \sigma^2 \exp\left(-\frac{1}{2l^2}|x_1 - x_2|\right)$$

Laplacian covariance function

Browninan motion

Ornstein-Uhlenbeck

$\Sigma =$ 



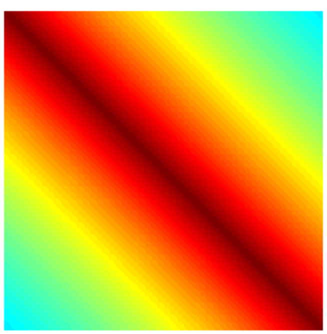
What effect does the form of the covariance function have?

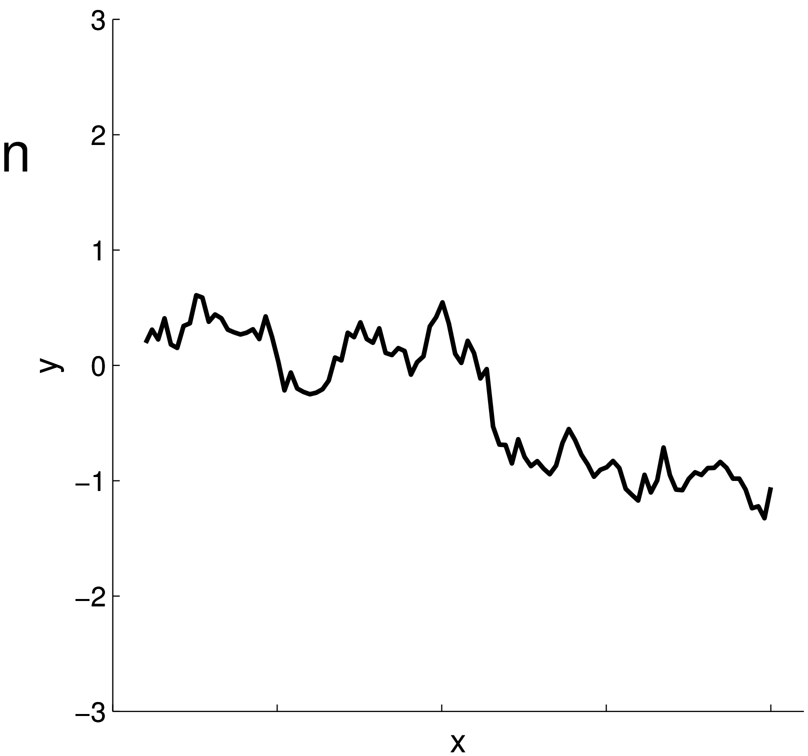
$$K(x_1, x_2) = \sigma^2 \exp\left(-\frac{1}{2l^2}|x_1 - x_2|\right)$$

Laplacian covariance function

Browninan motion

Ornstein-Uhlenbeck

$\Sigma =$ 



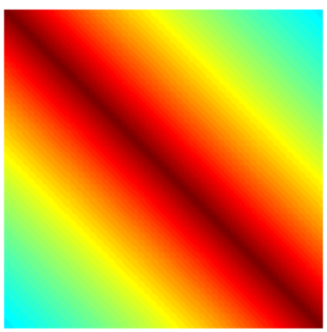
What effect does the form of the covariance function have?

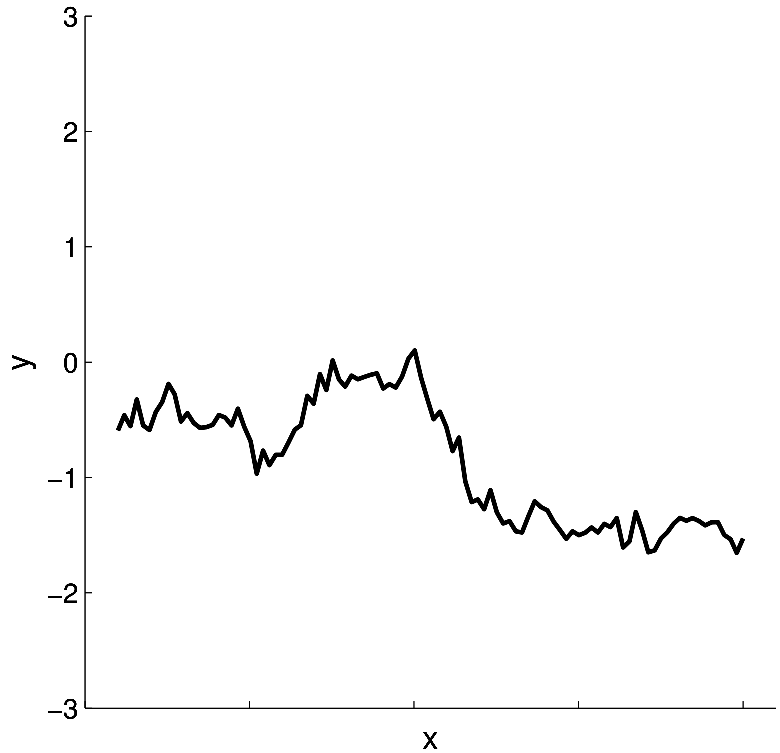
$$\mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) = \sigma^2 \exp\left(-\frac{1}{2\mathsf{l}^2}|\mathsf{x}_1 - \mathsf{x}_2|\right)$$

Laplacian covariance function

Browninan motion

Ornstein-Uhlenbeck

$\Sigma =$ 



What effect does the form of the covariance function have?

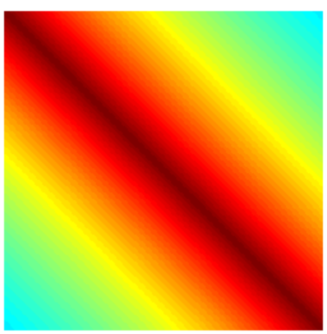
$$K(x_1, x_2) = \sigma^2 \exp\left(-\frac{1}{2l^2}|x_1 - x_2|\right)$$

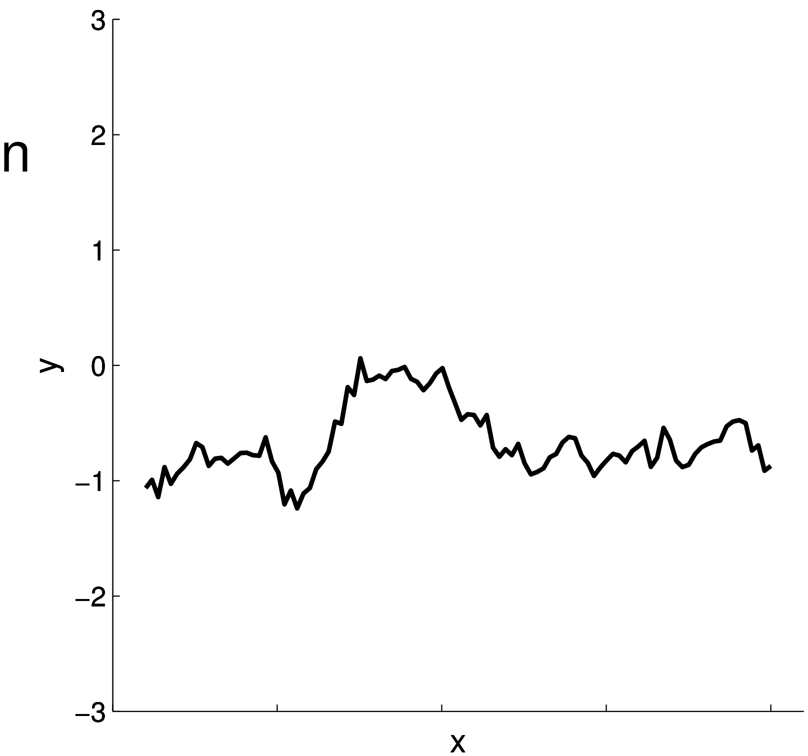
Laplacian covariance function

Browninan motion

Ornstein-Uhlenbeck

$\Sigma =$





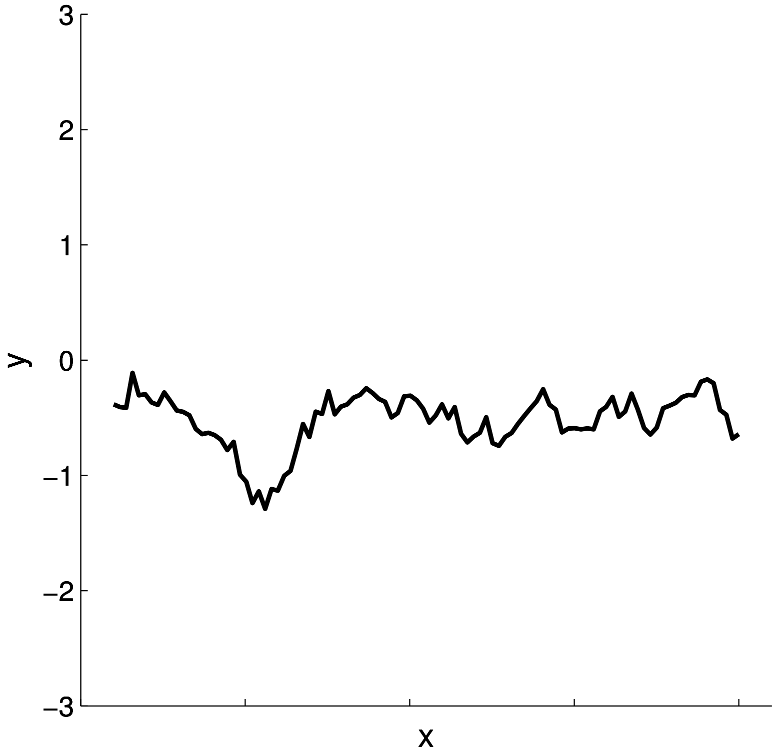
What effect does the form of the covariance function have?

$$K(x_1, x_2) = \sigma^2 \exp\left(-\frac{1}{2l^2}|x_1 - x_2|\right)$$

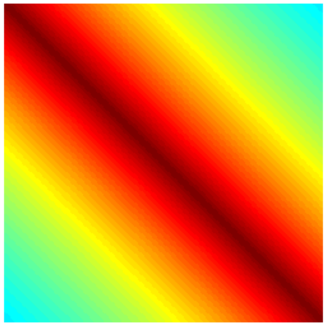
Laplacian covariance function

Browninan motion

Ornstein-Uhlenbeck



$\Sigma =$



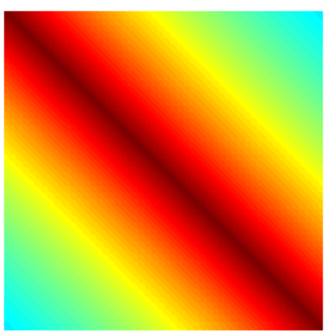
What effect does the form of the covariance function have?

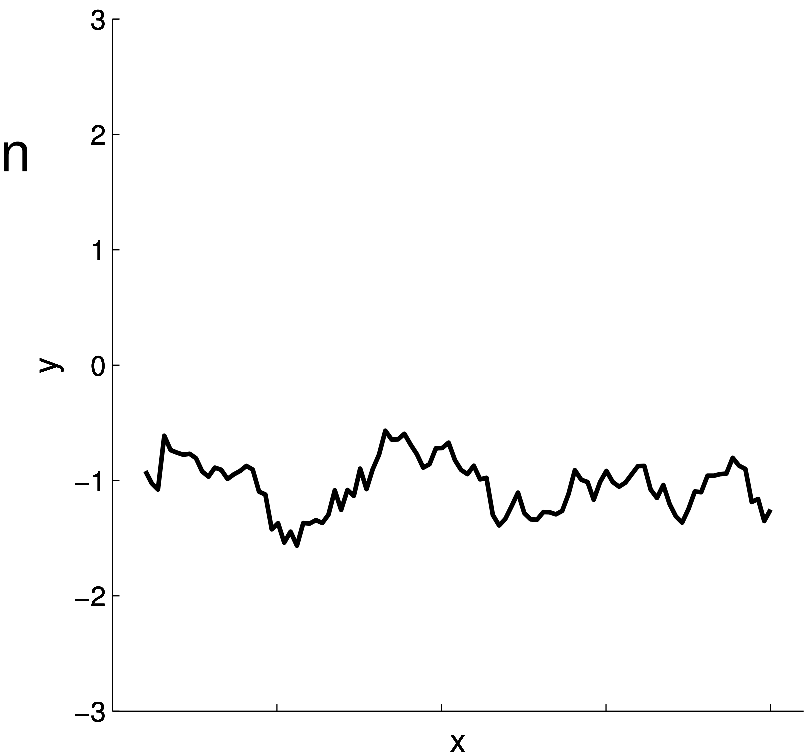
$$K(x_1, x_2) = \sigma^2 \exp\left(-\frac{1}{2l^2}|x_1 - x_2|\right)$$

Laplacian covariance function

Browninan motion

Ornstein-Uhlenbeck

$\Sigma =$ 



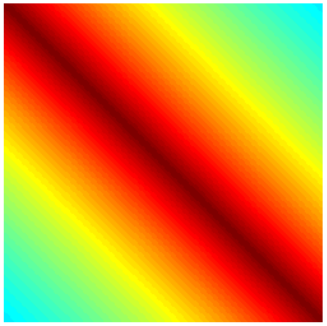
What effect does the form of the covariance function have?

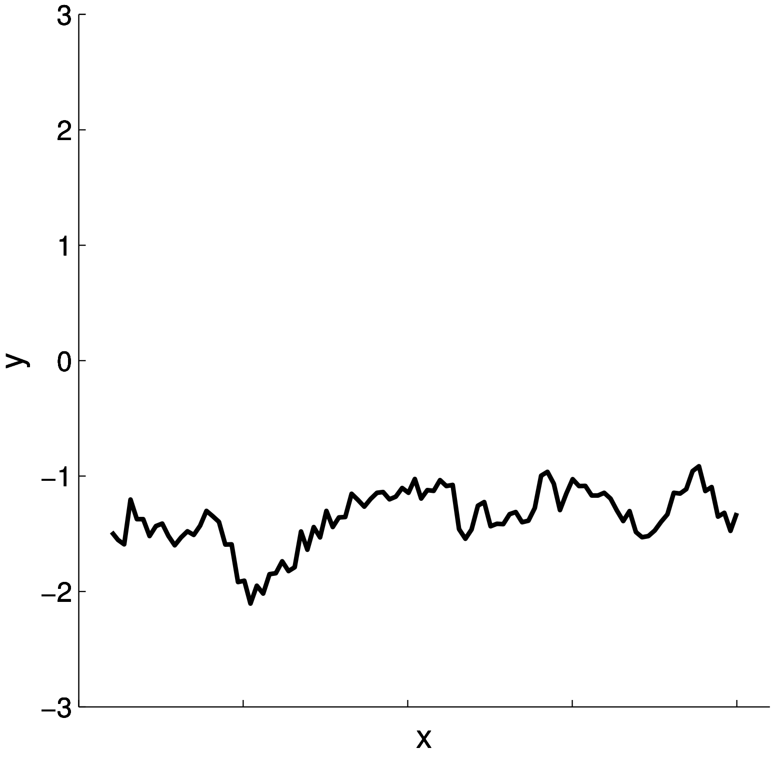
$$K(x_1, x_2) = \sigma^2 \exp\left(-\frac{1}{2l^2}|x_1 - x_2|\right)$$

Laplacian covariance function

Browninan motion

Ornstein-Uhlenbeck

$\Sigma =$ 



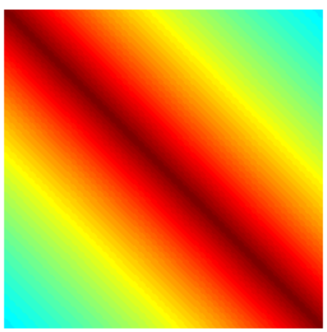
What effect does the form of the covariance function have?

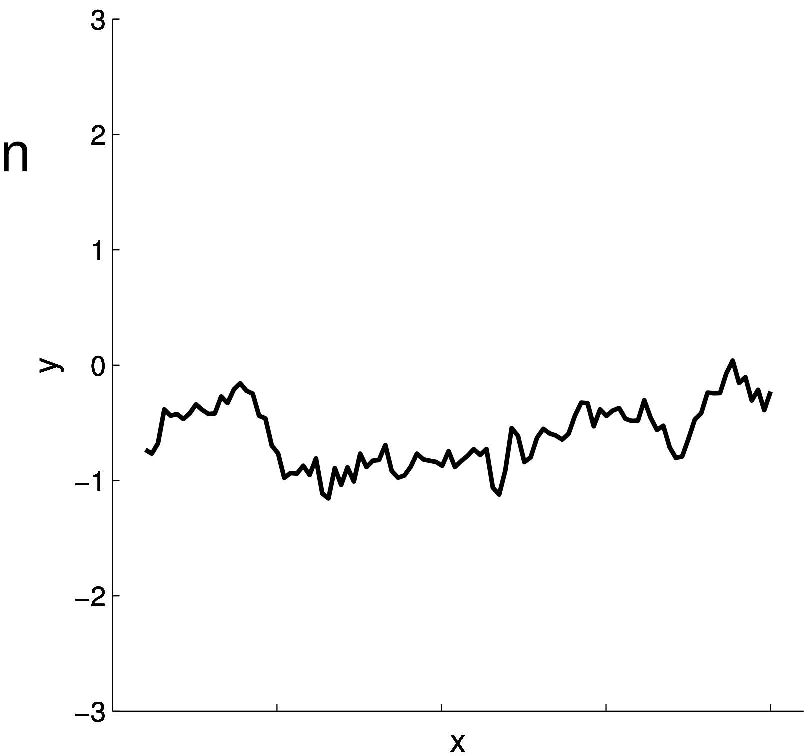
$$K(x_1, x_2) = \sigma^2 \exp\left(-\frac{1}{2l^2}|x_1 - x_2|\right)$$

Laplacian covariance function

Browninan motion

Ornstein-Uhlenbeck

$\Sigma =$ 



What effect does the form of the covariance function have?

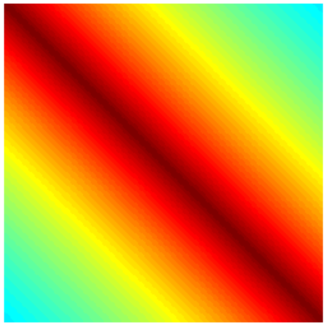
$$\mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) = \sigma^2 \exp\left(-\frac{1}{2\mathsf{l}^2}|\mathsf{x}_1 - \mathsf{x}_2|\right)$$

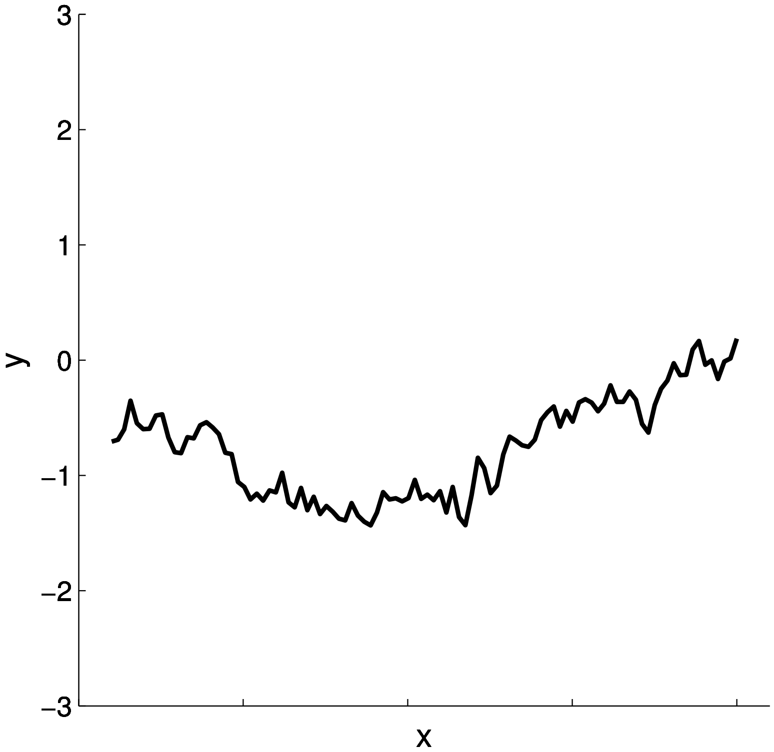
Laplacian covariance function

Browninan motion

Ornstein-Uhlenbeck

$\Sigma =$





What effect does the form of the covariance function have?

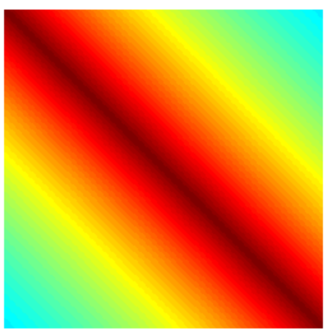
$$K(x_1, x_2) = \sigma^2 \exp\left(-\frac{1}{2l^2}|x_1 - x_2|\right)$$

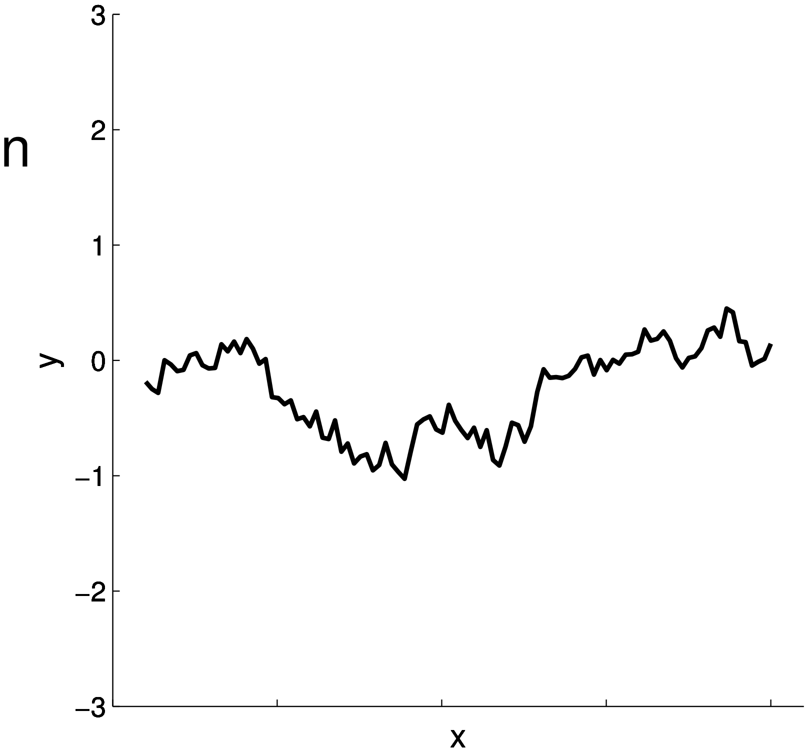
Laplacian covariance function

Browninan motion

Ornstein-Uhlenbeck

$\Sigma =$





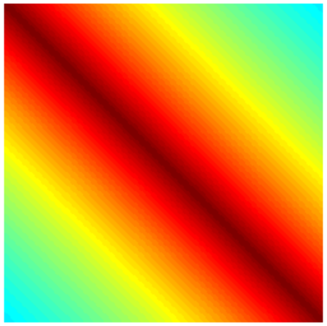
What effect does the form of the covariance function have?

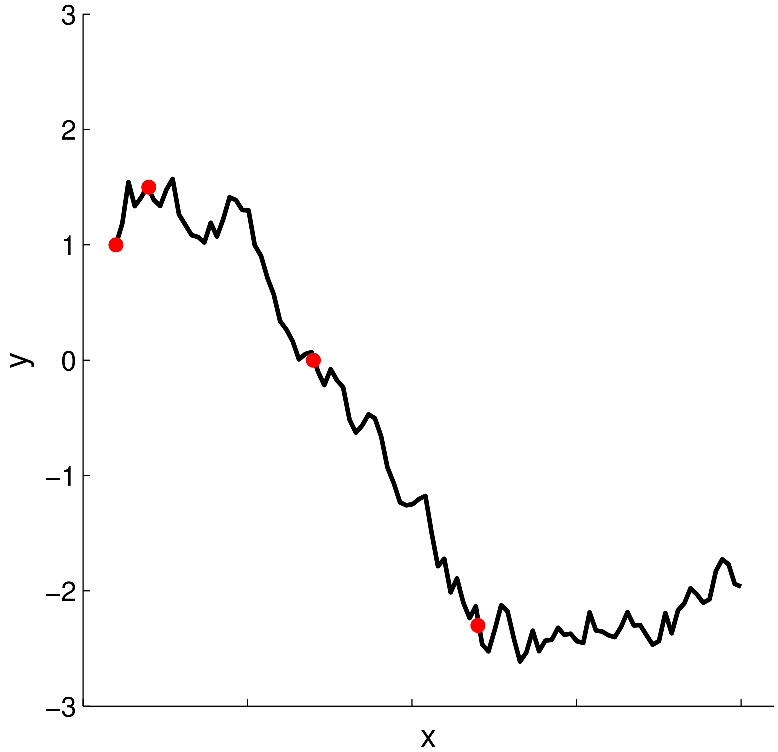
$$K(x_1, x_2) = \sigma^2 \exp\left(-\frac{1}{2l^2}|x_1 - x_2|\right)$$

Laplacian covariance function

Browninan motion

Ornstein-Uhlenbeck

$\Sigma =$ 



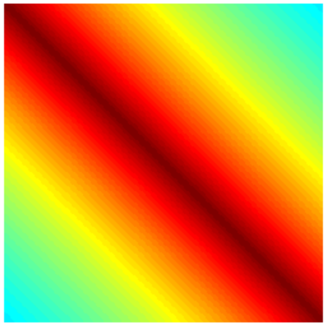
What effect does the form of the covariance function have?

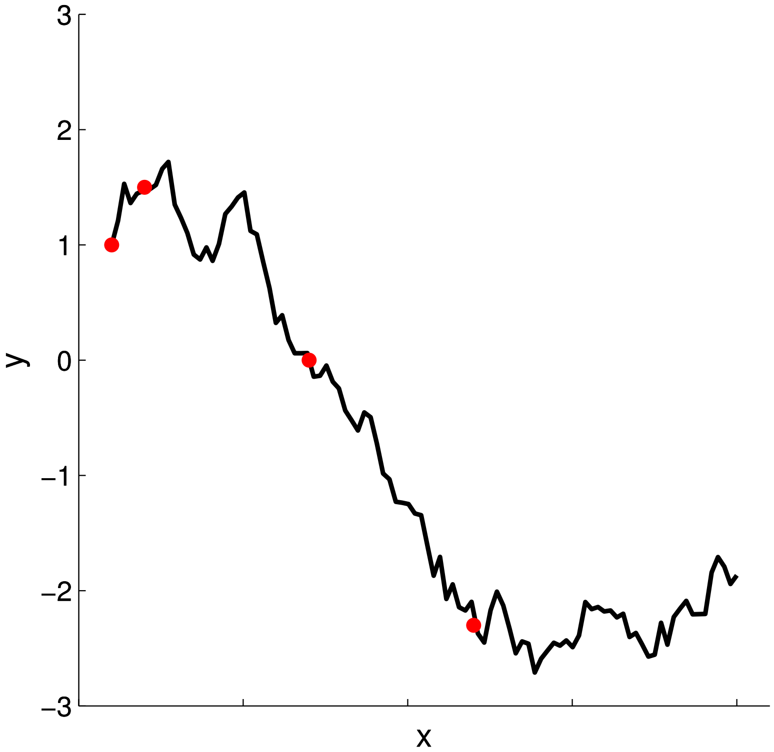
$$\mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) = \sigma^2 \exp\left(-\frac{1}{2\mathsf{l}^2}|\mathsf{x}_1 - \mathsf{x}_2|\right)$$

Laplacian covariance function

Browninan motion

Ornstein-Uhlenbeck

$\Sigma =$ 



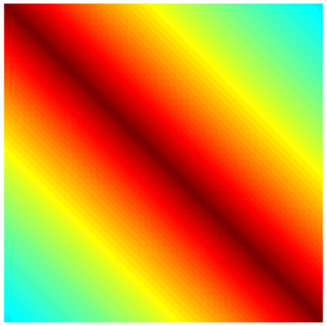
What effect does the form of the covariance function have?

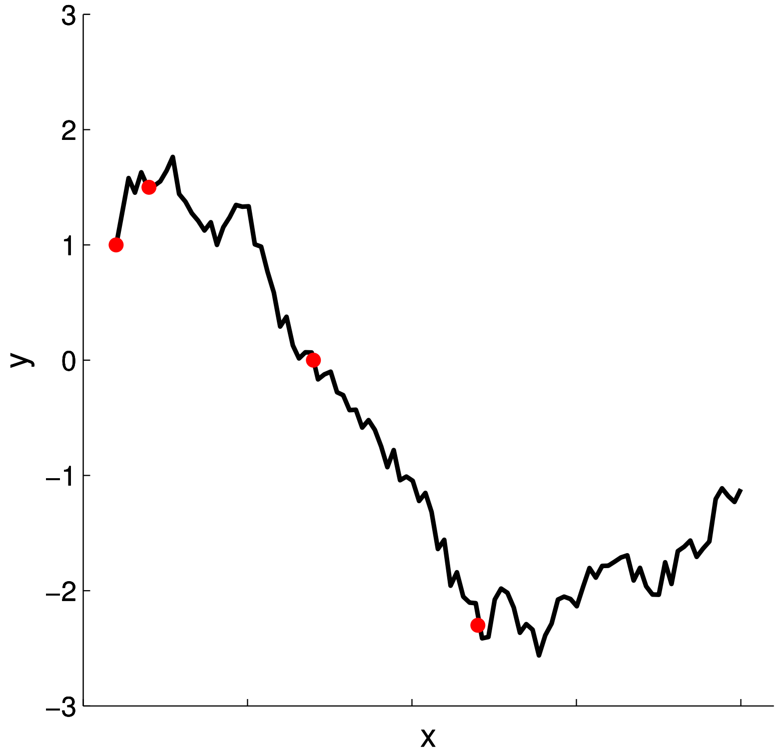
$$K(x_1, x_2) = \sigma^2 \exp\left(-\frac{1}{2l^2}|x_1 - x_2|\right)$$

Laplacian covariance function

Browninan motion

Ornstein-Uhlenbeck

$\Sigma =$ 



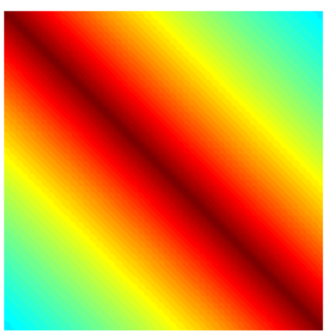
What effect does the form of the covariance function have?

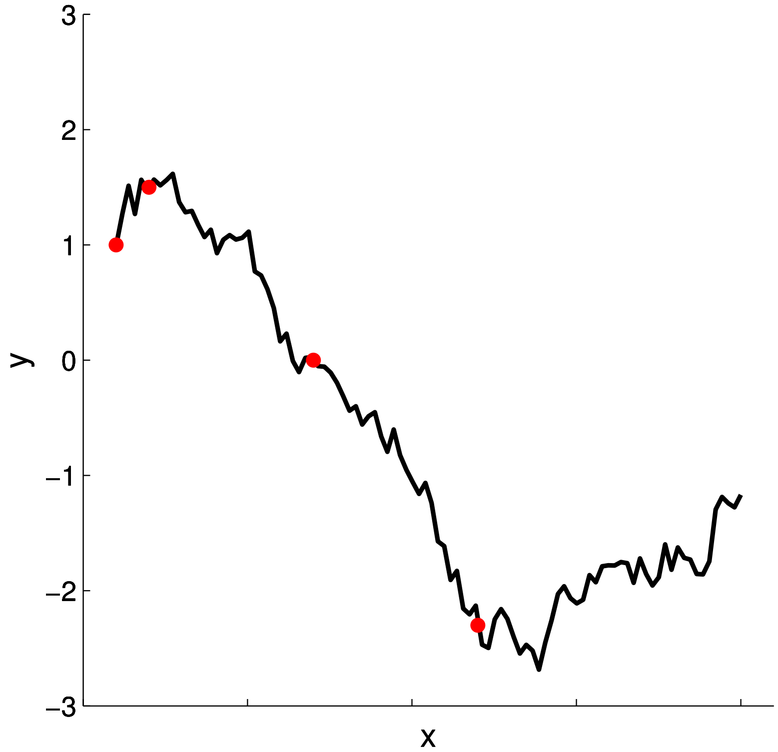
$$K(x_1, x_2) = \sigma^2 \exp\left(-\frac{1}{2l^2}|x_1 - x_2|\right)$$

Laplacian covariance function

Browninan motion

Ornstein-Uhlenbeck

$\Sigma =$ 



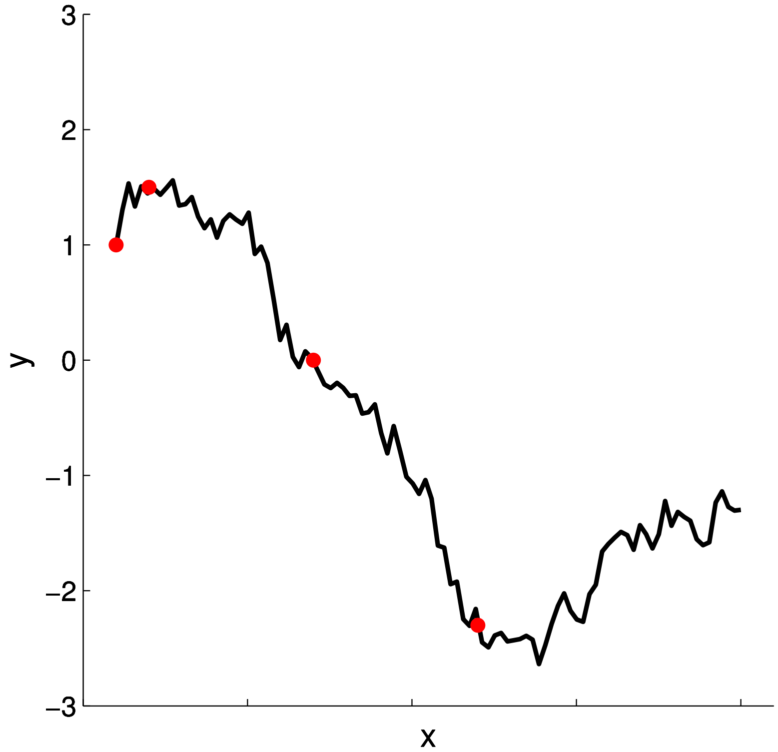
What effect does the form of the covariance function have?

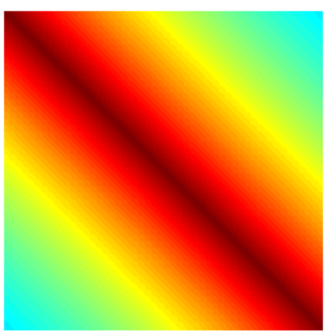
$$\mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) = \sigma^2 \exp\left(-\frac{1}{2\mathsf{l}^2}|\mathsf{x}_1 - \mathsf{x}_2|\right)$$

Laplacian covariance function

Browninan motion

Ornstein-Uhlenbeck



$\Sigma =$ 

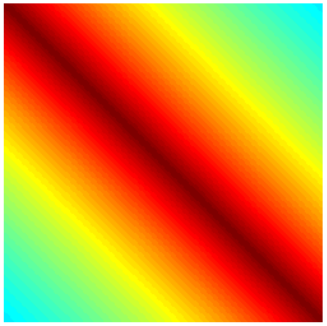
What effect does the form of the covariance function have?

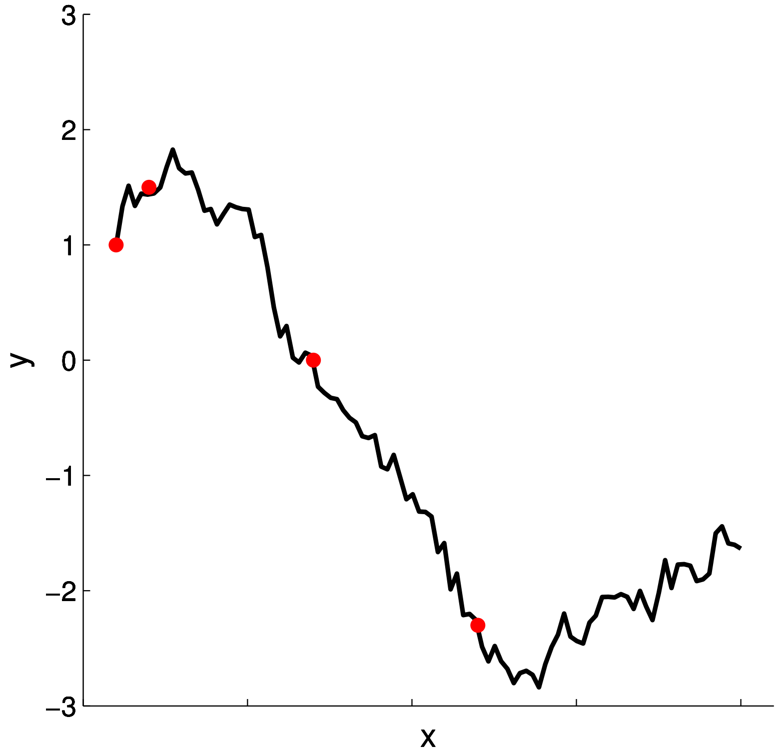
$$K(x_1, x_2) = \sigma^2 \exp\left(-\frac{1}{2l^2}|x_1 - x_2|\right)$$

Laplacian covariance function

Browninan motion

Ornstein-Uhlenbeck

$\Sigma =$ 



What effect does the form of the covariance function have?

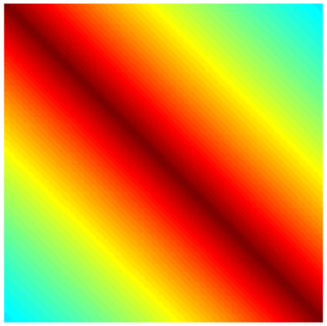
$$\mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) = \sigma^2 \exp\left(-\frac{1}{2\mathsf{l}^2}|\mathsf{x}_1 - \mathsf{x}_2|\right)$$

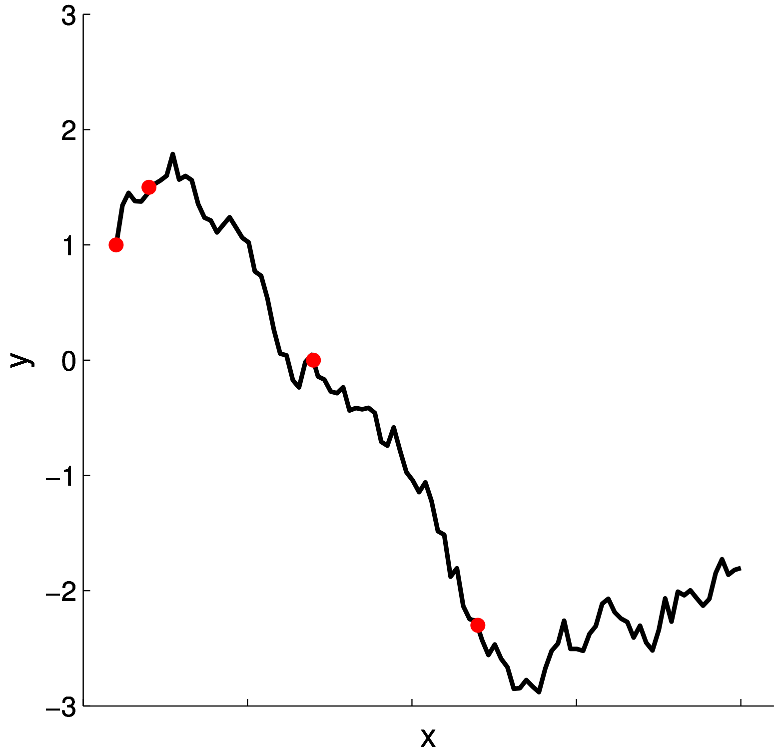
Laplacian covariance function

Browninan motion

Ornstein-Uhlenbeck

$\Sigma =$





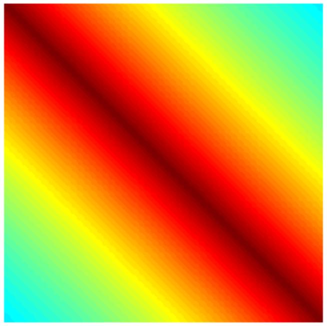
What effect does the form of the covariance function have?

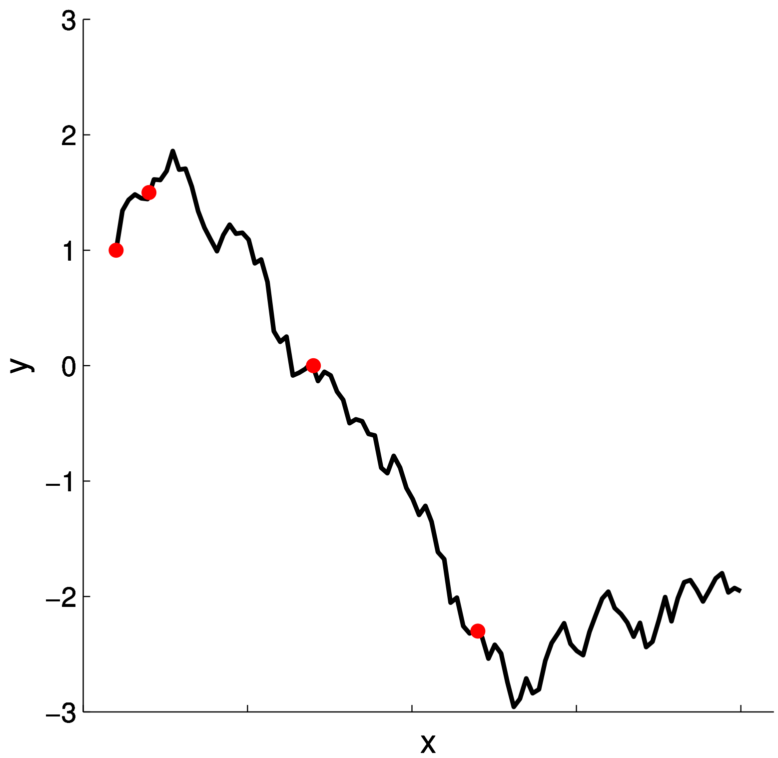
$$\mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) = \sigma^2 \exp\left(-\frac{1}{2\mathsf{l}^2}|\mathsf{x}_1 - \mathsf{x}_2|\right)$$

Laplacian covariance function

Browninan motion

Ornstein-Uhlenbeck

$\Sigma =$ 



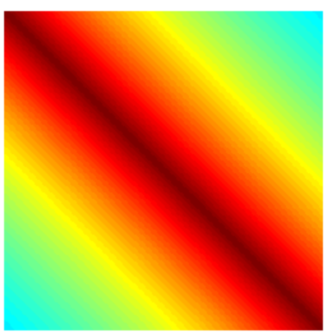
What effect does the form of the covariance function have?

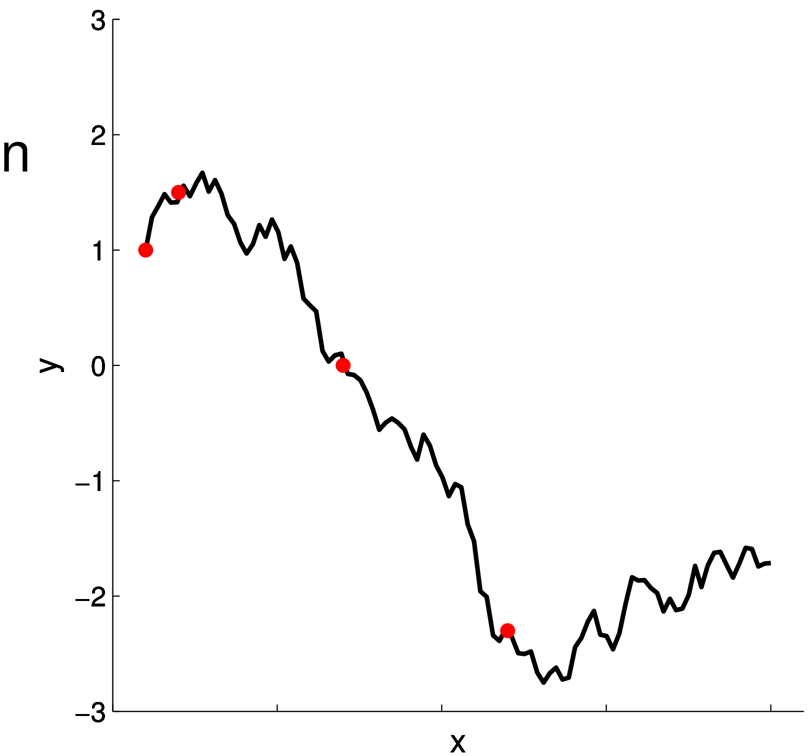
$$\mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) = \sigma^2 \exp\left(-\frac{1}{2\mathsf{l}^2}|\mathsf{x}_1 - \mathsf{x}_2|\right)$$

Laplacian covariance function

Browninan motion

Ornstein-Uhlenbeck

$\Sigma =$ 



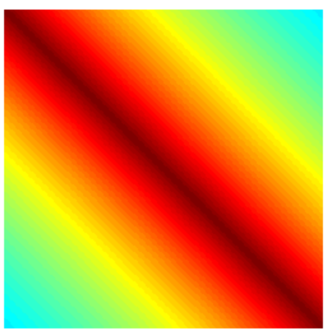
What effect does the form of the covariance function have?

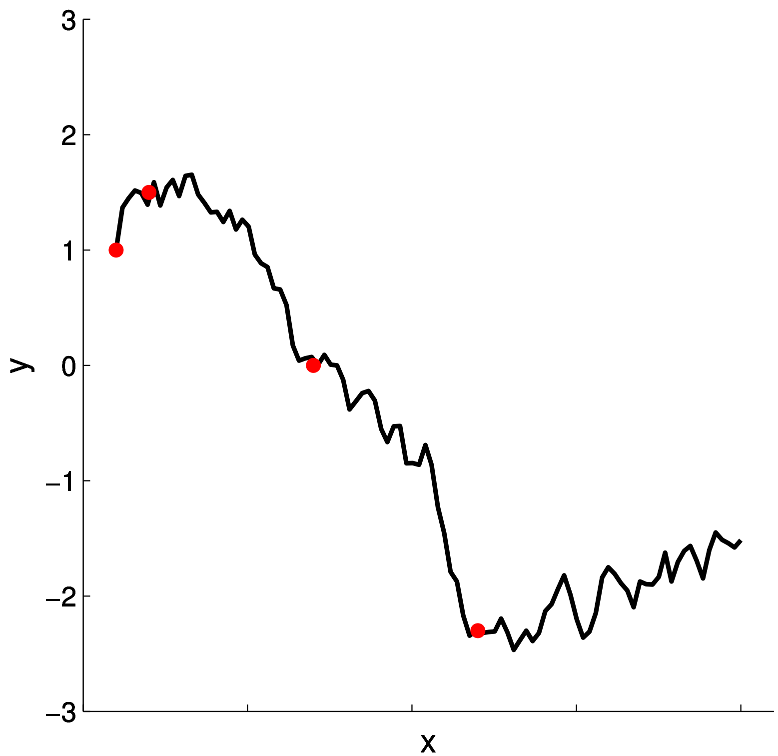
$$\mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) = \sigma^2 \exp\left(-\frac{1}{2\mathsf{l}^2}|\mathsf{x}_1 - \mathsf{x}_2|\right)$$

Laplacian covariance function

Browninan motion

Ornstein-Uhlenbeck

$\Sigma =$ 



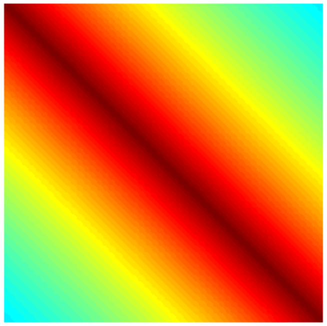
What effect does the form of the covariance function have?

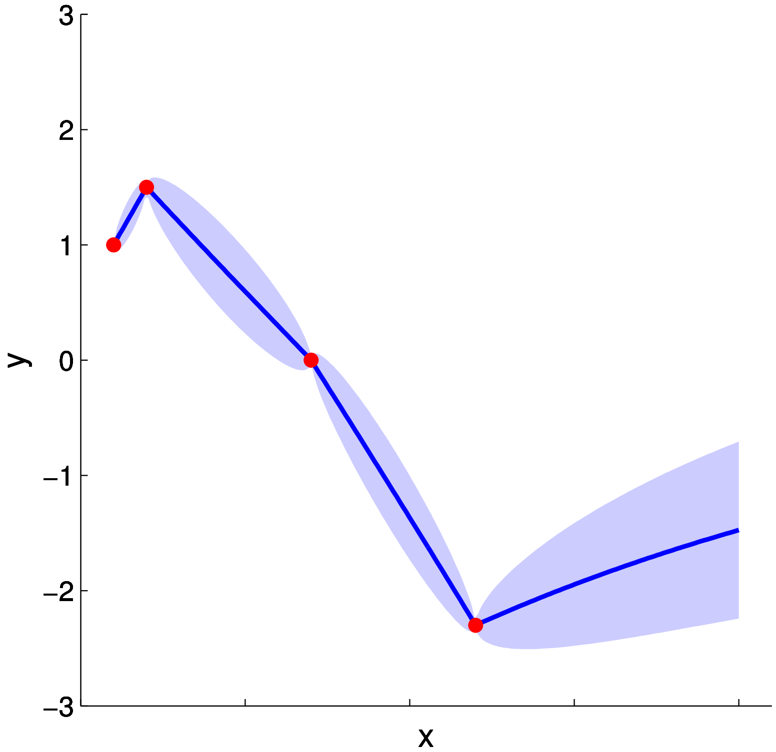
$$\mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) = \sigma^2 \exp\left(-\frac{1}{2\mathsf{l}^2}|\mathsf{x}_1 - \mathsf{x}_2|\right)$$

Laplacian covariance function

Browninan motion

Ornstein-Uhlenbeck

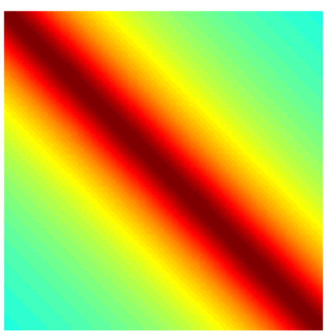
$\Sigma =$ 

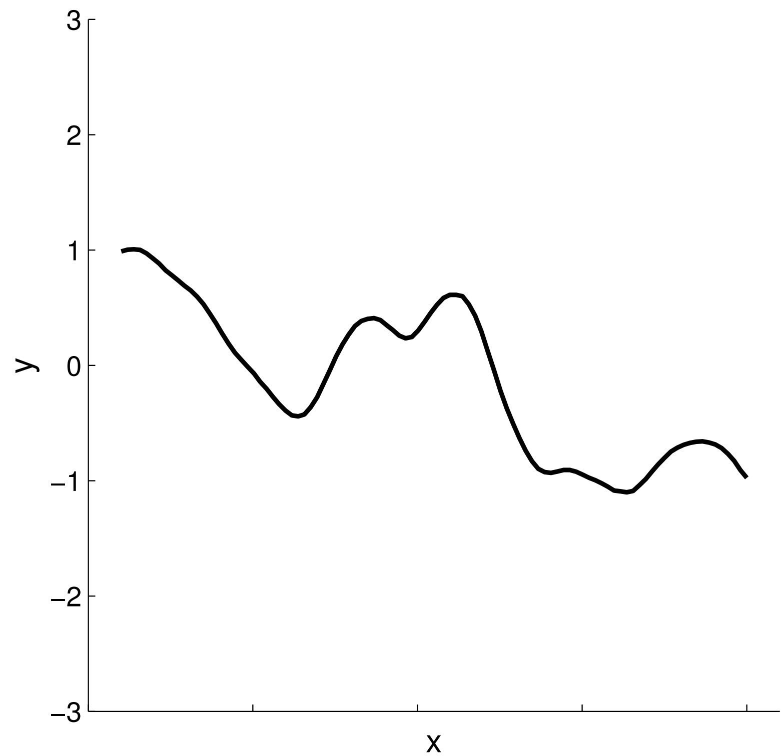


What effect does the form of the covariance function have?

$$K(x_1, x_2) = \sigma^2 \left(1 + \frac{1}{2\alpha l^2}|x_1 - x_2|\right)^{-\alpha}$$

Rational Quadratic

$\Sigma =$ 

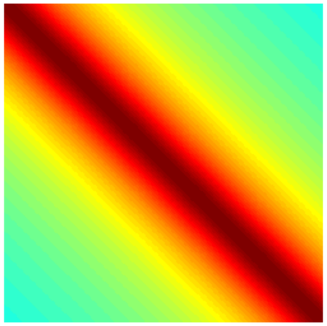


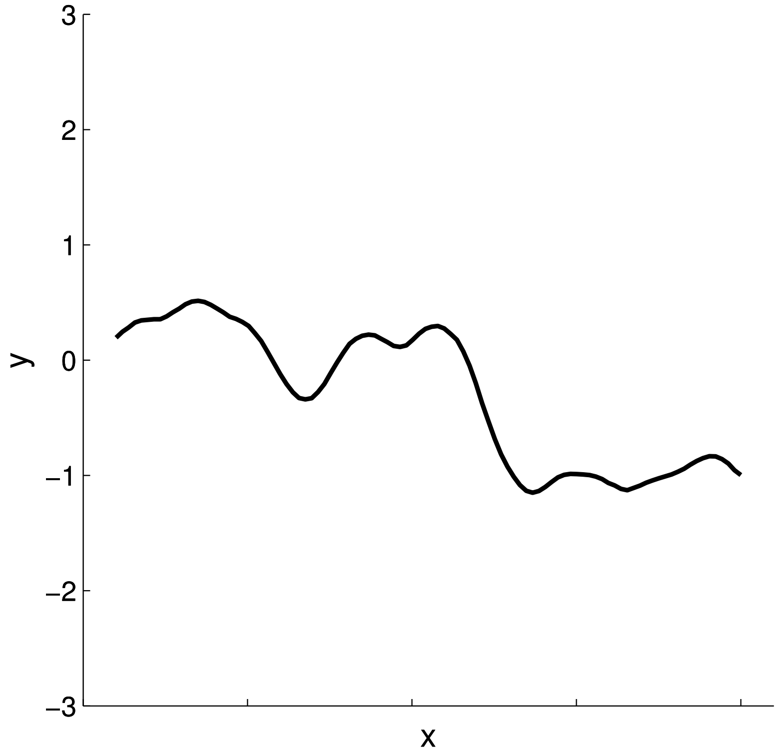
What effect does the form of the covariance function have?

$$K(x_1, x_2) = \sigma^2 \left(1 + \frac{1}{2\alpha l^2}|x_1 - x_2|\right)^{-\alpha}$$

Rational Quadratic

$\Sigma =$

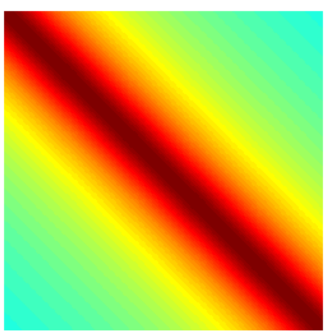


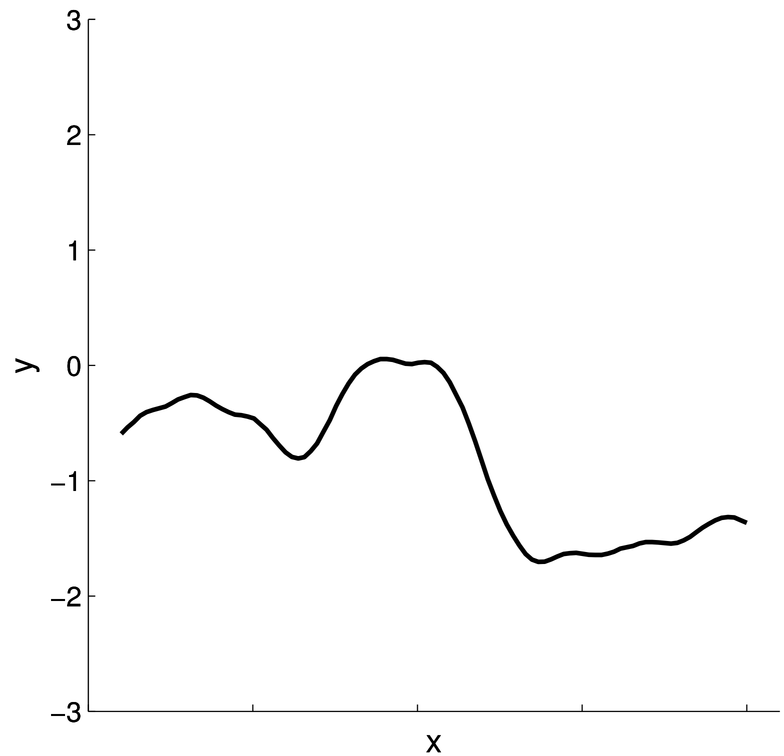


What effect does the form of the covariance function have?

$$\mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) = \sigma^2 \left(1 + \frac{1}{2\alpha \mathsf{l}^2} |\mathsf{x}_1 - \mathsf{x}_2|\right)^{-\alpha}$$

Rational Quadratic

$\Sigma =$ 

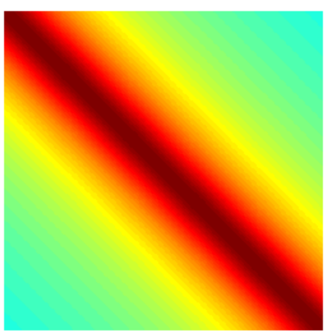


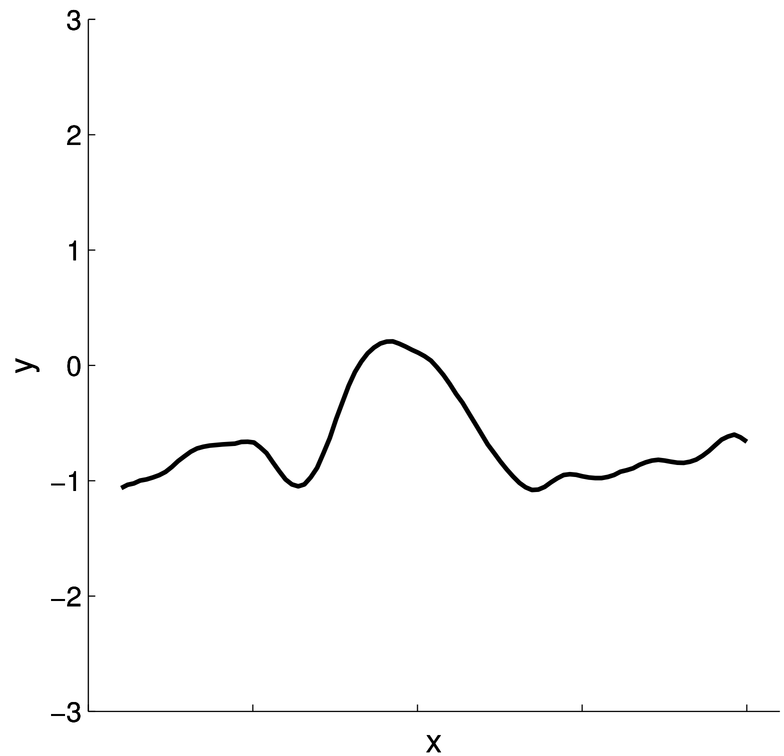
What effect does the form of the covariance function have?

$$\mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) = \sigma^2 \left(1 + \frac{1}{2\alpha \mathsf{l}^2} |\mathsf{x}_1 - \mathsf{x}_2|\right)^{-\alpha}$$

Rational Quadratic

$\Sigma =$

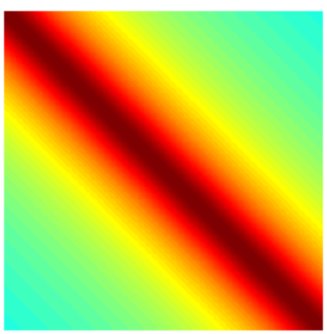


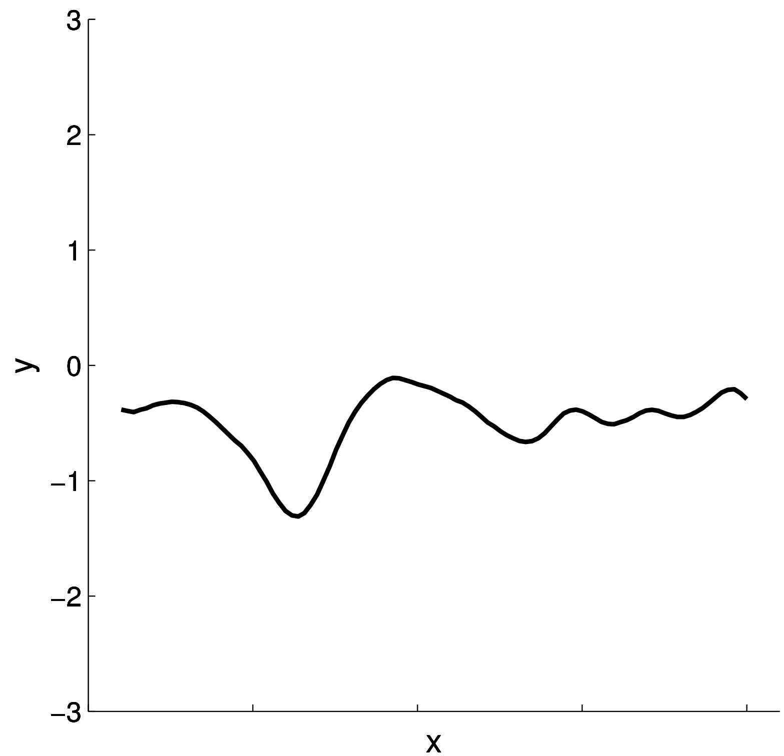


What effect does the form of the covariance function have?

$$\mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) = \sigma^2 \left(1 + \frac{1}{2\alpha \mathsf{l}^2} |\mathsf{x}_1 - \mathsf{x}_2|\right)^{-\alpha}$$

Rational Quadratic

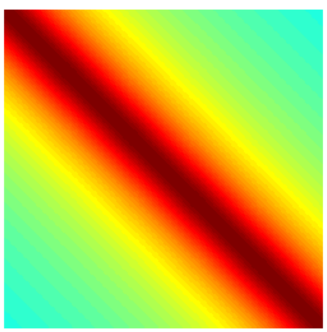
$\Sigma =$ 

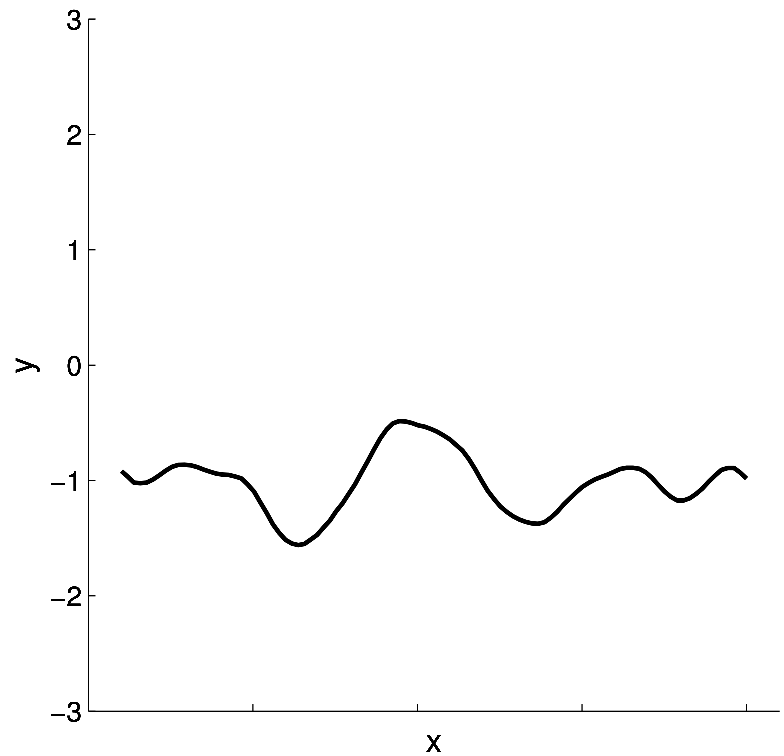


What effect does the form of the covariance function have?

$$\mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) = \sigma^2 \left(1 + \frac{1}{2\alpha \mathsf{l}^2} |\mathsf{x}_1 - \mathsf{x}_2|\right)^{-\alpha}$$

Rational Quadratic

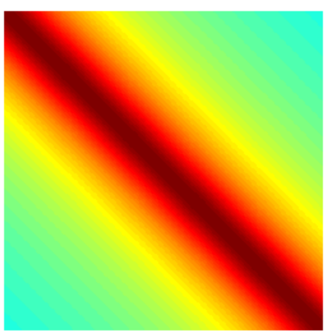
$\Sigma =$ 

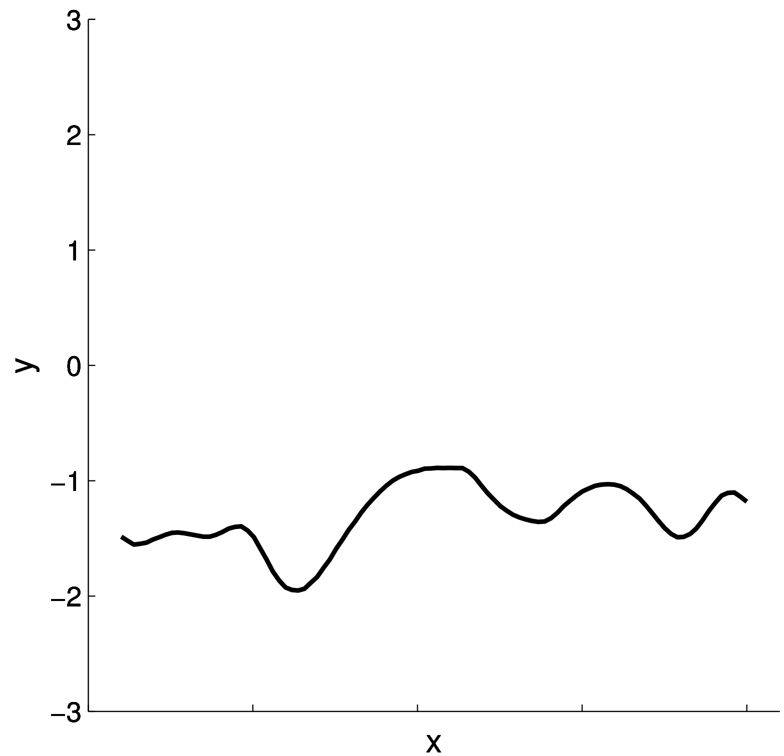


What effect does the form of the covariance function have?

$$\mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) = \sigma^2 \left(1 + \frac{1}{2\alpha \mathsf{l}^2}|\mathsf{x}_1 - \mathsf{x}_2|\right)^{-\alpha}$$

Rational Quadratic

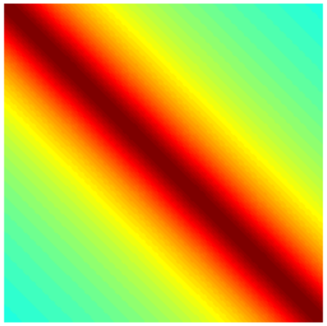
$\Sigma =$ 

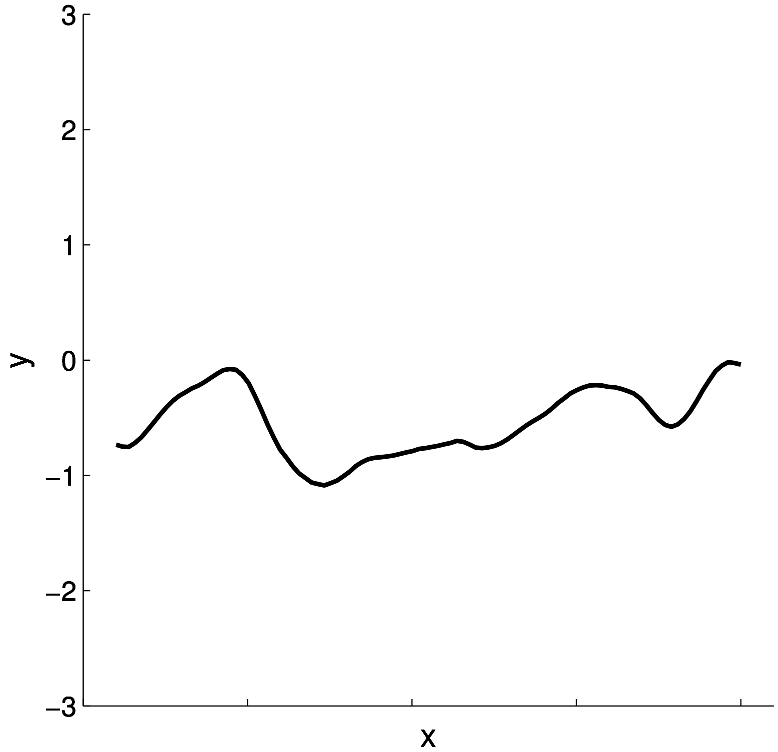


What effect does the form of the covariance function have?

$$K(x_1, x_2) = \sigma^2 \left(1 + \frac{1}{2\alpha l^2}|x_1 - x_2|\right)^{-\alpha}$$

Rational Quadratic

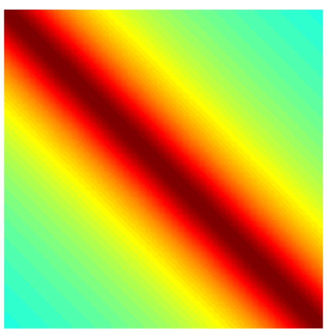
$\Sigma =$ 

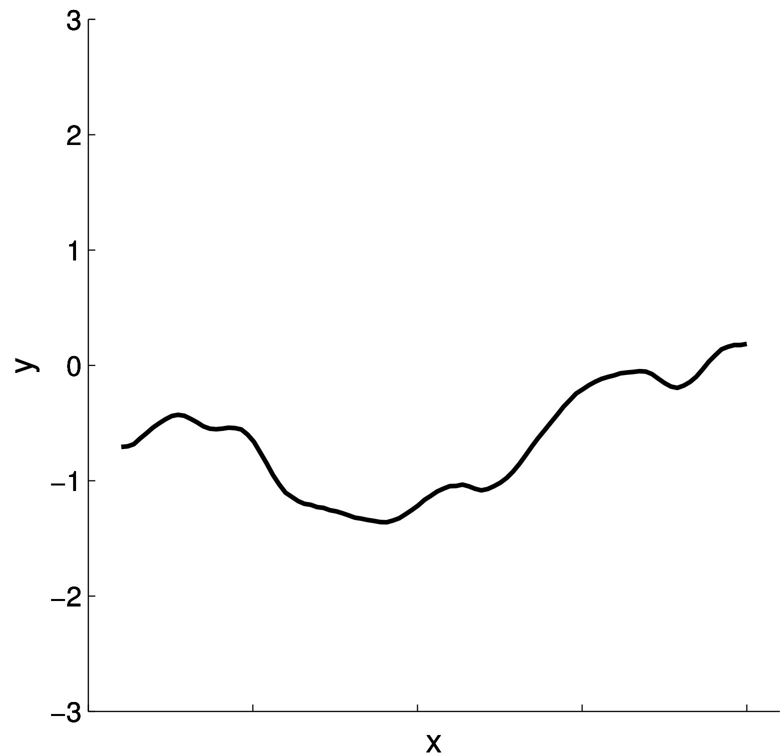


What effect does the form of the covariance function have?

$$\mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) = \sigma^2 \left(1 + \frac{1}{2\alpha \mathsf{l}^2} |\mathsf{x}_1 - \mathsf{x}_2|\right)^{-\alpha}$$

Rational Quadratic

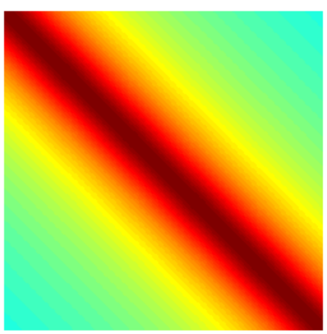
$\Sigma =$ 

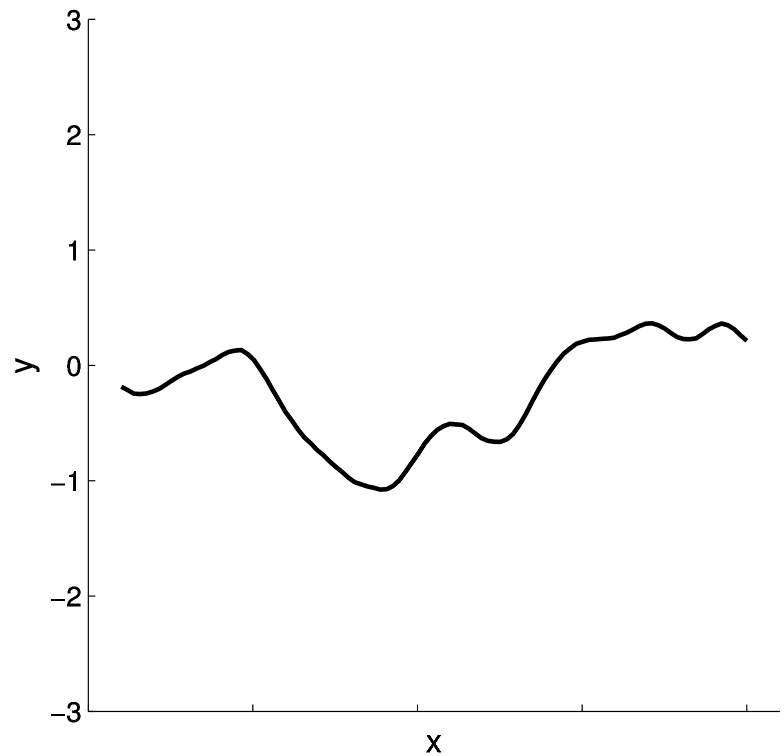


What effect does the form of the covariance function have?

$$K(x_1, x_2) = \sigma^2 \left(1 + \frac{1}{2\alpha l^2}|x_1 - x_2|\right)^{-\alpha}$$

Rational Quadratic

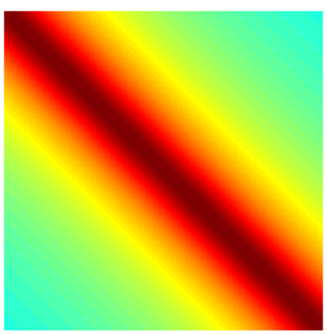
$\Sigma =$ 

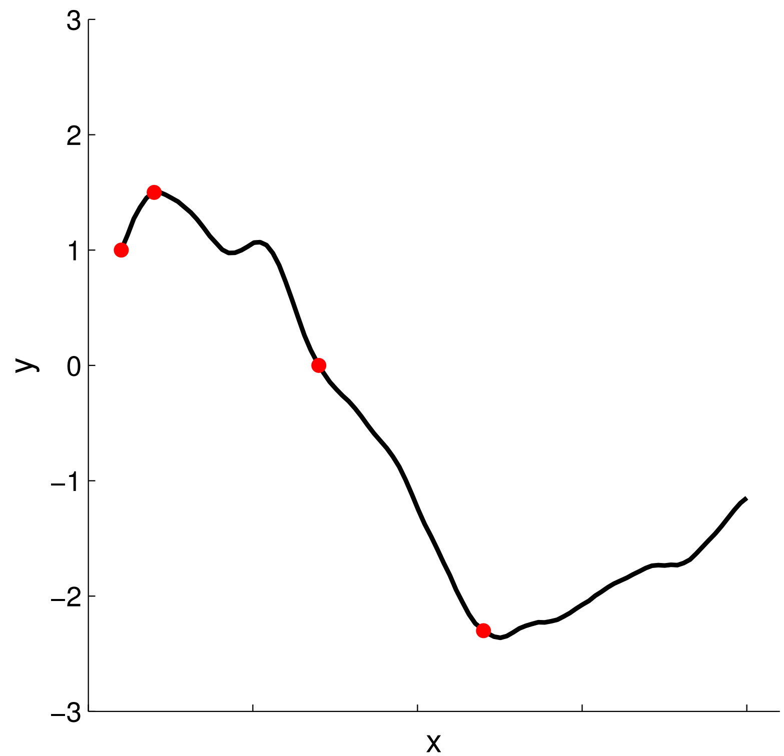


What effect does the form of the covariance function have?

$$K(x_1, x_2) = \sigma^2 \left(1 + \frac{1}{2\alpha l^2}|x_1 - x_2|\right)^{-\alpha}$$

Rational Quadratic

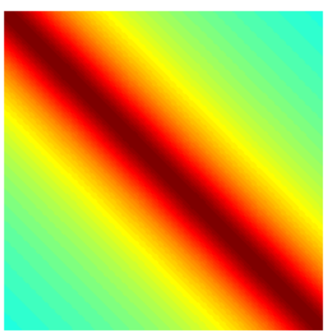
$\Sigma =$ 

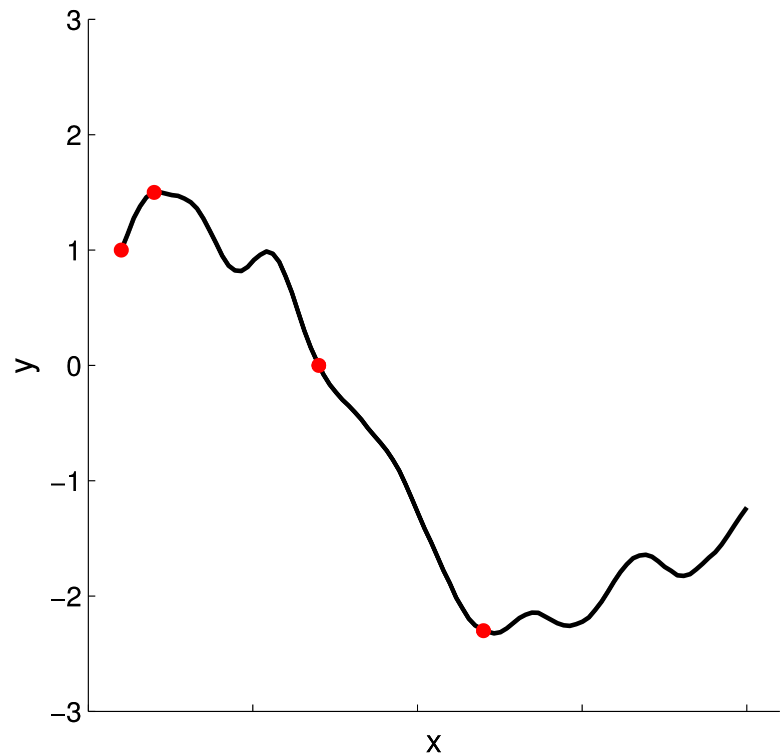


What effect does the form of the covariance function have?

$$K(x_1, x_2) = \sigma^2 \left(1 + \frac{1}{2\alpha l^2}|x_1 - x_2|\right)^{-\alpha}$$

Rational Quadratic

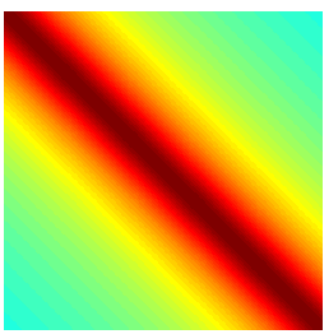
$\Sigma =$ 

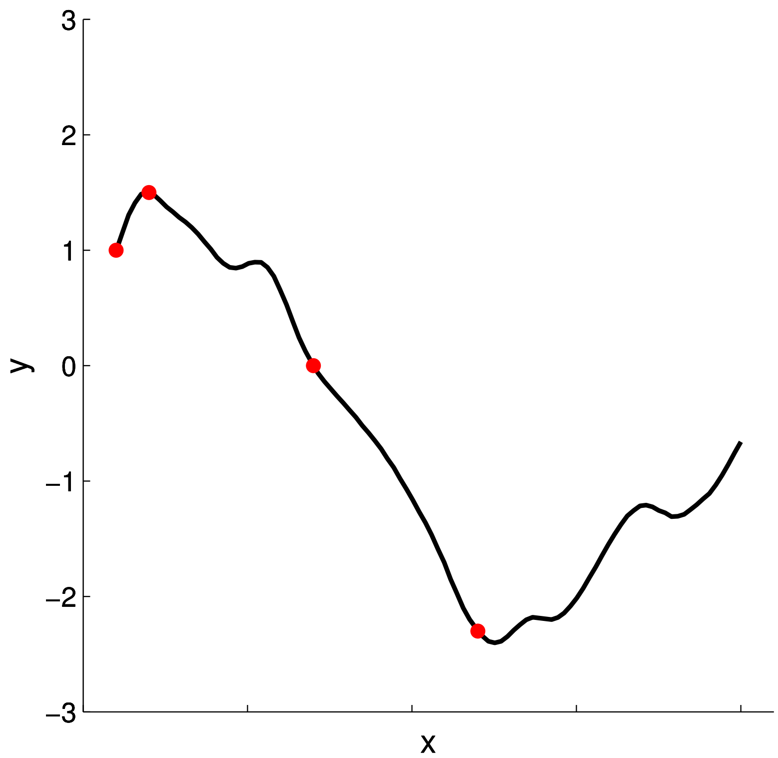


What effect does the form of the covariance function have?

$$K(x_1, x_2) = \sigma^2 \left(1 + \frac{1}{2\alpha l^2}|x_1 - x_2|\right)^{-\alpha}$$

Rational Quadratic

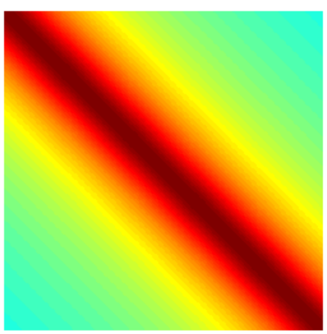
$\Sigma =$ 

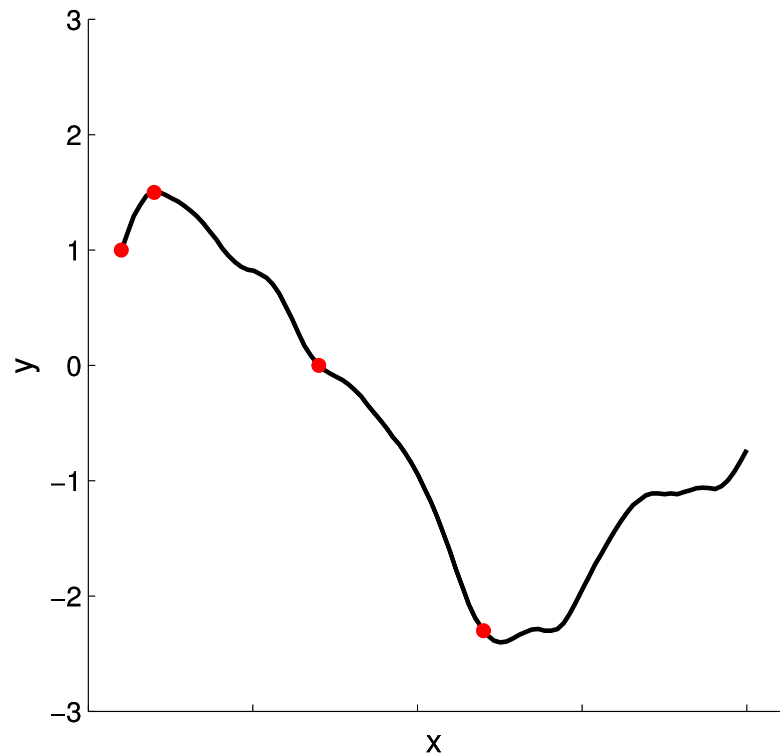


What effect does the form of the covariance function have?

$$K(x_1, x_2) = \sigma^2 \left(1 + \frac{1}{2\alpha l^2}|x_1 - x_2|\right)^{-\alpha}$$

Rational Quadratic

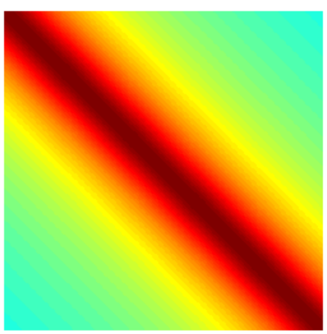
$\Sigma =$ 

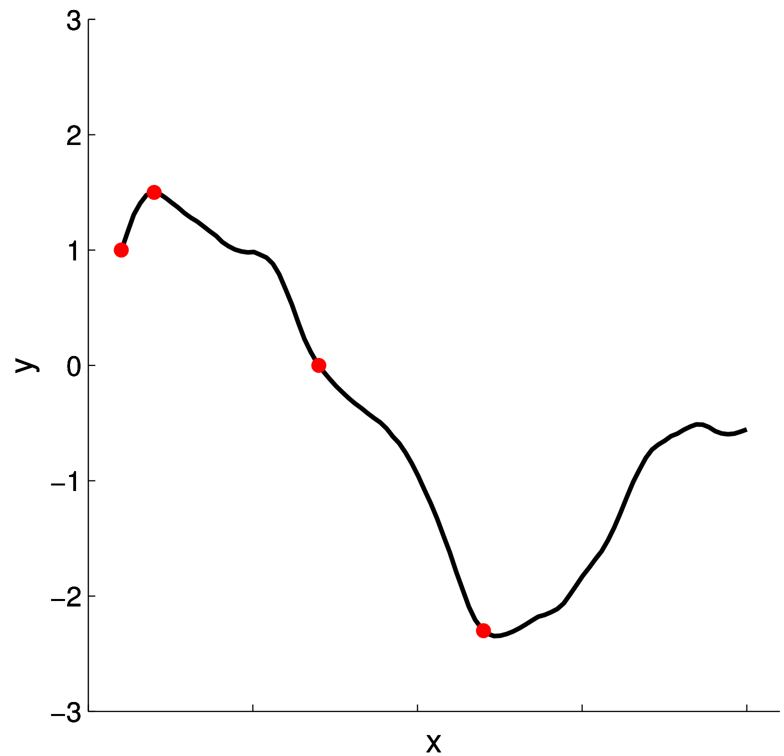


What effect does the form of the covariance function have?

$$K(x_1, x_2) = \sigma^2 \left(1 + \frac{1}{2\alpha l^2}|x_1 - x_2|\right)^{-\alpha}$$

Rational Quadratic

$\Sigma =$ 

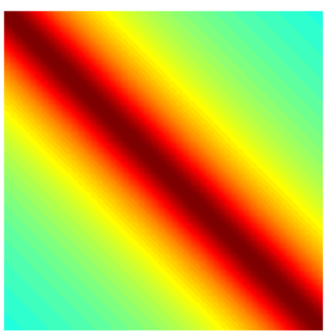


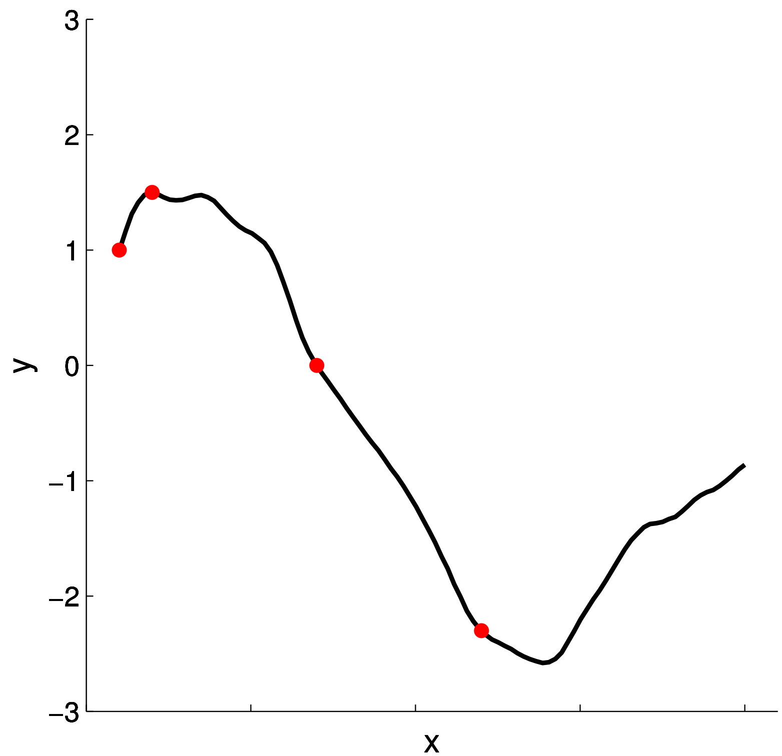
What effect does the form of the covariance function have?

$$K(x_1, x_2) = \sigma^2 \left(1 + \frac{1}{2\alpha l^2}|x_1 - x_2|\right)^{-\alpha}$$

Rational Quadratic

$\Sigma =$

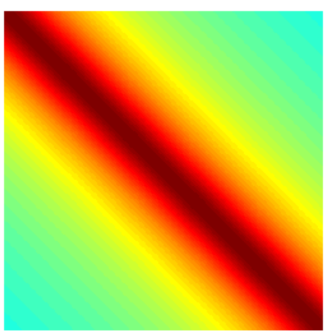


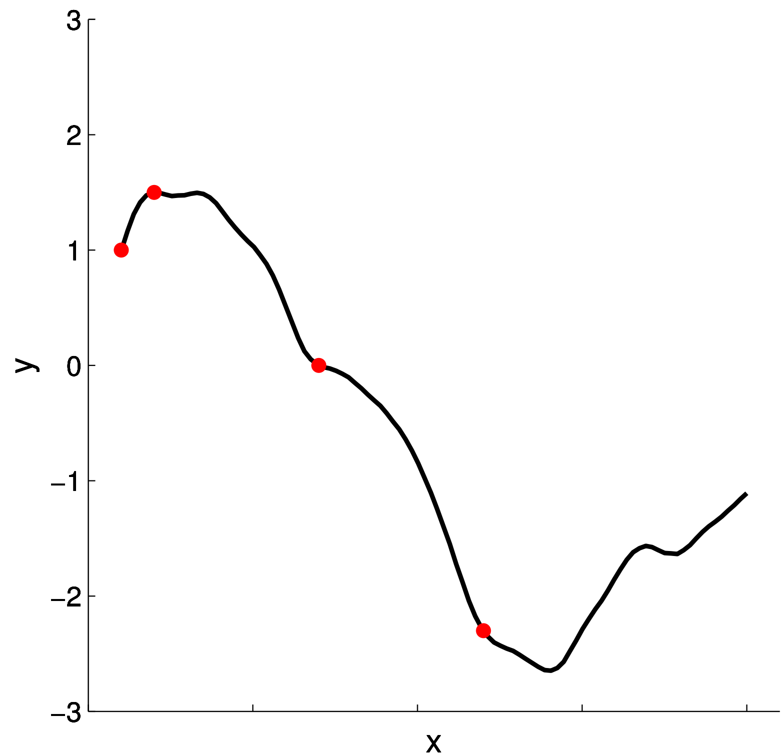


What effect does the form of the covariance function have?

$$\mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) = \sigma^2 \left(1 + \frac{1}{2\alpha \mathsf{l}^2} |\mathsf{x}_1 - \mathsf{x}_2|\right)^{-\alpha}$$

Rational Quadratic

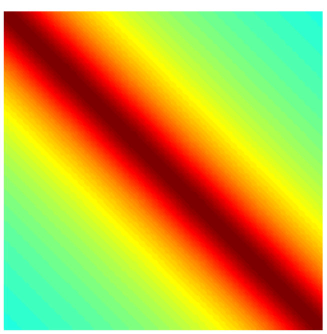
$\Sigma =$ 

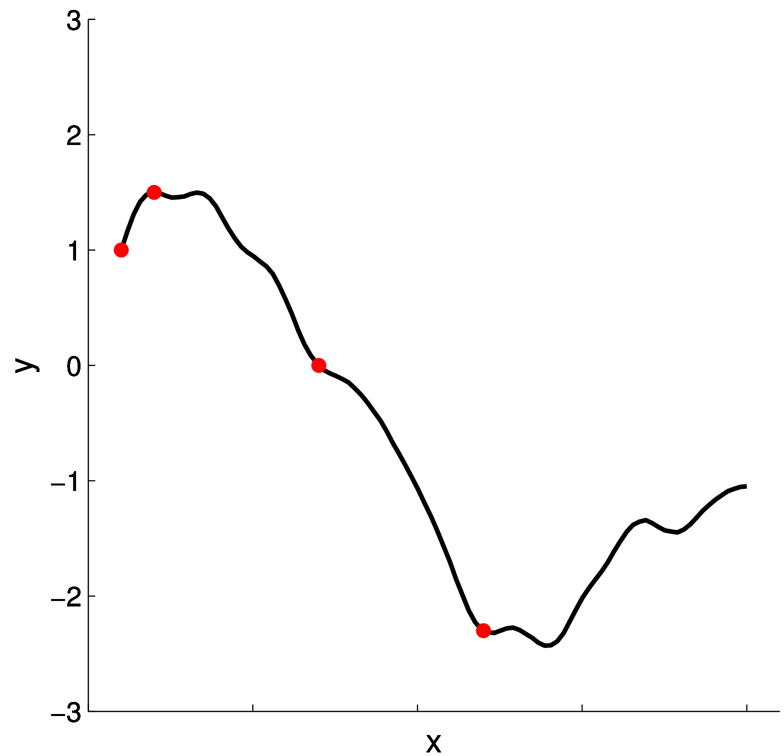


What effect does the form of the covariance function have?

$$K(x_1, x_2) = \sigma^2 \left(1 + \frac{1}{2\alpha l^2}|x_1 - x_2|\right)^{-\alpha}$$

Rational Quadratic

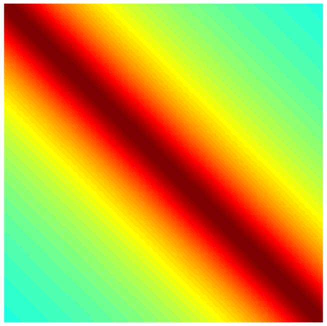
$\Sigma =$ 

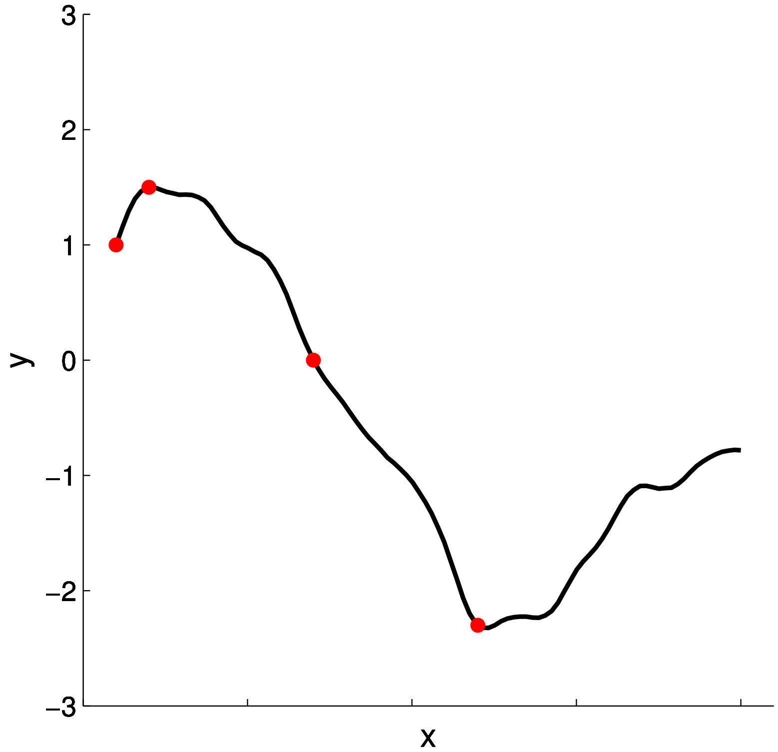


What effect does the form of the covariance function have?

$$K(x_1, x_2) = \sigma^2 \left(1 + \frac{1}{2\alpha l^2}|x_1 - x_2|\right)^{-\alpha}$$

Rational Quadratic

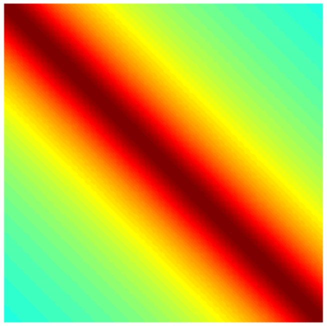
$\Sigma =$ 

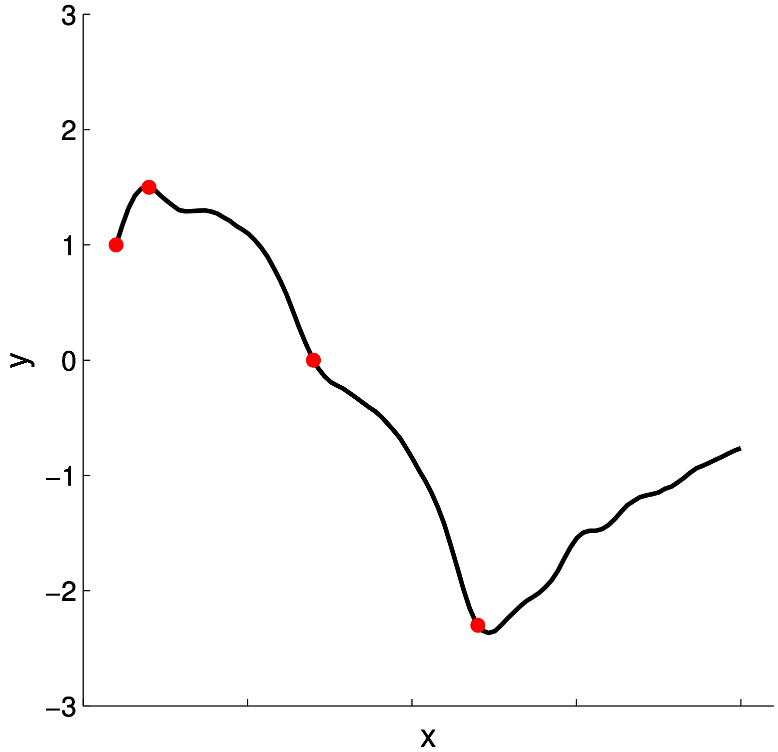


What effect does the form of the covariance function have?

$$\mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) = \sigma^2 \left(1 + \frac{1}{2\alpha \mathsf{l}^2}|\mathsf{x}_1 - \mathsf{x}_2|\right)^{-\alpha}$$

Rational Quadratic

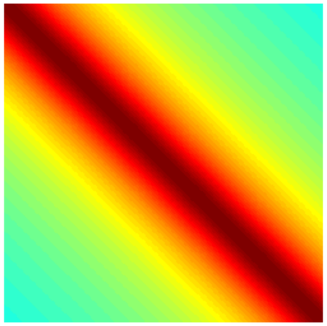
$\Sigma =$ 

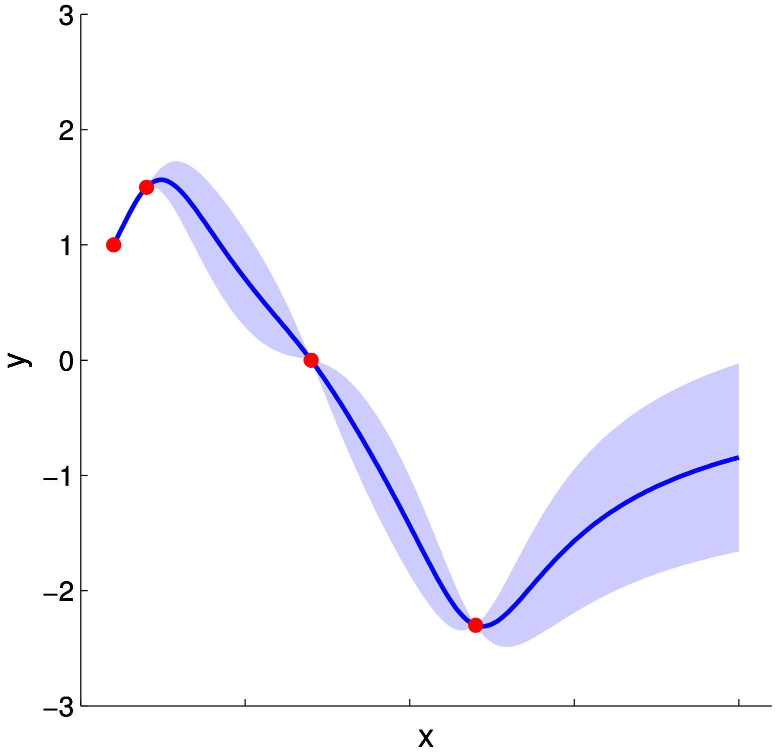


What effect does the form of the covariance function have?

$$\mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) = \sigma^2 \left(1 + \frac{1}{2\alpha \mathsf{l}^2}|\mathsf{x}_1 - \mathsf{x}_2|\right)^{-\alpha}$$

Rational Quadratic

$\Sigma =$ 

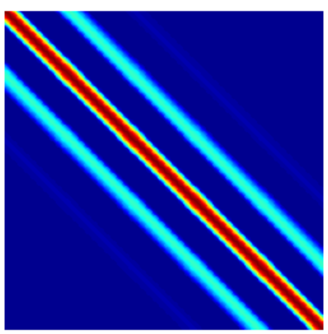


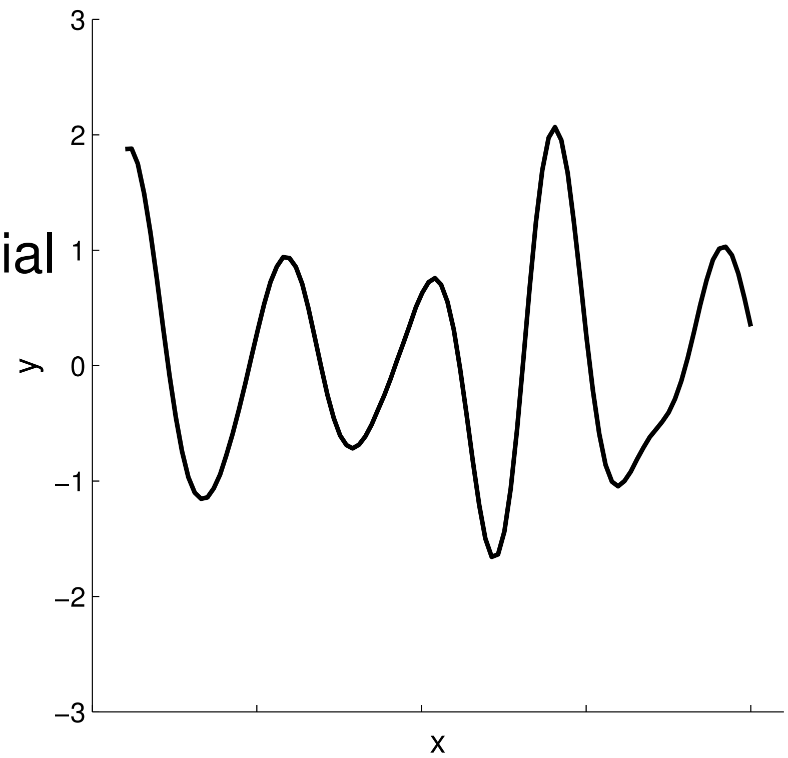
What effect does the form of the covariance function have?

$$K(x_1, x_2) = \sigma^2 \cos(\omega(x_1 - x_2)) \exp\left(-\frac{1}{2l^2}(x_1 - x_2)^2\right)$$

Periodic

sinusoid $\times$ squared exponential

$\Sigma =$ 

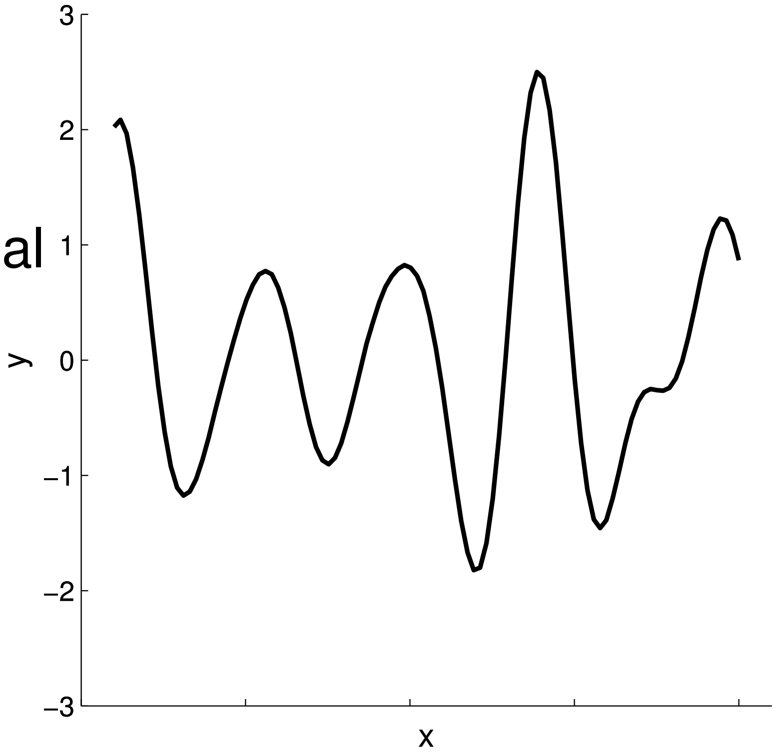


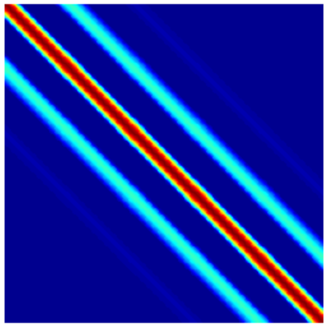
What effect does the form of the covariance function have?

$$\mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) = \sigma^2 \cos(\omega(\mathsf{x}_1 - \mathsf{x}_2)) \exp\left(-\frac{1}{2\mathsf{l}^2}(\mathsf{x}_1 - \mathsf{x}_2)^2\right)$$

Periodic

sinusoid $\times$ squared exponential



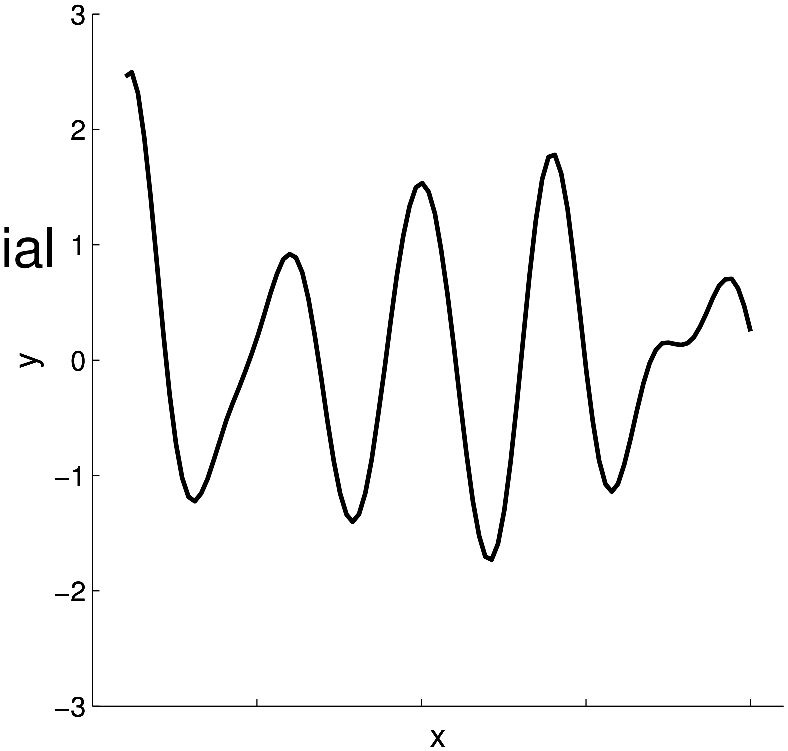
$\Sigma =$ 

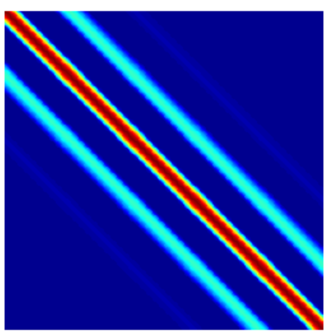
What effect does the form of the covariance function have?

$$\mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) = \sigma^2 \cos(\omega(\mathsf{x}_1 - \mathsf{x}_2)) \exp\left(-\frac{1}{2\mathsf{l}^2}(\mathsf{x}_1 - \mathsf{x}_2)^2\right)$$

Periodic

sinusoid $\times$ squared exponential



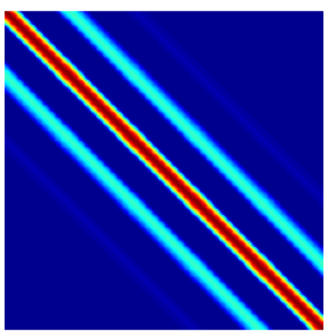
$\Sigma =$ 

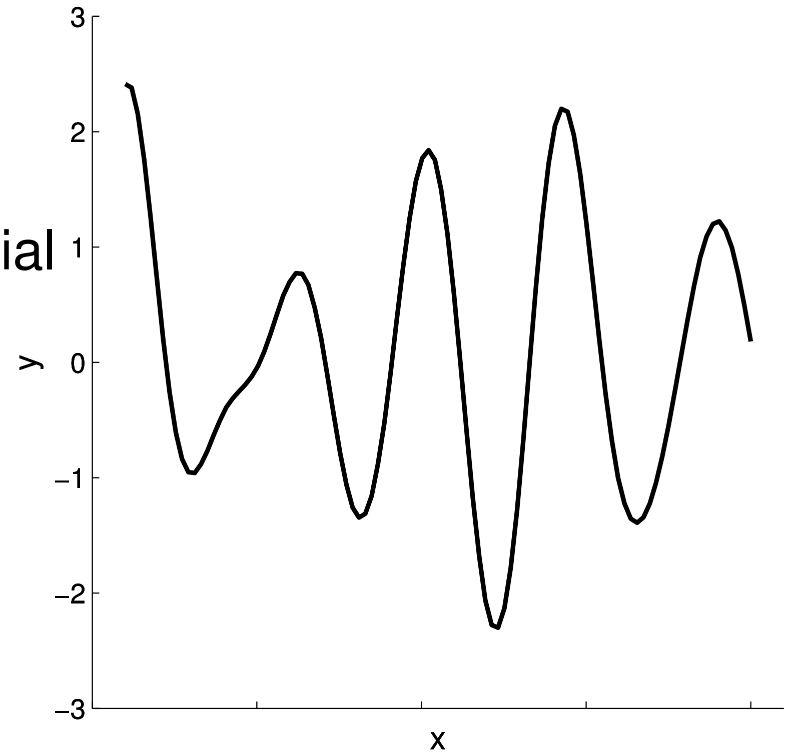
What effect does the form of the covariance function have?

$$K(x_1, x_2) = \sigma^2 \cos(\omega(x_1 - x_2)) \exp\left(-\frac{1}{2l^2}(x_1 - x_2)^2\right)$$

Periodic

sinusoid $\times$ squared exponential

$\Sigma =$ 

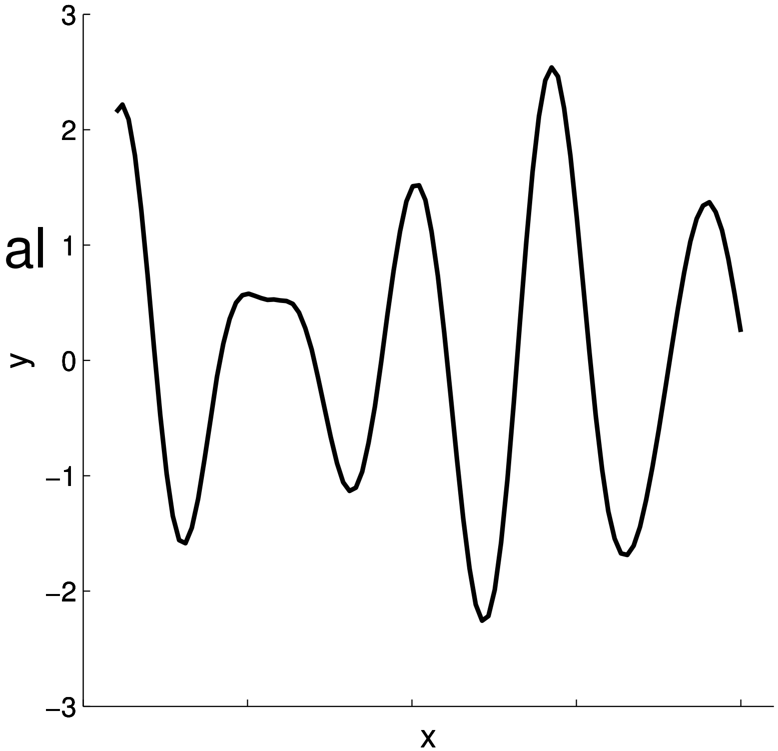


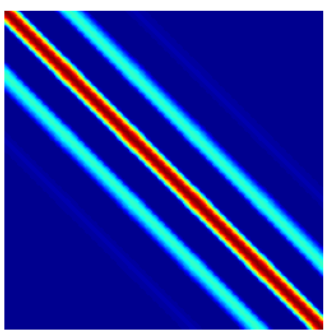
What effect does the form of the covariance function have?

$$\mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) = \sigma^2 \cos(\omega(\mathsf{x}_1 - \mathsf{x}_2)) \exp\left(-\frac{1}{2\mathsf{l}^2}(\mathsf{x}_1 - \mathsf{x}_2)^2\right)$$

Periodic

sinusoid $\times$ squared exponential



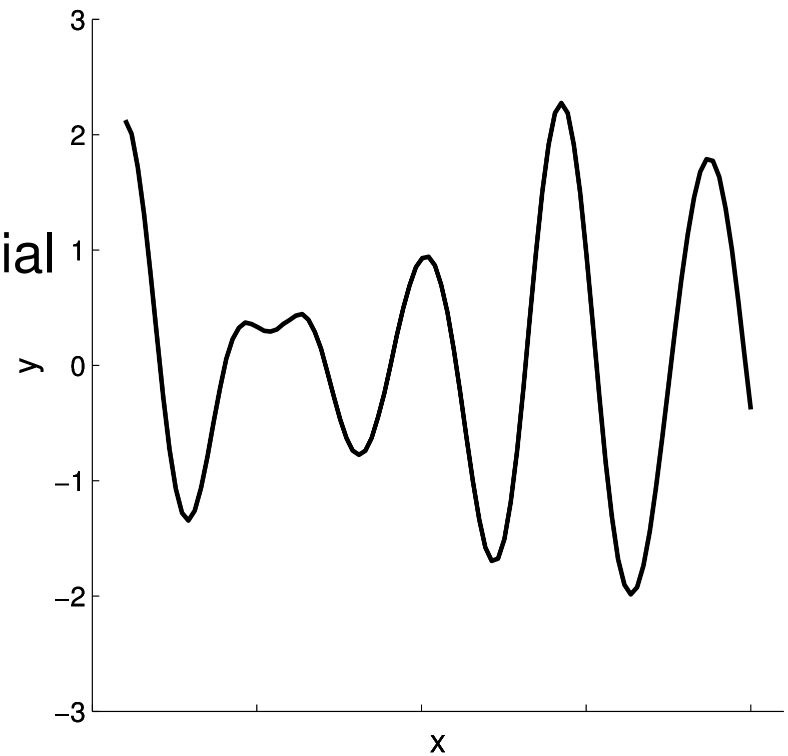
$\Sigma =$ 

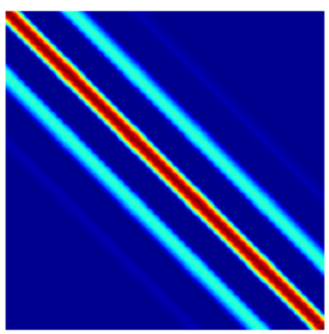
What effect does the form of the covariance function have?

$$K(x_1, x_2) = \sigma^2 \cos(\omega(x_1 - x_2)) \exp\left(-\frac{1}{2l^2}(x_1 - x_2)^2\right)$$

Periodic

sinusoid $\times$ squared exponential



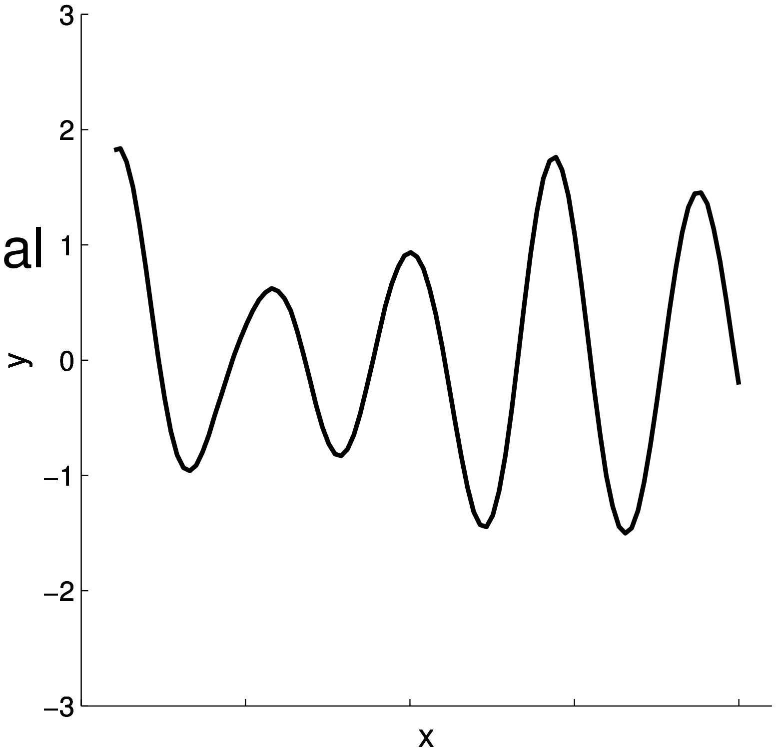
$\Sigma =$ 

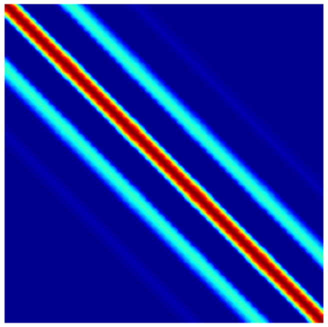
What effect does the form of the covariance function have?

$$\mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) = \sigma^2 \cos(\omega(\mathsf{x}_1 - \mathsf{x}_2)) \exp\left(-\frac{1}{2\mathsf{l}^2}(\mathsf{x}_1 - \mathsf{x}_2)^2\right)$$

Periodic

sinusoid $\times$ squared exponential



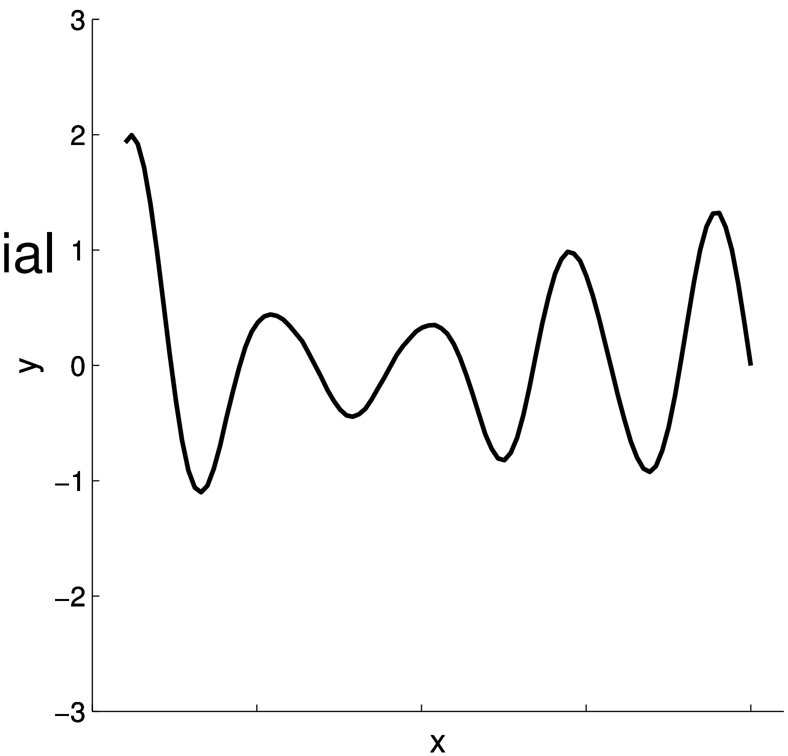
$\Sigma =$ 

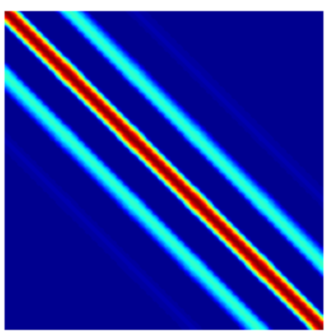
What effect does the form of the covariance function have?

$$\mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) = \sigma^2 \cos(\omega(\mathsf{x}_1 - \mathsf{x}_2)) \exp\left(-\frac{1}{2\mathsf{l}^2}(\mathsf{x}_1 - \mathsf{x}_2)^2\right)$$

Periodic

sinusoid $\times$ squared exponential



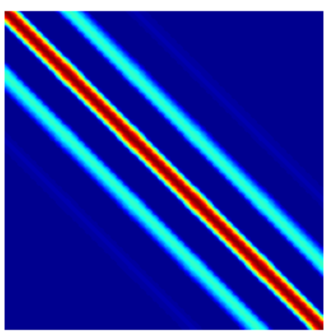
$\Sigma =$ 

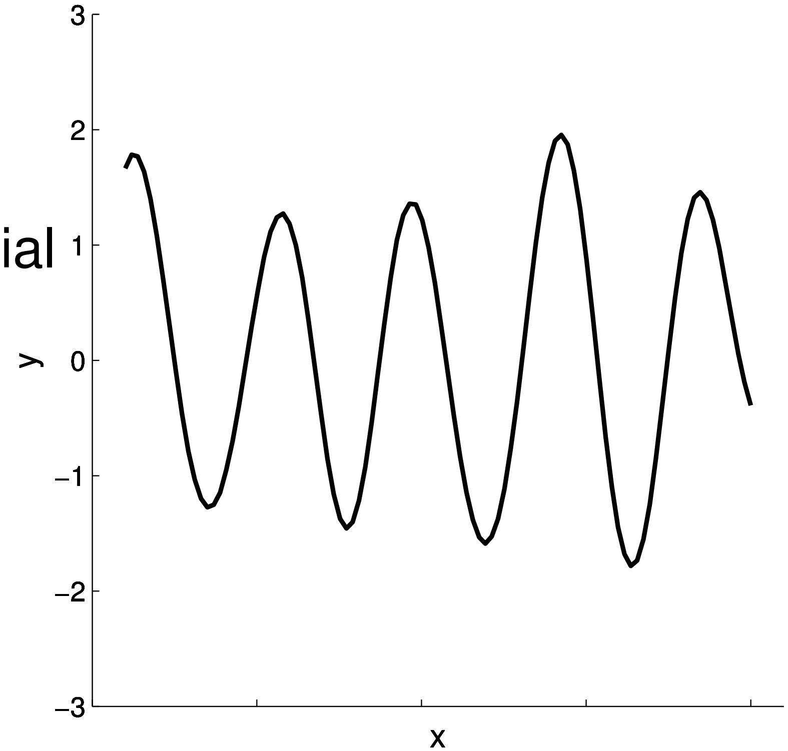
What effect does the form of the covariance function have?

$$\mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) = \sigma^2 \cos(\omega(\mathsf{x}_1 - \mathsf{x}_2)) \exp\left(-\frac{1}{2\mathsf{l}^2}(\mathsf{x}_1 - \mathsf{x}_2)^2\right)$$

Periodic

sinusoid $\times$ squared exponential

$\Sigma =$ 

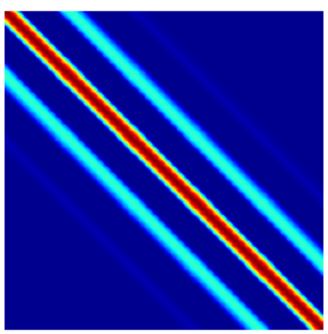


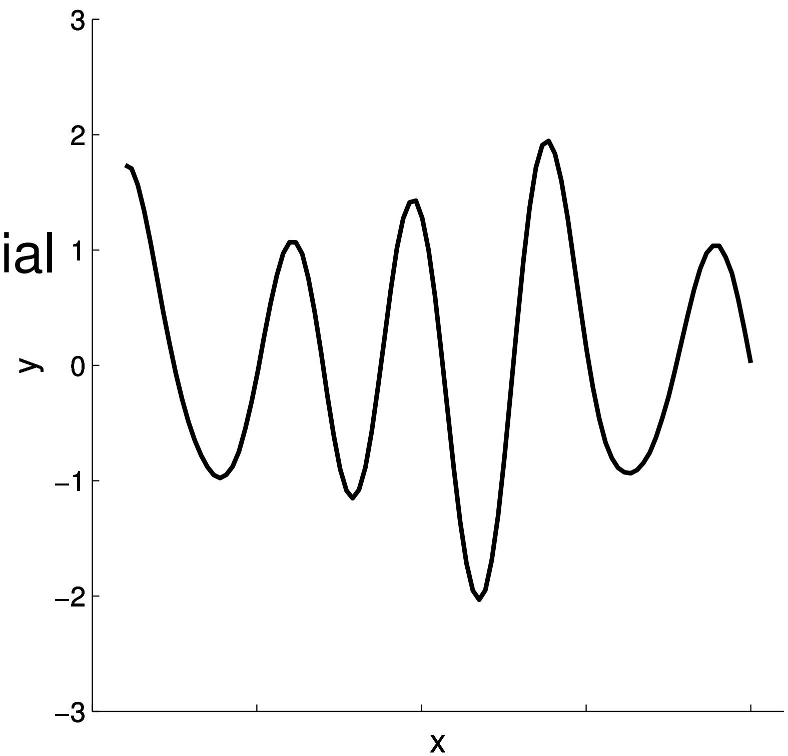
What effect does the form of the covariance function have?

$$\mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) = \sigma^2 \cos(\omega(\mathsf{x}_1 - \mathsf{x}_2)) \exp\left(-\frac{1}{2\mathsf{l}^2}(\mathsf{x}_1 - \mathsf{x}_2)^2\right)$$

Periodic

sinusoid $\times$ squared exponential

$\Sigma =$ 

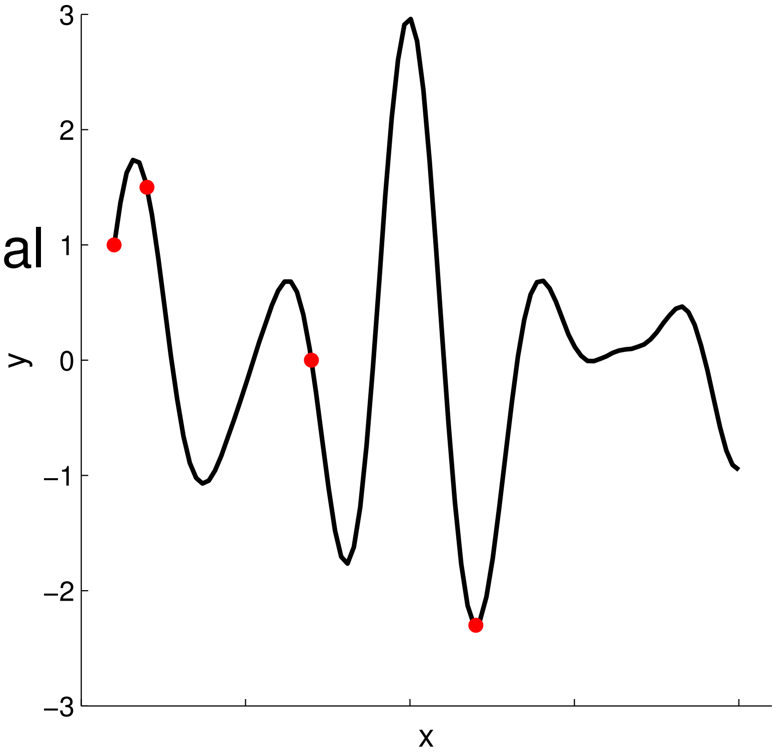


What effect does the form of the covariance function have?

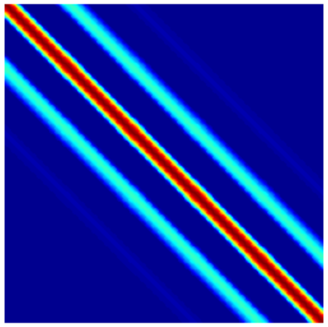
$$K(x_1, x_2) = \sigma^2 \cos(\omega(x_1 - x_2)) \exp\left(-\frac{1}{2l^2}(x_1 - x_2)^2\right)$$

Periodic

sinusoid $\times$ squared exponential



$\Sigma =$

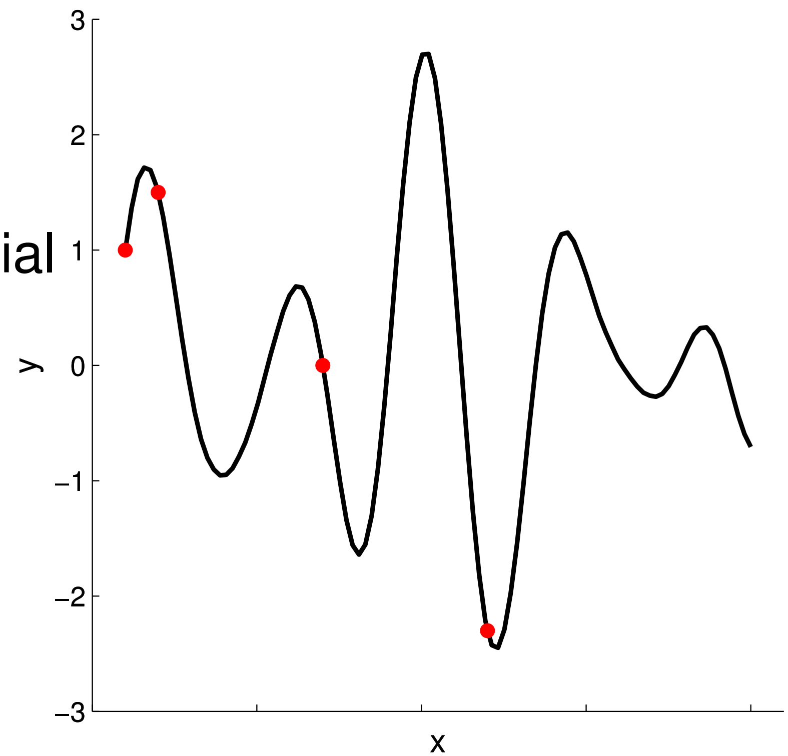


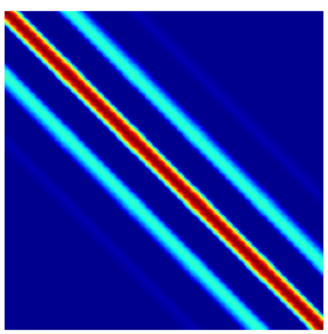
What effect does the form of the covariance function have?

$$K(x_1, x_2) = \sigma^2 \cos(\omega(x_1 - x_2)) \exp\left(-\frac{1}{2l^2}(x_1 - x_2)^2\right)$$

Periodic

sinusoid $\times$ squared exponential



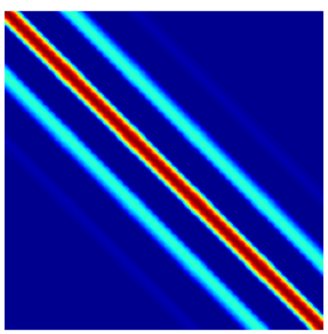
$\Sigma =$ 

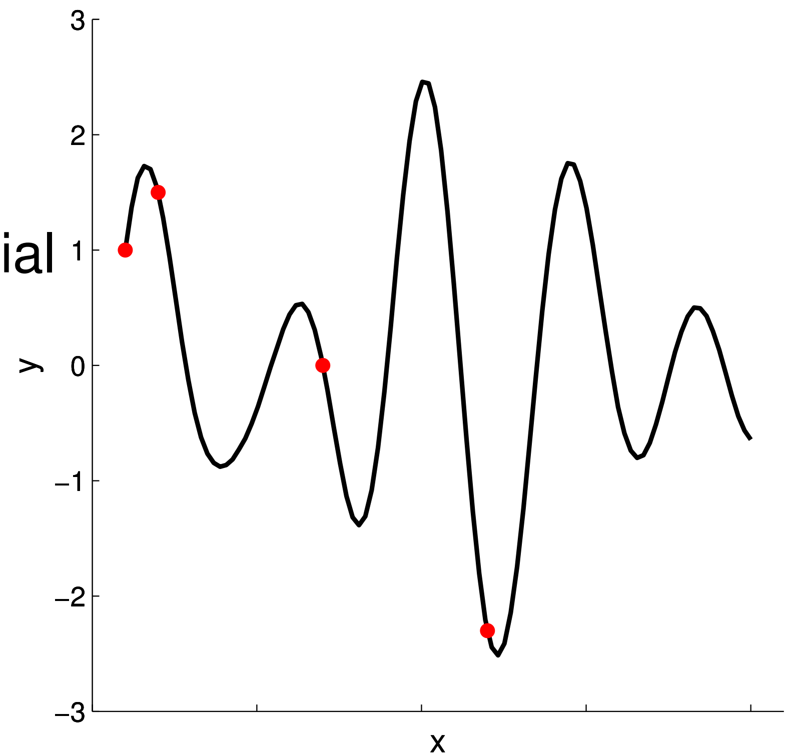
What effect does the form of the covariance function have?

$$\mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) = \sigma^2 \cos(\omega(\mathsf{x}_1 - \mathsf{x}_2)) \exp\left(-\frac{1}{2\mathsf{l}^2}(\mathsf{x}_1 - \mathsf{x}_2)^2\right)$$

Periodic

sinusoid $\times$ squared exponential

$\Sigma =$ 

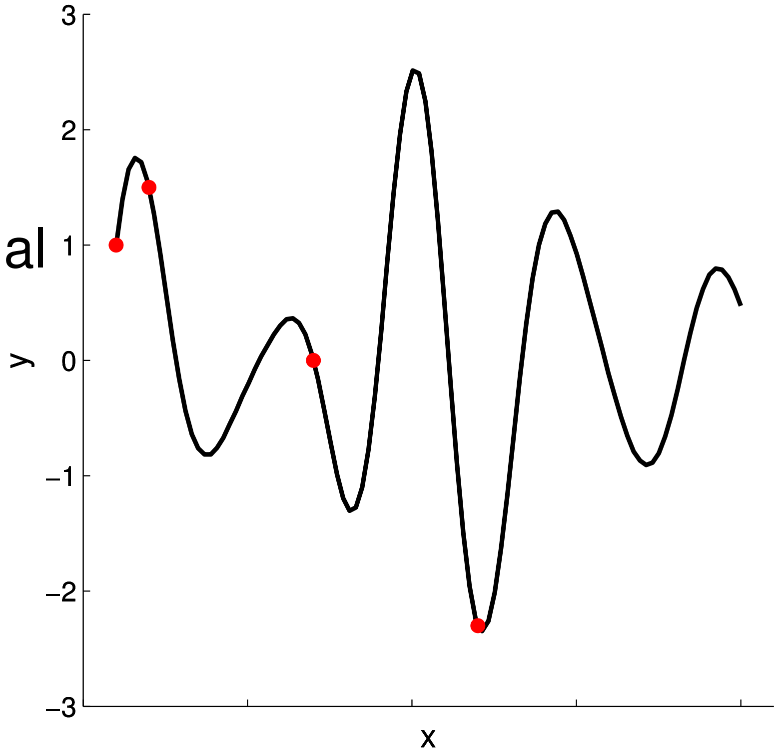


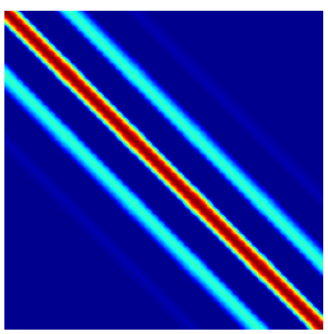
What effect does the form of the covariance function have?

$$K(x_1, x_2) = \sigma^2 \cos(\omega(x_1 - x_2)) \exp\left(-\frac{1}{2l^2}(x_1 - x_2)^2\right)$$

Periodic

sinusoid $\times$ squared exponential



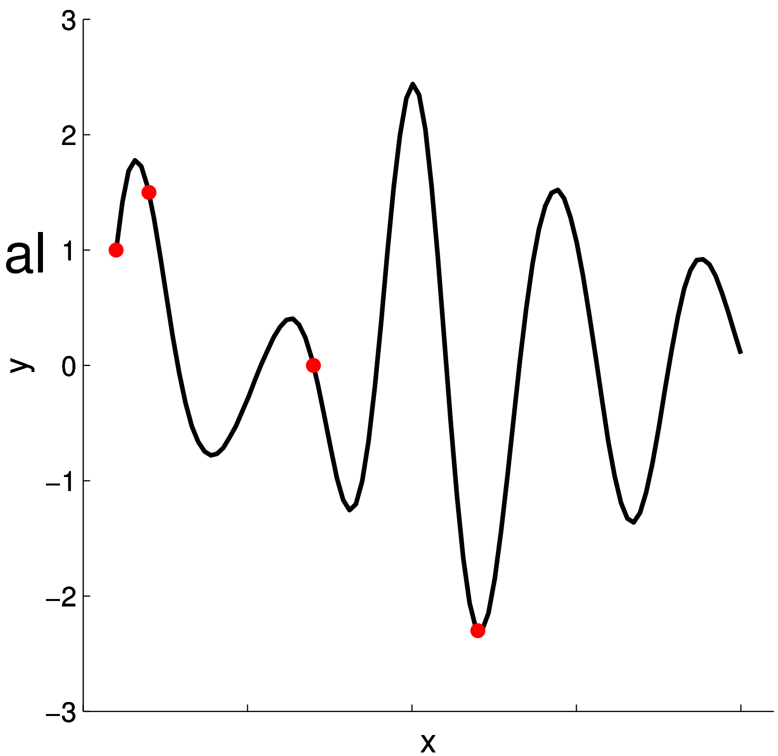
$\Sigma =$ 

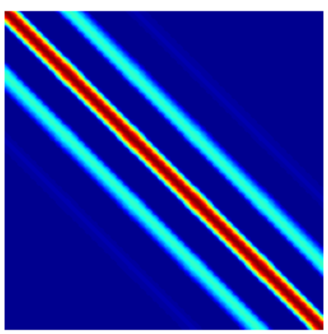
What effect does the form of the covariance function have?

$$K(x_1, x_2) = \sigma^2 \cos(\omega(x_1 - x_2)) \exp\left(-\frac{1}{2l^2}(x_1 - x_2)^2\right)$$

Periodic

sinusoid $\times$ squared exponential



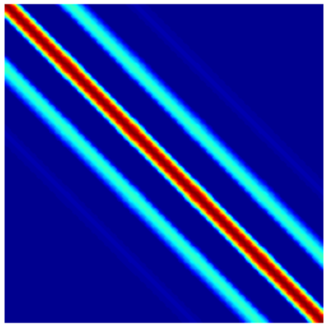
$\Sigma =$ 

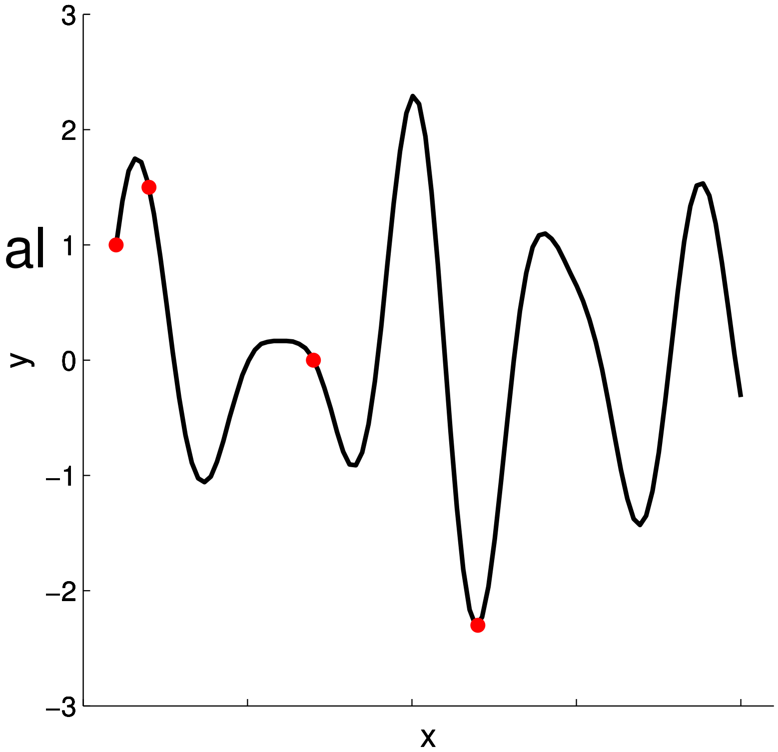
What effect does the form of the covariance function have?

$$K(x_1, x_2) = \sigma^2 \cos(\omega(x_1 - x_2)) \exp\left(-\frac{1}{2l^2}(x_1 - x_2)^2\right)$$

Periodic

sinusoid $\times$ squared exponential

$\Sigma =$ 

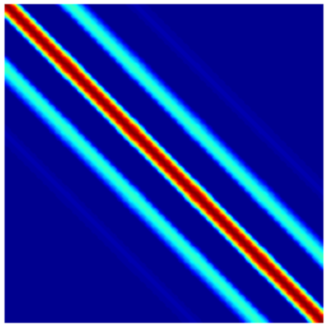


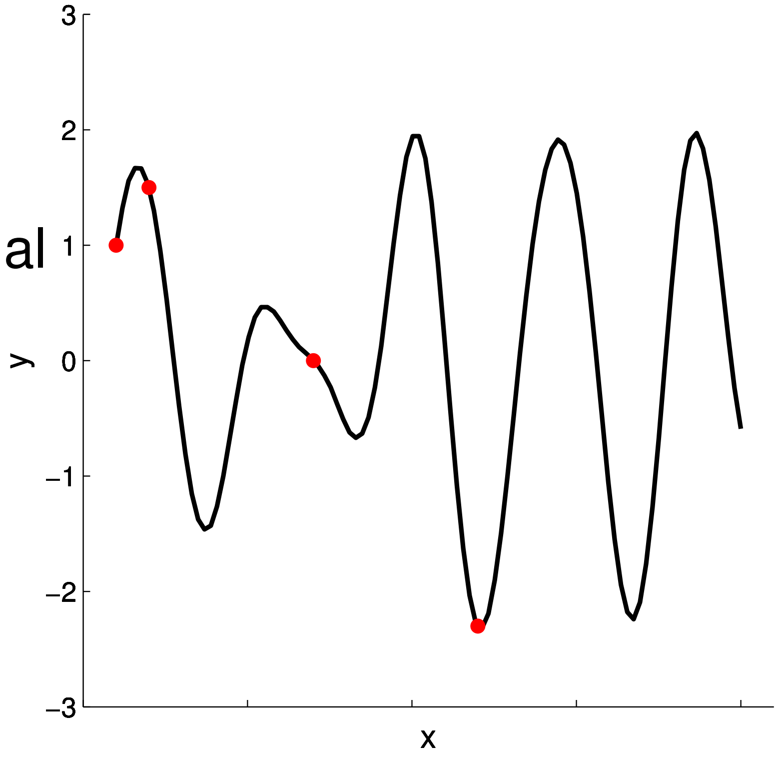
What effect does the form of the covariance function have?

$$K(x_1, x_2) = \sigma^2 \cos(\omega(x_1 - x_2)) \exp\left(-\frac{1}{2l^2}(x_1 - x_2)^2\right)$$

Periodic

sinusoid $\times$ squared exponential

$\Sigma =$ 

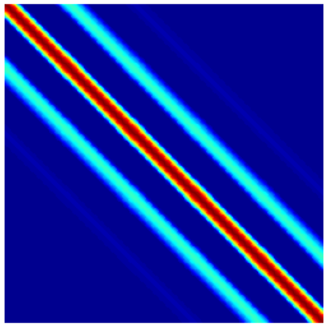


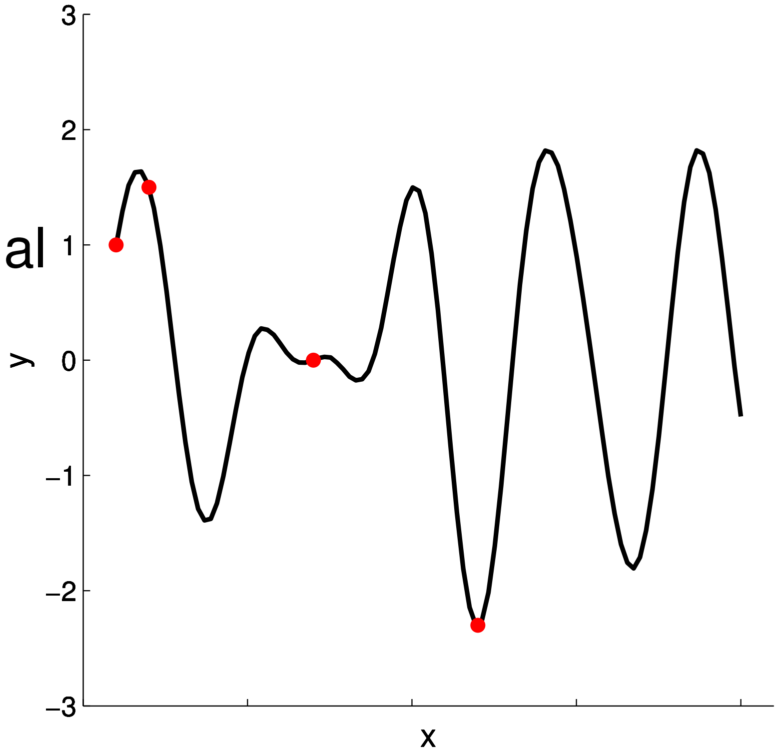
What effect does the form of the covariance function have?

$$\mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) = \sigma^2 \cos(\omega(\mathsf{x}_1 - \mathsf{x}_2)) \exp\left(-\frac{1}{2\mathsf{l}^2}(\mathsf{x}_1 - \mathsf{x}_2)^2\right)$$

Periodic

sinusoid $\times$ squared exponential

$\Sigma =$ 

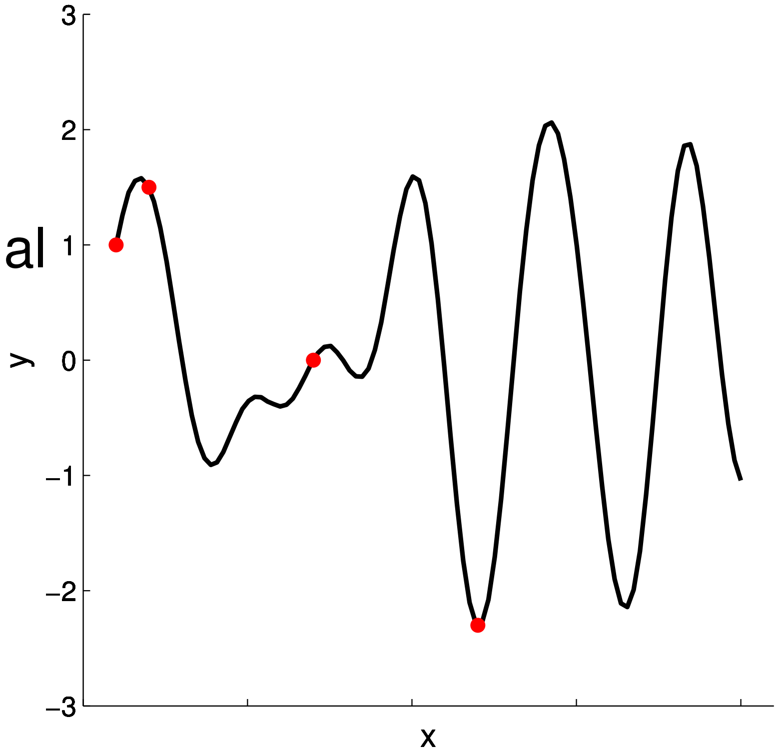


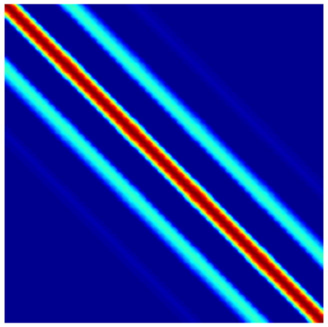
What effect does the form of the covariance function have?

$$\mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) = \sigma^2 \cos(\omega(\mathsf{x}_1 - \mathsf{x}_2)) \exp\left(-\frac{1}{2\mathsf{l}^2}(\mathsf{x}_1 - \mathsf{x}_2)^2\right)$$

Periodic

sinusoid $\times$ squared exponential



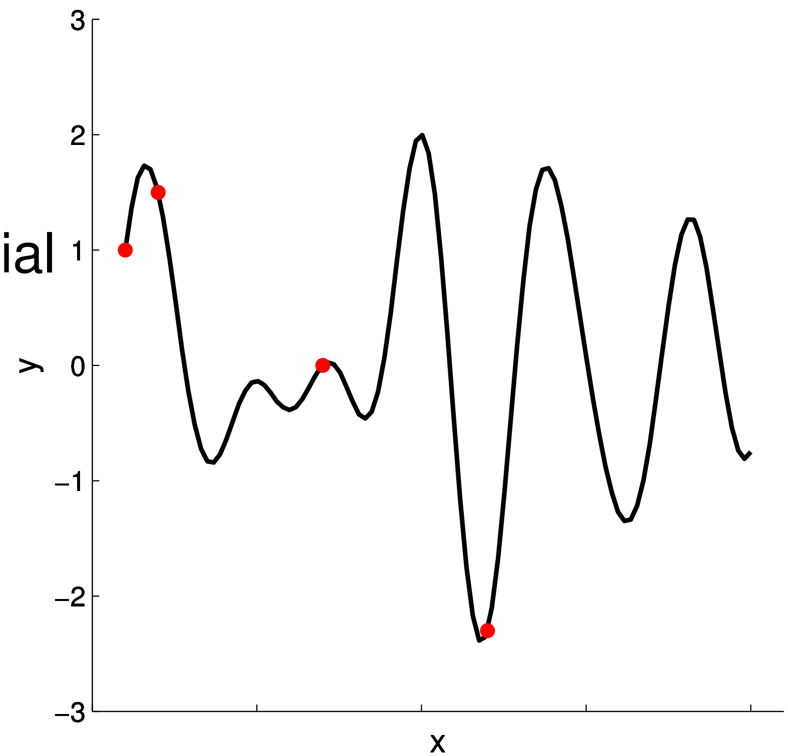
$\Sigma =$ 

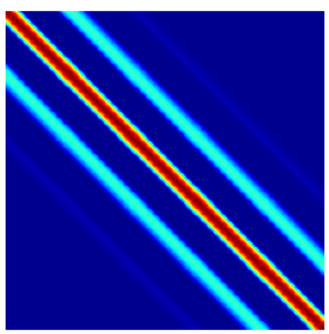
What effect does the form of the covariance function have?

$$\mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) = \sigma^2 \cos(\omega(\mathsf{x}_1 - \mathsf{x}_2)) \exp\left(-\frac{1}{2\mathsf{l}^2}(\mathsf{x}_1 - \mathsf{x}_2)^2\right)$$

Periodic

sinusoid $\times$ squared exponential



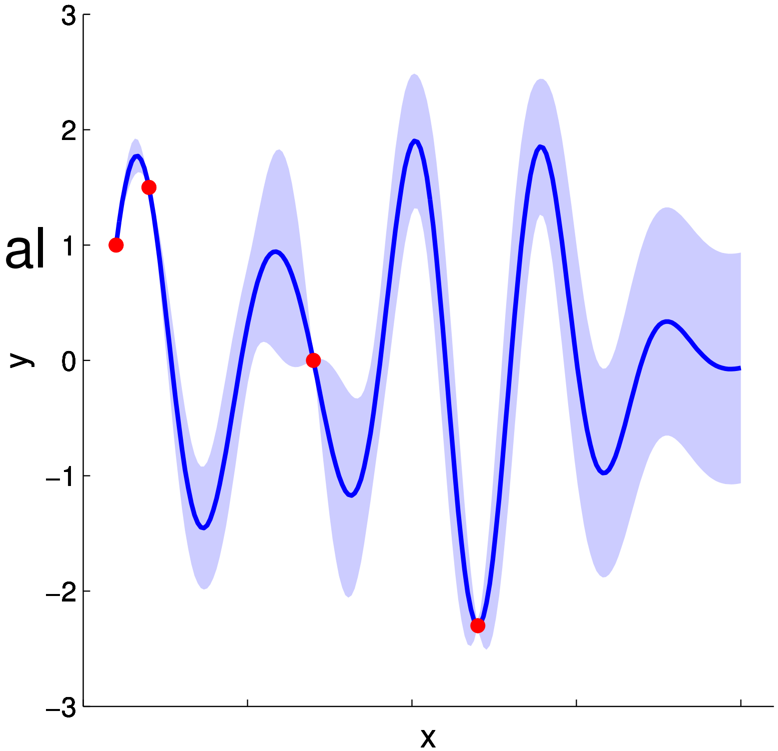
$\Sigma =$ 

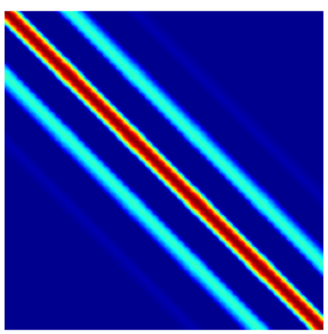
What effect does the form of the covariance function have?

$$\mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) = \sigma^2 \cos(\omega(\mathsf{x}_1 - \mathsf{x}_2)) \exp\left(-\frac{1}{2\mathsf{l}^2}(\mathsf{x}_1 - \mathsf{x}_2)^2\right)$$

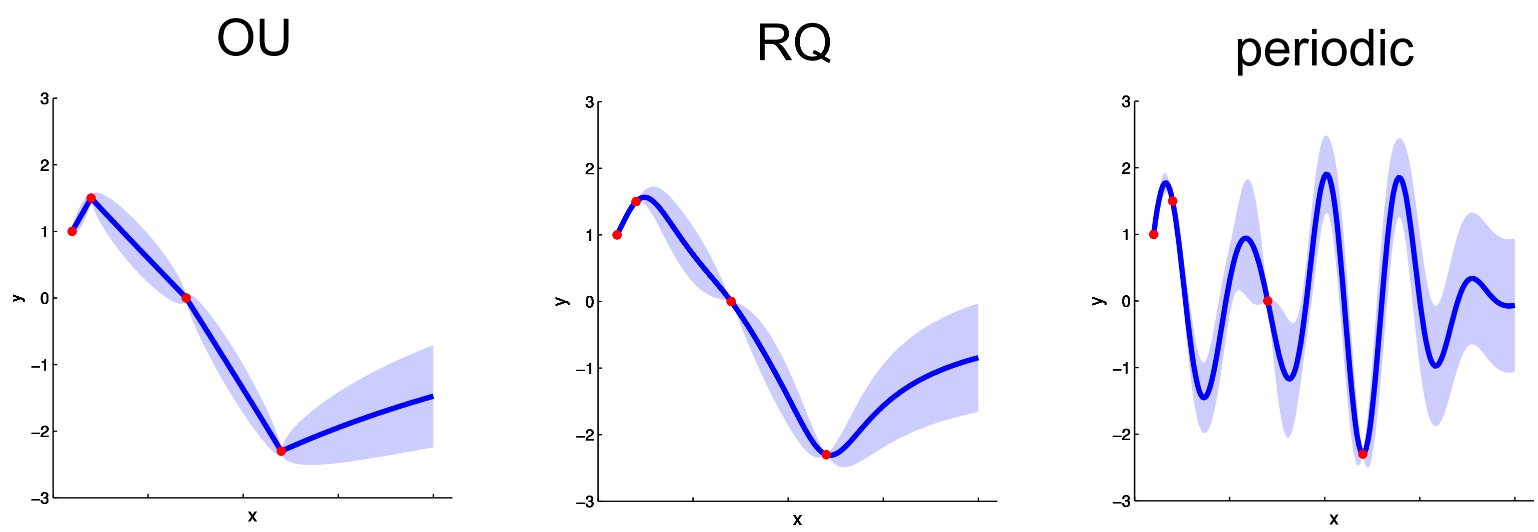
Periodic

sinusoid $\times$ squared exponential

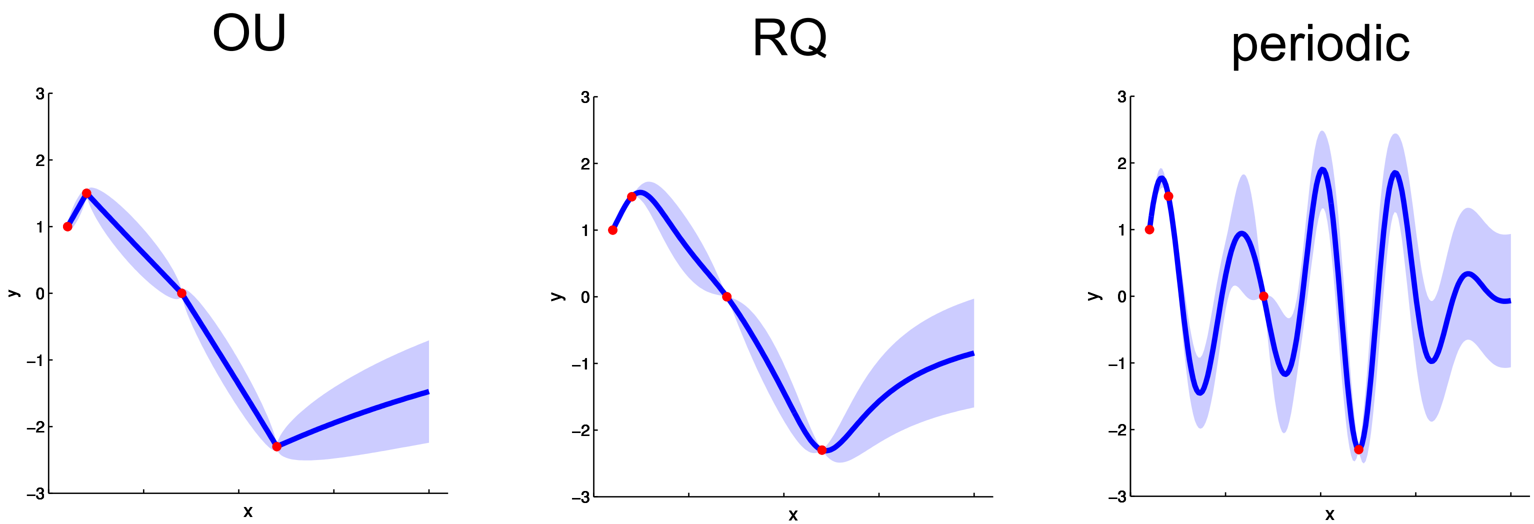


$\Sigma =$ 

The covariance function has a large effect



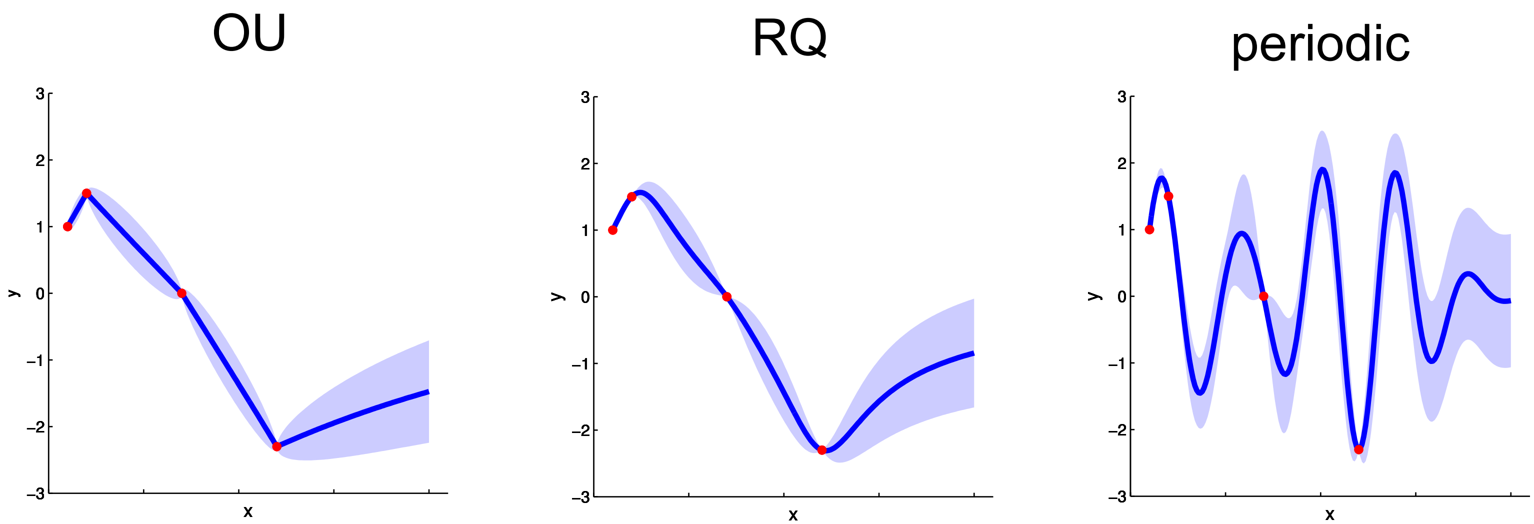
The covariance function has a large effect



Bayesian model comparison:

$$p(M|\mathbf{y}_{1:N}) = \frac{p(\mathbf{y}_{1:N}|M)p(M)}{\sum_{M'} p(\mathbf{y}_{1:N}|M')p(M')}$$

The covariance function has a large effect

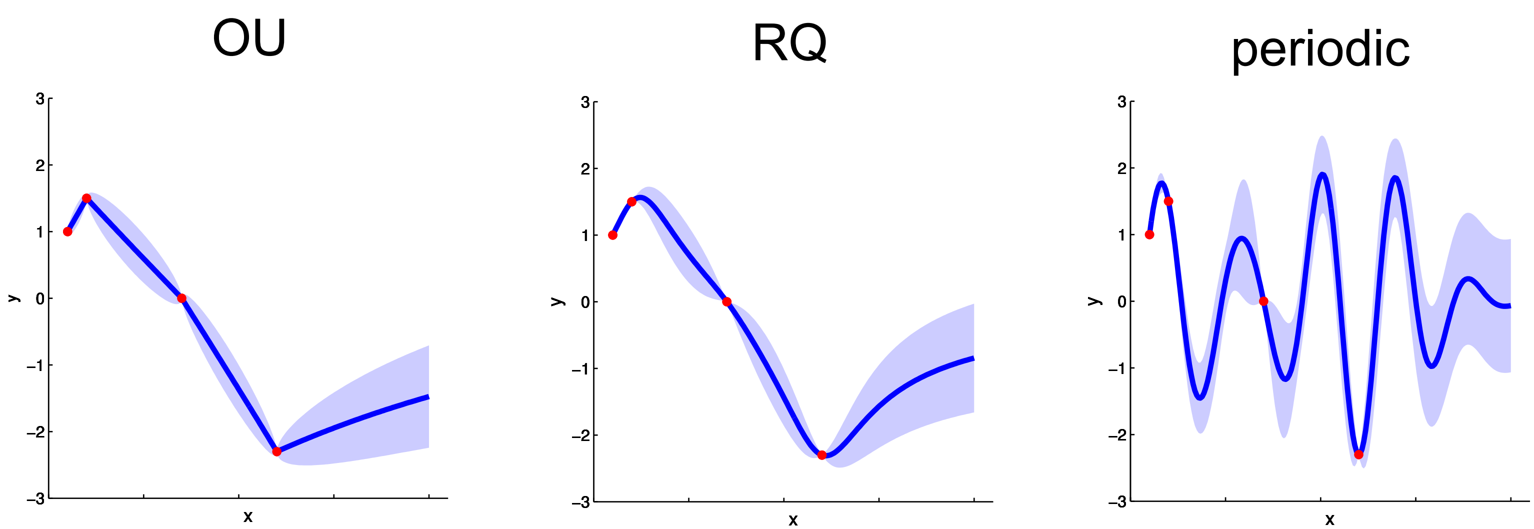


Bayesian model comparison:

$$p(M|\mathbf{y}_{1:N}) = \frac{p(\mathbf{y}_{1:N}|M)p(M)}{\sum_{M'} p(\mathbf{y}_{1:N}|M')p(M')}$$

prior over models

The covariance function has a large effect



Bayesian model comparison:

$$p(M|\mathbf{y}_{1:N}) = \frac{p(\mathbf{y}_{1:N}|M)p(M)}{\sum_{M'} p(\mathbf{y}_{1:N}|M')p(M')}$$

prior over models

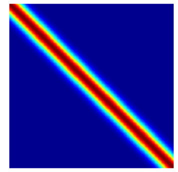
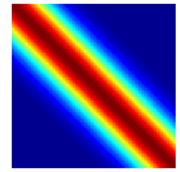
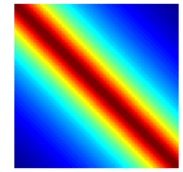
marginal likelihood $p(\mathbf{y}_{1:N}|M) = \int \mathrm{d}\theta \; p(\mathbf{y}_{1:N}|\theta, M)p(\theta|M)$

Making new covariance functions from old

(positive) linear combinations
of covariance functions

Making new covariance functions from old

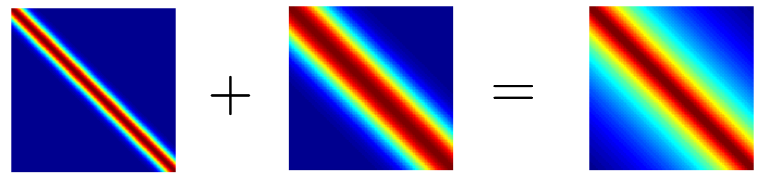
(positive) linear combinations
of covariance functions

 +  = 

e.g. scale mixture of SE = rational quadratic

Making new covariance functions from old

(positive) linear combinations of covariance functions

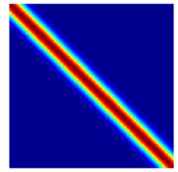
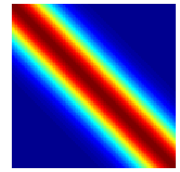
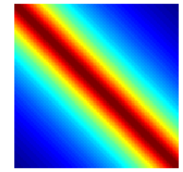


e.g. scale mixture of SE = rational quadratic

multiplication of covariance functions

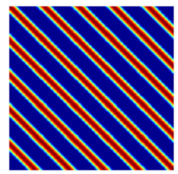
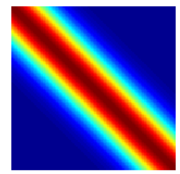
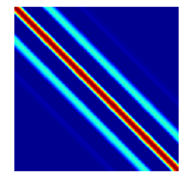
Making new covariance functions from old

(positive) linear combinations of covariance functions

 $+$  $=$ 

e.g. scale mixture of SE $=$ rational quadratic

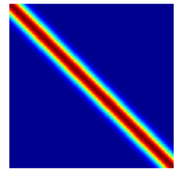
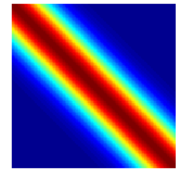
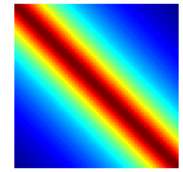
multiplication of covariance functions

 $\times$  $=$ 

e.g. periodic SE $\cos(\omega\Delta \mathsf{x})\exp(-\frac{1}{2l^2}\Delta \mathsf{x}^2)$

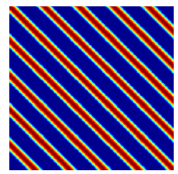
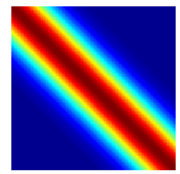
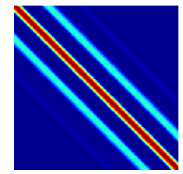
Making new covariance functions from old

(positive) linear combinations of covariance functions

 $+$  $=$ 

e.g. scale mixture of SE $=$ rational quadratic

multiplication of covariance functions

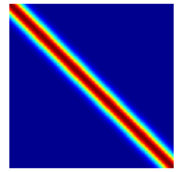
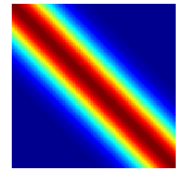
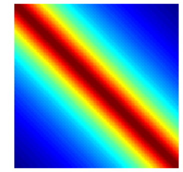
 $\times$  $=$ 

e.g. periodic SE $\cos(\omega \Delta \mathrm{x}) \exp(-\frac{1}{2l^2} \Delta \mathrm{x}^2)$

derivative of GP = GP

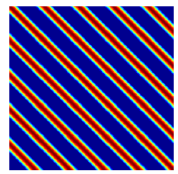
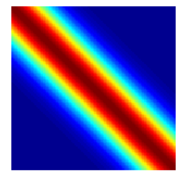
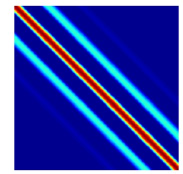
Making new covariance functions from old

(positive) linear combinations of covariance functions

 $+$  $=$ 

e.g. scale mixture of SE $=$ rational quadratic

multiplication of covariance functions

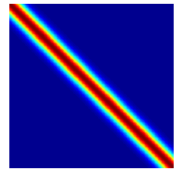
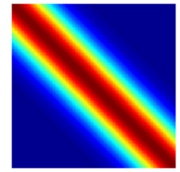
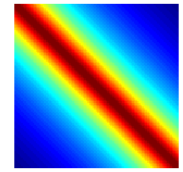
 $\times$  $=$ 

e.g. periodic SE $\cos(\omega \Delta x) \exp(-\frac{1}{2l^2} \Delta x^2)$

derivative of GP = GP $\frac{d}{dx} y(x)$

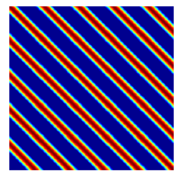
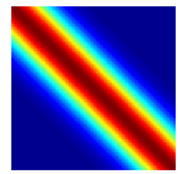
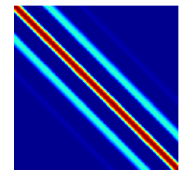
Making new covariance functions from old

(positive) linear combinations of covariance functions

 $+$  $=$ 

e.g. scale mixture of SE $=$ rational quadratic

multiplication of covariance functions

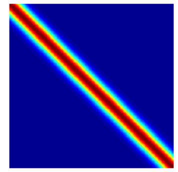
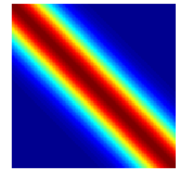
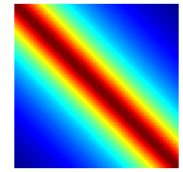
 $\times$  $=$ 

e.g. periodic SE $\cos(\omega\Delta \mathsf{x})\exp(-\frac{1}{2l^2}\Delta \mathsf{x}^2)$

derivative of GP = GP $\frac{\mathsf{d}}{\mathsf{dx}}\mathsf{y}(\mathsf{x}) = \frac{\mathsf{d}}{\mathsf{dx}}\sum_{k=1}^{\infty}\gamma_k \mathsf{g}_k(\mathsf{x})$

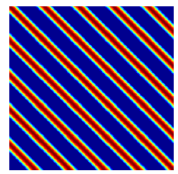
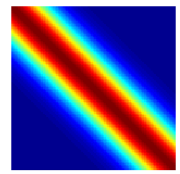
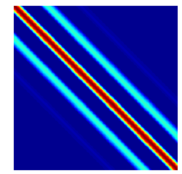
Making new covariance functions from old

(positive) linear combinations of covariance functions

 $+$  $=$ 

e.g. scale mixture of SE $=$ rational quadratic

multiplication of covariance functions

 $\times$  $=$ 

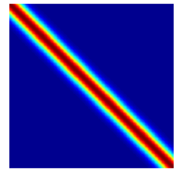
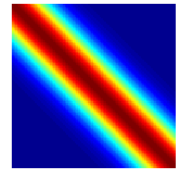
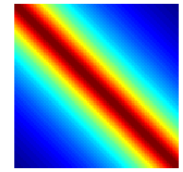
e.g. periodic SE $\cos(\omega \Delta \mathsf{x}) \exp(-\frac{1}{2l^2} \Delta \mathsf{x}^2)$

derivative of GP = GP

$$\frac{\mathsf{d}}{\mathsf{dx}} \mathsf{y}(\mathsf{x}) = \frac{\mathsf{d}}{\mathsf{dx}} \sum_{k=1}^{\infty} \gamma_k \mathsf{g}_k(\mathsf{x}) = \sum_{k=1}^{\infty} \gamma_k \frac{\mathsf{d}}{\mathsf{dx}} \mathsf{g}_k(\mathsf{x})$$

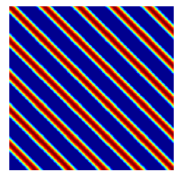
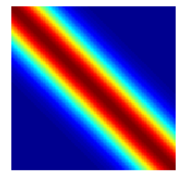
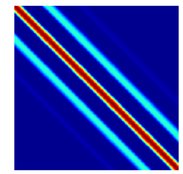
Making new covariance functions from old

(positive) linear combinations of covariance functions

 $+$  $=$ 

e.g. scale mixture of SE $=$ rational quadratic

multiplication of covariance functions

 $\times$  $=$ 

e.g. periodic SE $\cos(\omega \Delta \mathrm{x}) \exp(-\frac{1}{2l^2} \Delta \mathrm{x}^2)$

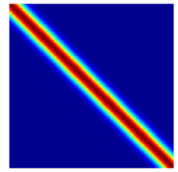
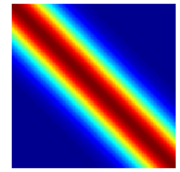
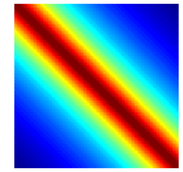
derivative of GP = GP

$$\frac{\mathrm{d}}{\mathrm{dx}} \mathrm{y}(\mathrm{x}) = \frac{\mathrm{d}}{\mathrm{dx}} \sum_{k=1}^{\infty} \gamma_k \mathrm{g}_k(\mathrm{x}) = \sum_{k=1}^{\infty} \gamma_k \frac{\mathrm{d}}{\mathrm{dx}} \mathrm{g}_k(\mathrm{x})$$

new basis: $\mathrm{g}'_k(\mathrm{x}) = \frac{\mathrm{d}}{\mathrm{dx}} \mathrm{g}_k(\mathrm{x})$

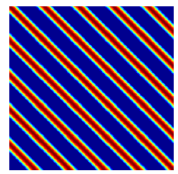
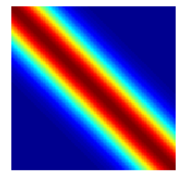
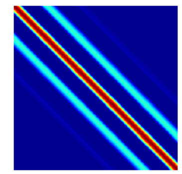
Making new covariance functions from old

(positive) linear combinations of covariance functions

 $+$  $=$ 

e.g. scale mixture of SE $=$ rational quadratic

multiplication of covariance functions

 $\times$  $=$ 

e.g. periodic SE $\cos(\omega \Delta \mathsf{x}) \exp(-\frac{1}{2l^2} \Delta \mathsf{x}^2)$

derivative of GP = GP

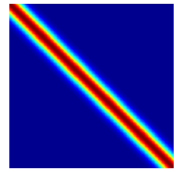
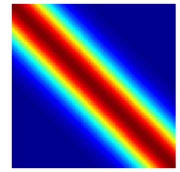
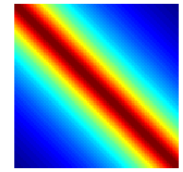
$$\frac{\mathsf{d}}{\mathsf{dx}} \mathsf{y}(\mathsf{x}) = \frac{\mathsf{d}}{\mathsf{dx}} \sum_{k=1}^{\infty} \gamma_k \mathsf{g}_k(\mathsf{x}) = \sum_{k=1}^{\infty} \gamma_k \frac{\mathsf{d}}{\mathsf{dx}} \mathsf{g}_k(\mathsf{x})$$

new basis: $\mathsf{g}'_k(\mathsf{x}) = \frac{\mathsf{d}}{\mathsf{dx}} \mathsf{g}_k(\mathsf{x})$

new covariance: $\mathsf{K}'(\mathsf{x}, \mathsf{x}') = \frac{\mathsf{d}}{\mathsf{dx}} \frac{\mathsf{d}}{\mathsf{dx}'} \mathsf{K}(\mathsf{x}, \mathsf{x}')$

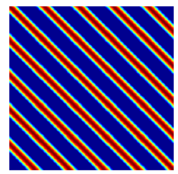
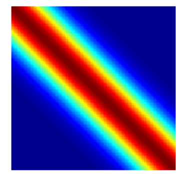
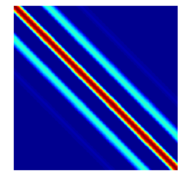
Making new covariance functions from old

(positive) linear combinations of covariance functions

 $+$  $=$ 

e.g. scale mixture of SE $=$ rational quadratic

multiplication of covariance functions

 $\times$  $=$ 

e.g. periodic SE $\cos(\omega \Delta \mathsf{x}) \exp(-\frac{1}{2l^2} \Delta \mathsf{x}^2)$

derivative of GP = GP

$$\frac{\mathsf{d}}{\mathsf{dx}} \mathsf{y}(\mathsf{x}) = \frac{\mathsf{d}}{\mathsf{dx}} \sum_{k=1}^{\infty} \gamma_k \mathsf{g}_k(\mathsf{x}) = \sum_{k=1}^{\infty} \gamma_k \frac{\mathsf{d}}{\mathsf{dx}} \mathsf{g}_k(\mathsf{x})$$

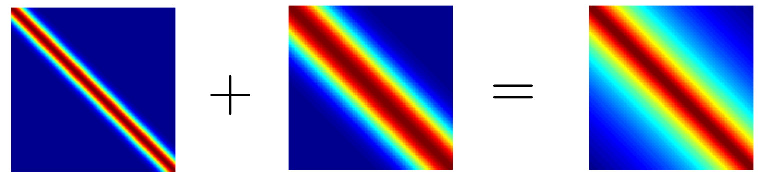
new basis: $\mathsf{g}'_k(\mathsf{x}) = \frac{\mathsf{d}}{\mathsf{dx}} \mathsf{g}_k(\mathsf{x})$

new covariance: $\mathsf{K}'(\mathsf{x}, \mathsf{x}') = \frac{\mathsf{d}}{\mathsf{dx}} \frac{\mathsf{d}}{\mathsf{dx}'} \mathsf{K}(\mathsf{x}, \mathsf{x}')$

integral of GP = GP

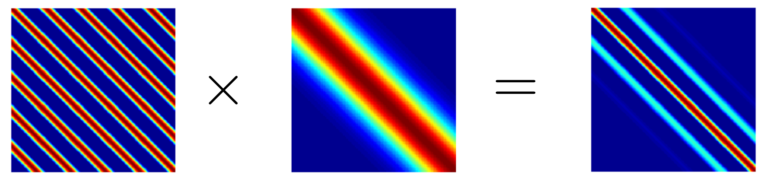
Making new covariance functions from old

(positive) linear combinations of covariance functions



e.g. scale mixture of SE = rational quadratic

multiplication of covariance functions



e.g. periodic SE $\cos(\omega \Delta \mathrm{x}) \exp(-\frac{1}{2l^2} \Delta \mathrm{x}^2)$

derivative of GP = GP

$$\frac{\mathrm{d}}{\mathrm{dx}} \mathrm{y}(\mathrm{x}) = \frac{\mathrm{d}}{\mathrm{dx}} \sum_{k=1}^{\infty} \gamma_k \mathrm{g}_k(\mathrm{x}) = \sum_{k=1}^{\infty} \gamma_k \frac{\mathrm{d}}{\mathrm{dx}} \mathrm{g}_k(\mathrm{x})$$

new basis: $\mathrm{g}'_k(\mathrm{x}) = \frac{\mathrm{d}}{\mathrm{dx}} \mathrm{g}_k(\mathrm{x})$

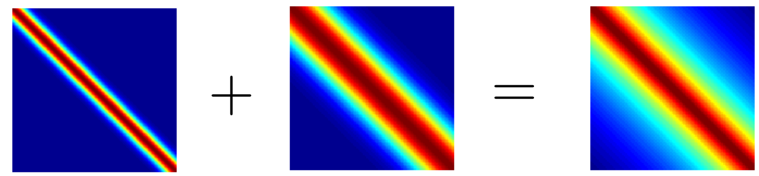
new covariance: $\mathrm{K}'(\mathrm{x}, \mathrm{x}') = \frac{\mathrm{d}}{\mathrm{dx}} \frac{\mathrm{d}}{\mathrm{dx}'} \mathrm{K}(\mathrm{x}, \mathrm{x}')$

integral of GP = GP

$$\int \mathrm{dx}\; \mathrm{y}(\mathrm{x}) = \sum_{k=1}^{\infty} \gamma_k \int \mathrm{dx}\; \mathrm{g}_k(\mathrm{x})$$

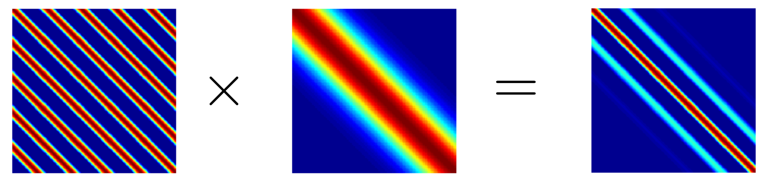
Making new covariance functions from old

(positive) linear combinations of covariance functions



e.g. scale mixture of SE $=$ rational quadratic

multiplication of covariance functions



e.g. periodic SE $\cos(\omega \Delta \mathrm{x}) \exp(-\frac{1}{2l^2} \Delta \mathrm{x}^2)$

derivative of GP = GP

$$\frac{\mathrm{d}}{\mathrm{dx}} \mathrm{y}(\mathrm{x}) = \frac{\mathrm{d}}{\mathrm{dx}} \sum_{k=1}^{\infty} \gamma_k \mathrm{g}_k(\mathrm{x}) = \sum_{k=1}^{\infty} \gamma_k \frac{\mathrm{d}}{\mathrm{dx}} \mathrm{g}_k(\mathrm{x})$$

new basis: $\mathrm{g}'_k(\mathrm{x}) = \frac{\mathrm{d}}{\mathrm{dx}} \mathrm{g}_k(\mathrm{x})$

new covariance: $\mathrm{K}'(\mathrm{x}, \mathrm{x}') = \frac{\mathrm{d}}{\mathrm{dx}} \frac{\mathrm{d}}{\mathrm{dx}'} \mathrm{K}(\mathrm{x}, \mathrm{x}')$

integral of GP = GP

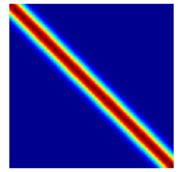
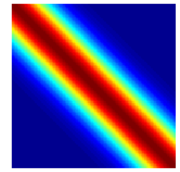
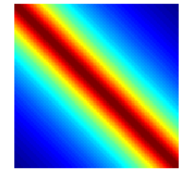
$$\int \mathrm{dx}\ \mathrm{y}(\mathrm{x}) = \sum_{k=1}^{\infty} \gamma_k \int \mathrm{dx}\ \mathrm{g}_k(\mathrm{x})$$

new basis: $\mathrm{g}'_k(\mathrm{x}) = \int \mathrm{dx}\ \mathrm{g}_k(\mathrm{x})$

new covariance: $\mathrm{K}'(\mathrm{x}, \mathrm{x}') = \int \int \mathrm{dx dx}'\ \mathrm{K}(\mathrm{x}, \mathrm{x}')$

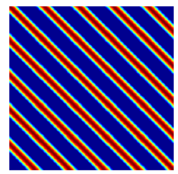
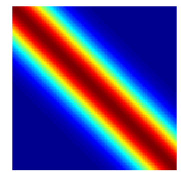
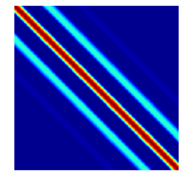
Making new covariance functions from old

(positive) linear combinations of covariance functions

 $+$  $=$ 

e.g. scale mixture of SE $=$ rational quadratic

multiplication of covariance functions

 $\times$  $=$ 

e.g. periodic SE $\cos(\omega \Delta \mathrm{x}) \exp(-\frac{1}{2l^2} \Delta \mathrm{x}^2)$

derivative of GP = GP

integral of GP = GP

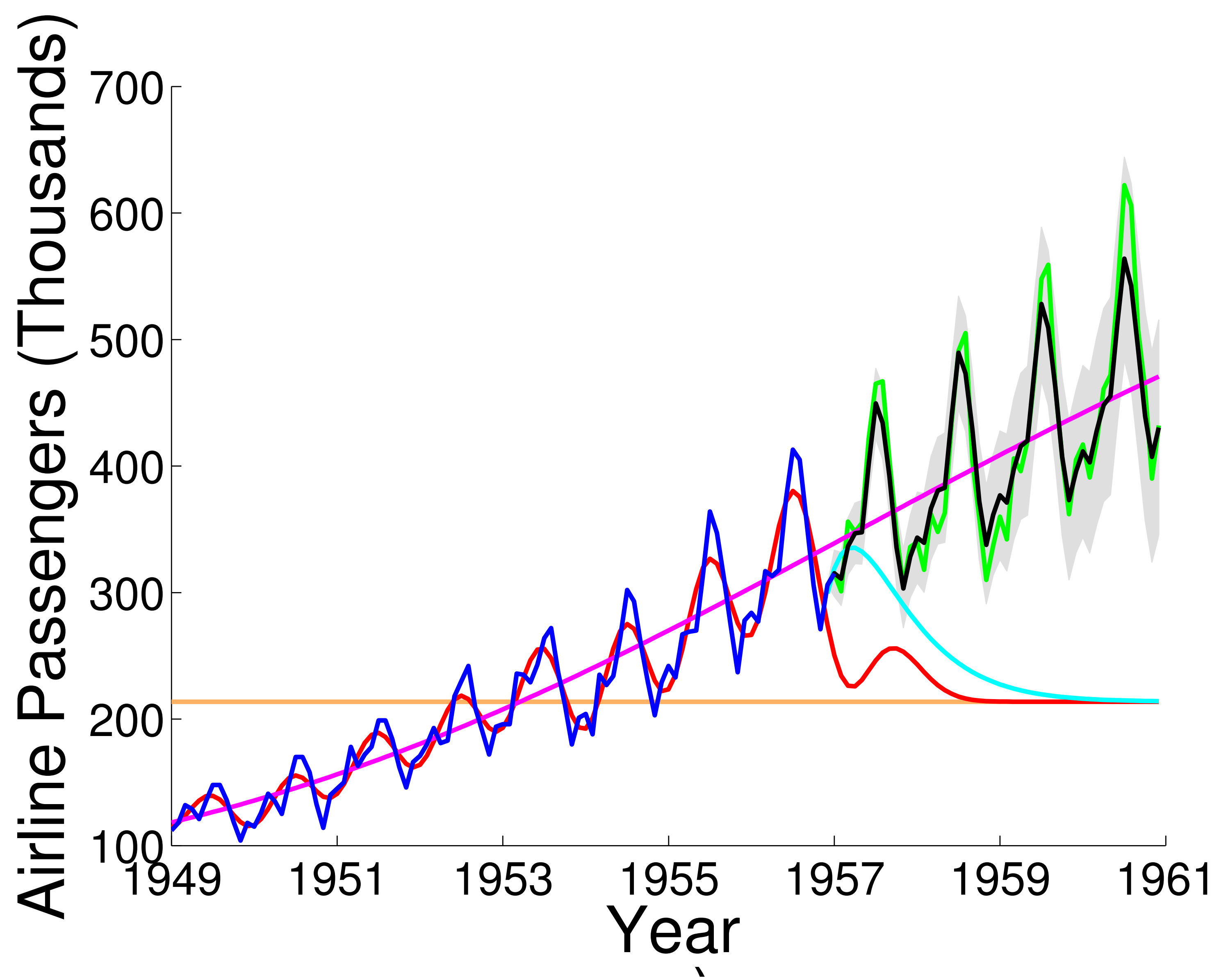
filtering a GP = GP

$\mathrm{V(x)} \otimes \mathrm{y(x)}$

$\mathrm{K'(x, x')} = \mathrm{V(x)} \otimes \mathrm{K(x, x')} \otimes \mathrm{V(x')}$

Meta-kernel, learn which is right directly

$$k(\tau) = \sum_{q=1}^{Q} w_q \prod_{p=1}^{P} \exp\{-2\pi^2 \tau_p^2 v_q^{(p)}\} \cos(2\pi \tau_p \mu_q^{(p)}).$$



How do we choose the hyper-parameters?

idea: use probability distributions to represent plausibility of hyper-parameters (uncertainty) given the data

How do we choose the hyper-parameters?

idea: use probability distributions to represent plausibility of hyper-parameters (uncertainty) given the data

$$p(\theta|\mathbf{y}_{1:N}) = \frac{p(\mathbf{y}_{1:N}|\theta)p(\theta)}{p(\mathbf{y}_{1:N})}$$ (Bayes' Rule)

How do we choose the hyper-parameters?

idea: use probability distributions to represent plausibility of hyper-parameters (uncertainty) given the data

what we know after seeing the data $\propto$ what the data tell us (likelihood) $\times$ what we knew before seeing the data (prior)

$$p(\theta|\mathbf{y}_{1:N}) = \frac{p(\mathbf{y}_{1:N}|\theta)p(\theta)}{p(\mathbf{y}_{1:N})}$$ (Bayes' Rule)

How do we choose the hyper-parameters?

idea: use probability distributions to represent plausibility of hyper-parameters (uncertainty) given the data

what we know after seeing the data $\propto$ what the data tell us (likelihood) $\times$ what we knew before seeing the data (prior)

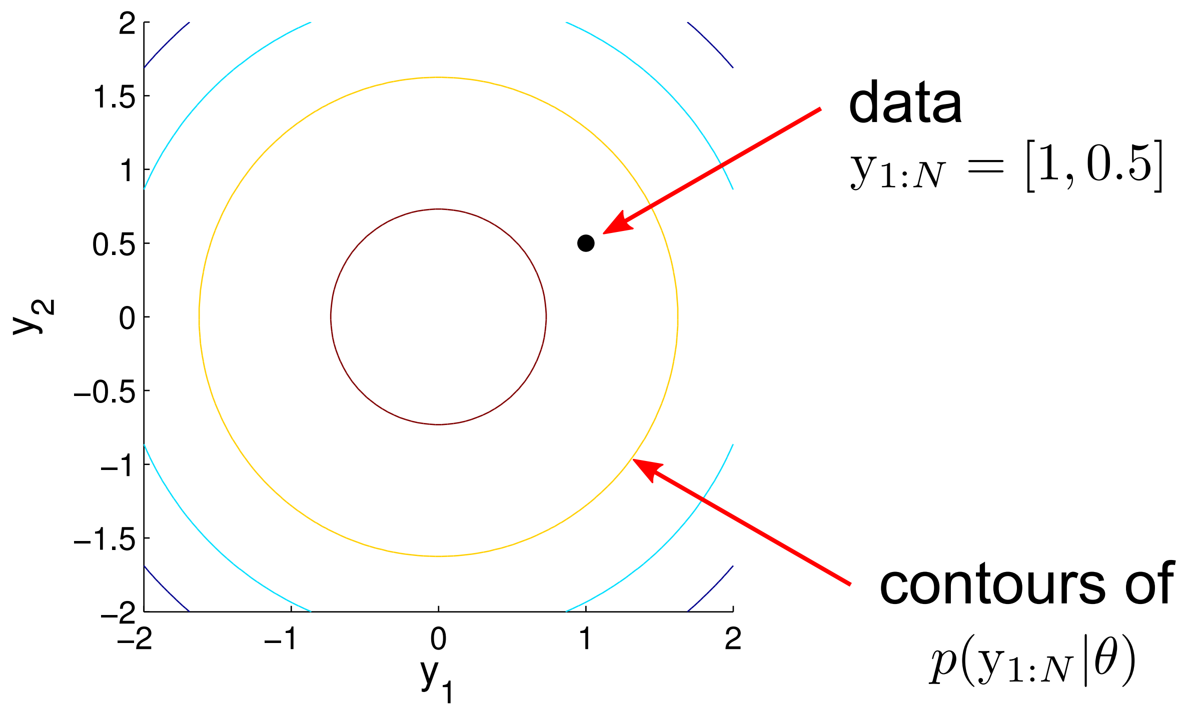
$$p(\theta|\mathbf{y}_{1:N}) = \frac{p(\mathbf{y}_{1:N}|\theta)p(\theta)}{p(\mathbf{y}_{1:N})}$$ (Bayes' Rule)

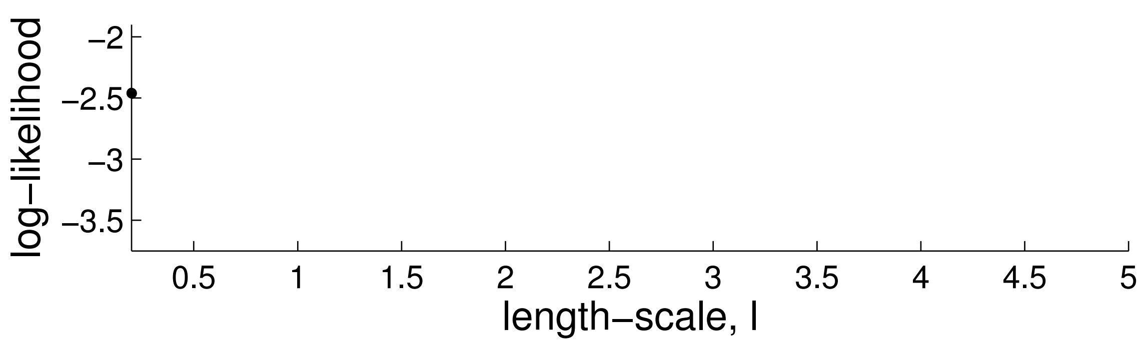
$p(\mathbf{y}_{1:N}|\theta)$ = likelihood of the parameters

= how well did θ predict the data we observed

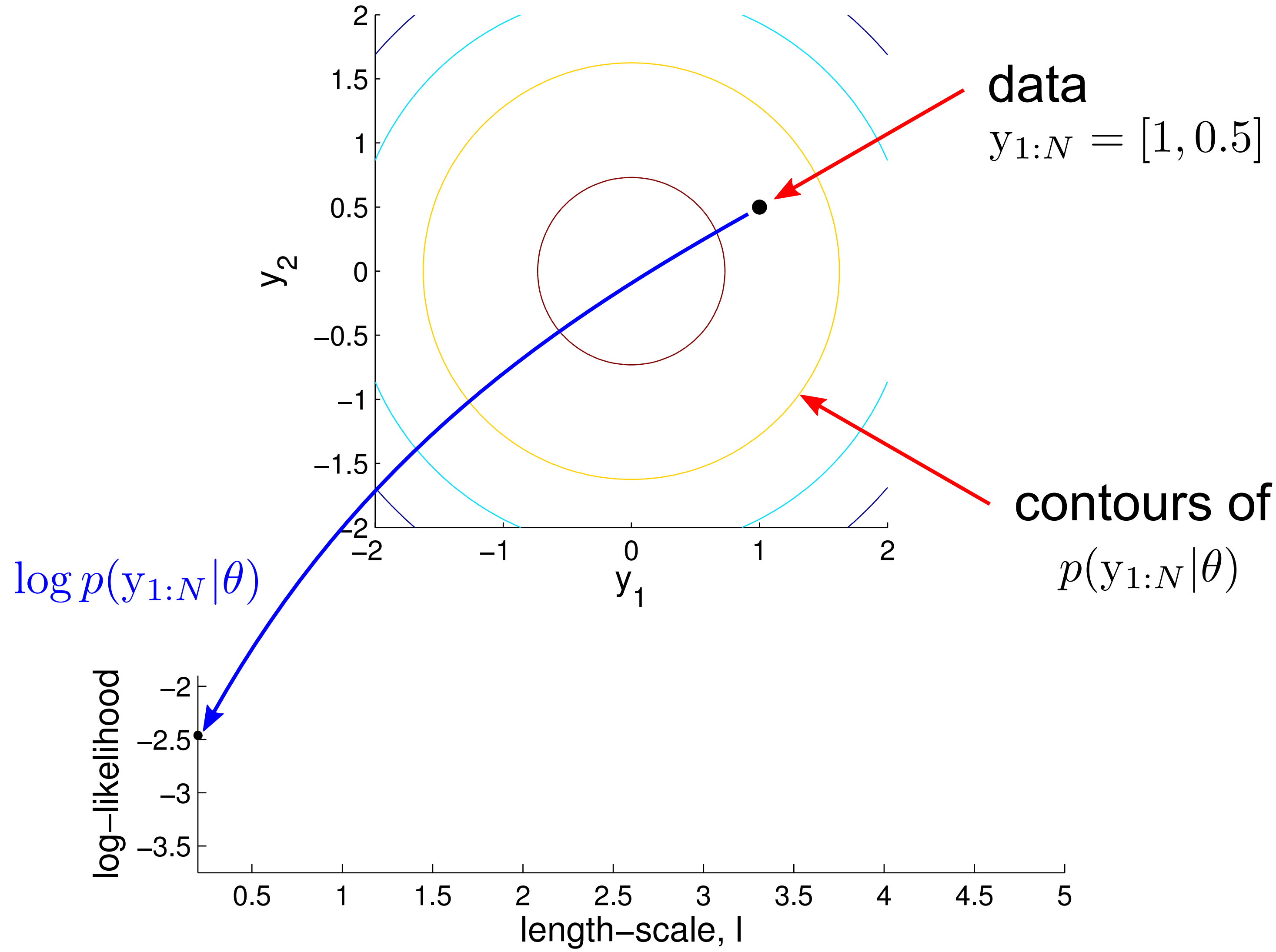
$$p(\mathbf{y}_{1:N}|\theta) = \frac{1}{\det(2\pi\Sigma(\theta))^{-1/2}} \exp\left(-\frac{1}{2}\mathbf{y}_{1:N}^{\top}\Sigma^{-1}(\theta)\mathbf{y}_{1:N}\right)$$

How do we choose the hyper-parameters?

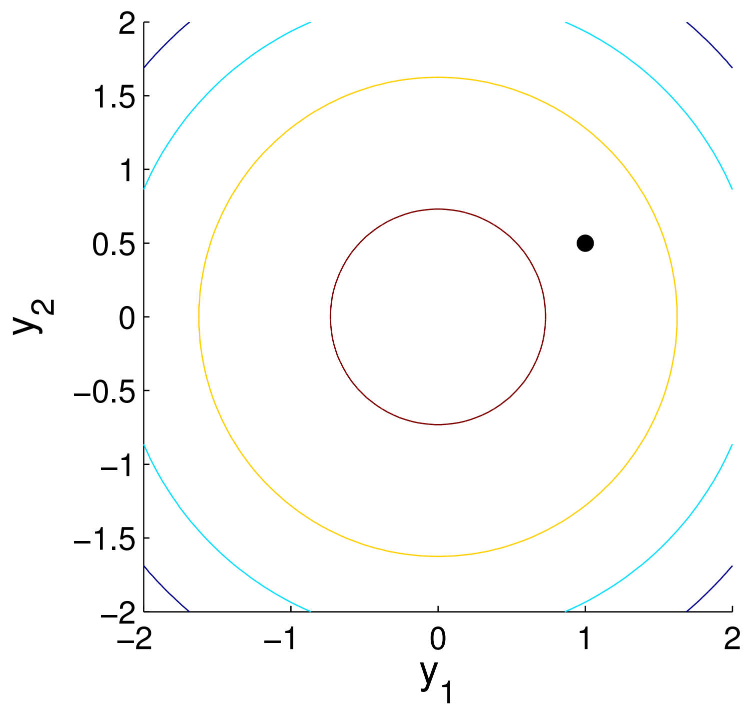


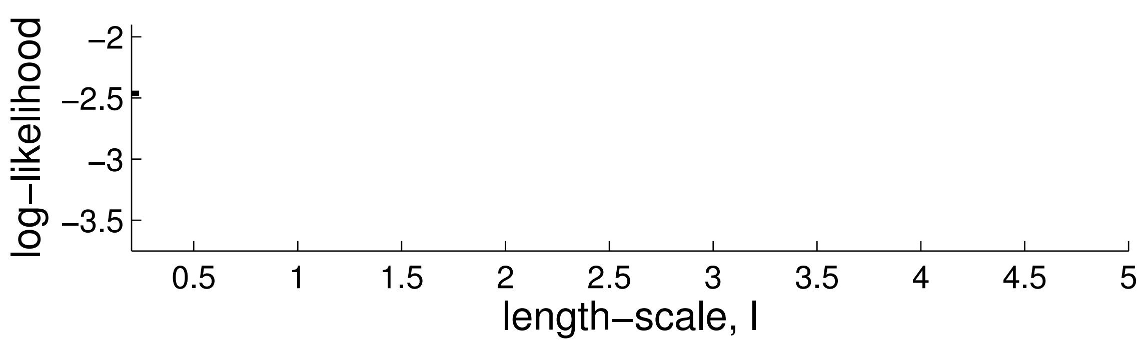


How do we choose the hyper-parameters?

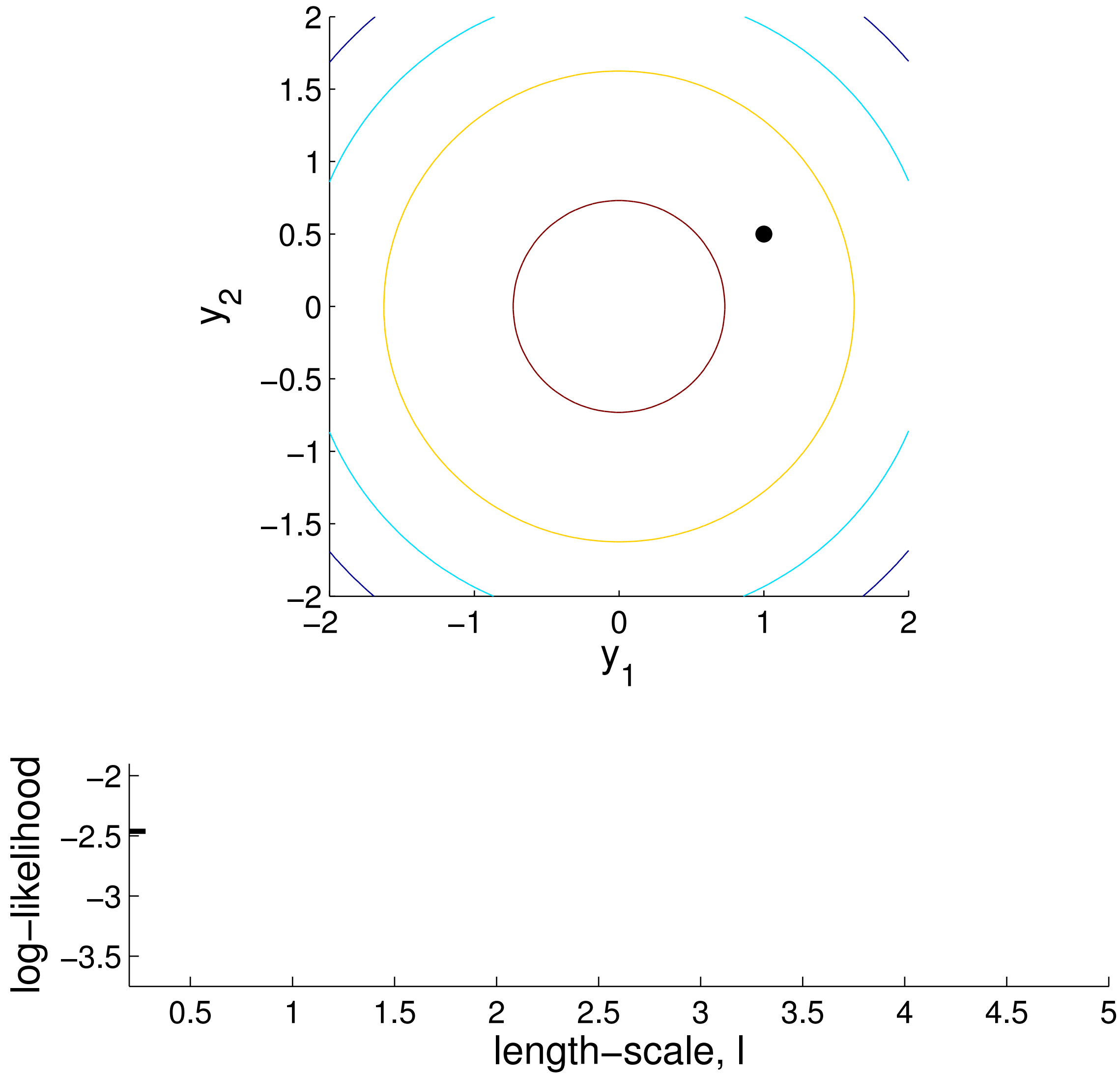


How do we choose the hyper-parameters?

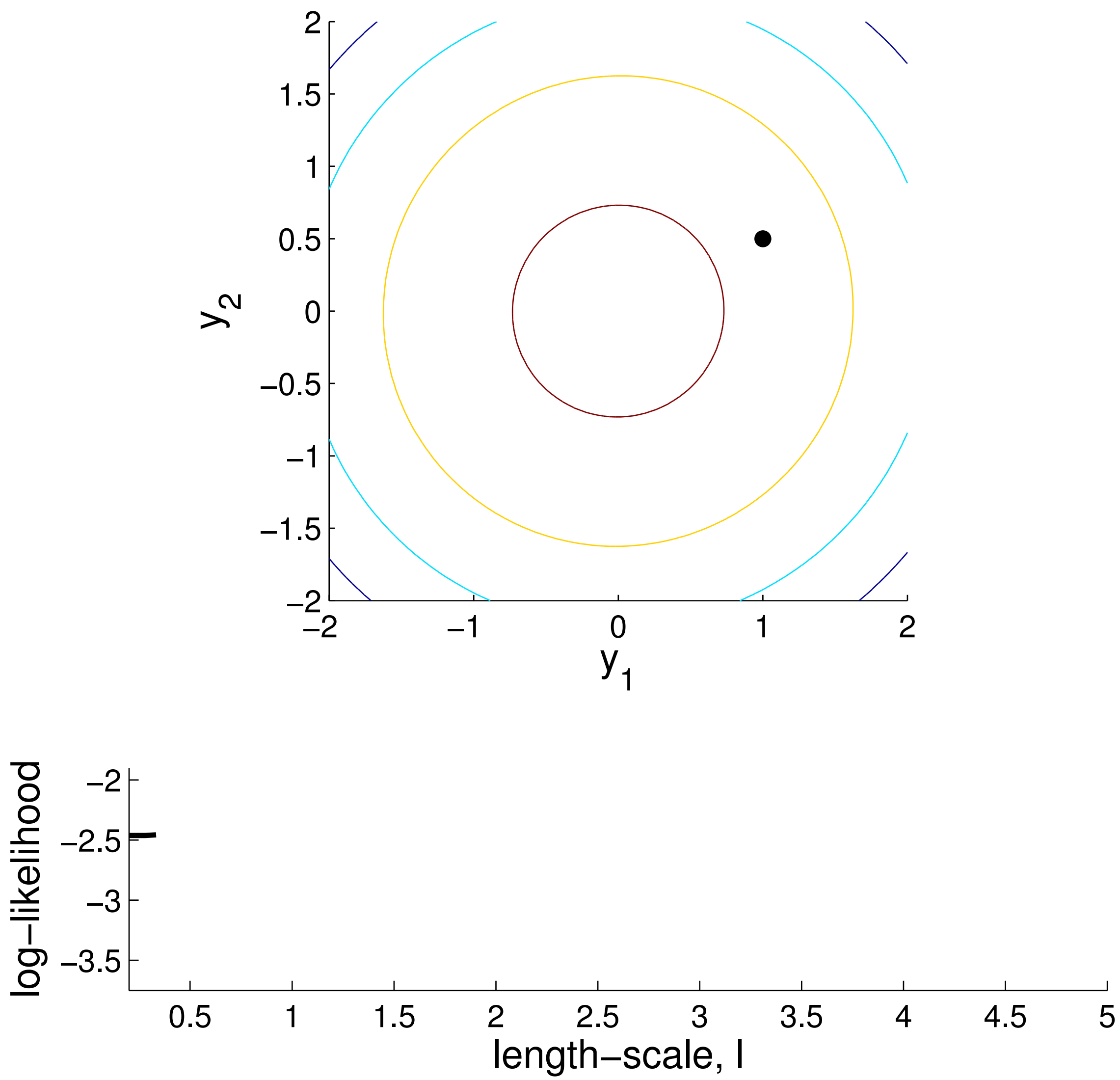




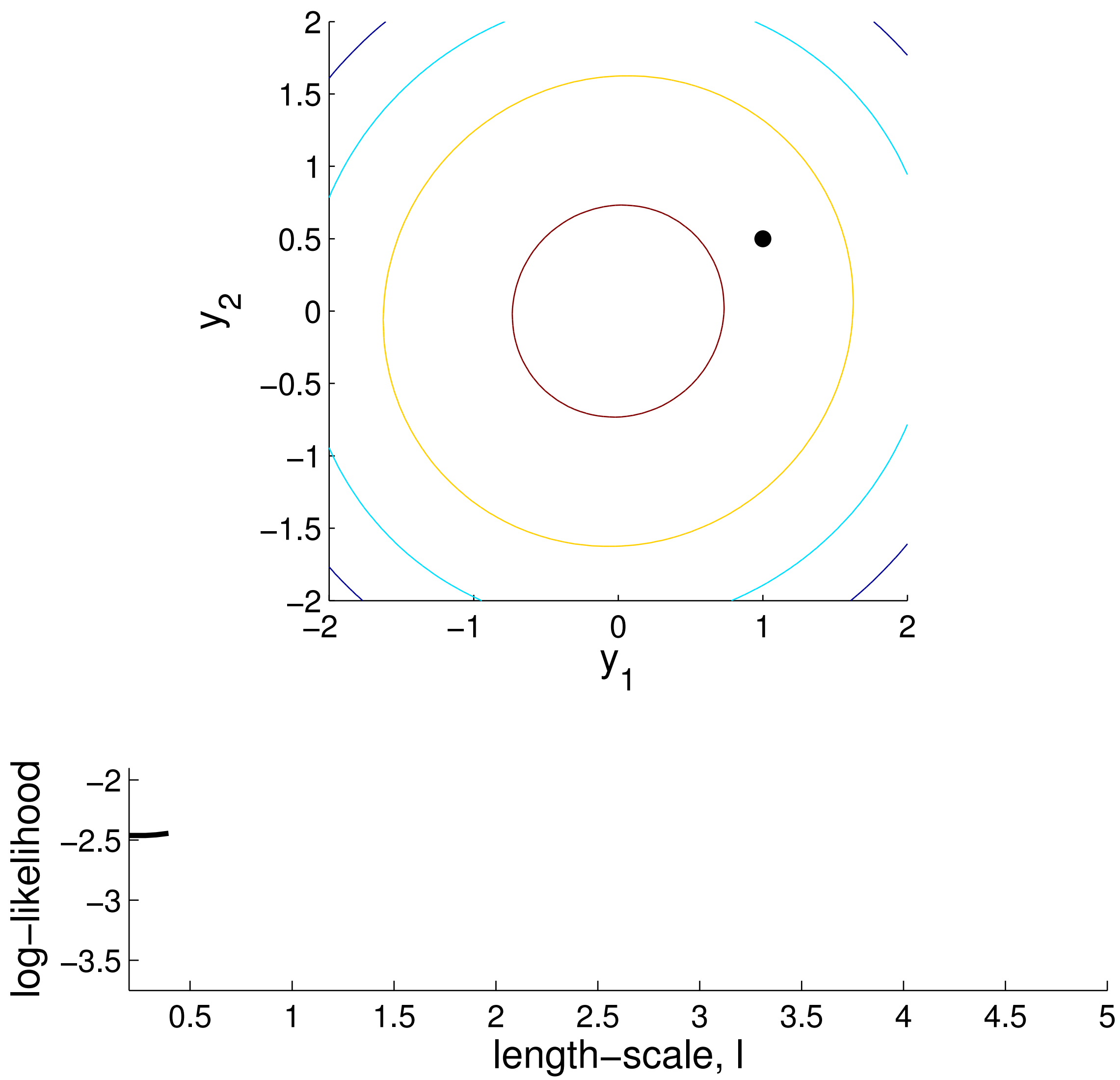
How do we choose the hyper-parameters?



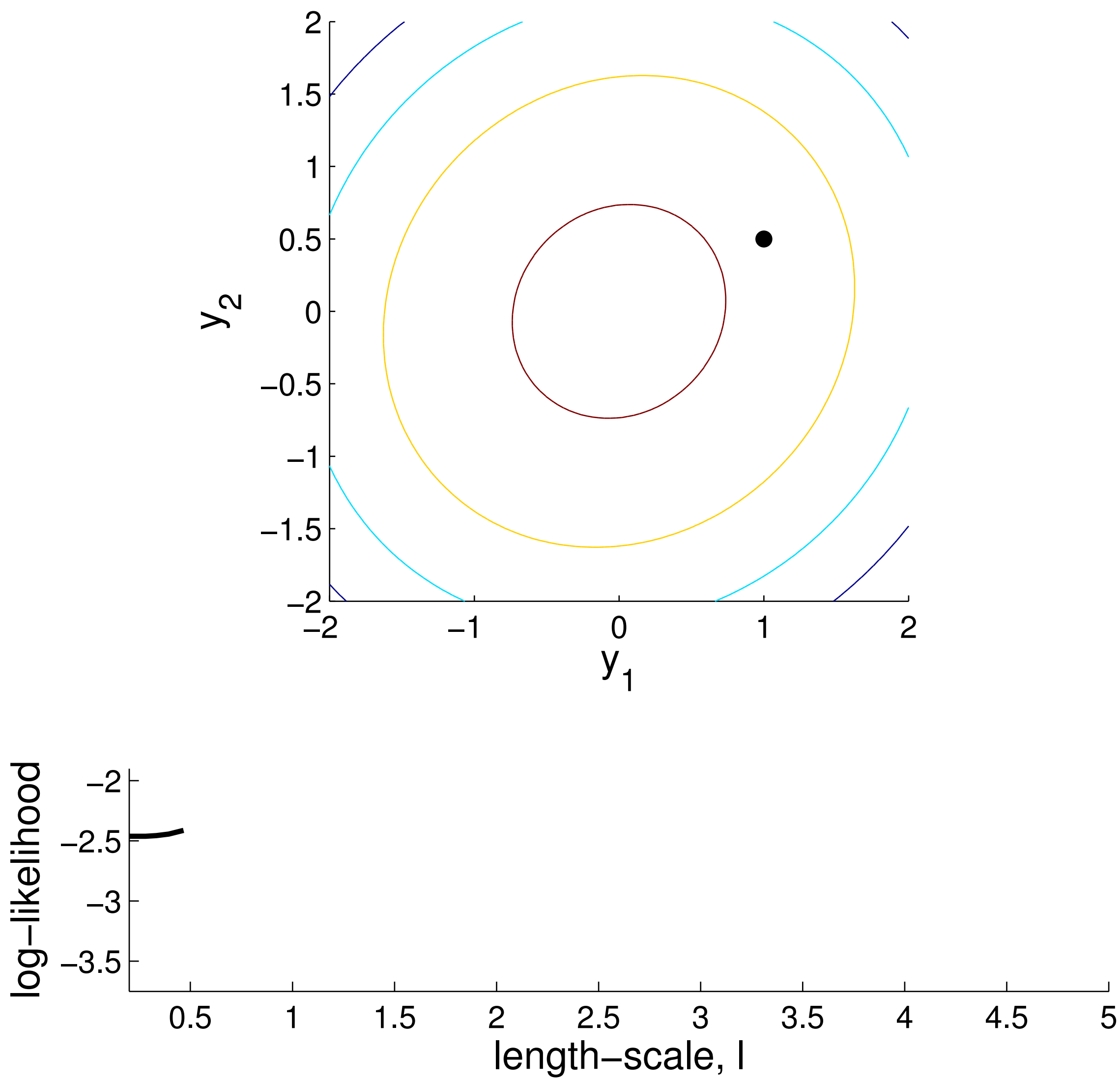
How do we choose the hyper-parameters?



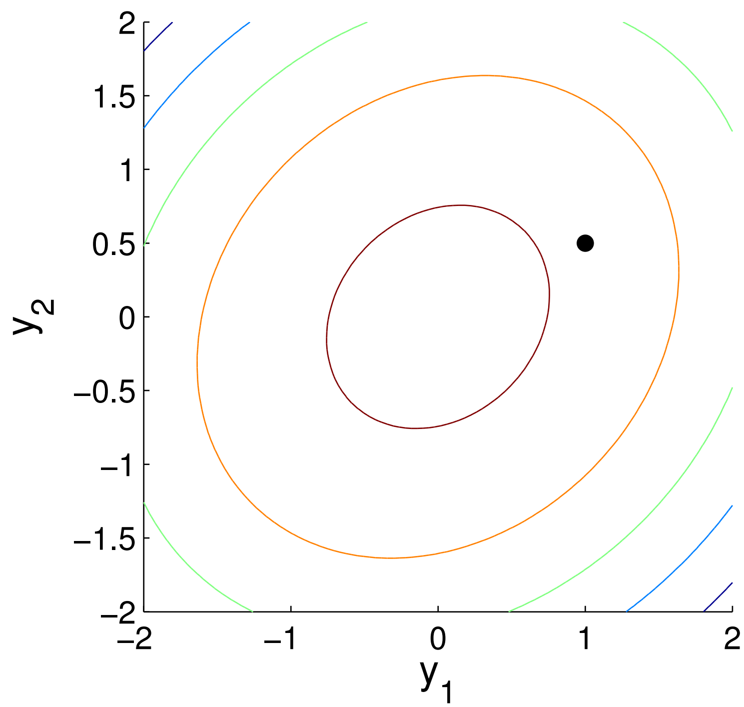
How do we choose the hyper-parameters?

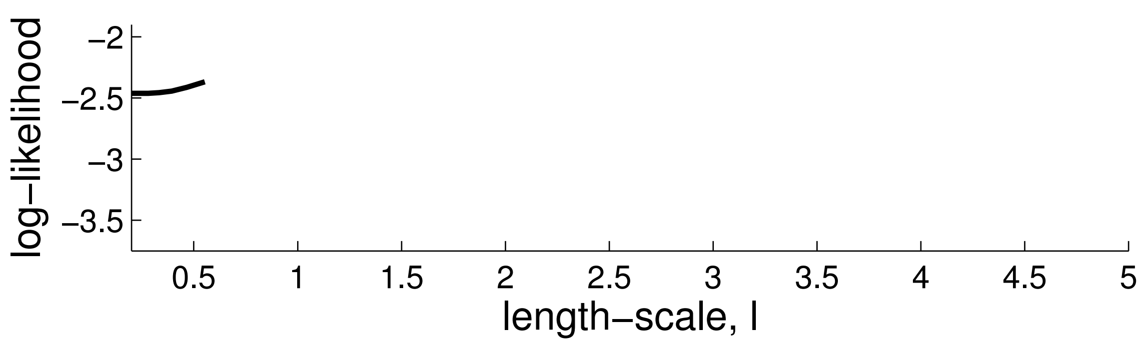


How do we choose the hyper-parameters?

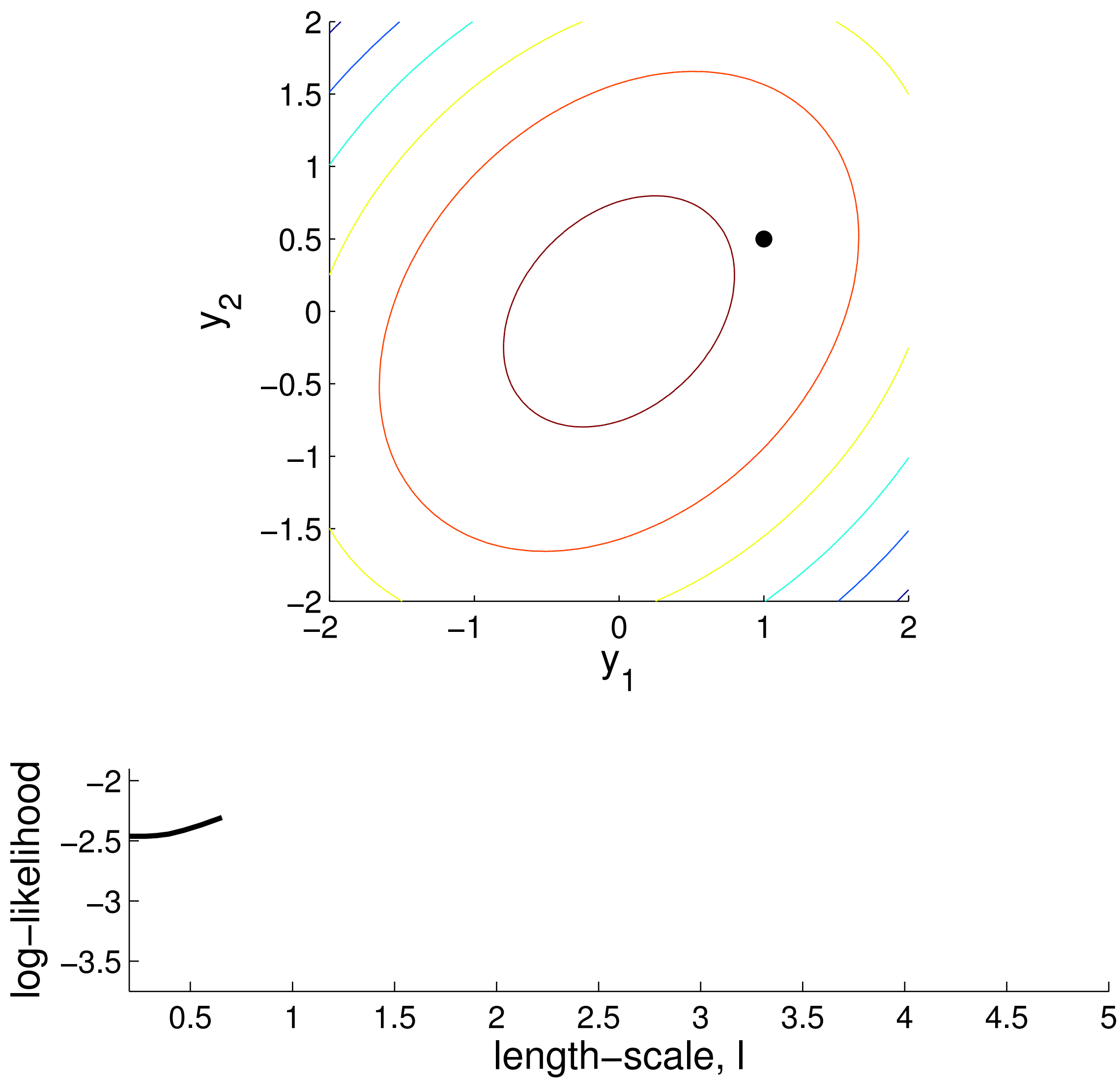


How do we choose the hyper-parameters?

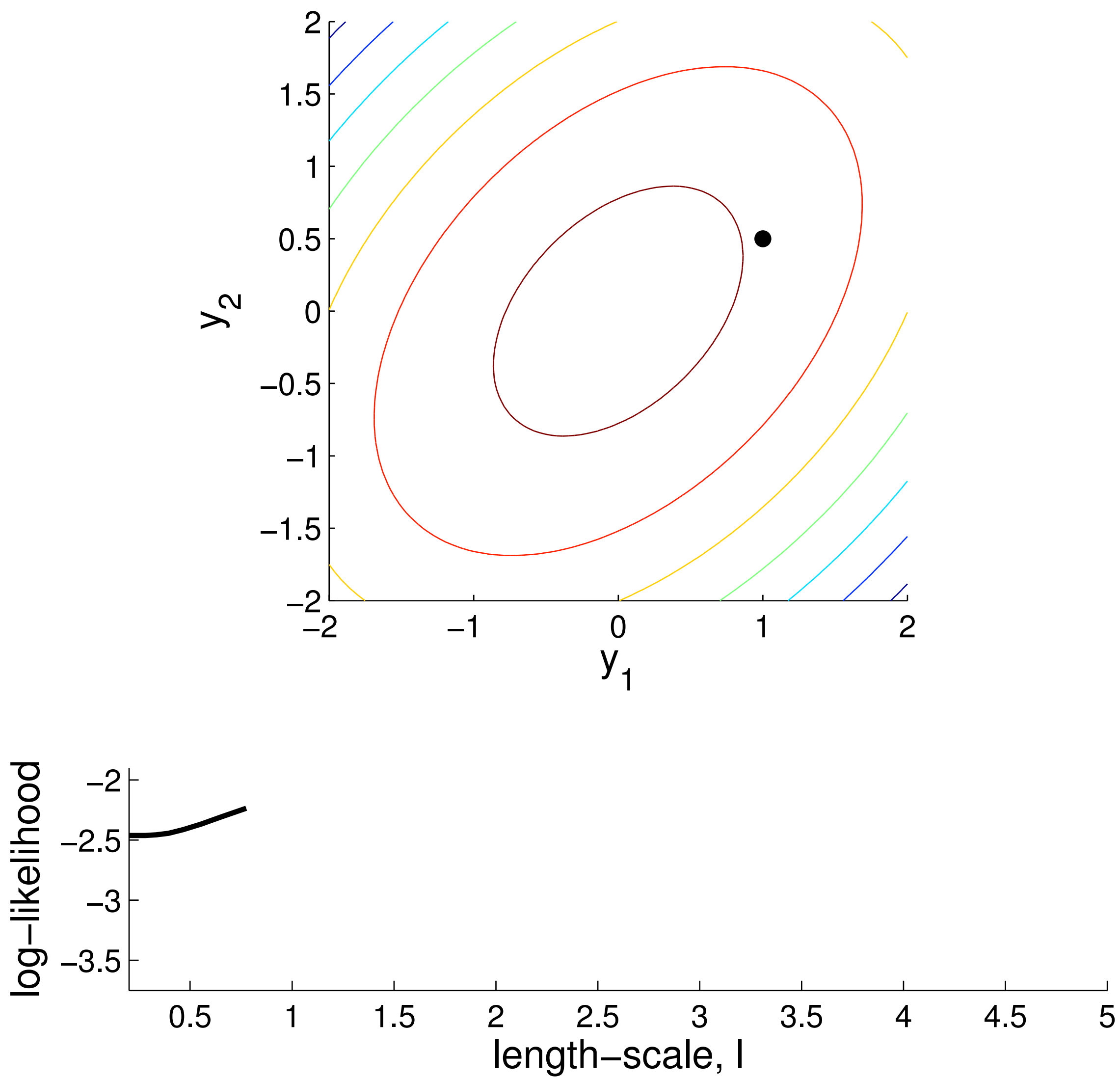




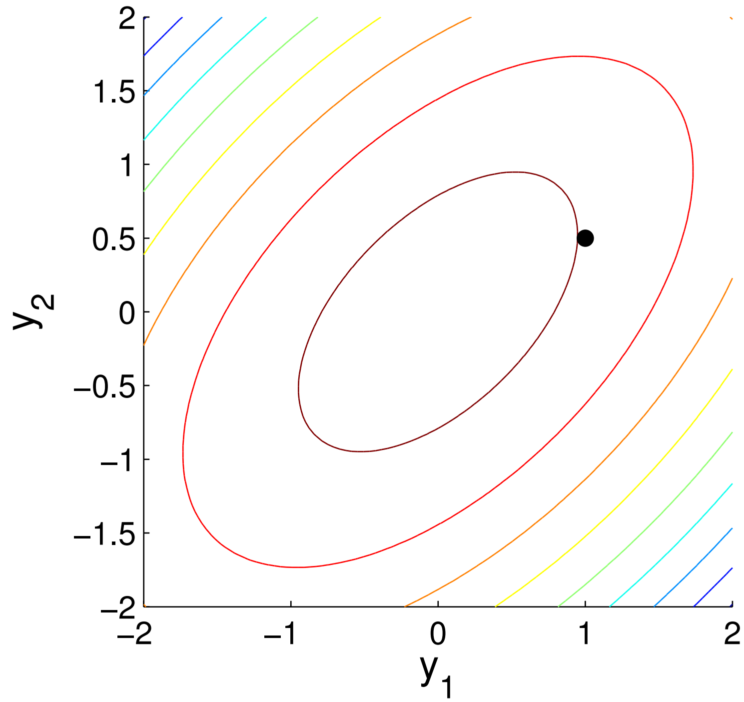
How do we choose the hyper-parameters?

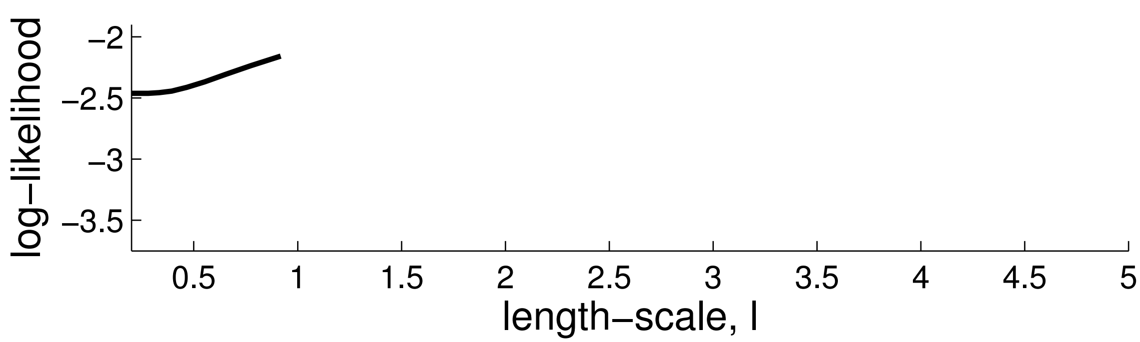


How do we choose the hyper-parameters?

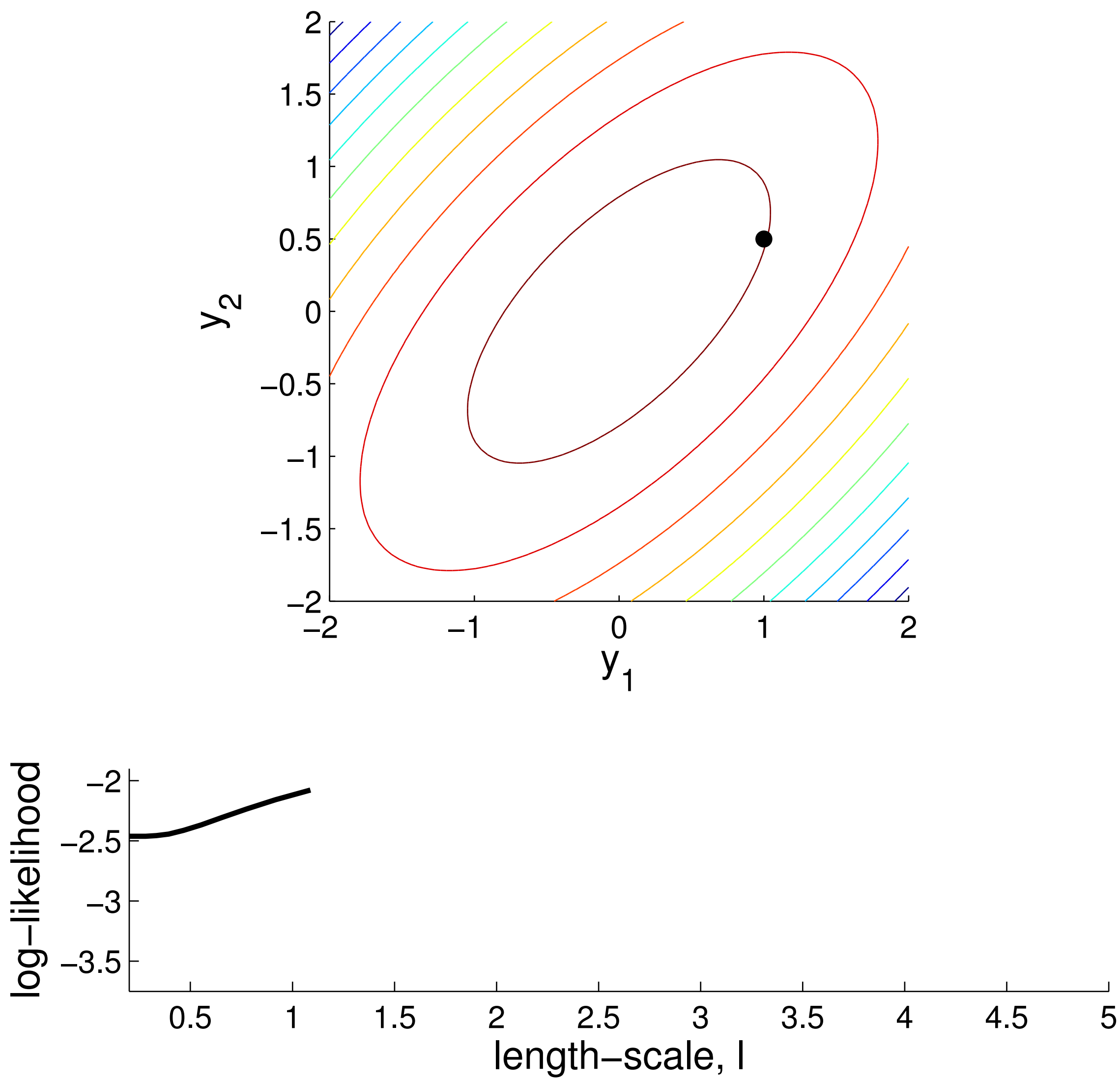


How do we choose the hyper-parameters?

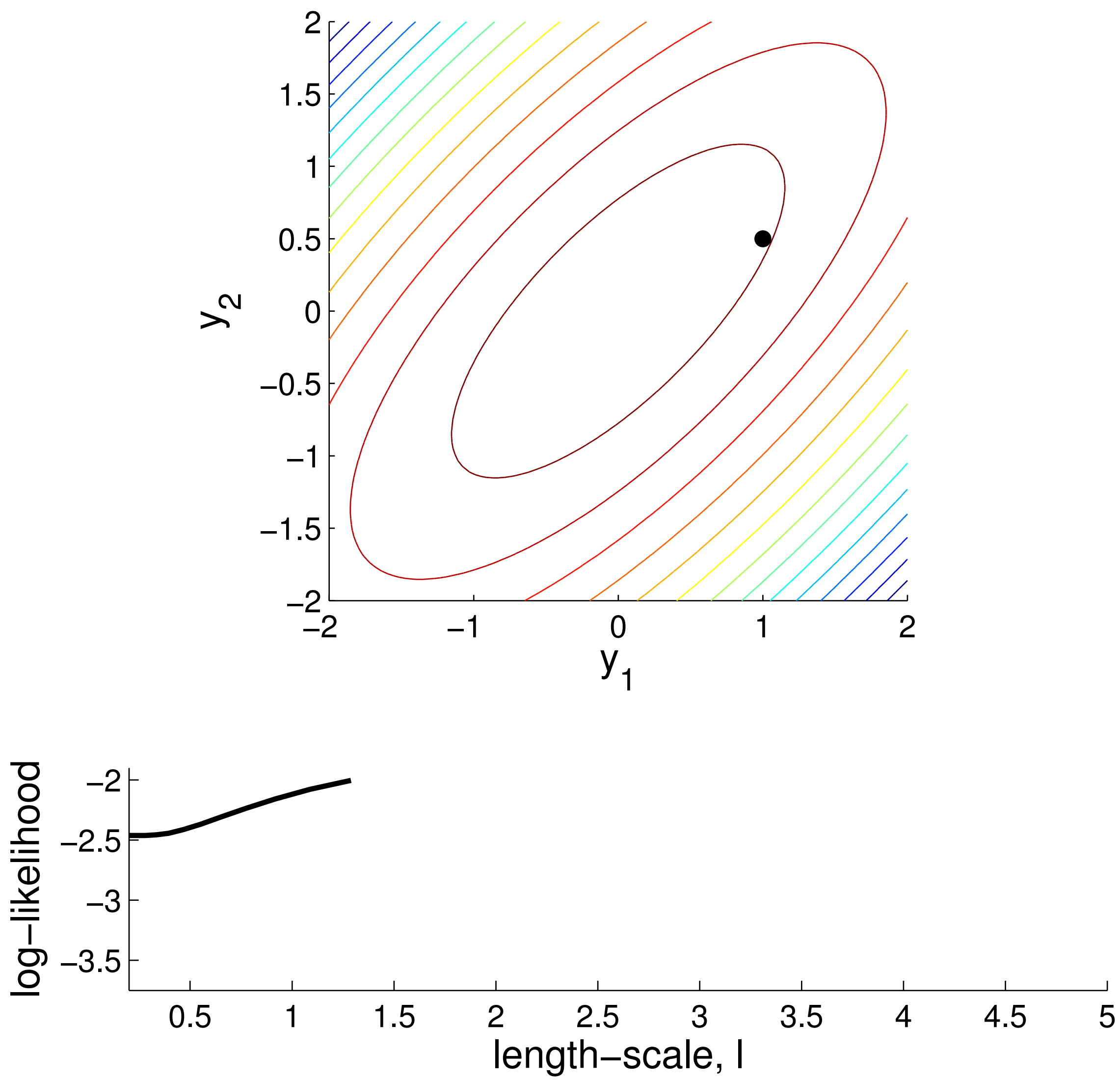




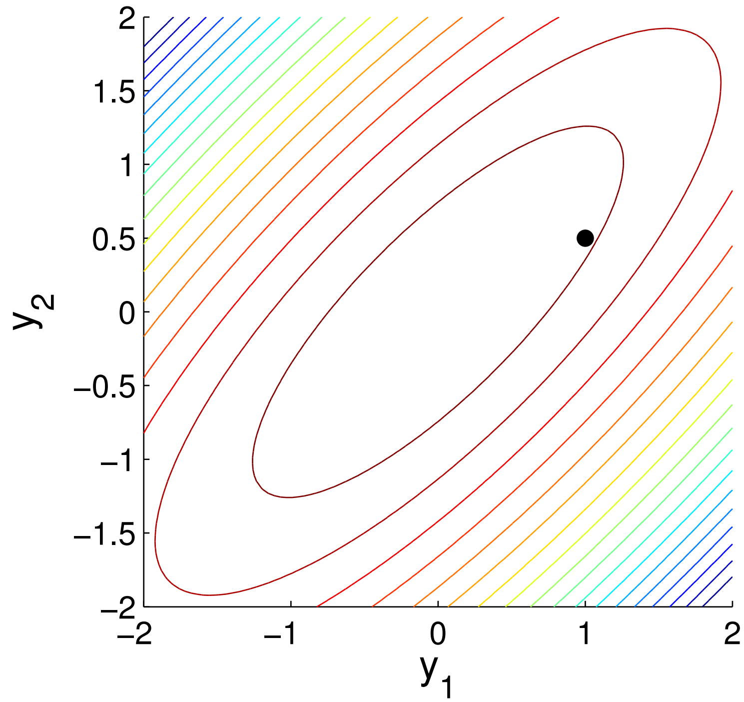
How do we choose the hyper-parameters?

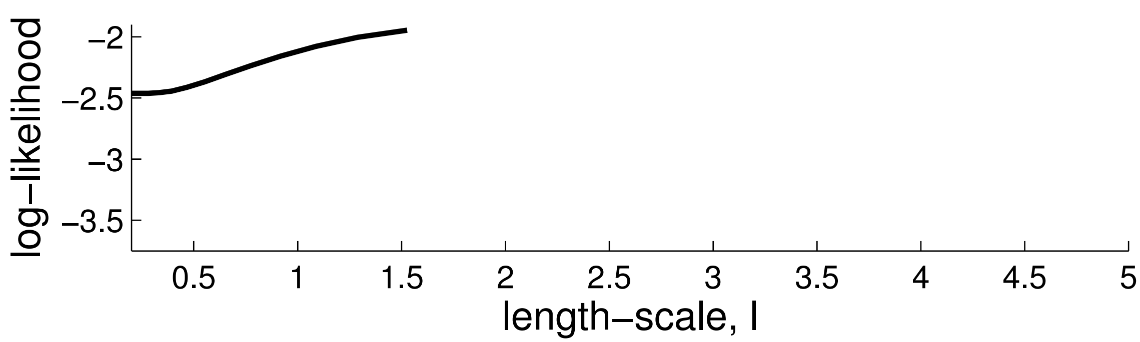


How do we choose the hyper-parameters?

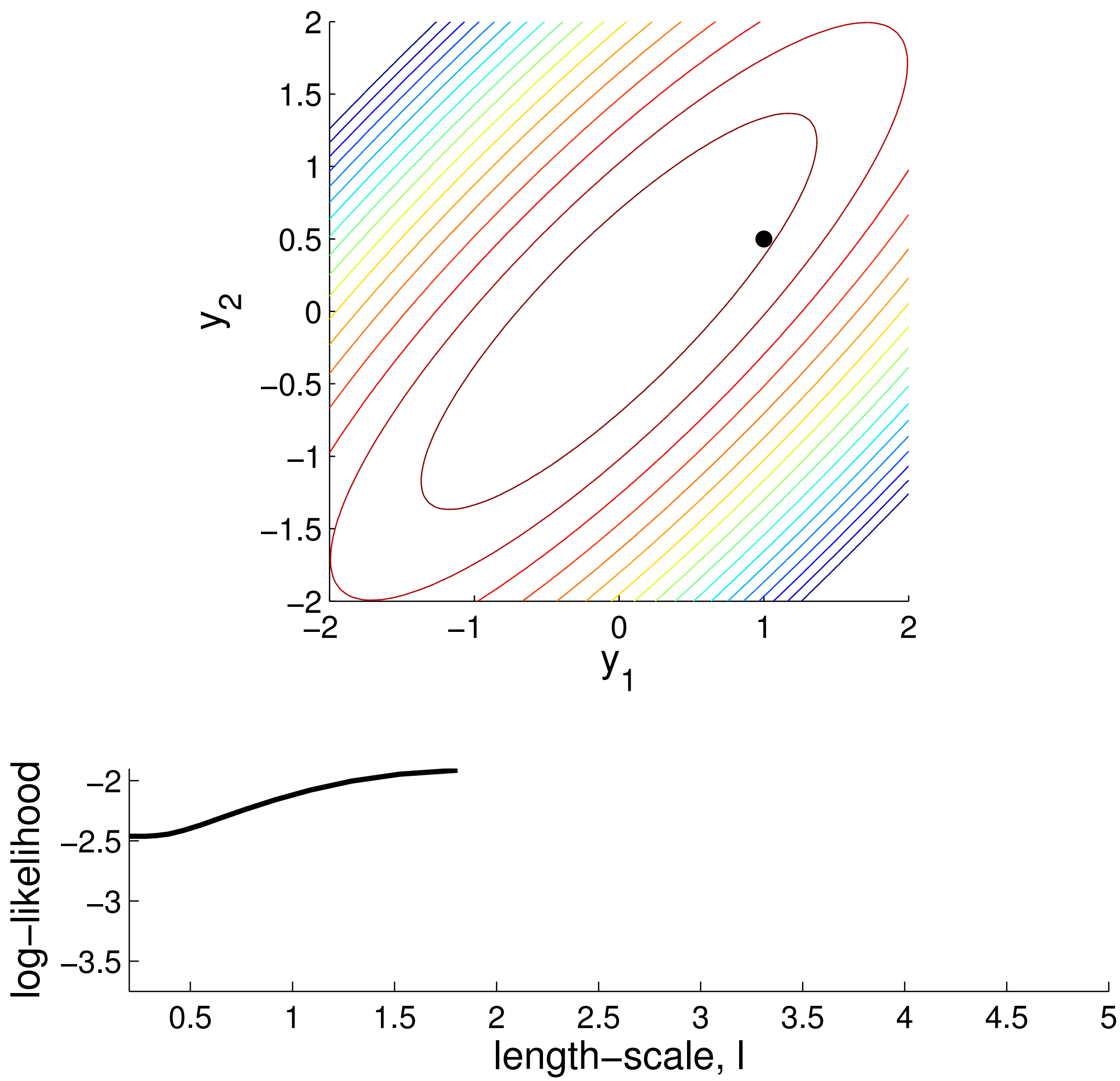


How do we choose the hyper-parameters?

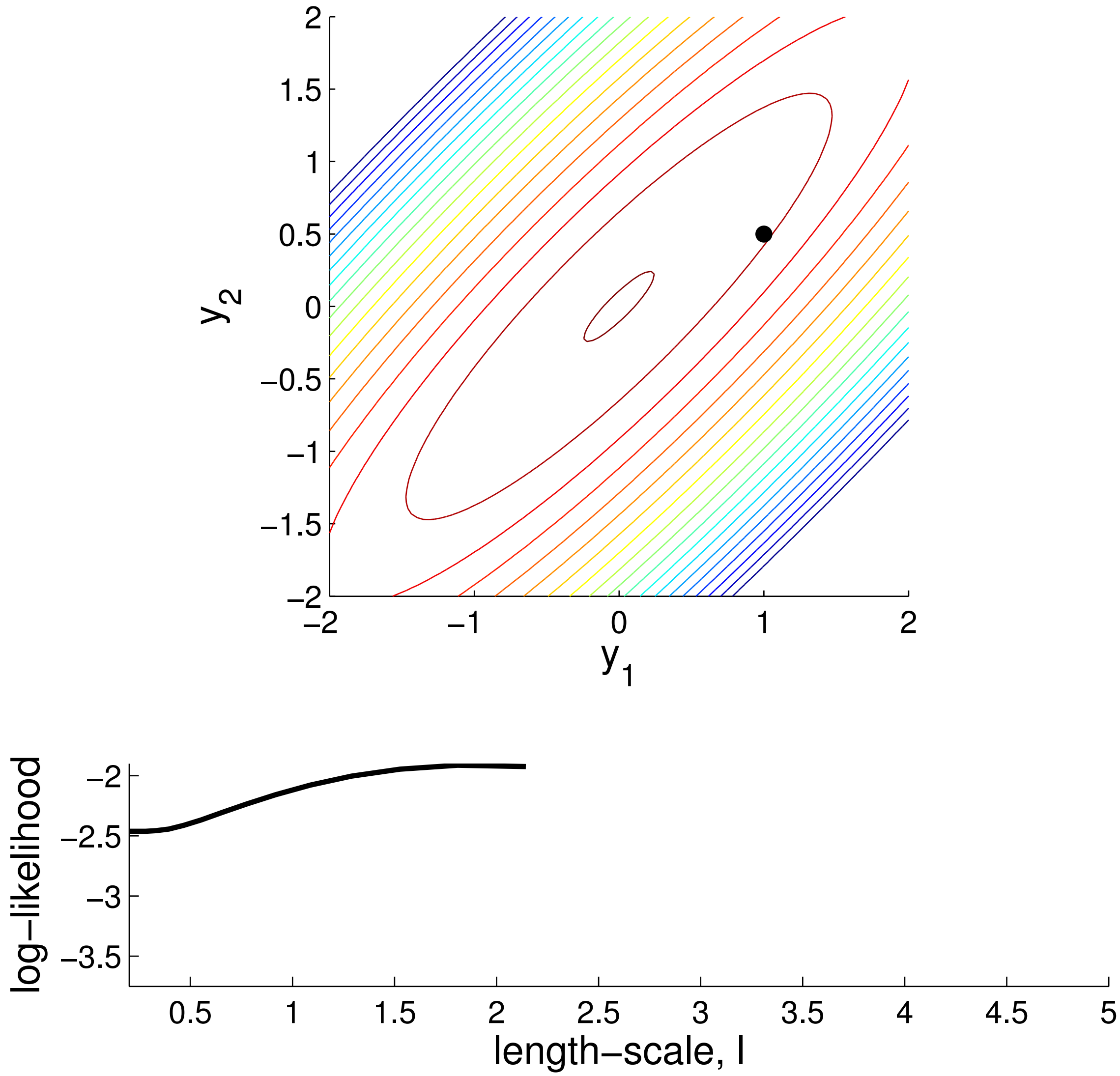




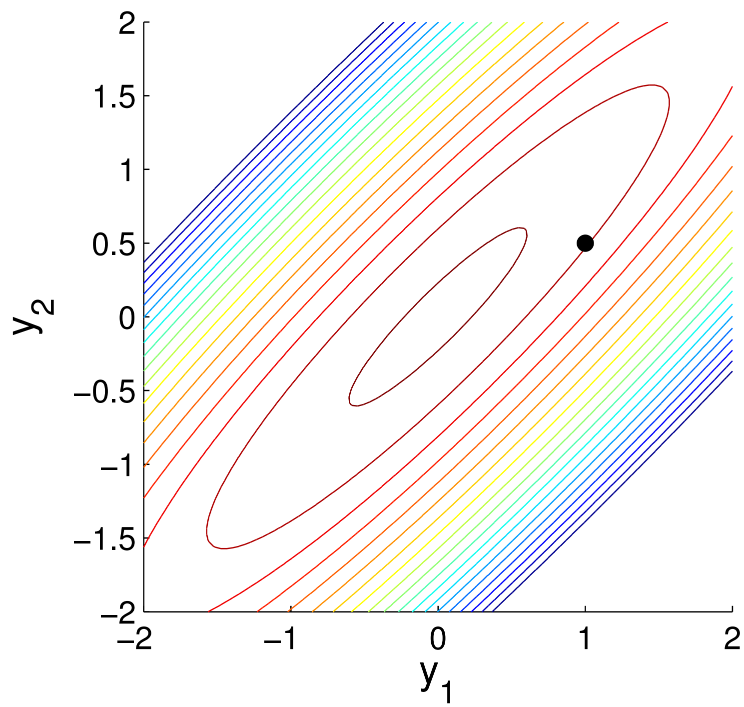
How do we choose the hyper-parameters?

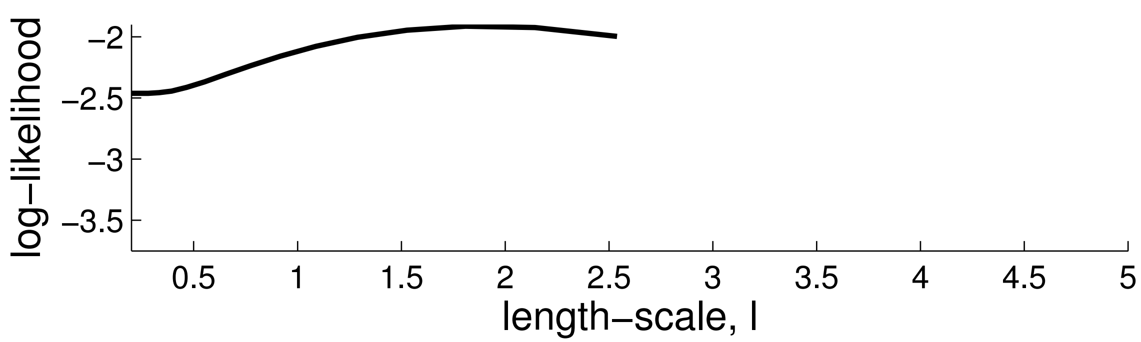


How do we choose the hyper-parameters?

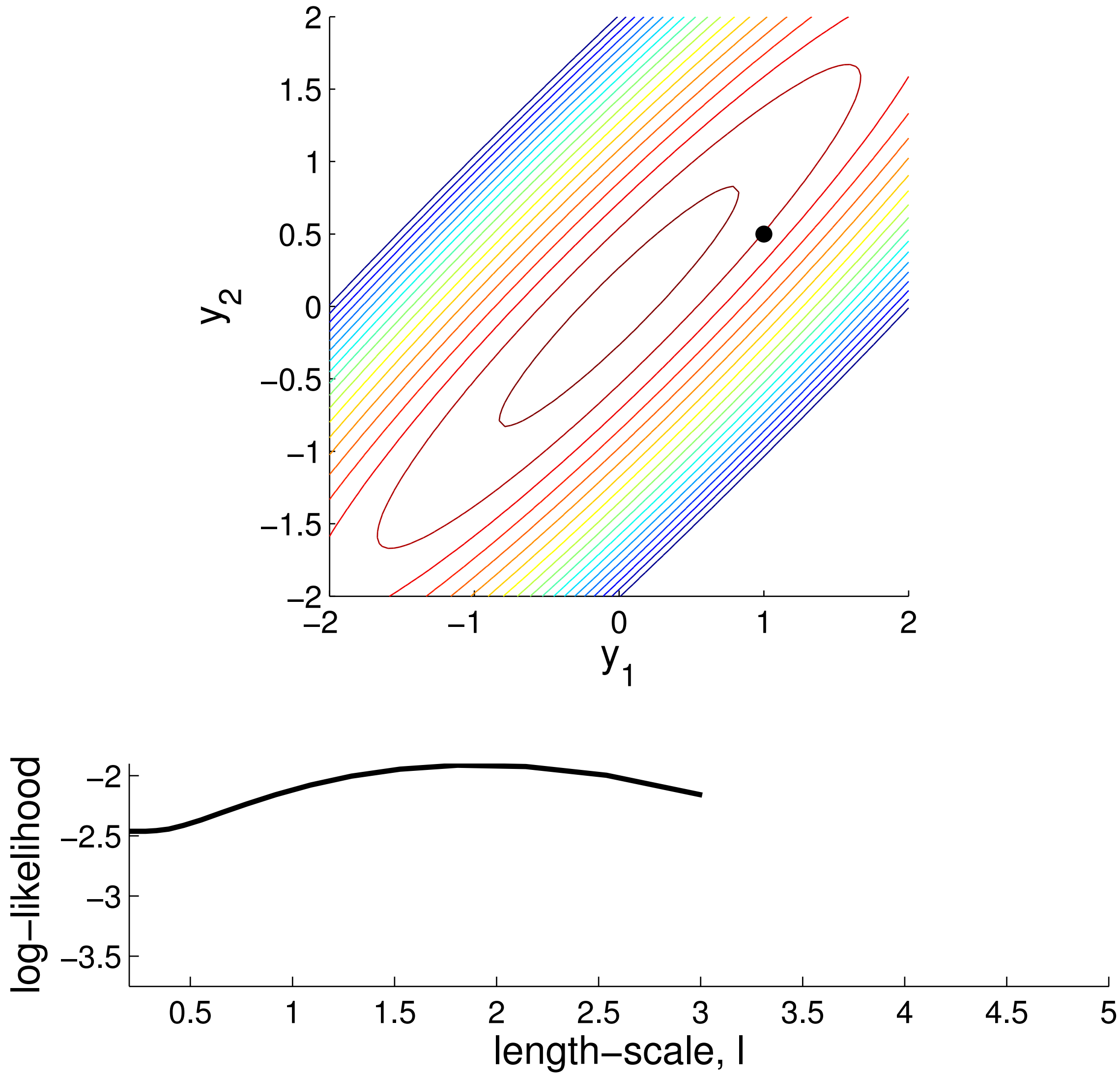


How do we choose the hyper-parameters?

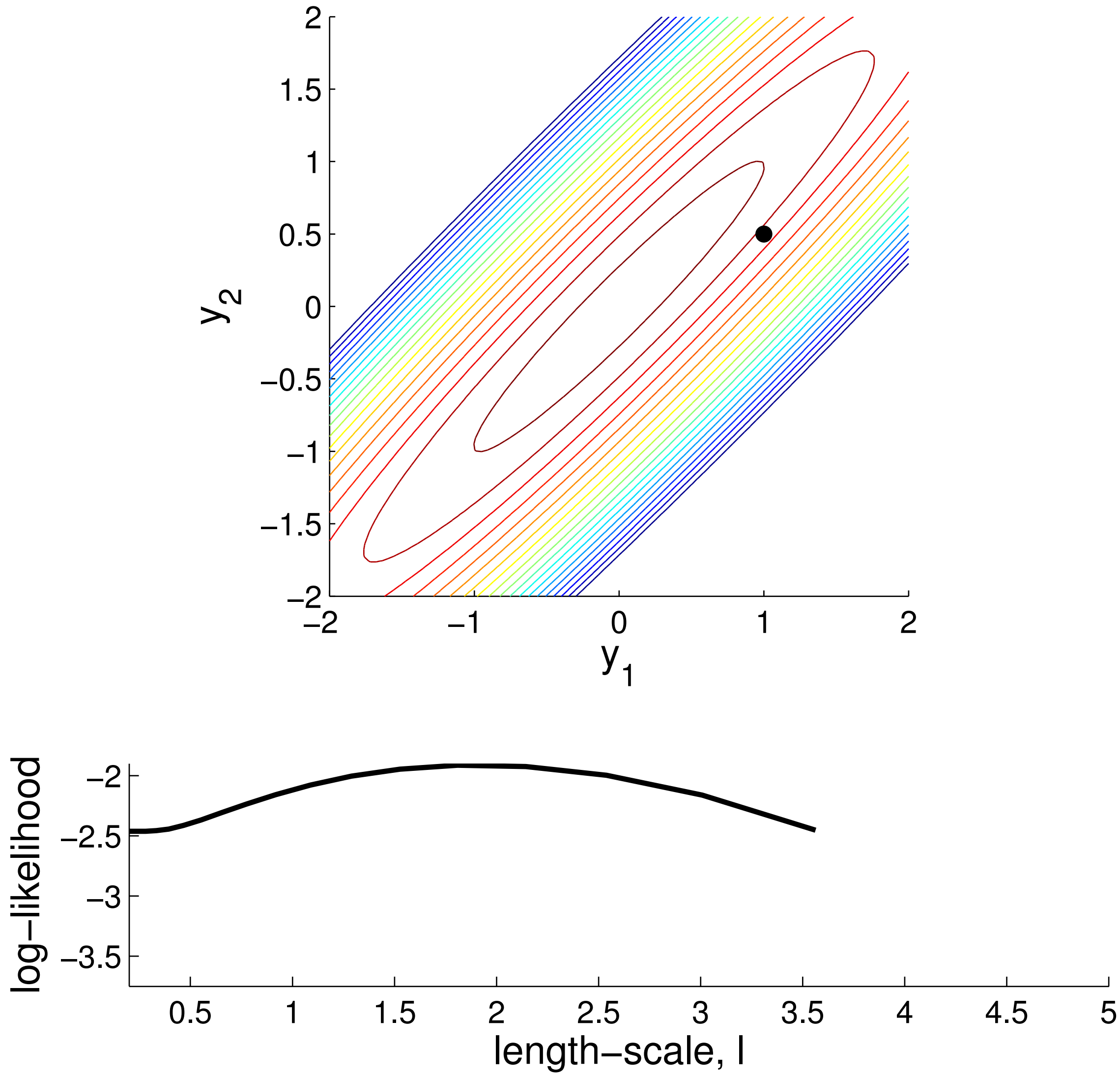




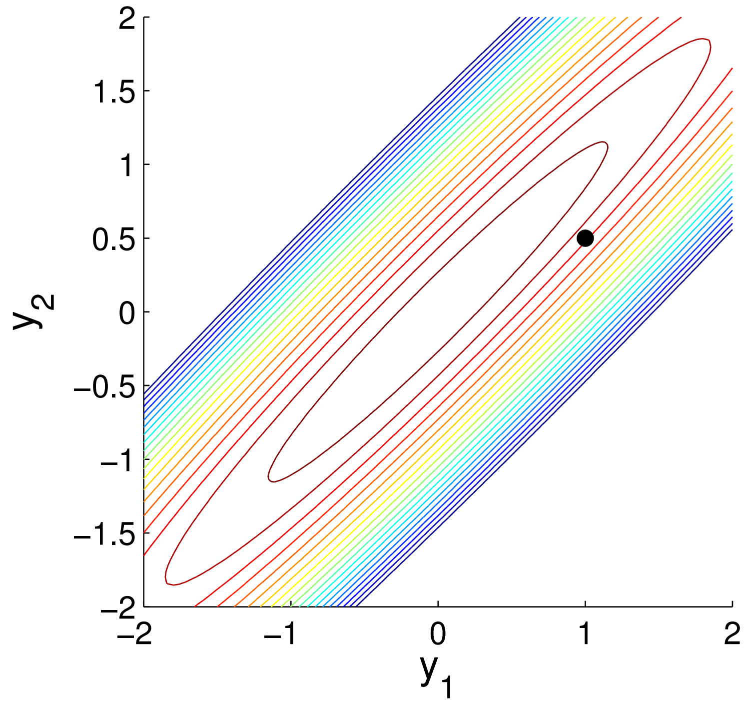
How do we choose the hyper-parameters?

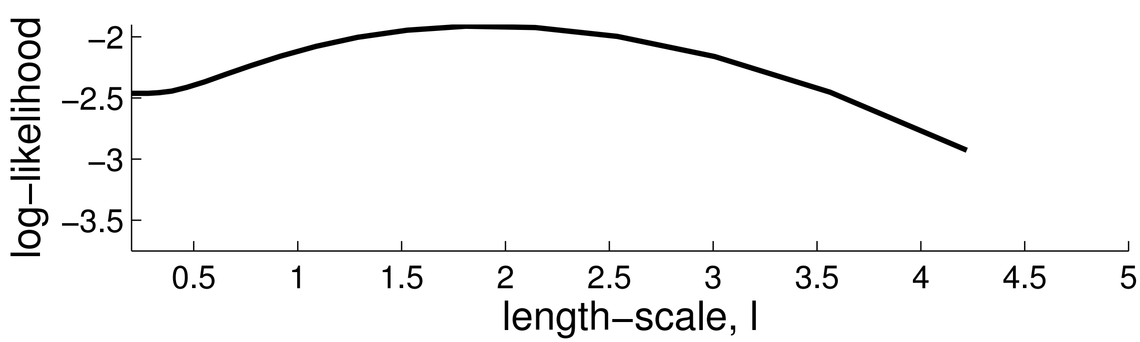


How do we choose the hyper-parameters?

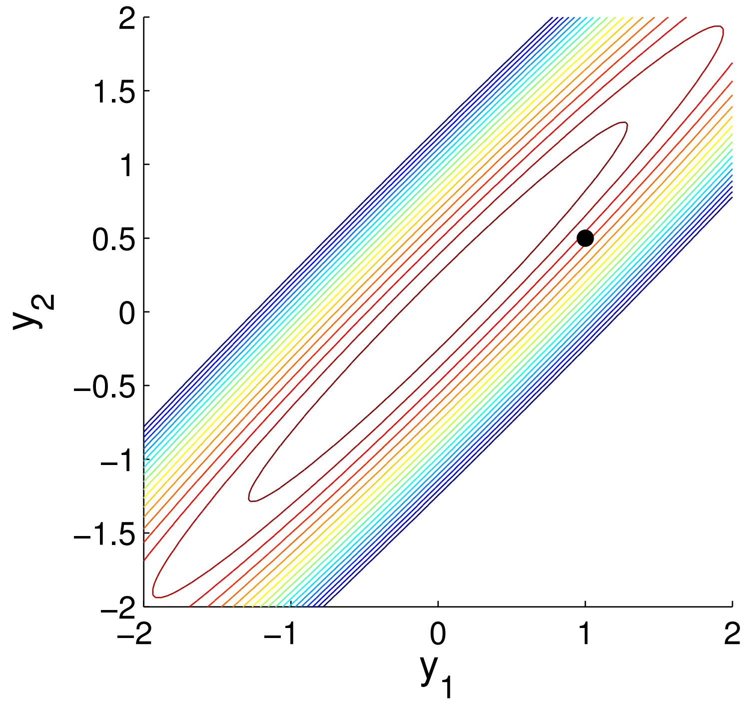


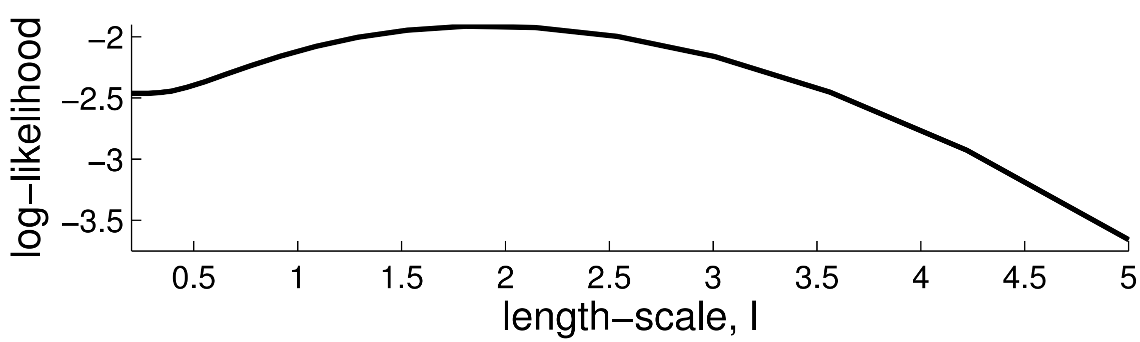
How do we choose the hyper-parameters?



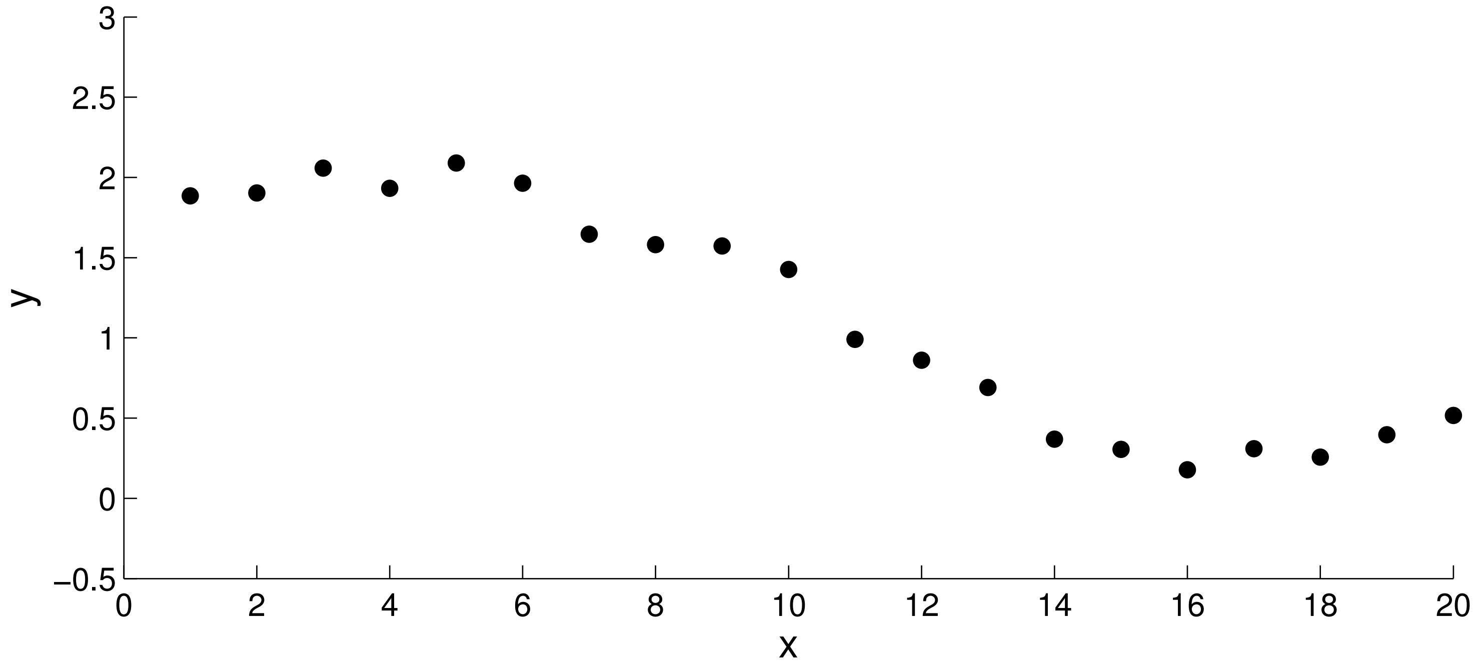


How do we choose the hyper-parameters?

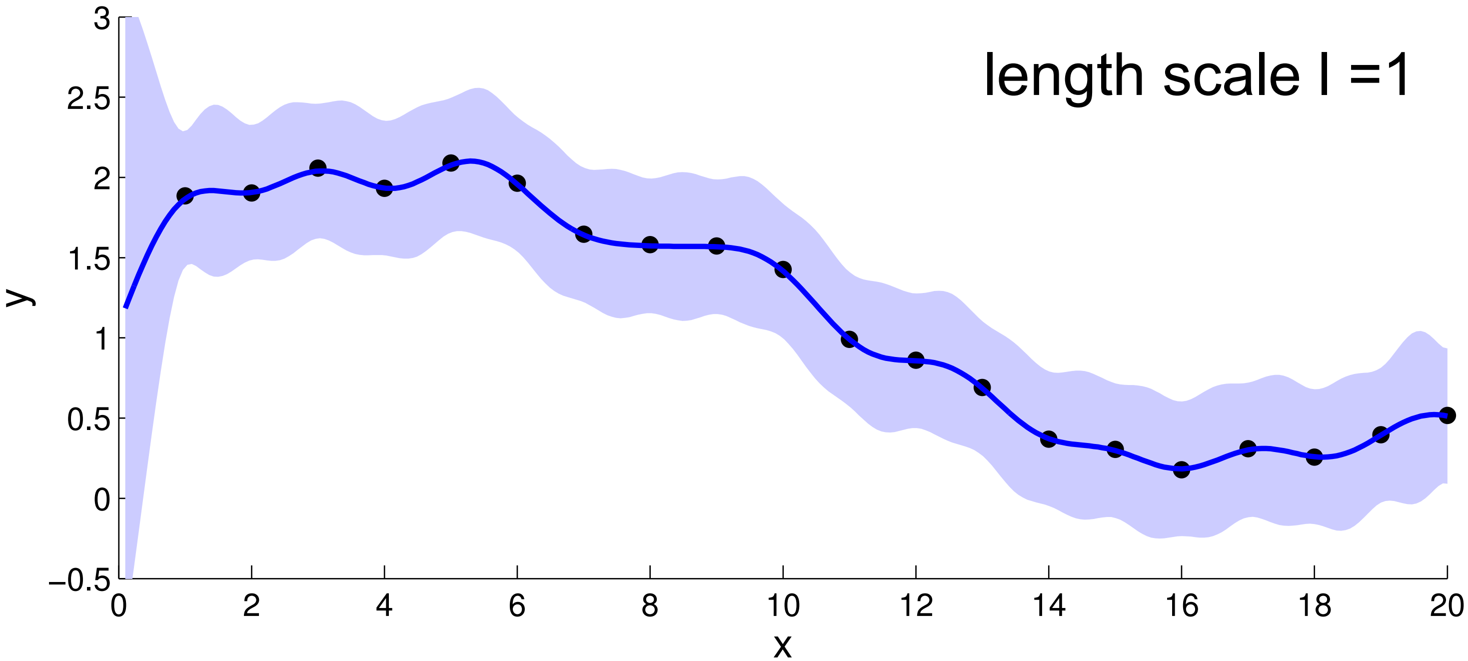




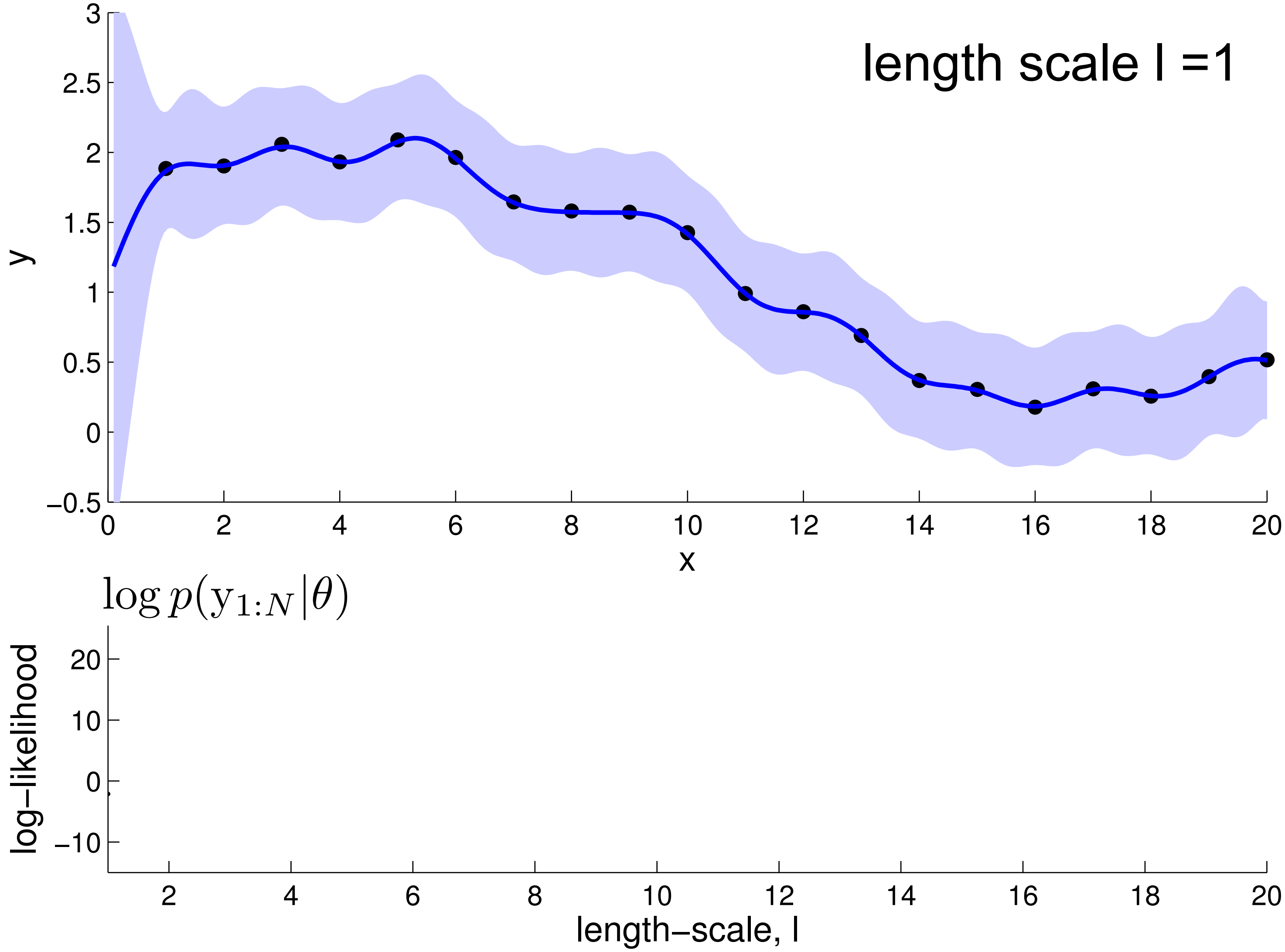
How do we choose the hyper-parameters?



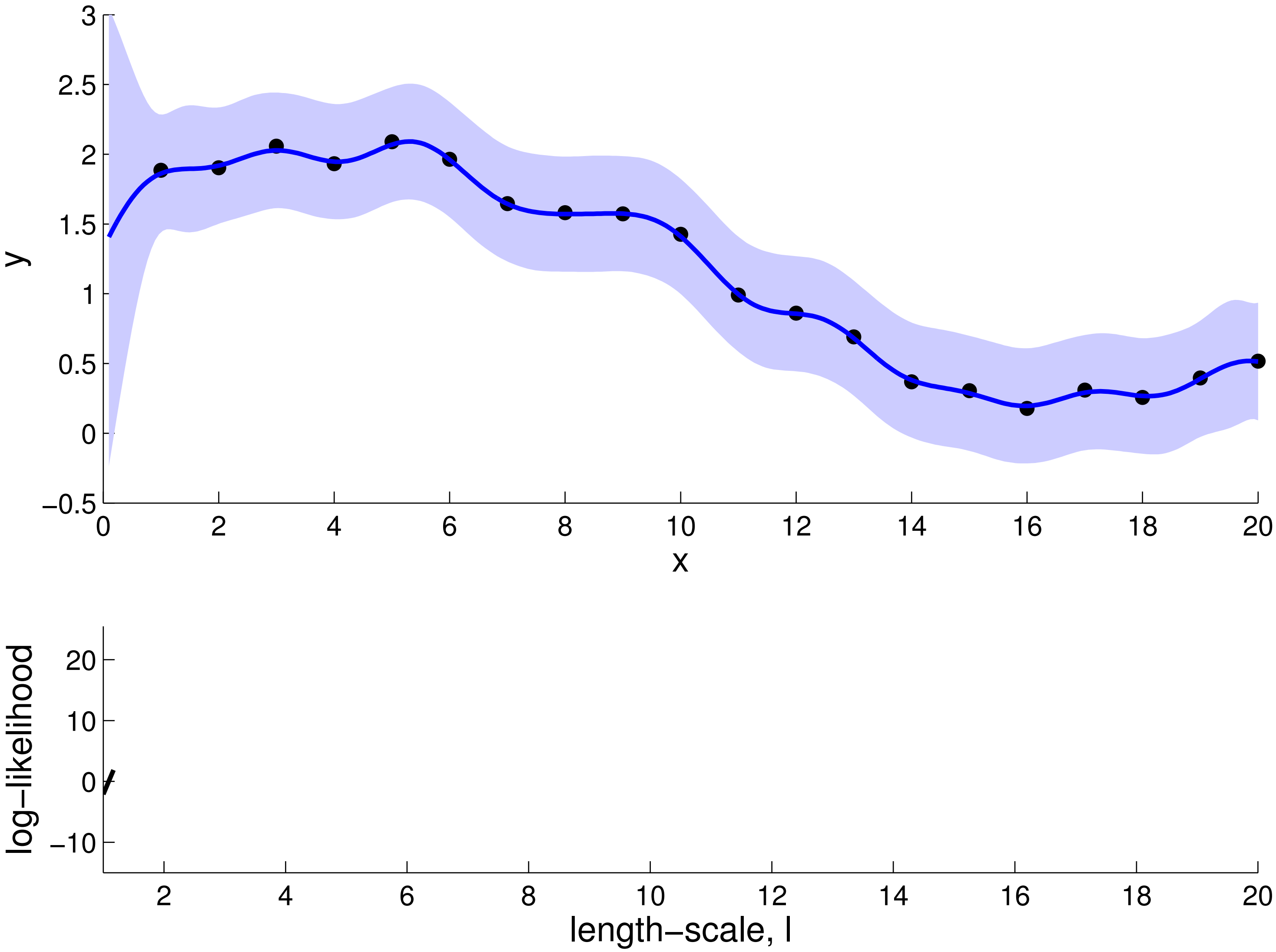
How do we choose the hyper-parameters?



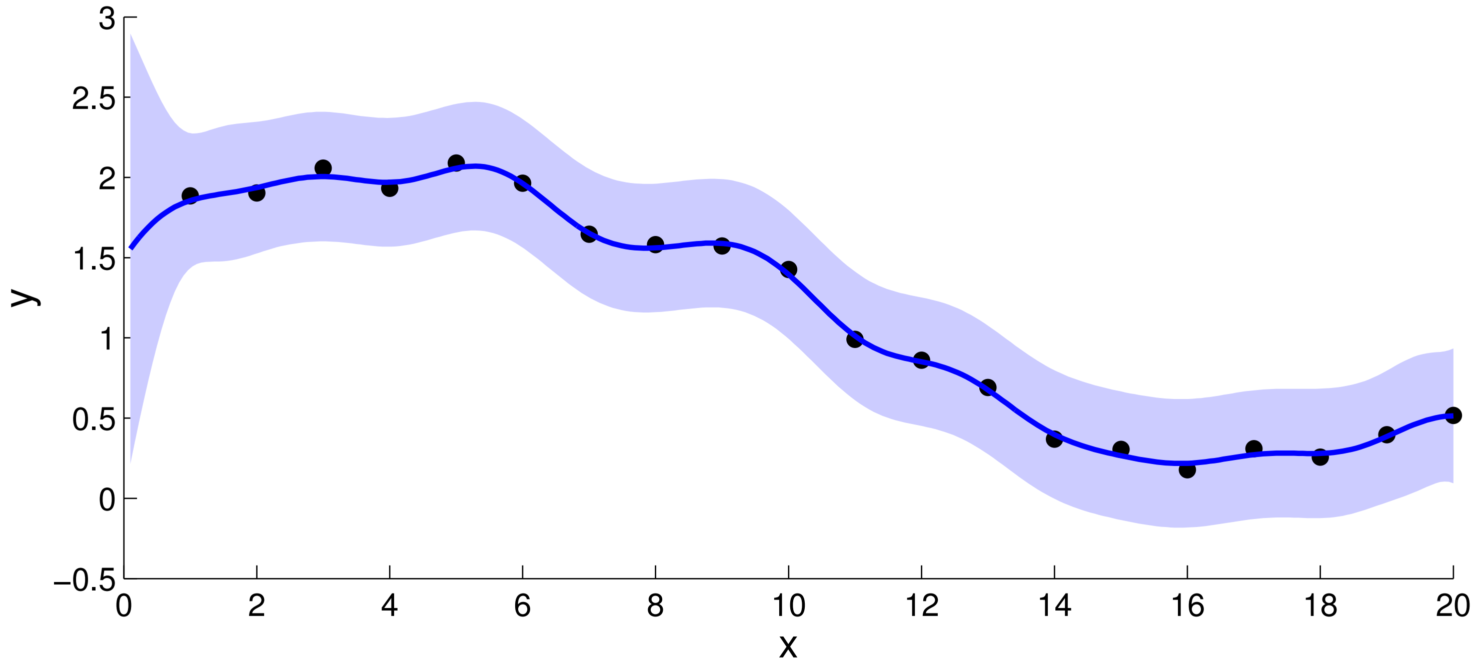
How do we choose the hyper-parameters?

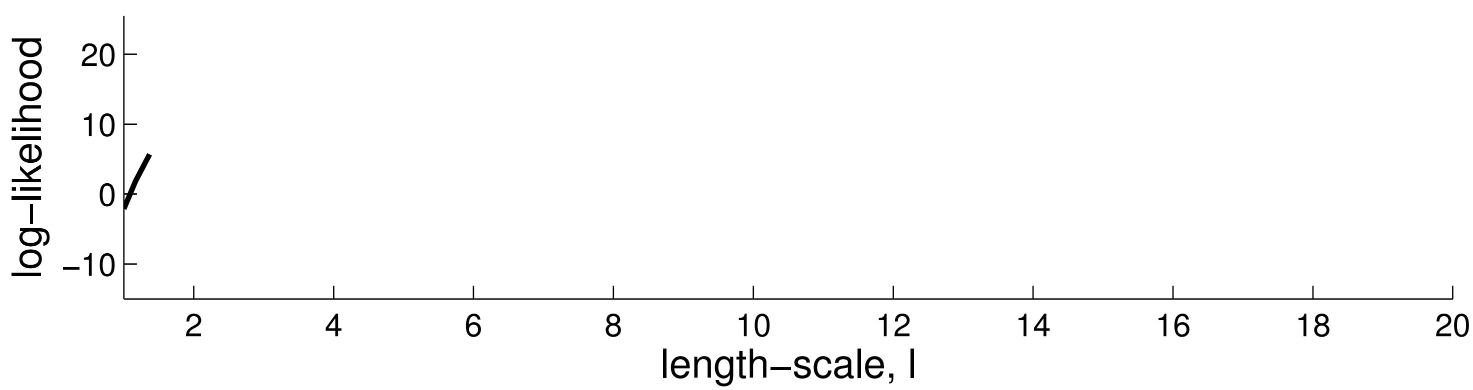


How do we choose the hyper-parameters?

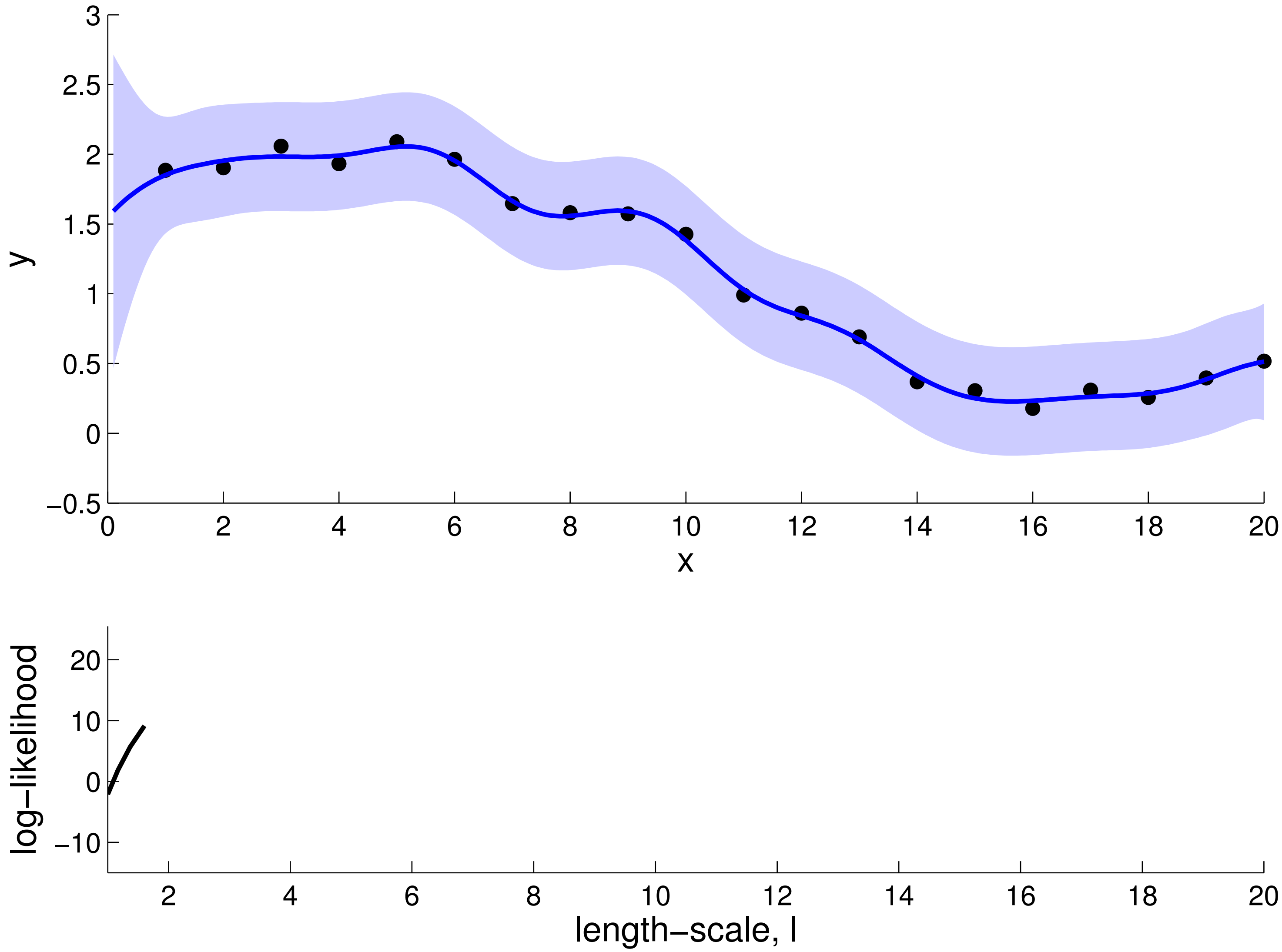


How do we choose the hyper-parameters?

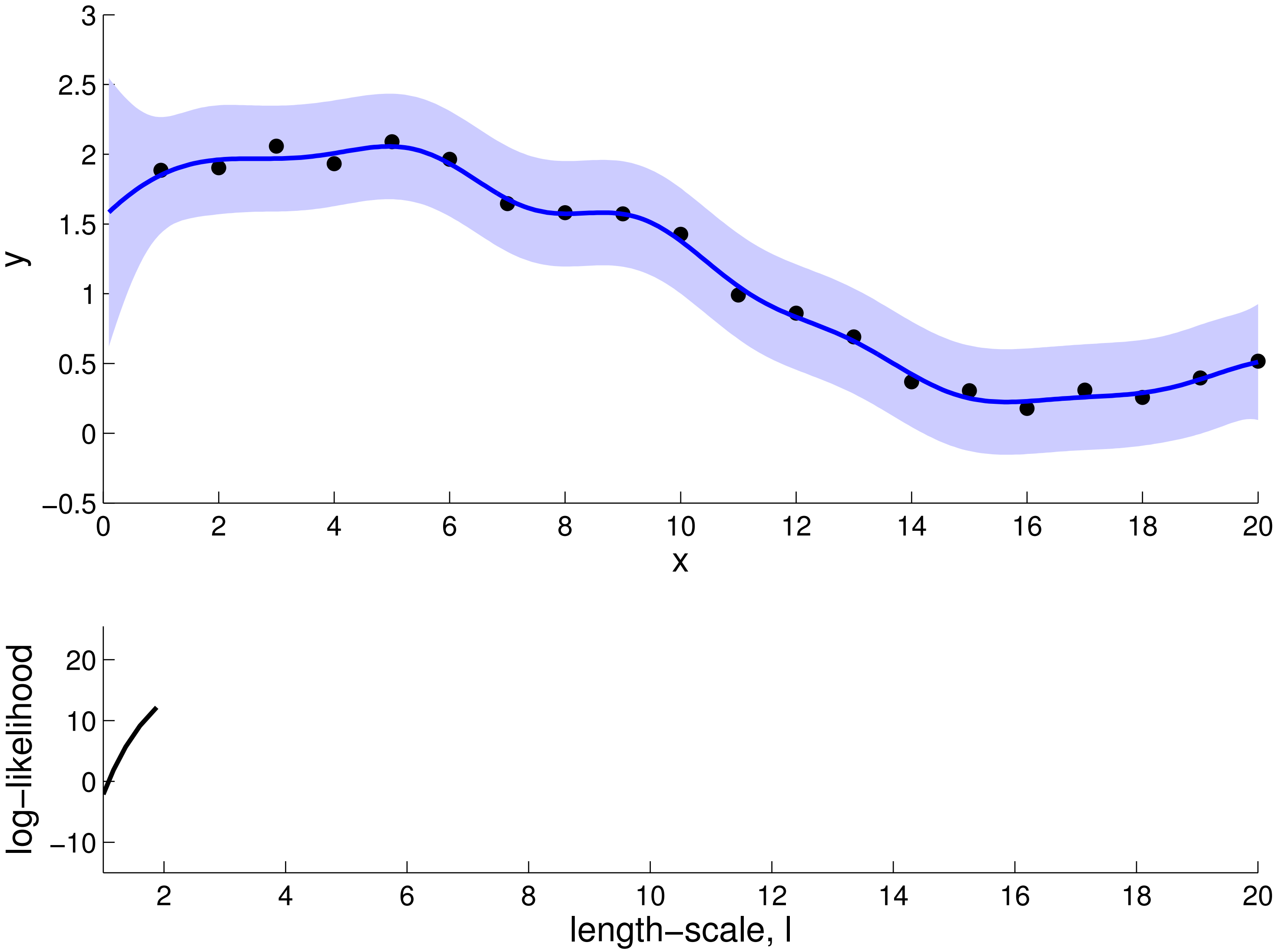




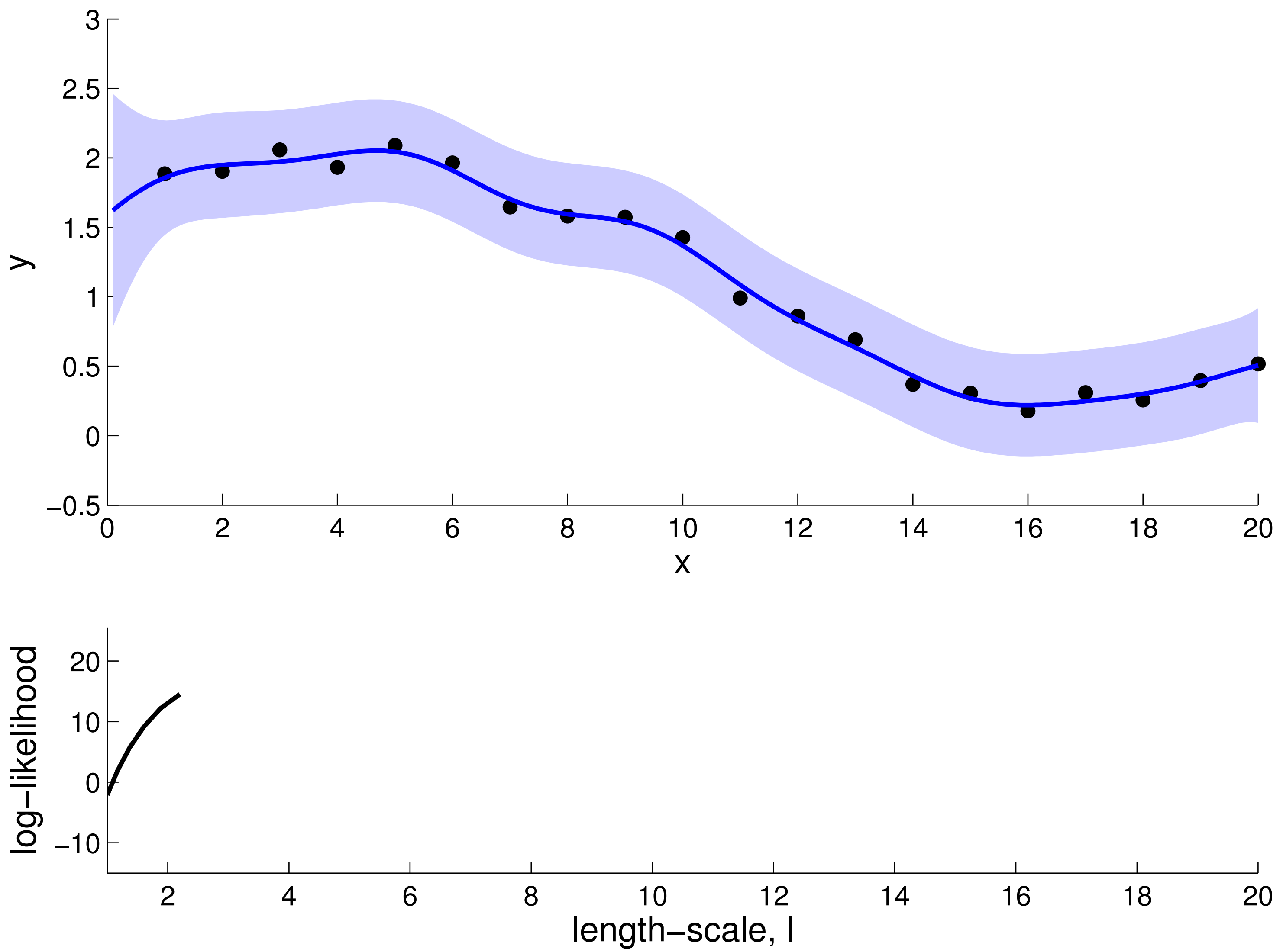
How do we choose the hyper-parameters?



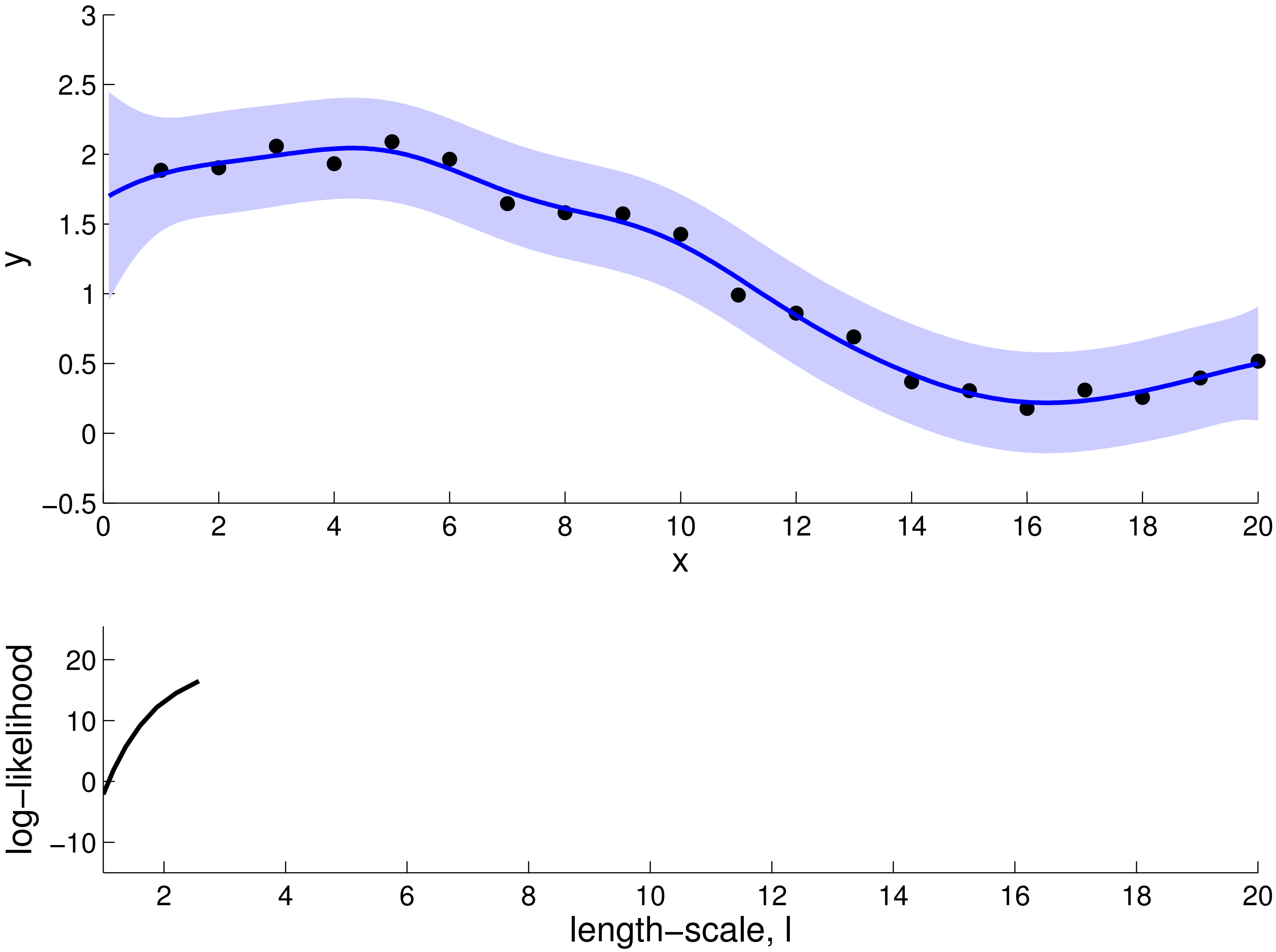
How do we choose the hyper-parameters?



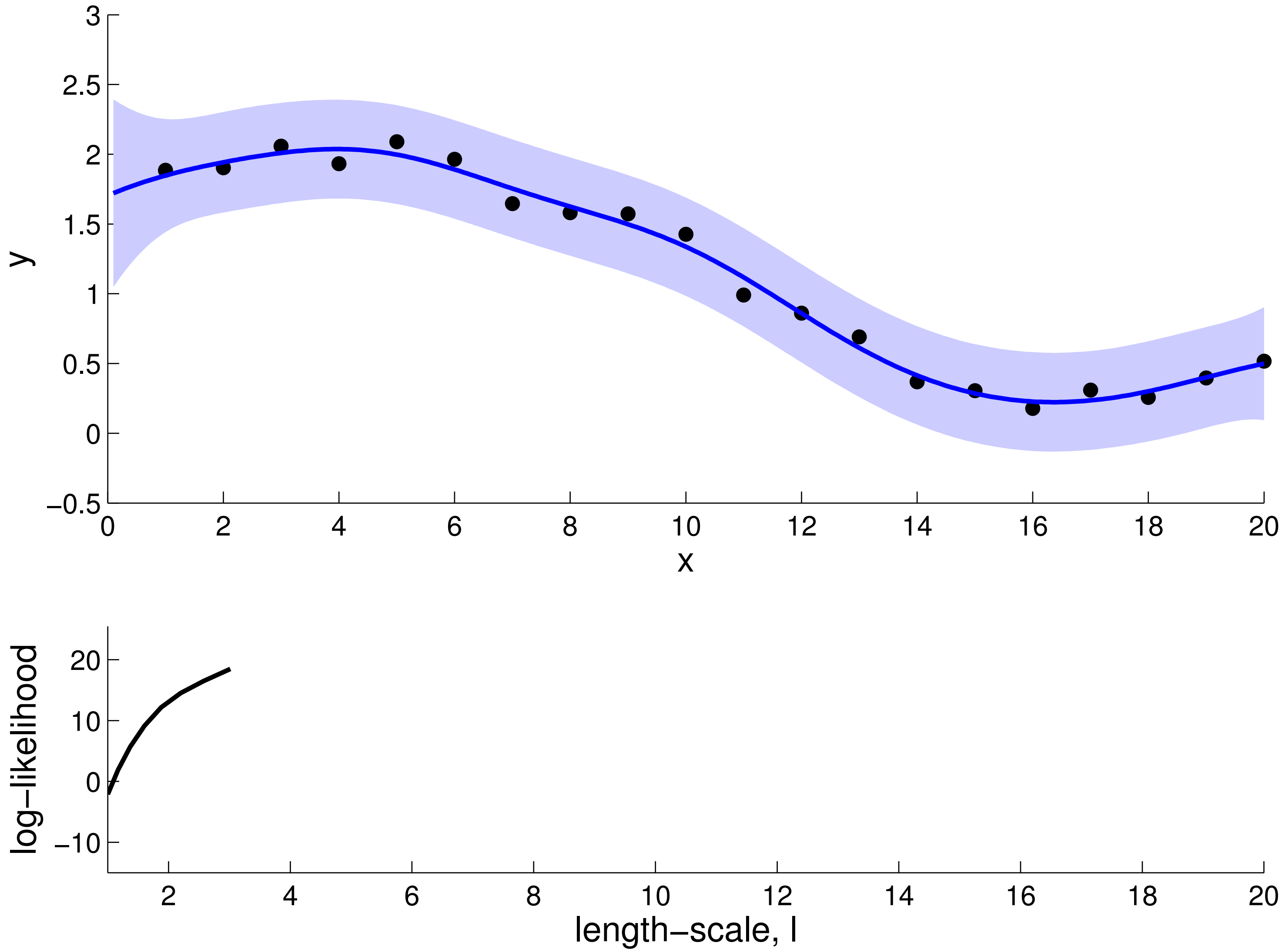
How do we choose the hyper-parameters?



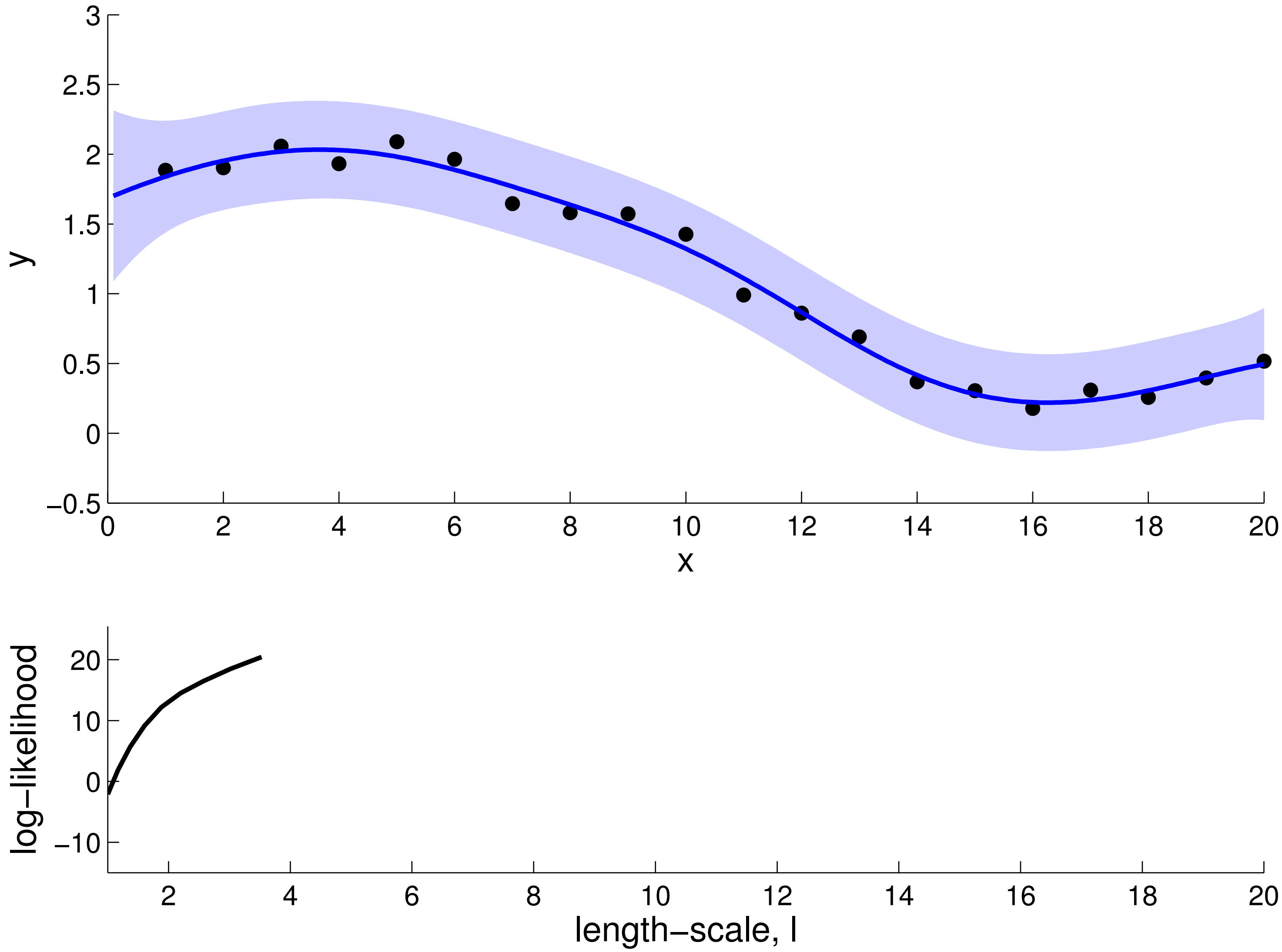
How do we choose the hyper-parameters?



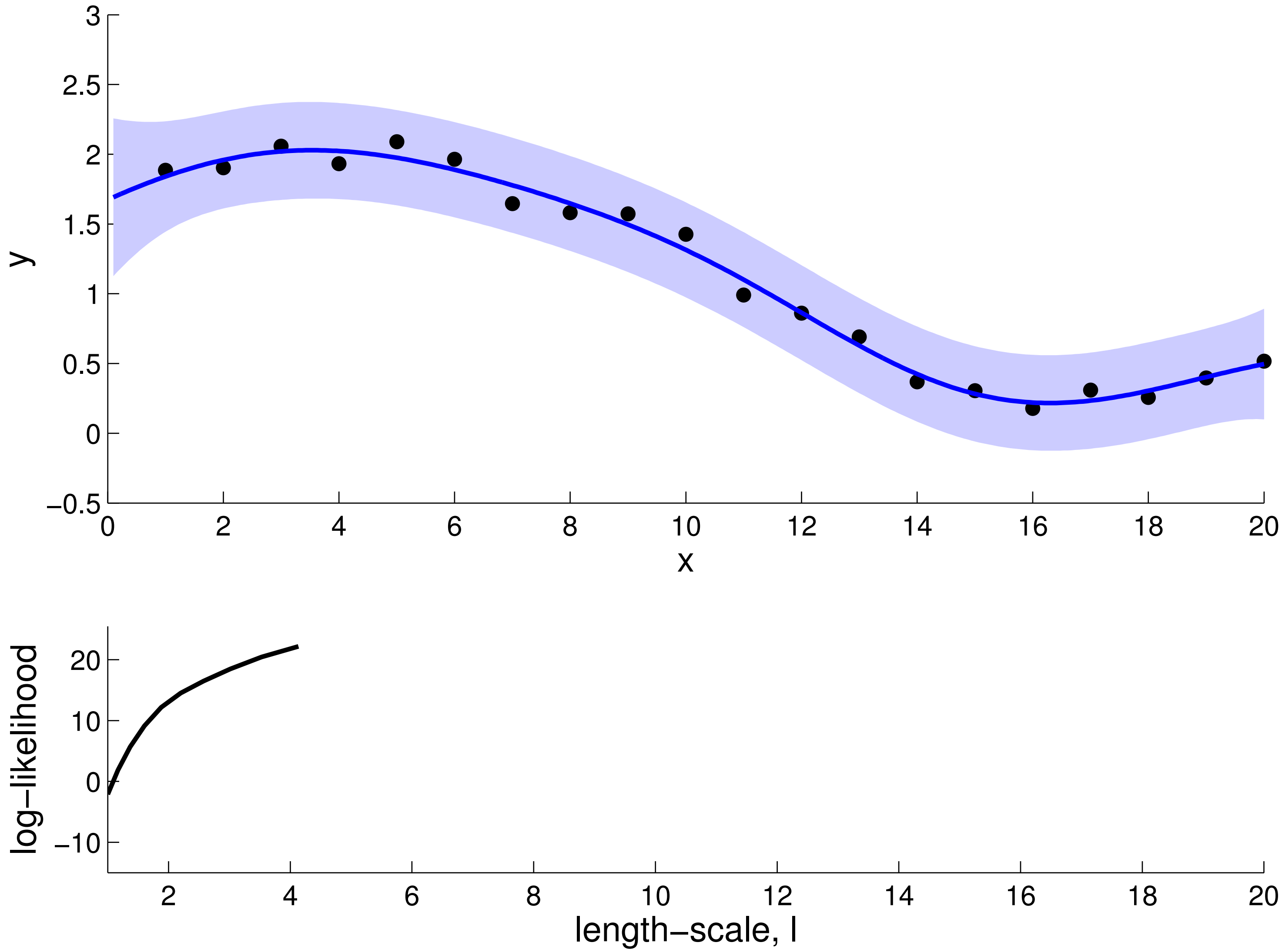
How do we choose the hyper-parameters?



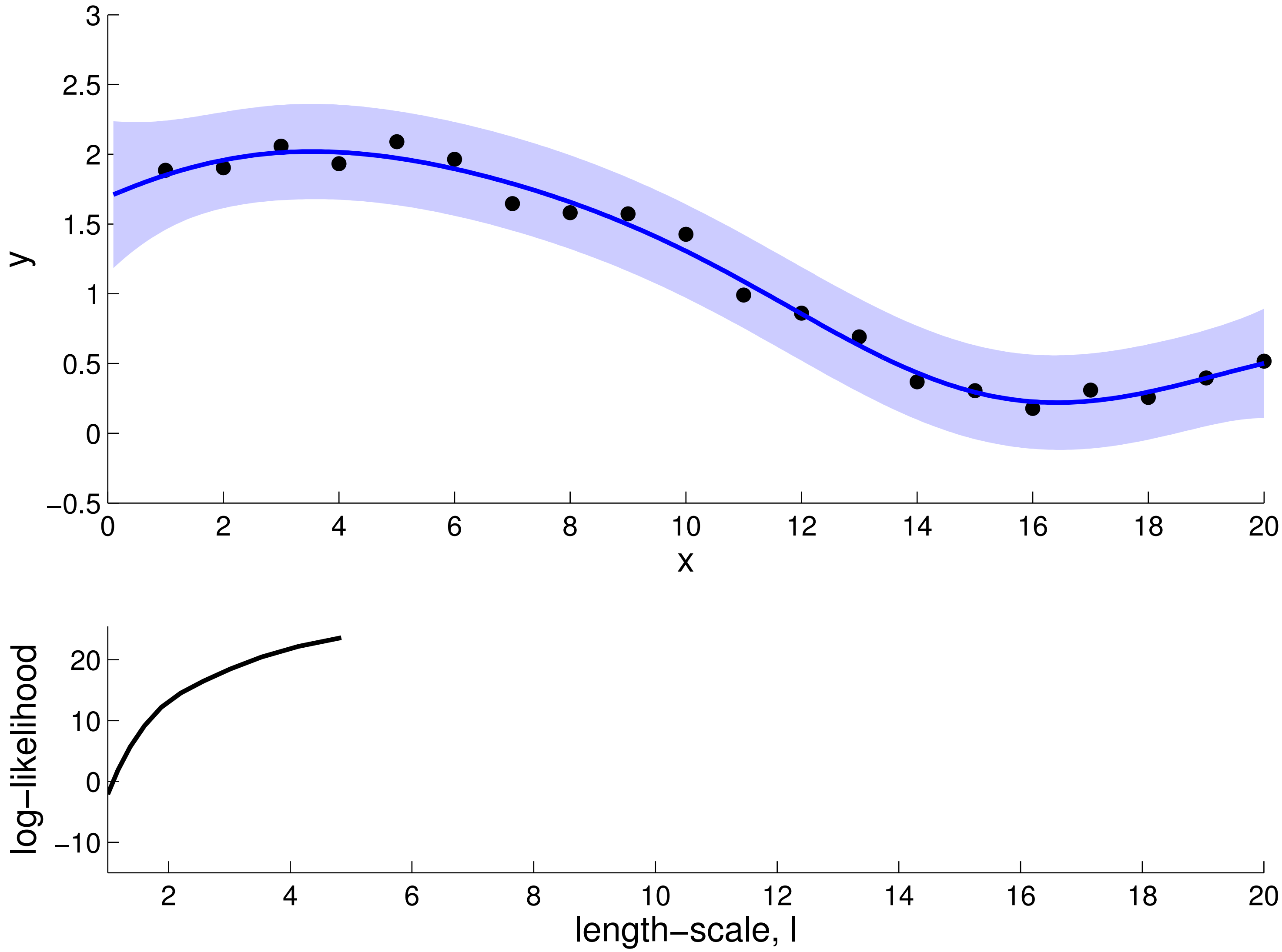
How do we choose the hyper-parameters?



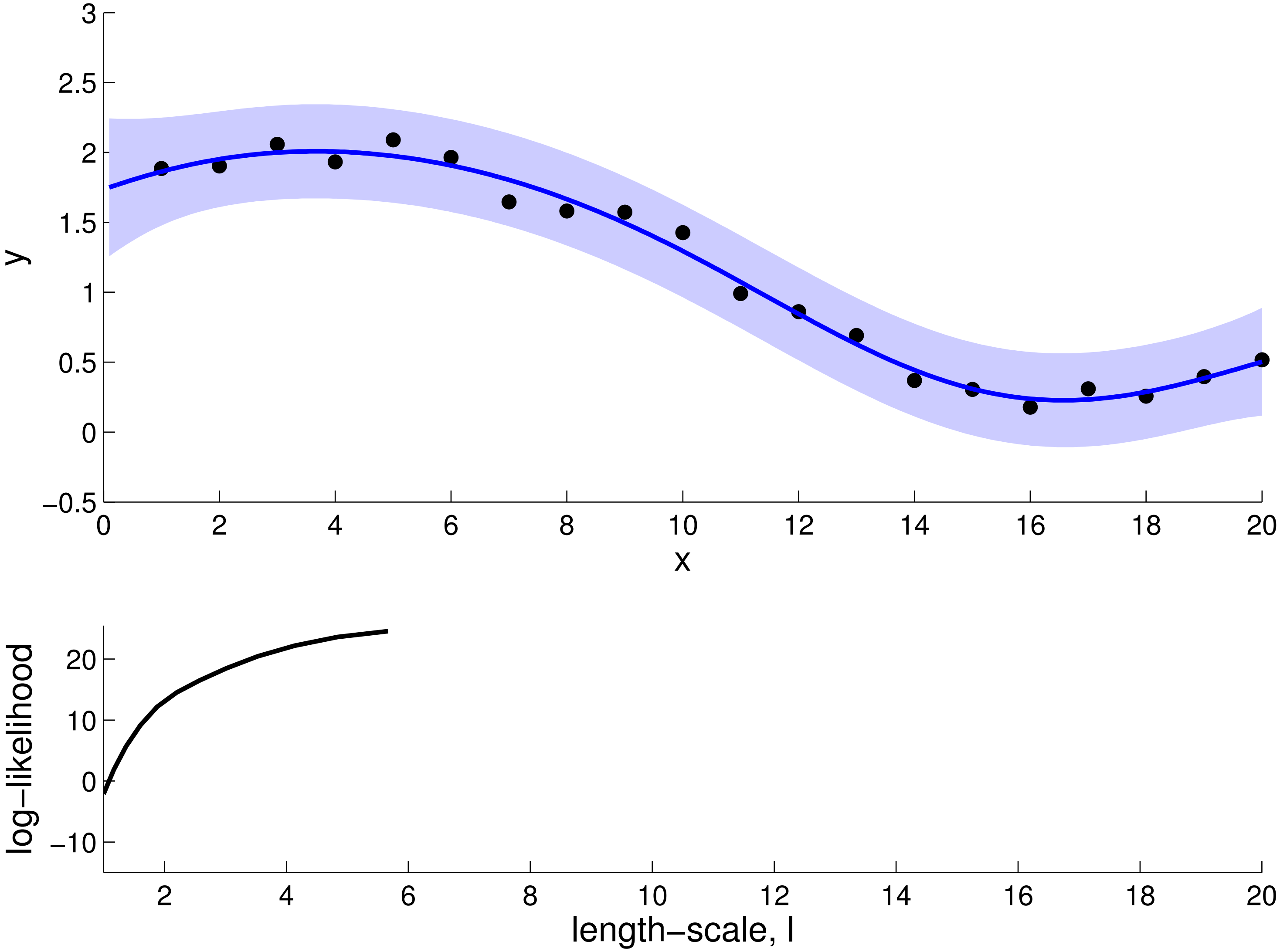
How do we choose the hyper-parameters?



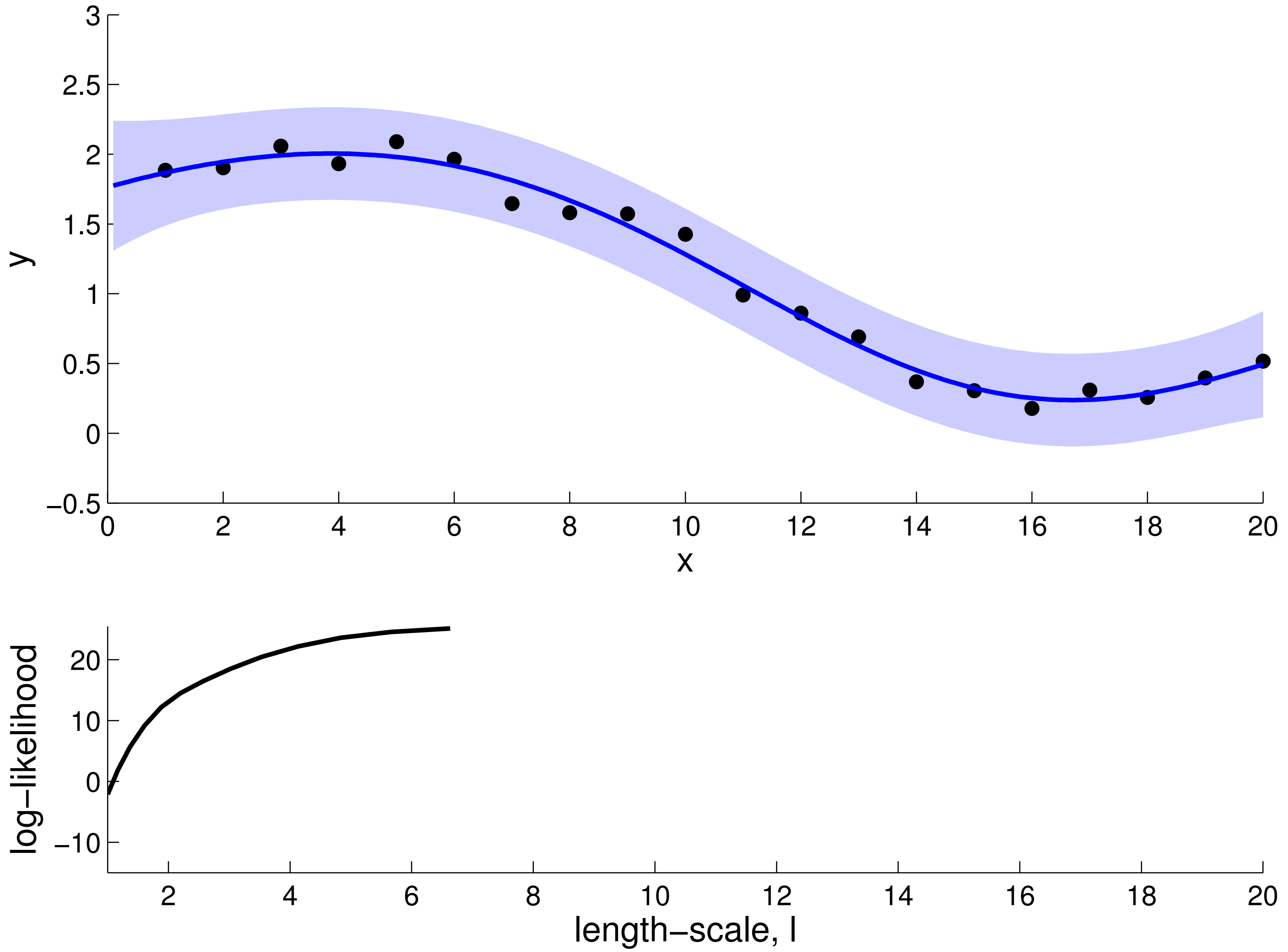
How do we choose the hyper-parameters?



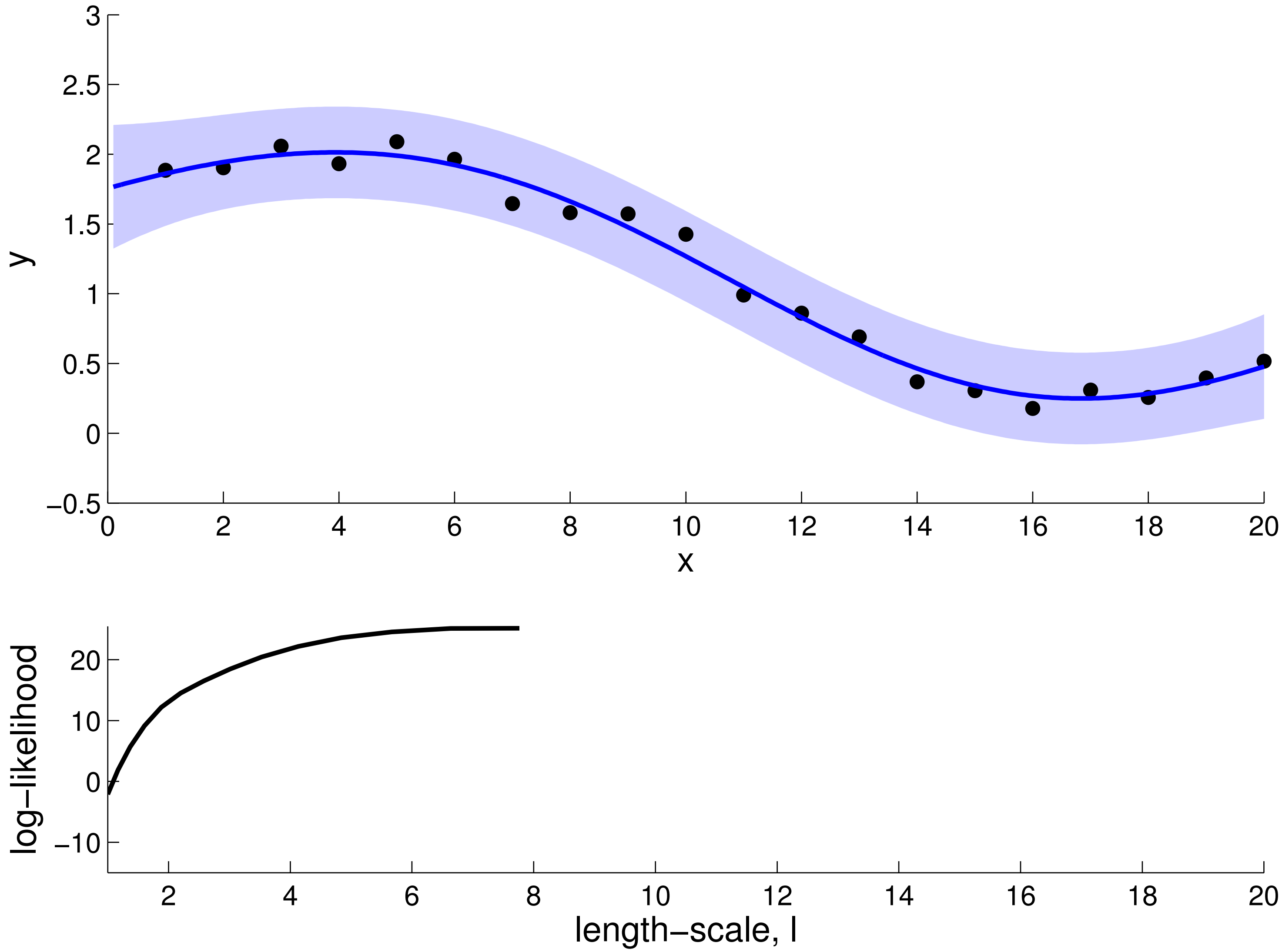
How do we choose the hyper-parameters?



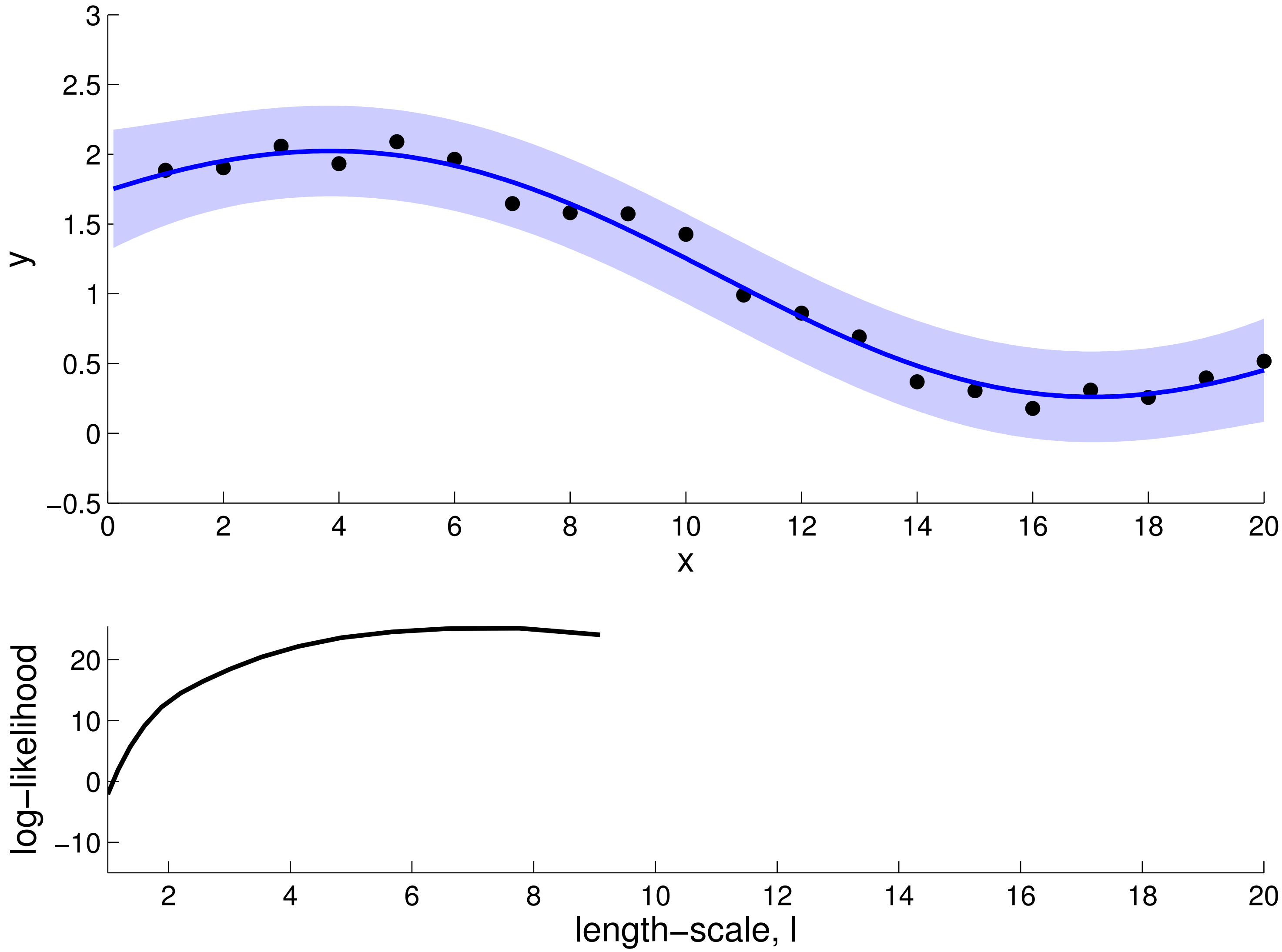
How do we choose the hyper-parameters?



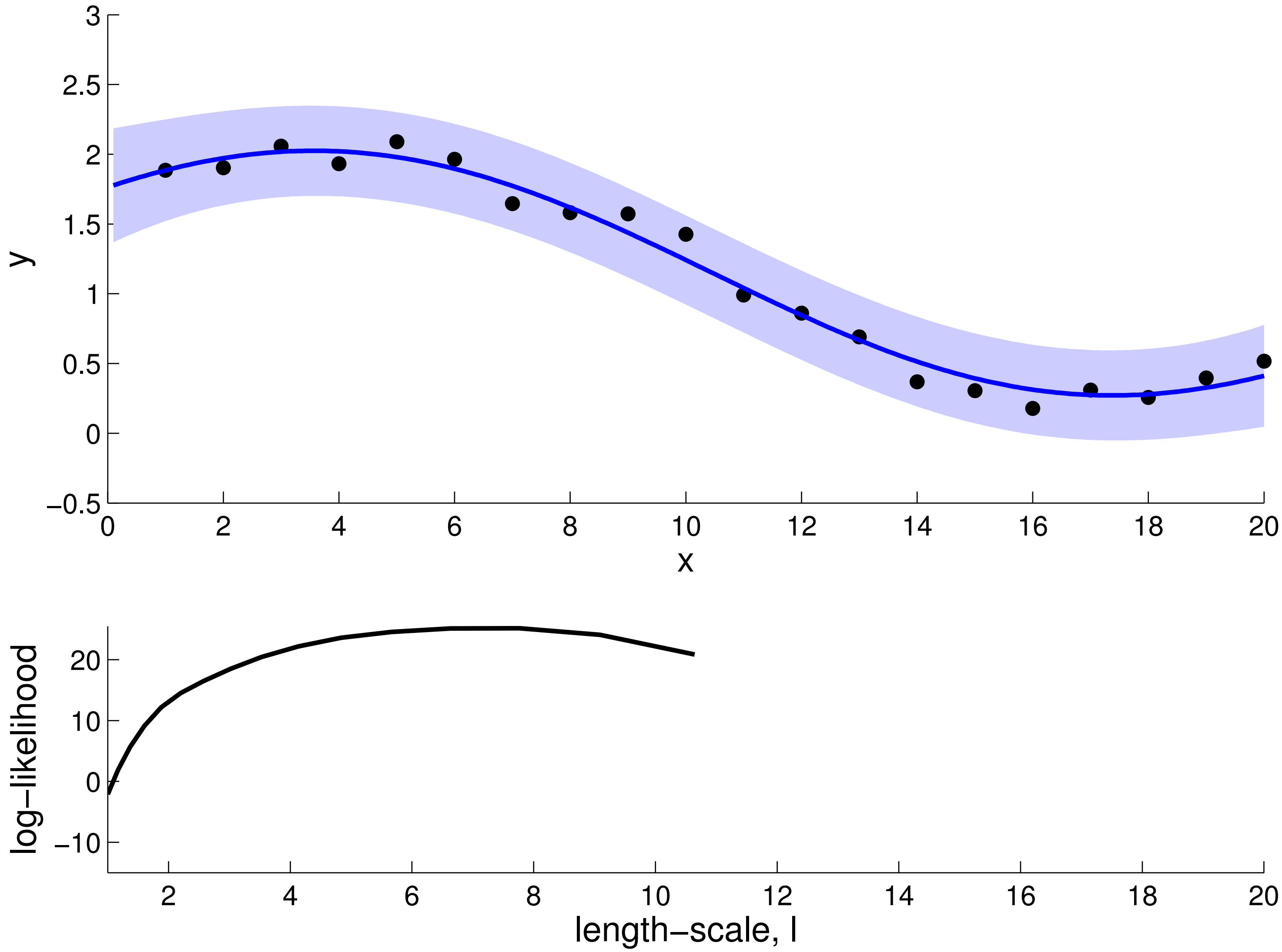
How do we choose the hyper-parameters?



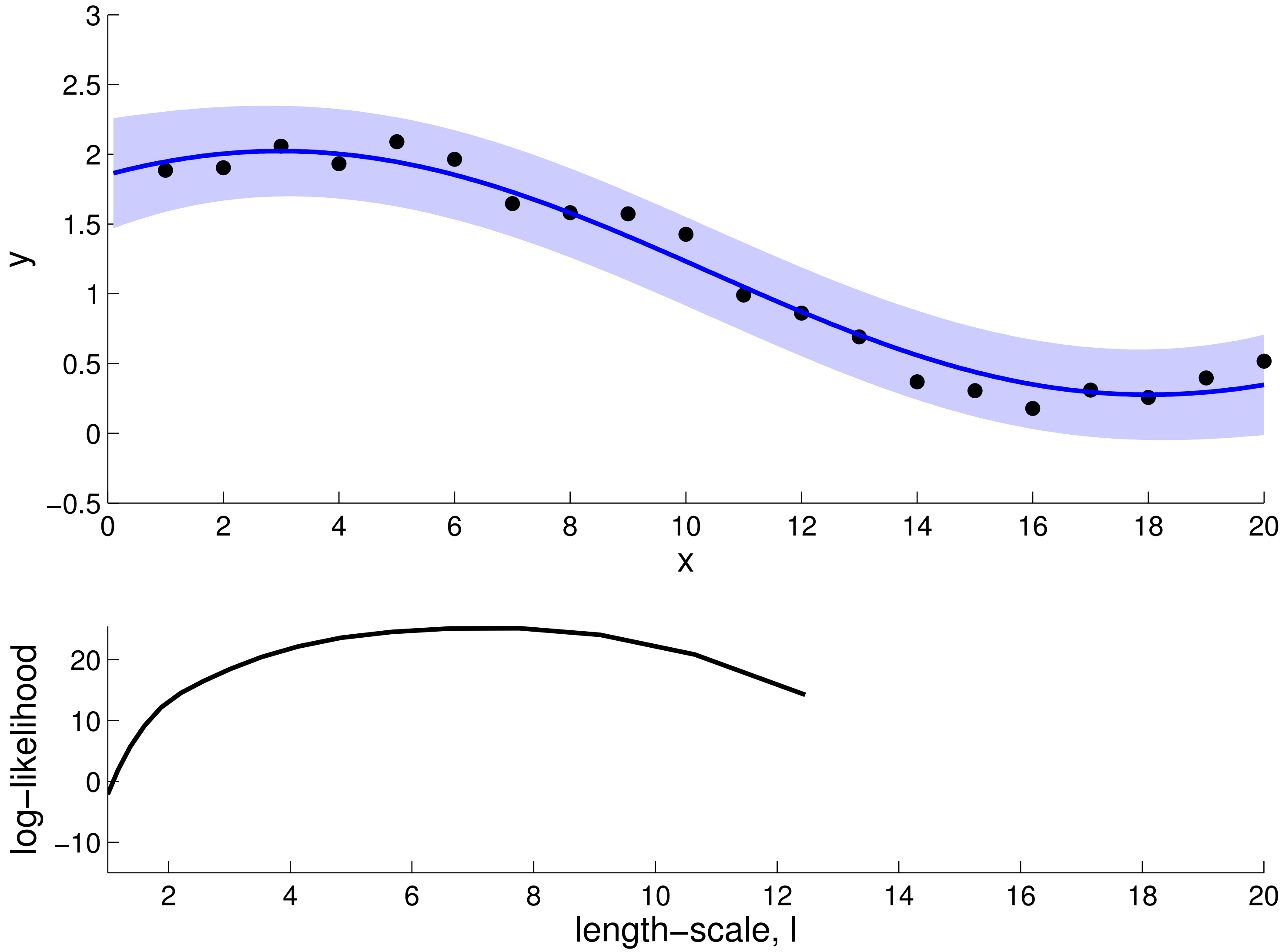
How do we choose the hyper-parameters?



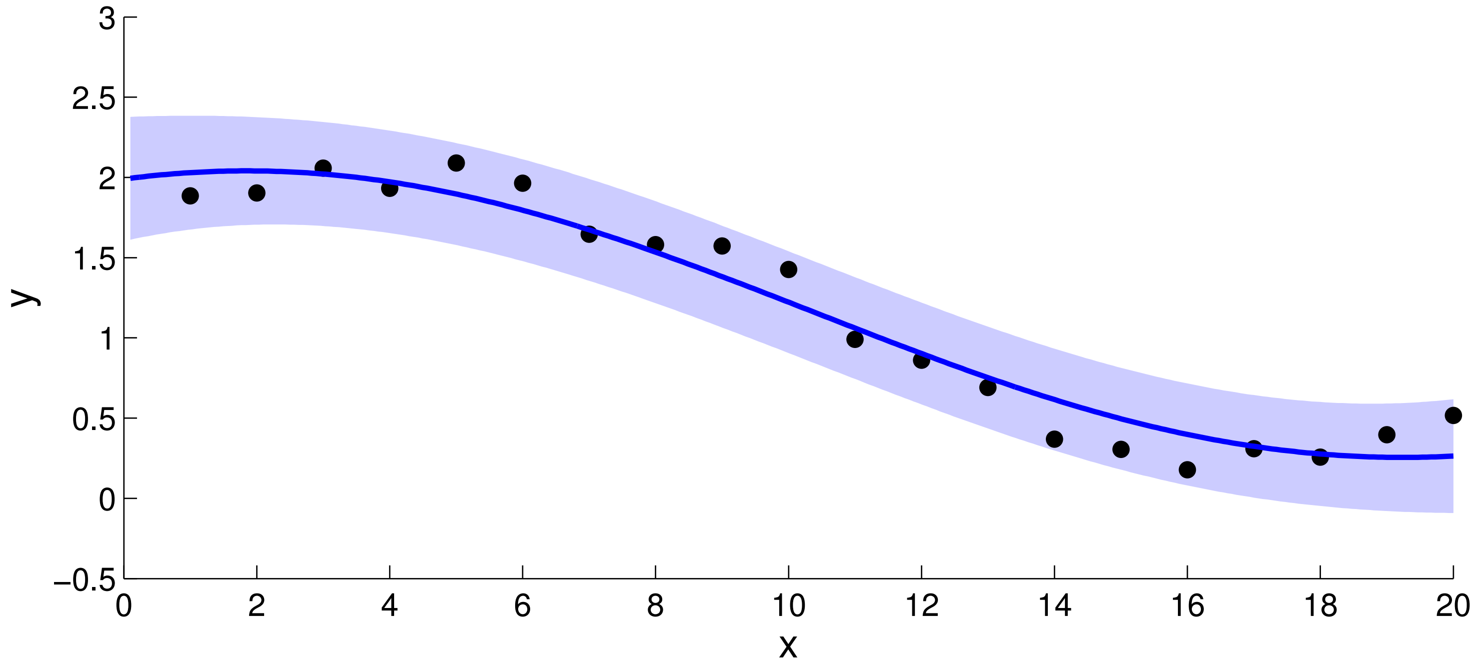
How do we choose the hyper-parameters?

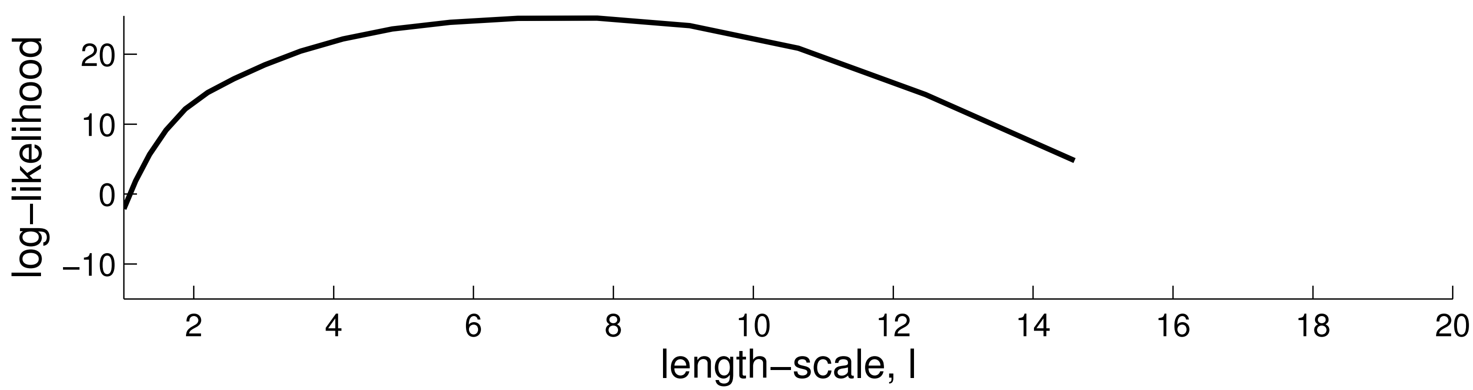


How do we choose the hyper-parameters?

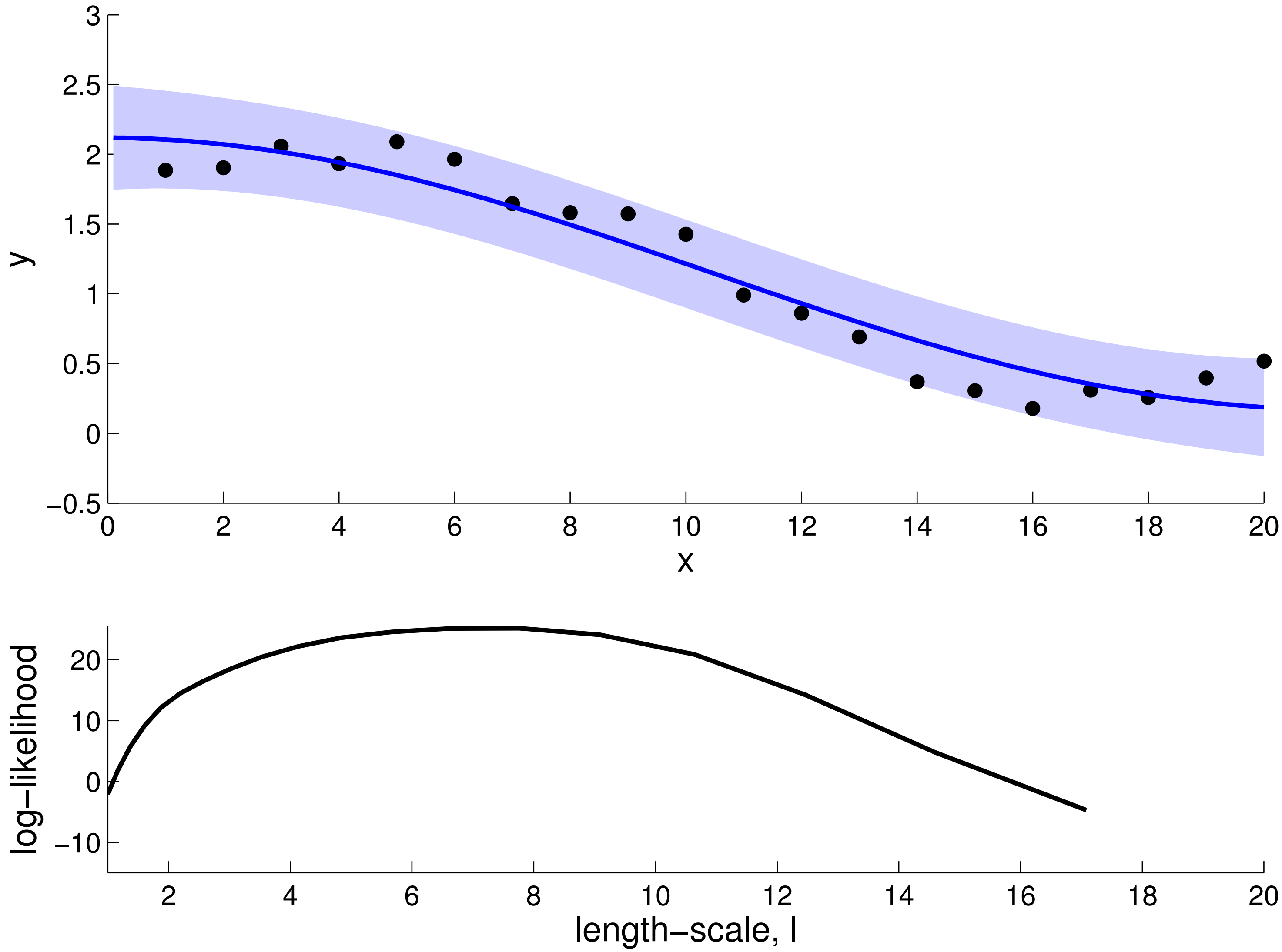


How do we choose the hyper-parameters?

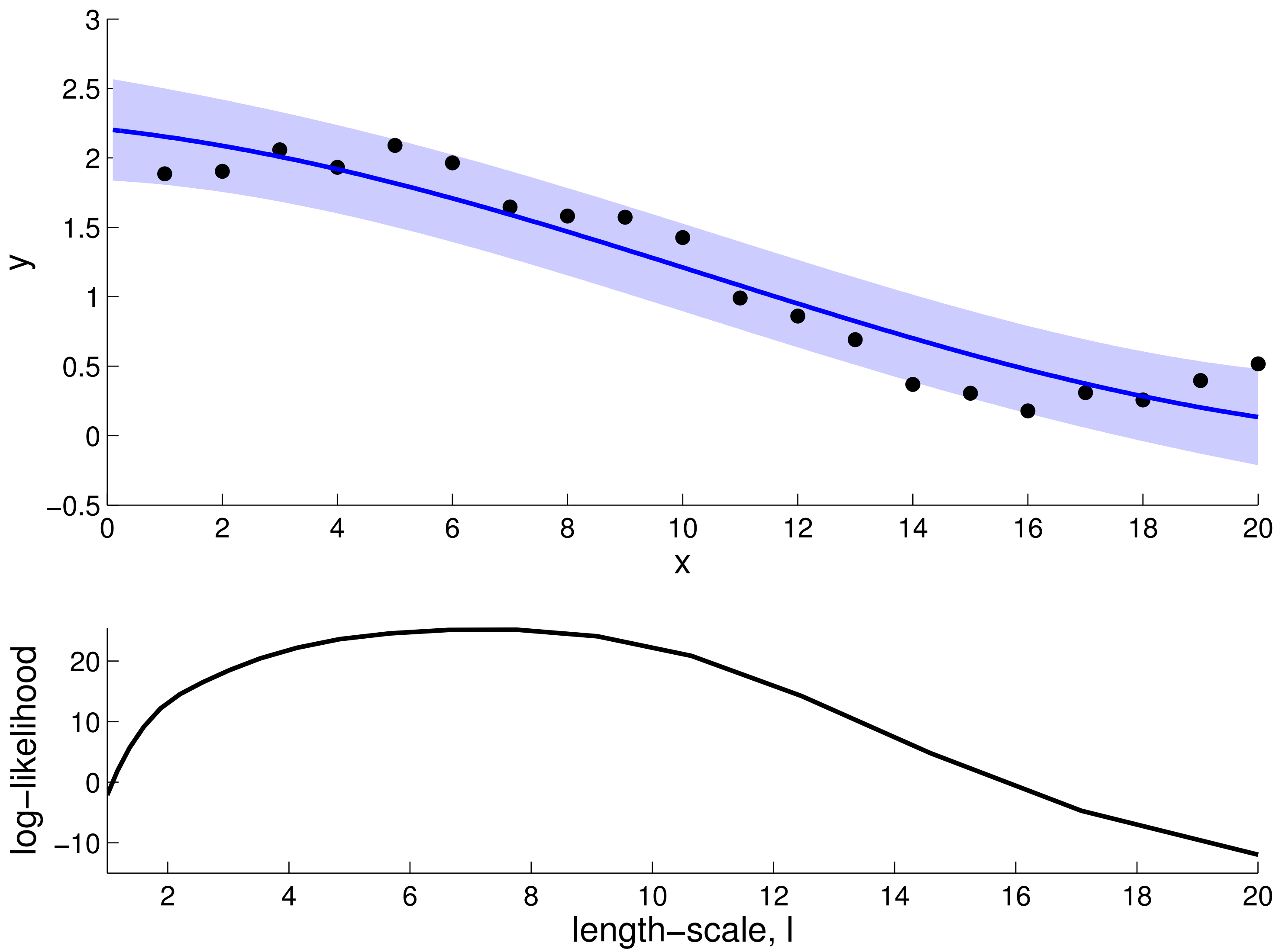




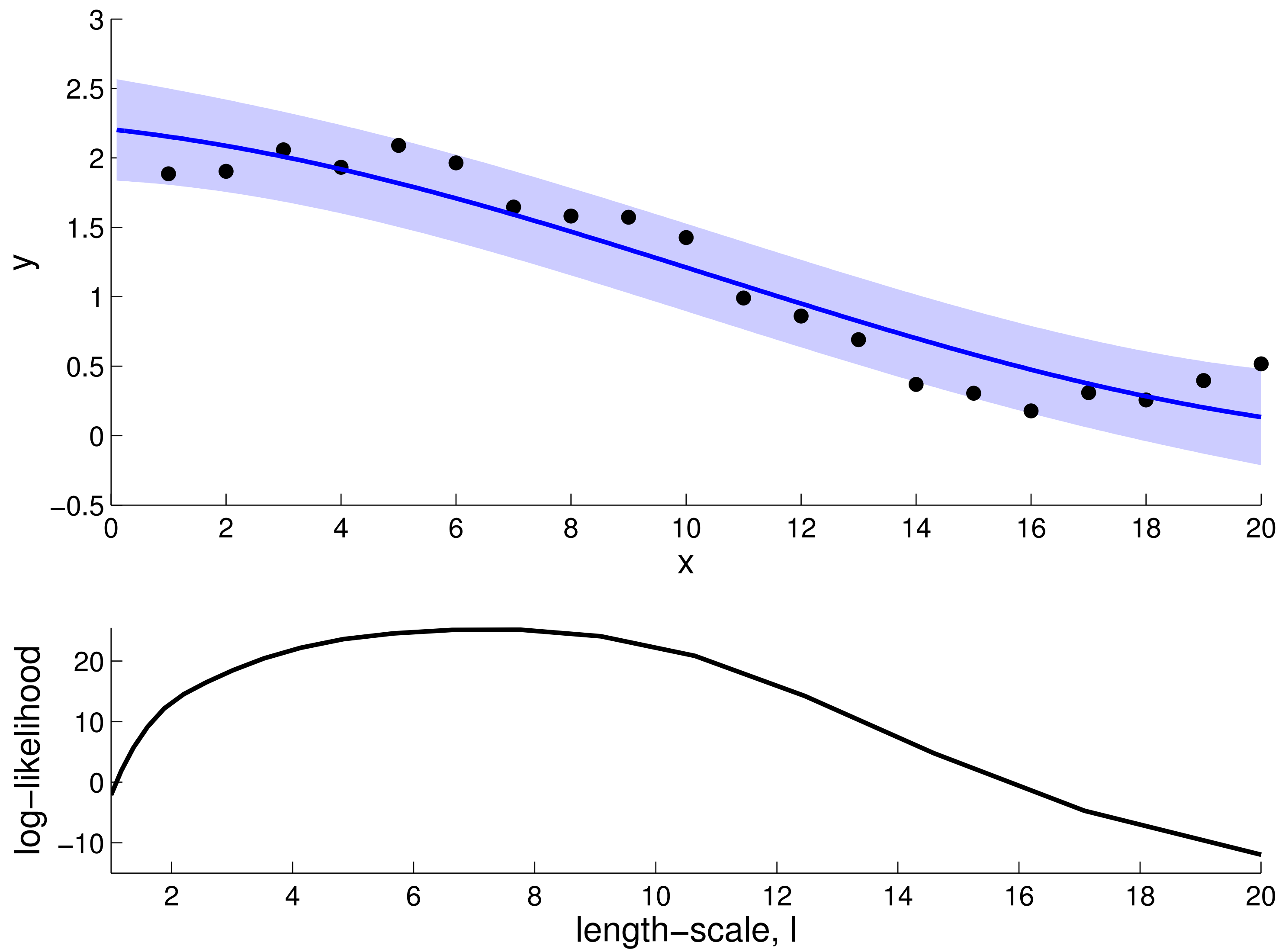
How do we choose the hyper-parameters?



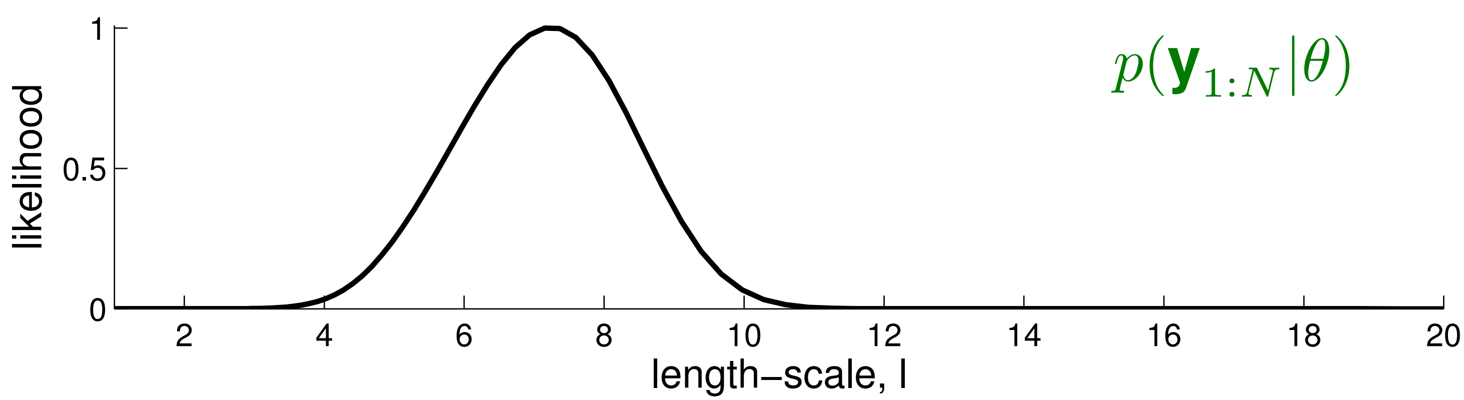
How do we choose the hyper-parameters?



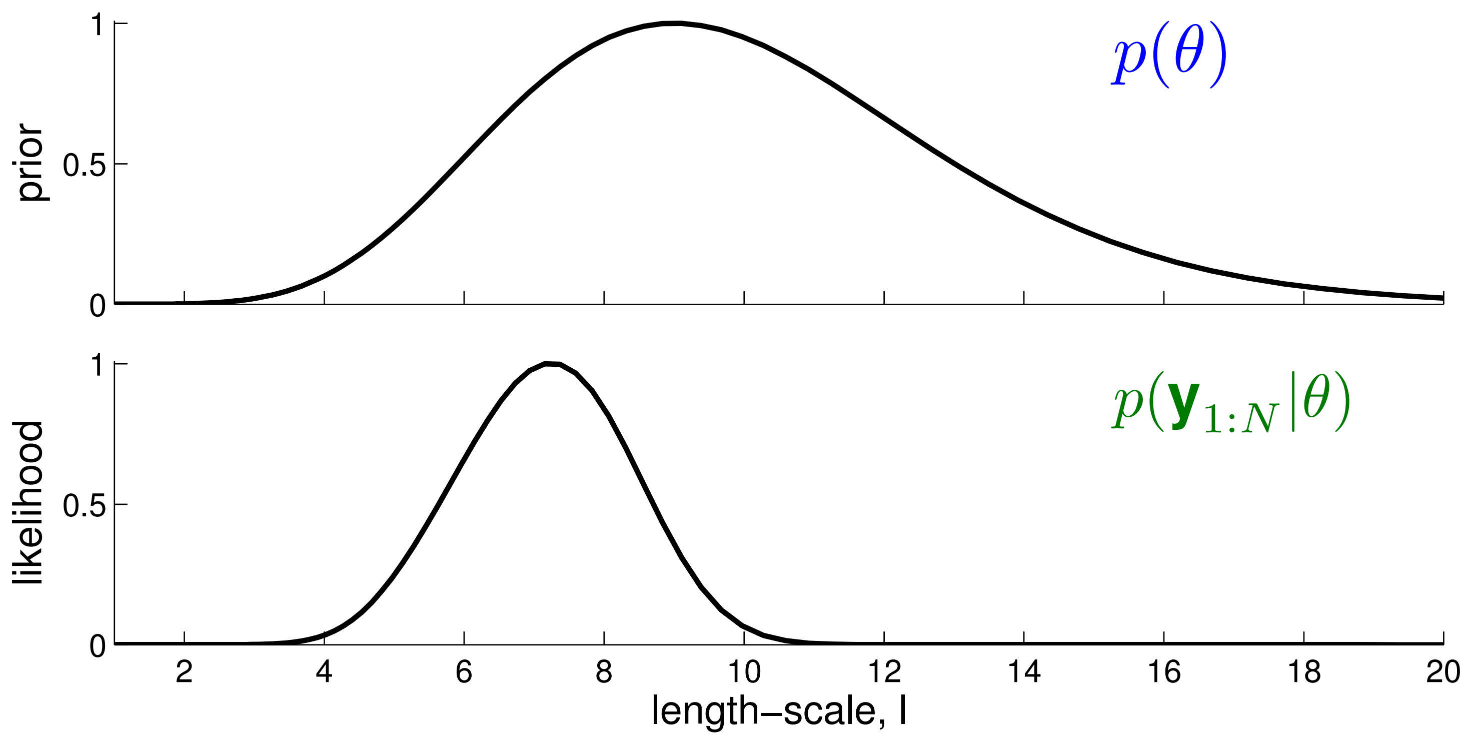
How do we choose the hyper-parameters?



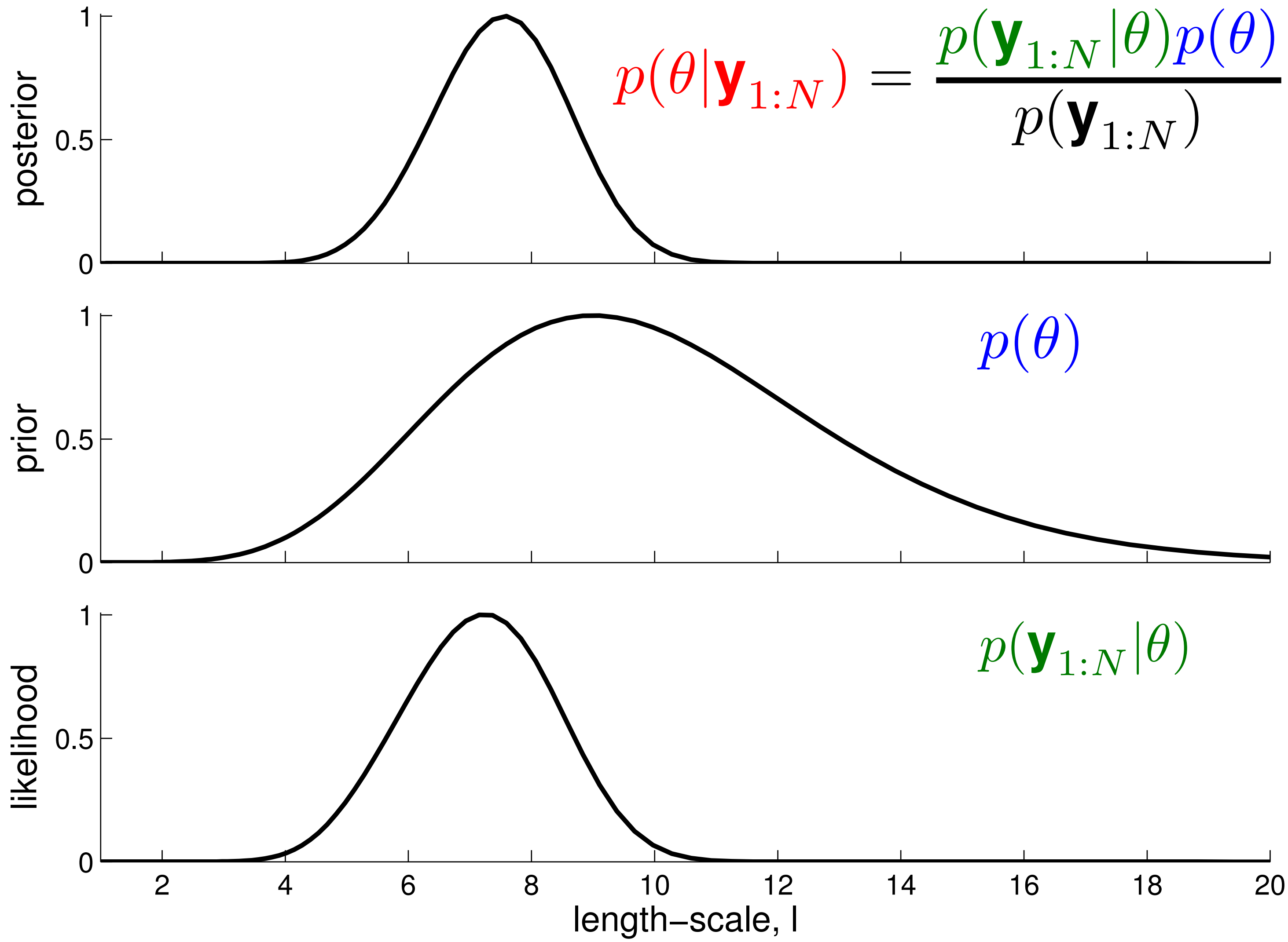
How do we choose the hyper-parameters?



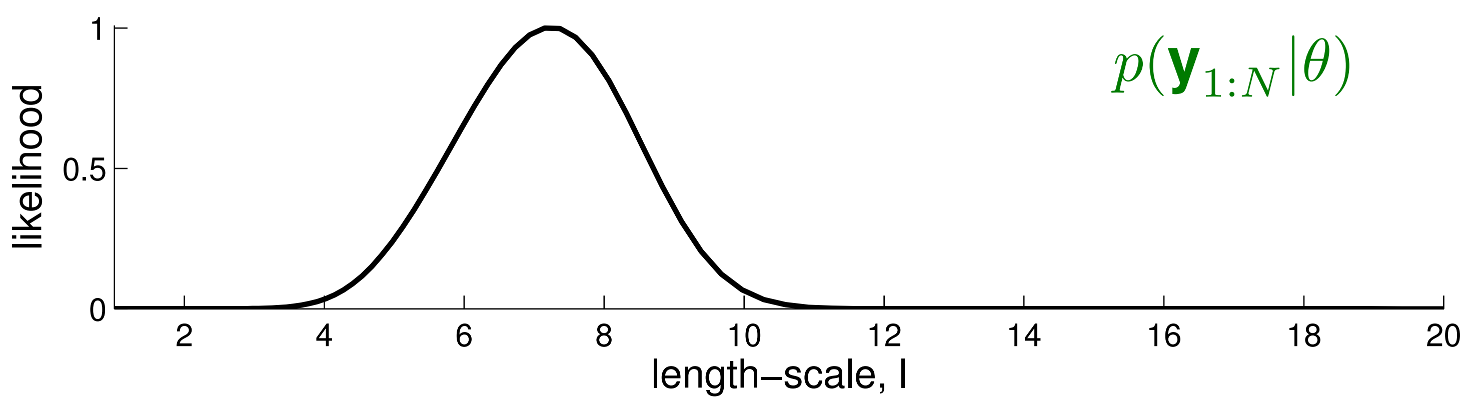
How do we choose the hyper-parameters?



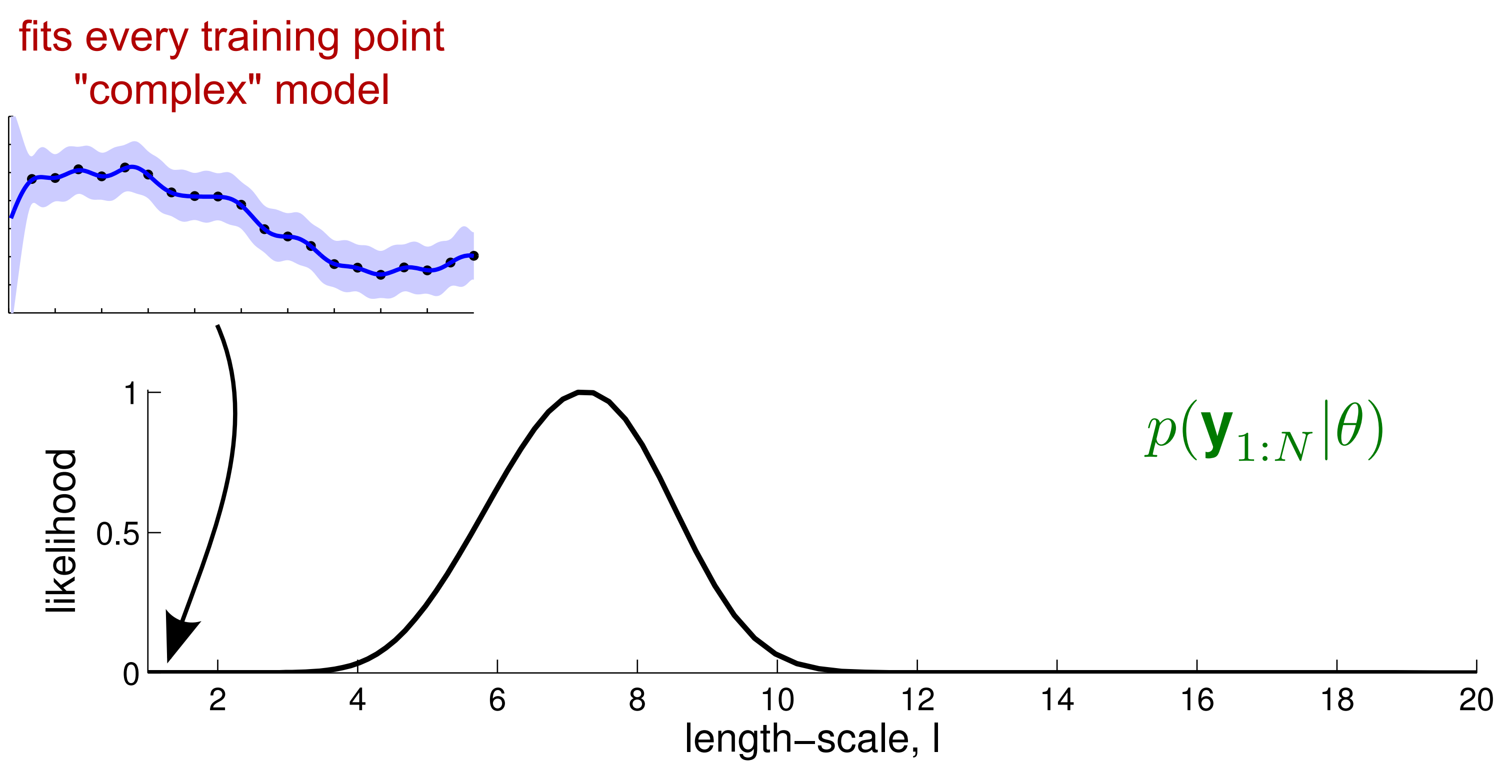
How do we choose the hyper-parameters?



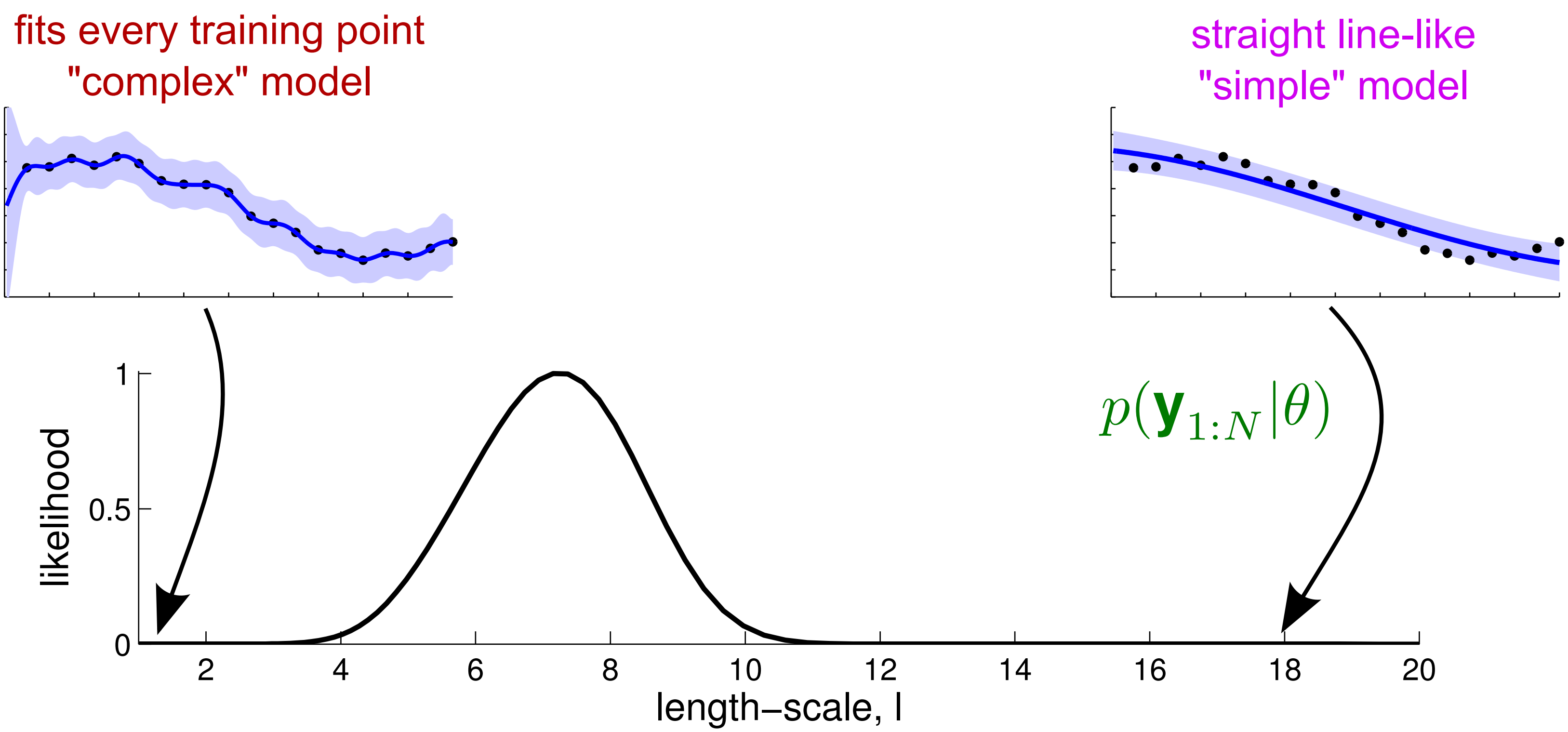
Why does Bayesian inference work?



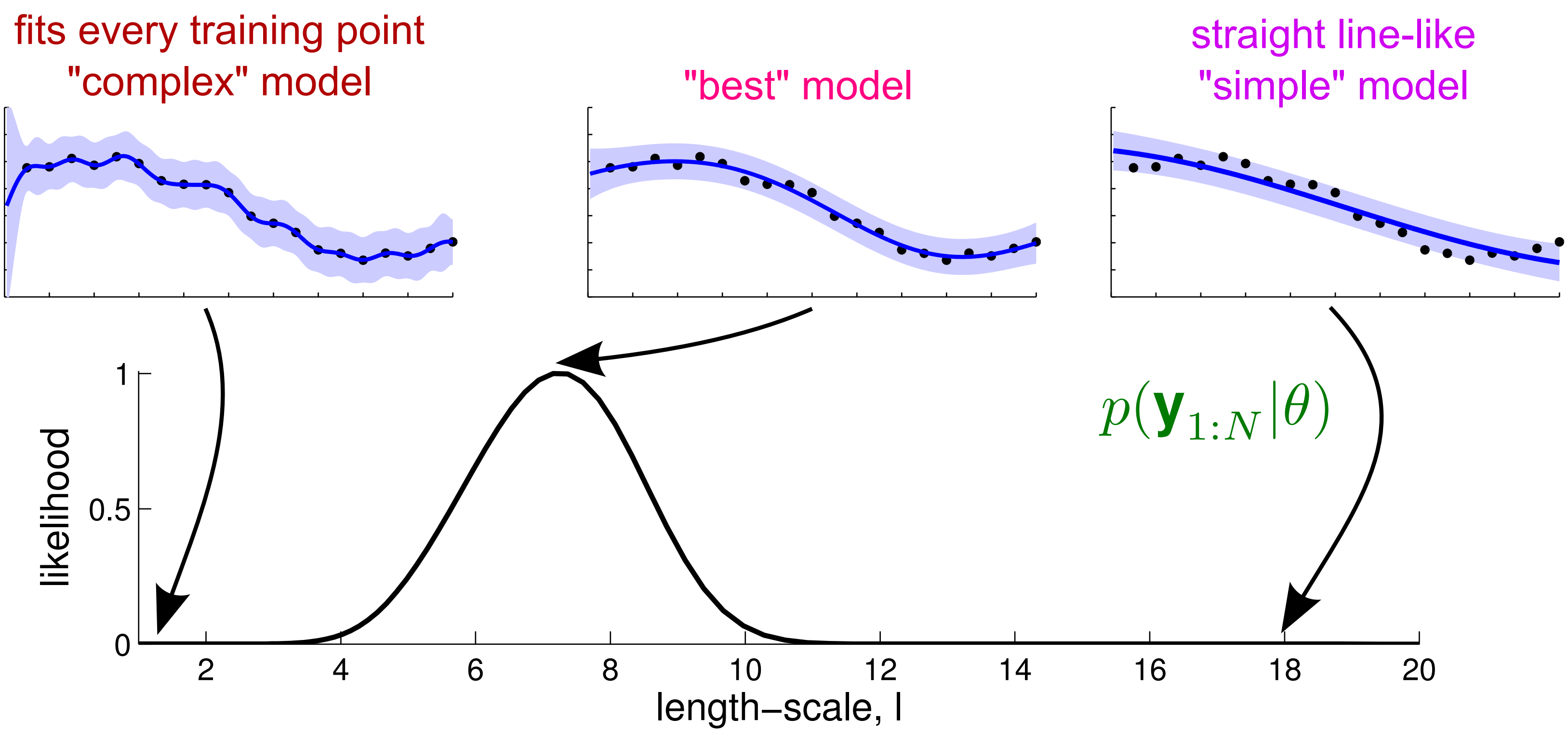
Why does Bayesian inference work?



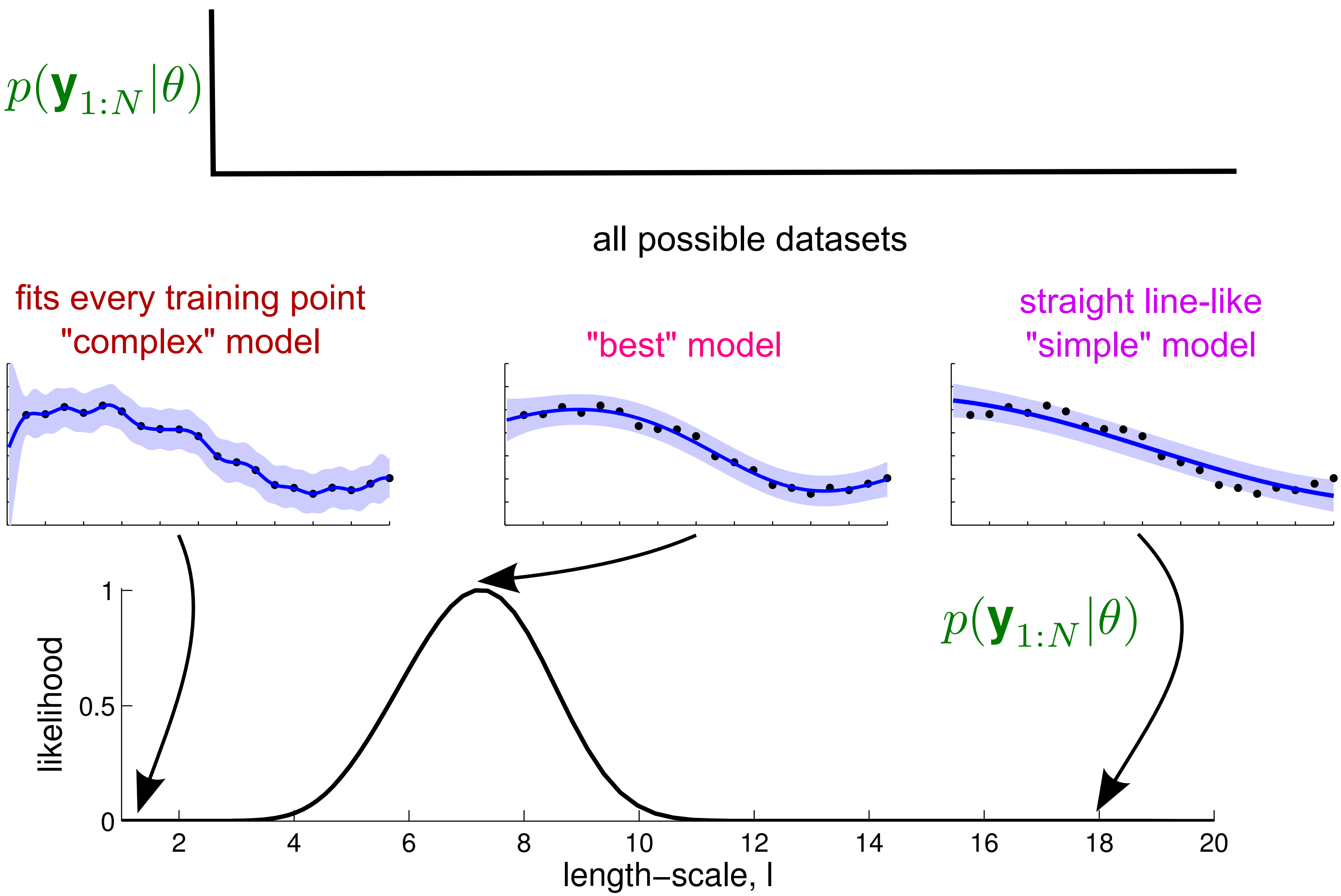
Why does Bayesian inference work?



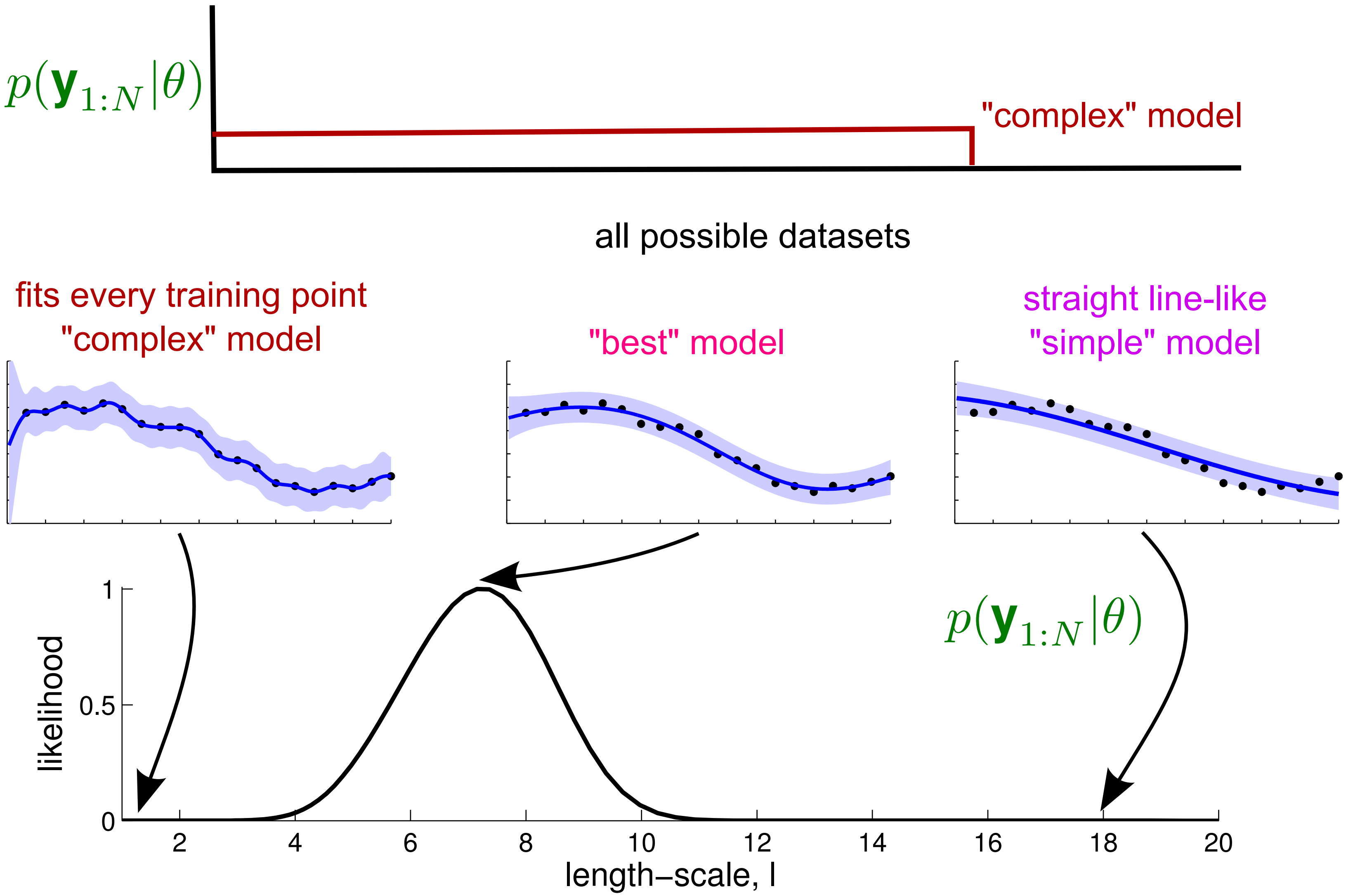
Why does Bayesian inference work?



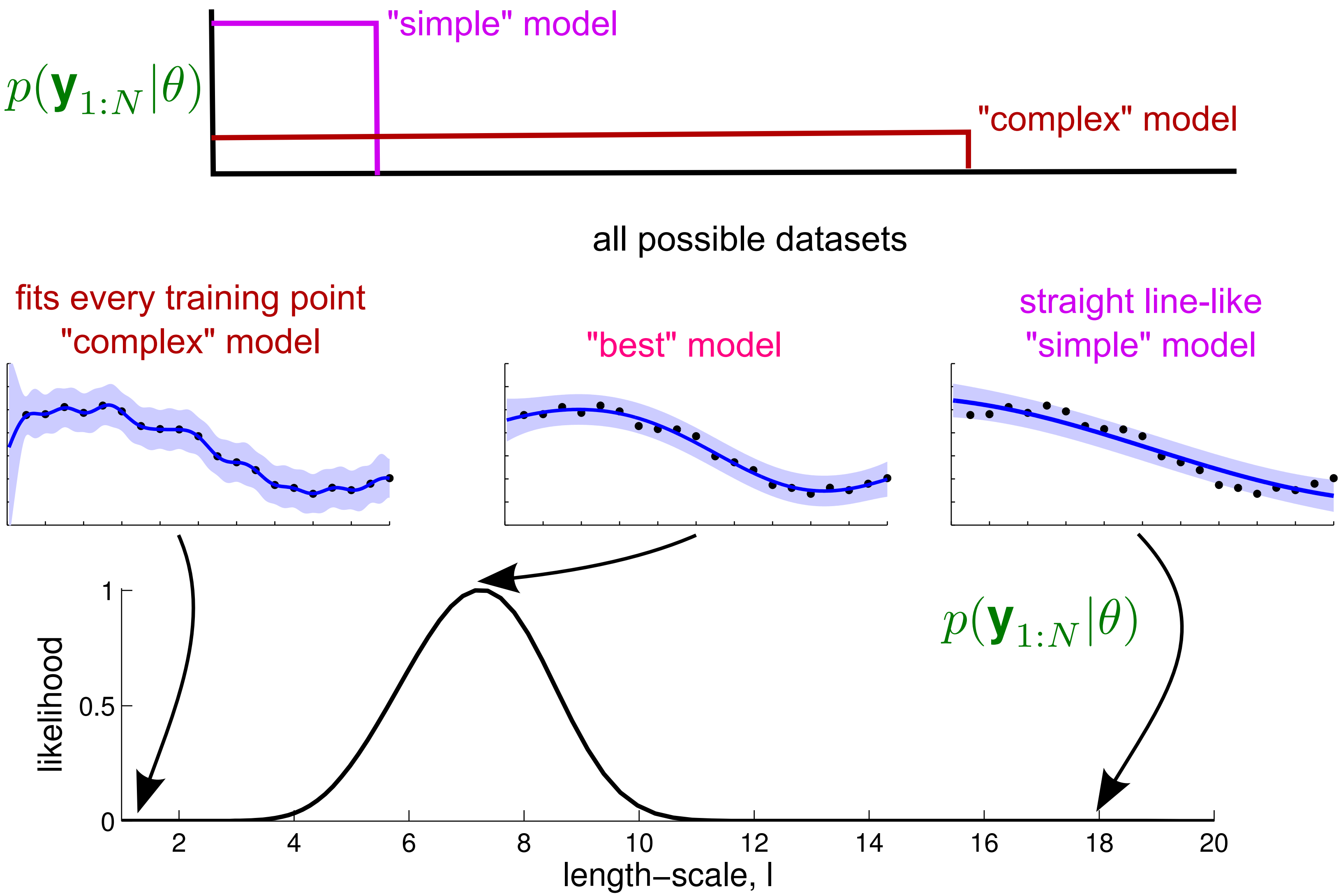
Why does Bayesian inference work?



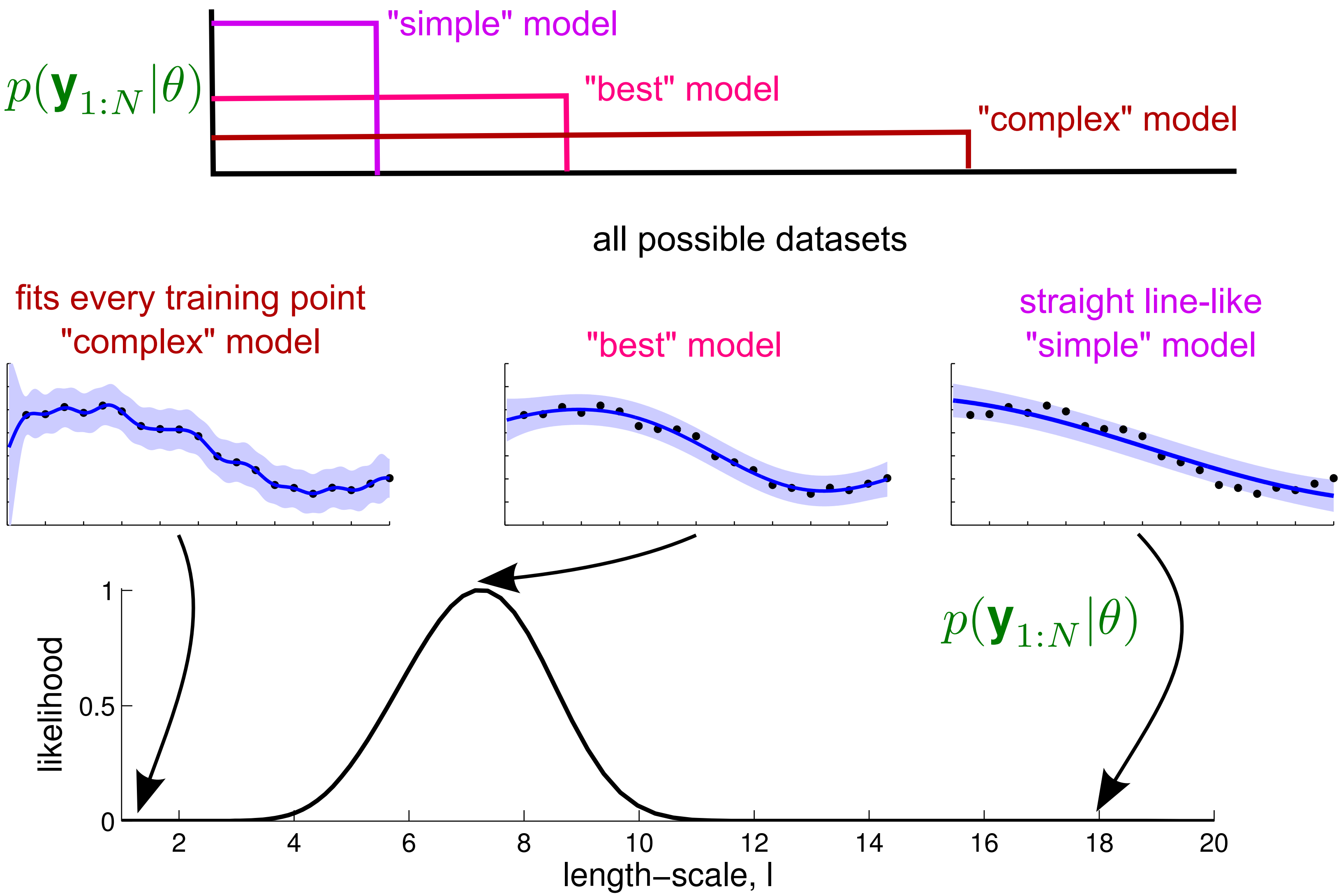
Why does Bayesian inference work?



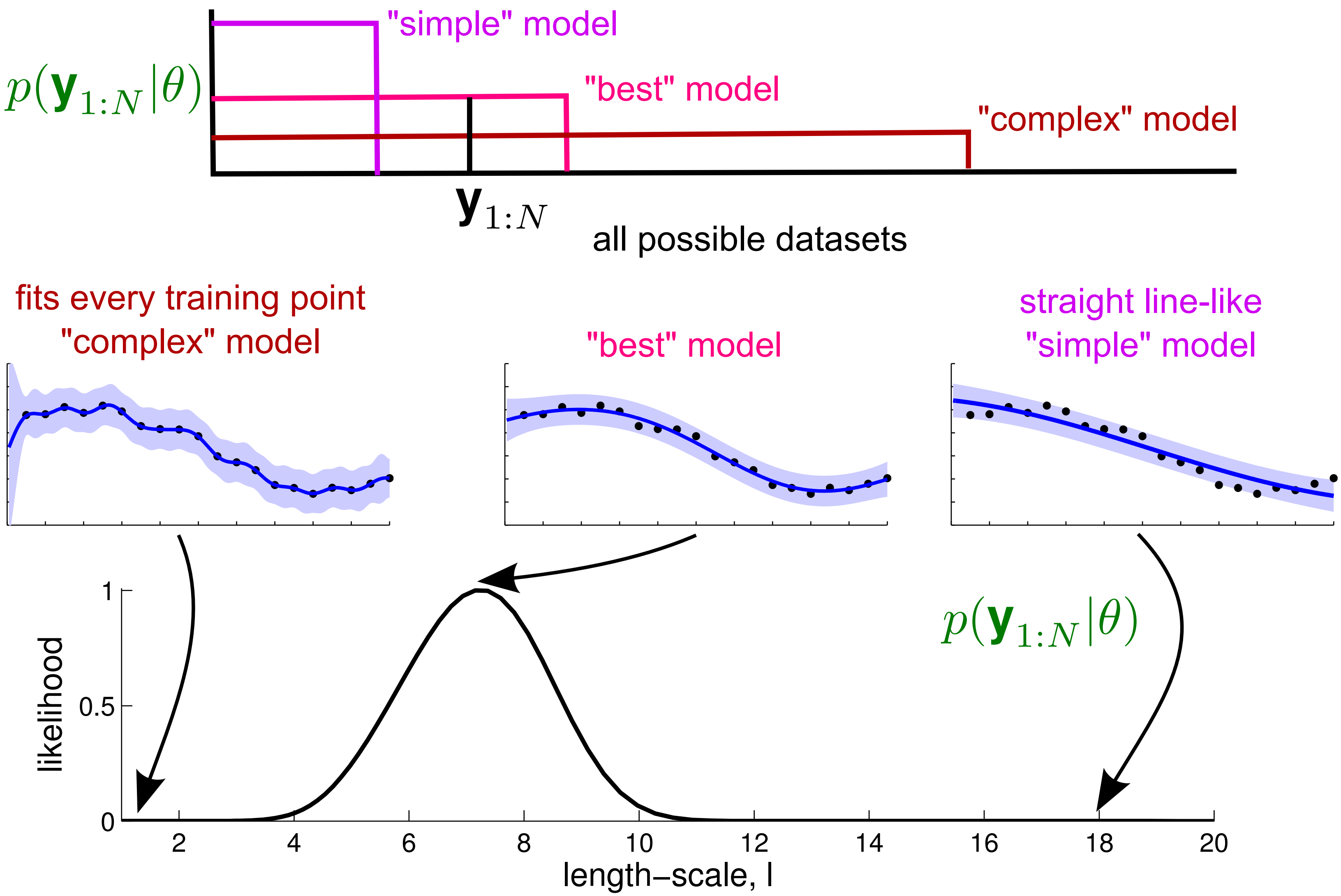
Why does Bayesian inference work?



Why does Bayesian inference work?



Why does Bayesian inference work? Occam's Razor.



Summary

- Gaussian process: **collection of random variables, any finite subset of which are Gaussian distributed**
- Easy to use
  - Predictions correspond to models with infinite numbers of parameters
- GPs have many standard methods as special cases
- Problem: N^3 complexity
  - approximation methods for $N > 2000$ or special covariance functions
- **Great reference**: Rasmussen & Williams `www.gaussianprocess.org/`

Beyond regression

GPs useful when ever a prior over functions is required

- dimensionality reduction
- time-series models (Kalman filter)
- clustering
- active learning
- reinforcement learning
- ...