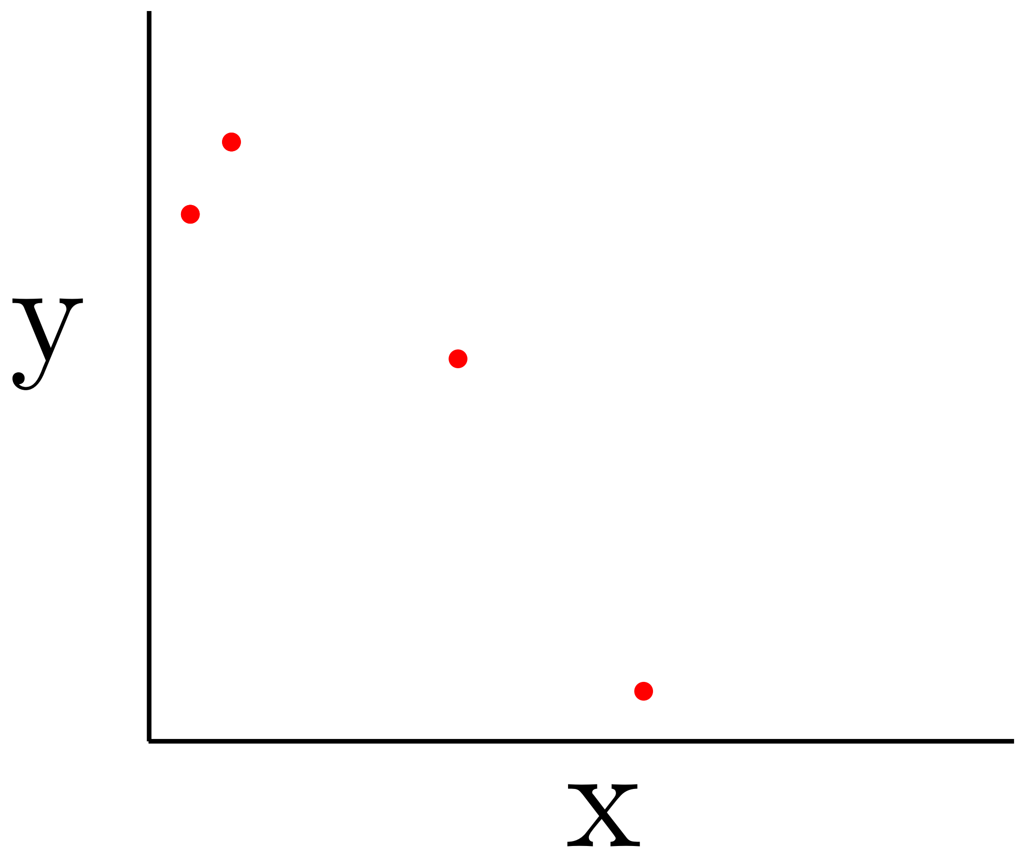
An introduction to Gaussian processes for probabilistic inference

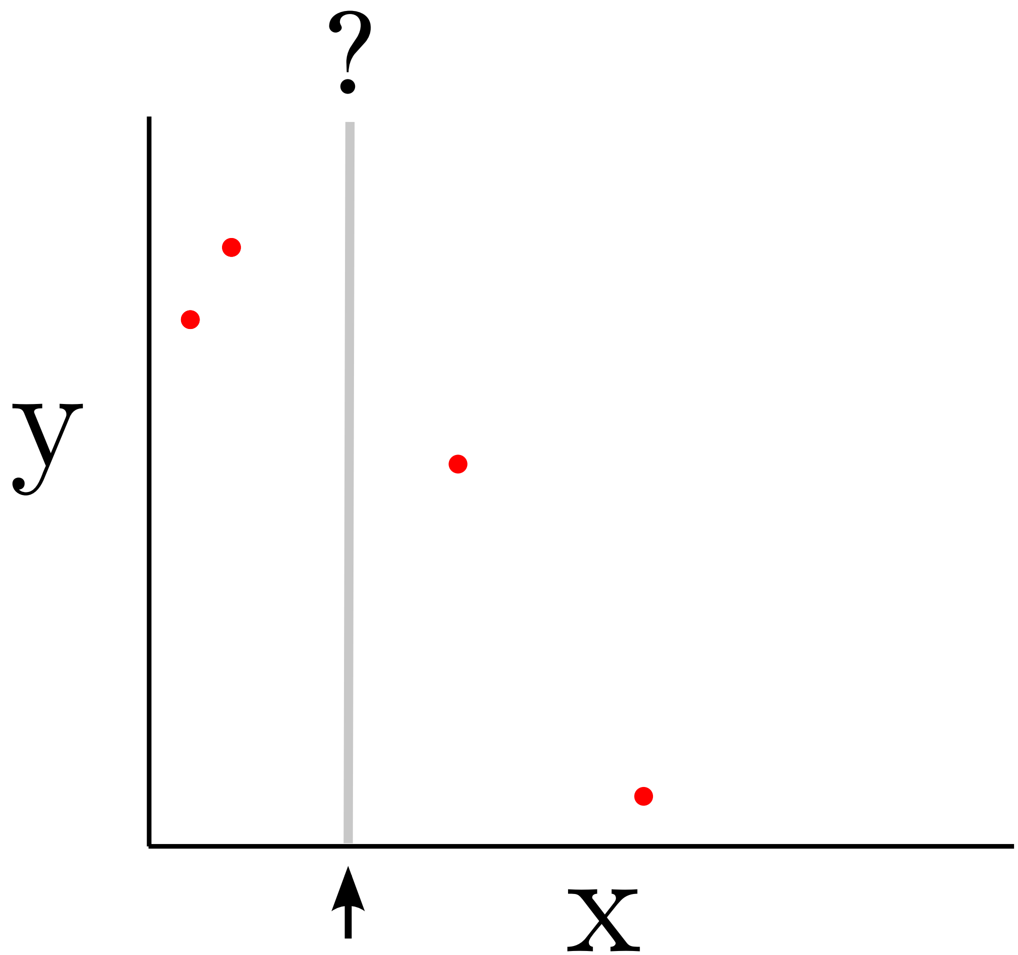
Dr. Richard E. Turner (ret26@cam.ac.uk)

Computational and Biological Learning Lab, Department of Engineering,
University of Cambridge

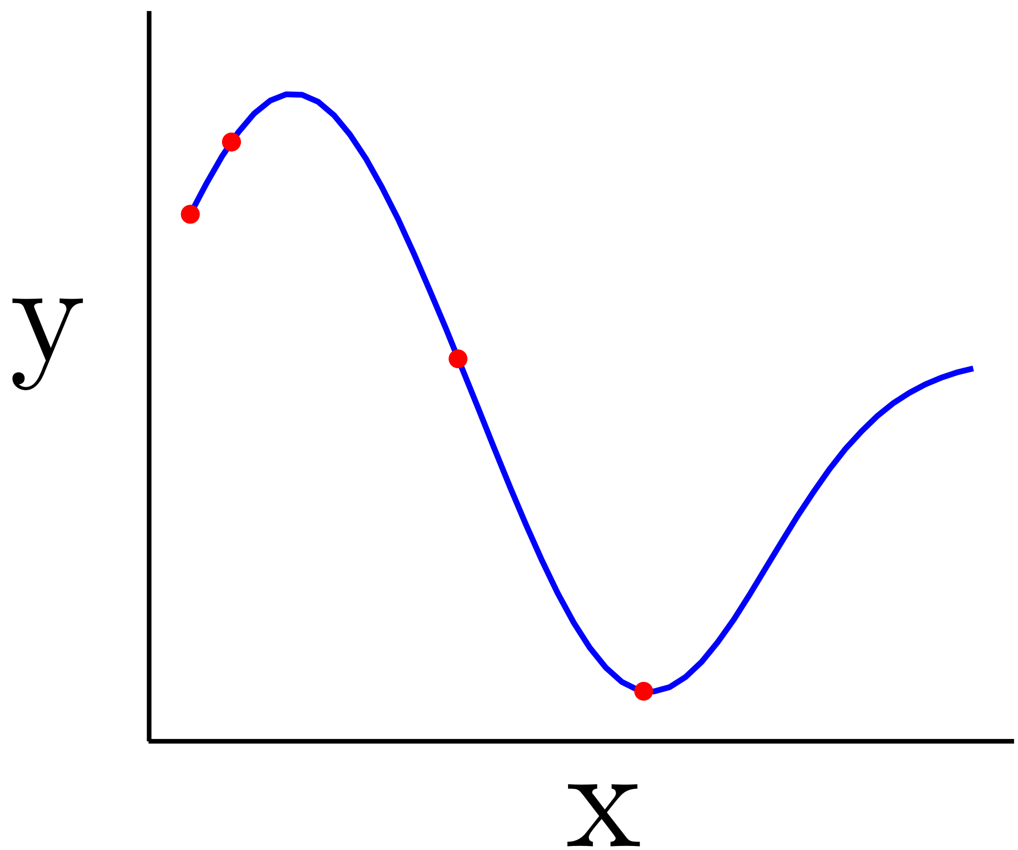
Motivation: non-linear regression



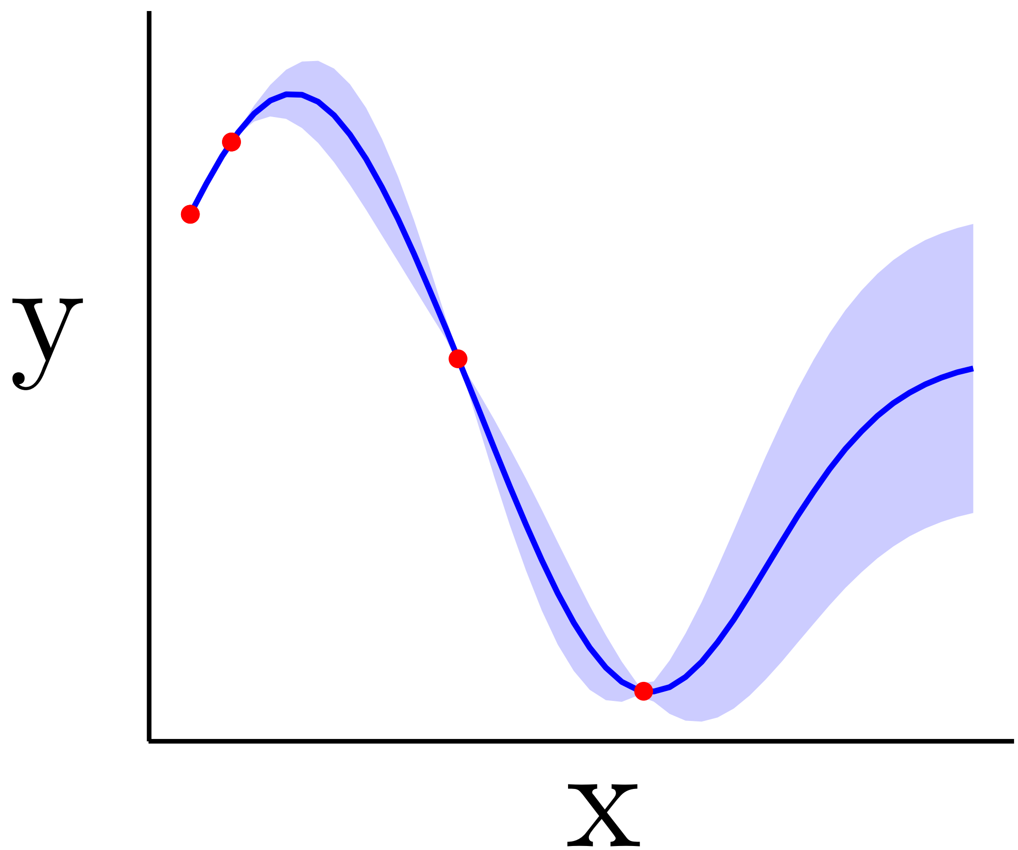
Motivation: non-linear regression



Motivation: non-linear regression

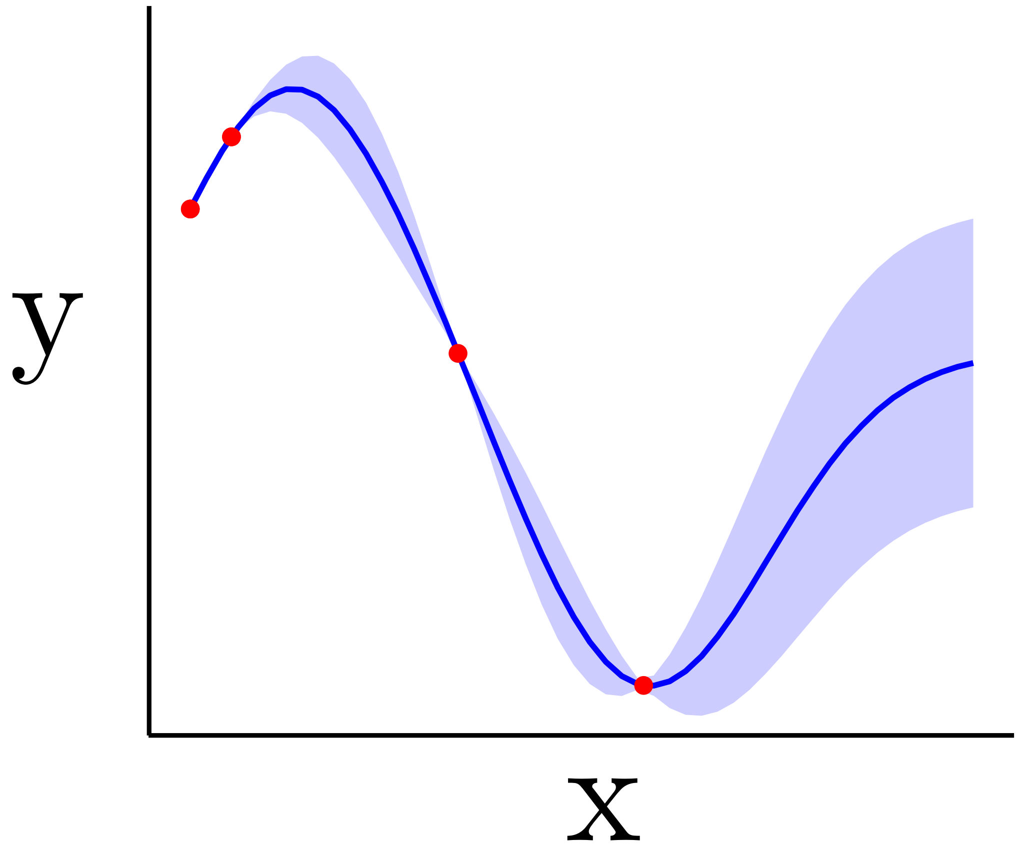


Motivation: non-linear regression



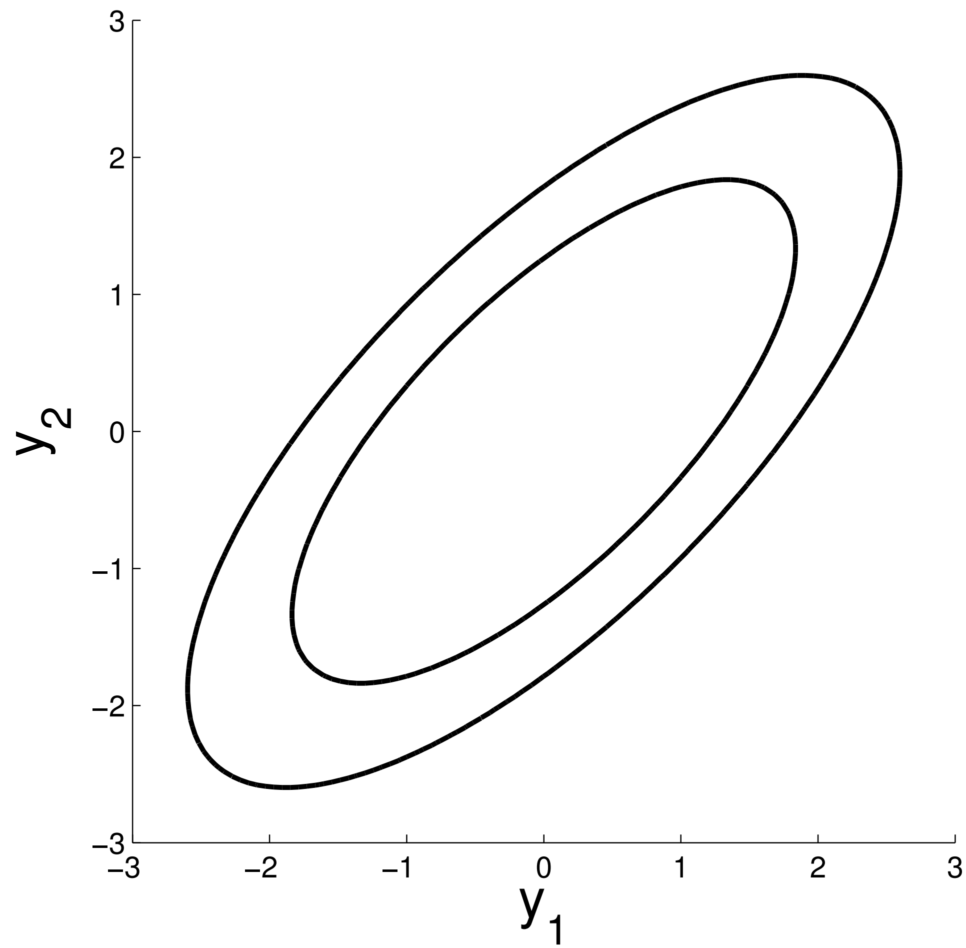
Motivation: non-linear regression

Can we do this with a plain old Gaussian?



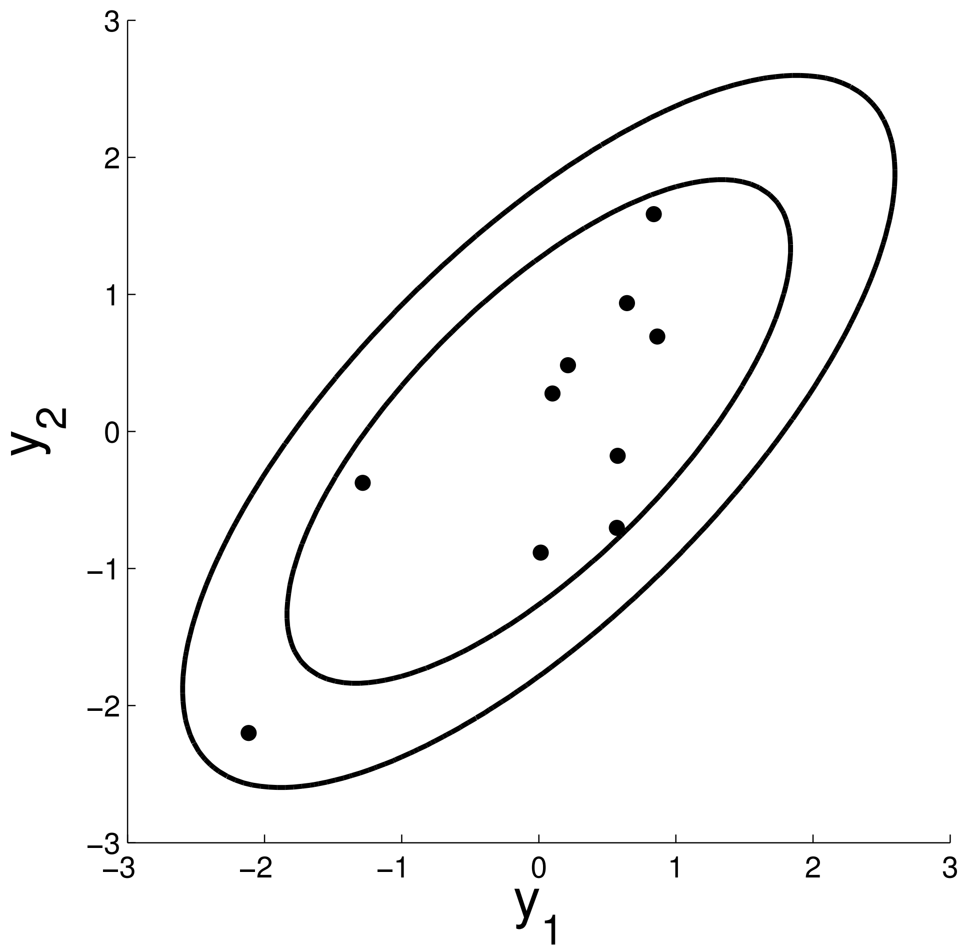
Gaussian distribution

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



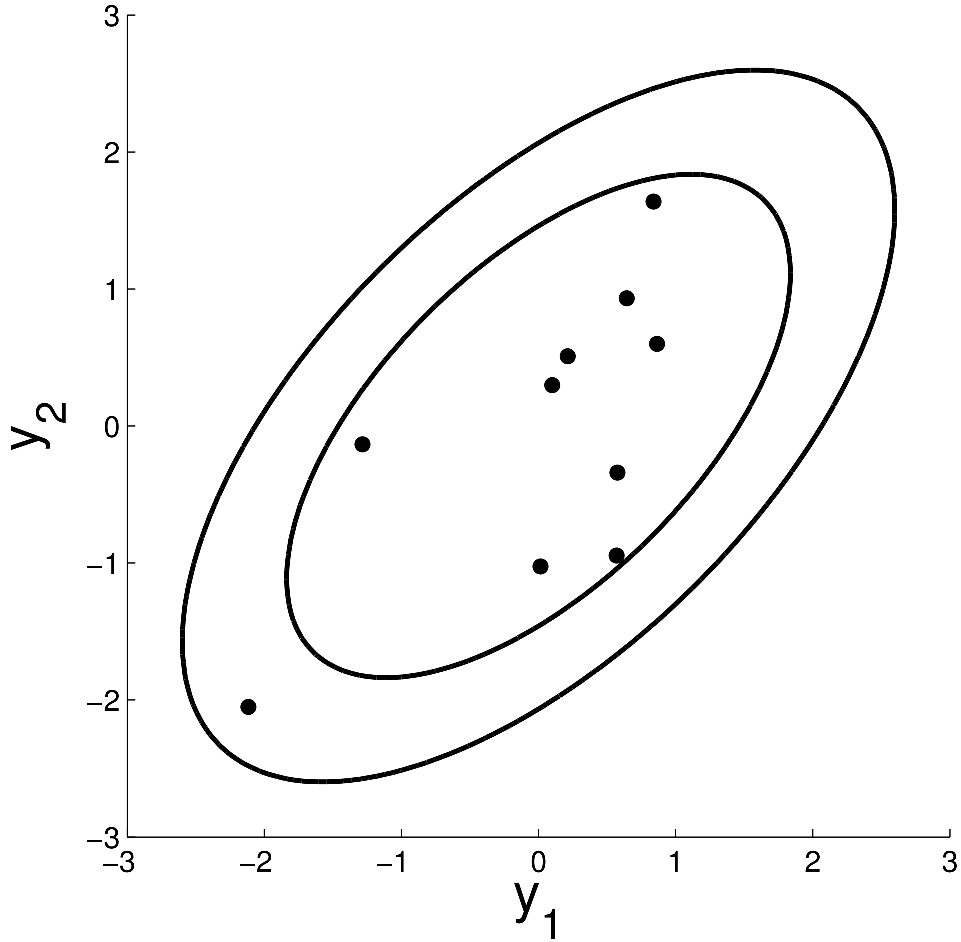
Gaussian distribution

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



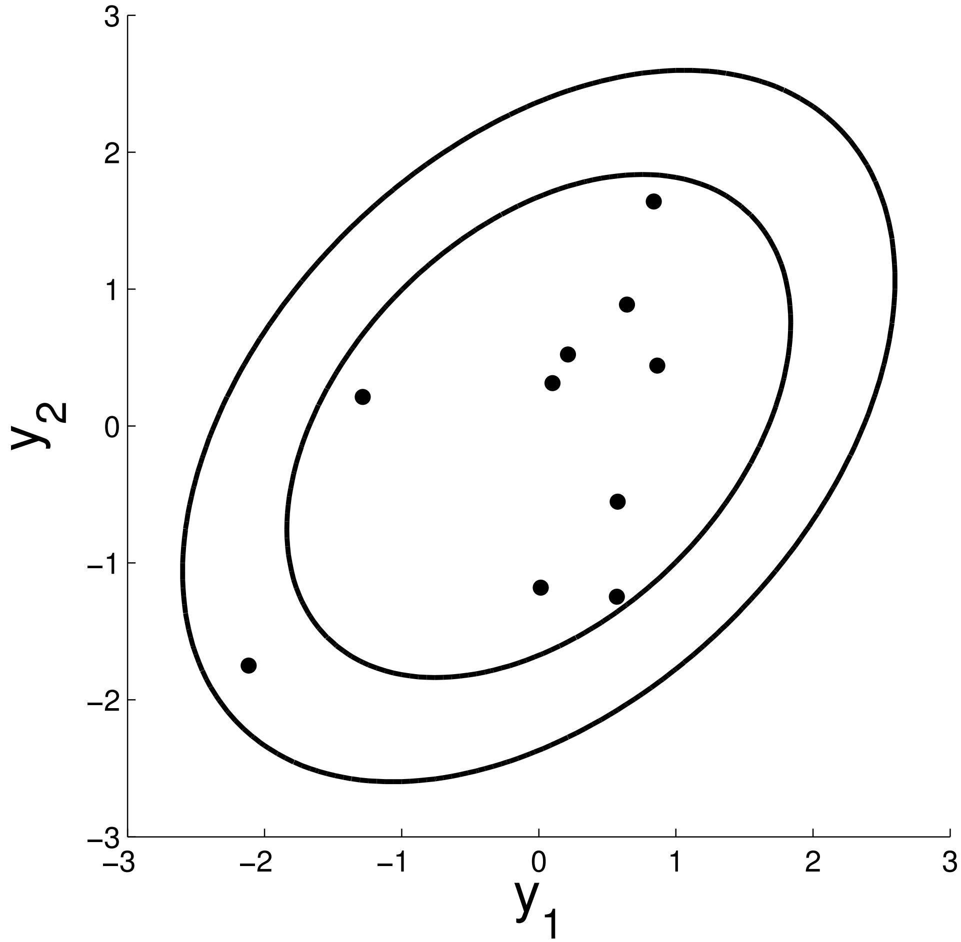
Gaussian distribution

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



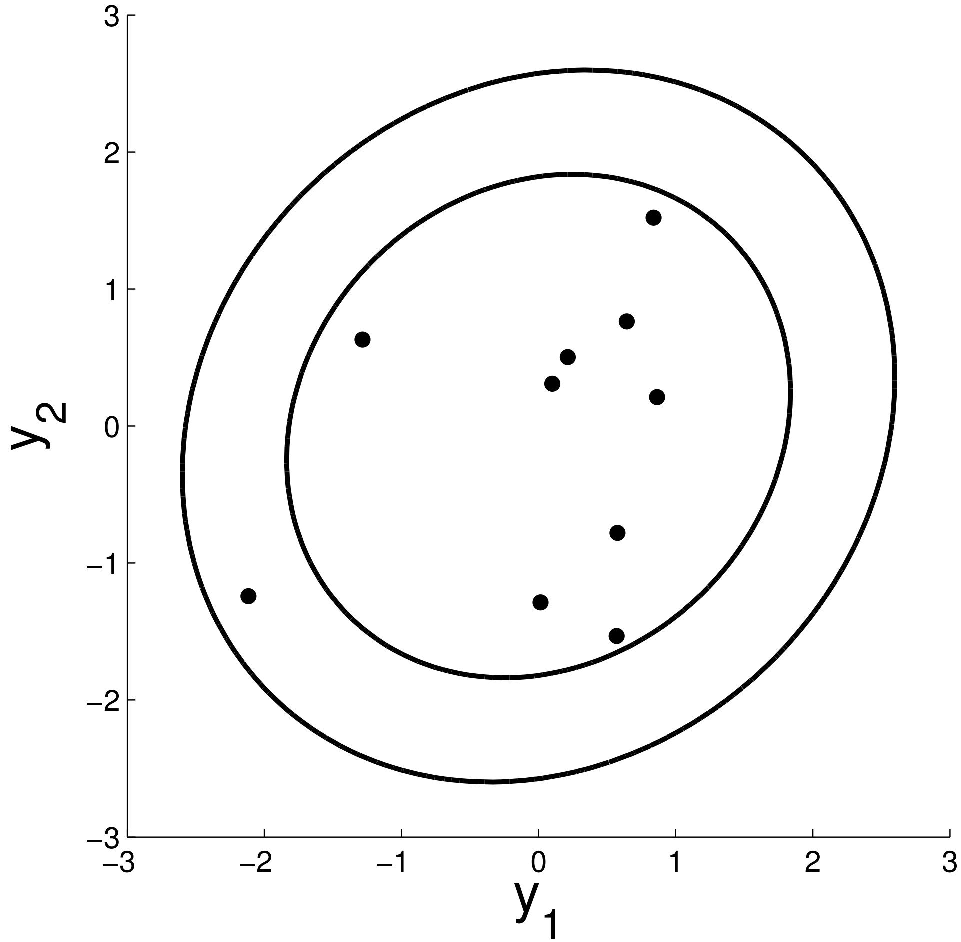
Gaussian distribution

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



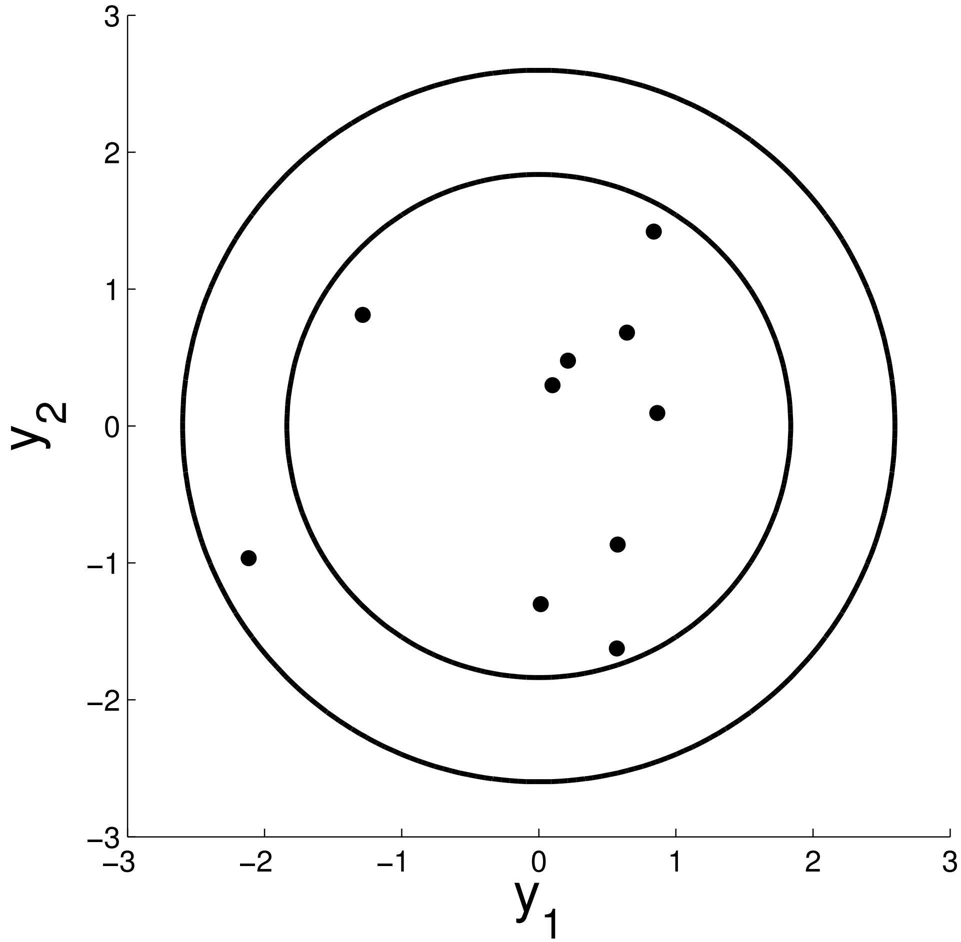
Gaussian distribution

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



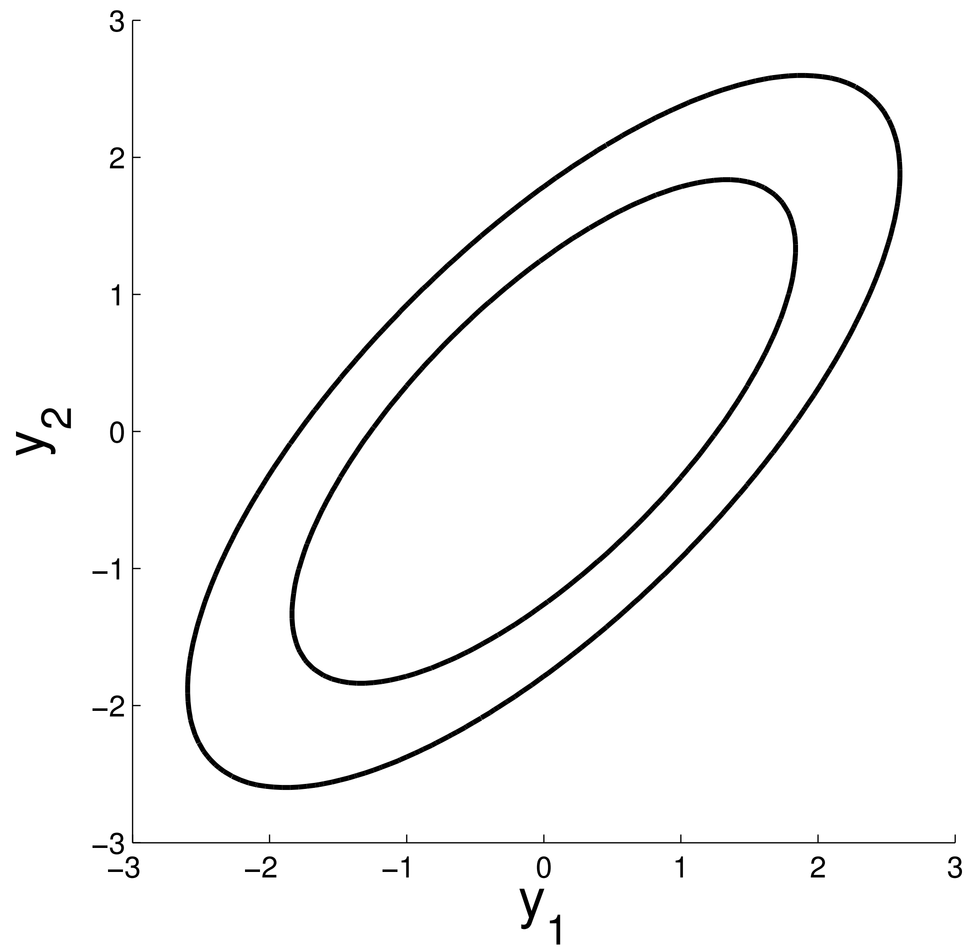
Gaussian distribution

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



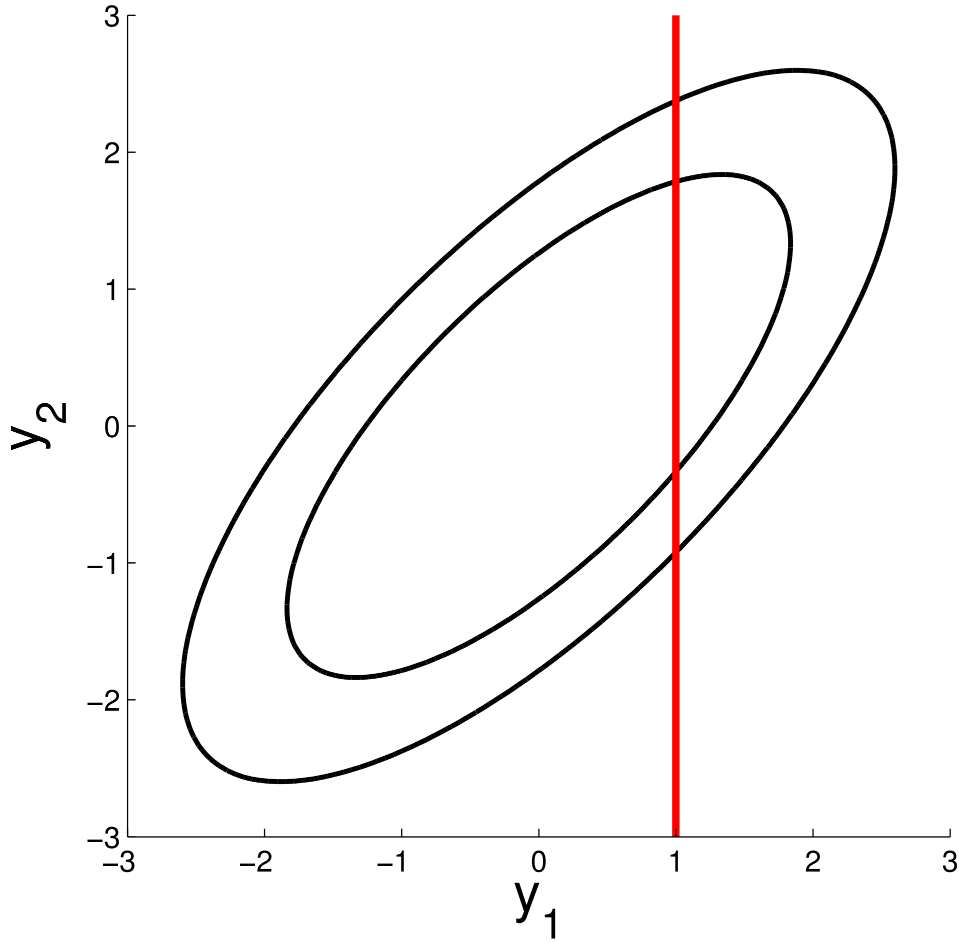
Gaussian distribution

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



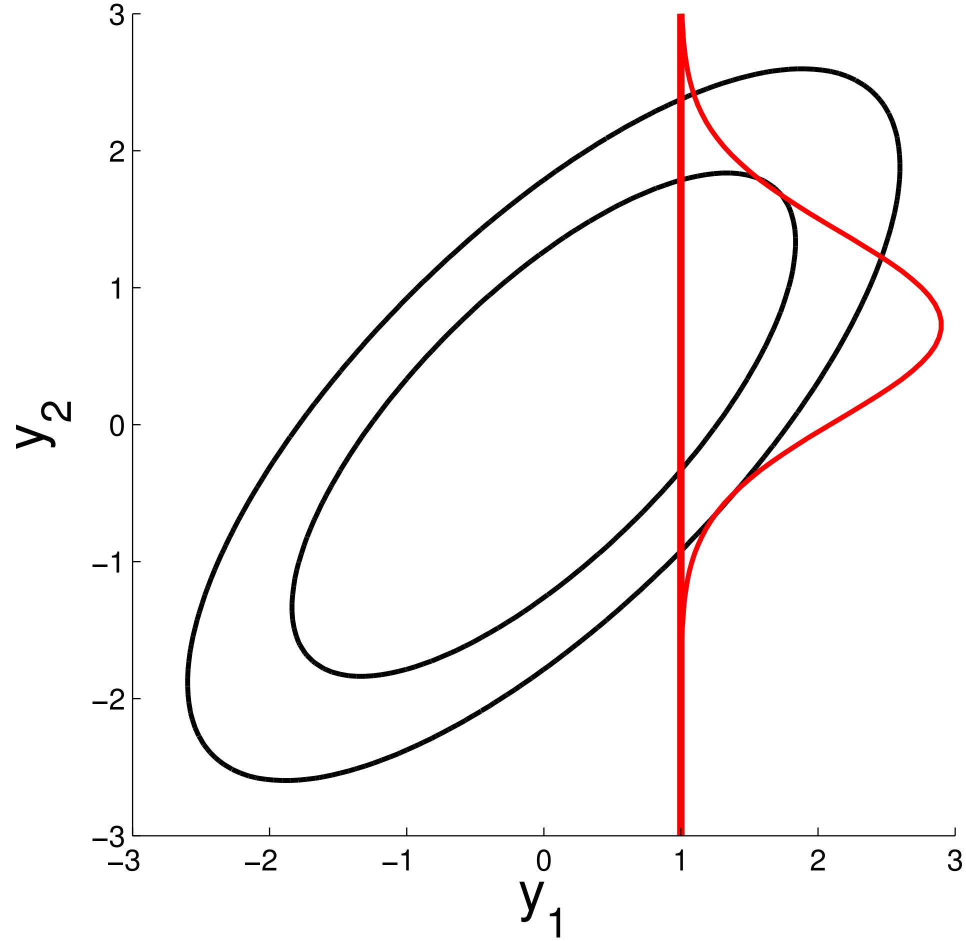
Gaussian distribution

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



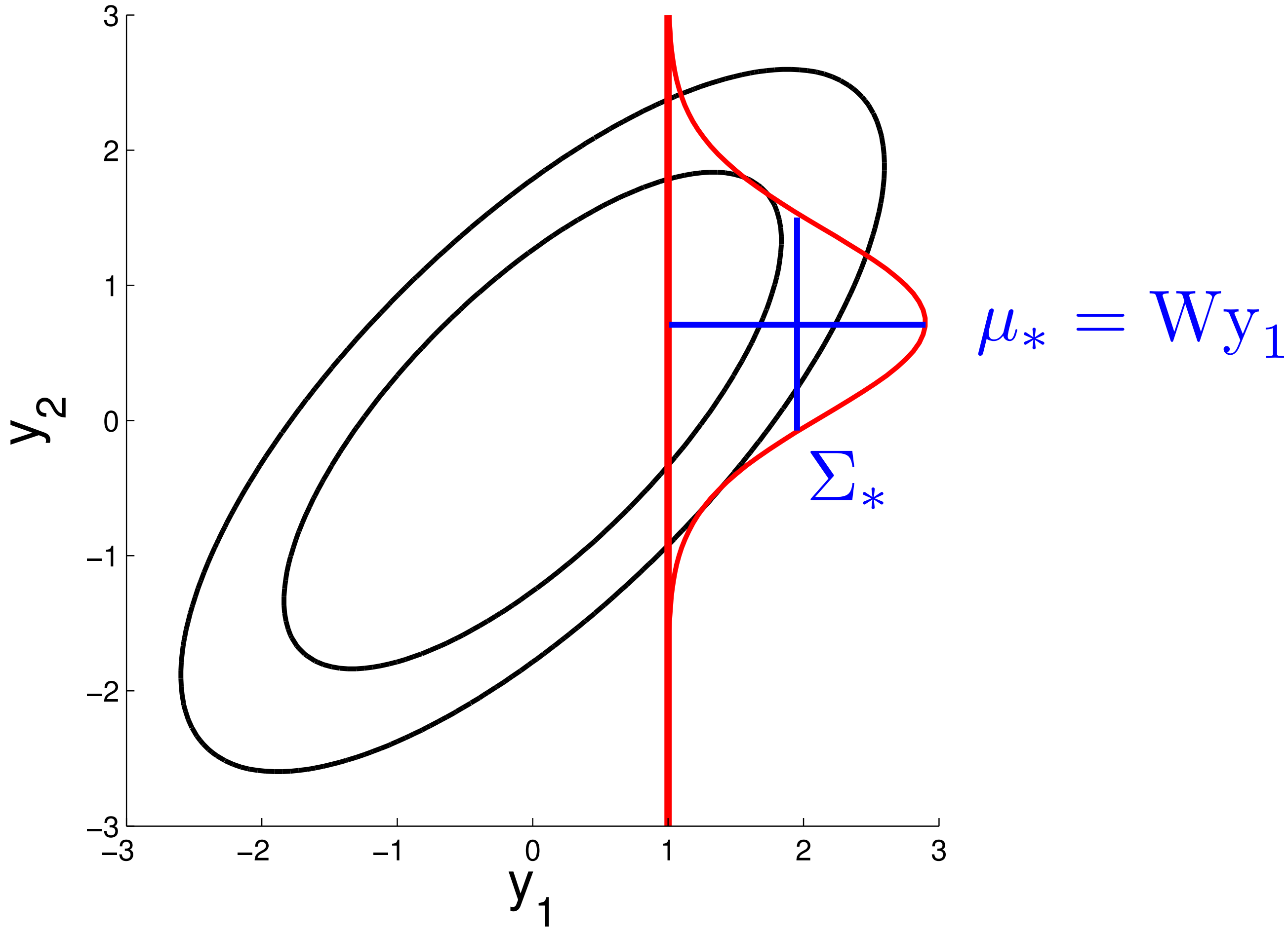
Gaussian distribution

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



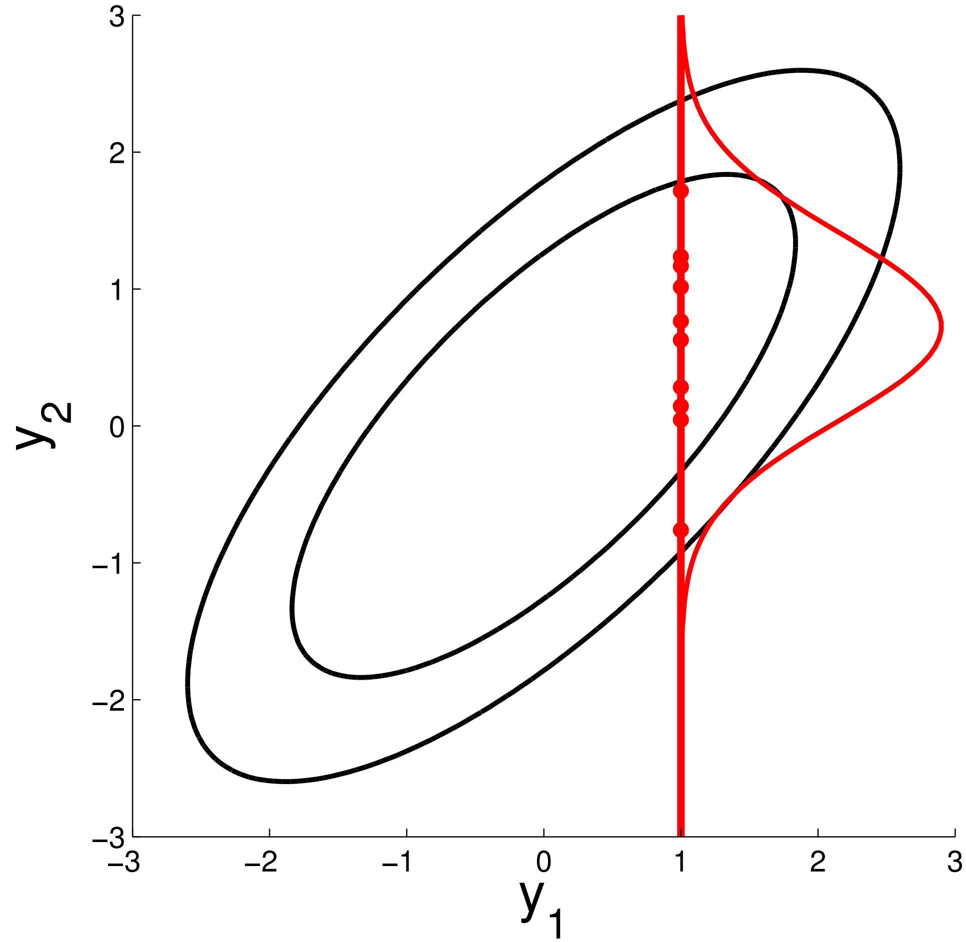
Gaussian distribution

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



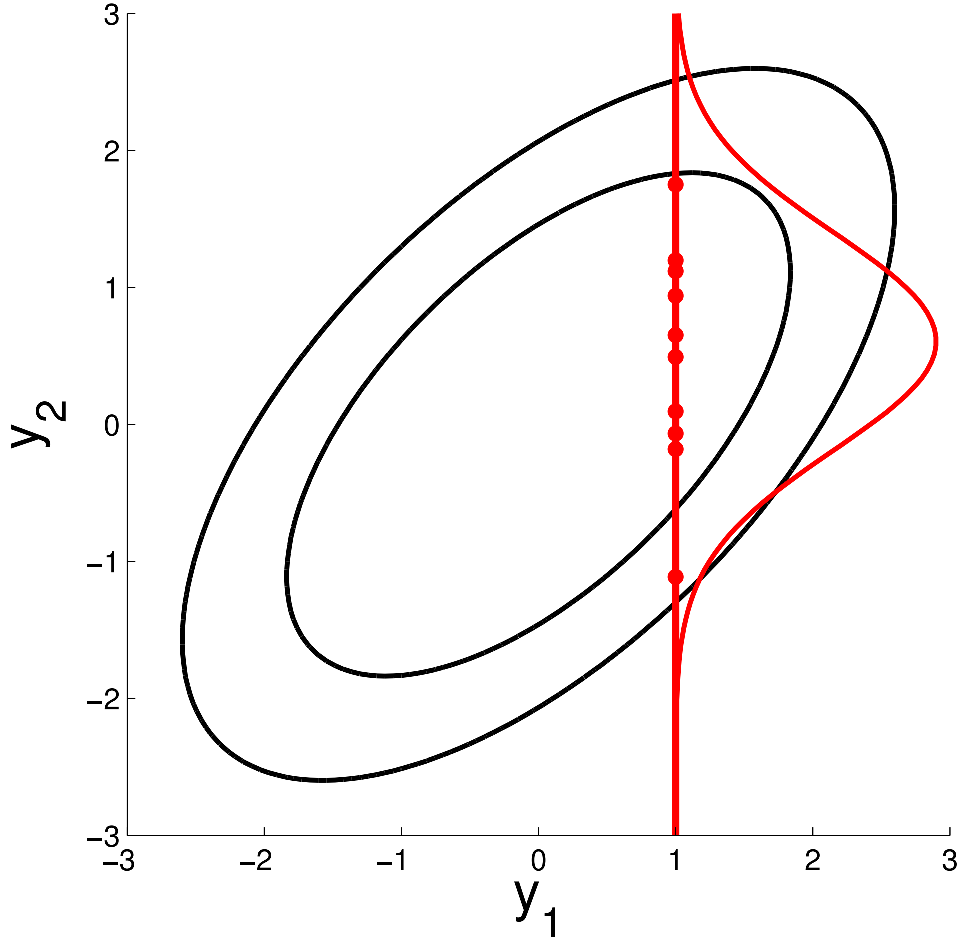
Gaussian distribution

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



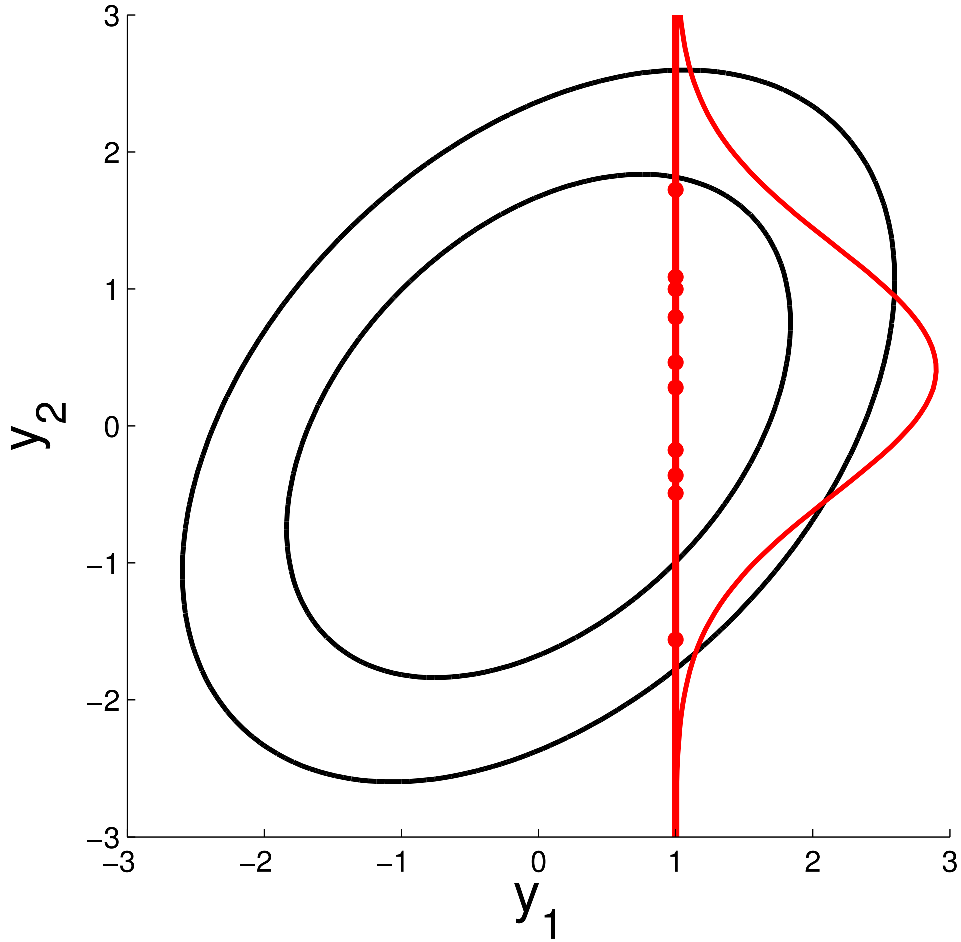
Gaussian distribution

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



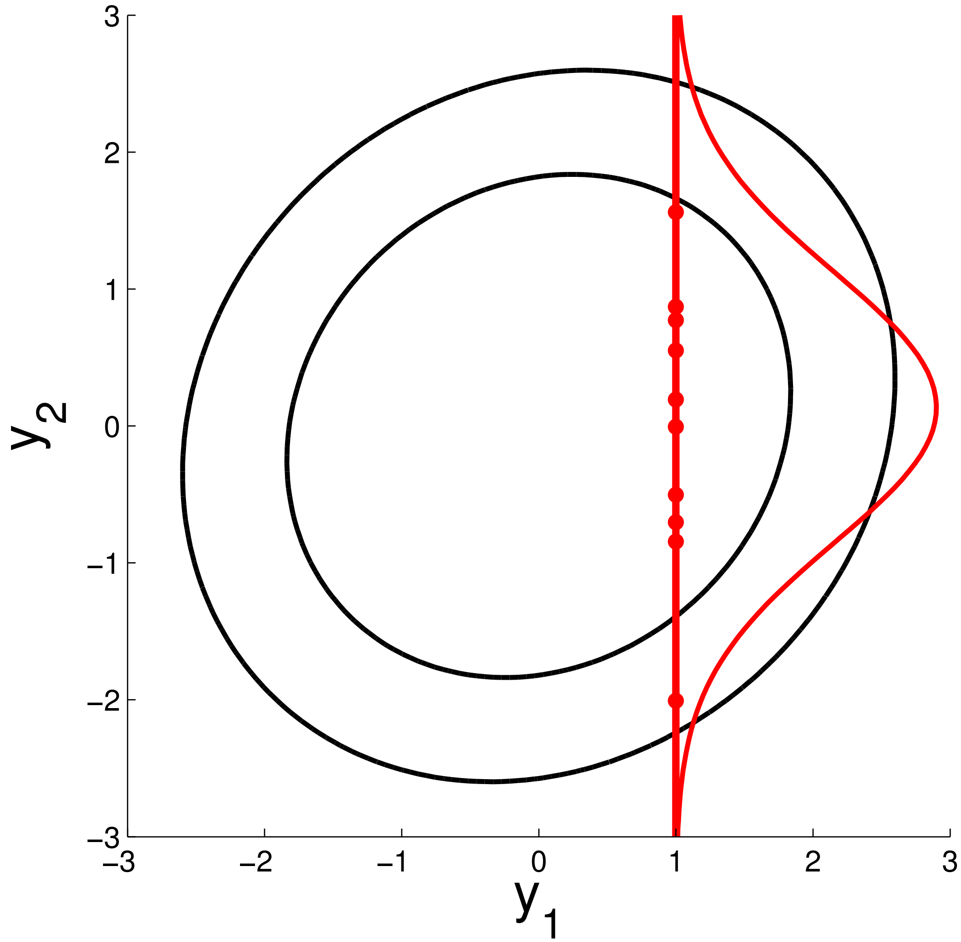
Gaussian distribution

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



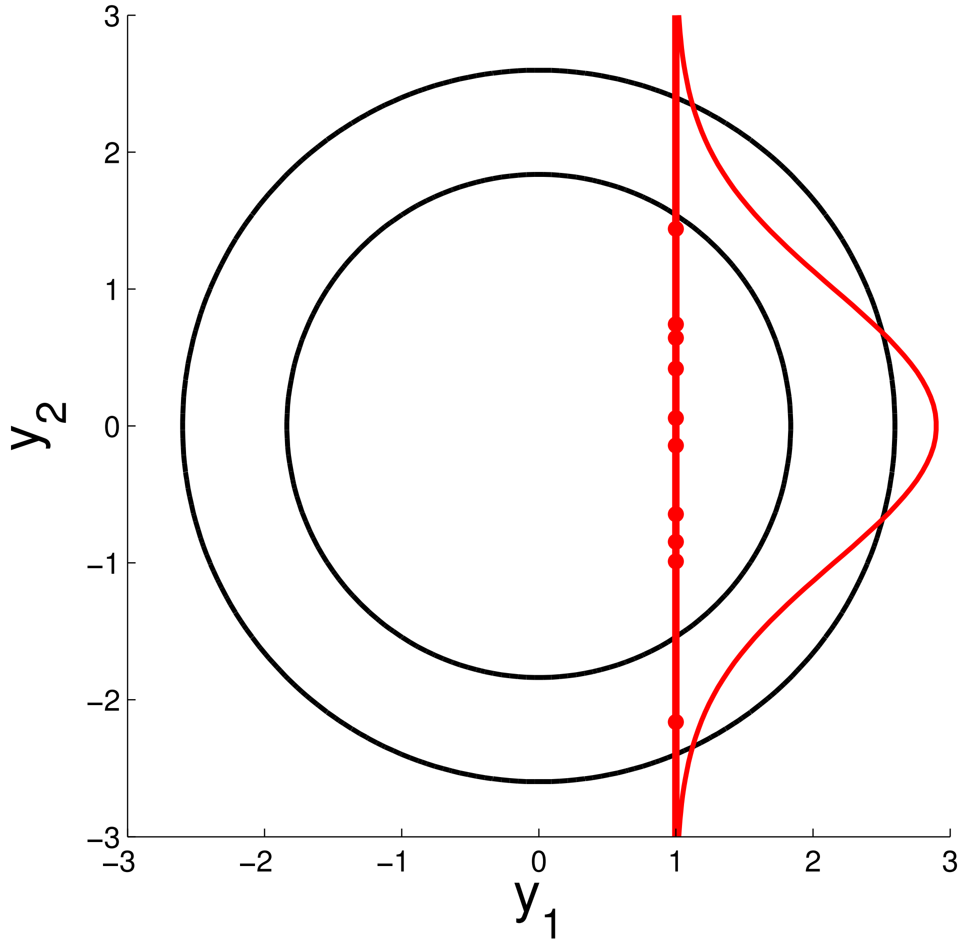
Gaussian distribution

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

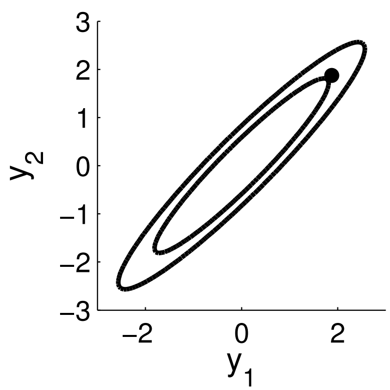


Gaussian distribution

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

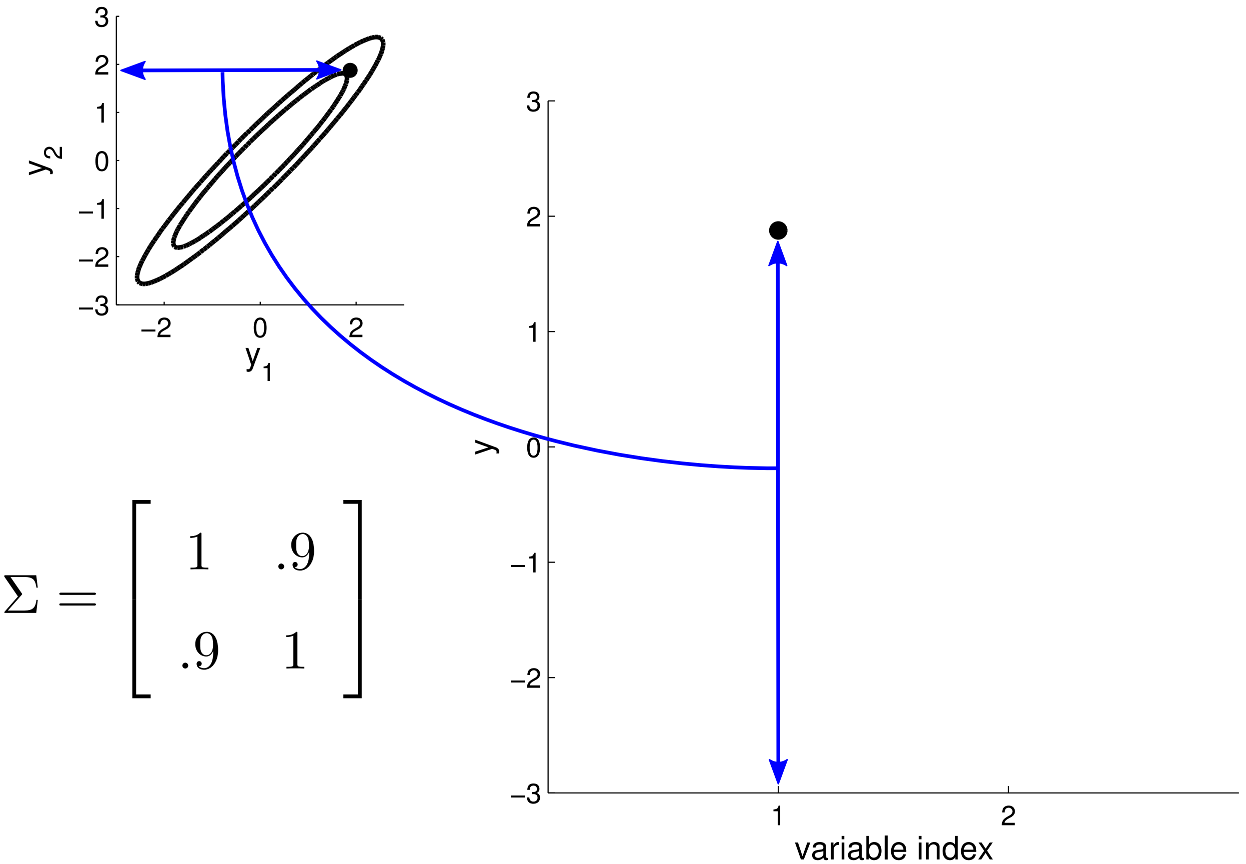


New visualisation



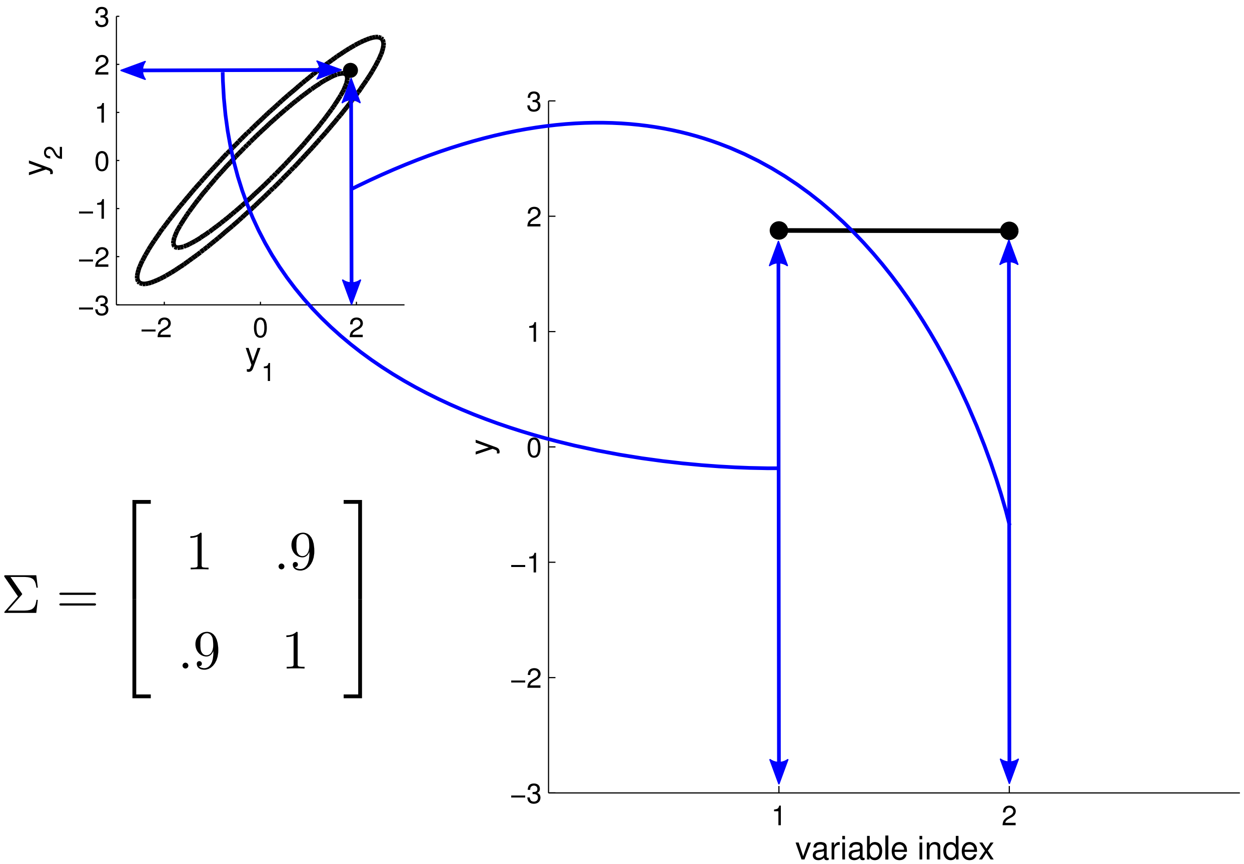
$$\Sigma = \begin{bmatrix} 1 & .9 \\ .9 & 1 \end{bmatrix}$$

New visualisation



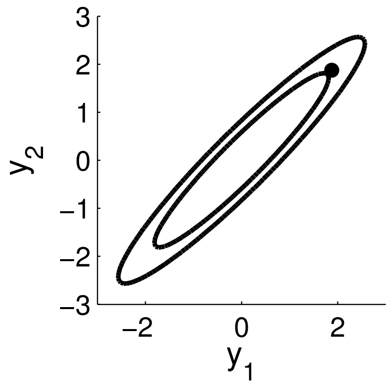
$$\Sigma = \begin{bmatrix} 1 & .9 \\ .9 & 1 \end{bmatrix}$$

New visualisation

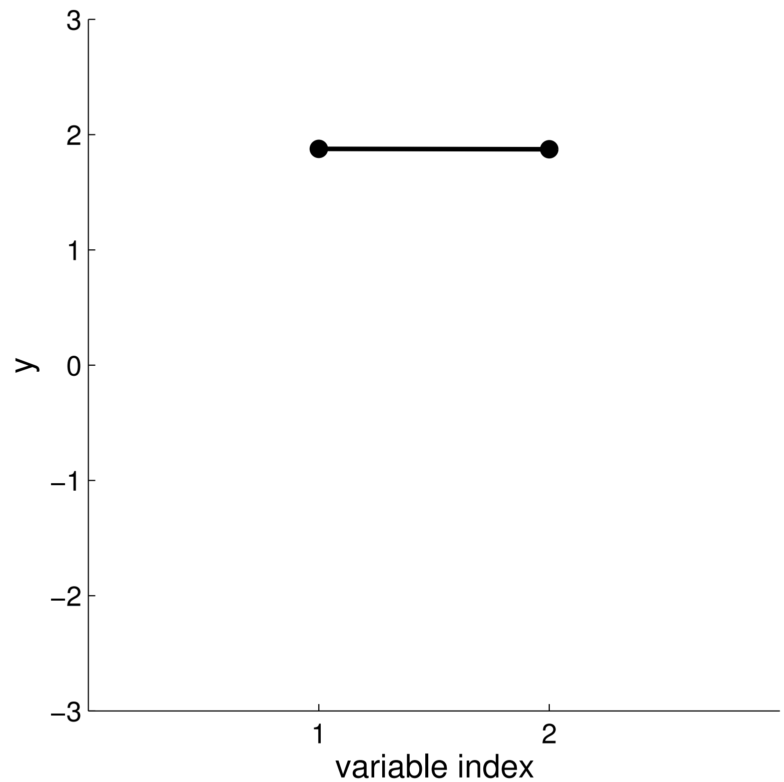


$$\Sigma = \begin{bmatrix} 1 & .9 \\ .9 & 1 \end{bmatrix}$$

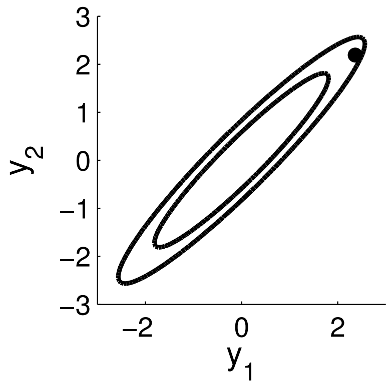
New visualisation



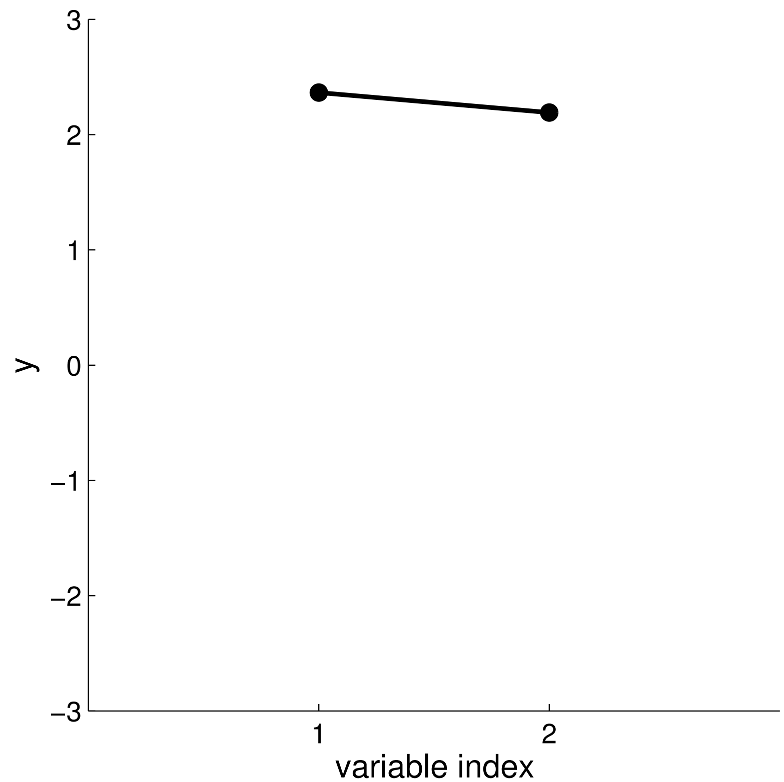
$$\Sigma = \begin{bmatrix} 1 & .9 \\ .9 & 1 \end{bmatrix}$$



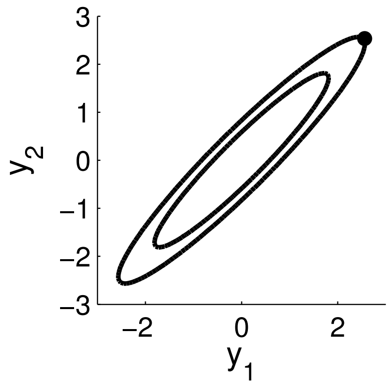
New visualisation



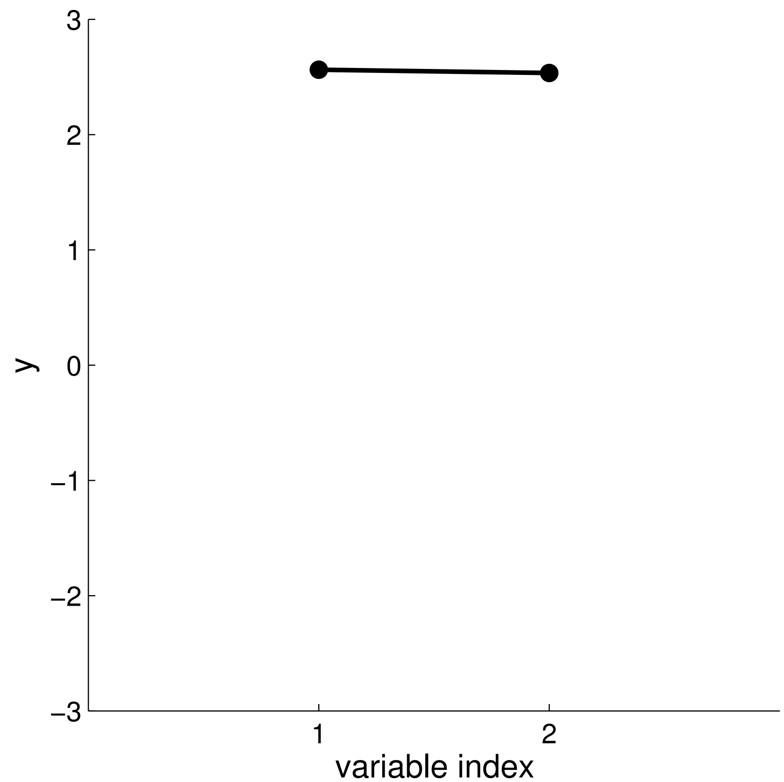
$$\Sigma = \begin{bmatrix} 1 & .9 \\ .9 & 1 \end{bmatrix}$$



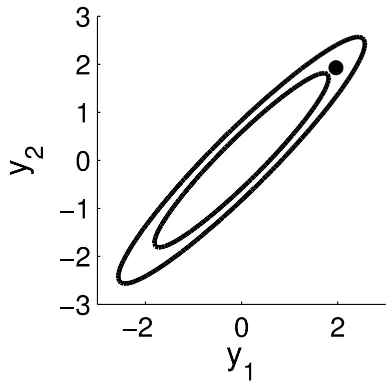
New visualisation



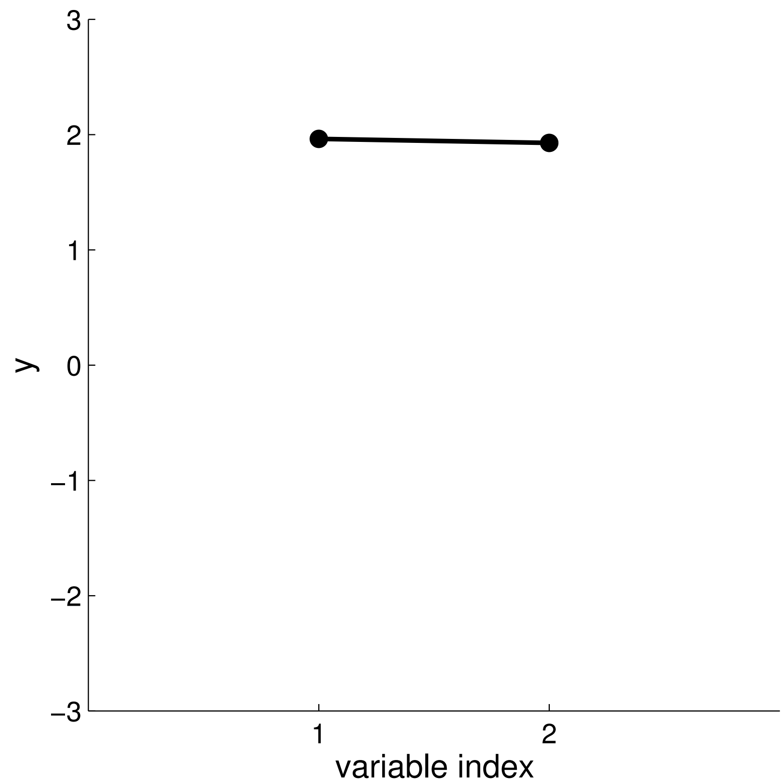
$$\Sigma = \begin{bmatrix} 1 & .9 \\ .9 & 1 \end{bmatrix}$$



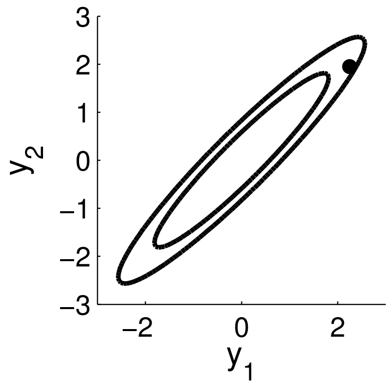
New visualisation



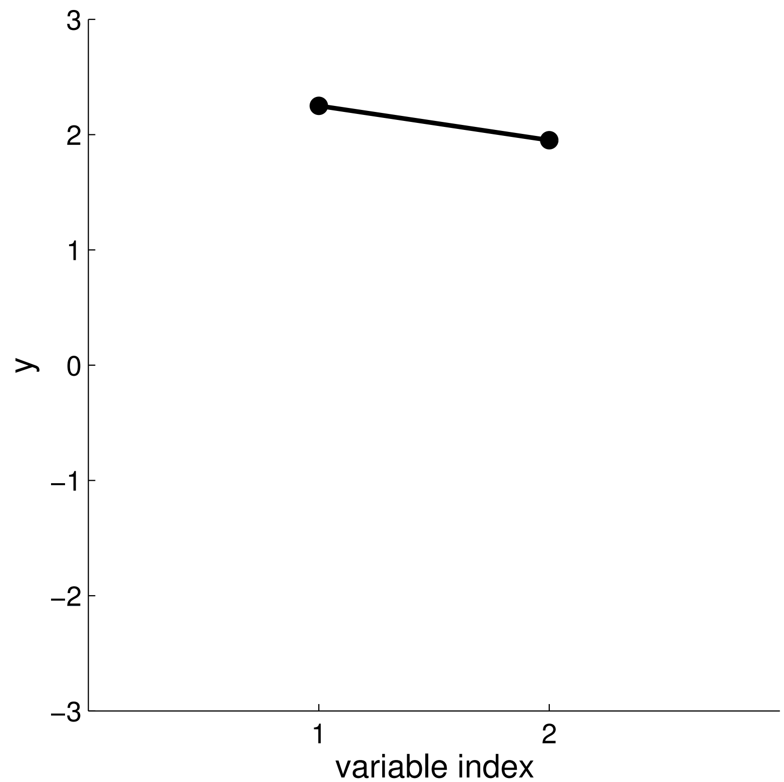
$$\Sigma = \begin{bmatrix} 1 & .9 \\ .9 & 1 \end{bmatrix}$$



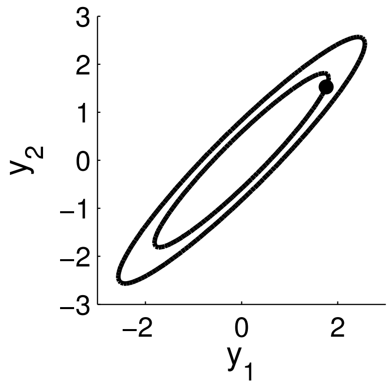
New visualisation



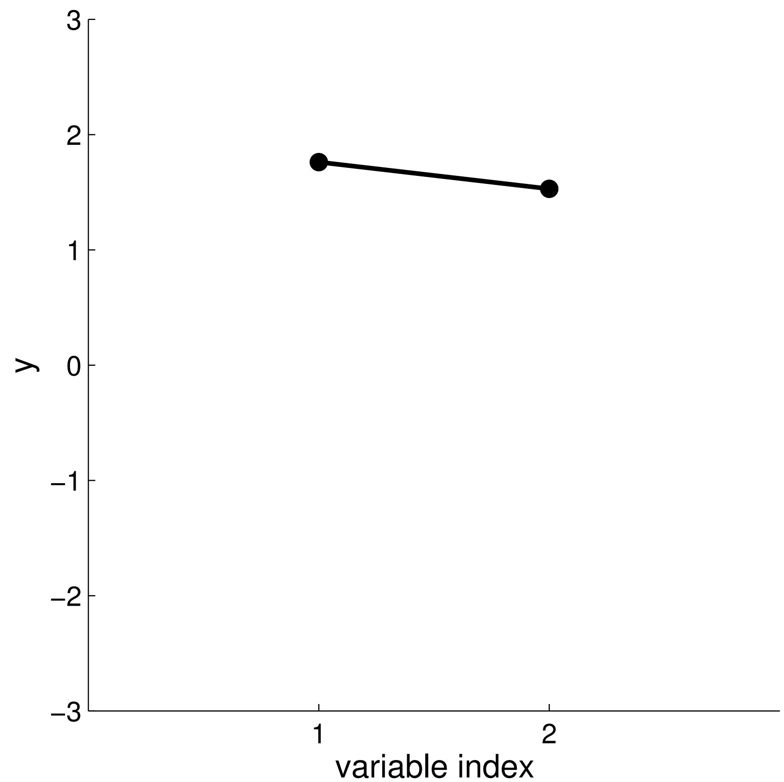
$$\Sigma = \begin{bmatrix} 1 & .9 \\ .9 & 1 \end{bmatrix}$$



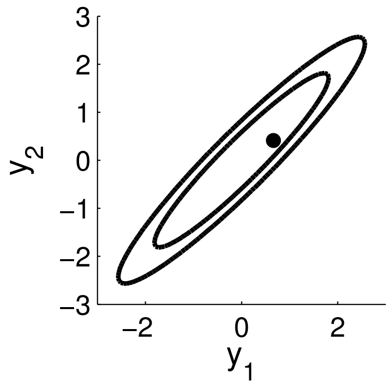
New visualisation



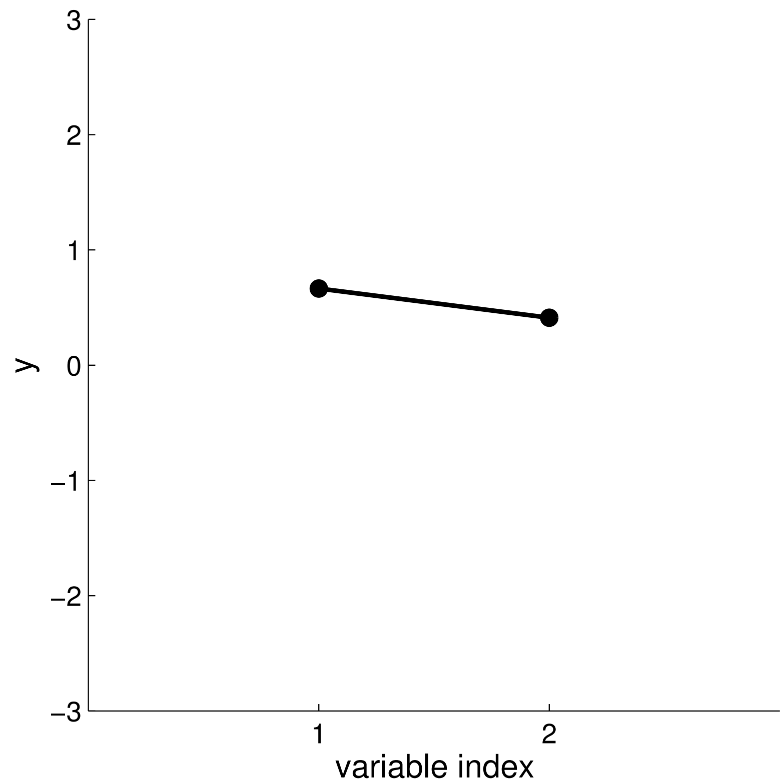
$$\Sigma = \begin{bmatrix} 1 & .9 \\ .9 & 1 \end{bmatrix}$$



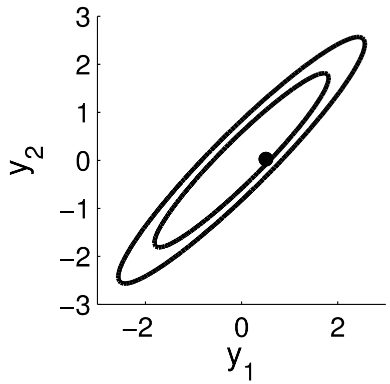
New visualisation



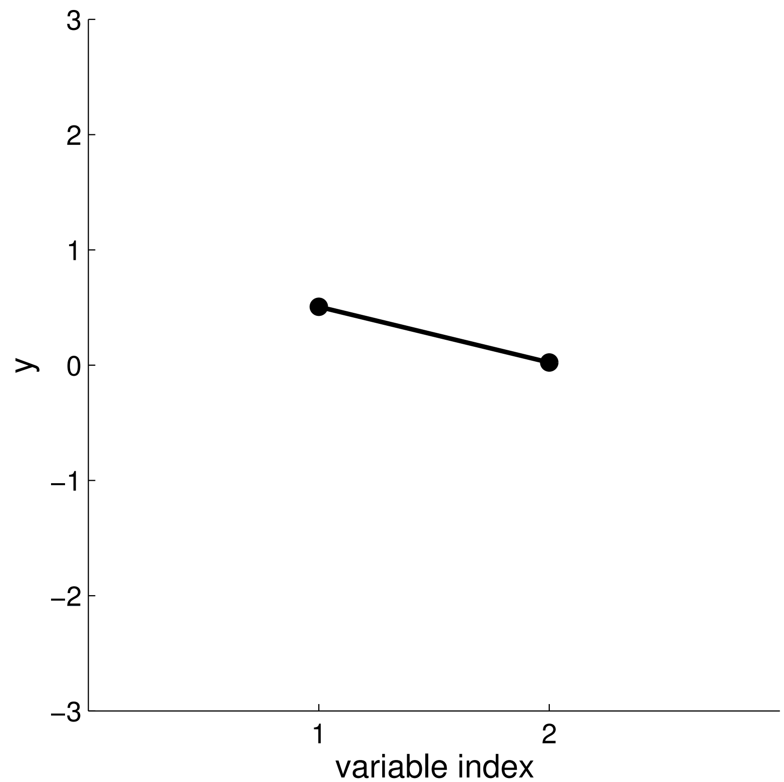
$$\Sigma = \begin{bmatrix} 1 & .9 \\ .9 & 1 \end{bmatrix}$$



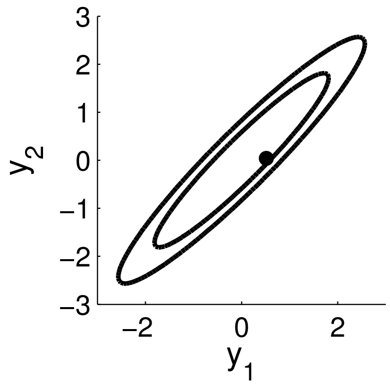
New visualisation



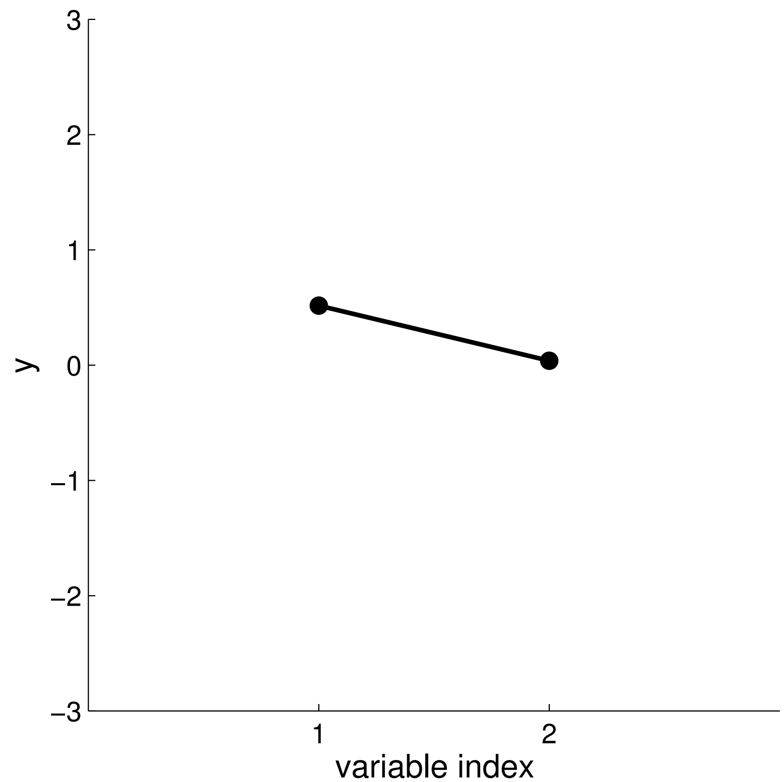
$$\Sigma = \begin{bmatrix} 1 & .9 \\ .9 & 1 \end{bmatrix}$$



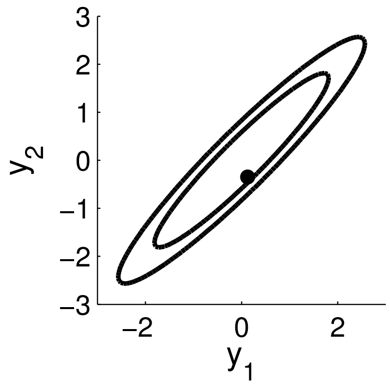
New visualisation



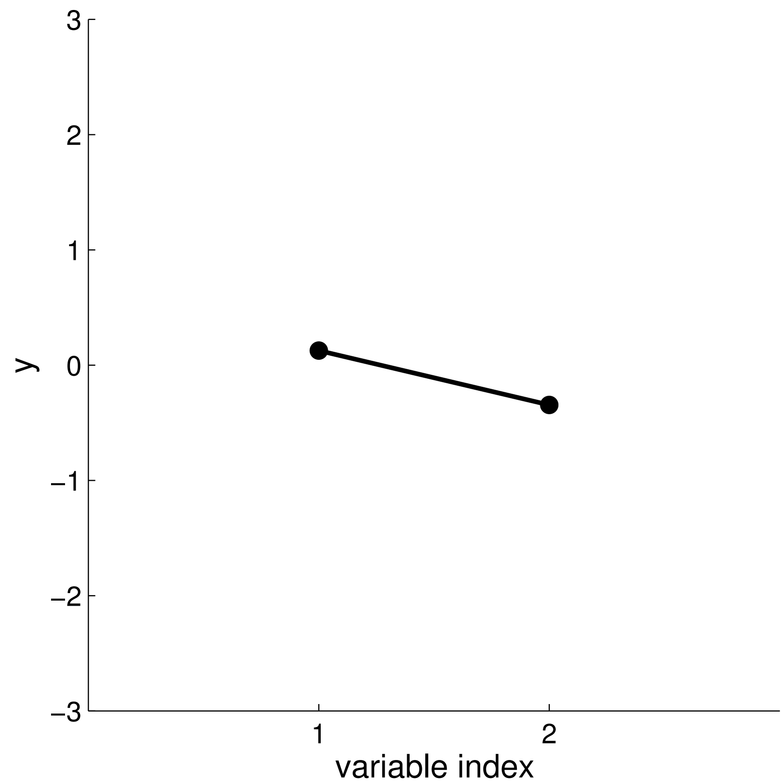
$$\Sigma = \begin{bmatrix} 1 & .9 \\ .9 & 1 \end{bmatrix}$$



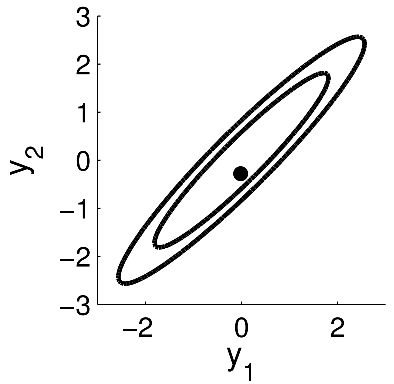
New visualisation



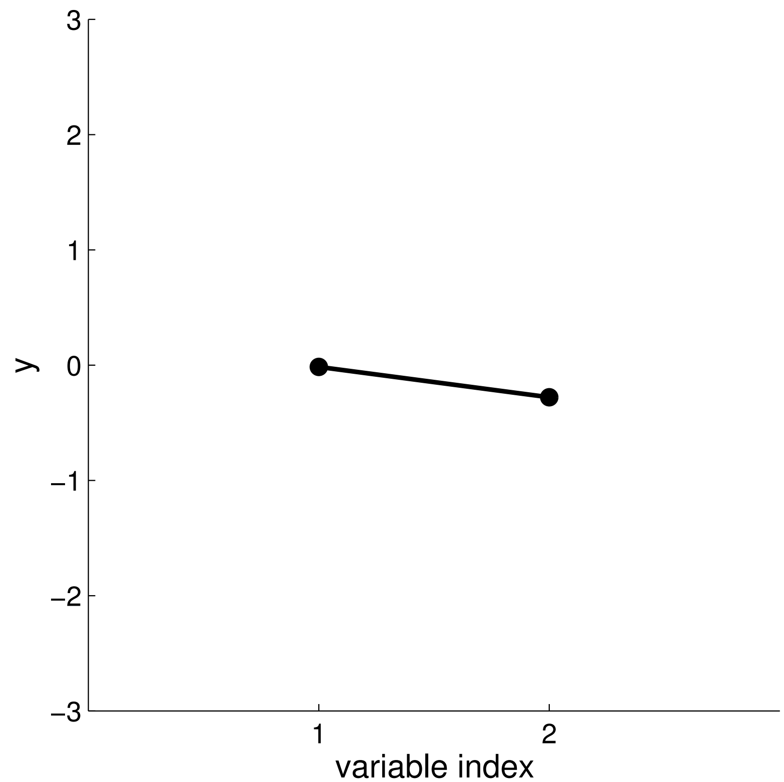
$$\Sigma = \begin{bmatrix} 1 & .9 \\ .9 & 1 \end{bmatrix}$$



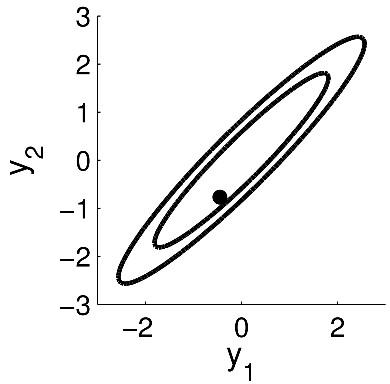
New visualisation



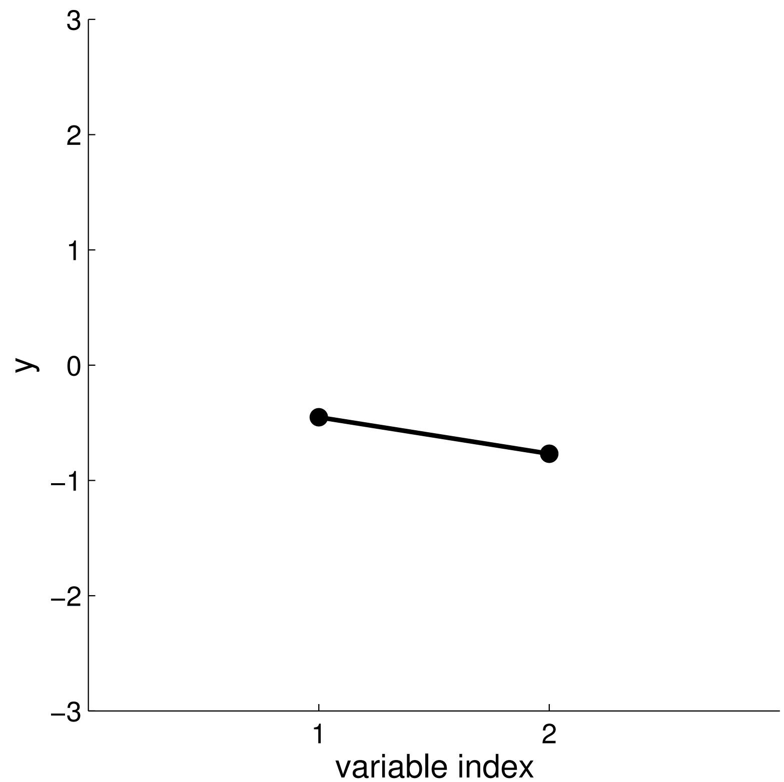
$$\Sigma = \begin{bmatrix} 1 & .9 \\ .9 & 1 \end{bmatrix}$$



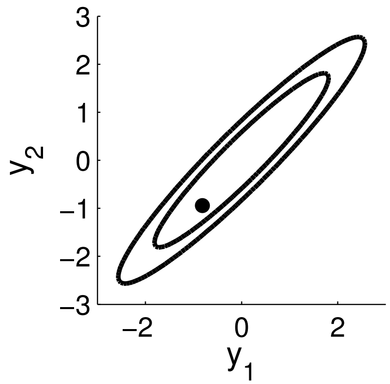
New visualisation



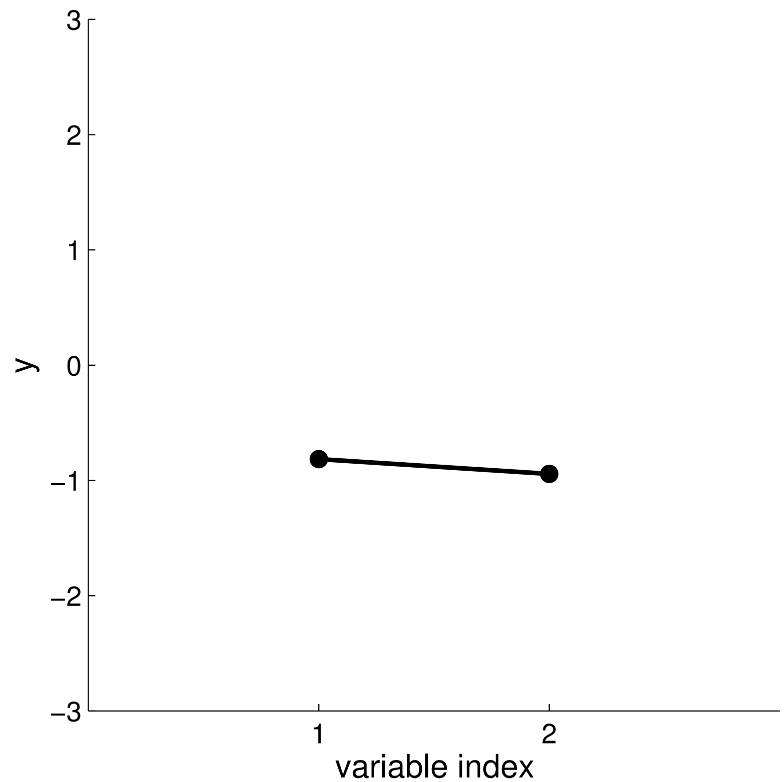
$$\Sigma = \begin{bmatrix} 1 & .9 \\ .9 & 1 \end{bmatrix}$$



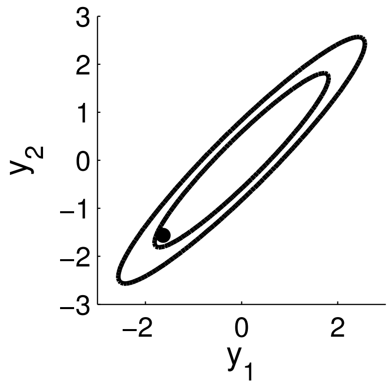
New visualisation



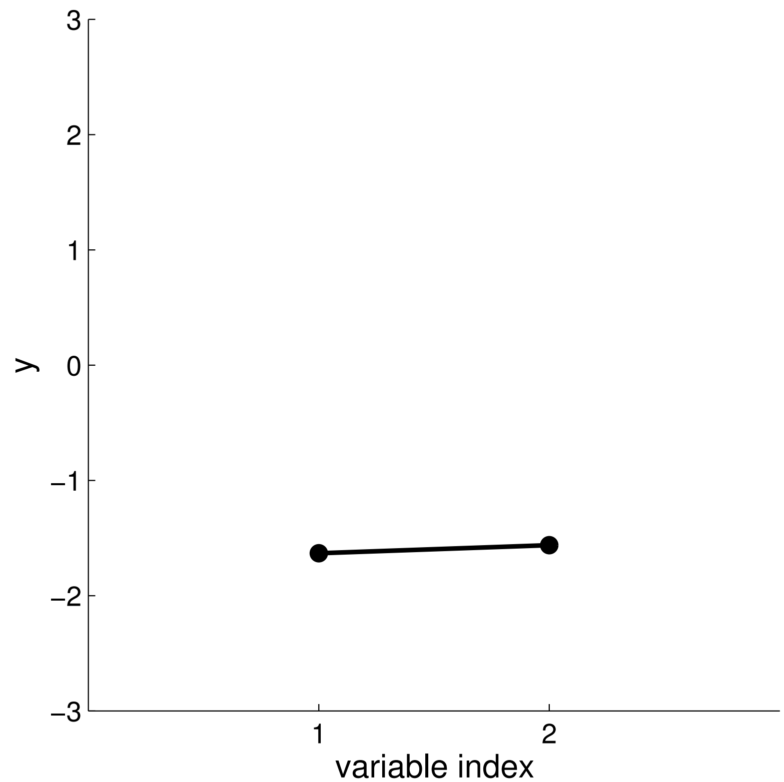
$$\Sigma = \begin{bmatrix} 1 & .9 \\ .9 & 1 \end{bmatrix}$$



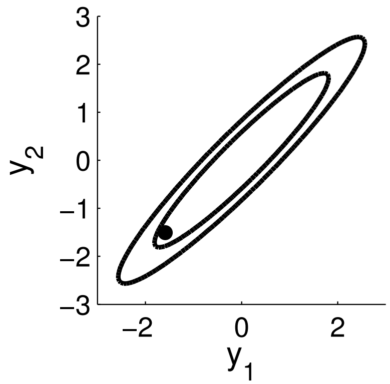
New visualisation



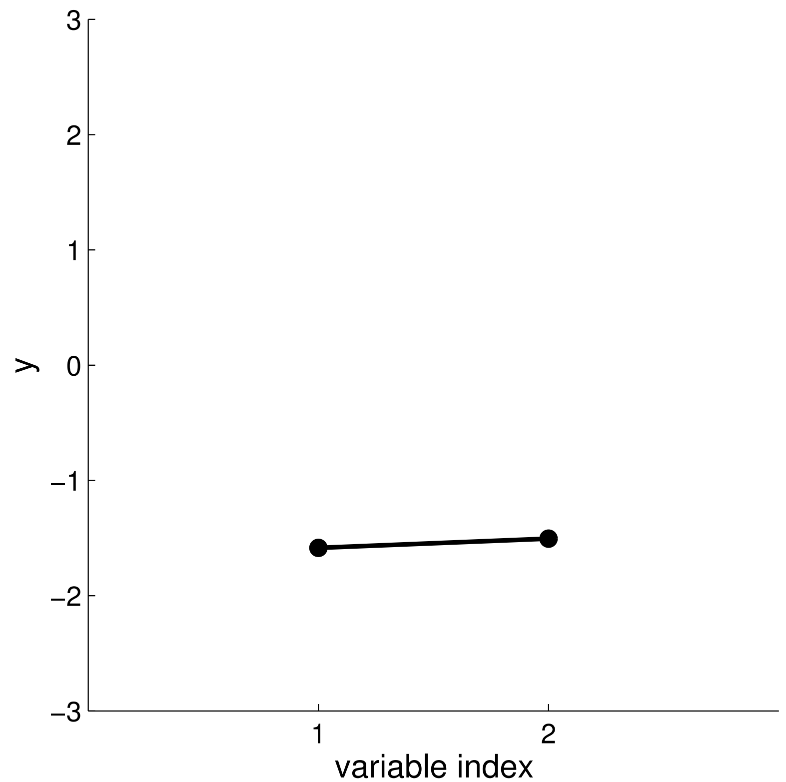
$$\Sigma = \begin{bmatrix} 1 & .9 \\ .9 & 1 \end{bmatrix}$$



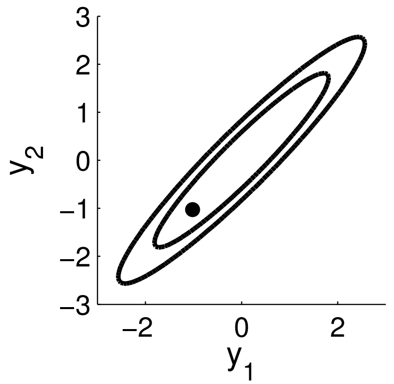
New visualisation



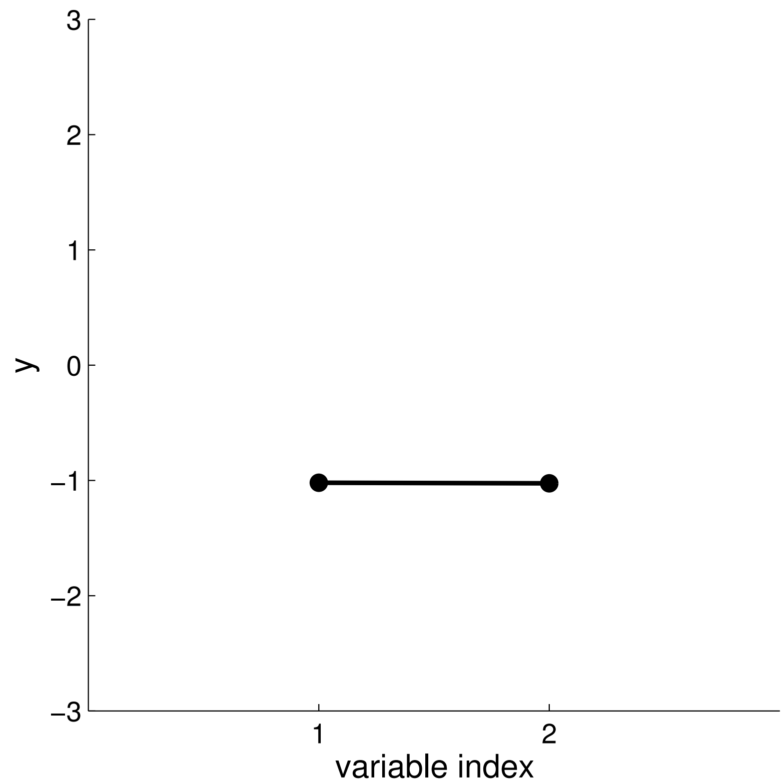
$$\Sigma = \begin{bmatrix} 1 & .9 \\ .9 & 1 \end{bmatrix}$$



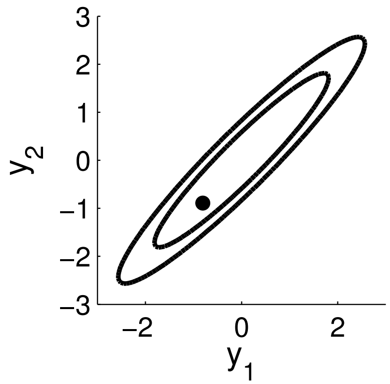
New visualisation



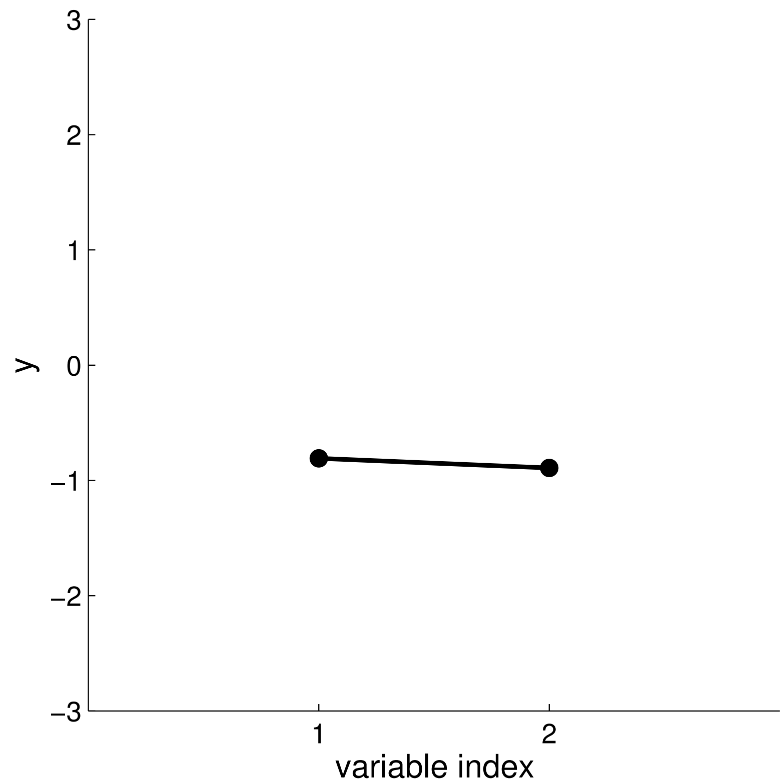
$$\Sigma = \begin{bmatrix} 1 & .9 \\ .9 & 1 \end{bmatrix}$$



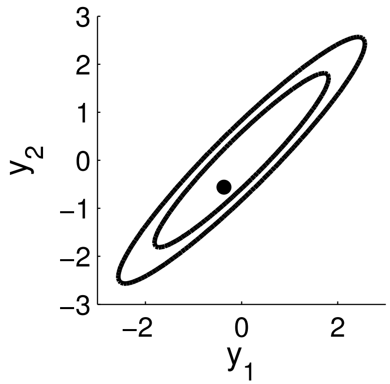
New visualisation



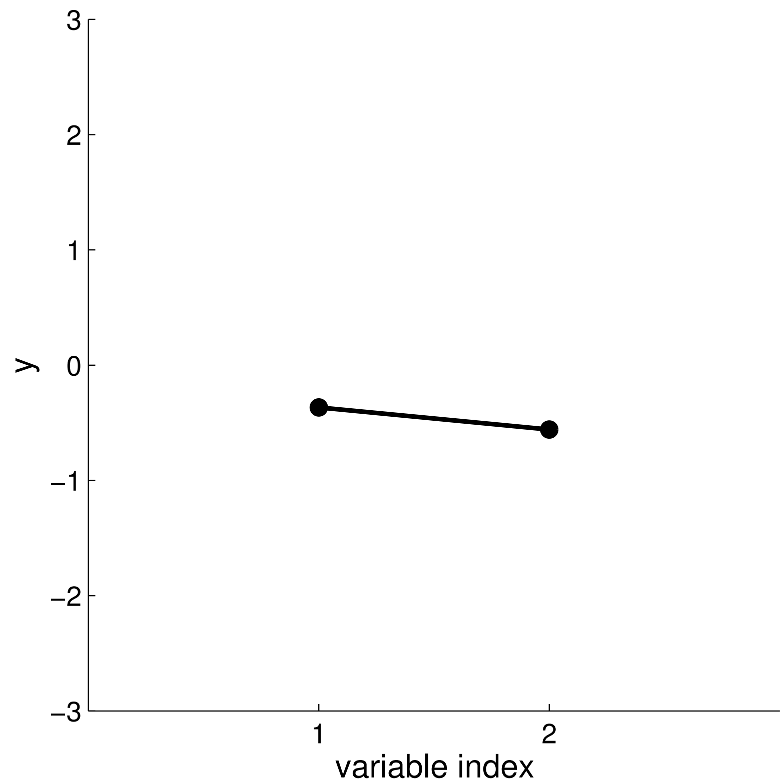
$$\Sigma = \begin{bmatrix} 1 & .9 \\ .9 & 1 \end{bmatrix}$$



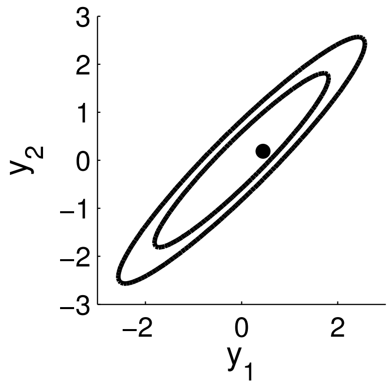
New visualisation



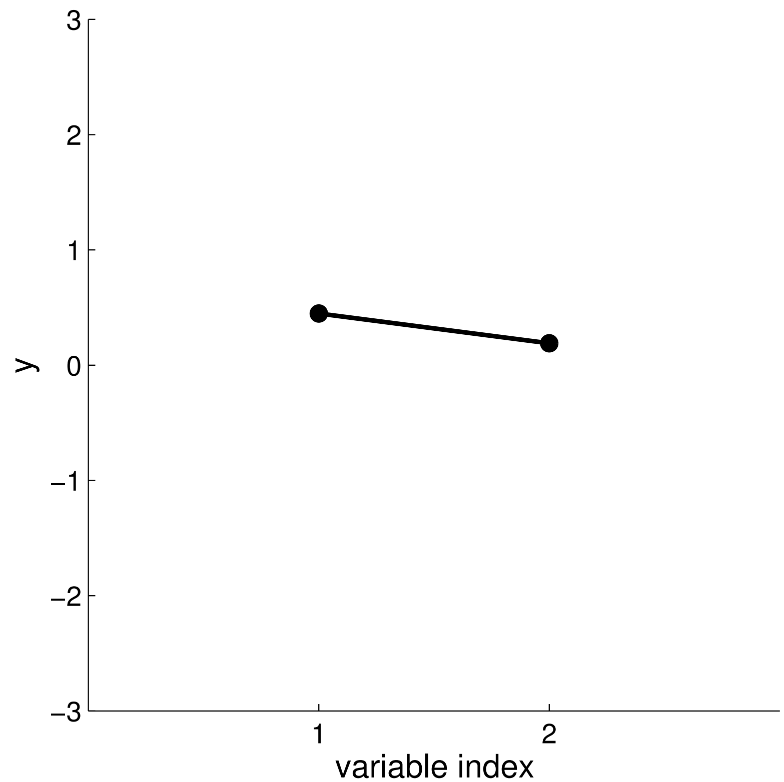
$$\Sigma = \begin{bmatrix} 1 & .9 \\ .9 & 1 \end{bmatrix}$$



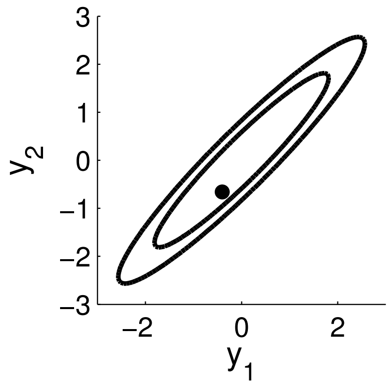
New visualisation



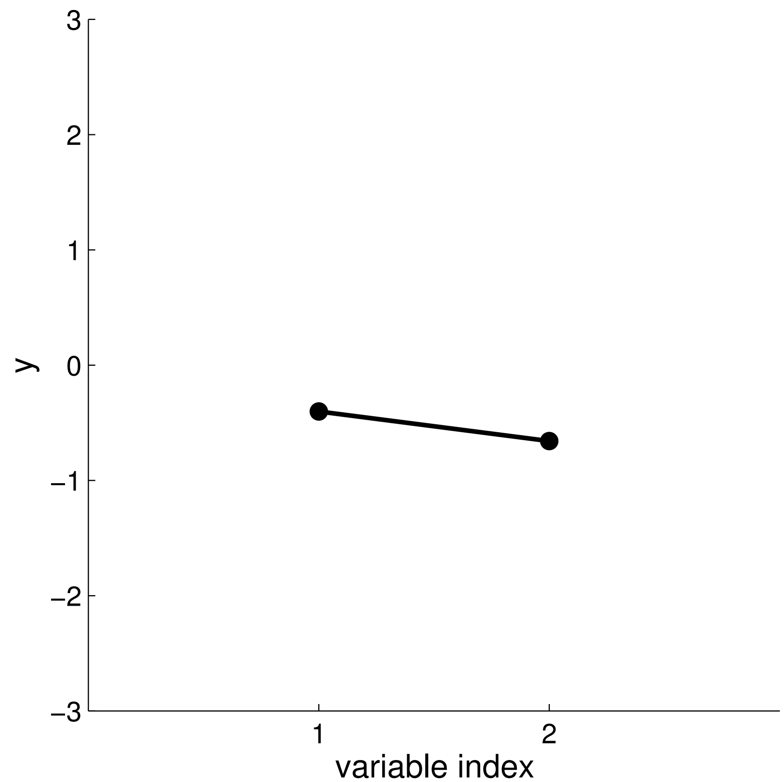
$$\Sigma = \begin{bmatrix} 1 & .9 \\ .9 & 1 \end{bmatrix}$$



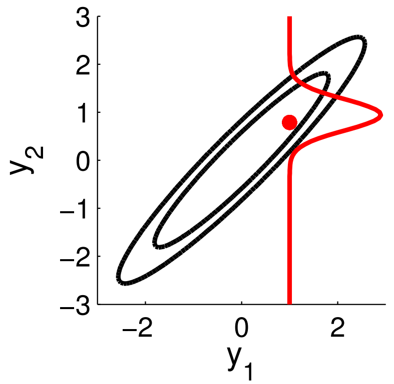
New visualisation



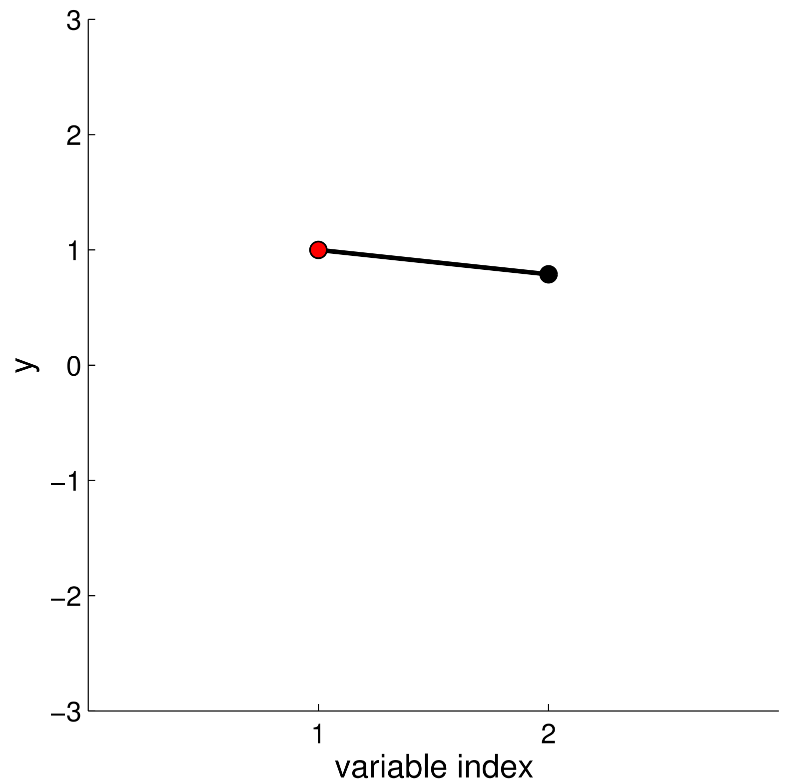
$$\Sigma = \begin{bmatrix} 1 & .9 \\ .9 & 1 \end{bmatrix}$$



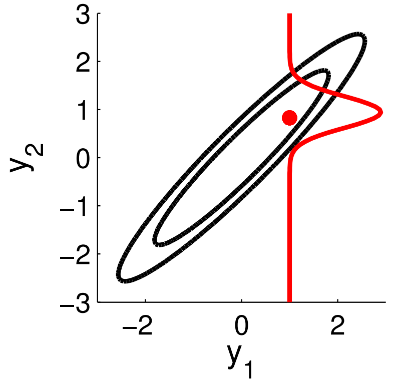
New visualisation



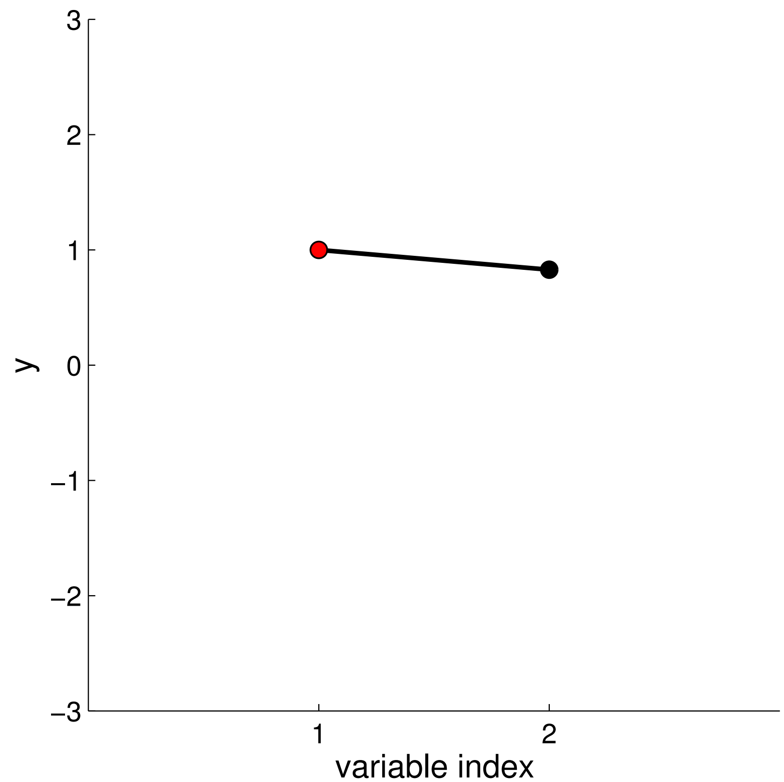
$$\Sigma = \begin{bmatrix} 1 & .9 \\ .9 & 1 \end{bmatrix}$$



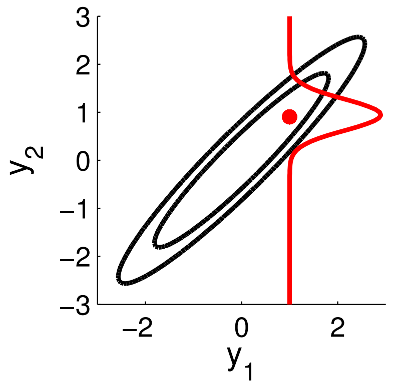
New visualisation



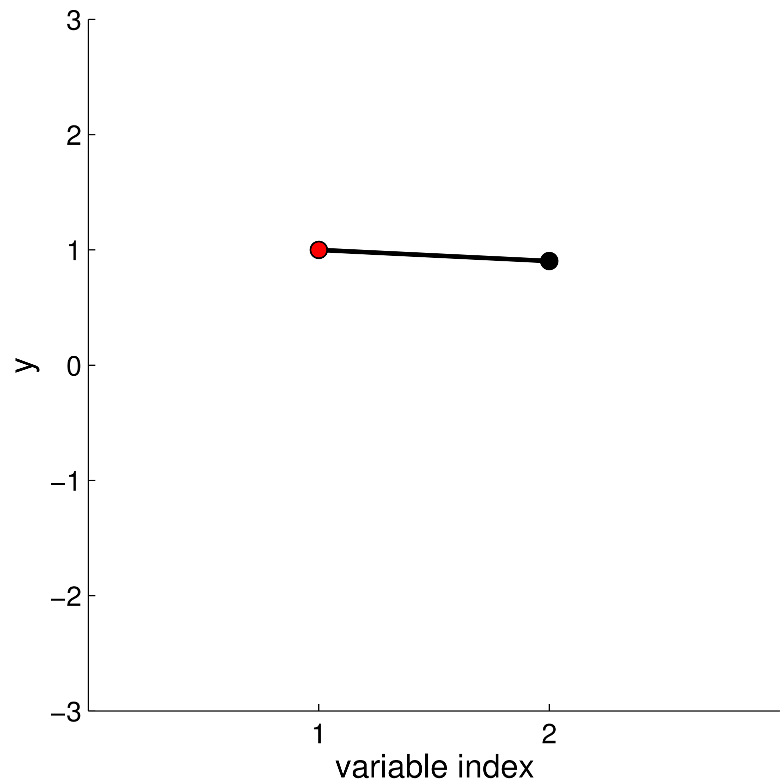
$$\Sigma = \begin{bmatrix} 1 & .9 \\ .9 & 1 \end{bmatrix}$$



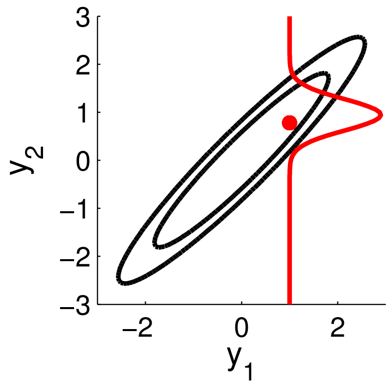
New visualisation



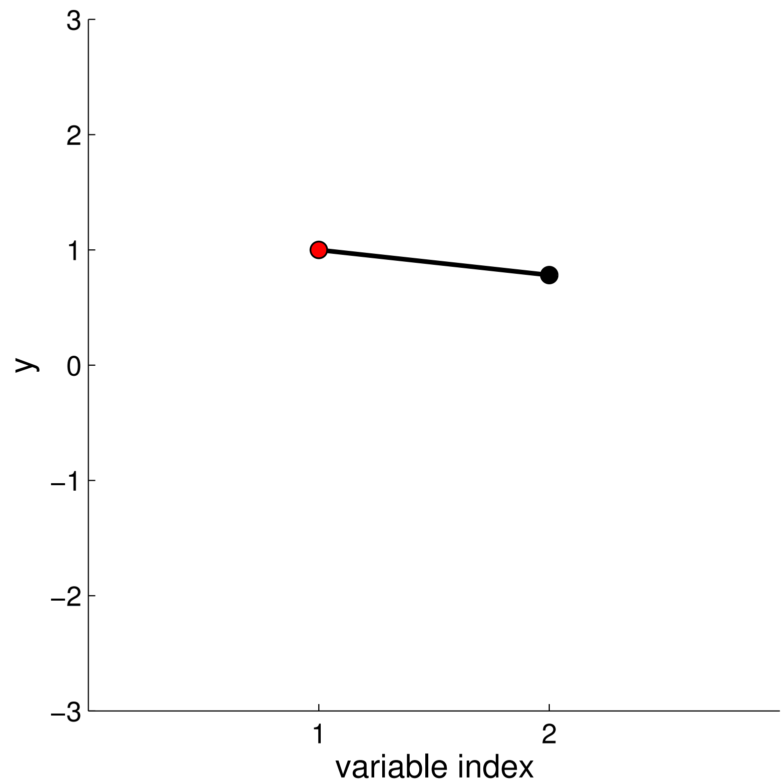
$$\Sigma = \begin{bmatrix} 1 & .9 \\ .9 & 1 \end{bmatrix}$$



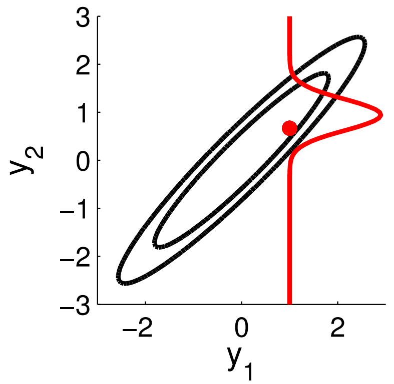
New visualisation



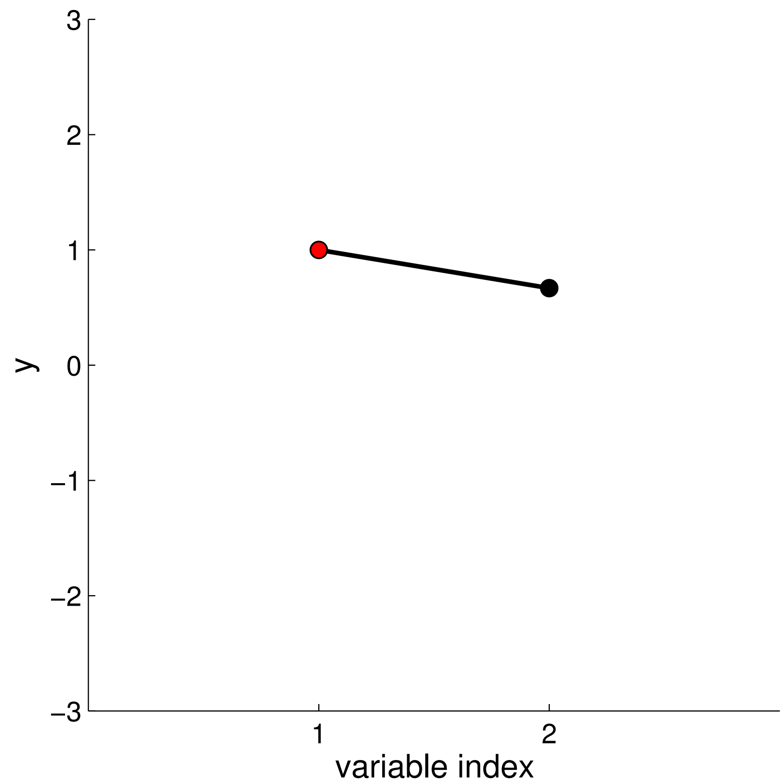
$$\Sigma = \begin{bmatrix} 1 & .9 \\ .9 & 1 \end{bmatrix}$$



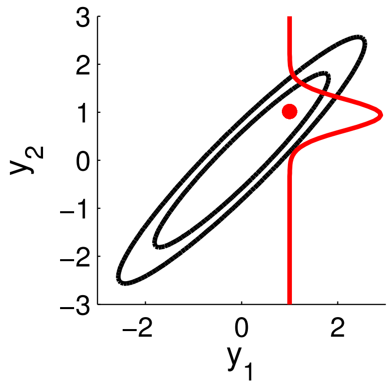
New visualisation



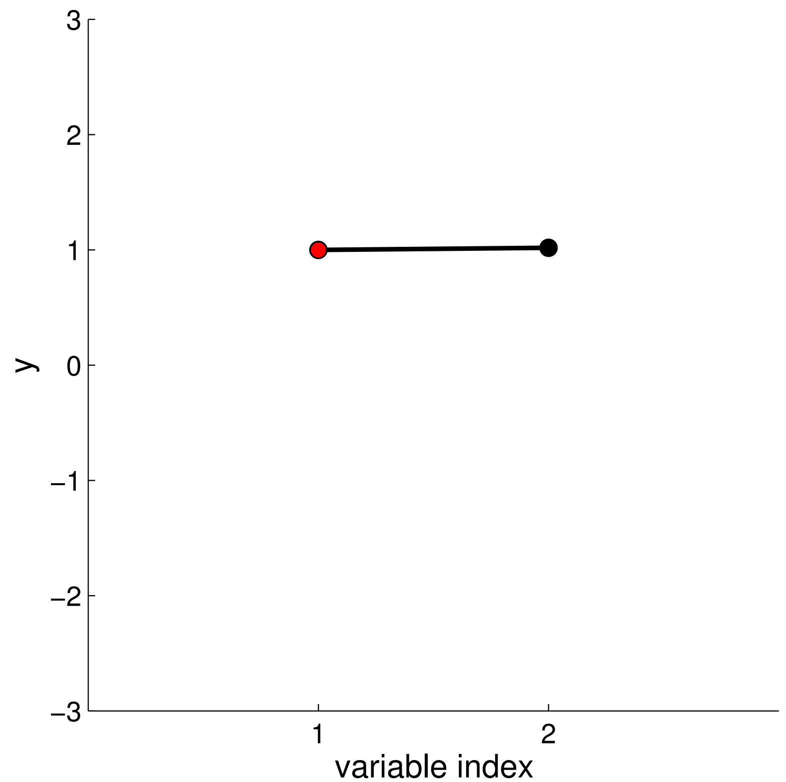
$$\Sigma = \begin{bmatrix} 1 & .9 \\ .9 & 1 \end{bmatrix}$$



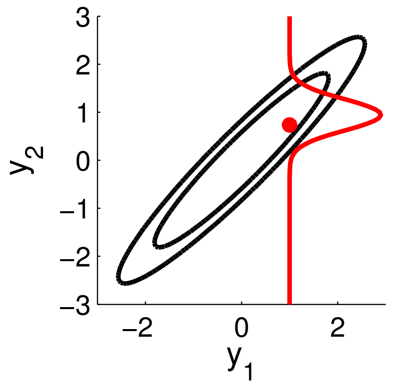
New visualisation



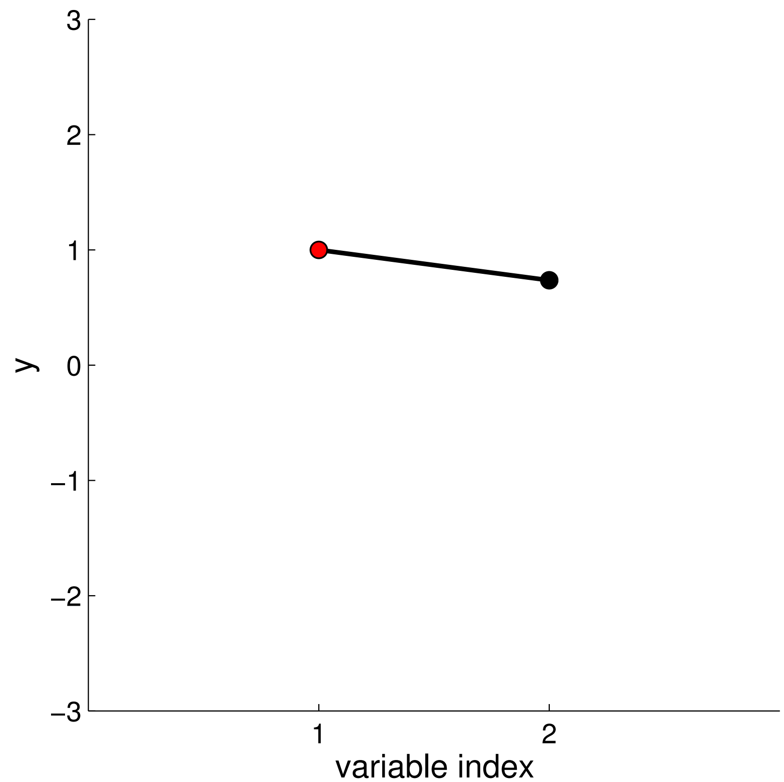
$$\Sigma = \begin{bmatrix} 1 & .9 \\ .9 & 1 \end{bmatrix}$$



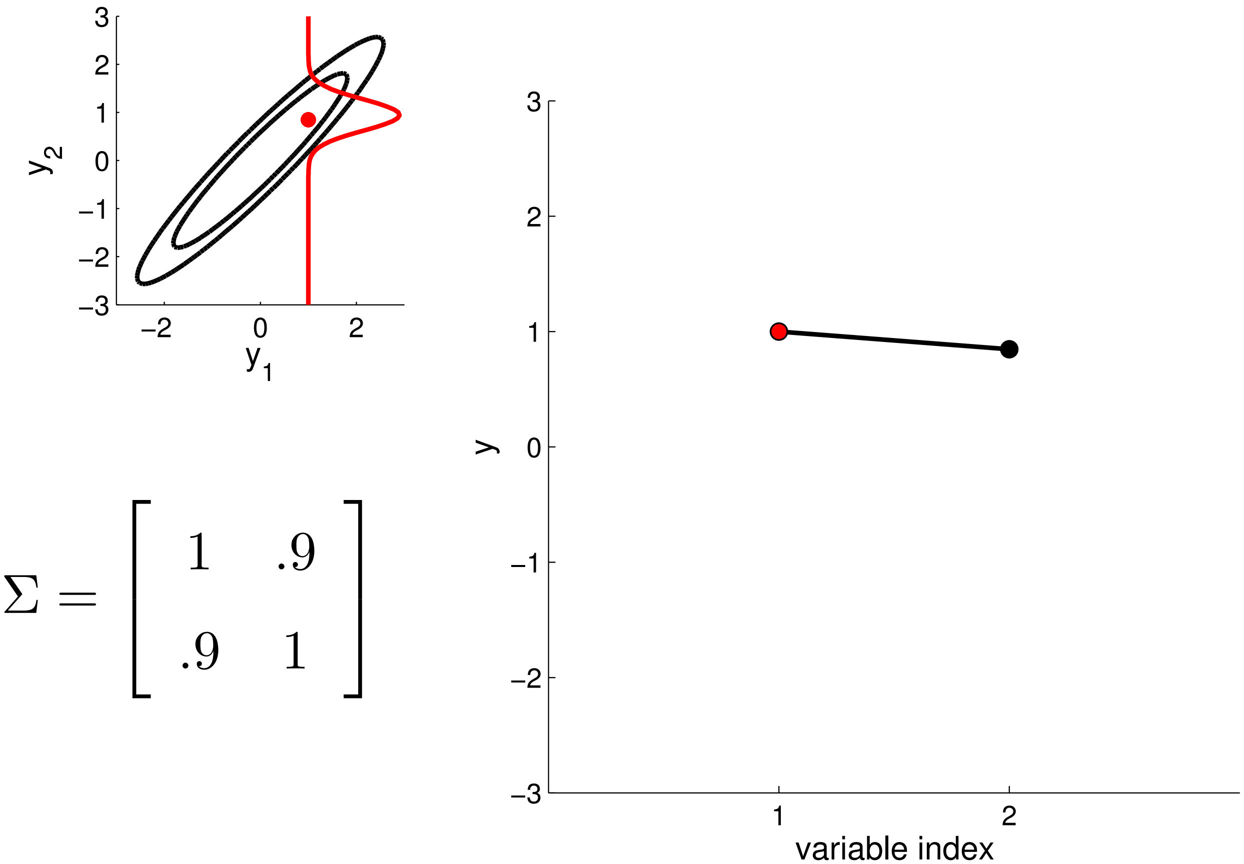
New visualisation



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

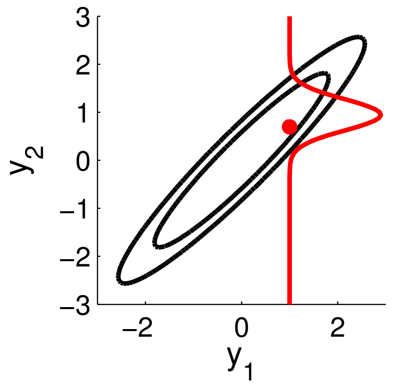


New visualisation

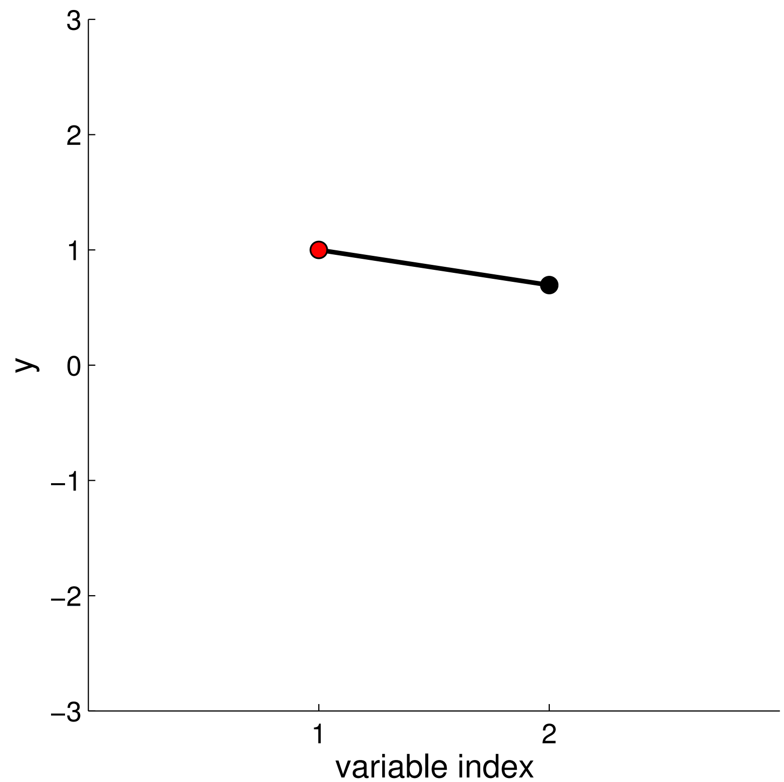


$$\Sigma = \begin{bmatrix} 1 & .9 \\ .9 & 1 \end{bmatrix}$$

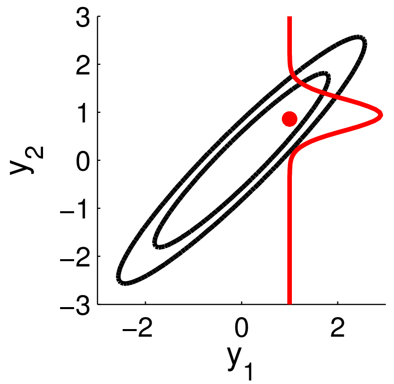
New visualisation



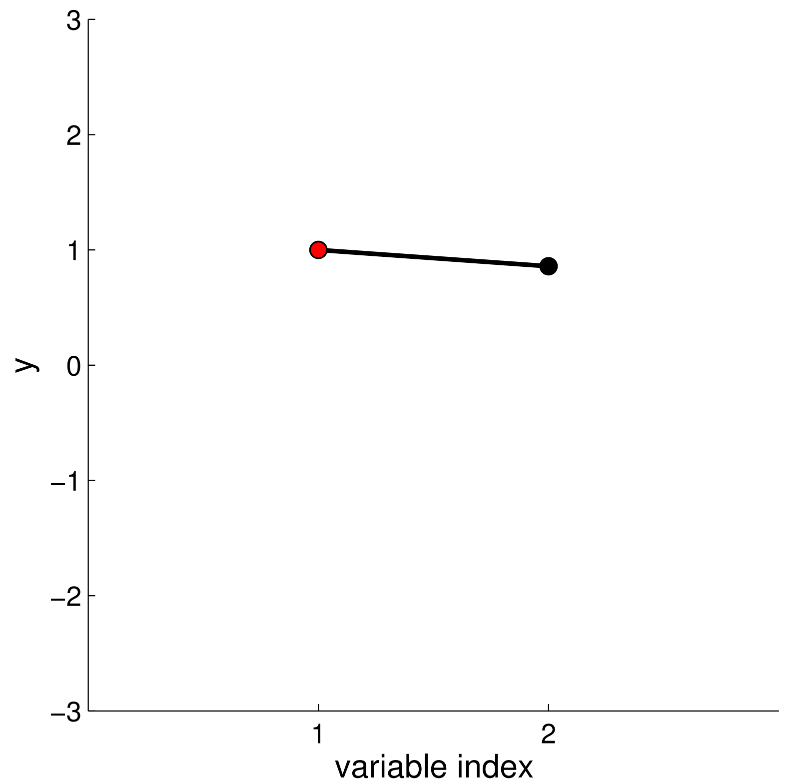
$$\Sigma = \begin{bmatrix} 1 & .9 \\ .9 & 1 \end{bmatrix}$$



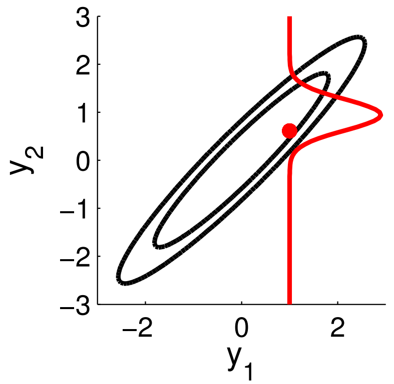
New visualisation



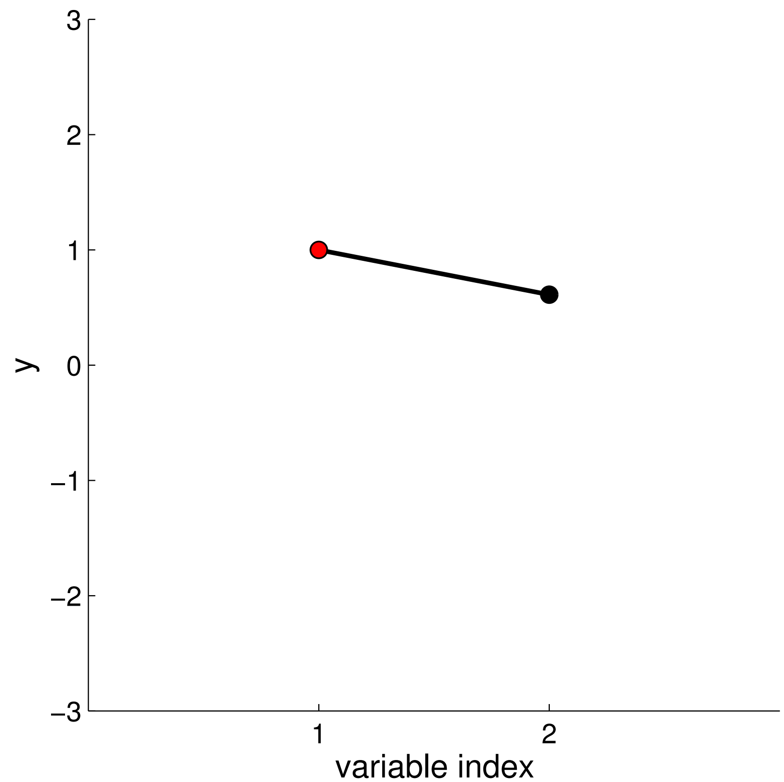
$$\Sigma = \begin{bmatrix} 1 & .9 \\ .9 & 1 \end{bmatrix}$$



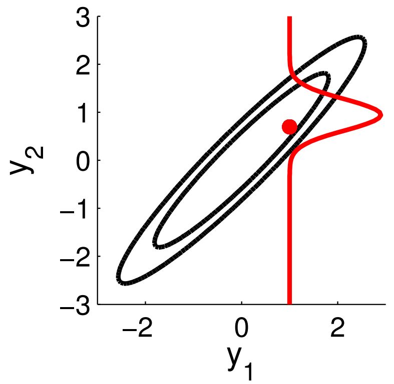
New visualisation



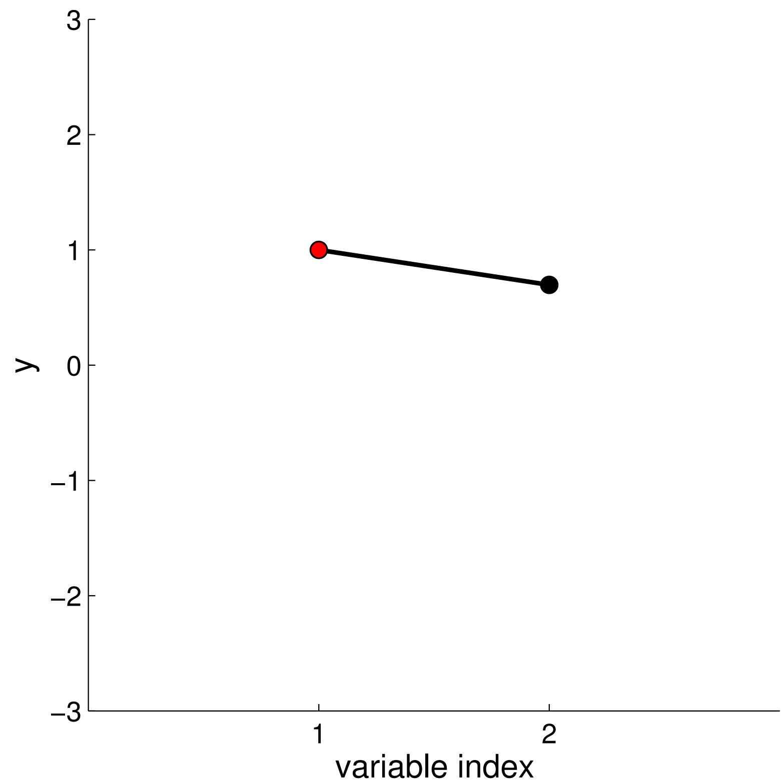
$$\Sigma = \begin{bmatrix} 1 & .9 \\ .9 & 1 \end{bmatrix}$$



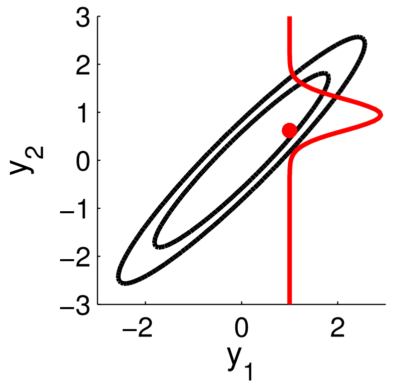
New visualisation



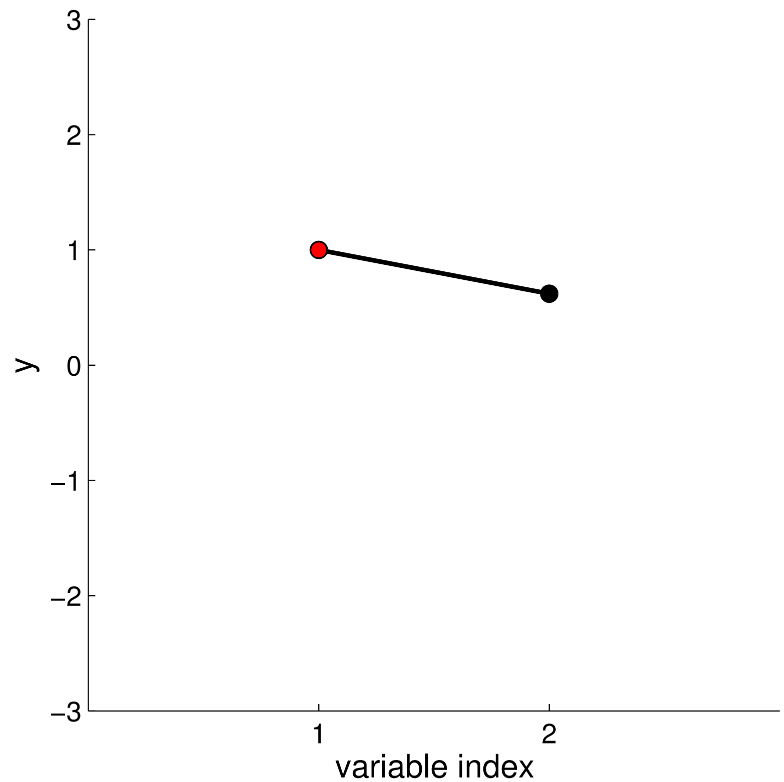
$$\Sigma = \begin{bmatrix} 1 & .9 \\ .9 & 1 \end{bmatrix}$$



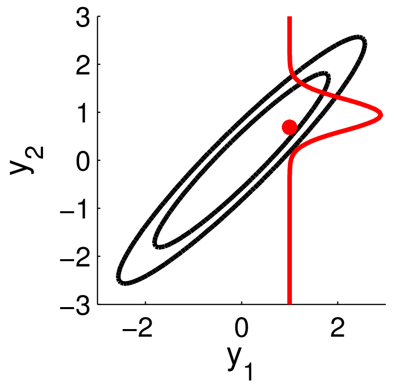
New visualisation



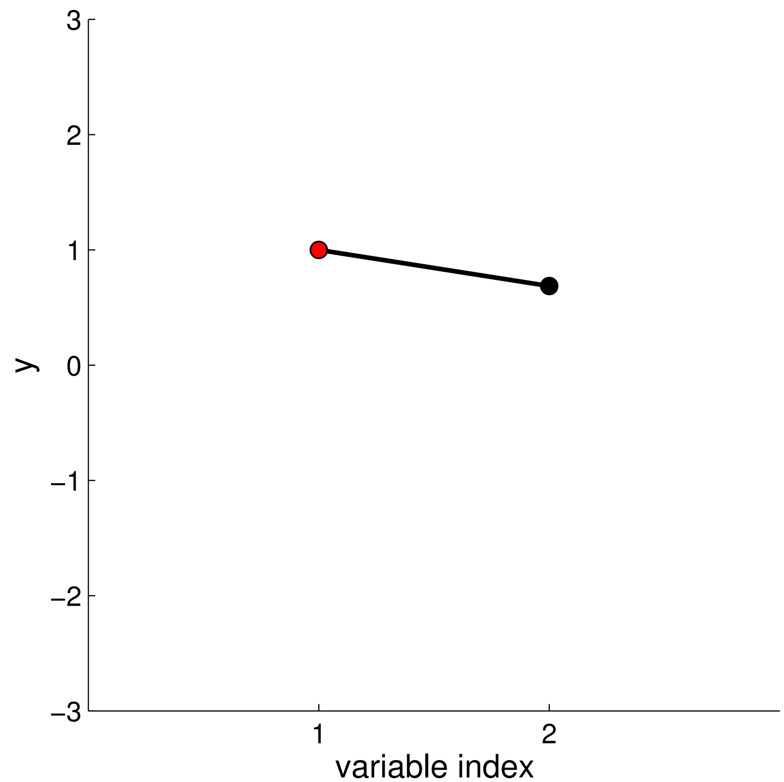
$$\Sigma = \begin{bmatrix} 1 & .9 \\ .9 & 1 \end{bmatrix}$$



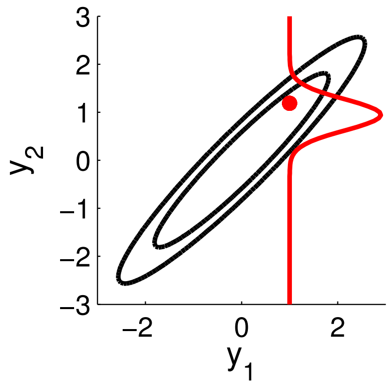
New visualisation



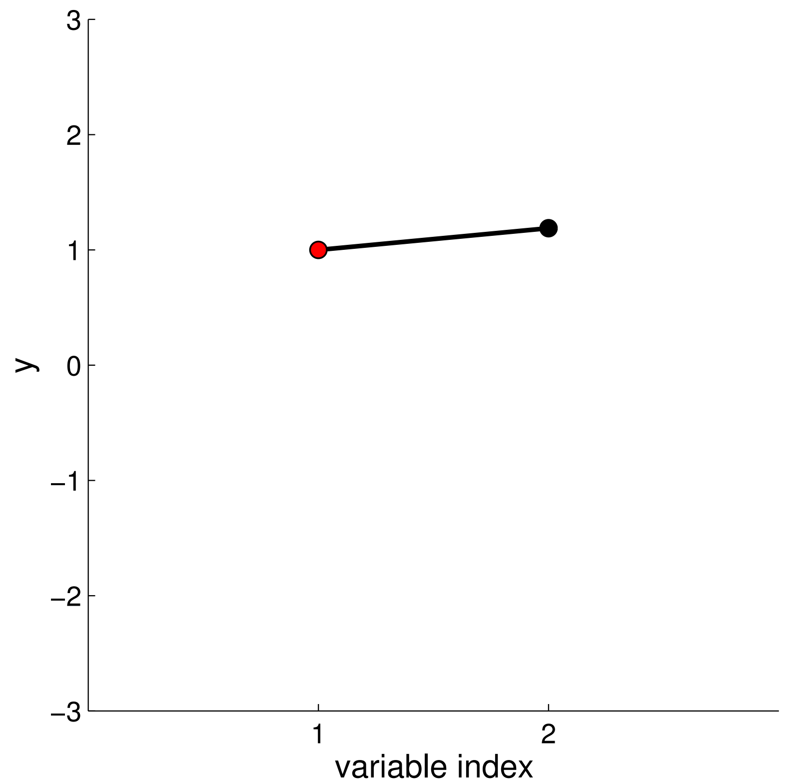
$$\Sigma = \begin{bmatrix} 1 & .9 \\ .9 & 1 \end{bmatrix}$$



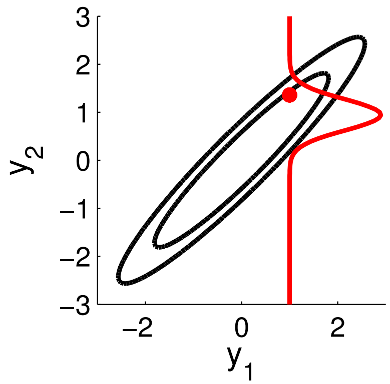
New visualisation



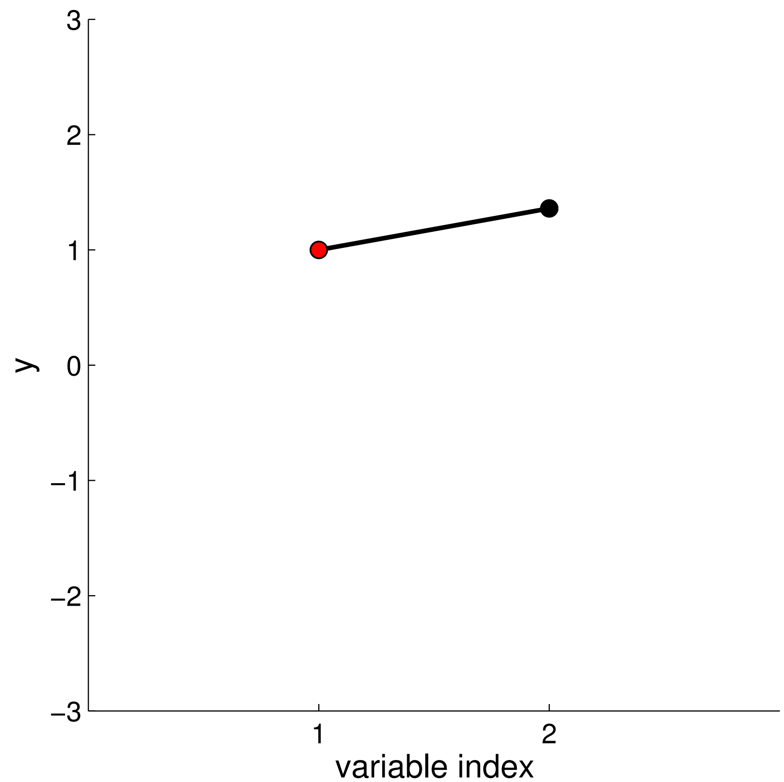
$$\Sigma = \begin{bmatrix} 1 & .9 \\ .9 & 1 \end{bmatrix}$$



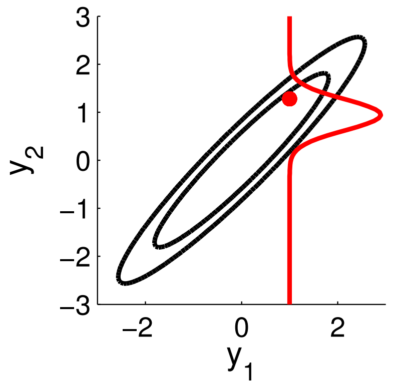
New visualisation



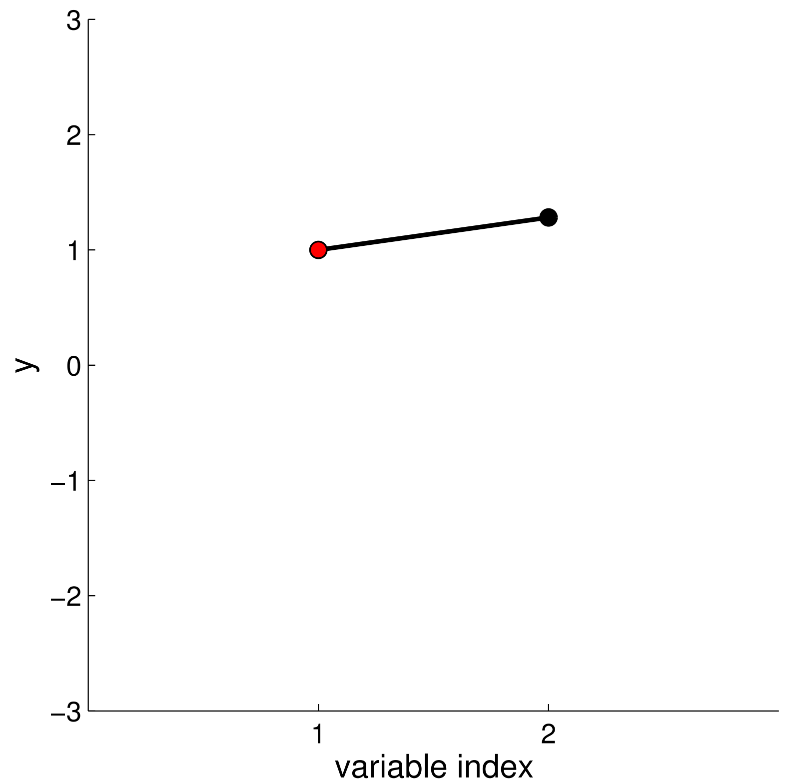
$$\Sigma = \begin{bmatrix} 1 & .9 \\ .9 & 1 \end{bmatrix}$$



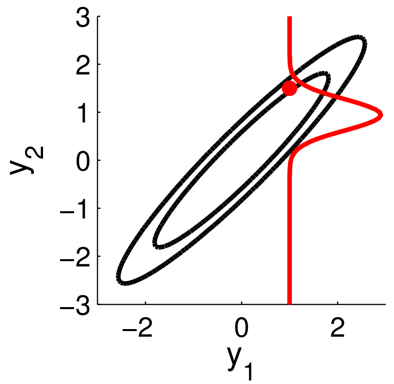
New visualisation



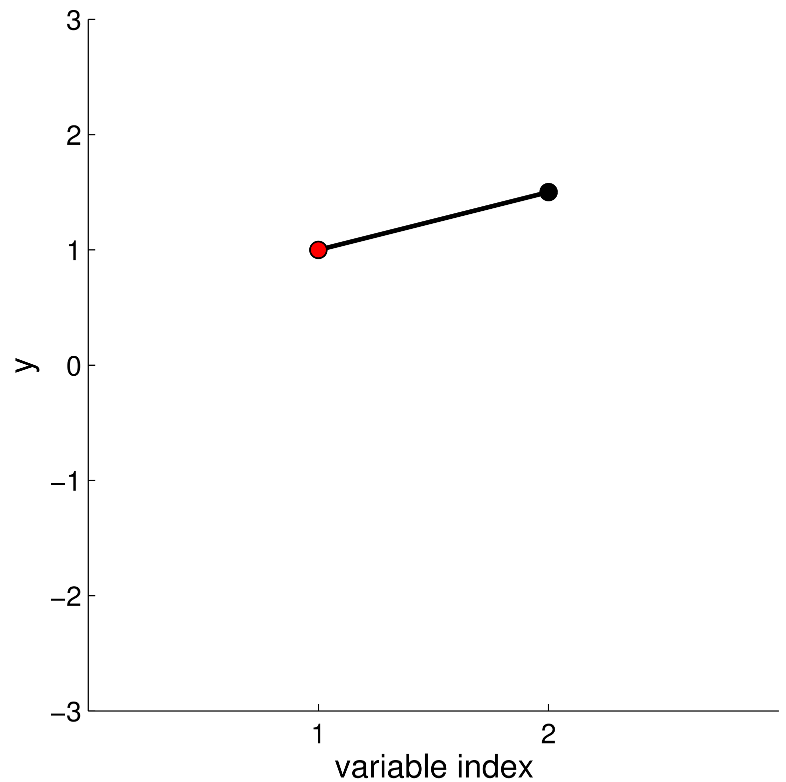
$$\Sigma = \begin{bmatrix} 1 & .9 \\ .9 & 1 \end{bmatrix}$$



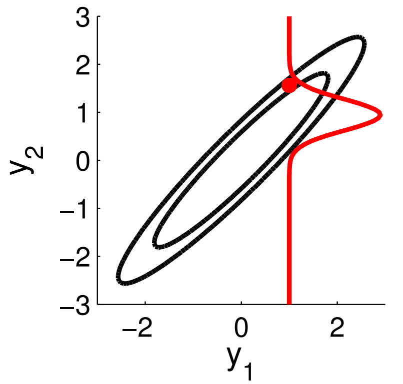
New visualisation



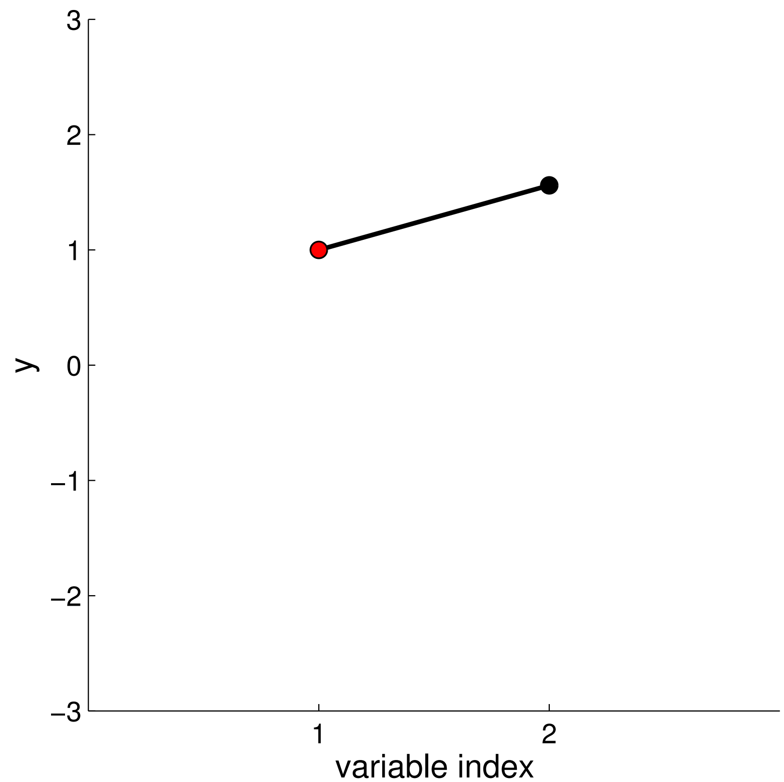
$$\Sigma = \begin{bmatrix} 1 & .9 \\ .9 & 1 \end{bmatrix}$$



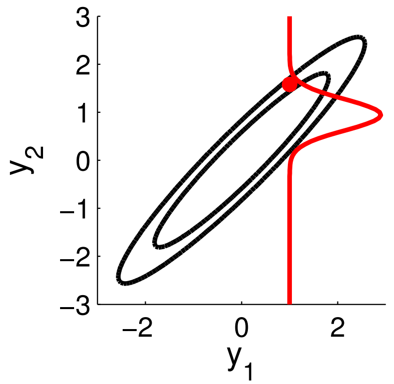
New visualisation



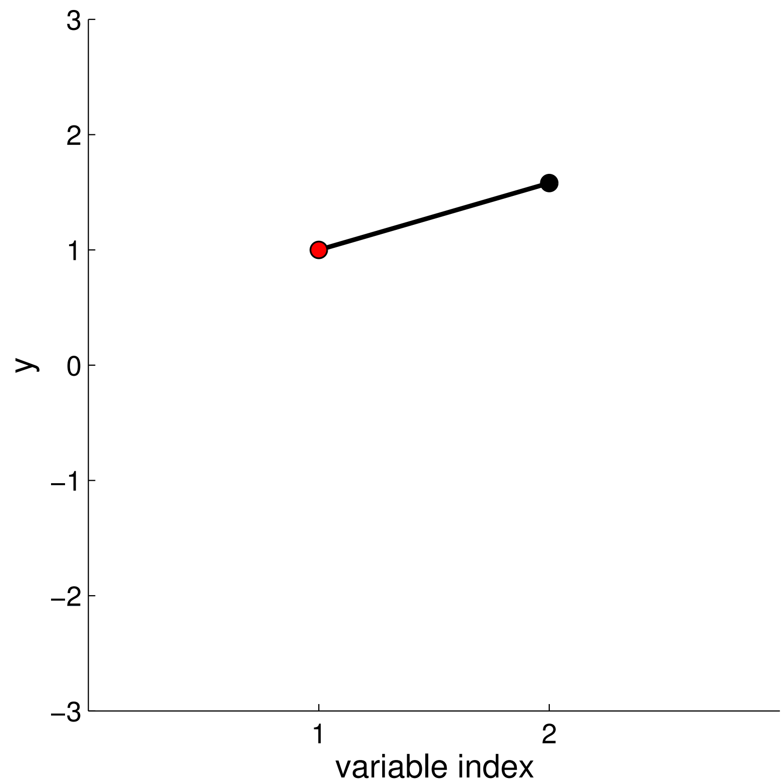
$$\Sigma = \begin{bmatrix} 1 & .9 \\ .9 & 1 \end{bmatrix}$$



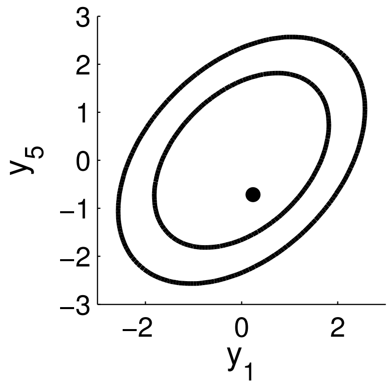
New visualisation

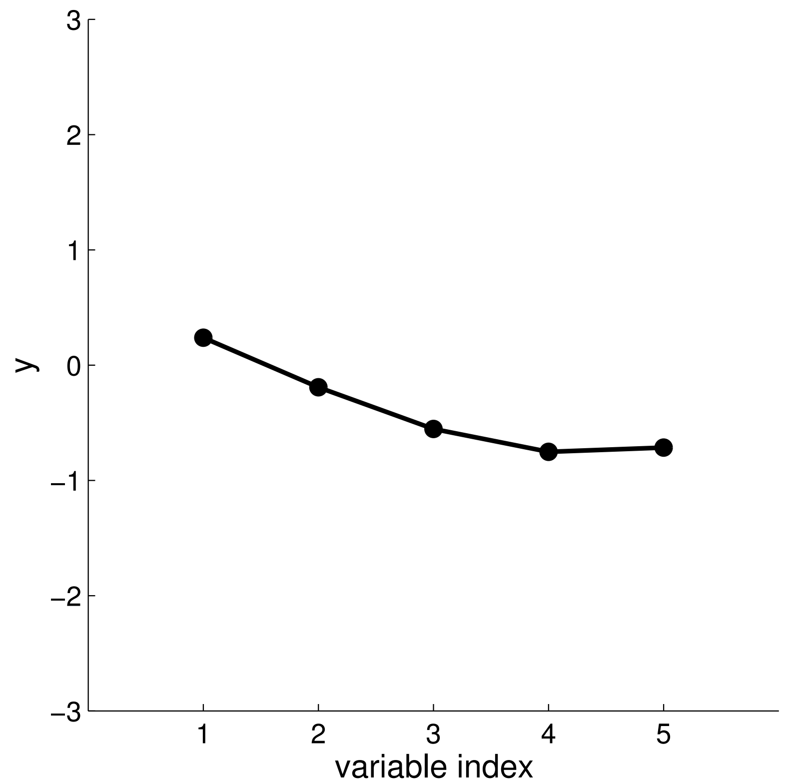


$$\Sigma = \begin{bmatrix} 1 & .9 \\ .9 & 1 \end{bmatrix}$$



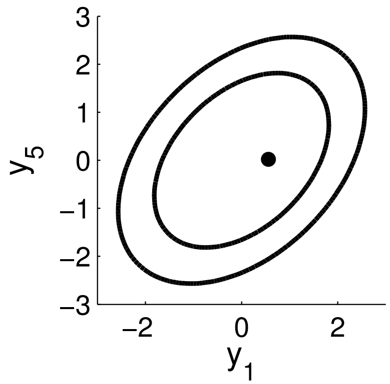
New visualisation

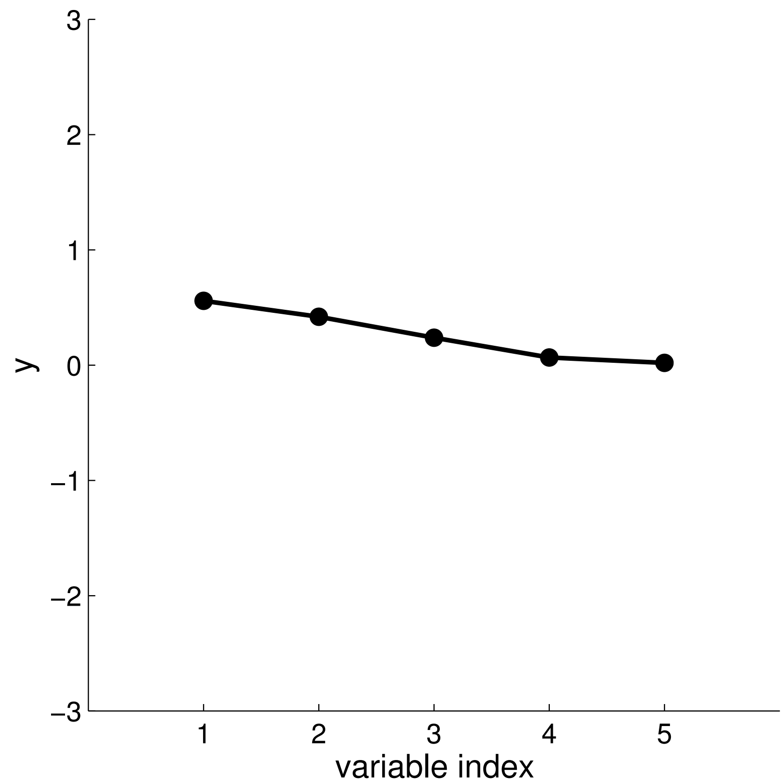




$$\Sigma = \begin{bmatrix} 1 & .9 & .8 & .6 & .4 \\ .9 & 1 & .9 & .8 & .6 \\ .8 & .9 & 1 & .9 & .8 \\ .6 & .8 & .9 & 1 & .9 \\ .4 & .6 & .8 & .9 & 1 \end{bmatrix}$$

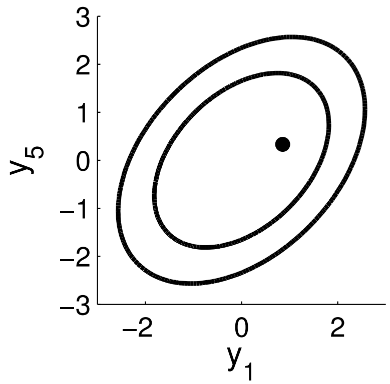
New visualisation



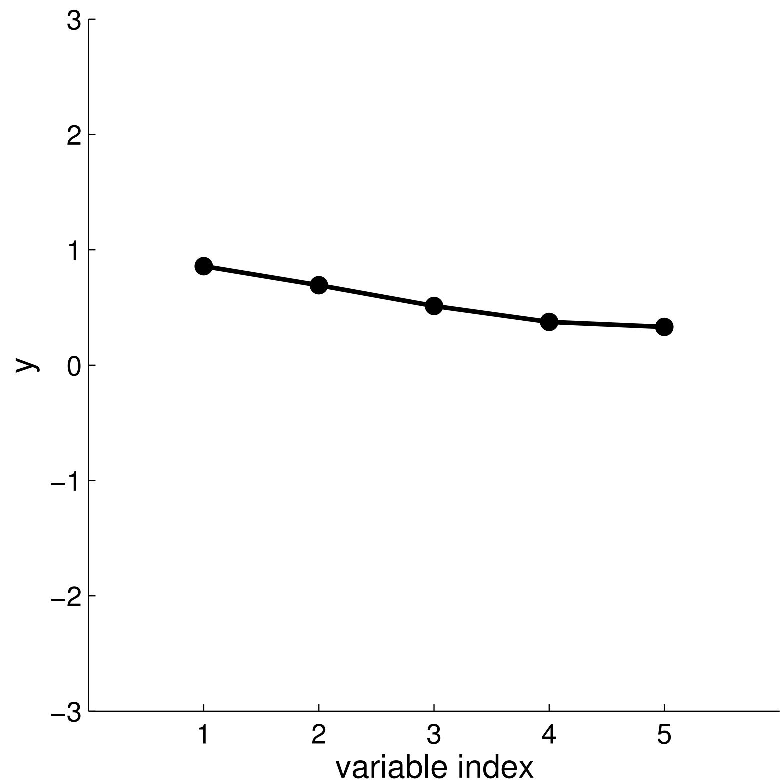


$$\Sigma = \begin{bmatrix} 1 & .9 & .8 & .6 & .4 \\ .9 & 1 & .9 & .8 & .6 \\ .8 & .9 & 1 & .9 & .8 \\ .6 & .8 & .9 & 1 & .9 \\ .4 & .6 & .8 & .9 & 1 \end{bmatrix}$$

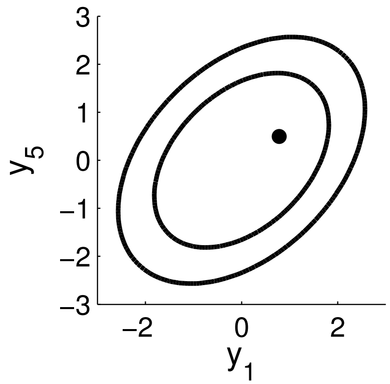
New visualisation

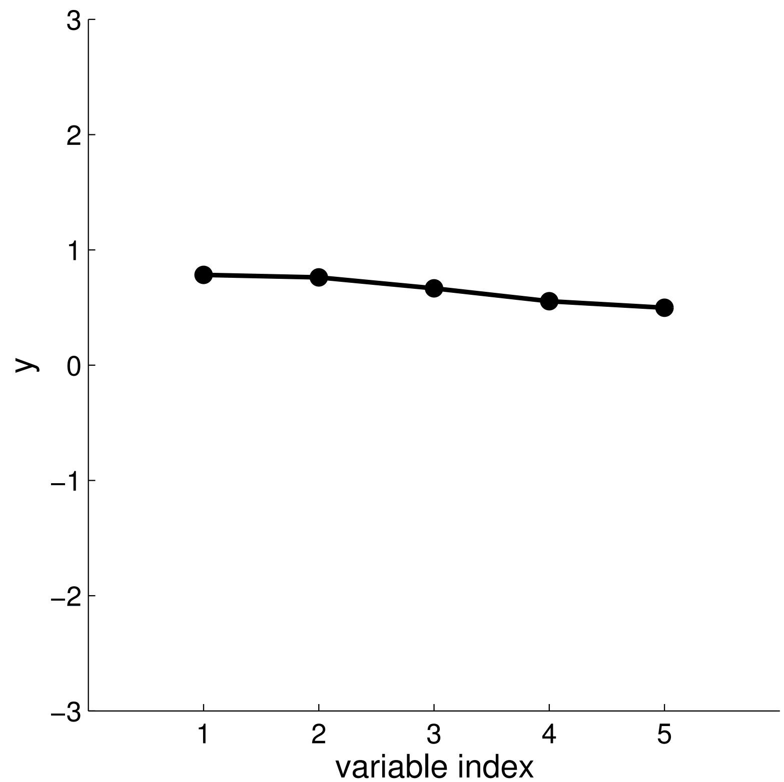


$$\Sigma = \begin{bmatrix} 1 & .9 & .8 & .6 & .4 \\ .9 & 1 & .9 & .8 & .6 \\ .8 & .9 & 1 & .9 & .8 \\ .6 & .8 & .9 & 1 & .9 \\ .4 & .6 & .8 & .9 & 1 \end{bmatrix}$$



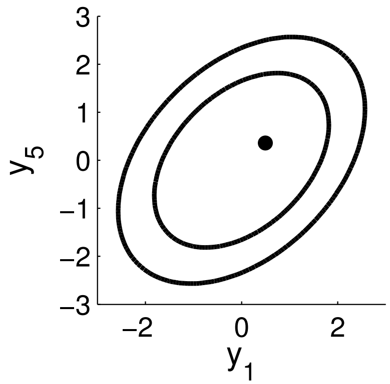
New visualisation

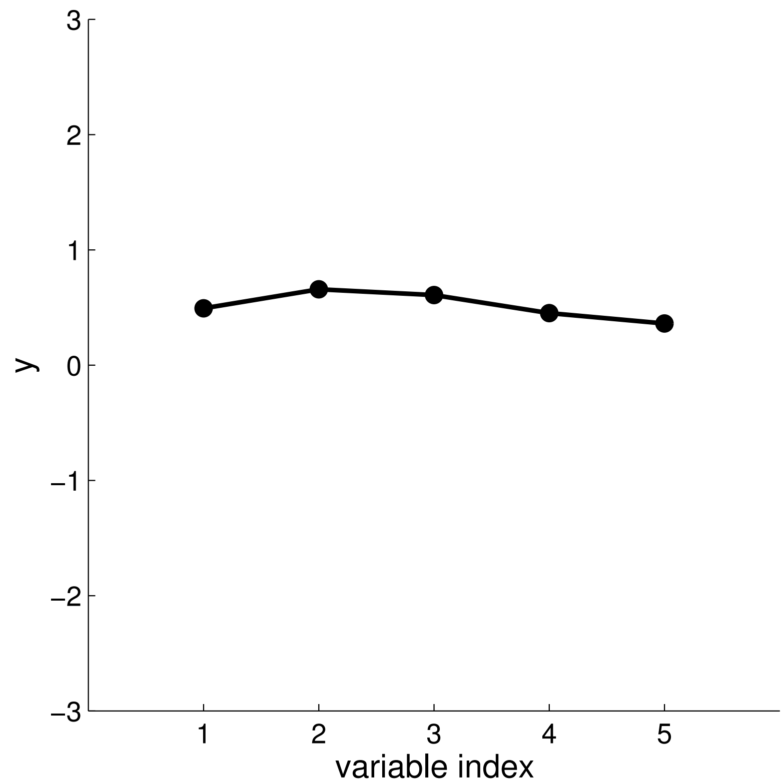




$$\Sigma = \begin{bmatrix} 1 & .9 & .8 & .6 & .4 \\ .9 & 1 & .9 & .8 & .6 \\ .8 & .9 & 1 & .9 & .8 \\ .6 & .8 & .9 & 1 & .9 \\ .4 & .6 & .8 & .9 & 1 \end{bmatrix}$$

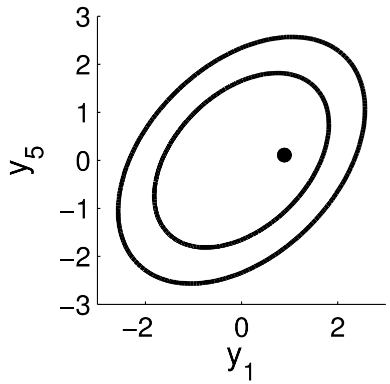
New visualisation

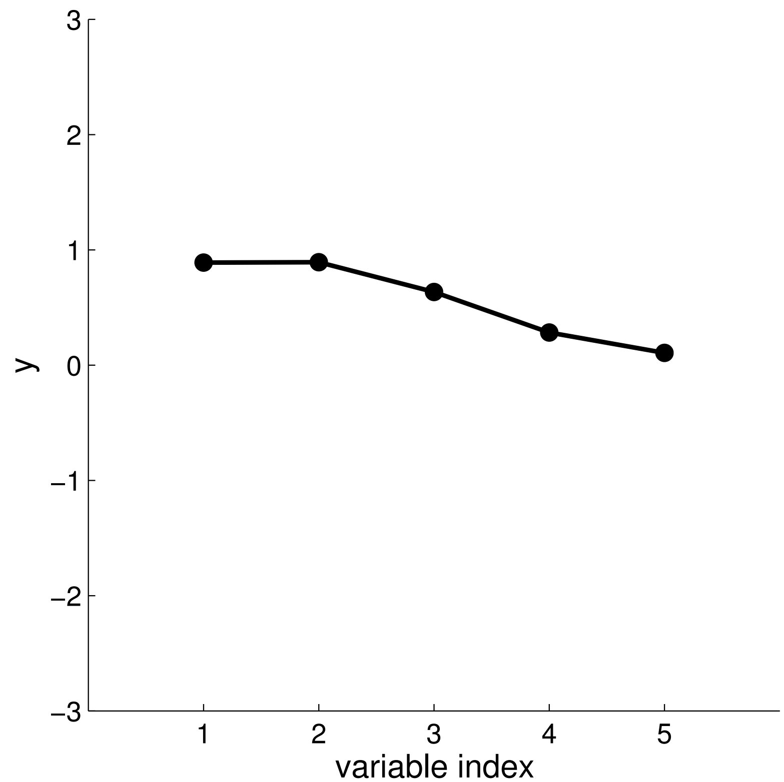




$$\Sigma = \begin{bmatrix} 1 & .9 & .8 & .6 & .4 \\ .9 & 1 & .9 & .8 & .6 \\ .8 & .9 & 1 & .9 & .8 \\ .6 & .8 & .9 & 1 & .9 \\ .4 & .6 & .8 & .9 & 1 \end{bmatrix}$$

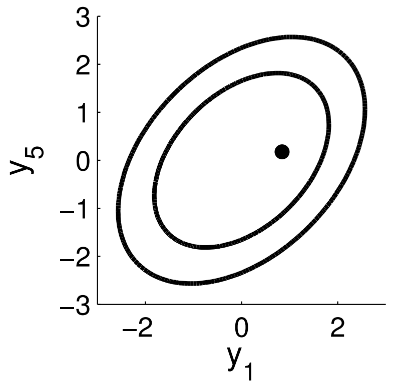
New visualisation

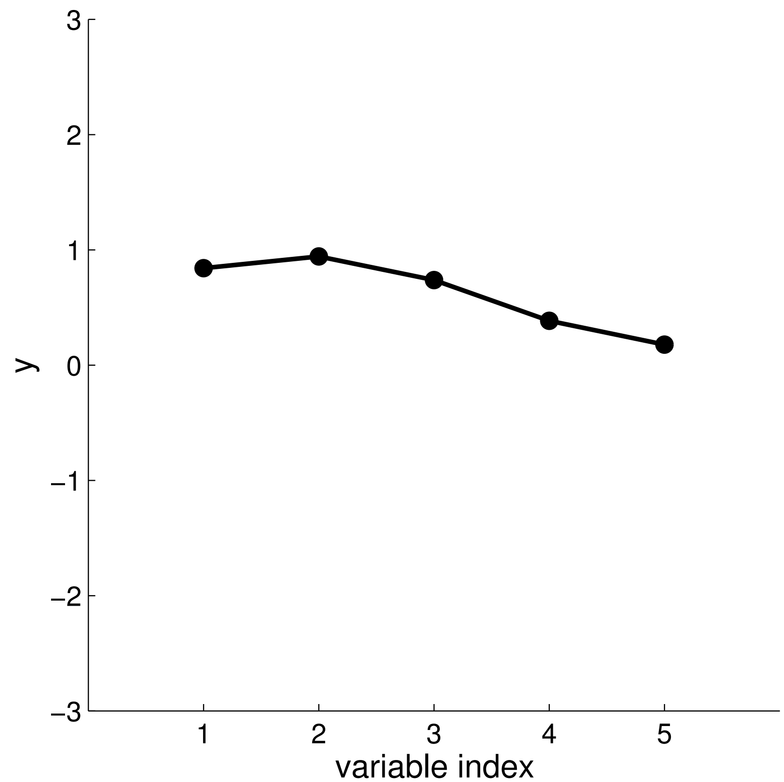




$$\Sigma = \begin{bmatrix} 1 & .9 & .8 & .6 & .4 \\ .9 & 1 & .9 & .8 & .6 \\ .8 & .9 & 1 & .9 & .8 \\ .6 & .8 & .9 & 1 & .9 \\ .4 & .6 & .8 & .9 & 1 \end{bmatrix}$$

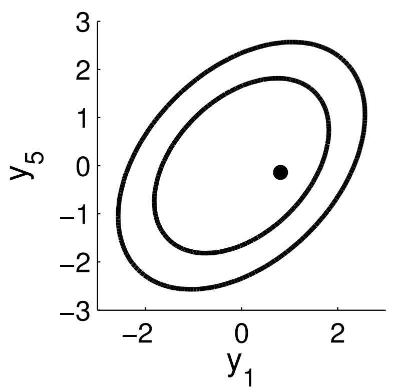
New visualisation

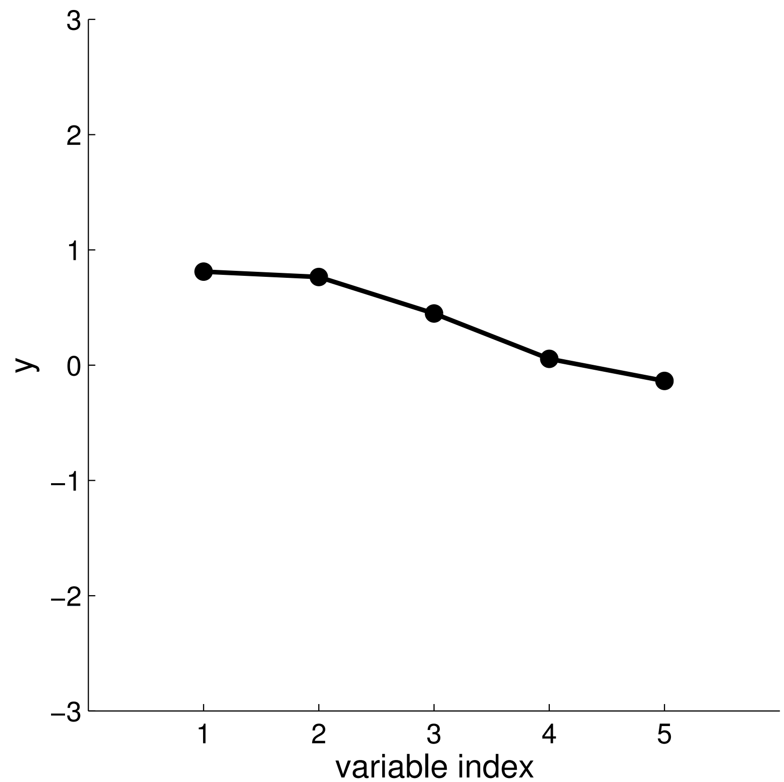




$$\Sigma = \begin{bmatrix} 1 & .9 & .8 & .6 & .4 \\ .9 & 1 & .9 & .8 & .6 \\ .8 & .9 & 1 & .9 & .8 \\ .6 & .8 & .9 & 1 & .9 \\ .4 & .6 & .8 & .9 & 1 \end{bmatrix}$$

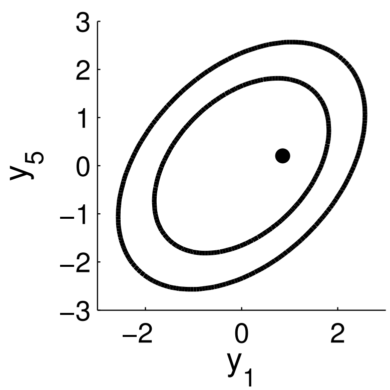
New visualisation

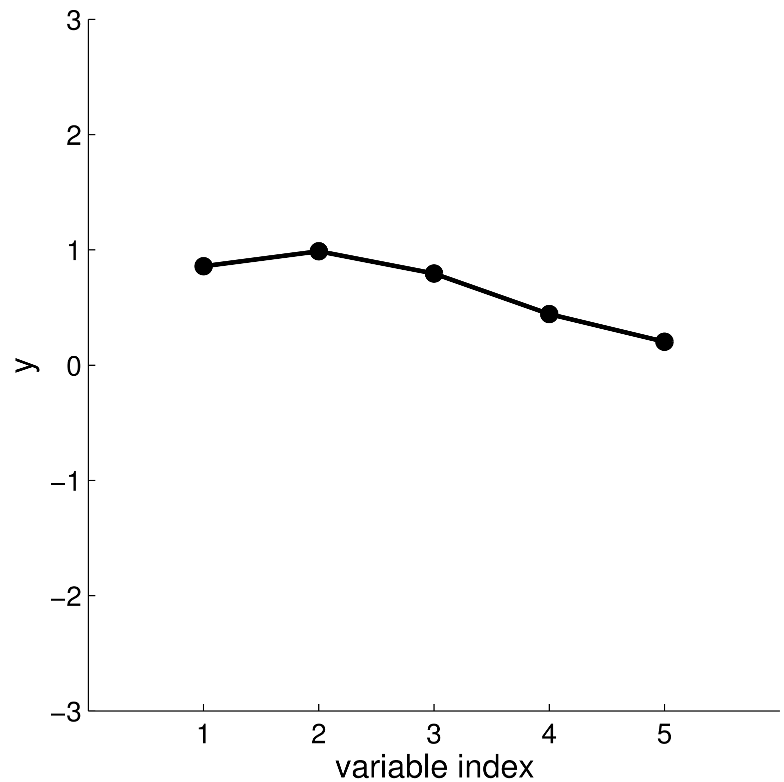




$$\Sigma = \begin{bmatrix} 1 & .9 & .8 & .6 & .4 \\ .9 & 1 & .9 & .8 & .6 \\ .8 & .9 & 1 & .9 & .8 \\ .6 & .8 & .9 & 1 & .9 \\ .4 & .6 & .8 & .9 & 1 \end{bmatrix}$$

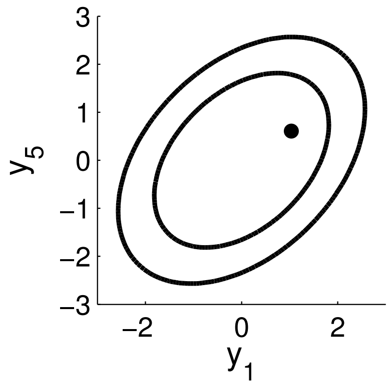
New visualisation

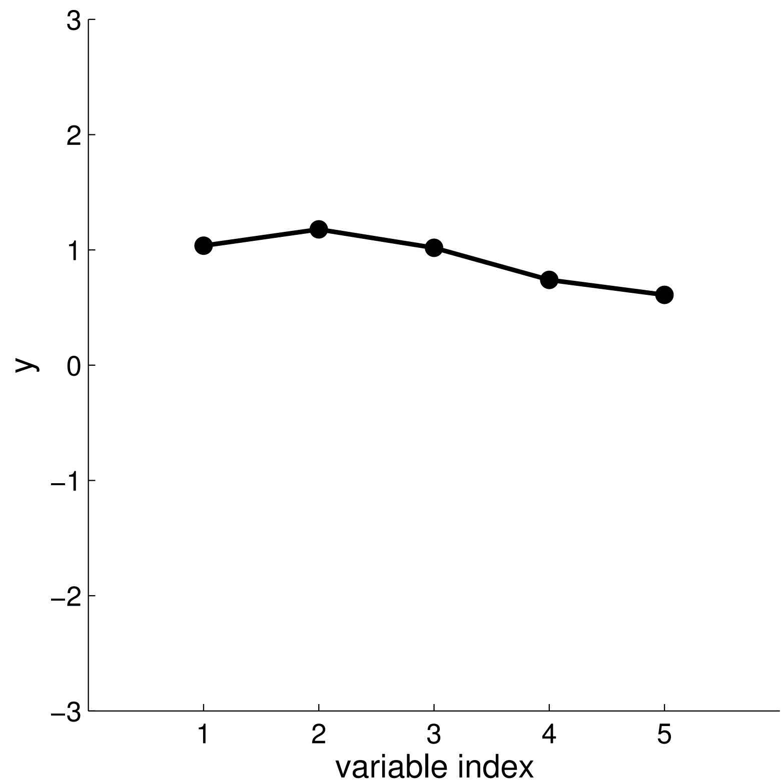




$$\Sigma = \begin{bmatrix} 1 & .9 & .8 & .6 & .4 \\ .9 & 1 & .9 & .8 & .6 \\ .8 & .9 & 1 & .9 & .8 \\ .6 & .8 & .9 & 1 & .9 \\ .4 & .6 & .8 & .9 & 1 \end{bmatrix}$$

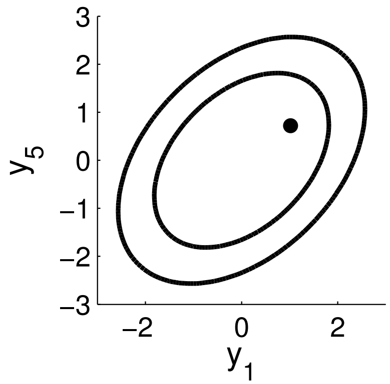
New visualisation

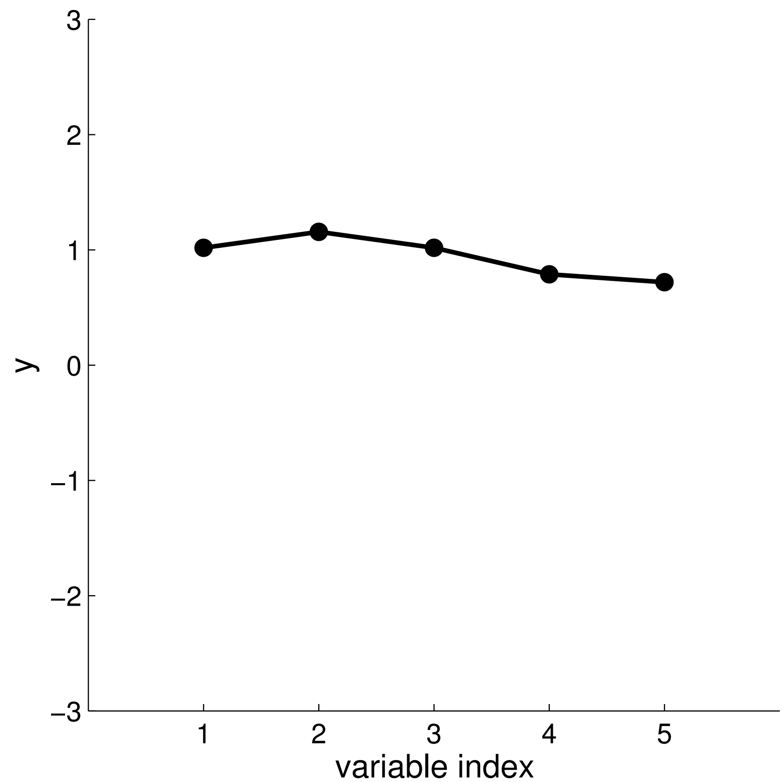




$$\Sigma = \begin{bmatrix} 1 & .9 & .8 & .6 & .4 \\ .9 & 1 & .9 & .8 & .6 \\ .8 & .9 & 1 & .9 & .8 \\ .6 & .8 & .9 & 1 & .9 \\ .4 & .6 & .8 & .9 & 1 \end{bmatrix}$$

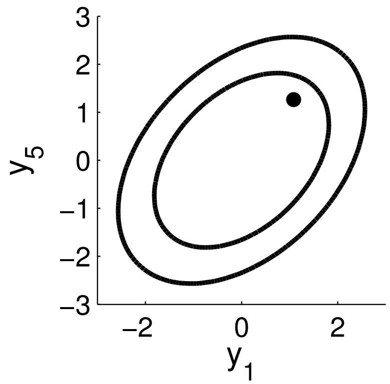
New visualisation

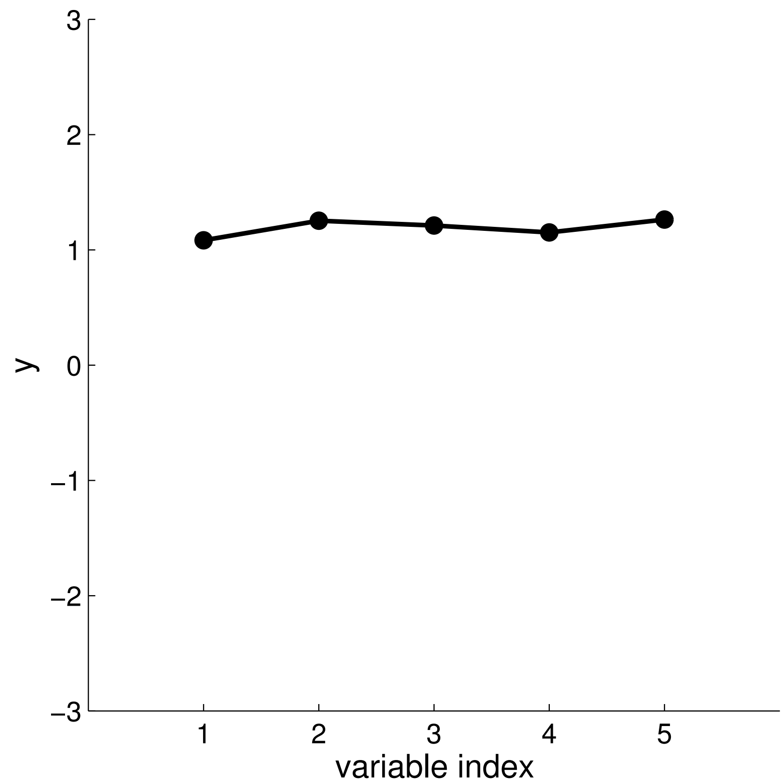




$$\Sigma = \begin{bmatrix} 1 & .9 & .8 & .6 & .4 \\ .9 & 1 & .9 & .8 & .6 \\ .8 & .9 & 1 & .9 & .8 \\ .6 & .8 & .9 & 1 & .9 \\ .4 & .6 & .8 & .9 & 1 \end{bmatrix}$$

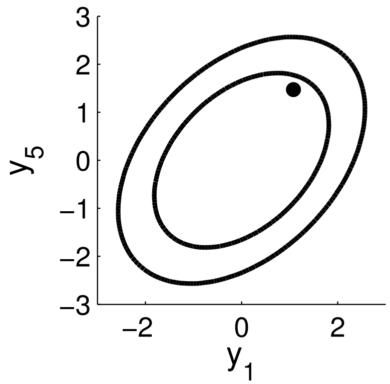
New visualisation

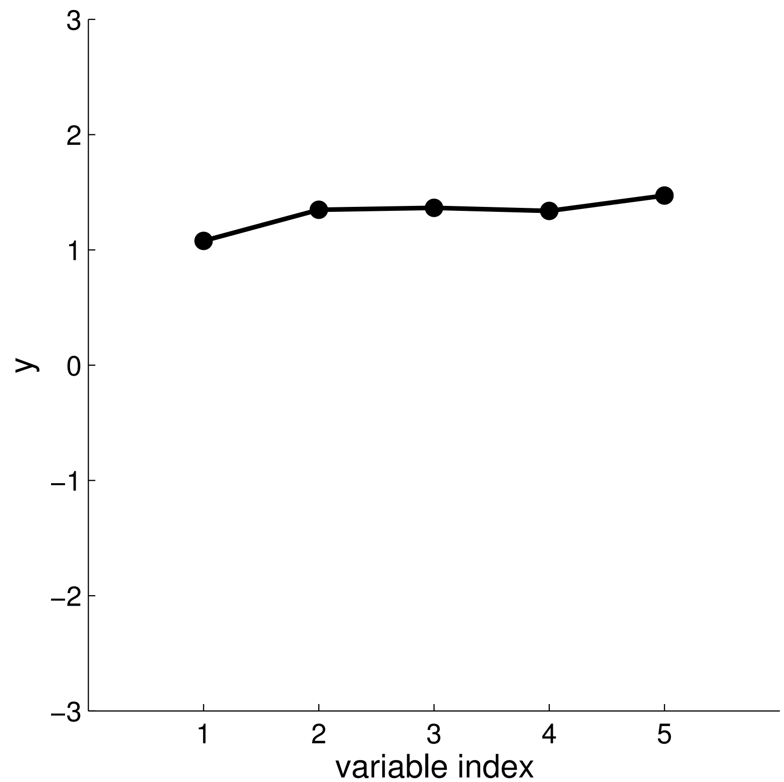




$$\Sigma = \begin{bmatrix} 1 & .9 & .8 & .6 & .4 \\ .9 & 1 & .9 & .8 & .6 \\ .8 & .9 & 1 & .9 & .8 \\ .6 & .8 & .9 & 1 & .9 \\ .4 & .6 & .8 & .9 & 1 \end{bmatrix}$$

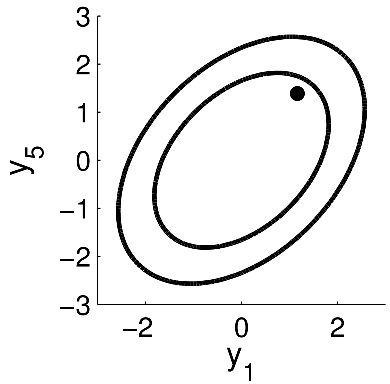
New visualisation

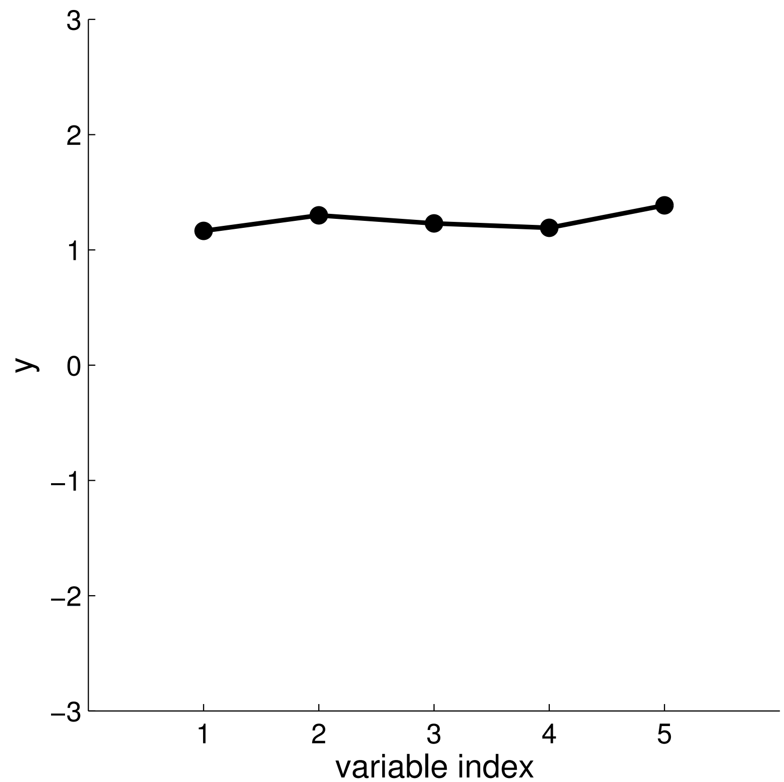




$$\Sigma = \begin{bmatrix} 1 & .9 & .8 & .6 & .4 \\ .9 & 1 & .9 & .8 & .6 \\ .8 & .9 & 1 & .9 & .8 \\ .6 & .8 & .9 & 1 & .9 \\ .4 & .6 & .8 & .9 & 1 \end{bmatrix}$$

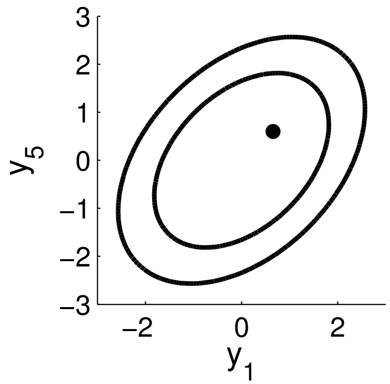
New visualisation

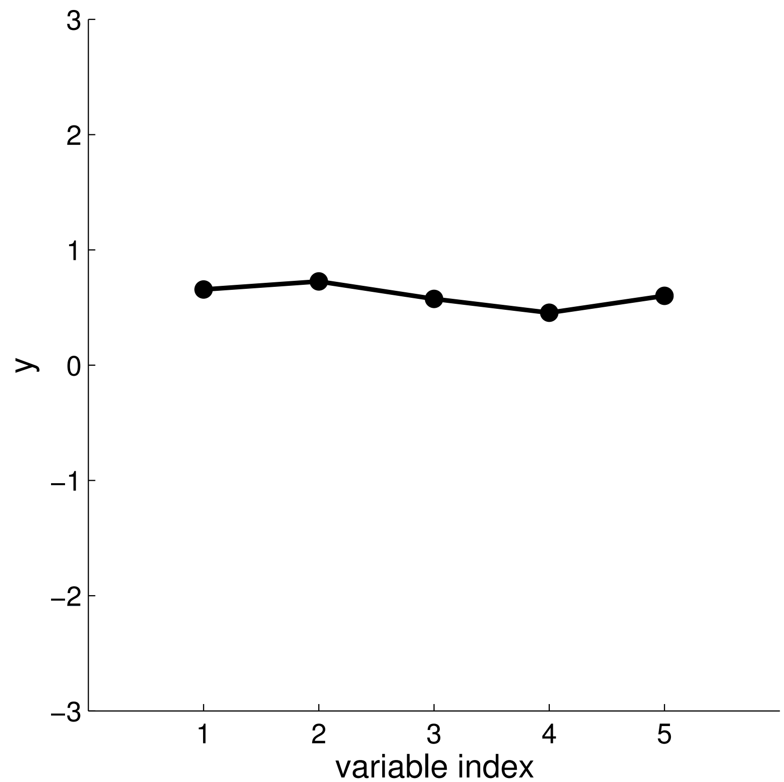




$$\Sigma = \begin{bmatrix} 1 & .9 & .8 & .6 & .4 \\ .9 & 1 & .9 & .8 & .6 \\ .8 & .9 & 1 & .9 & .8 \\ .6 & .8 & .9 & 1 & .9 \\ .4 & .6 & .8 & .9 & 1 \end{bmatrix}$$

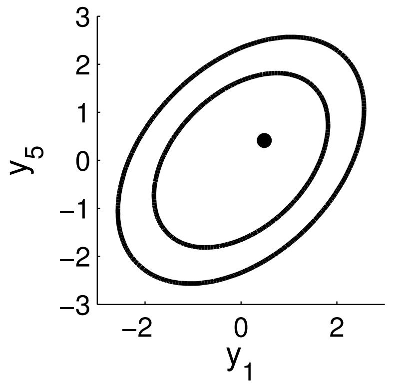
New visualisation

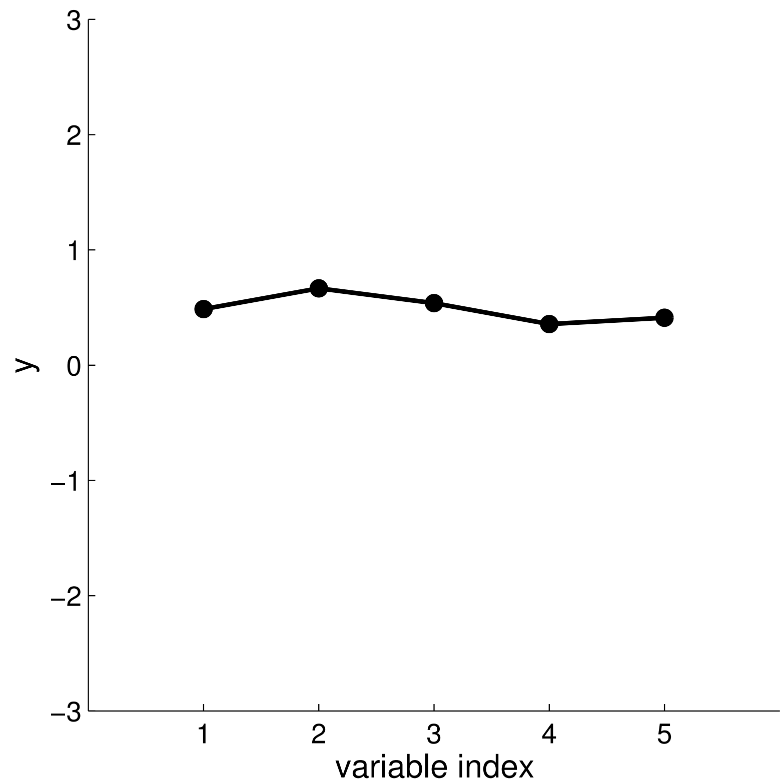




$$\Sigma = \begin{bmatrix} 1 & .9 & .8 & .6 & .4 \\ .9 & 1 & .9 & .8 & .6 \\ .8 & .9 & 1 & .9 & .8 \\ .6 & .8 & .9 & 1 & .9 \\ .4 & .6 & .8 & .9 & 1 \end{bmatrix}$$

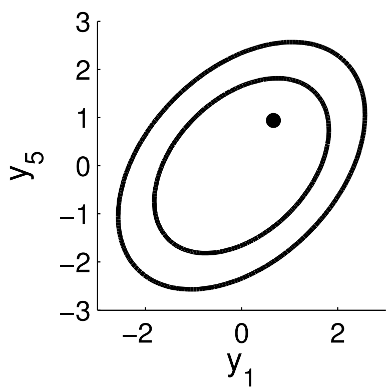
New visualisation

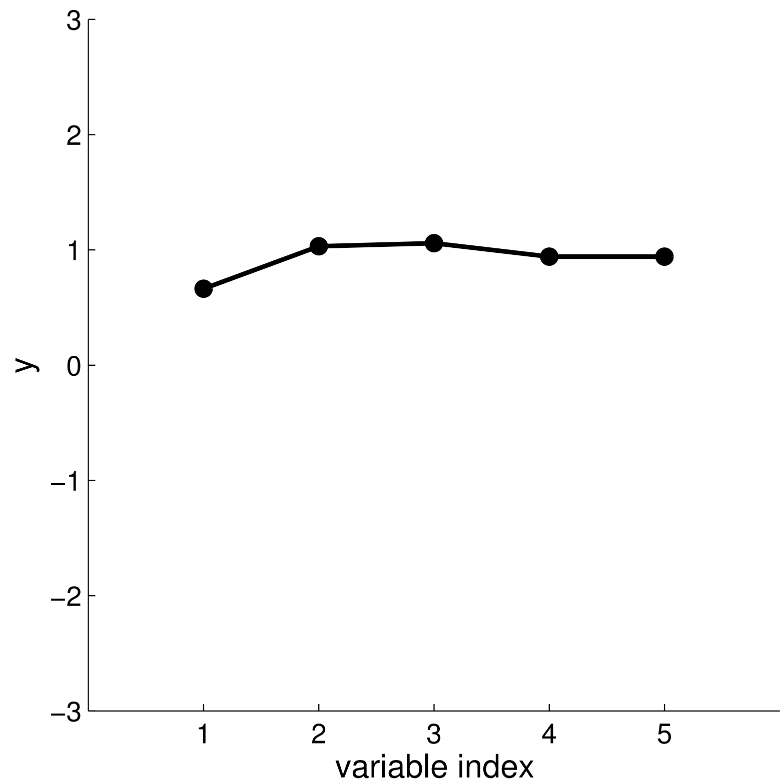




$$\Sigma = \begin{bmatrix} 1 & .9 & .8 & .6 & .4 \\ .9 & 1 & .9 & .8 & .6 \\ .8 & .9 & 1 & .9 & .8 \\ .6 & .8 & .9 & 1 & .9 \\ .4 & .6 & .8 & .9 & 1 \end{bmatrix}$$

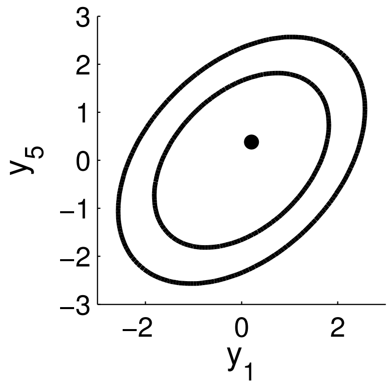
New visualisation

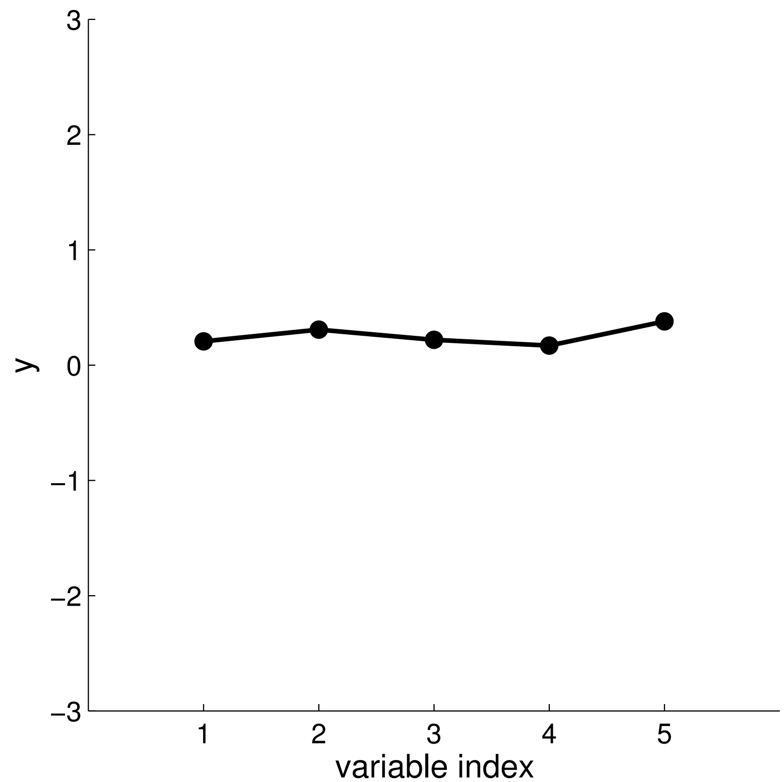




$$\Sigma = \begin{bmatrix} 1 & .9 & .8 & .6 & .4 \\ .9 & 1 & .9 & .8 & .6 \\ .8 & .9 & 1 & .9 & .8 \\ .6 & .8 & .9 & 1 & .9 \\ .4 & .6 & .8 & .9 & 1 \end{bmatrix}$$

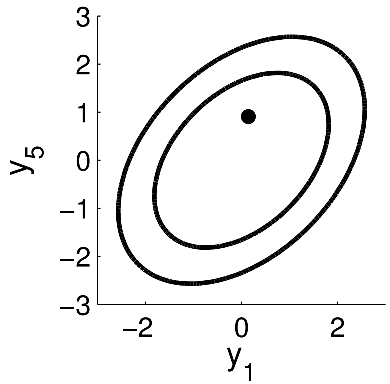
New visualisation

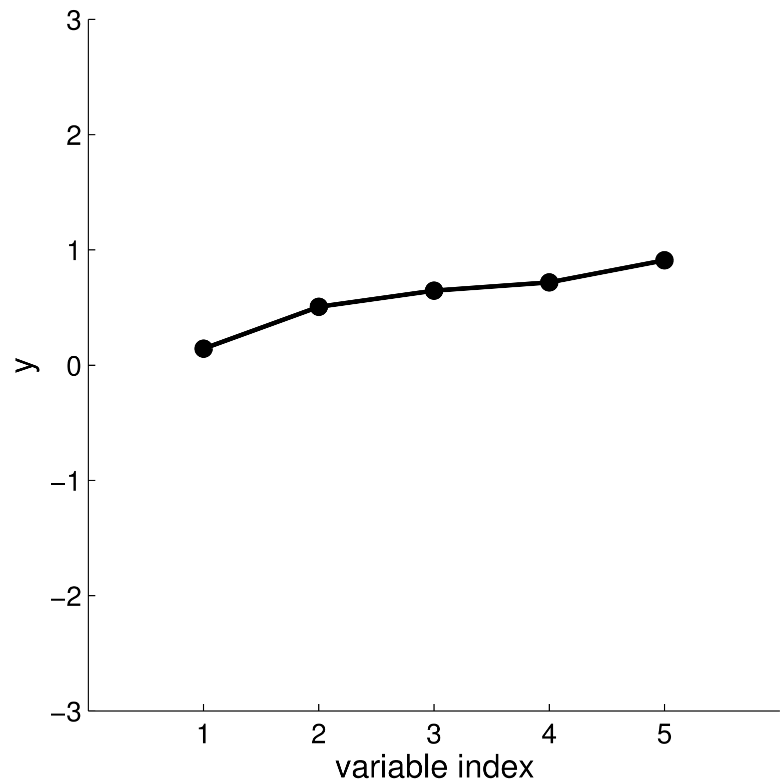




$$\Sigma = \begin{bmatrix} 1 & .9 & .8 & .6 & .4 \\ .9 & 1 & .9 & .8 & .6 \\ .8 & .9 & 1 & .9 & .8 \\ .6 & .8 & .9 & 1 & .9 \\ .4 & .6 & .8 & .9 & 1 \end{bmatrix}$$

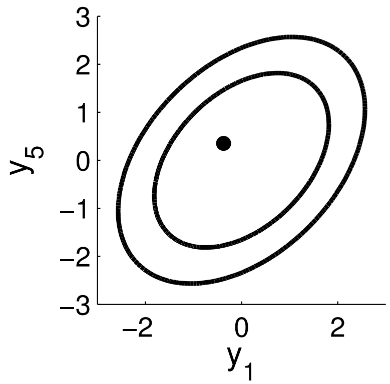
New visualisation



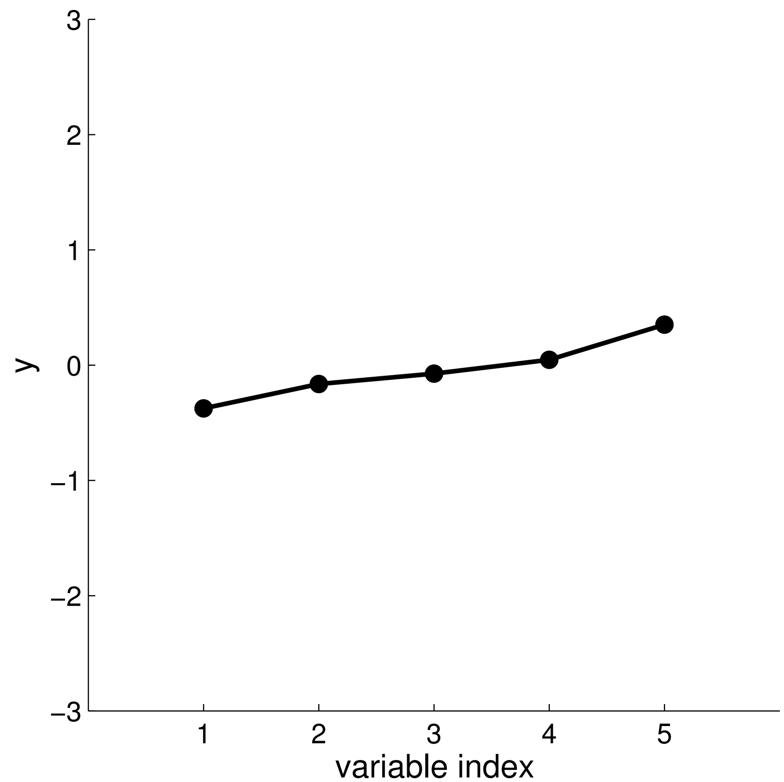


$$\Sigma = \begin{bmatrix} 1 & .9 & .8 & .6 & .4 \\ .9 & 1 & .9 & .8 & .6 \\ .8 & .9 & 1 & .9 & .8 \\ .6 & .8 & .9 & 1 & .9 \\ .4 & .6 & .8 & .9 & 1 \end{bmatrix}$$

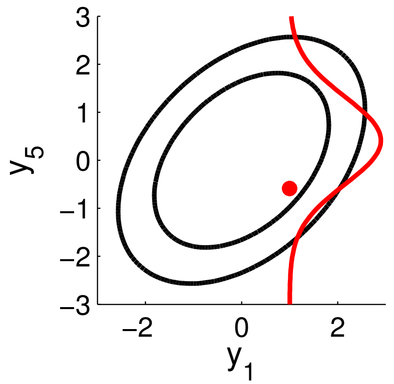
New visualisation

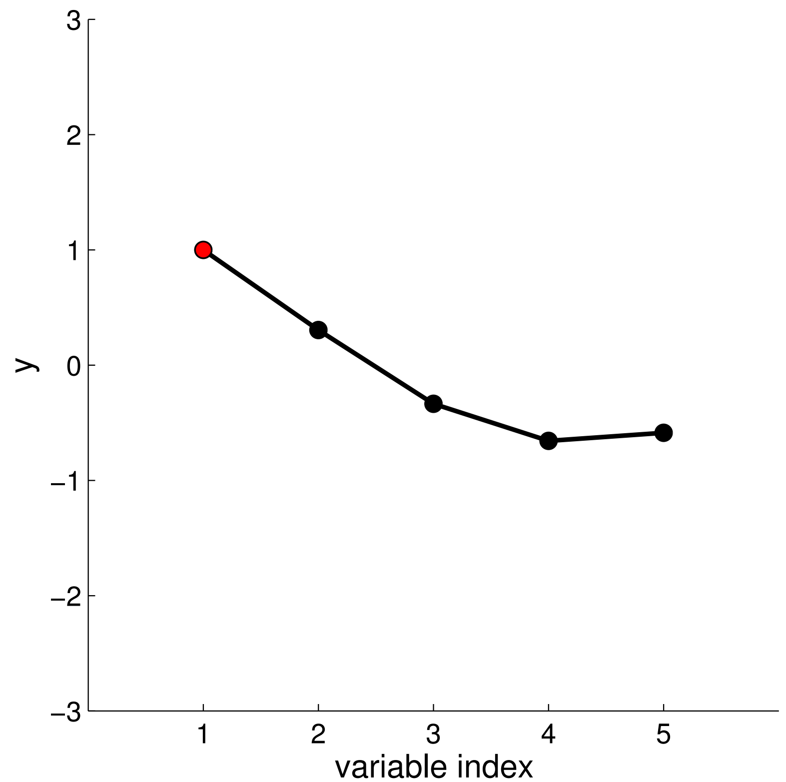


$$\Sigma = \begin{bmatrix} 1 & .9 & .8 & .6 & .4 \\ .9 & 1 & .9 & .8 & .6 \\ .8 & .9 & 1 & .9 & .8 \\ .6 & .8 & .9 & 1 & .9 \\ .4 & .6 & .8 & .9 & 1 \end{bmatrix}$$



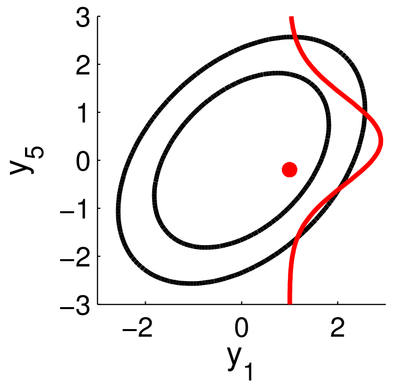
New visualisation

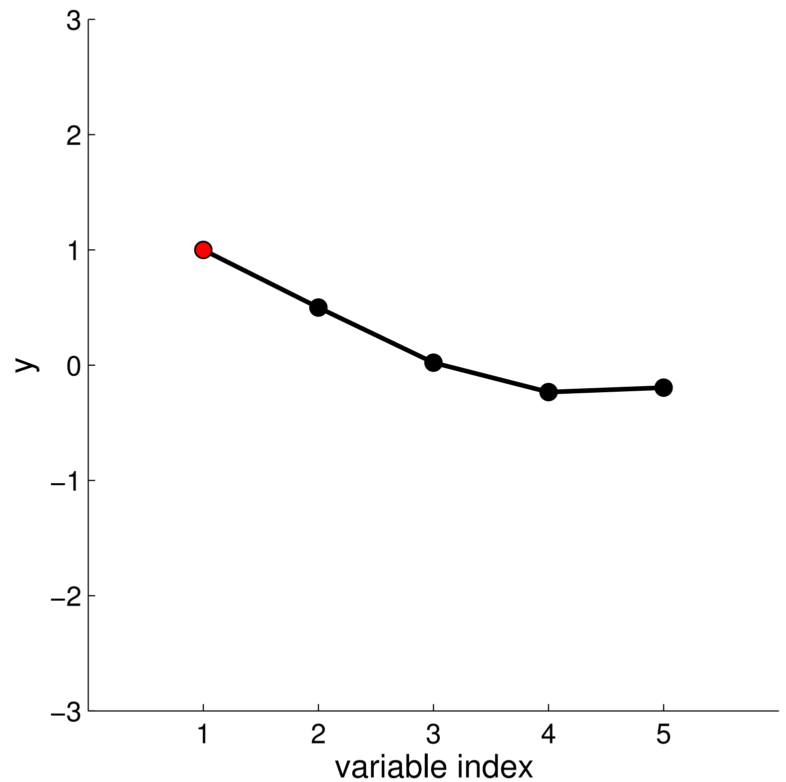




$$\Sigma = \begin{bmatrix} 1 & .9 & .8 & .6 & .4 \\ .9 & 1 & .9 & .8 & .6 \\ .8 & .9 & 1 & .9 & .8 \\ .6 & .8 & .9 & 1 & .9 \\ .4 & .6 & .8 & .9 & 1 \end{bmatrix}$$

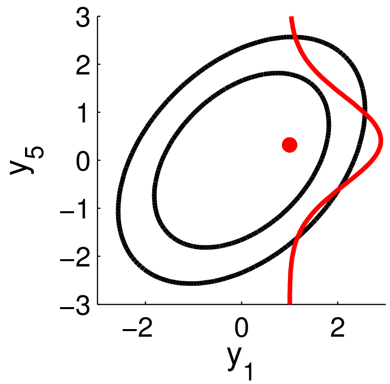
New visualisation



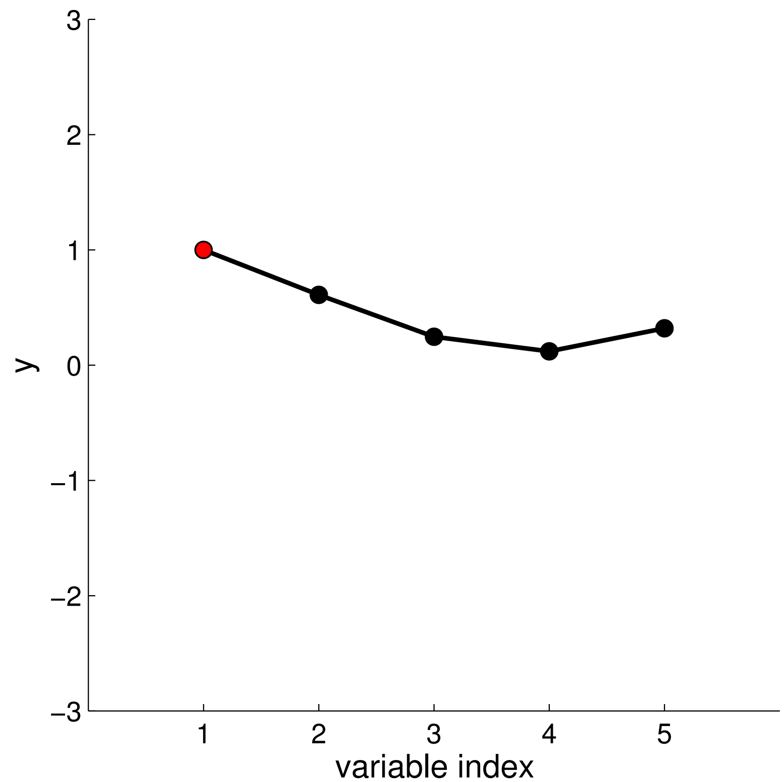


$$\Sigma = \begin{bmatrix} 1 & .9 & .8 & .6 & .4 \\ .9 & 1 & .9 & .8 & .6 \\ .8 & .9 & 1 & .9 & .8 \\ .6 & .8 & .9 & 1 & .9 \\ .4 & .6 & .8 & .9 & 1 \end{bmatrix}$$

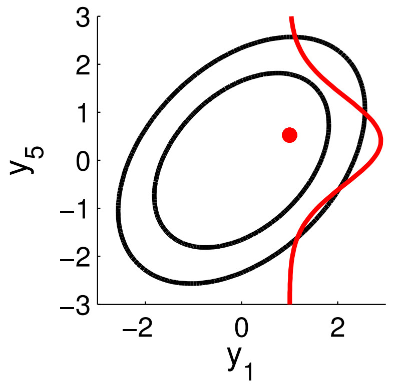
New visualisation

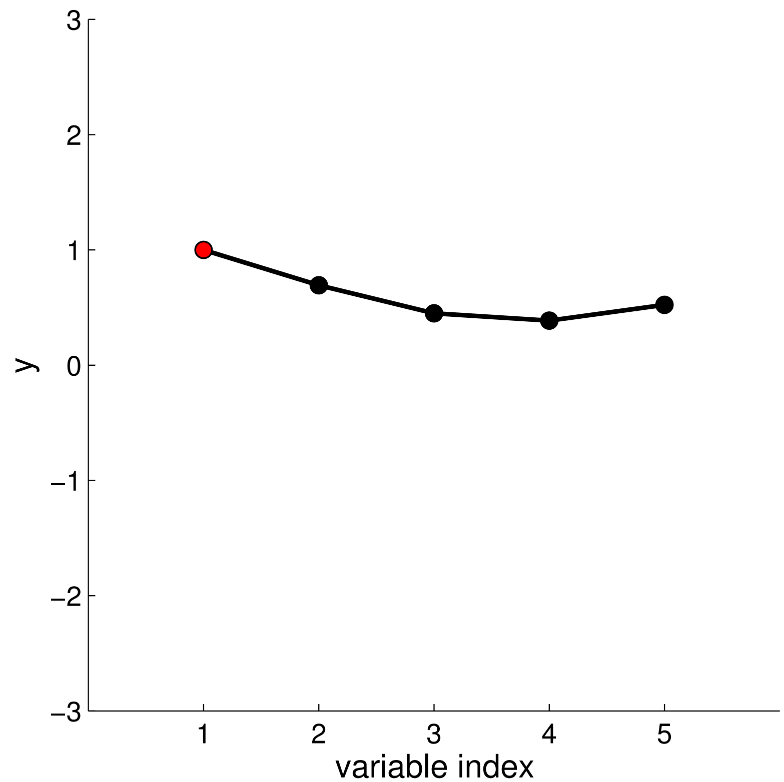


$$\Sigma = \begin{bmatrix} 1 & .9 & .8 & .6 & .4 \\ .9 & 1 & .9 & .8 & .6 \\ .8 & .9 & 1 & .9 & .8 \\ .6 & .8 & .9 & 1 & .9 \\ .4 & .6 & .8 & .9 & 1 \end{bmatrix}$$



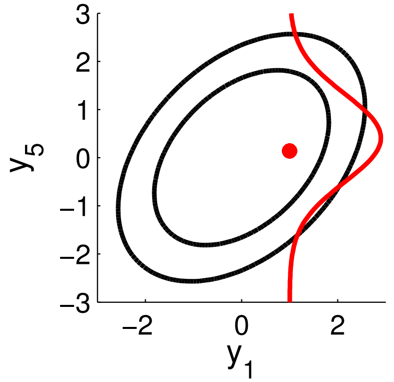
New visualisation

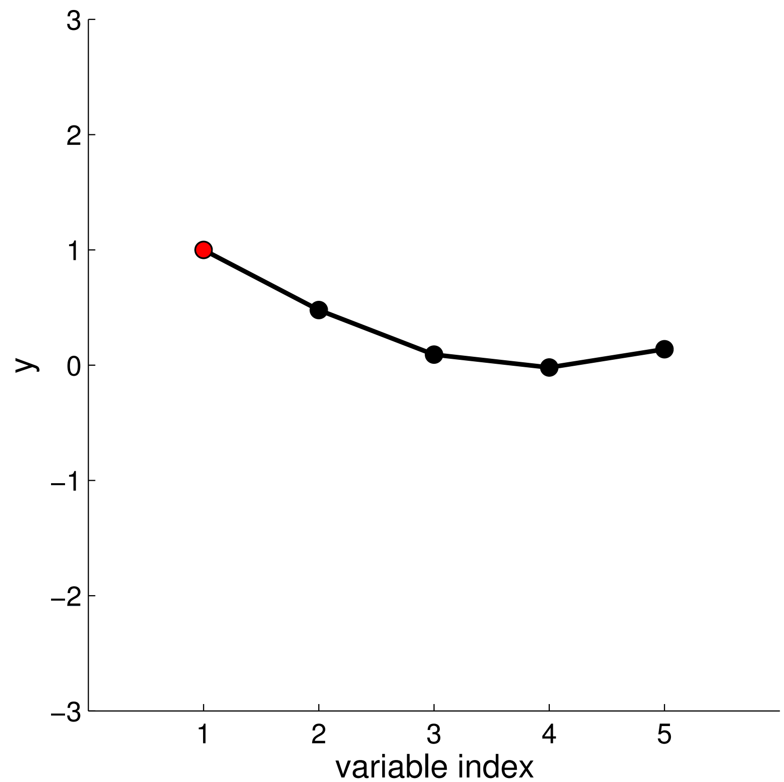




$$\Sigma = \begin{bmatrix} 1 & .9 & .8 & .6 & .4 \\ .9 & 1 & .9 & .8 & .6 \\ .8 & .9 & 1 & .9 & .8 \\ .6 & .8 & .9 & 1 & .9 \\ .4 & .6 & .8 & .9 & 1 \end{bmatrix}$$

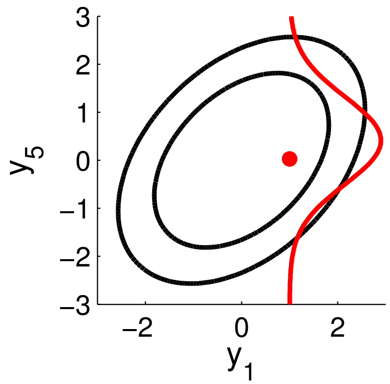
New visualisation

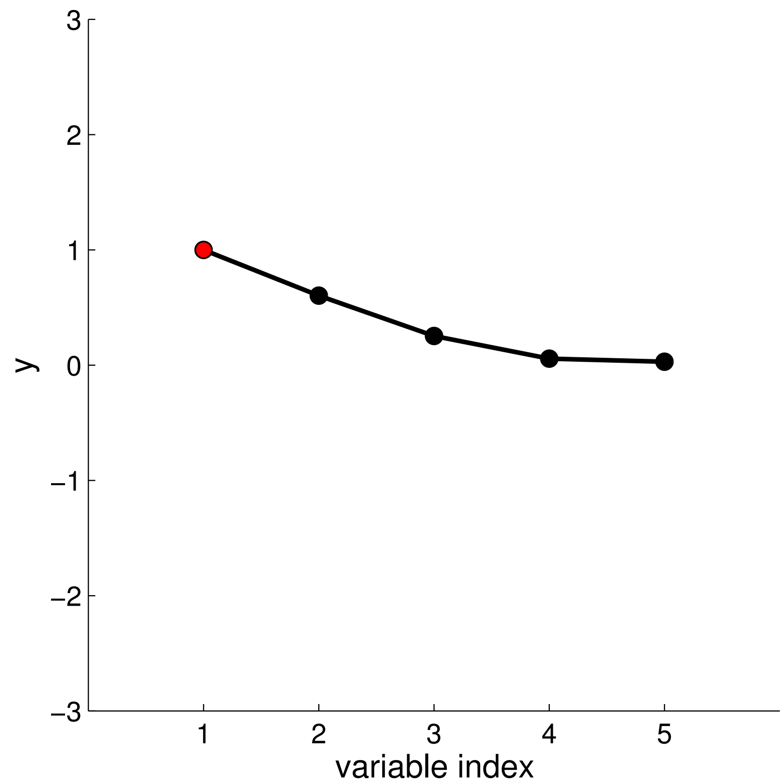




$$\Sigma = \begin{bmatrix} 1 & .9 & .8 & .6 & .4 \\ .9 & 1 & .9 & .8 & .6 \\ .8 & .9 & 1 & .9 & .8 \\ .6 & .8 & .9 & 1 & .9 \\ .4 & .6 & .8 & .9 & 1 \end{bmatrix}$$

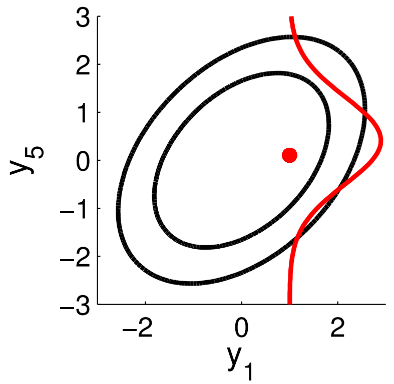
New visualisation

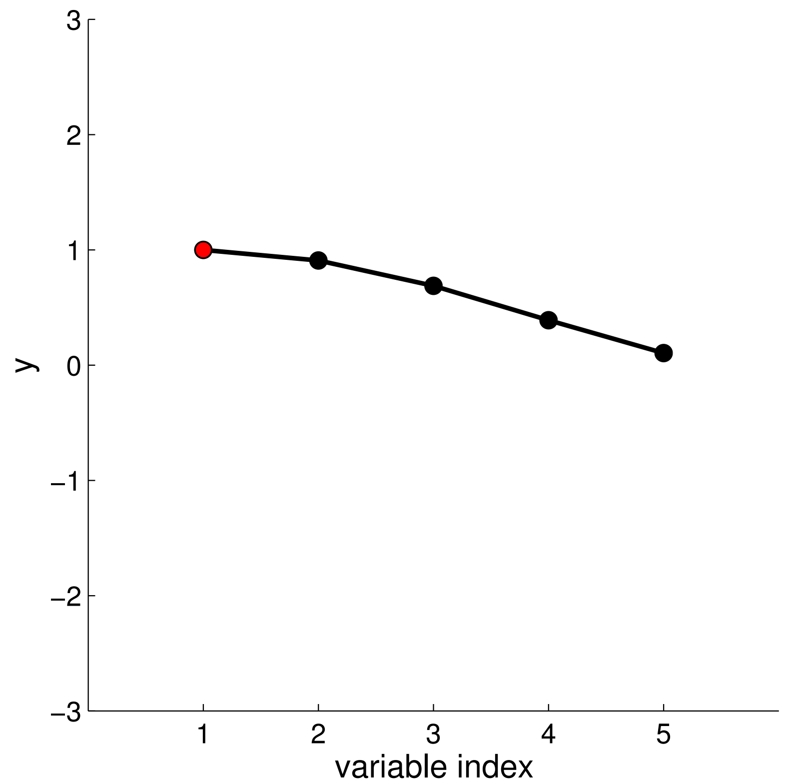




$$\Sigma = \begin{bmatrix} 1 & .9 & .8 & .6 & .4 \\ .9 & 1 & .9 & .8 & .6 \\ .8 & .9 & 1 & .9 & .8 \\ .6 & .8 & .9 & 1 & .9 \\ .4 & .6 & .8 & .9 & 1 \end{bmatrix}$$

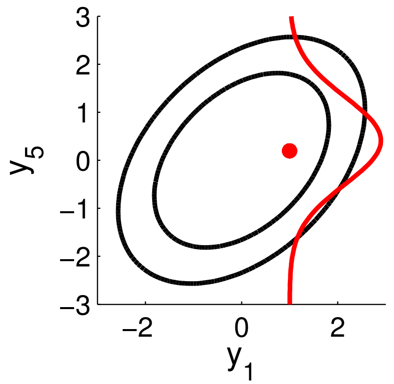
New visualisation



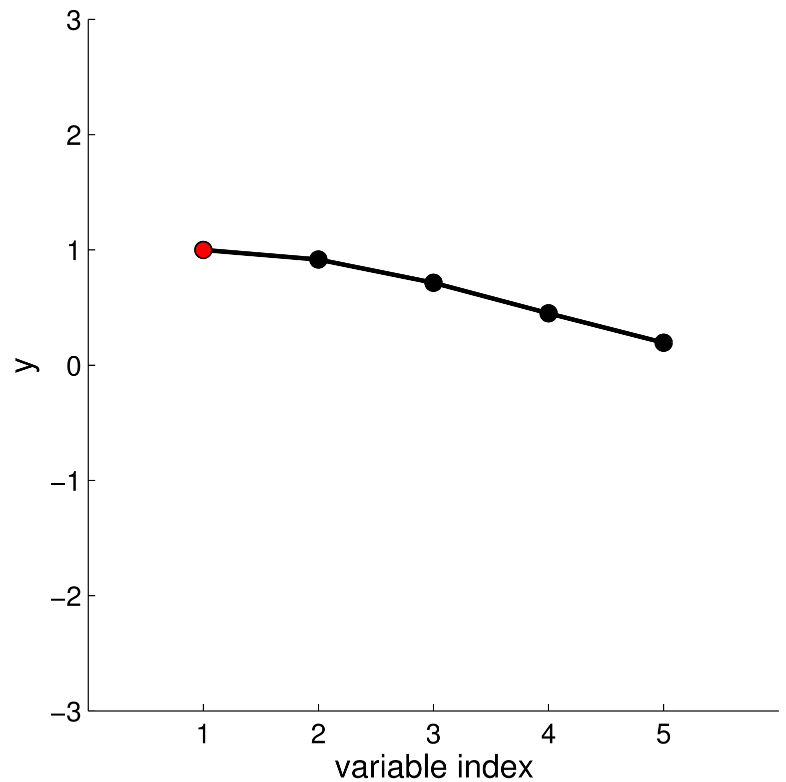


$$\Sigma = \begin{bmatrix} 1 & .9 & .8 & .6 & .4 \\ .9 & 1 & .9 & .8 & .6 \\ .8 & .9 & 1 & .9 & .8 \\ .6 & .8 & .9 & 1 & .9 \\ .4 & .6 & .8 & .9 & 1 \end{bmatrix}$$

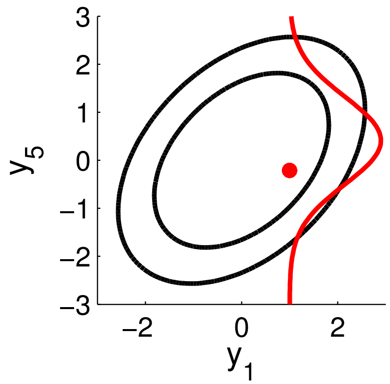
New visualisation

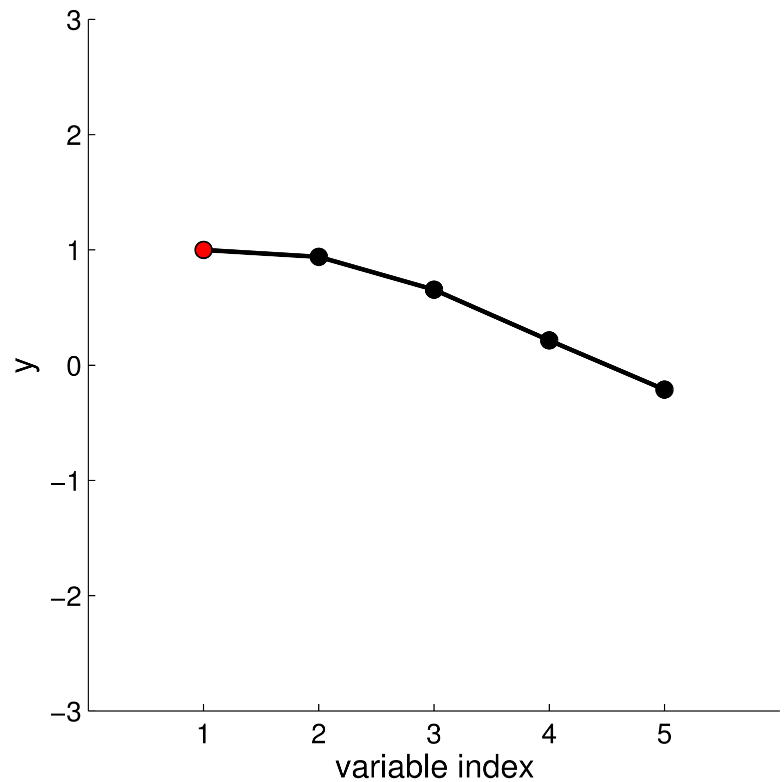


$$\Sigma = \begin{bmatrix} 1 & .9 & .8 & .6 & .4 \\ .9 & 1 & .9 & .8 & .6 \\ .8 & .9 & 1 & .9 & .8 \\ .6 & .8 & .9 & 1 & .9 \\ .4 & .6 & .8 & .9 & 1 \end{bmatrix}$$



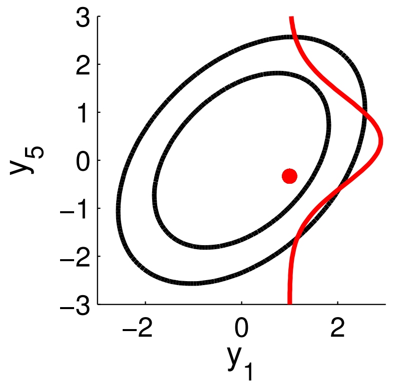
New visualisation

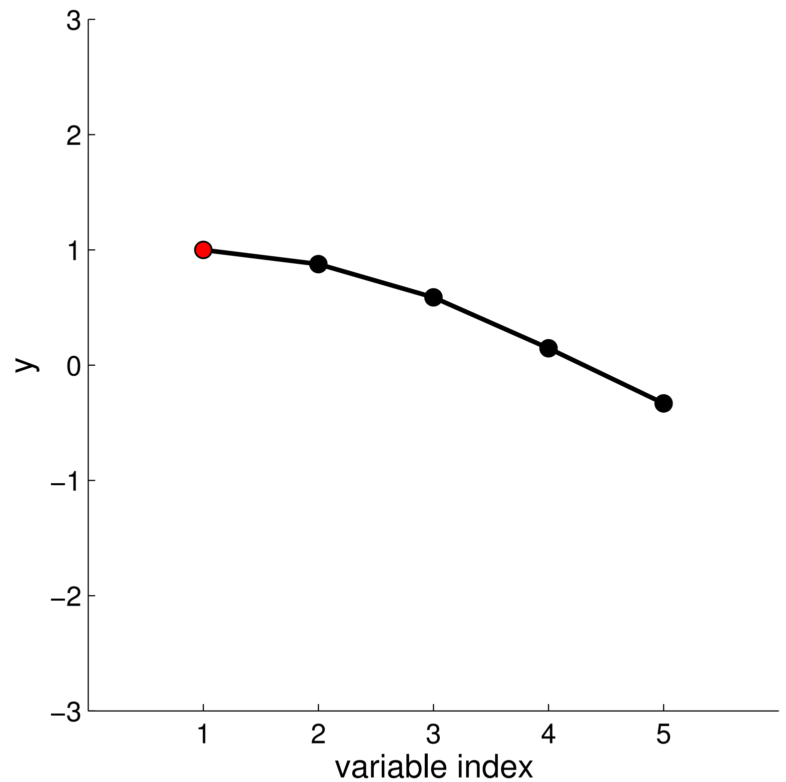




$$\Sigma = \begin{bmatrix} 1 & .9 & .8 & .6 & .4 \\ .9 & 1 & .9 & .8 & .6 \\ .8 & .9 & 1 & .9 & .8 \\ .6 & .8 & .9 & 1 & .9 \\ .4 & .6 & .8 & .9 & 1 \end{bmatrix}$$

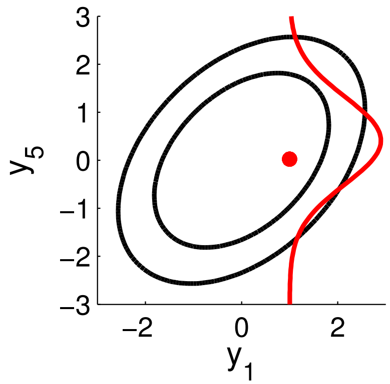
New visualisation

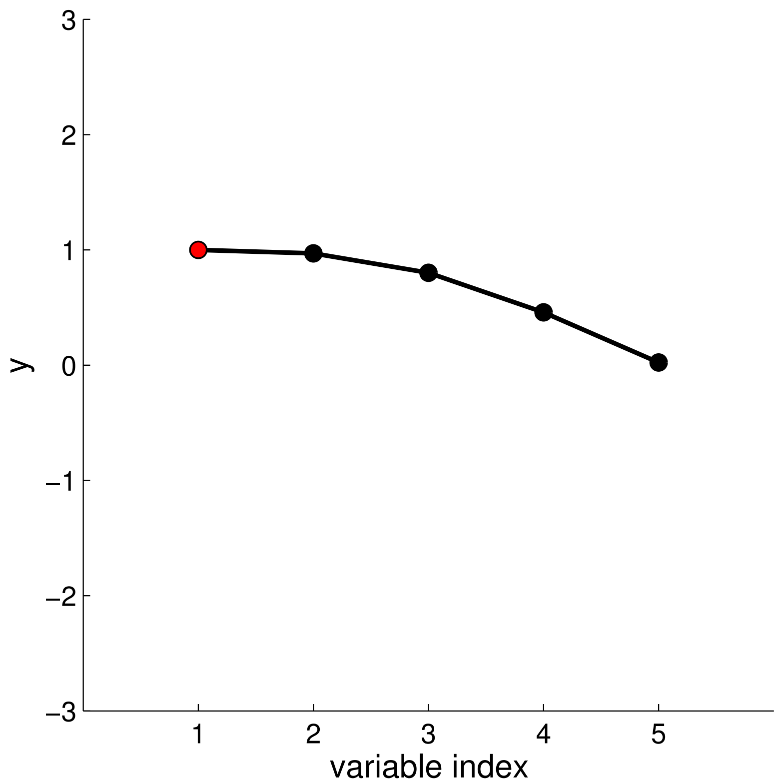




$$\Sigma = \begin{bmatrix} 1 & .9 & .8 & .6 & .4 \\ .9 & 1 & .9 & .8 & .6 \\ .8 & .9 & 1 & .9 & .8 \\ .6 & .8 & .9 & 1 & .9 \\ .4 & .6 & .8 & .9 & 1 \end{bmatrix}$$

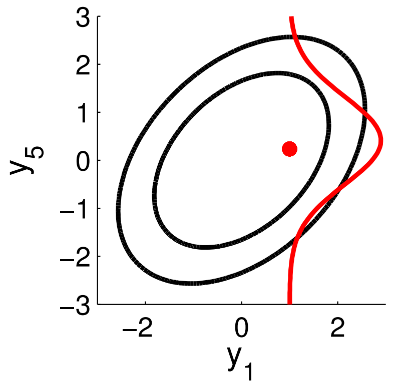
New visualisation

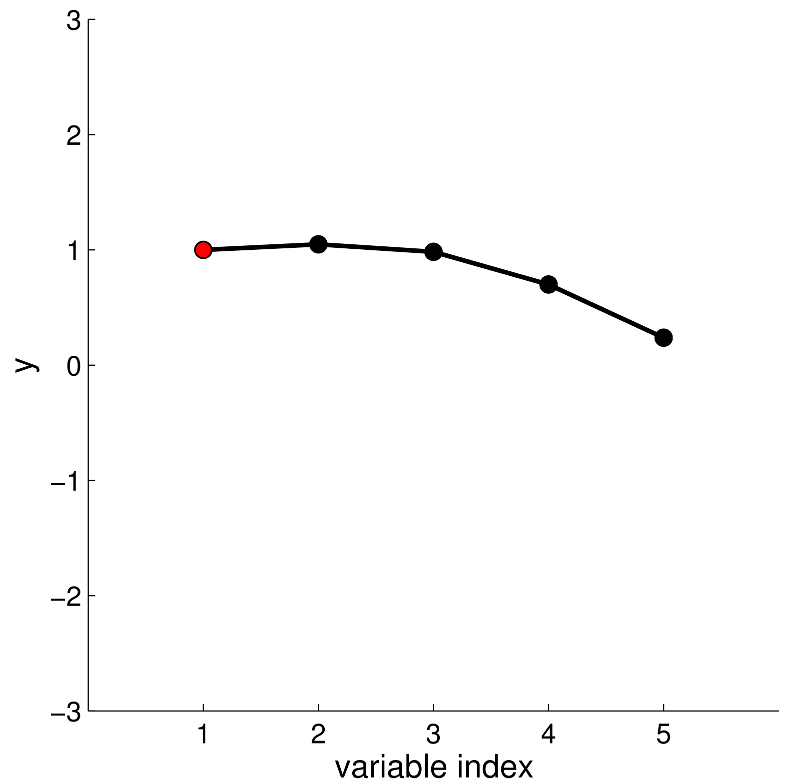




$$\Sigma = \begin{bmatrix} 1 & .9 & .8 & .6 & .4 \\ .9 & 1 & .9 & .8 & .6 \\ .8 & .9 & 1 & .9 & .8 \\ .6 & .8 & .9 & 1 & .9 \\ .4 & .6 & .8 & .9 & 1 \end{bmatrix}$$

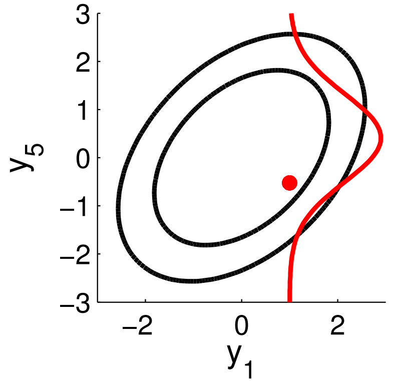
New visualisation



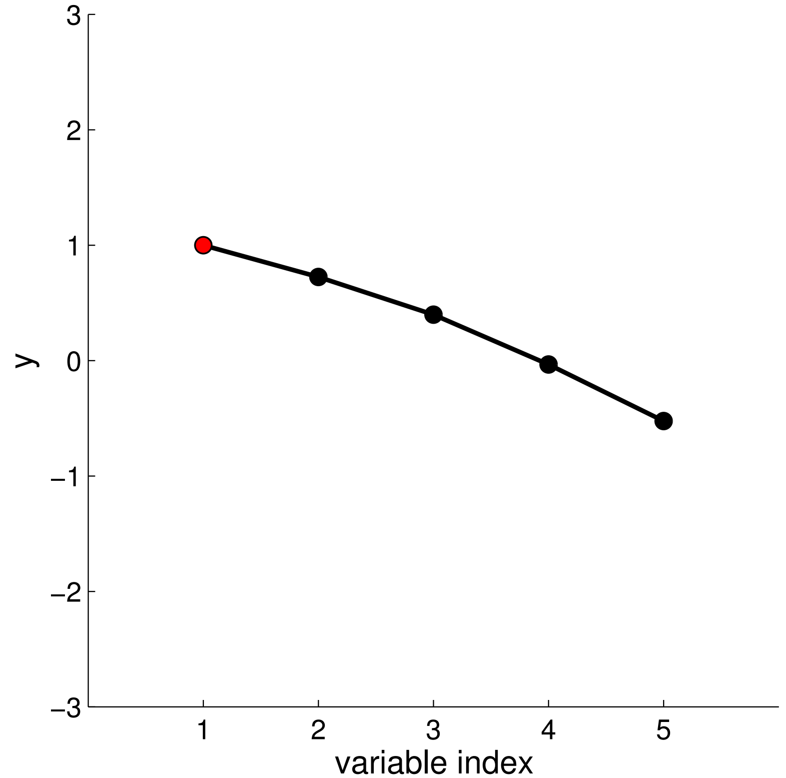


$$\Sigma = \begin{bmatrix} 1 & .9 & .8 & .6 & .4 \\ .9 & 1 & .9 & .8 & .6 \\ .8 & .9 & 1 & .9 & .8 \\ .6 & .8 & .9 & 1 & .9 \\ .4 & .6 & .8 & .9 & 1 \end{bmatrix}$$

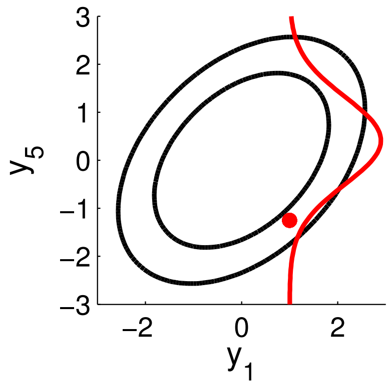
New visualisation

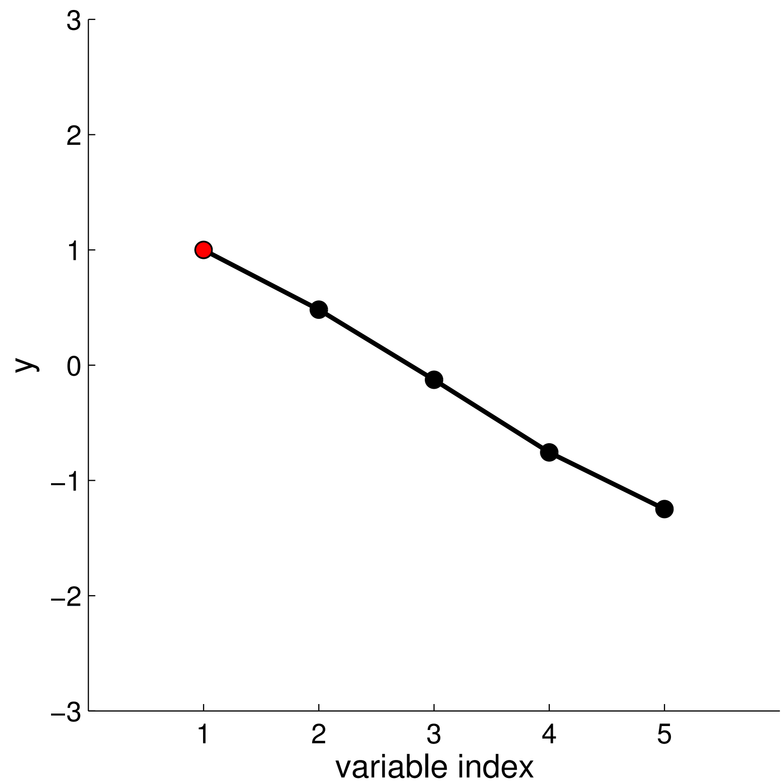


$$\Sigma = \begin{bmatrix} 1 & .9 & .8 & .6 & .4 \\ .9 & 1 & .9 & .8 & .6 \\ .8 & .9 & 1 & .9 & .8 \\ .6 & .8 & .9 & 1 & .9 \\ .4 & .6 & .8 & .9 & 1 \end{bmatrix}$$



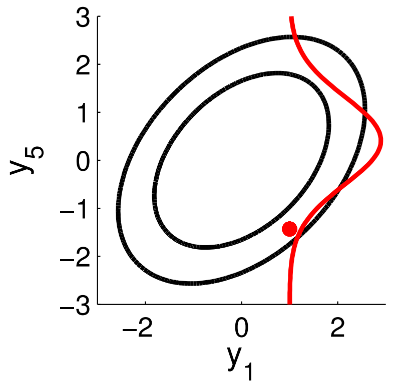
New visualisation

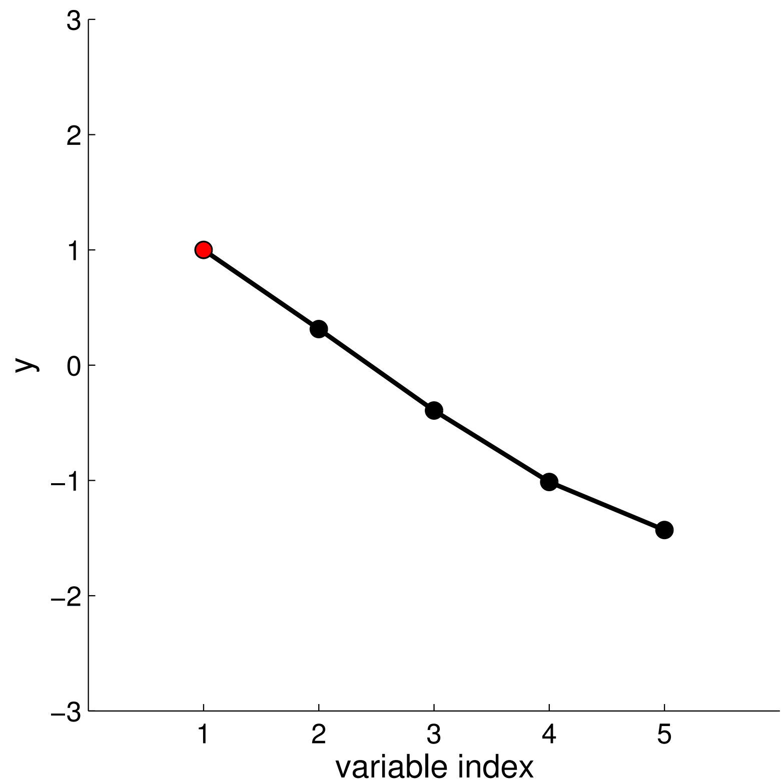




$$\Sigma = \begin{bmatrix} 1 & .9 & .8 & .6 & .4 \\ .9 & 1 & .9 & .8 & .6 \\ .8 & .9 & 1 & .9 & .8 \\ .6 & .8 & .9 & 1 & .9 \\ .4 & .6 & .8 & .9 & 1 \end{bmatrix}$$

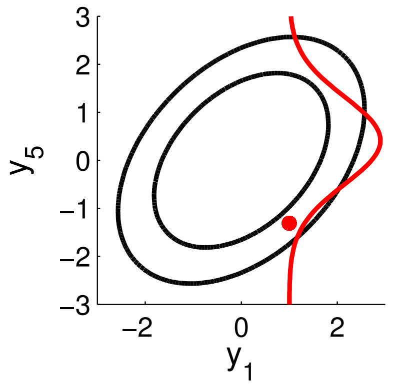
New visualisation

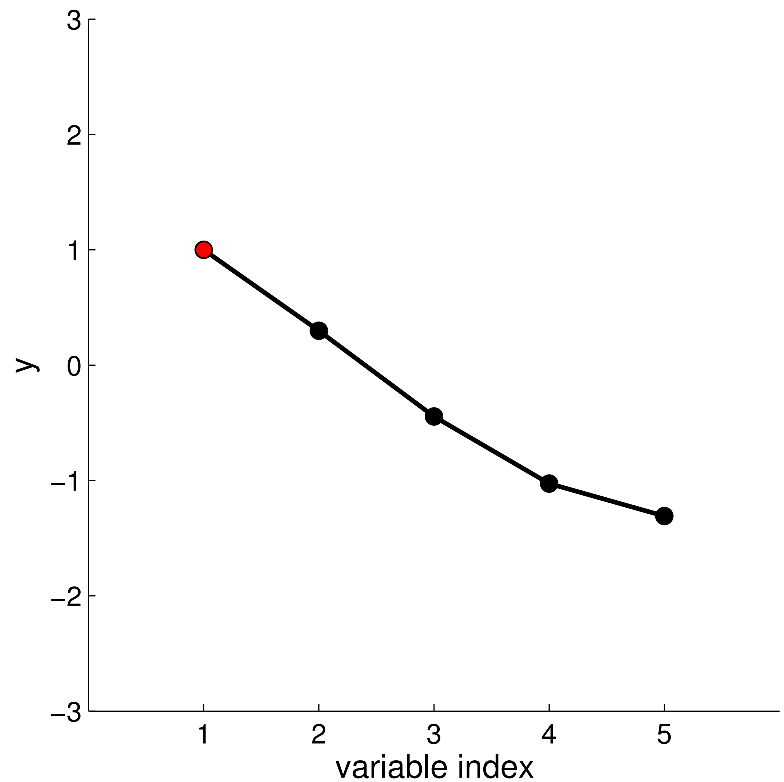




$$\Sigma = \begin{bmatrix} 1 & .9 & .8 & .6 & .4 \\ .9 & 1 & .9 & .8 & .6 \\ .8 & .9 & 1 & .9 & .8 \\ .6 & .8 & .9 & 1 & .9 \\ .4 & .6 & .8 & .9 & 1 \end{bmatrix}$$

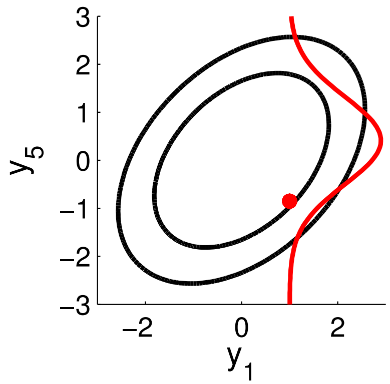
New visualisation



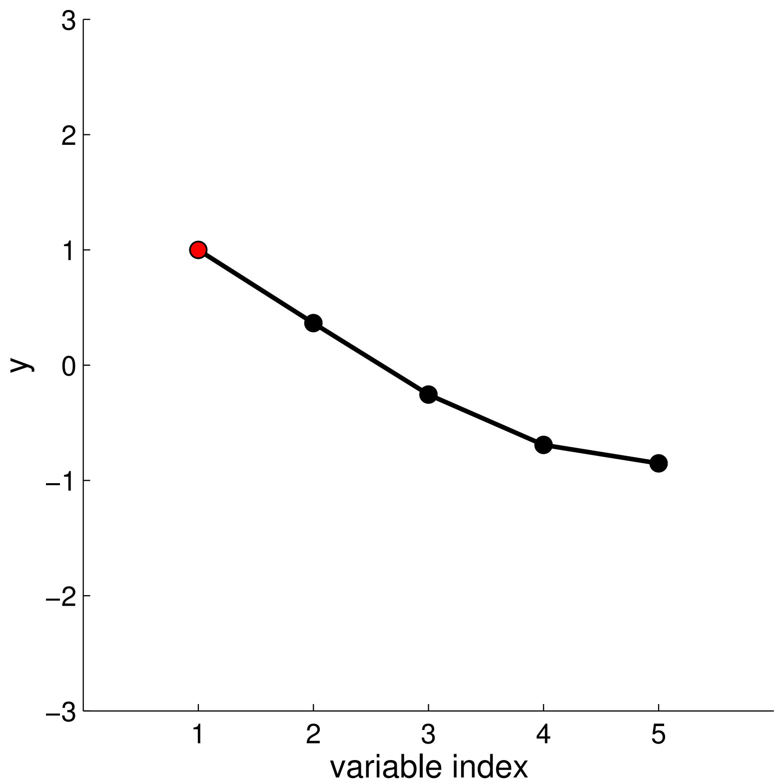


$$\Sigma = \begin{bmatrix} 1 & .9 & .8 & .6 & .4 \\ .9 & 1 & .9 & .8 & .6 \\ .8 & .9 & 1 & .9 & .8 \\ .6 & .8 & .9 & 1 & .9 \\ .4 & .6 & .8 & .9 & 1 \end{bmatrix}$$

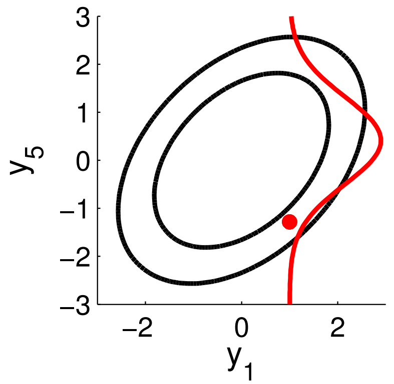
New visualisation

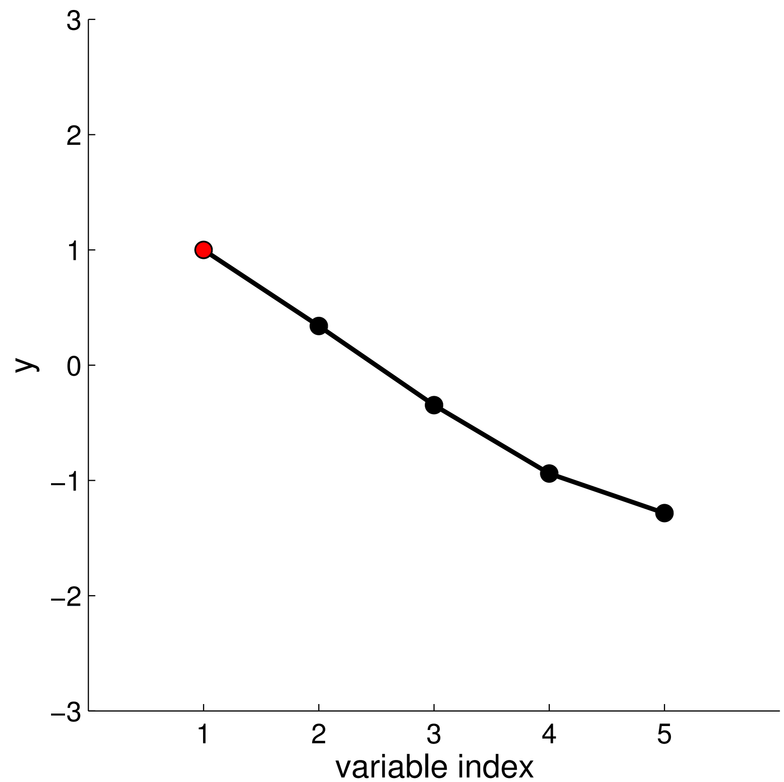


$$\Sigma = \begin{bmatrix} 1 & .9 & .8 & .6 & .4 \\ .9 & 1 & .9 & .8 & .6 \\ .8 & .9 & 1 & .9 & .8 \\ .6 & .8 & .9 & 1 & .9 \\ .4 & .6 & .8 & .9 & 1 \end{bmatrix}$$



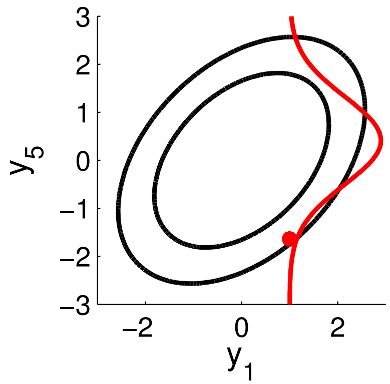
New visualisation

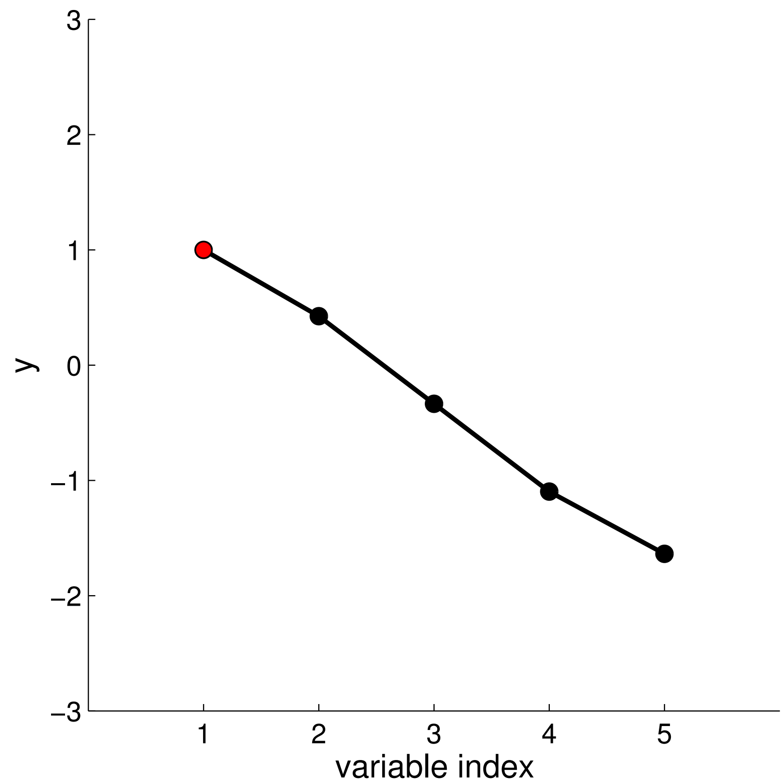




$$\Sigma = \begin{bmatrix} 1 & .9 & .8 & .6 & .4 \\ .9 & 1 & .9 & .8 & .6 \\ .8 & .9 & 1 & .9 & .8 \\ .6 & .8 & .9 & 1 & .9 \\ .4 & .6 & .8 & .9 & 1 \end{bmatrix}$$

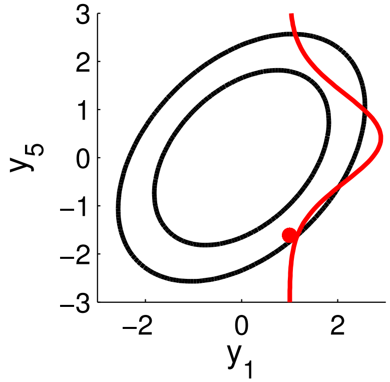
New visualisation

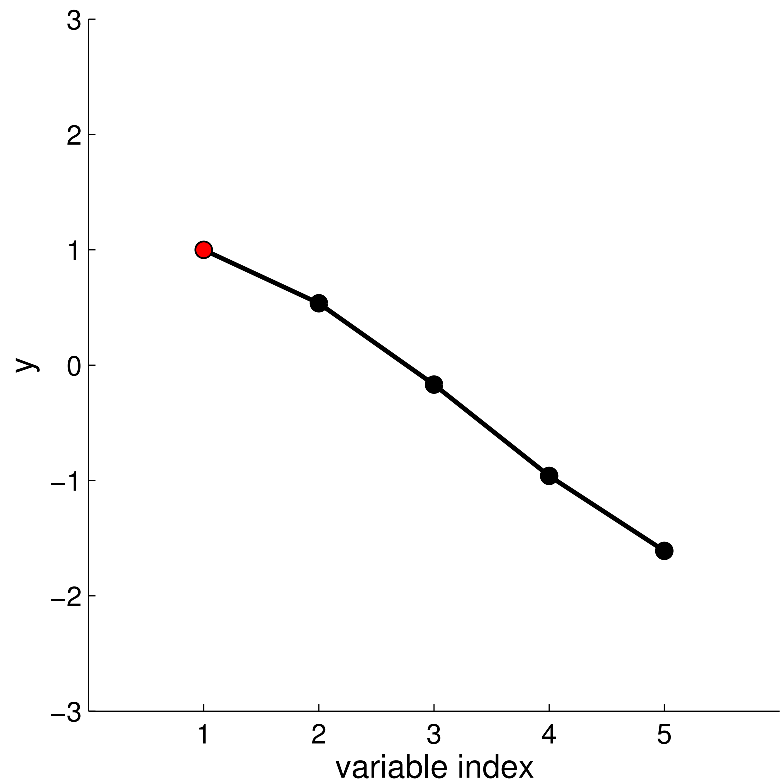




$$\Sigma = \begin{bmatrix} 1 & .9 & .8 & .6 & .4 \\ .9 & 1 & .9 & .8 & .6 \\ .8 & .9 & 1 & .9 & .8 \\ .6 & .8 & .9 & 1 & .9 \\ .4 & .6 & .8 & .9 & 1 \end{bmatrix}$$

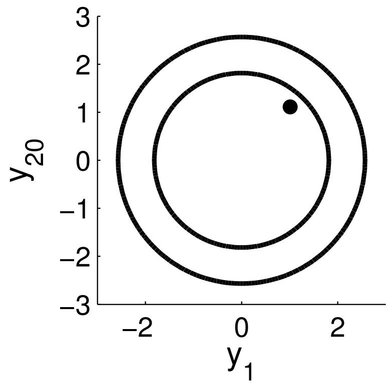
New visualisation

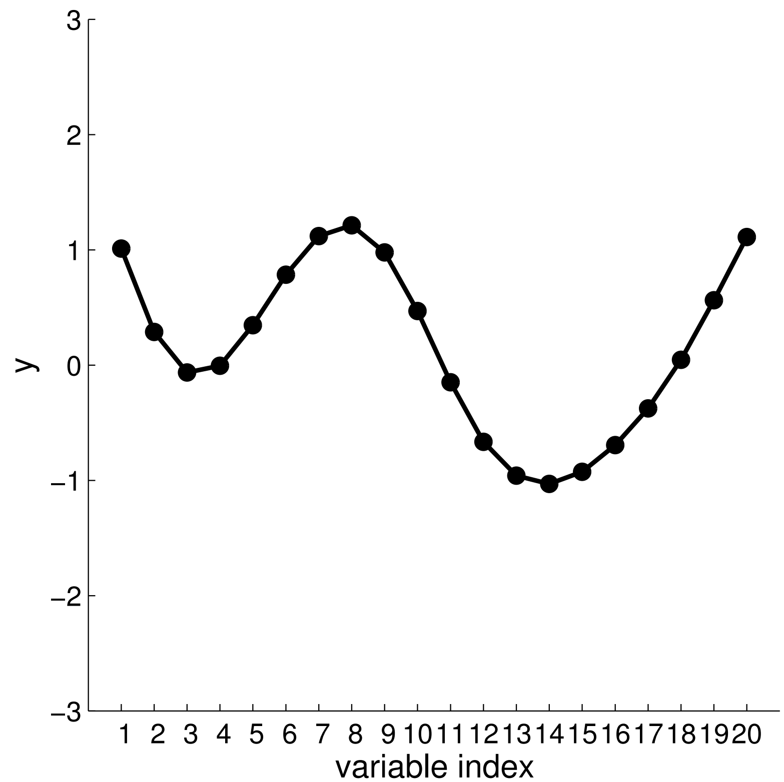




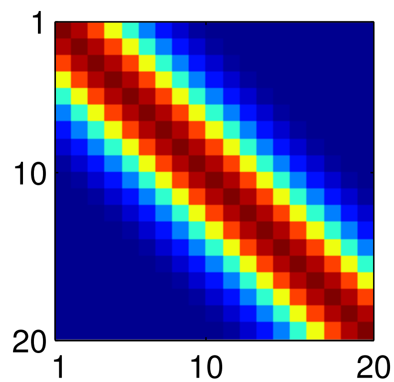
$$\Sigma = \begin{bmatrix} 1 & .9 & .8 & .6 & .4 \\ .9 & 1 & .9 & .8 & .6 \\ .8 & .9 & 1 & .9 & .8 \\ .6 & .8 & .9 & 1 & .9 \\ .4 & .6 & .8 & .9 & 1 \end{bmatrix}$$

New visualisation

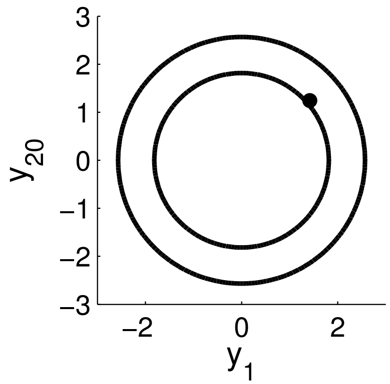




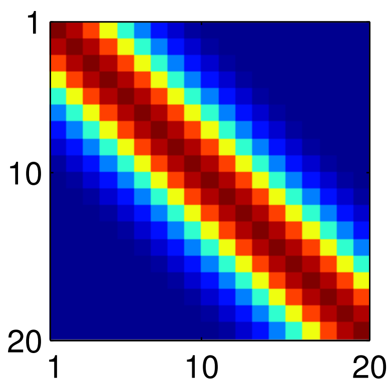
$$\Sigma =$$

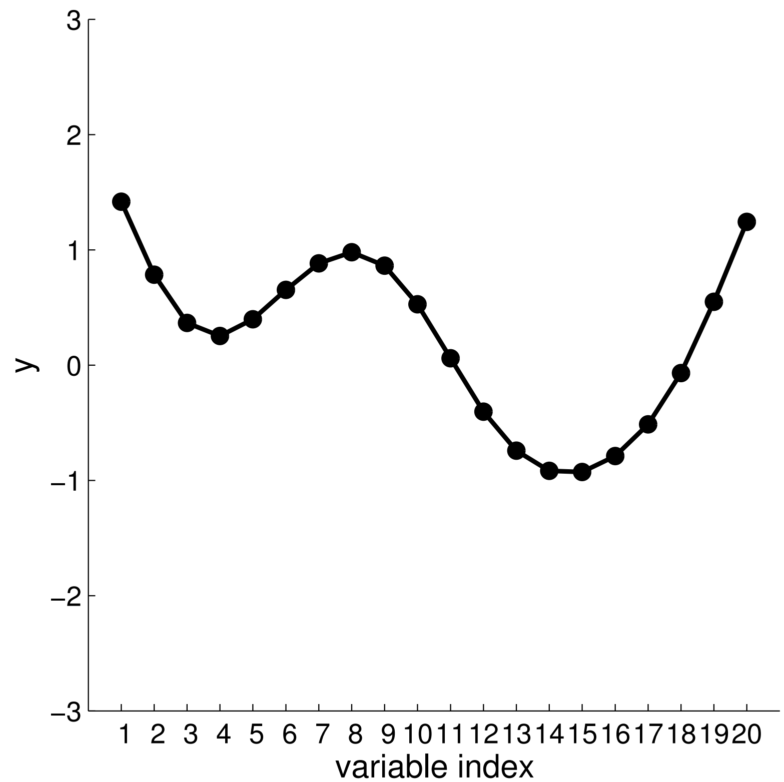


New visualisation

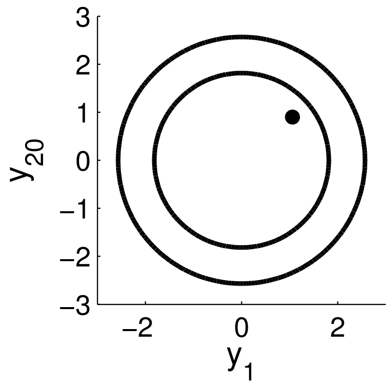


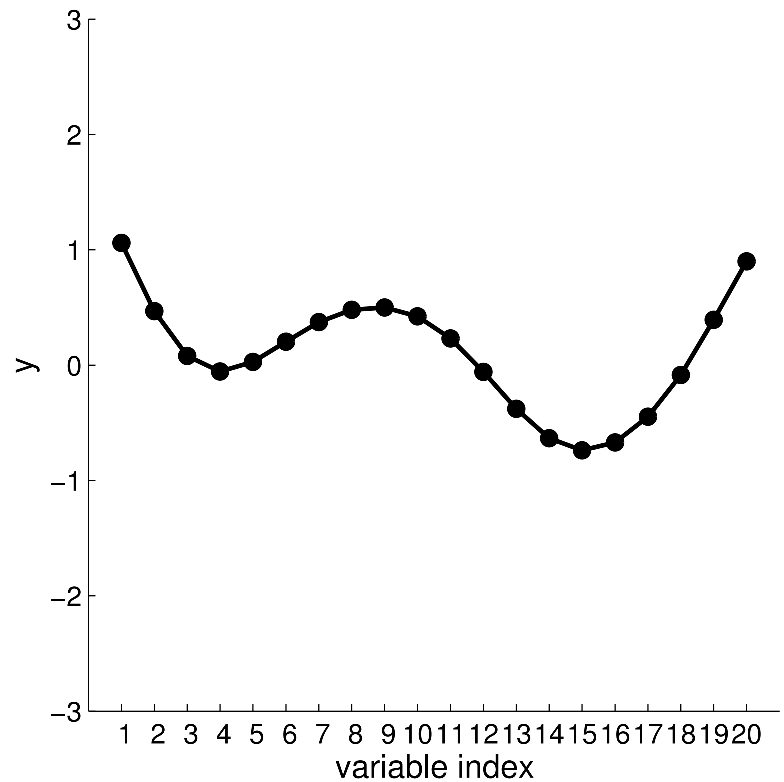
$\Sigma =$



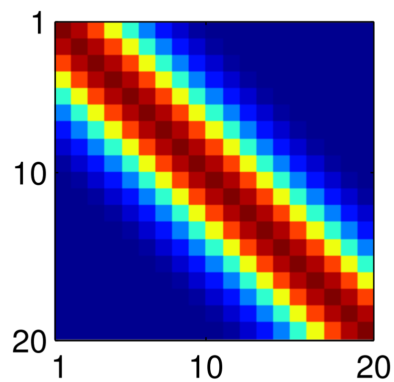


New visualisation

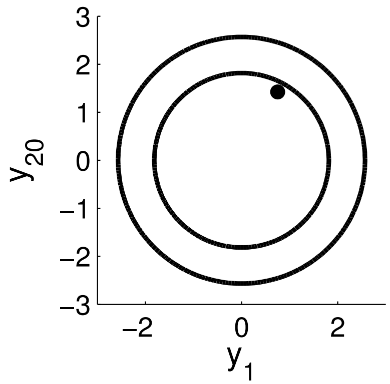


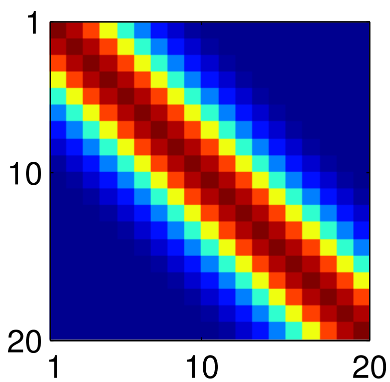


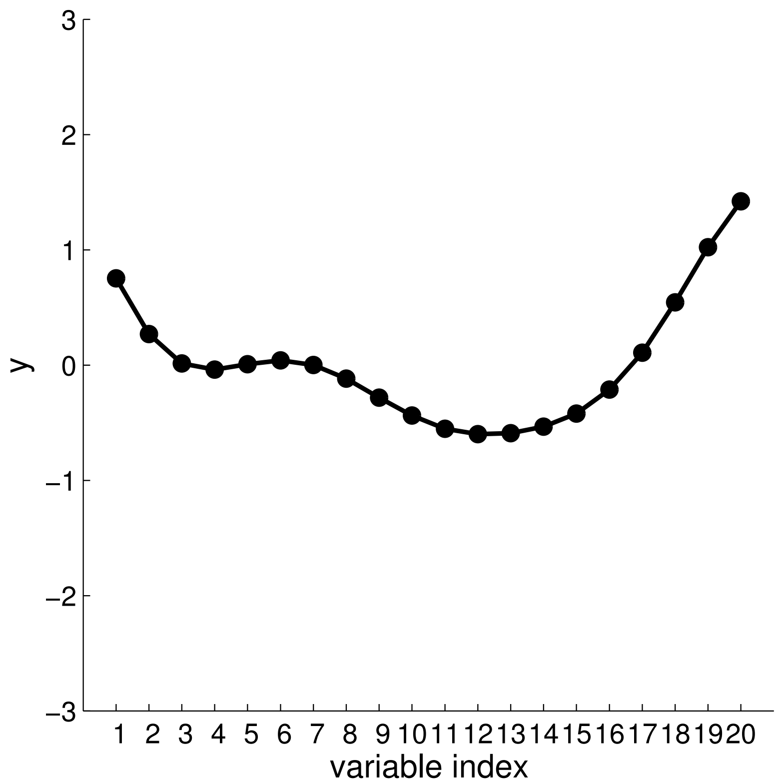
$\Sigma =$



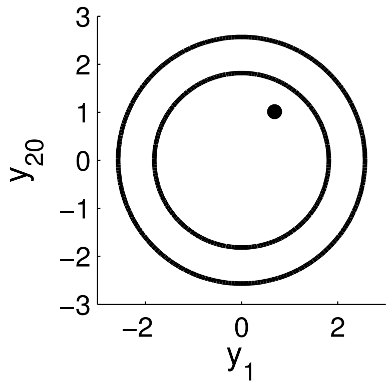
New visualisation

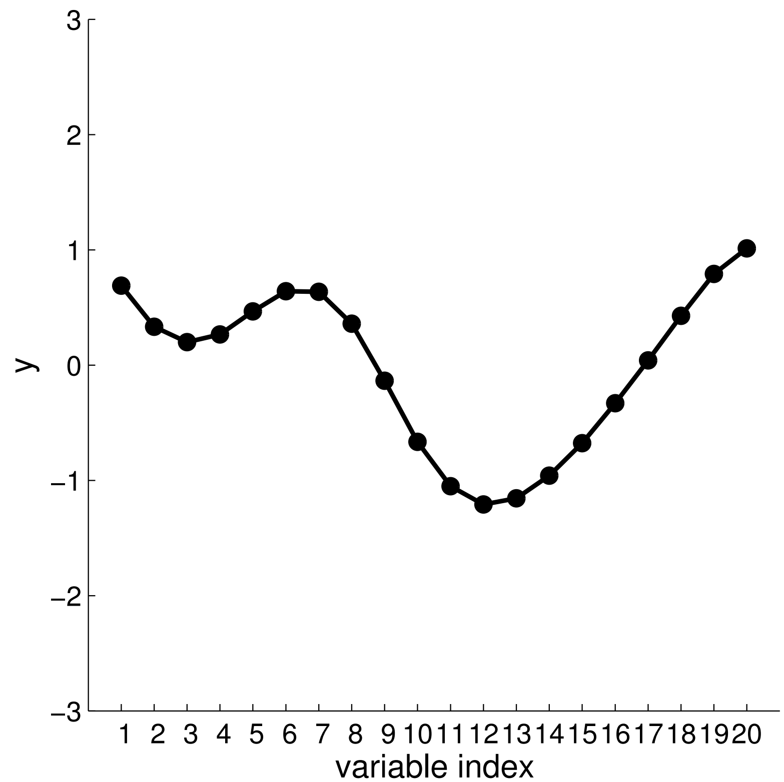


$\Sigma =$ 

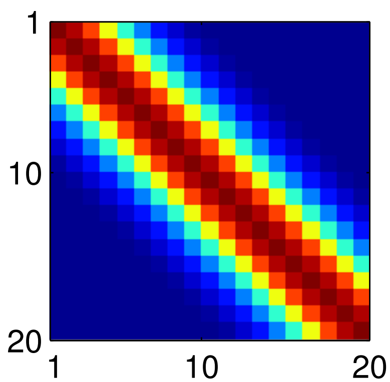


New visualisation

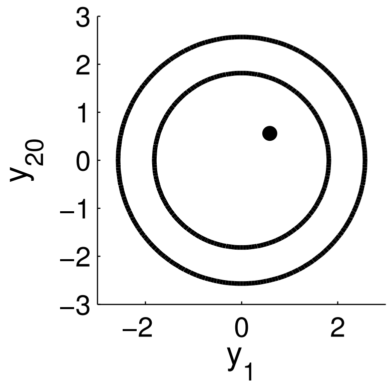


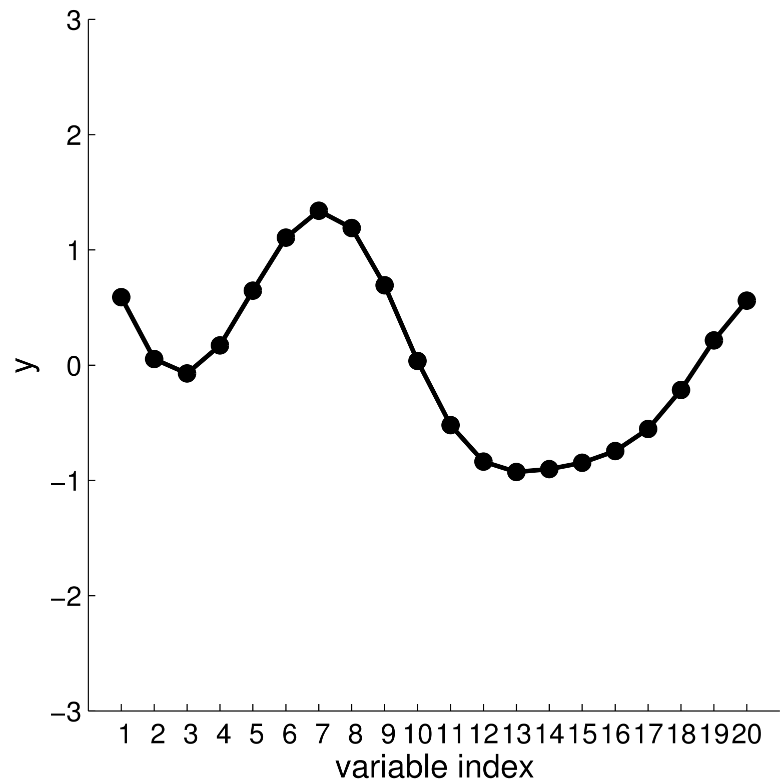


$\Sigma =$

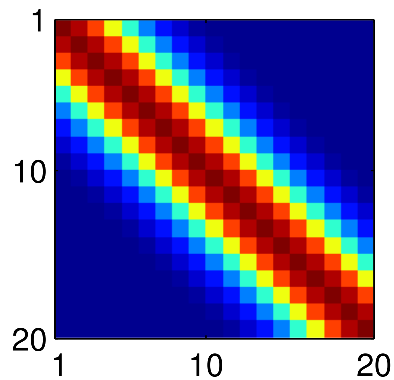


New visualisation

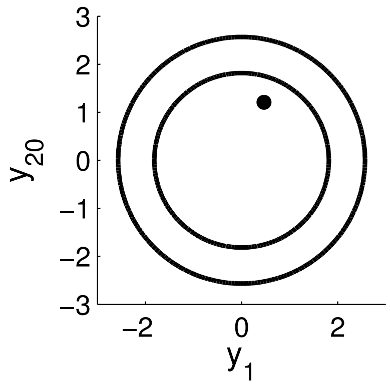




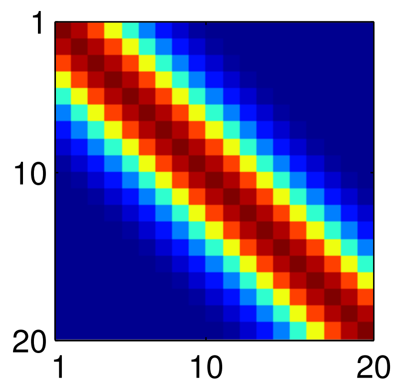
$\Sigma =$

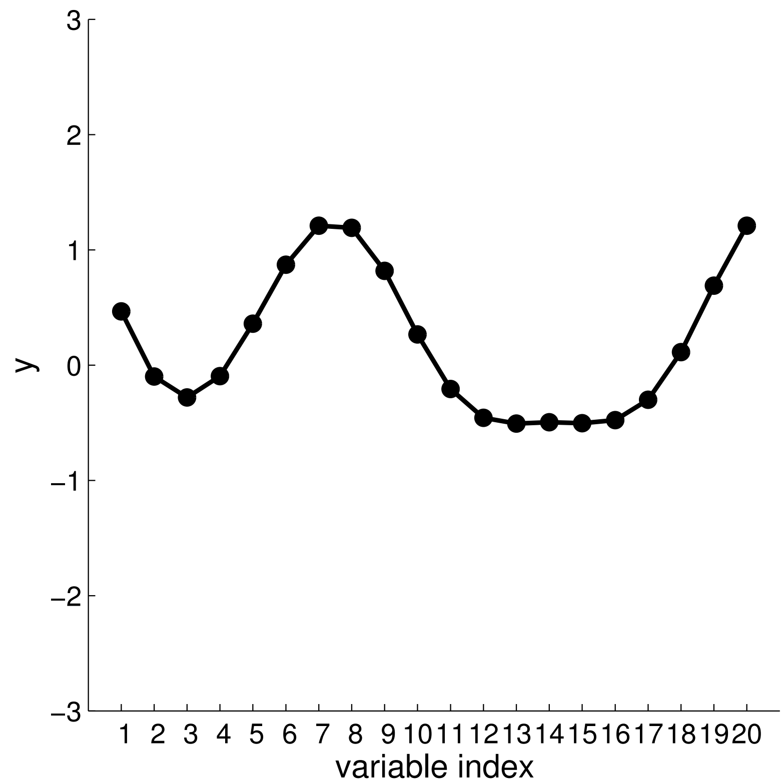


New visualisation

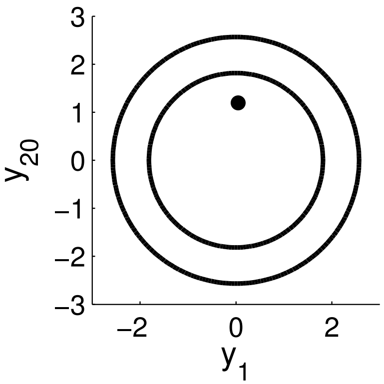


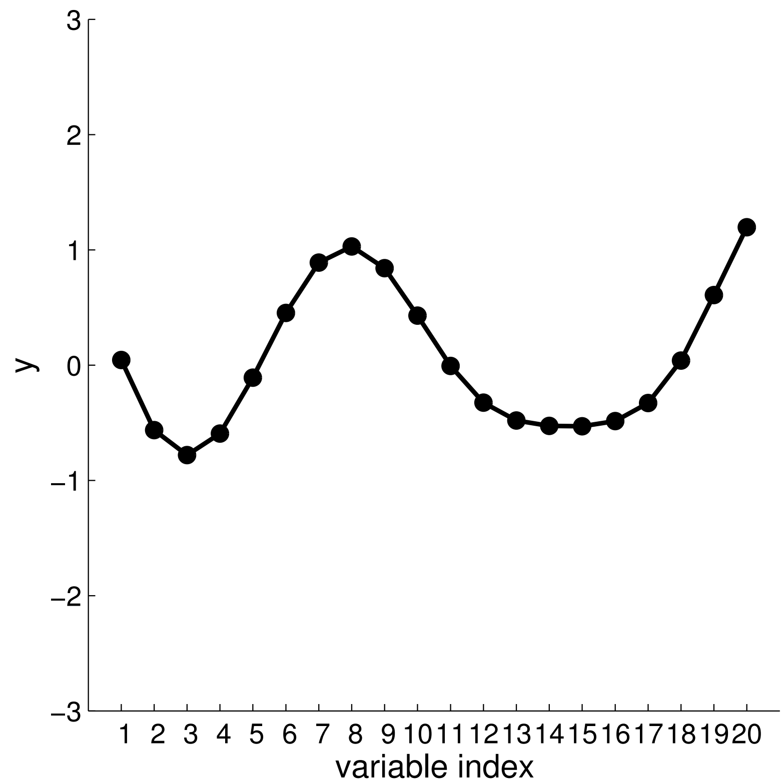
$\Sigma =$



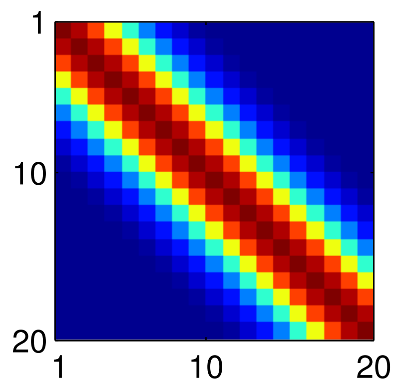


New visualisation

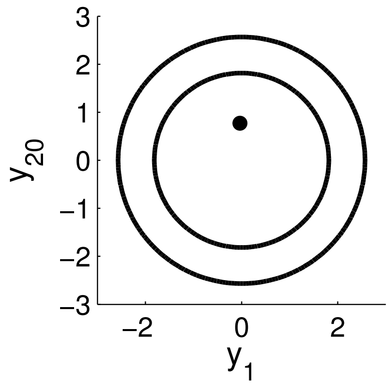


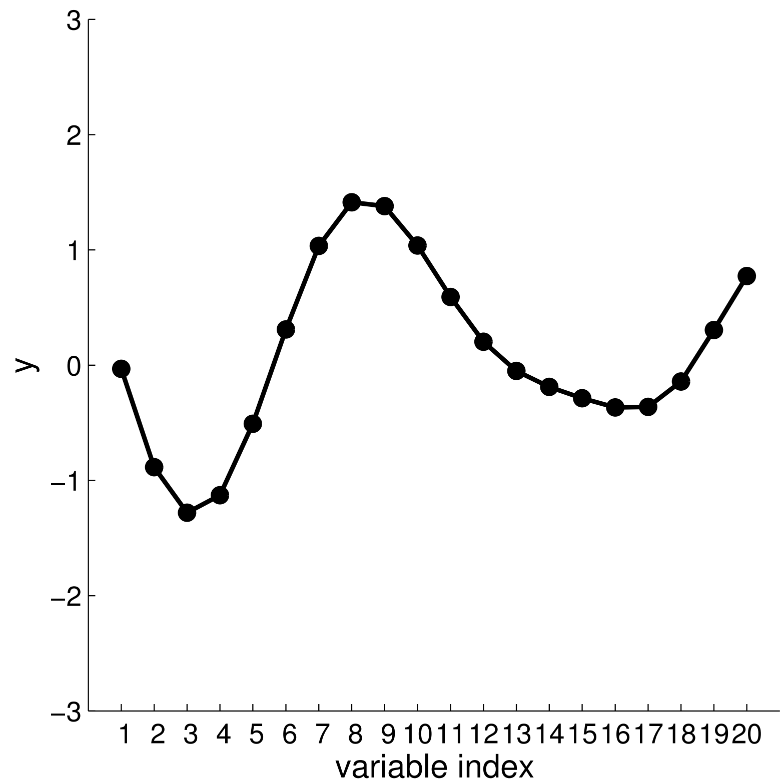


$$\Sigma =$$

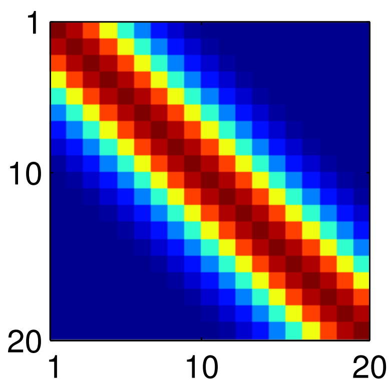


New visualisation

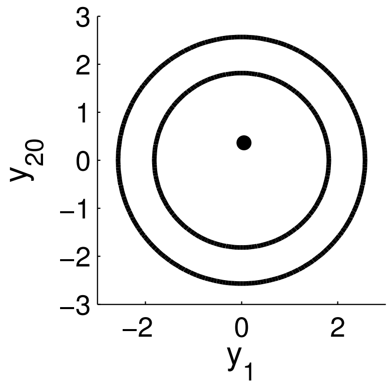


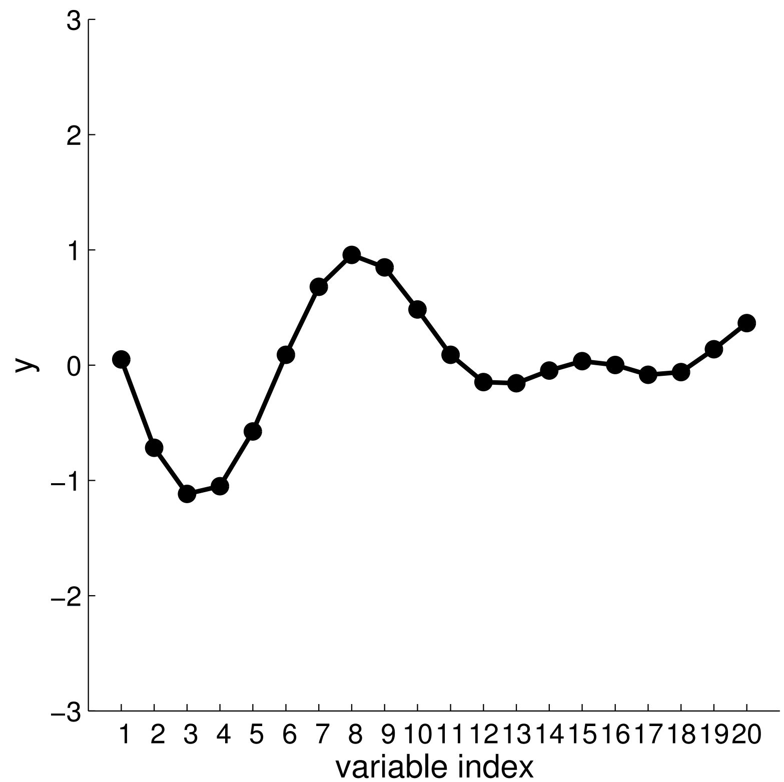


$$\Sigma =$$

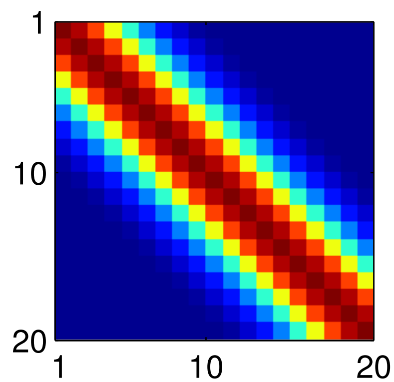


New visualisation

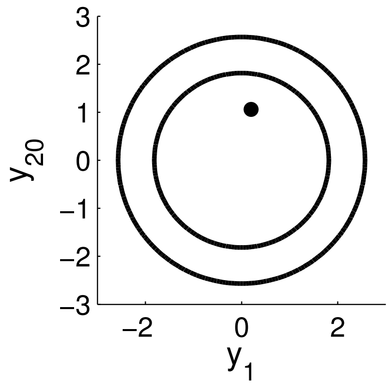


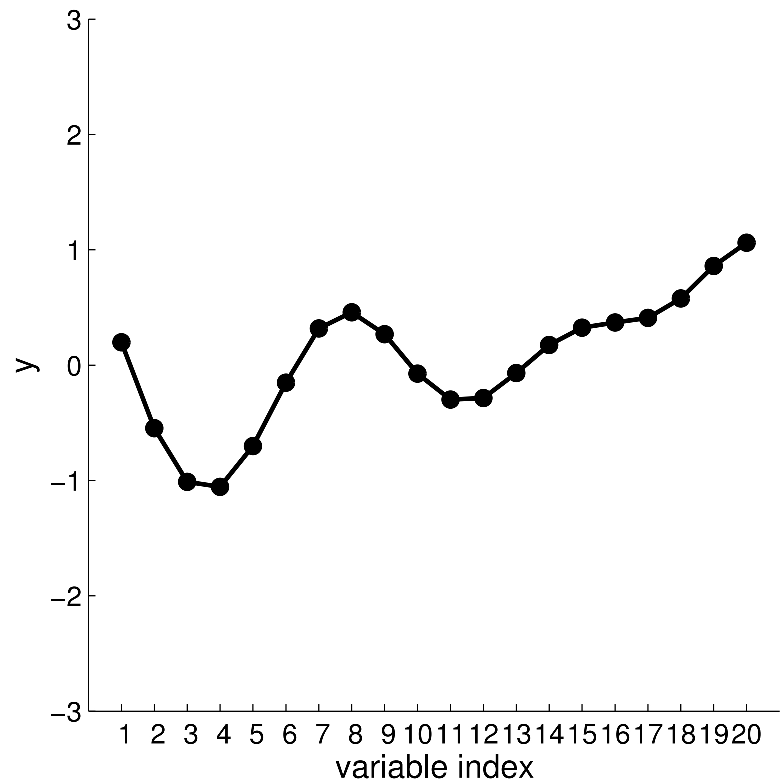


$\Sigma =$

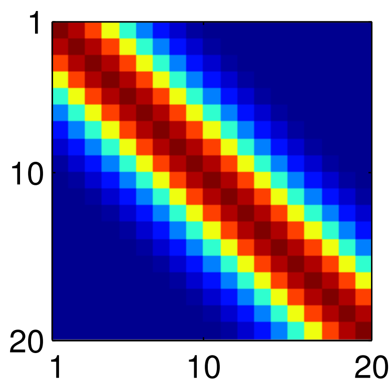


New visualisation

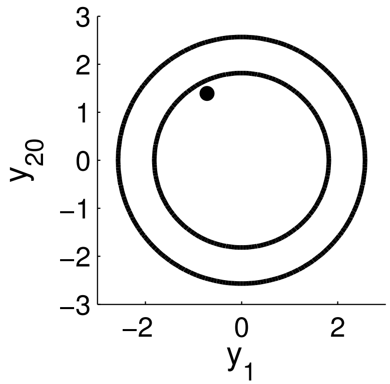


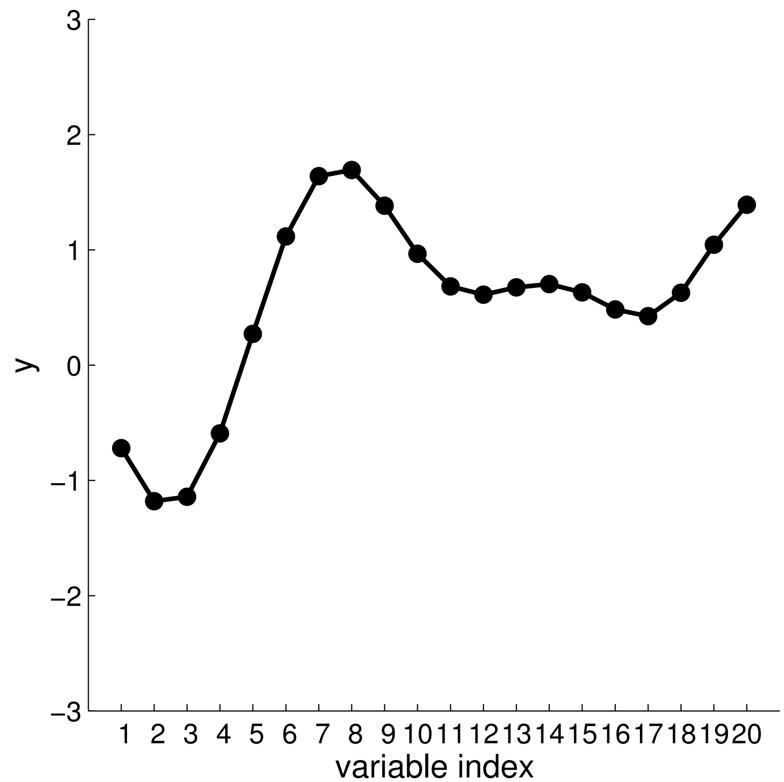


$\Sigma =$

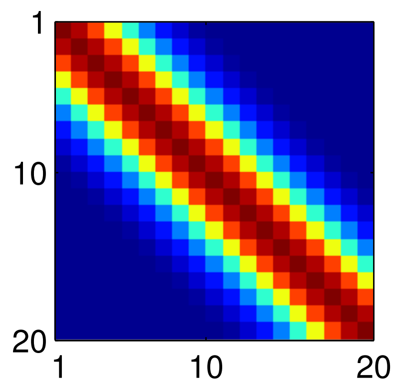


New visualisation

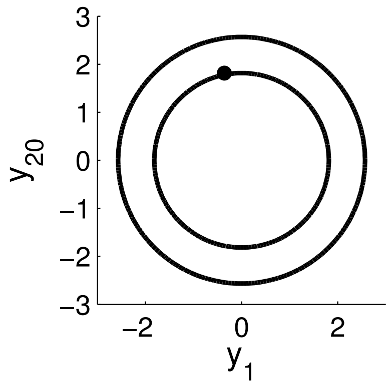


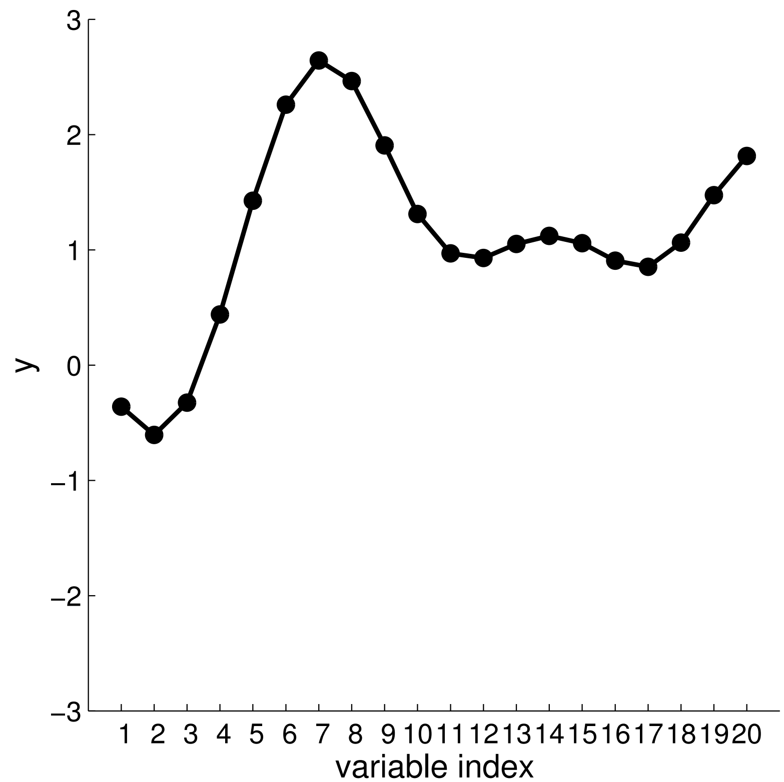


$\Sigma =$

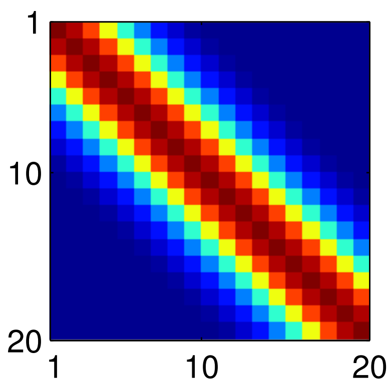


New visualisation

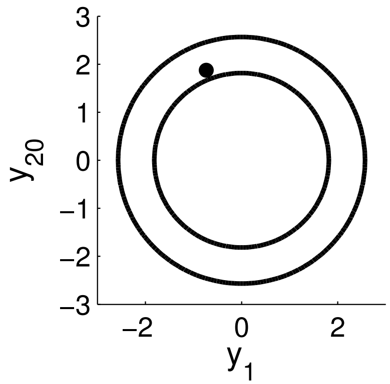




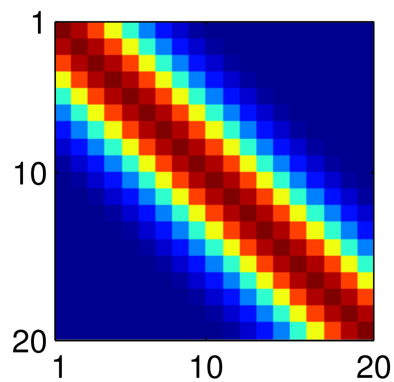
$\Sigma =$



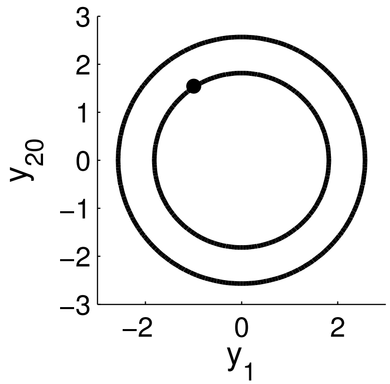
New visualisation

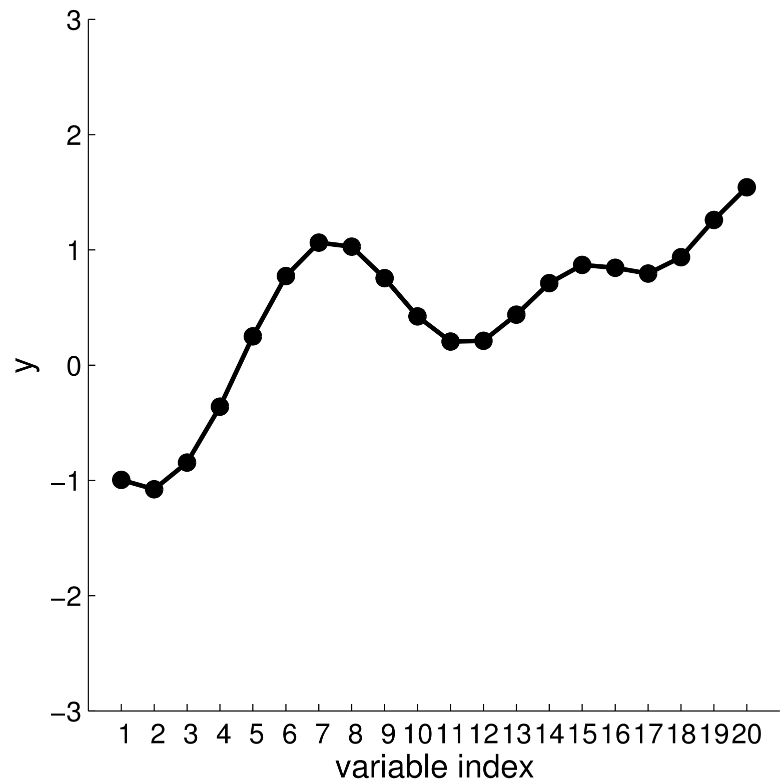


$$\Sigma =$$

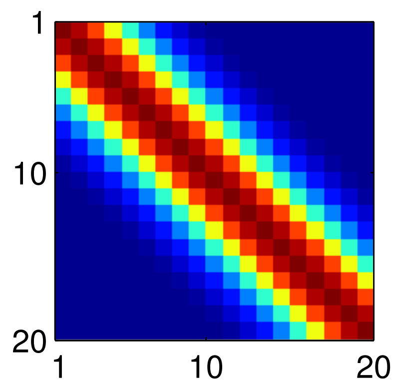


New visualisation

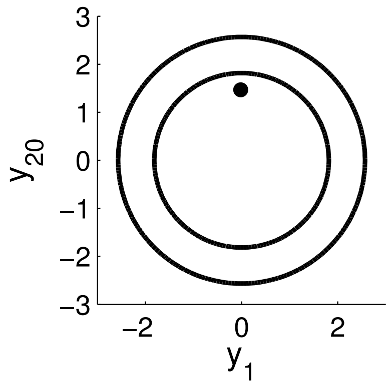


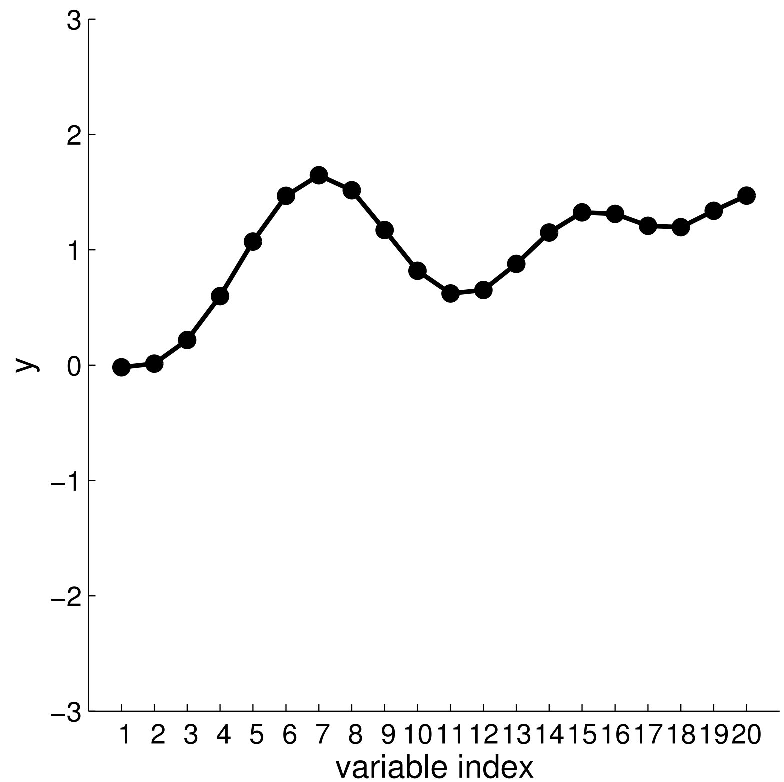


$\Sigma =$

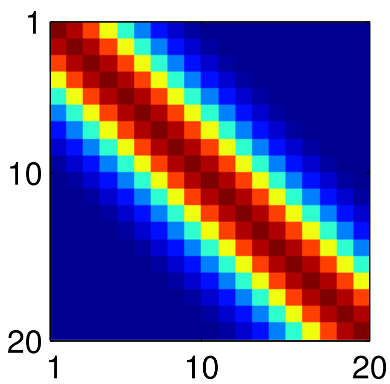


New visualisation

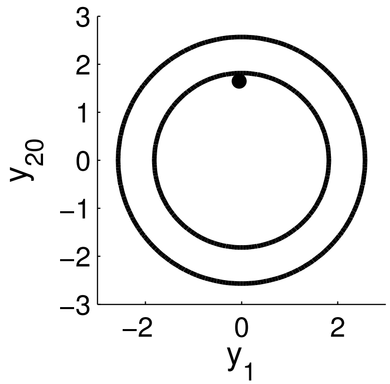


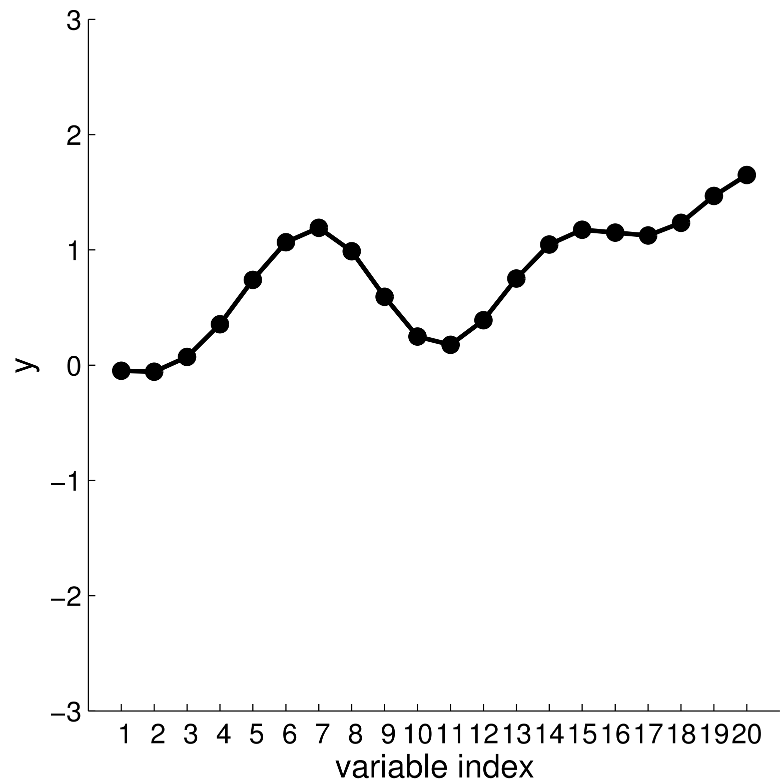


$\Sigma =$

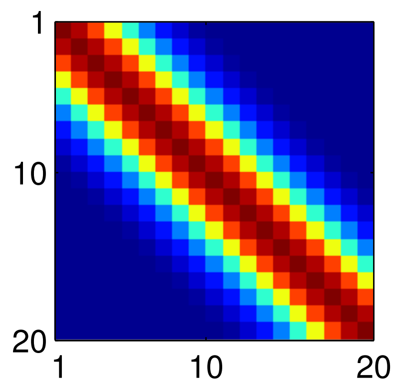


New visualisation

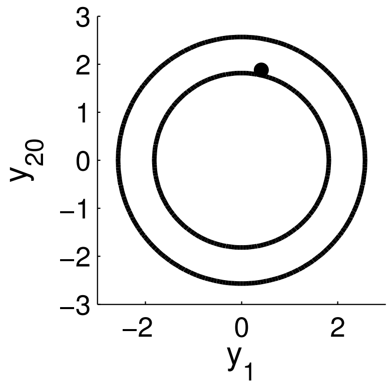


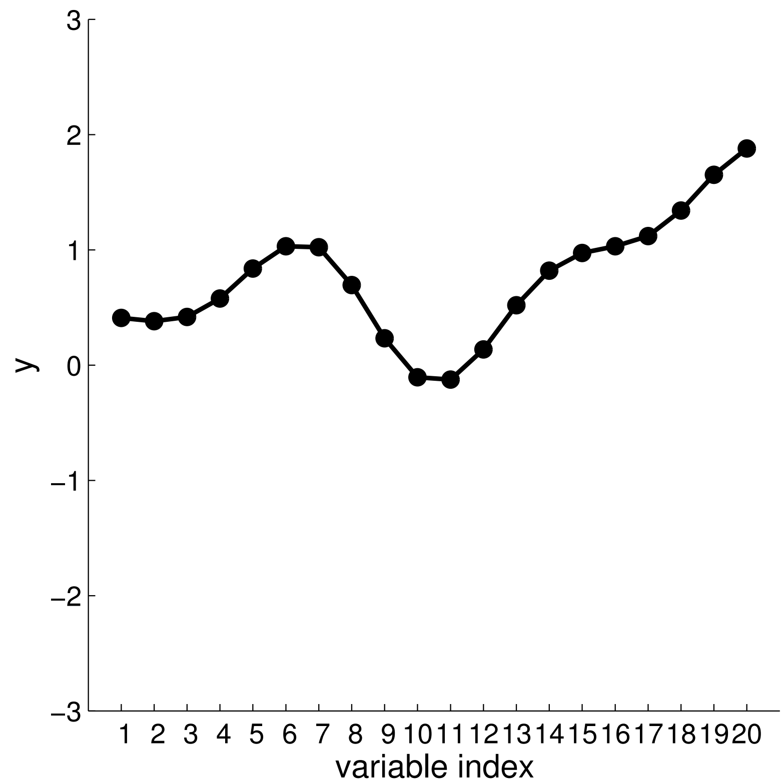


$$\Sigma =$$

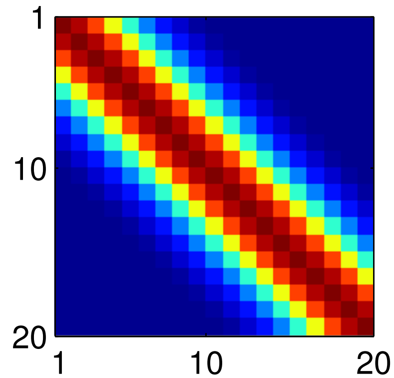


New visualisation

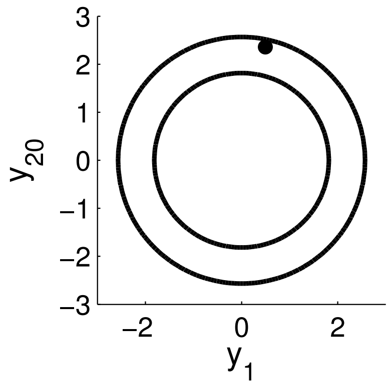


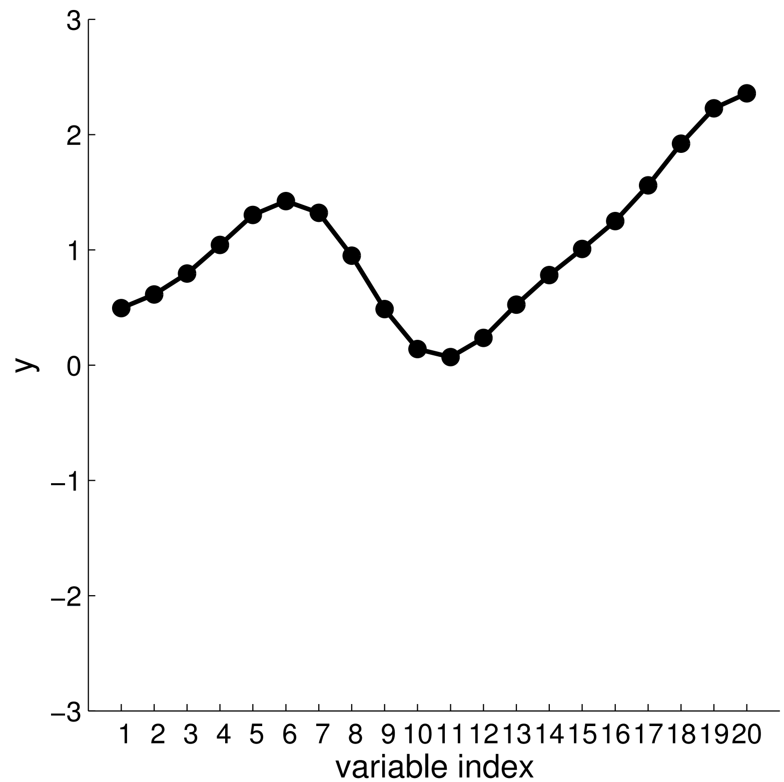


$\Sigma =$

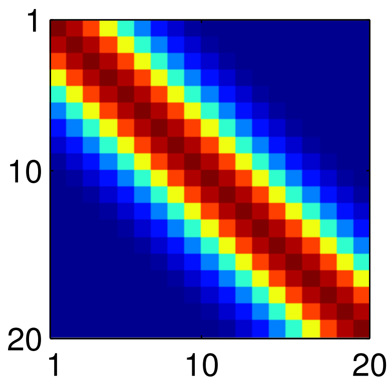


New visualisation

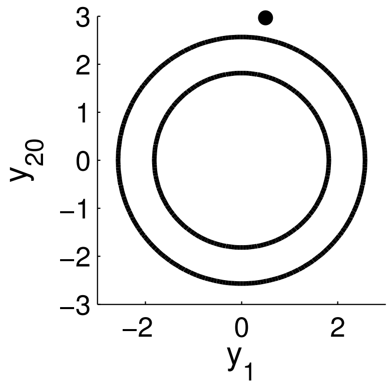


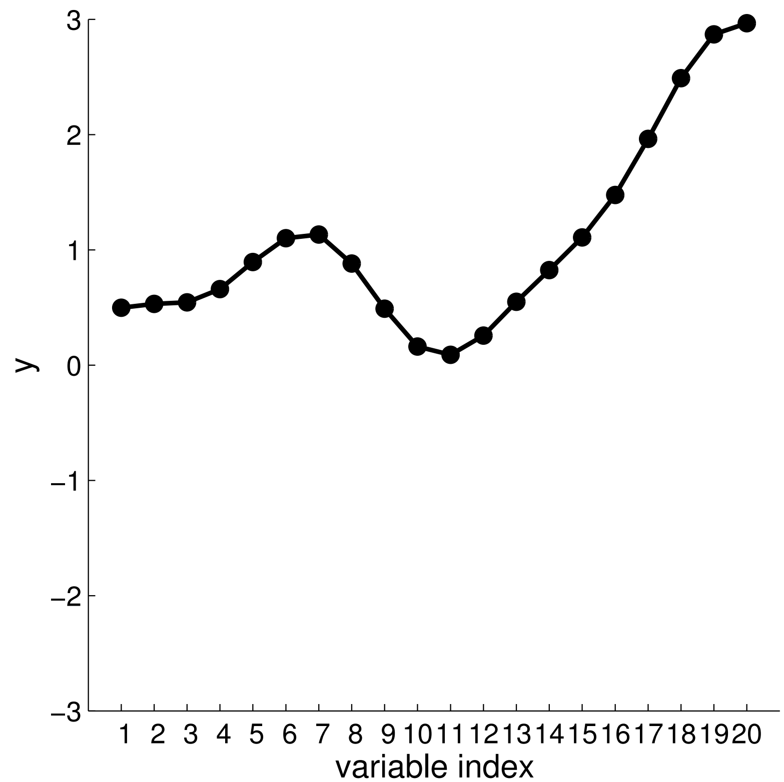


$$\Sigma =$$

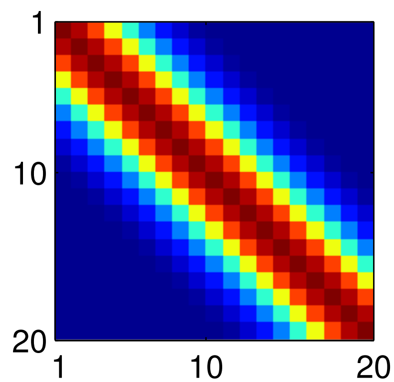


New visualisation

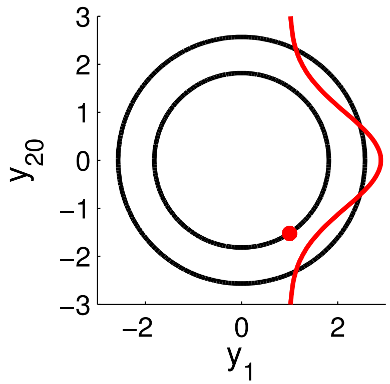




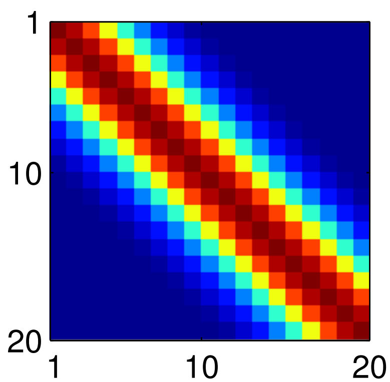
$\Sigma =$

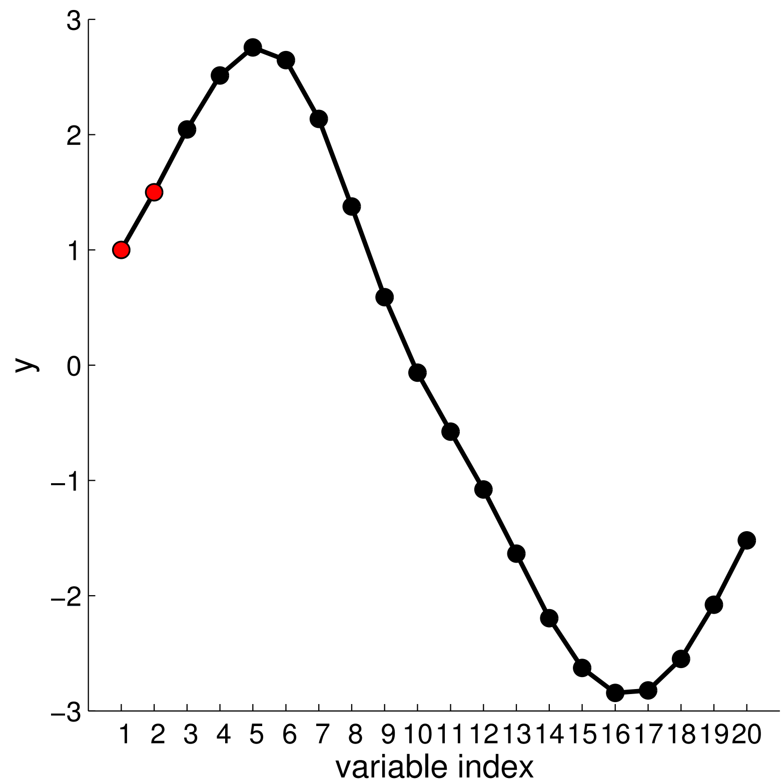


New visualisation

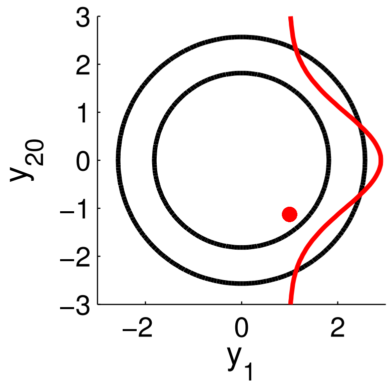


$\Sigma =$

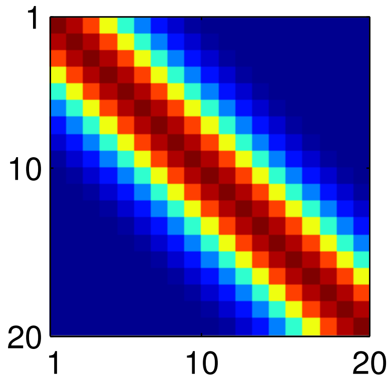


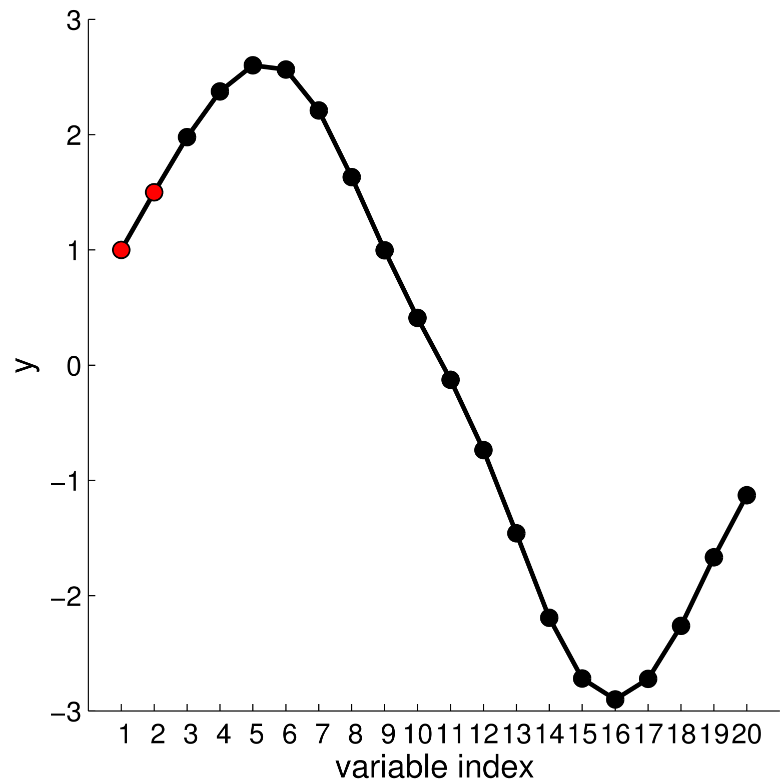


New visualisation

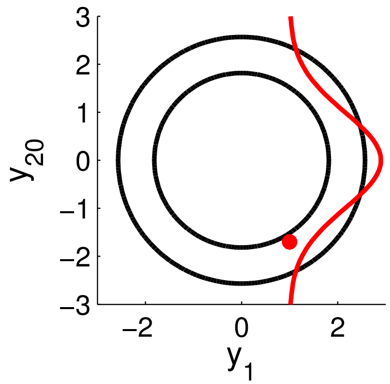


$$\Sigma =$$

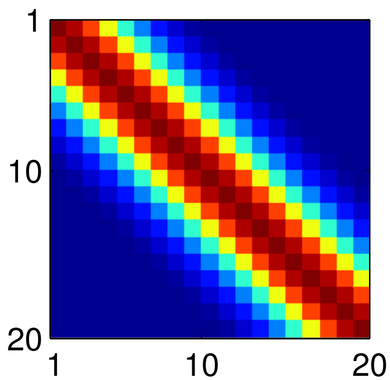


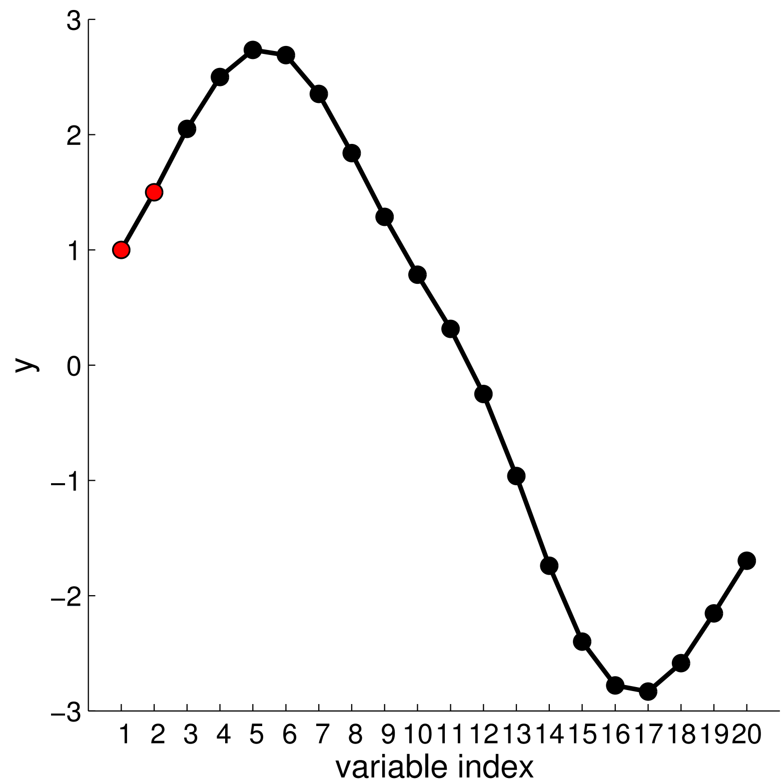


New visualisation

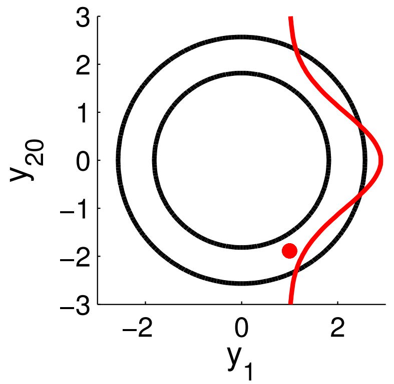


$\Sigma =$

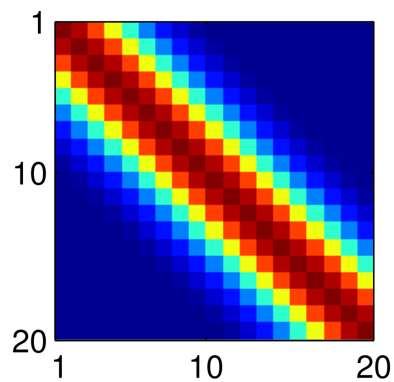


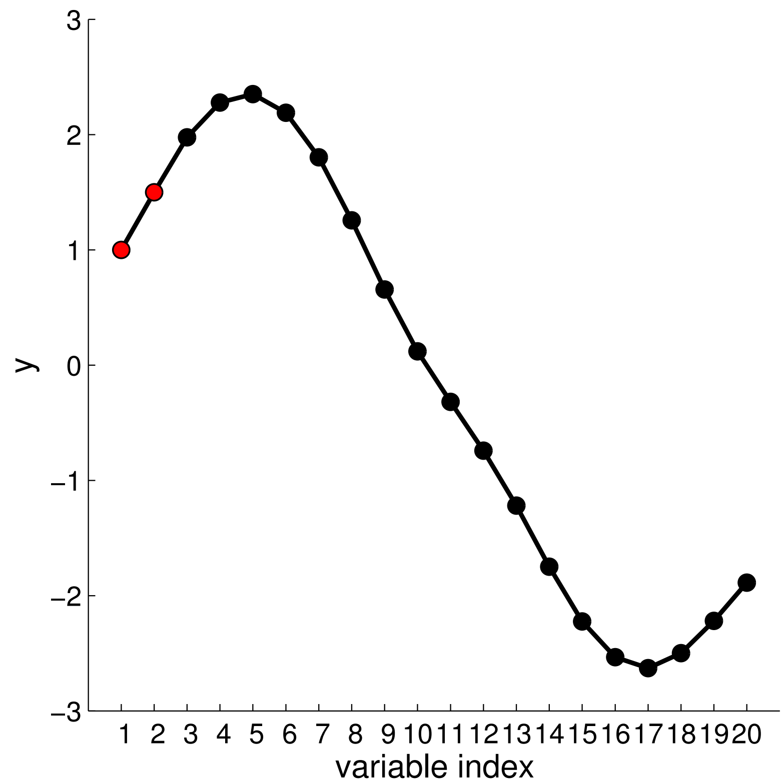


New visualisation

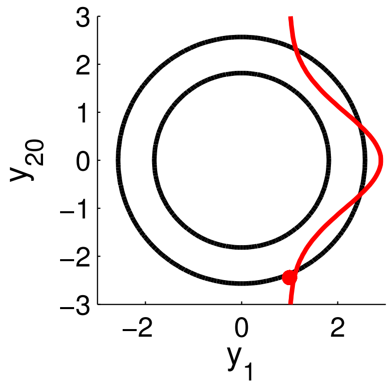


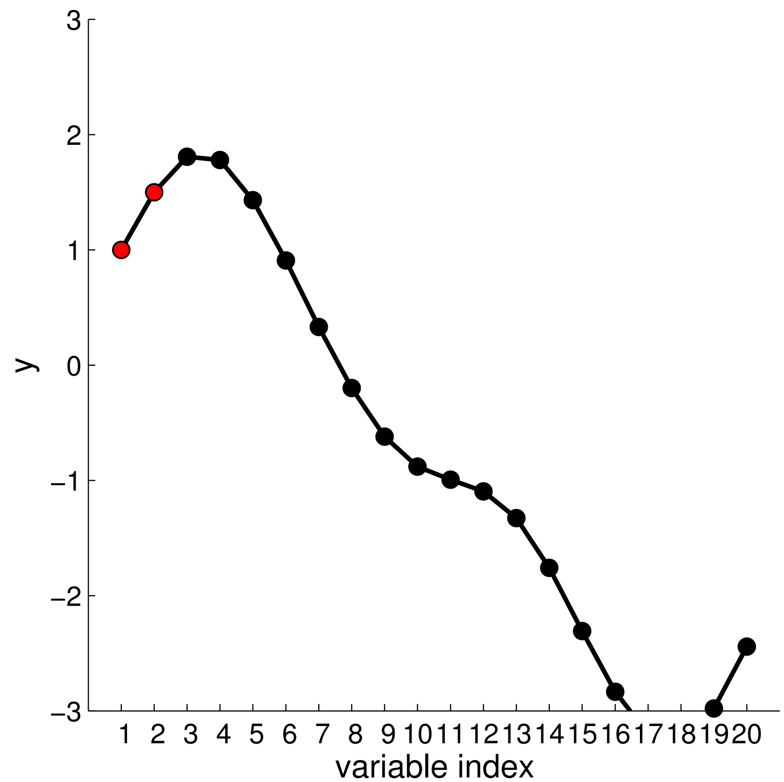
$\Sigma =$



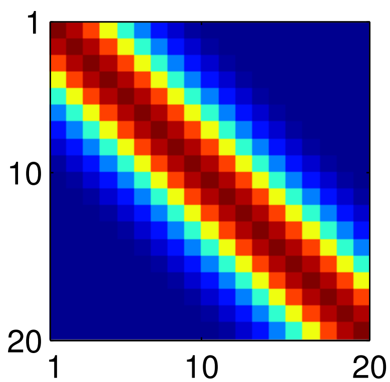


New visualisation

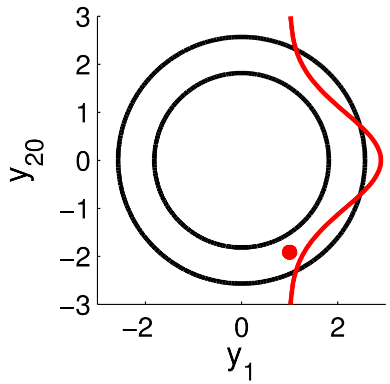


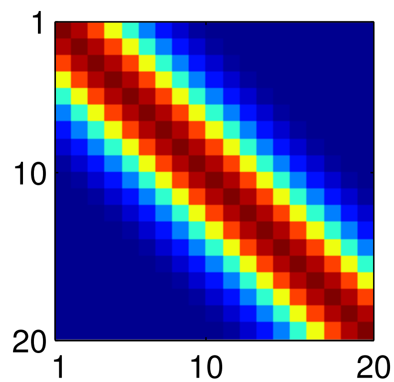


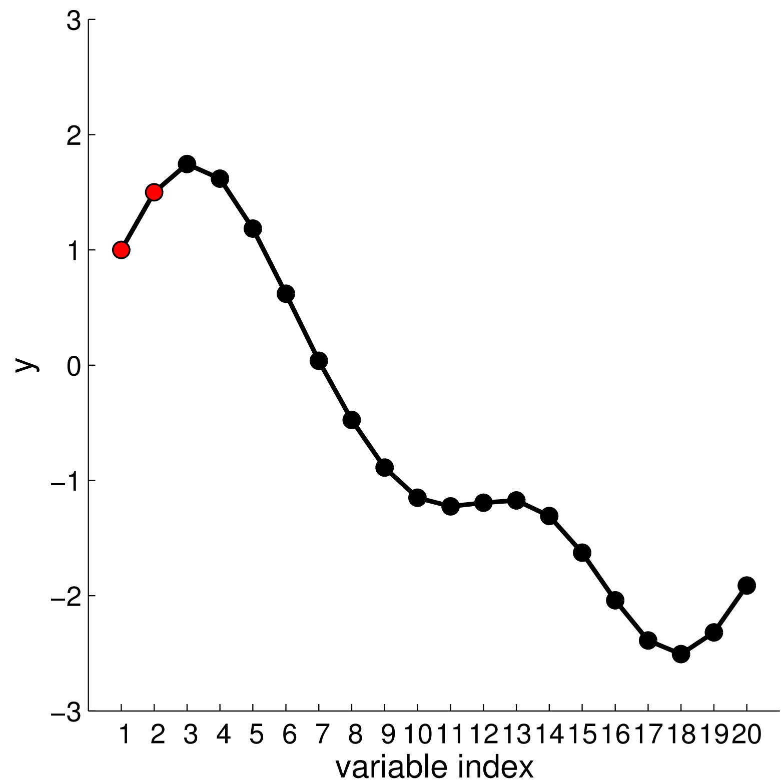
$\Sigma =$



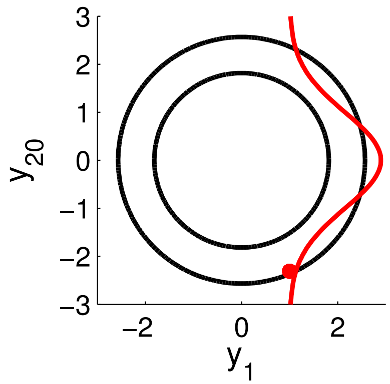
New visualisation

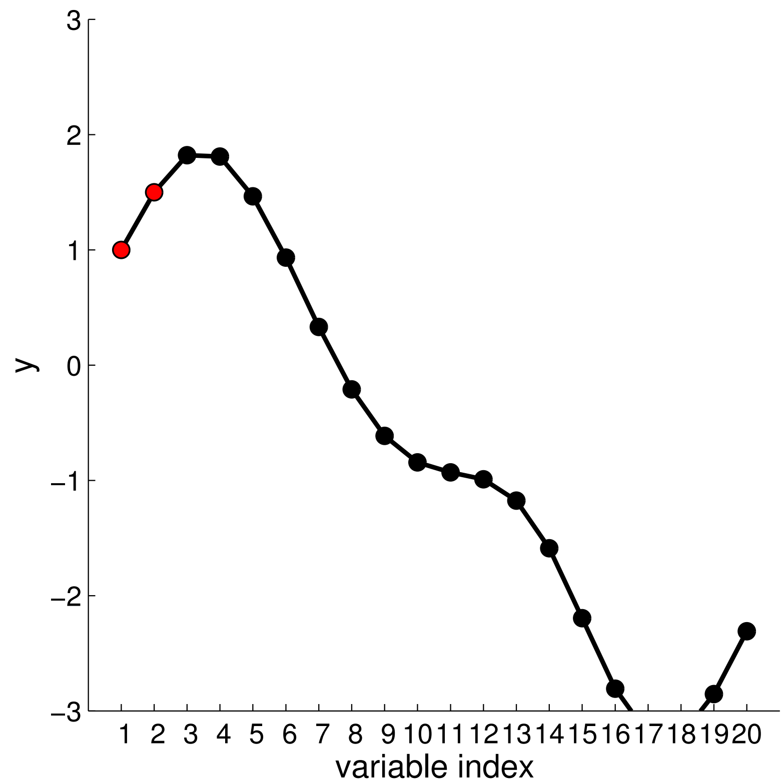


$\Sigma =$ 

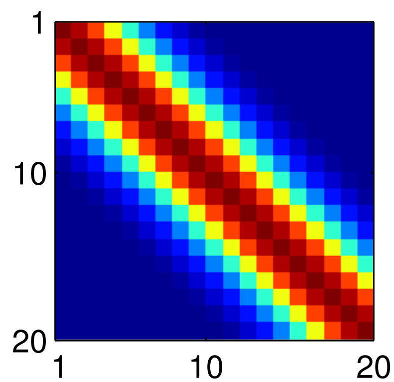


New visualisation

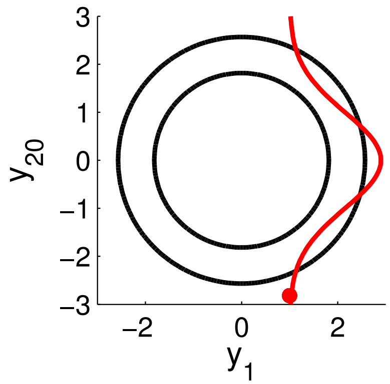


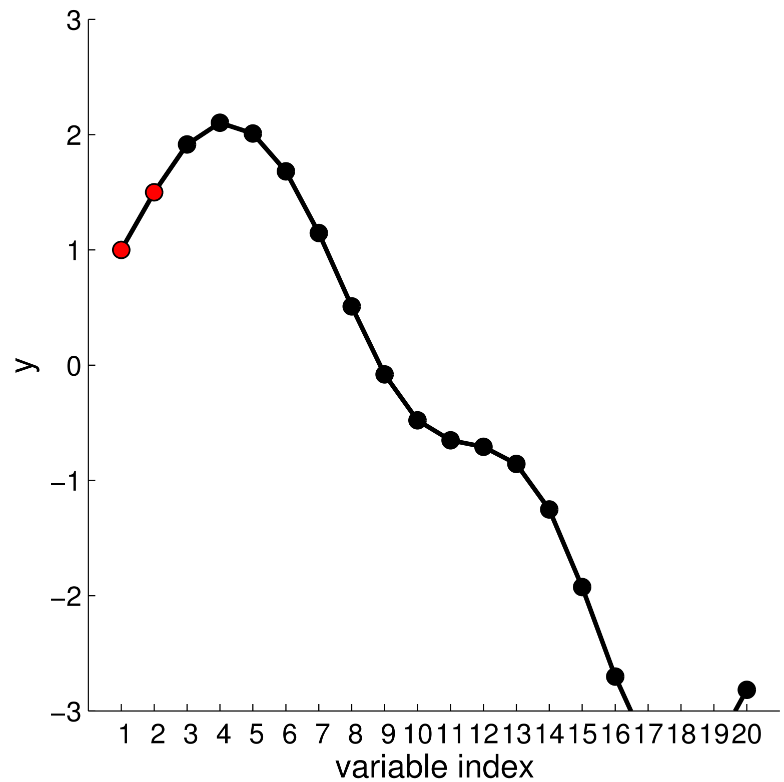


$\Sigma =$

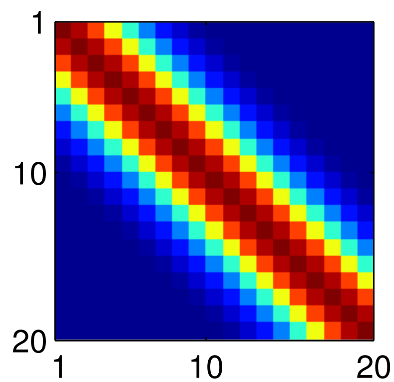


New visualisation

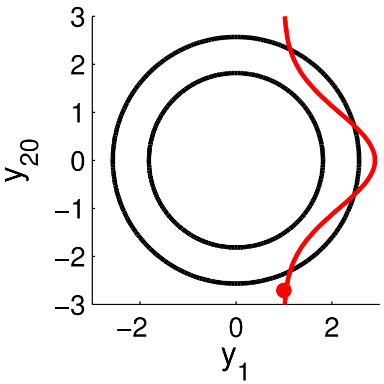


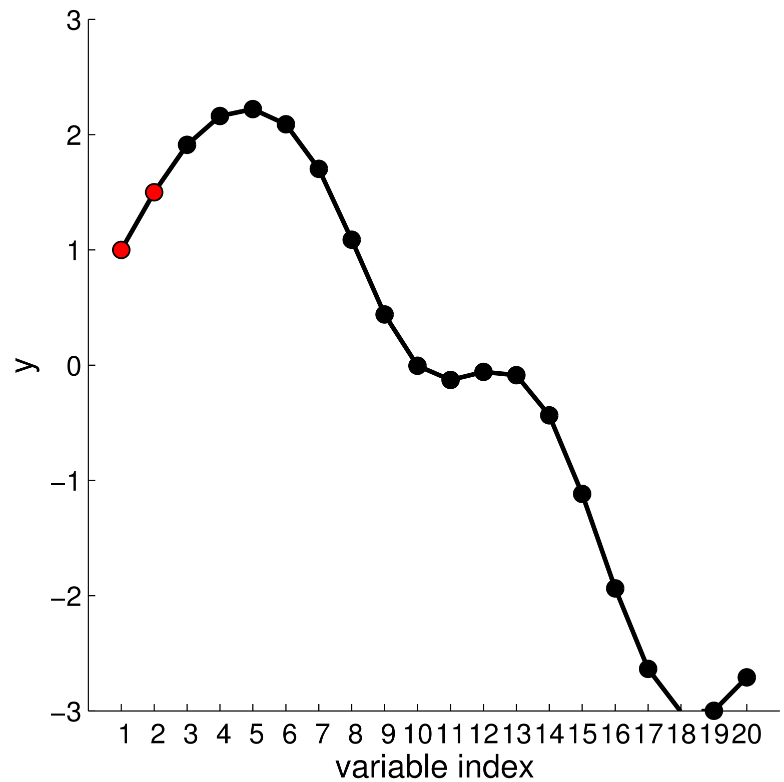


$\Sigma =$

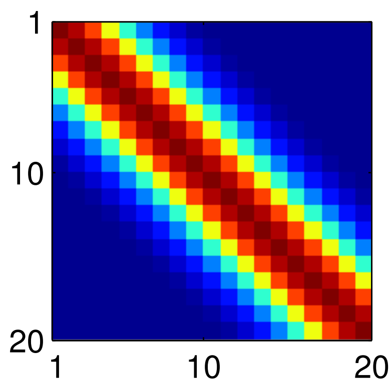


New visualisation

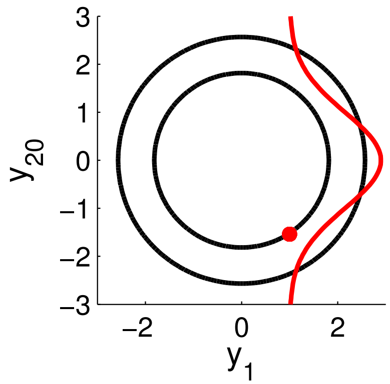


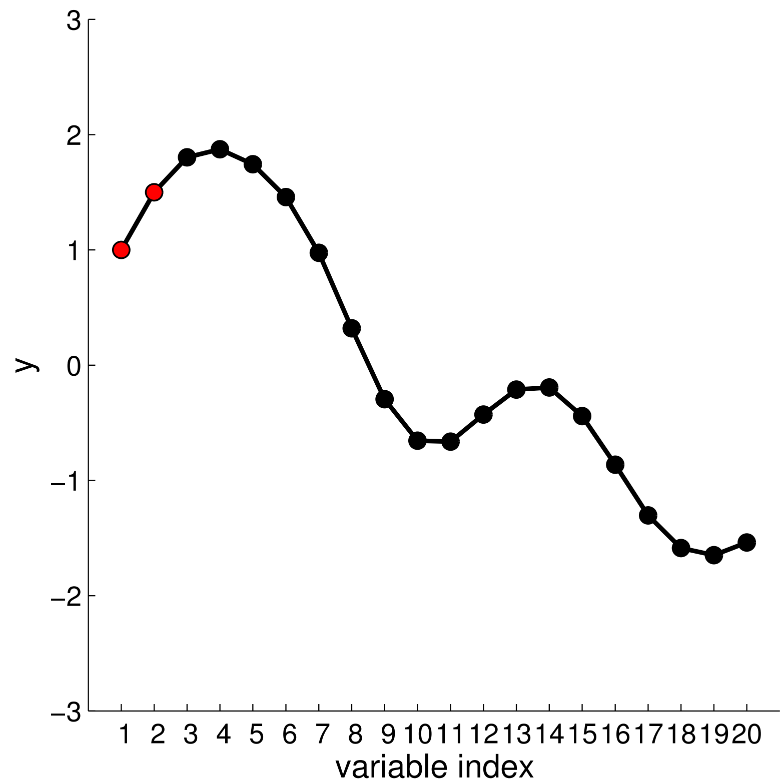


$$\Sigma =$$

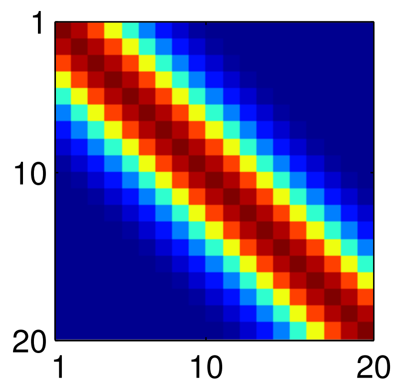


New visualisation

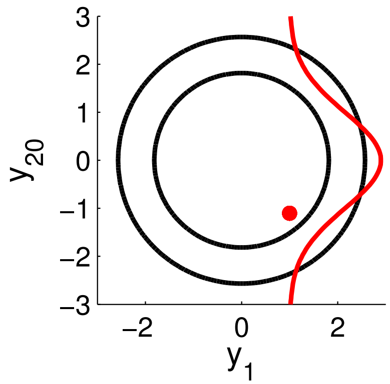


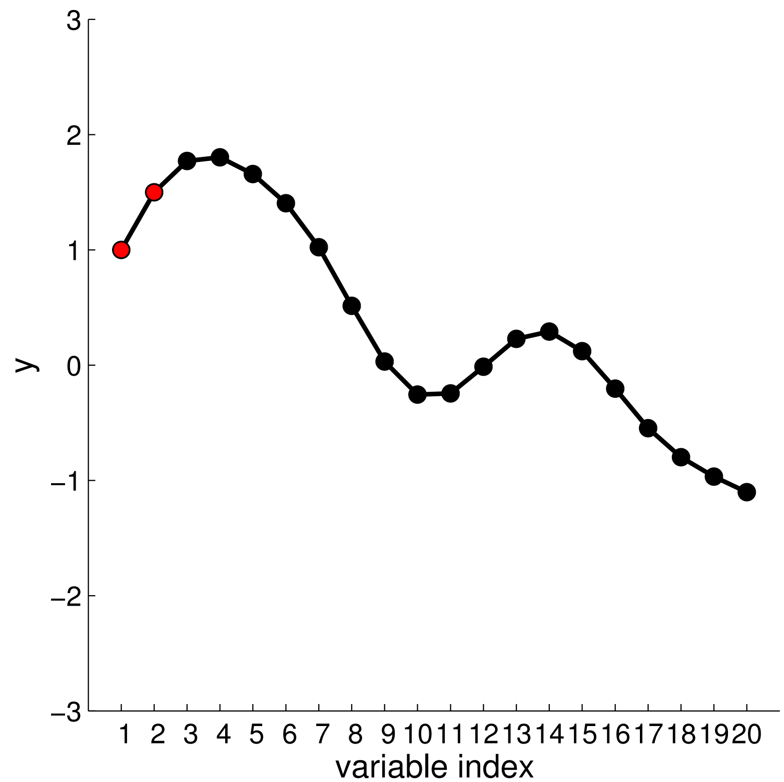


$\Sigma =$

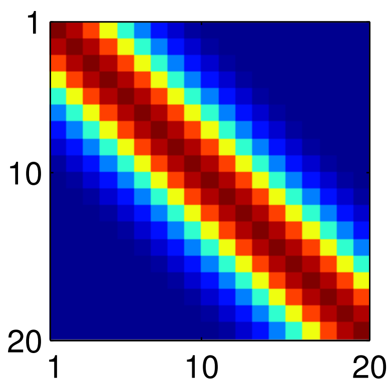


New visualisation

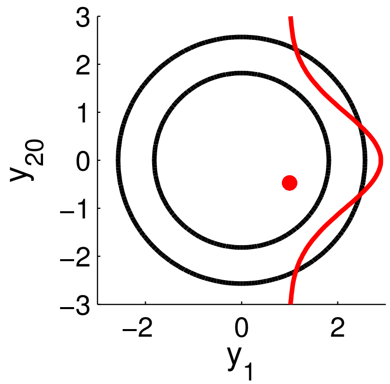


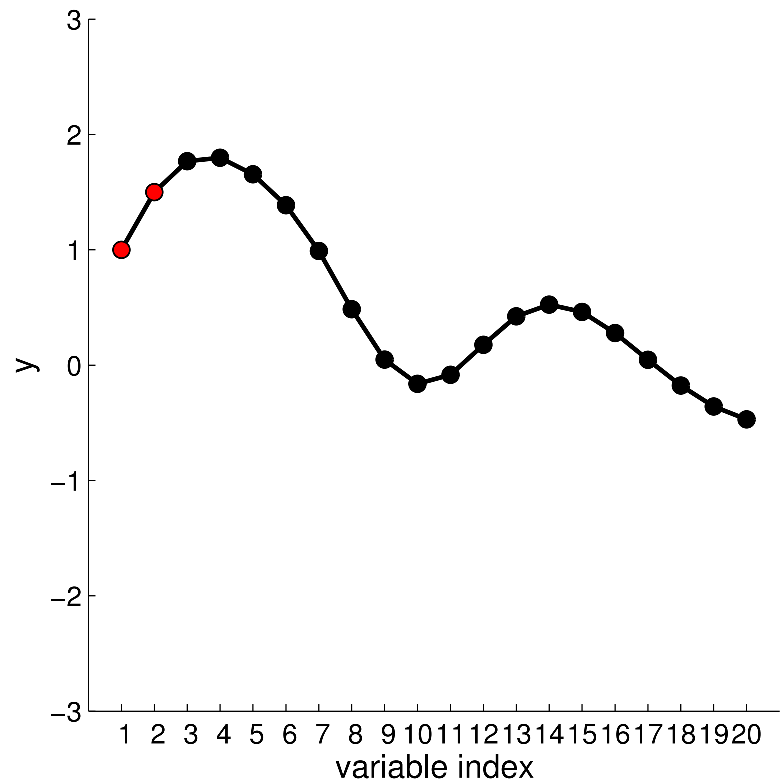


$$\Sigma =$$

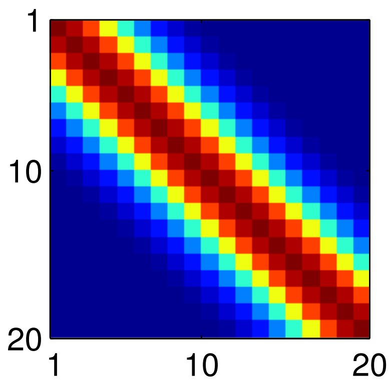


New visualisation

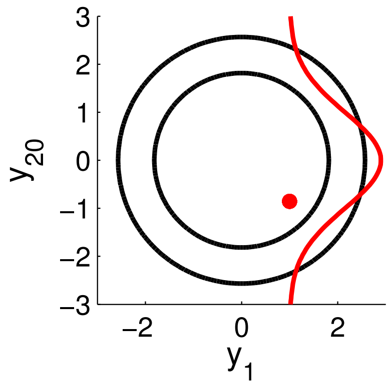


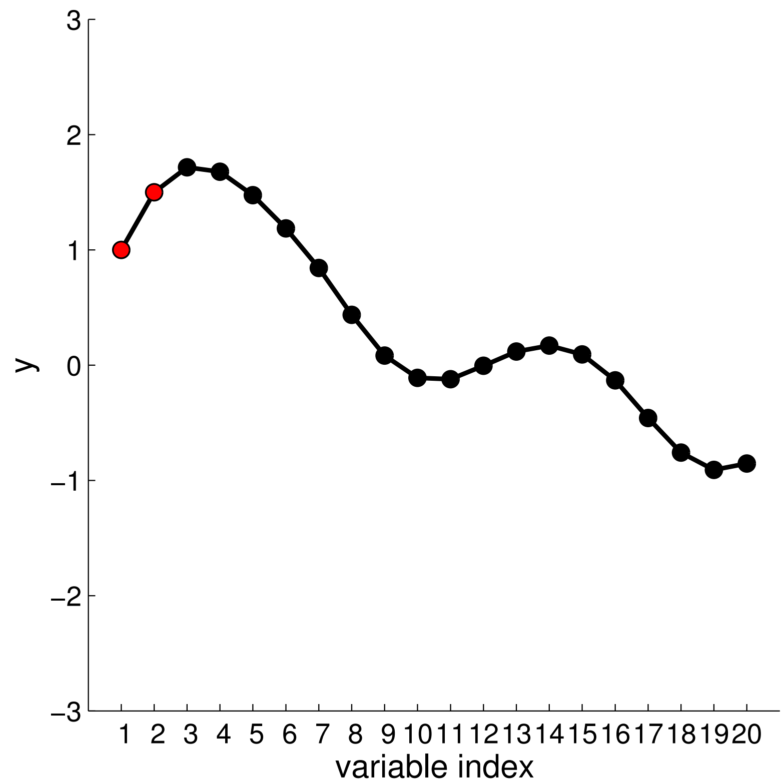


$$\Sigma =$$

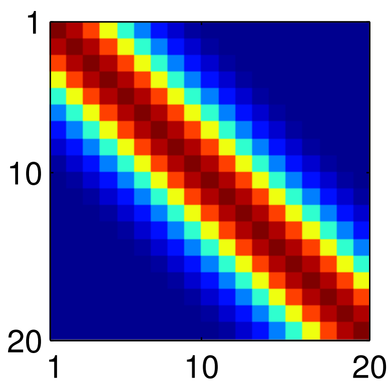


New visualisation

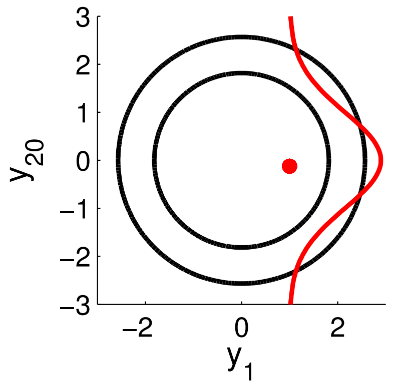


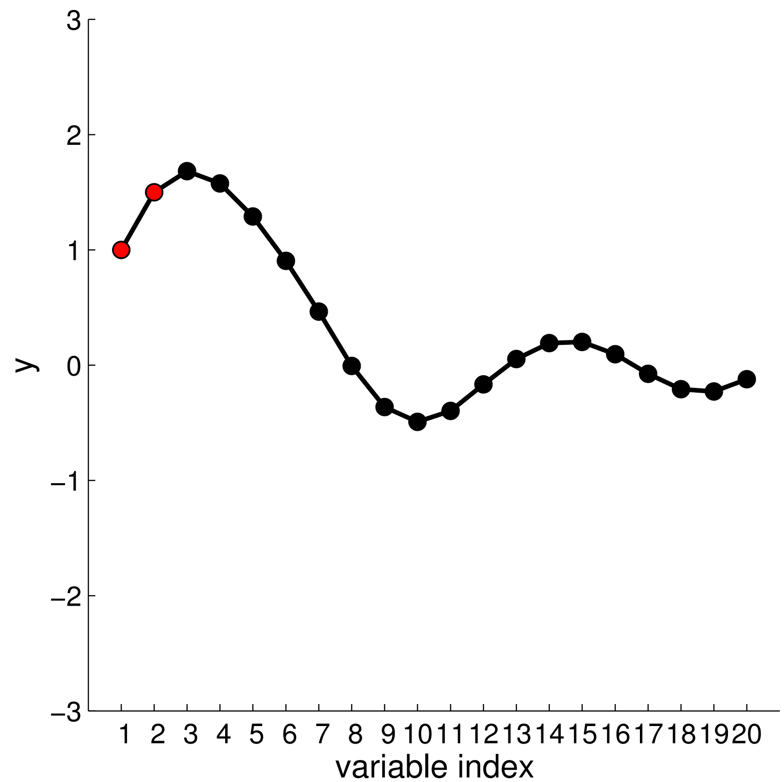


$$\Sigma =$$

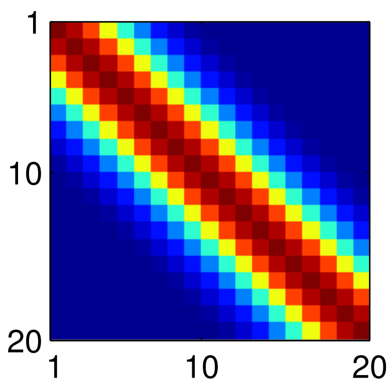


New visualisation

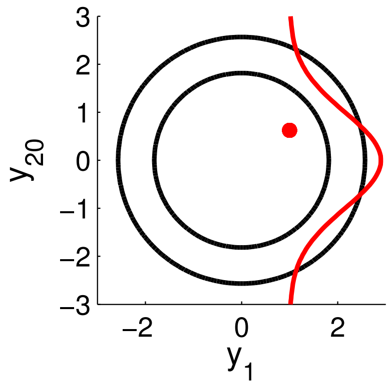


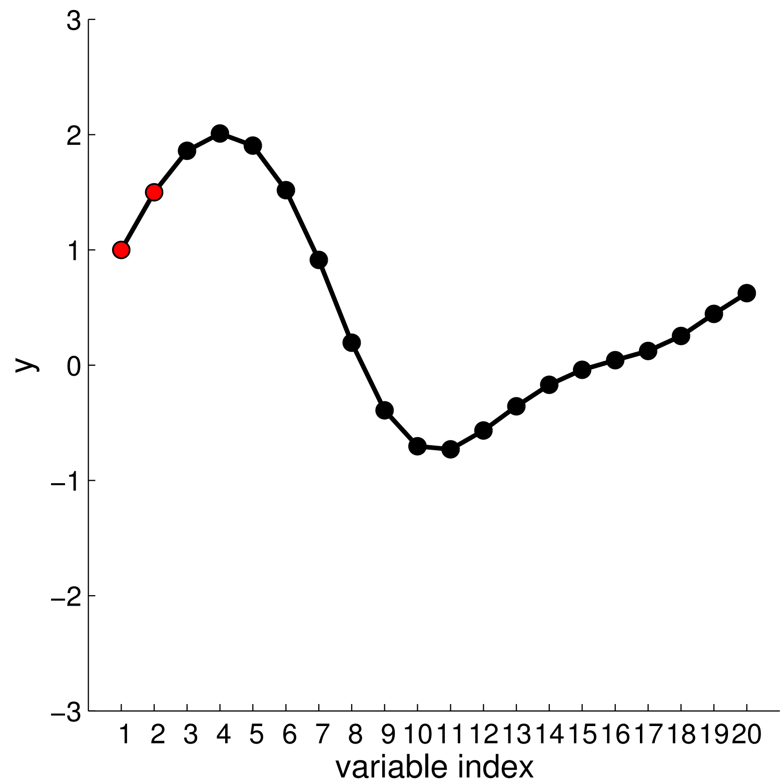


$\Sigma =$

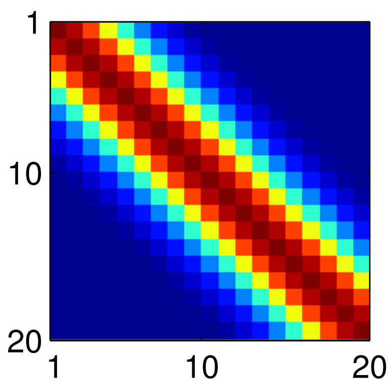


New visualisation

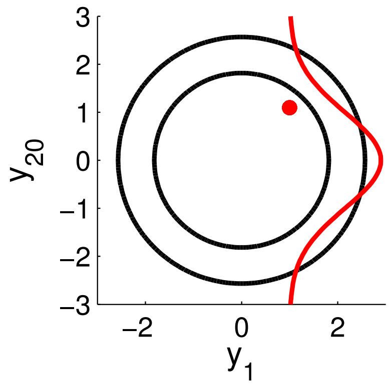


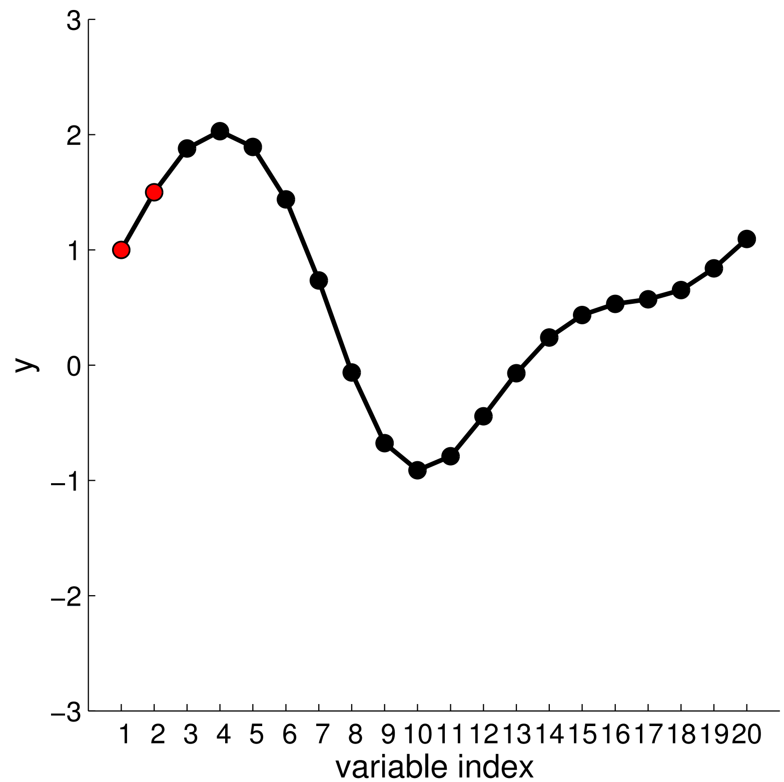


$\Sigma =$

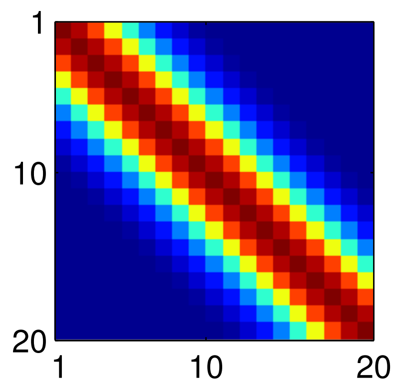


New visualisation

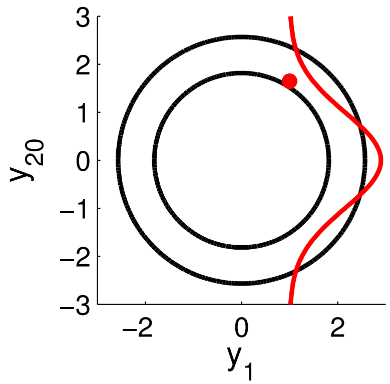


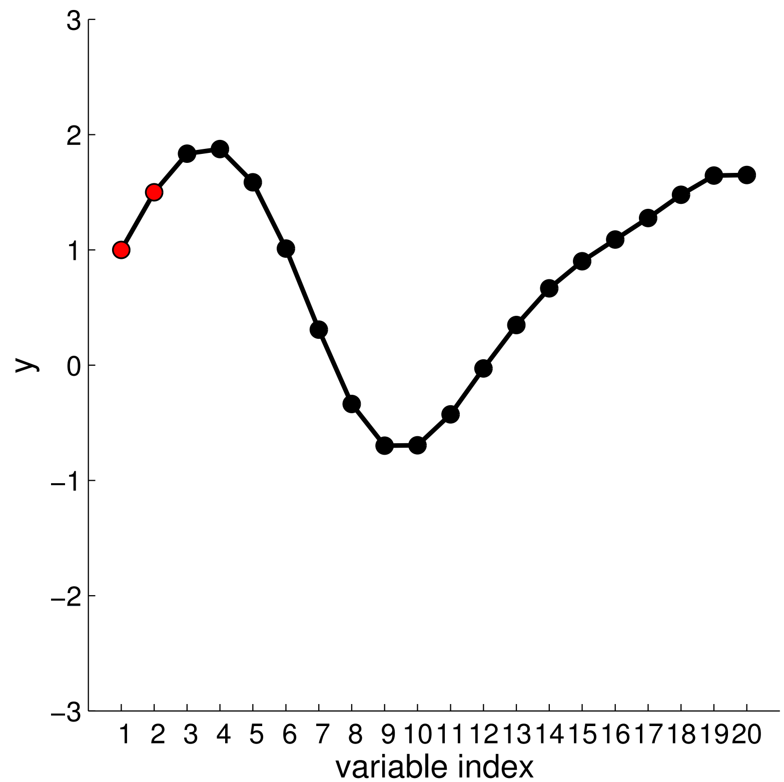


$\Sigma =$

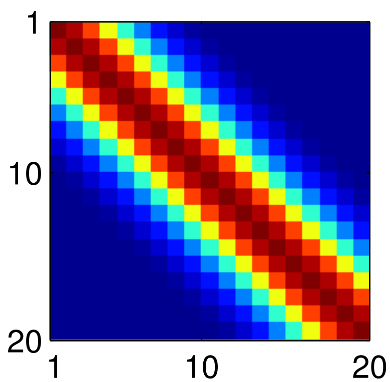


New visualisation

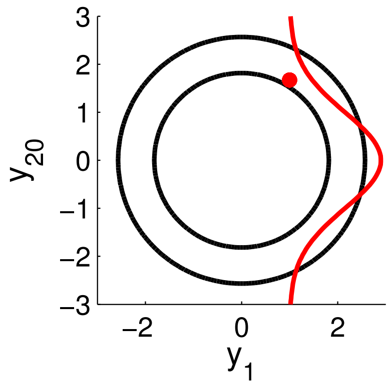


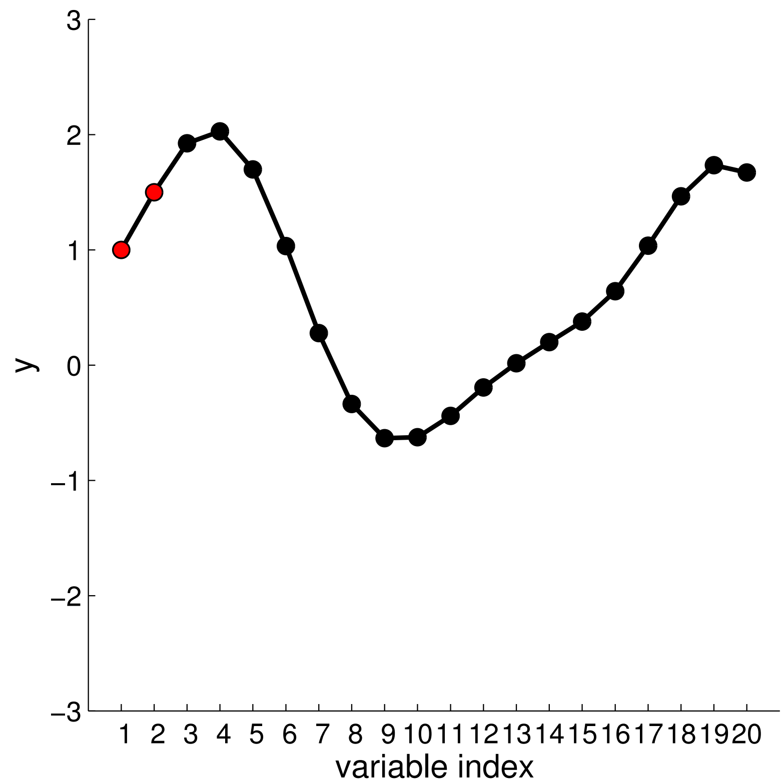


$$\Sigma =$$

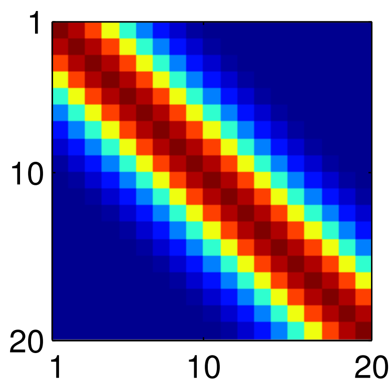


New visualisation

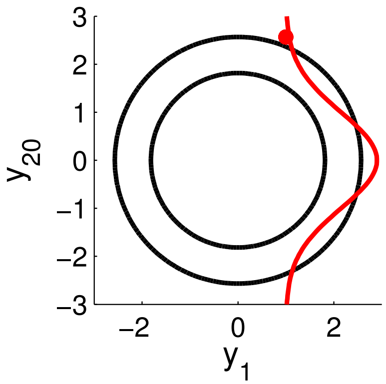




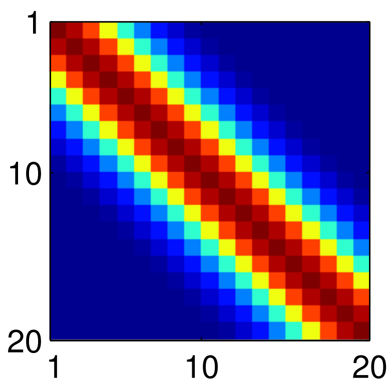
$\Sigma =$

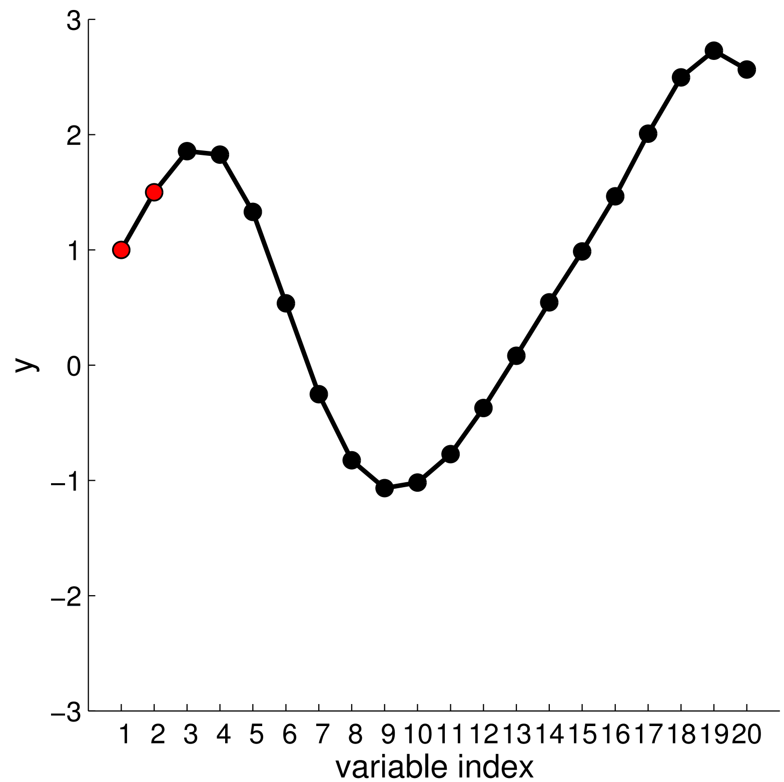


New visualisation

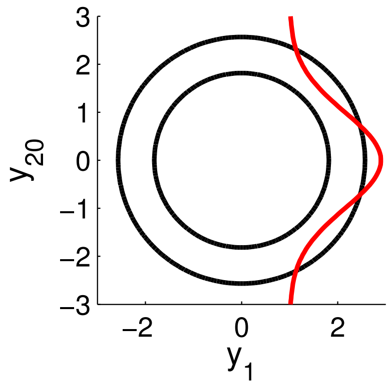


$$\Sigma =$$

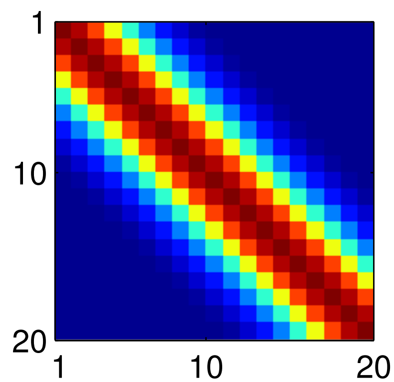


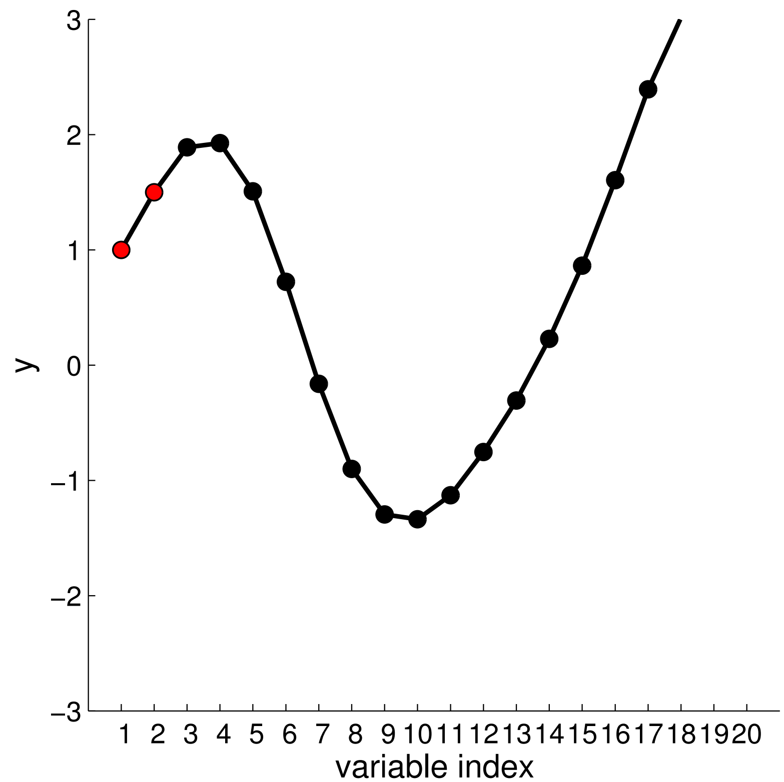


New visualisation



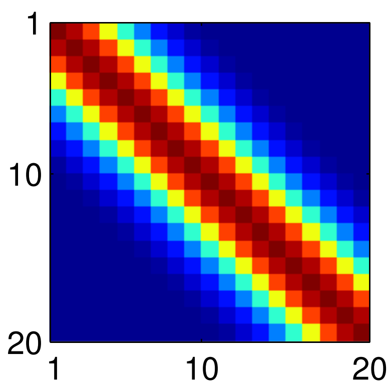
$$\Sigma =$$

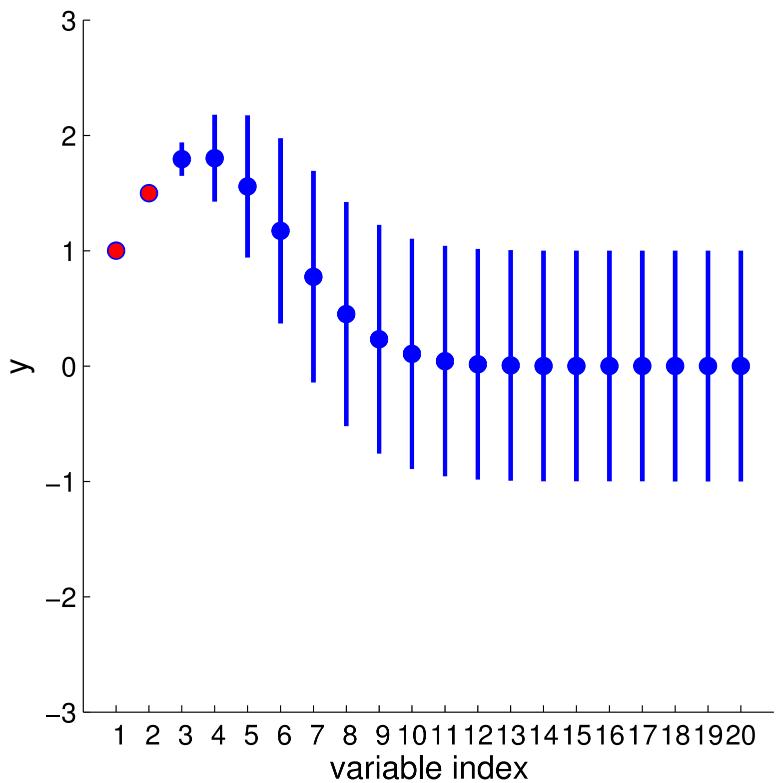




Regression using Gaussians

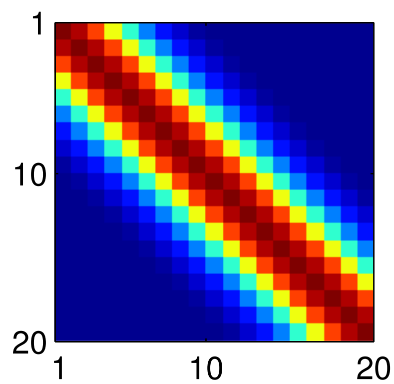
$\Sigma =$

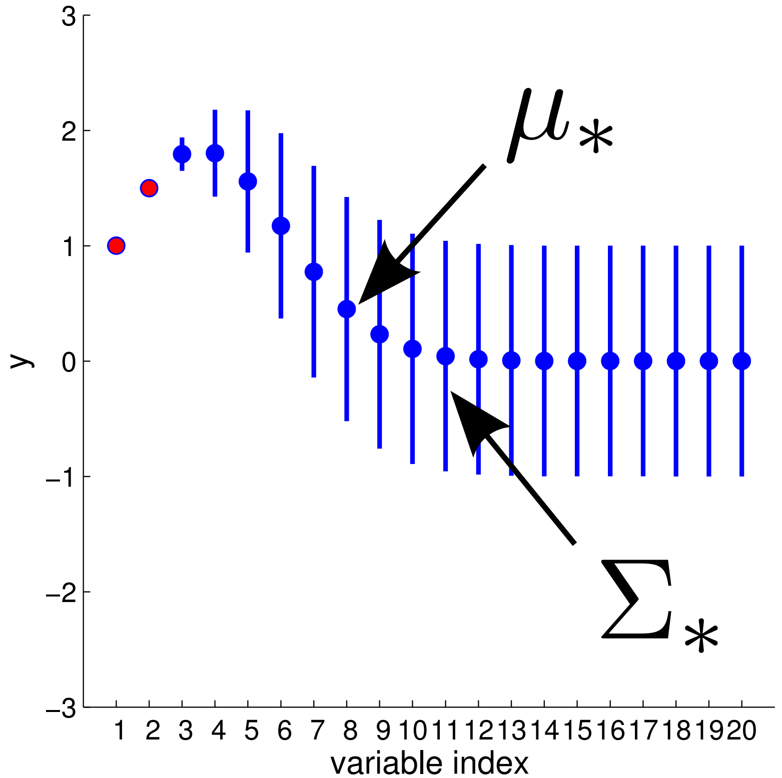




Regression using Gaussians

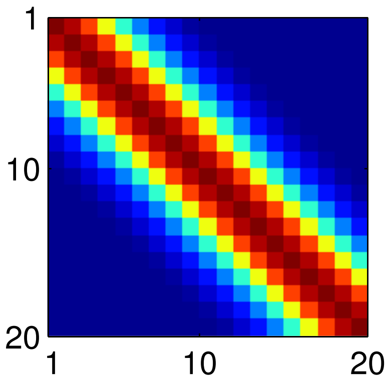
$\Sigma =$

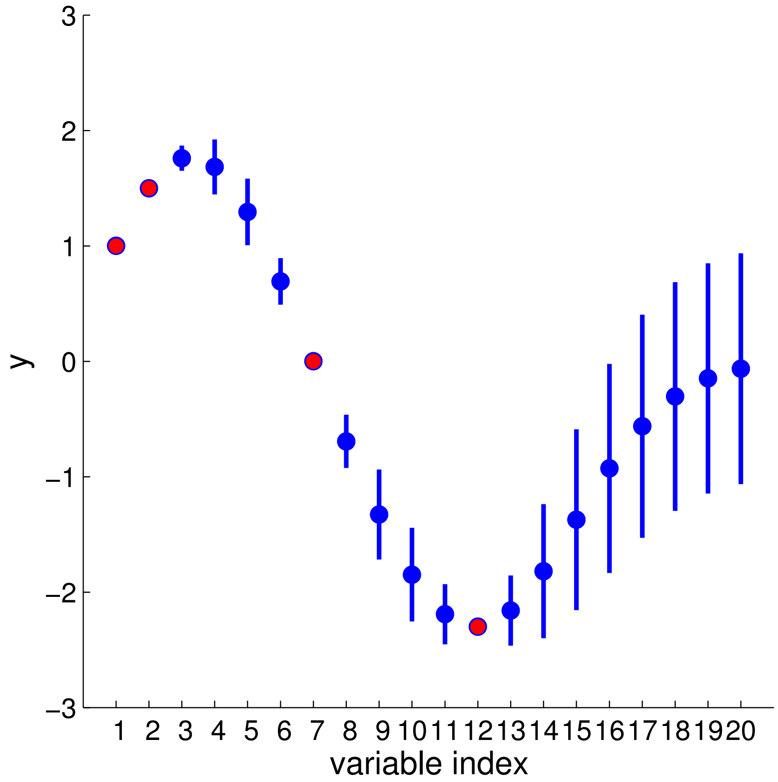




Regression using Gaussians

$\Sigma =$



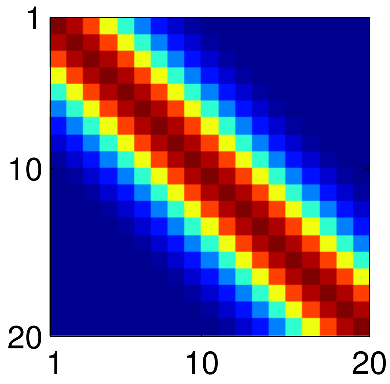


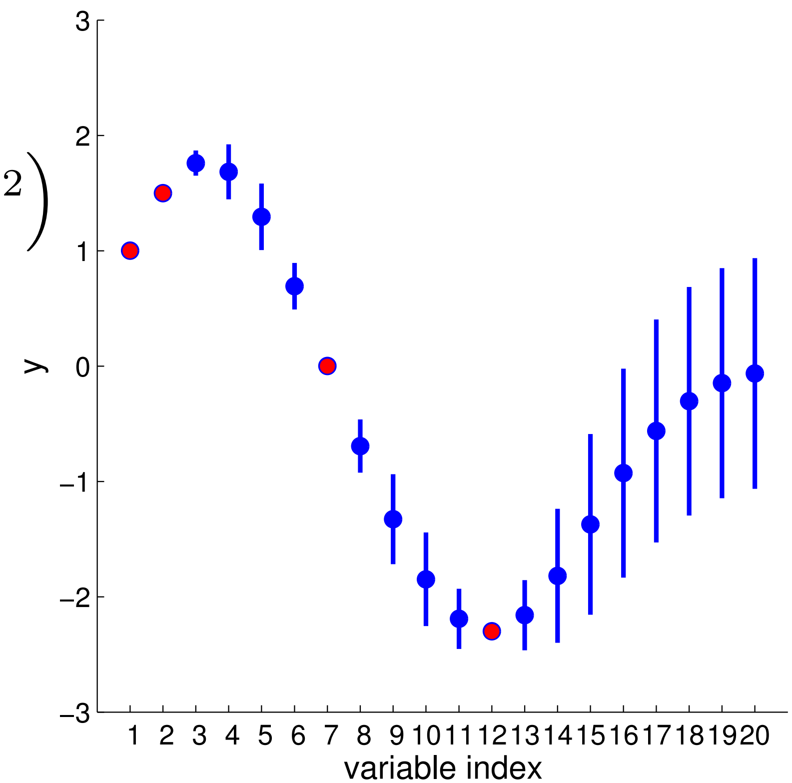
Regression using Gaussians

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

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

$\Sigma =$

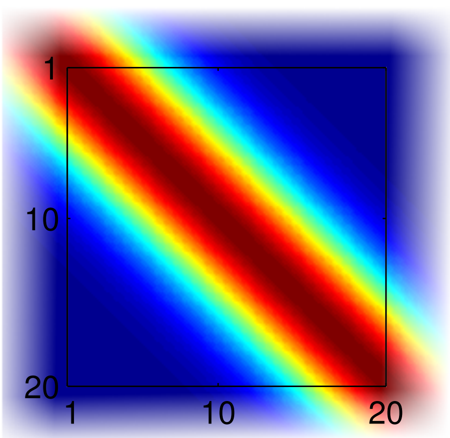


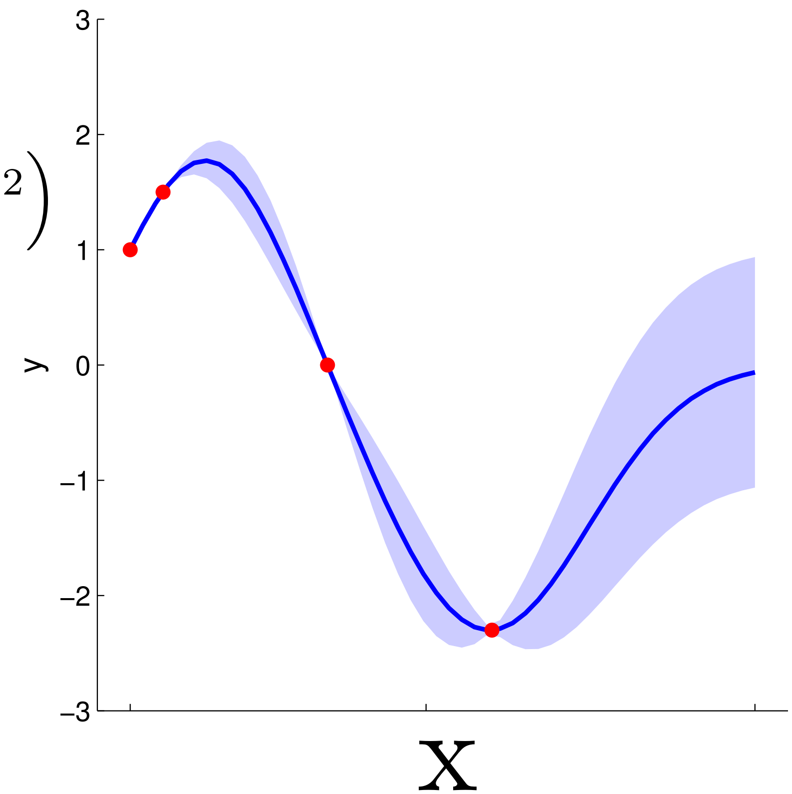


Regression: probabilistic inference in function space

$$\Sigma(x_1, x_2) = K(x_1, x_2) + I\sigma_y^2$$

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

$\Sigma =$ 



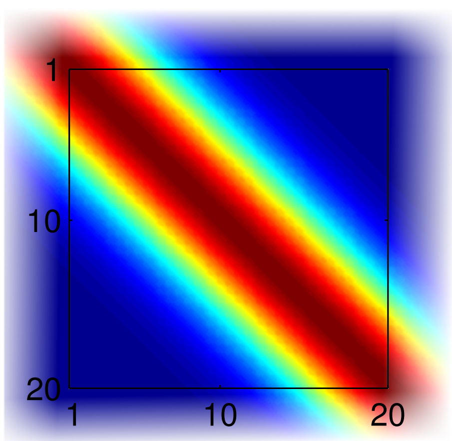
Regression: probabilistic inference in function space

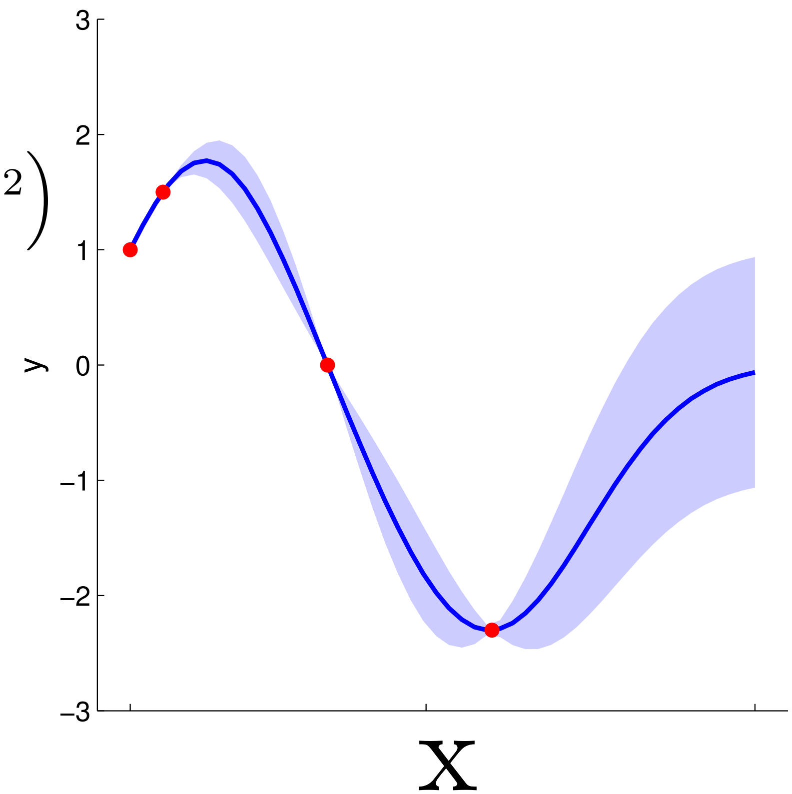
Non-parametric (∞-parametric)

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

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

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

$\Sigma =$ 



Parametric model

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

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

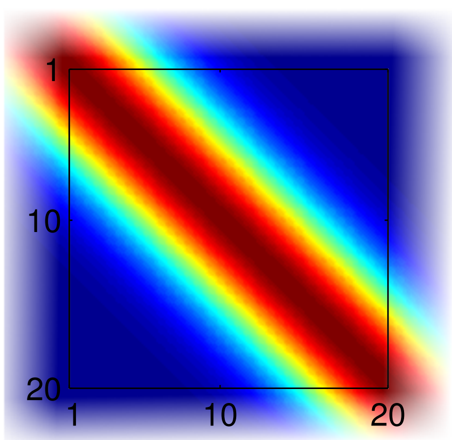
Regression: probabilistic inference in function space

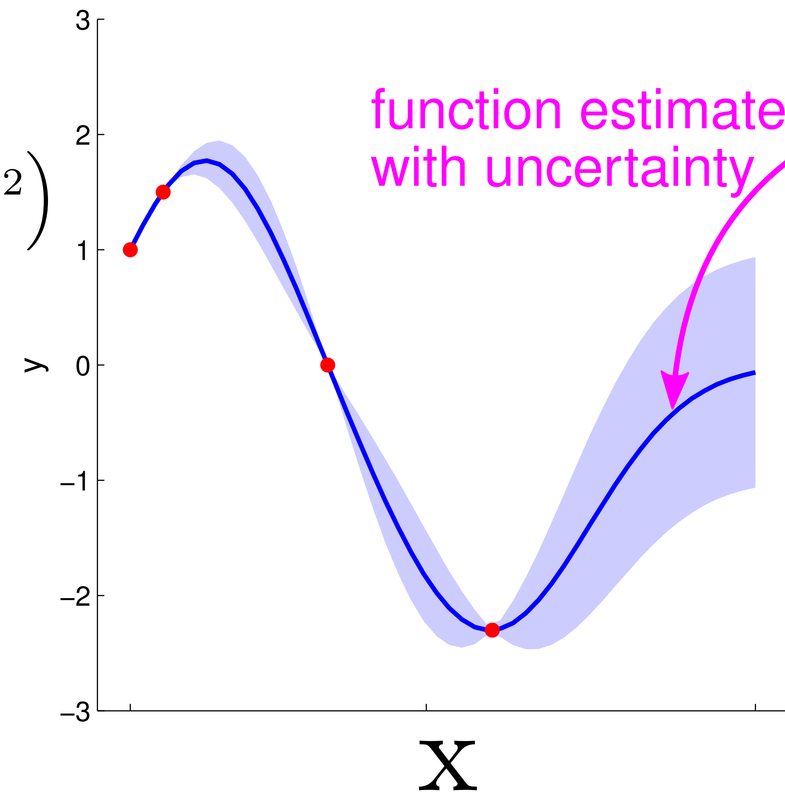
Non-parametric (∞-parametric)

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

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

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

$\Sigma =$ 



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

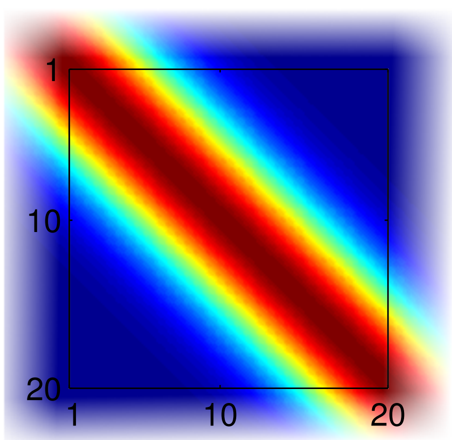
Non-parametric (∞-parametric)

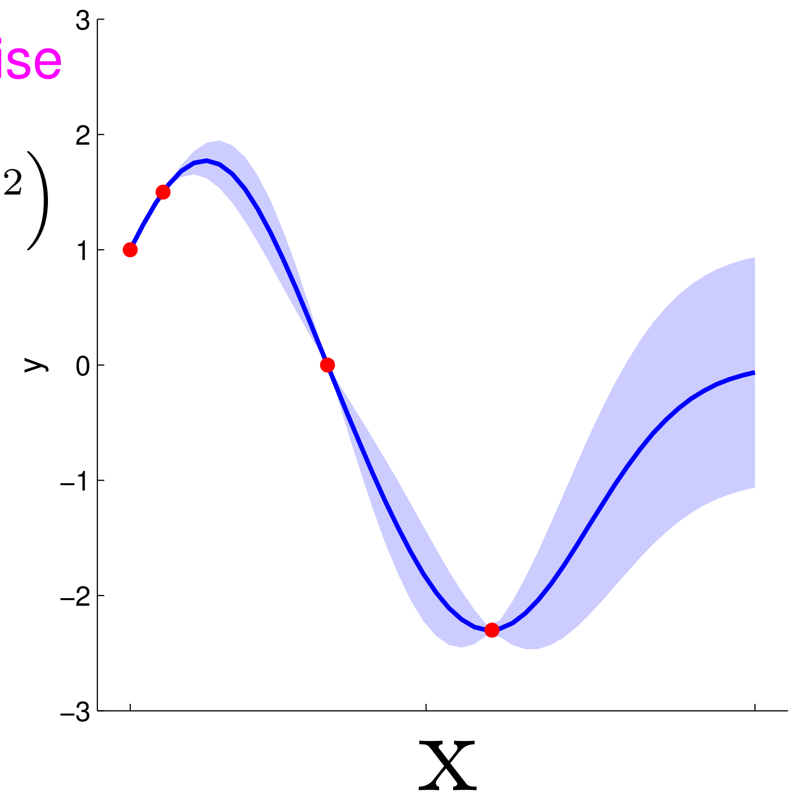
$$p(\mathsf{y}|\theta) = \mathcal{N}(0, \Sigma)$$

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

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

$\Sigma =$





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

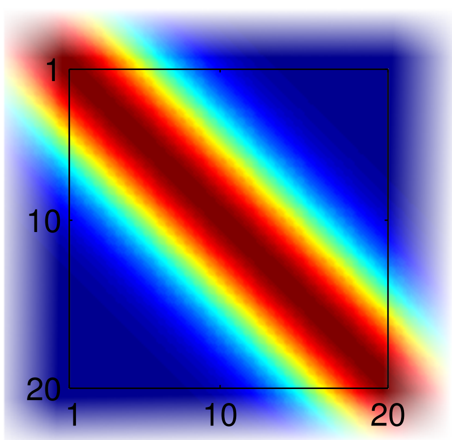
Non-parametric (∞-parametric)

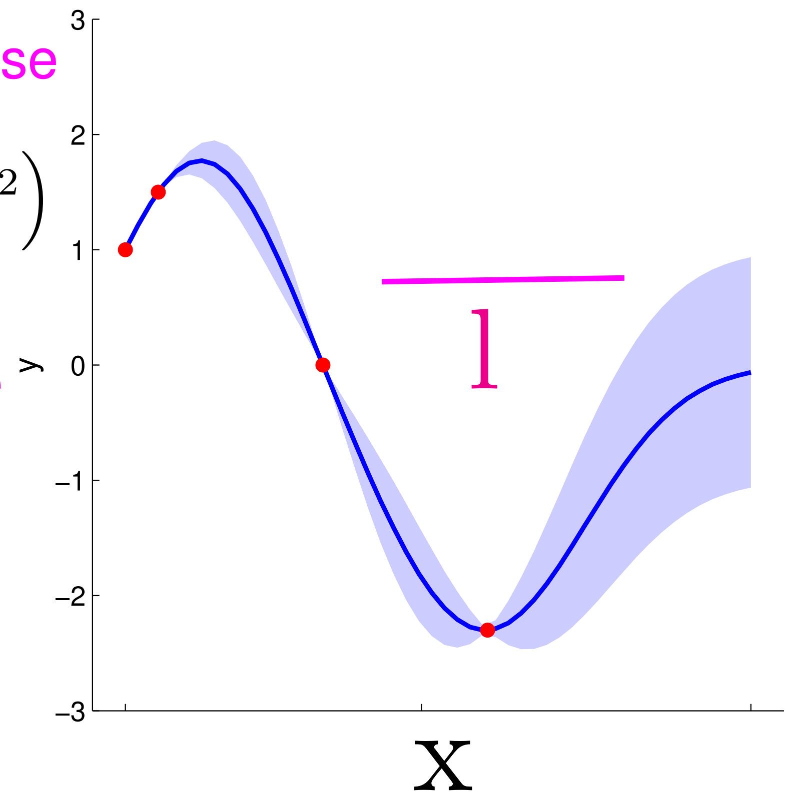
$$p(\mathsf{y}|\theta) = \mathcal{N}(0, \Sigma)$$

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

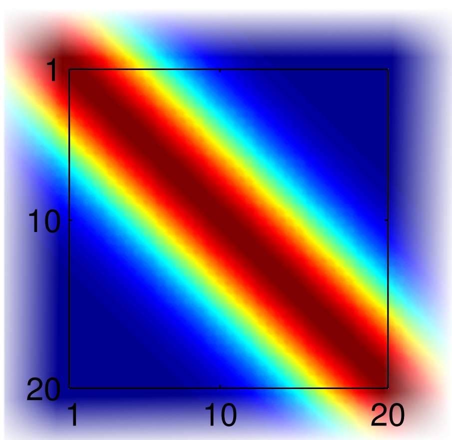
$$p(\mathsf{y}|\theta) = \mathcal{N}(0, \Sigma)$$

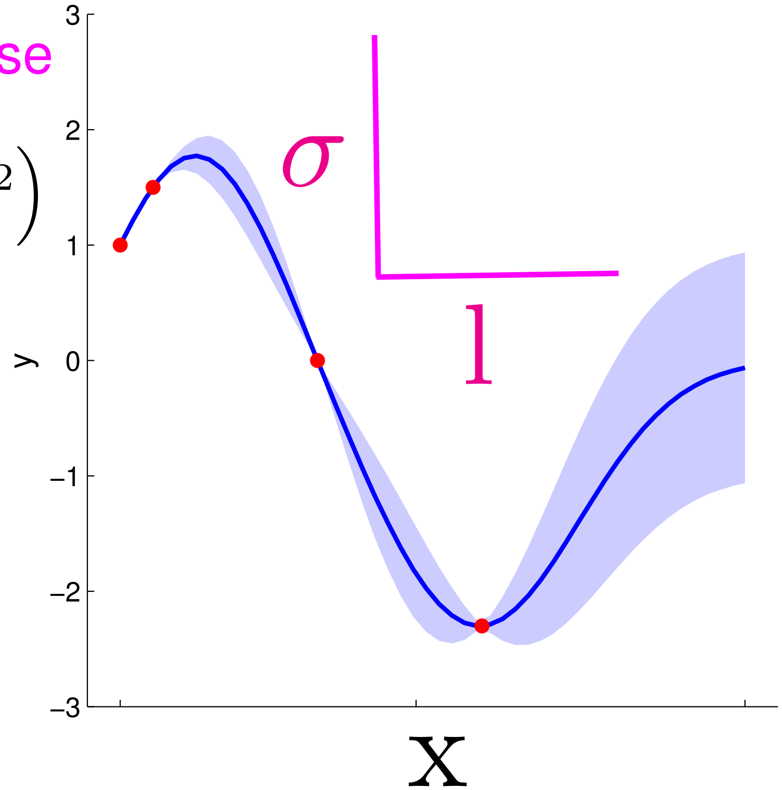
$$\Sigma(\mathsf{x}_1, \mathsf{x}_2) = \mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) + \mathsf{I}\sigma_{\mathsf{y}}^2$$ ← noise

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

↑ 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:

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

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:

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

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

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:

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

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

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:

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

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

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:

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

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

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

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

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

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

Q2. Why do we use covariances and not precisions to express modelling choices?

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:

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

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

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

Q2. Why do we use covariances and not precisions to express modelling choices?

$$\mathcal{P} = \Sigma^{-1}$$

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:

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

$$p(\mathbf{y}_1, \mathbf{y}_2) = \mathcal{N}\left(\begin{bmatrix} \mathbf{a} \\ \mathbf{b} \end{bmatrix}, \begin{bmatrix} A & B \\ B^T & C \end{bmatrix} \right) \implies p(\mathbf{y}_1) = \mathcal{N}(\mathbf{a}, A) \implies \mathcal{P} = A^{-1}$$

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

Q2. Why do we use covariances and not precisions to express modelling choices?

$$\mathcal{P} = \Sigma^{-1}$$

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:

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

$$p(\mathbf{y}_1, \mathbf{y}_2) = \mathcal{N}\left(\begin{bmatrix} \mathbf{a} \\ \mathbf{b} \end{bmatrix}, \begin{bmatrix} A & B \\ B^\top & C \end{bmatrix} \right) \implies p(\mathbf{y}_1) = \mathcal{N}(\mathbf{a}, A) \implies \mathcal{P} = A^{-1}$$

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

Q2. Why do we use covariances and not precisions to express modelling choices?

$$\mathcal{P} = \Sigma^{-1}$$

$$\begin{bmatrix} A & B \\ B^\top & C \end{bmatrix}^{-1} = \begin{bmatrix} (A - B^\top C^{-1} B)^{-1} & -A^{-1}B(C - BA^{-1}B^\top)^{-1} \\ -C^{-1}B(A - B^\top C^{-1}B)^{-1} & (C - BA^{-1}B^\top)^{-1} \end{bmatrix}$$

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:

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

$$p(\mathbf{y}_1, \mathbf{y}_2) = \mathcal{N}\left(\begin{bmatrix} \mathbf{a} \\ \mathbf{b} \end{bmatrix}, \begin{bmatrix} A & B \\ B^\top & C \end{bmatrix} \right) \implies p(\mathbf{y}_1) = \mathcal{N}(\mathbf{a}, A) \implies \mathcal{P} = A^{-1}$$

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

Q2. Why do we use covariances and not precisions to express modelling choices?

$$\mathcal{P} = \Sigma^{-1}$$

not the same

$$\begin{bmatrix} A & B \\ B^\top & C \end{bmatrix}^{-1} = \begin{bmatrix} (A - B^\top C^{-1} B)^{-1} & -A^{-1}B(C - BA^{-1}B^\top)^{-1} \\ -C^{-1}B(A - B^\top C^{-1}B)^{-1} & (C - BA^{-1}B^\top)^{-1} \end{bmatrix}$$

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:

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

$$p(\mathbf{y}_1, \mathbf{y}_2) = \mathcal{N}\left(\begin{bmatrix} \mathbf{a} \\ \mathbf{b} \end{bmatrix}, \begin{bmatrix} A & B \\ B^\top & C \end{bmatrix}\right) \implies p(\mathbf{y}_1) = \mathcal{N}(\mathbf{a}, A) \implies \mathcal{P} = A^{-1}$$

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

Q2. Why do we use covariances and not precisions to express modelling choices?

not the same

$$\mathcal{P} = \Sigma^{-1}$$

$$\begin{bmatrix} A & B \\ B^\top & C \end{bmatrix}^{-1} = \begin{bmatrix} (A - B^\top C^{-1} B)^{-1} & -A^{-1}B(C - BA^{-1}B^\top)^{-1} \\ -C^{-1}B(A - B^\top C^{-1}B)^{-1} & (C - BA^{-1}B^\top)^{-1} \end{bmatrix}$$

$\implies$ Entries in a precision matrix depend on what other data we are considering

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:

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

$$p(\mathbf{y}_1, \mathbf{y}_2) = \mathcal{N}\left(\begin{bmatrix} \mathbf{a} \\ \mathbf{b} \end{bmatrix}, \begin{bmatrix} \mathsf{A} & \mathsf{B} \\ \mathsf{B}^\top & \mathsf{C} \end{bmatrix}\right) \implies p(\mathbf{y}_1) = \mathcal{N}(\mathbf{a}, \mathsf{A}) \implies \mathcal{P} = \mathsf{A}^{-1}$$

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

Q2. Why do we use covariances and not precisions to express modelling choices?

$$\mathcal{P} = \Sigma^{-1}$$

not the same

$$\begin{bmatrix} \mathsf{A} & \mathsf{B} \\ \mathsf{B}^\top & \mathsf{C} \end{bmatrix}^{-1} = \begin{bmatrix} (\mathsf{A} - \mathsf{B}^\top \mathsf{C}^{-1} \mathsf{B})^{-1} & -\mathsf{A}^{-1}\mathsf{B}(\mathsf{C} - \mathsf{B}\mathsf{A}^{-1}\mathsf{B}^\top)^{-1} \\ -\mathsf{C}^{-1}\mathsf{B}(\mathsf{A} - \mathsf{B}^\top\mathsf{C}^{-1}\mathsf{B})^{-1} & (\mathsf{C} - \mathsf{B}\mathsf{A}^{-1}\mathsf{B}^\top)^{-1} \end{bmatrix}$$

$\implies$ Entries in a precision matrix depend on what other data we are considering

Precisions tell us about the conditional independence relationships between pairs of variables (are $\mathbf{y}_1$ and $\mathbf{y}_2$ independent given all other variables).

Mathematical Foundations: Regression

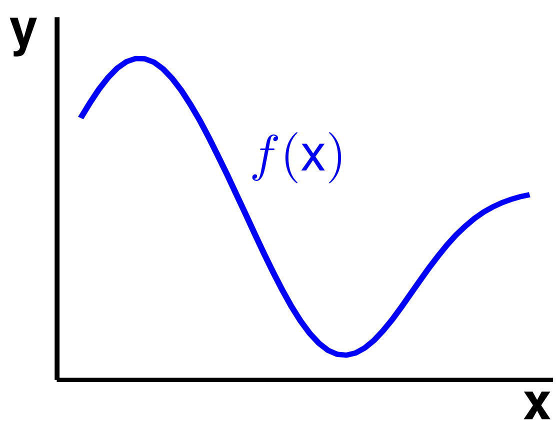
Q3. What's the formal justification for how we were using GPs for regression?

Mathematical Foundations: Regression

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

Generative model (like non-linear regression)

$$y(x) = f(x) + \epsilon\sigma_y$$



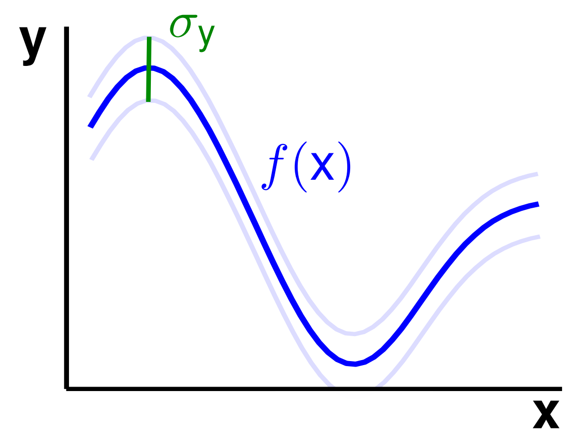
Mathematical Foundations: Regression

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

Generative model (like non-linear regression)

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

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



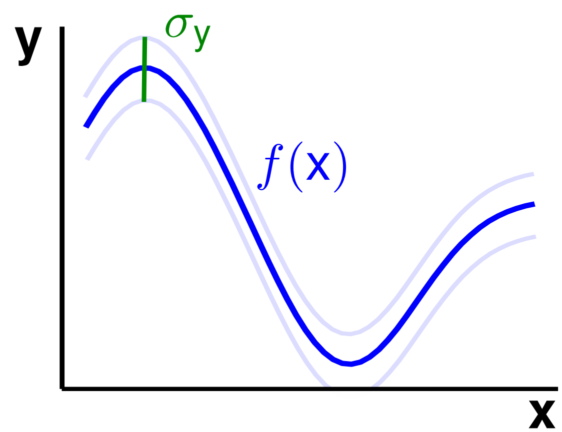
Mathematical Foundations: Regression

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

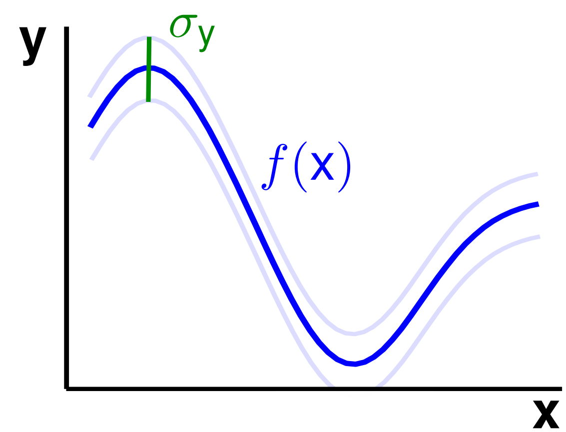
$$p(f(\mathsf{x})|\theta) = \mathcal{GP}(0, \mathsf{K}(\mathsf{x}, \mathsf{x}'))$$

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

(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)

$$y(x) = f(x) + \epsilon\sigma_y$$

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

place GP prior over the non-linear function

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

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

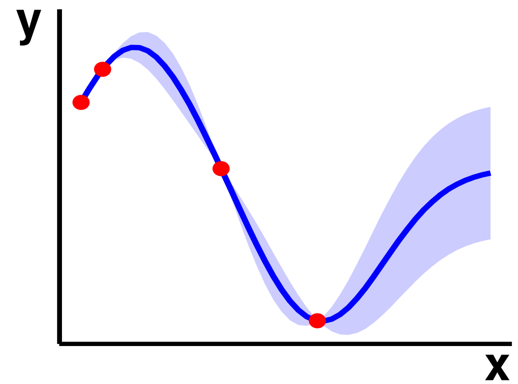
(smoothly wiggling functions expected)

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

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

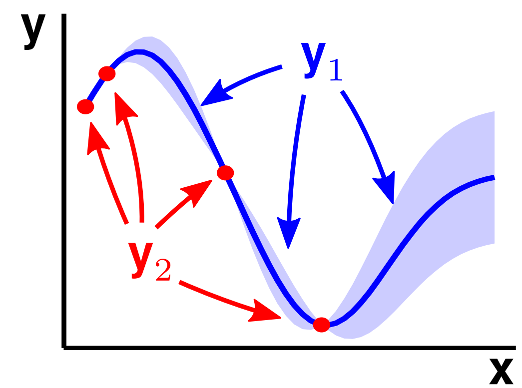
Mathematical foundations: Prediction

Q4. How do we make predictions?



Mathematical foundations: Prediction

Q4. How do we make predictions?



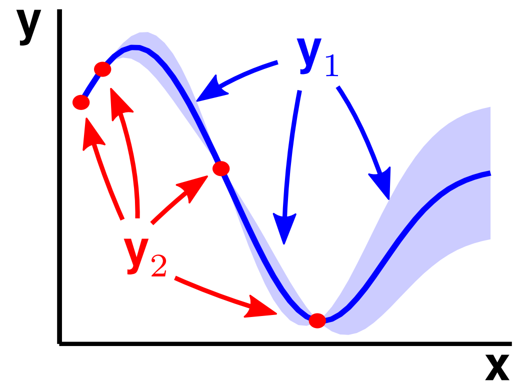
$$p(\mathbf{y}_1|\mathbf{y}_2) = \frac{p(\mathbf{y}_1, \mathbf{y}_2)}{p(\mathbf{y}_2)}$$

Mathematical foundations: Prediction

Q4. How do we make predictions?

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

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



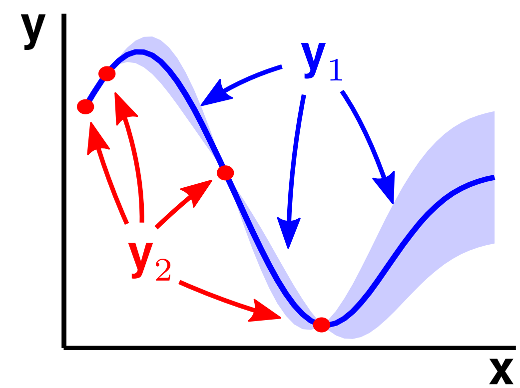
Mathematical foundations: Prediction

Q4. How do we make predictions?

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

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

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



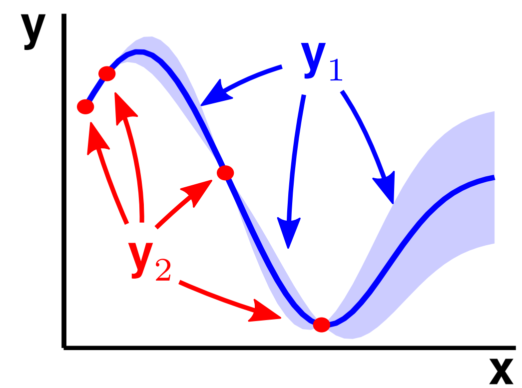
Mathematical foundations: Prediction

Q4. How do we make predictions?

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

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

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



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

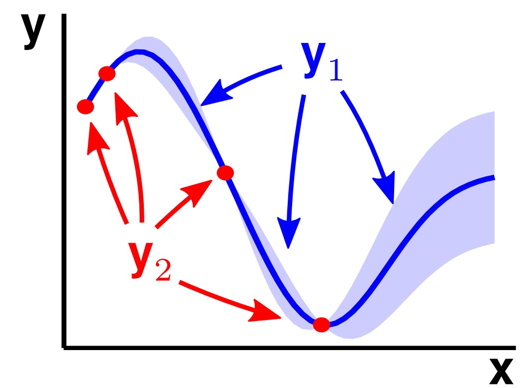
Mathematical foundations: Prediction

Q4. How do we make predictions?

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

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

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



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

predictive mean

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

$$= \mathsf{B}\mathsf{C}^{-1}\mathbf{y}_2$$

$$= \mathsf{W}\mathbf{y}_2$$

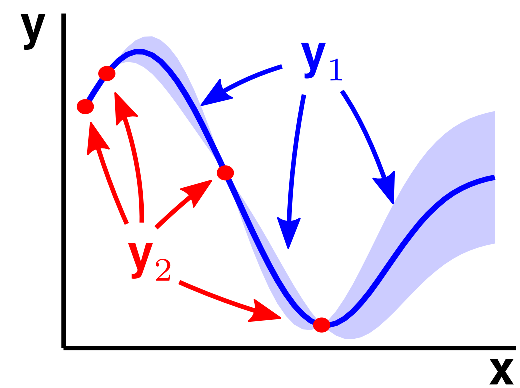
linear in the data

Mathematical foundations: Prediction

Q4. How do we make predictions?

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

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



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

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

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

predictive uncertainty = prior uncertainty − reduction in uncertainty

predictions more confident than prior

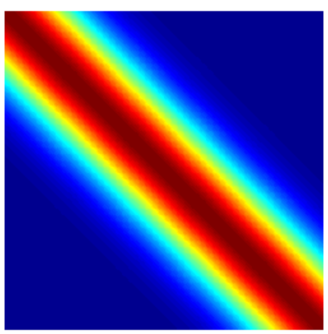
What effect do the hyper-parameters have?

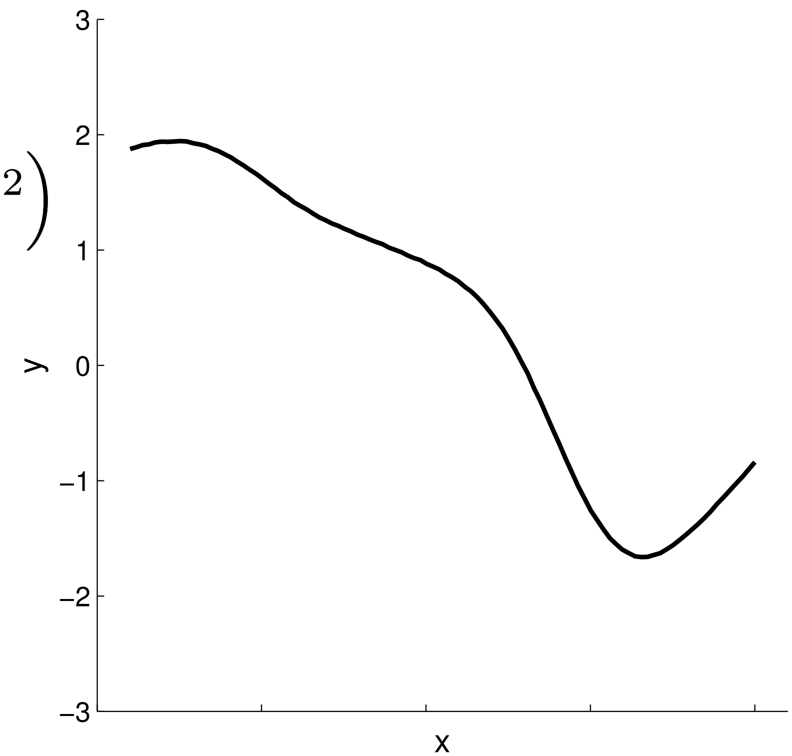
Non-parametric (∞-parametric)

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

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

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

$\Sigma =$ 



Parametric model

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

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

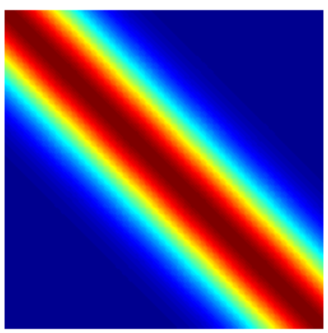
What effect do the hyper-parameters have?

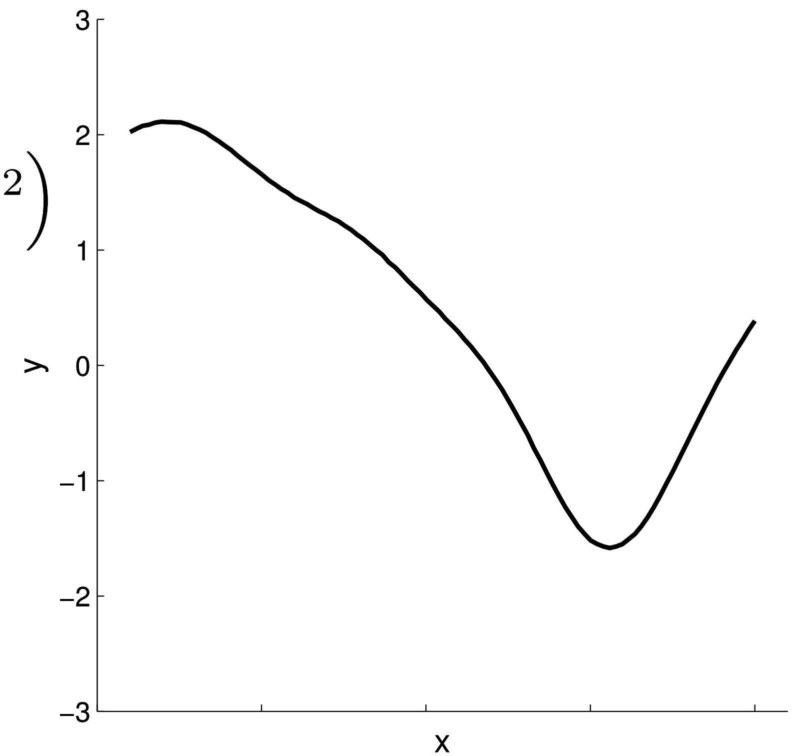
Non-parametric (∞-parametric)

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

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

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

$\Sigma =$ 



Parametric model

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

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

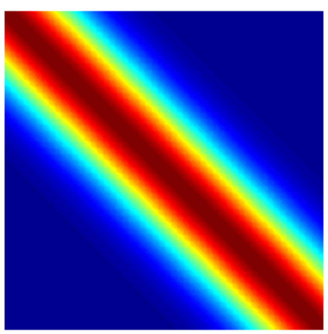
What effect do the hyper-parameters have?

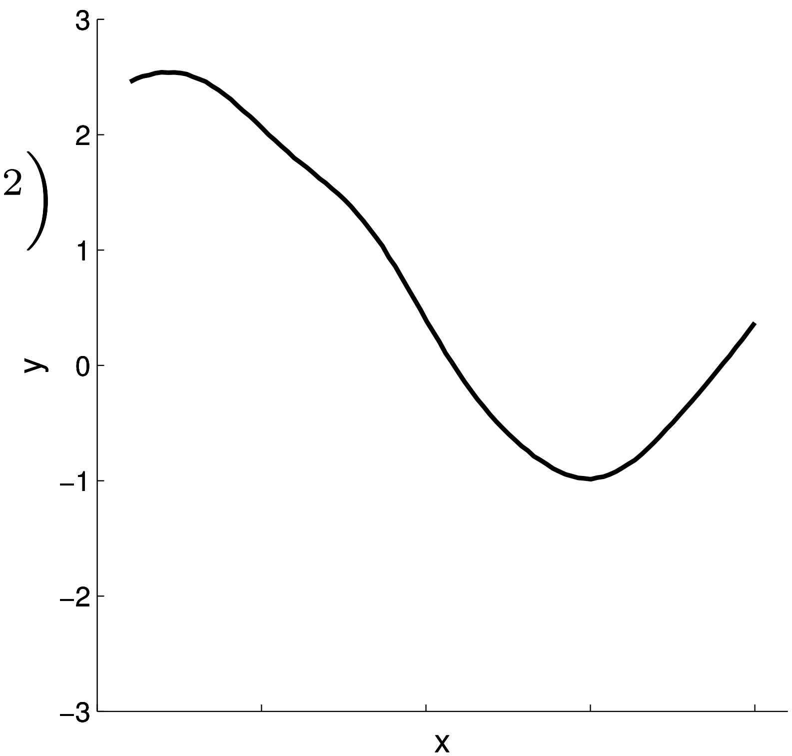
Non-parametric (∞-parametric)

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

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

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

$\Sigma =$ 



Parametric model

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

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

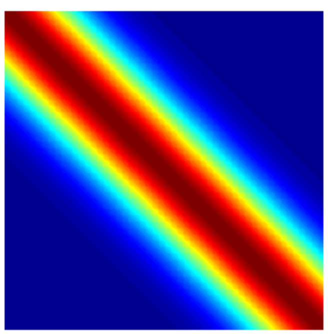
What effect do the hyper-parameters have?

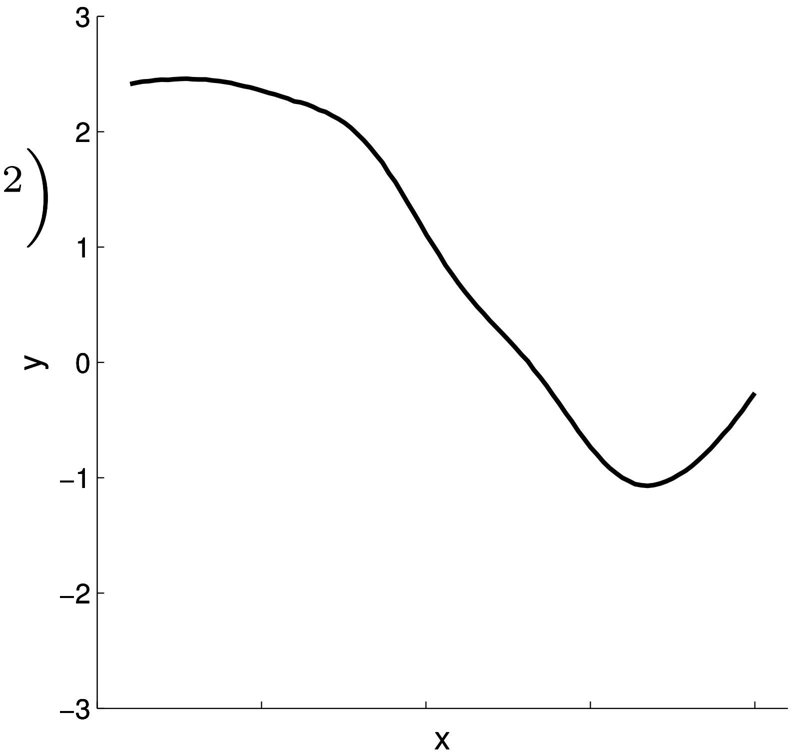
Non-parametric (∞-parametric)

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

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

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

$\Sigma =$ 



Parametric model

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

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

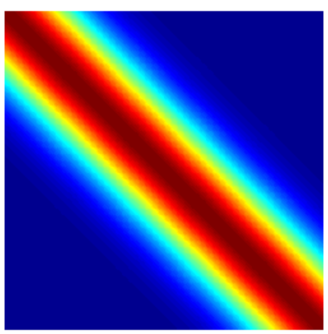
What effect do the hyper-parameters have?

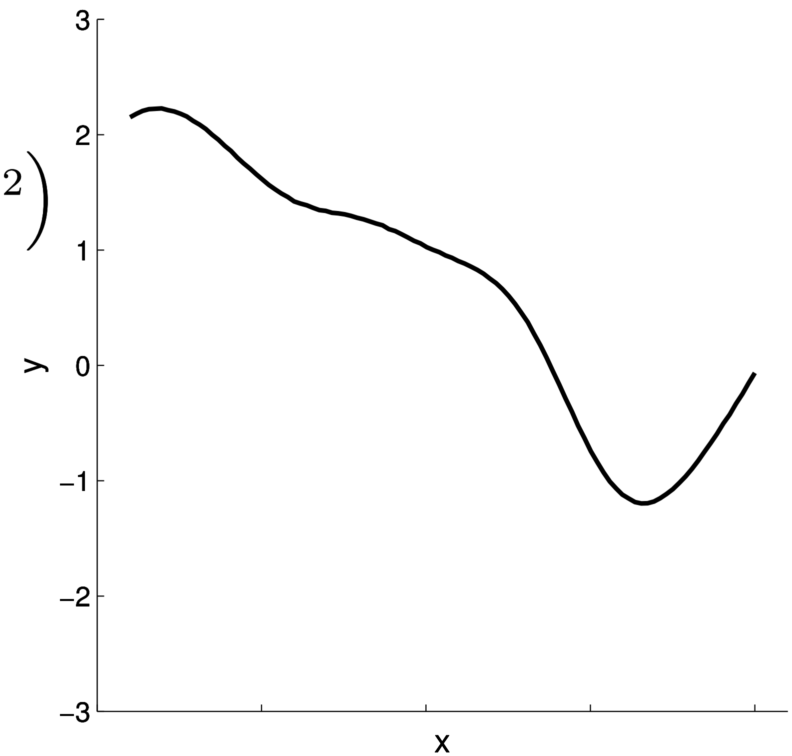
Non-parametric (∞-parametric)

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

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

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

$\Sigma =$ 



Parametric model

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

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

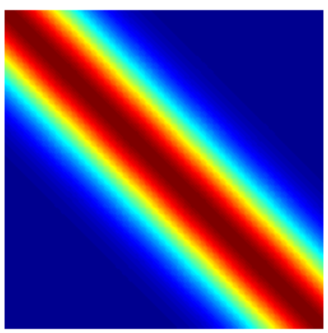
What effect do the hyper-parameters have?

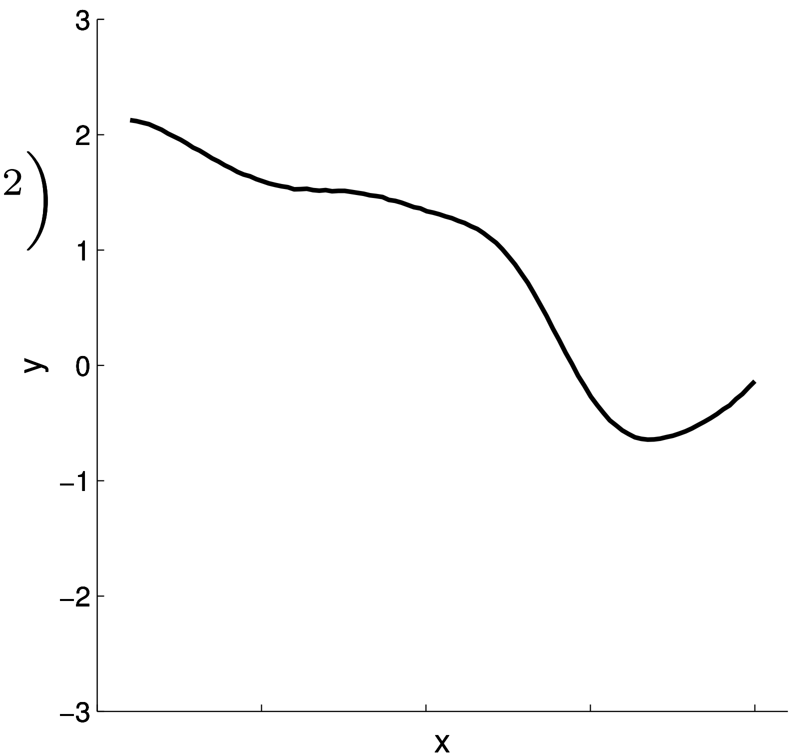
Non-parametric (∞-parametric)

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

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

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

$\Sigma =$ 



Parametric model

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

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

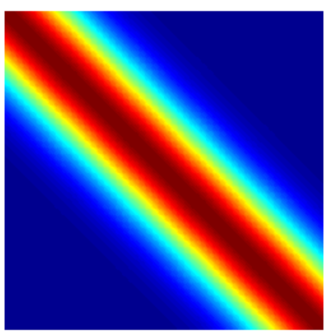
What effect do the hyper-parameters have?

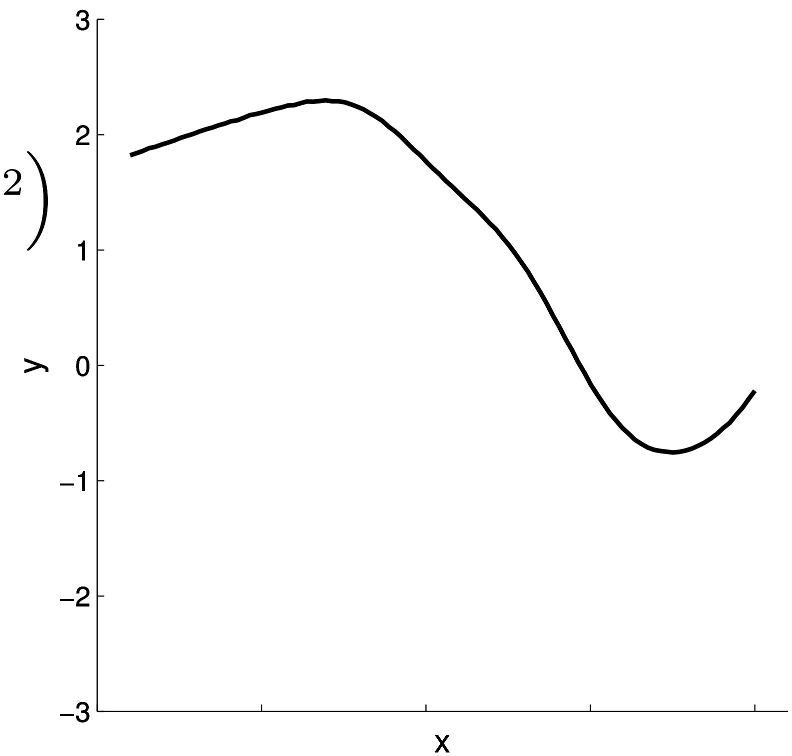
Non-parametric (∞-parametric)

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

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

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

$\Sigma =$ 



Parametric model

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

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

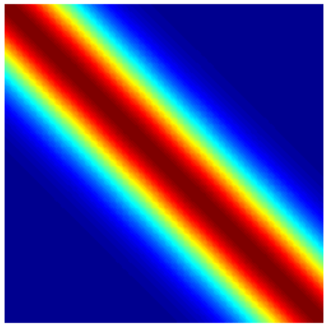
What effect do the hyper-parameters have?

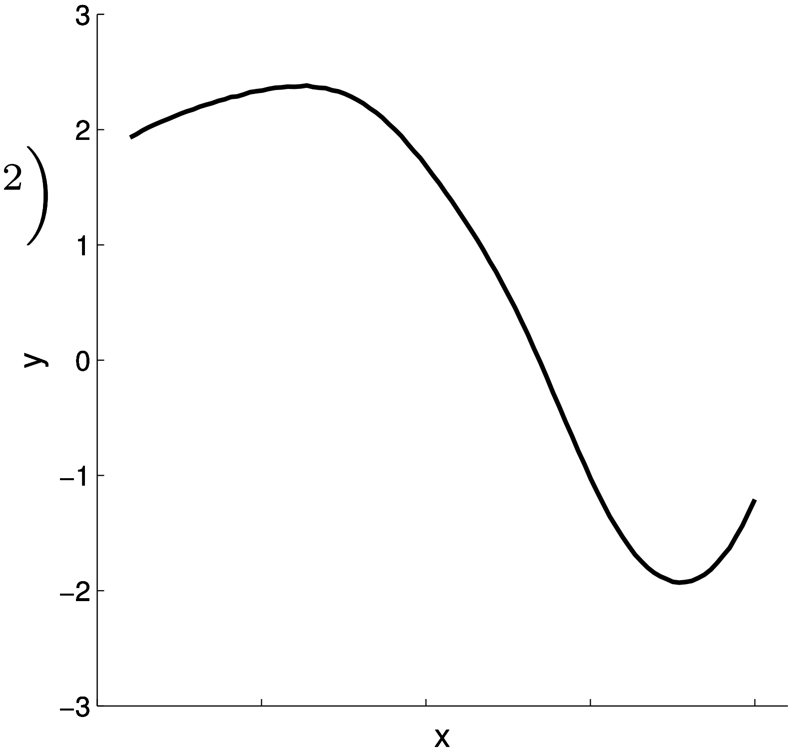
Non-parametric (∞-parametric)

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

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

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

$\Sigma =$ 



Parametric model

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

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

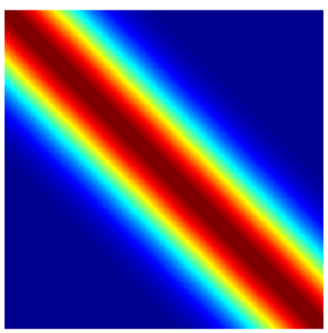
What effect do the hyper-parameters have?

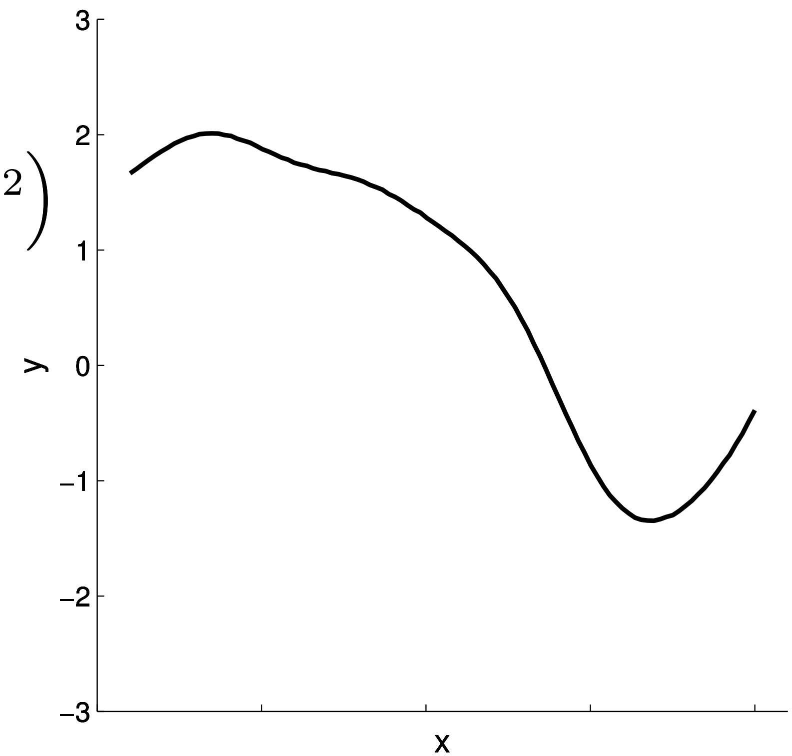
Non-parametric (∞-parametric)

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

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

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

$\Sigma =$ 



Parametric model

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

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

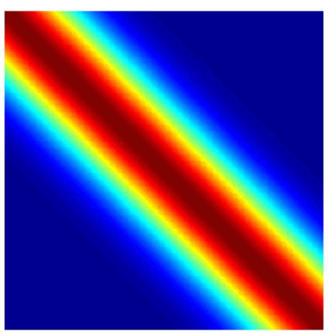
What effect do the hyper-parameters have?

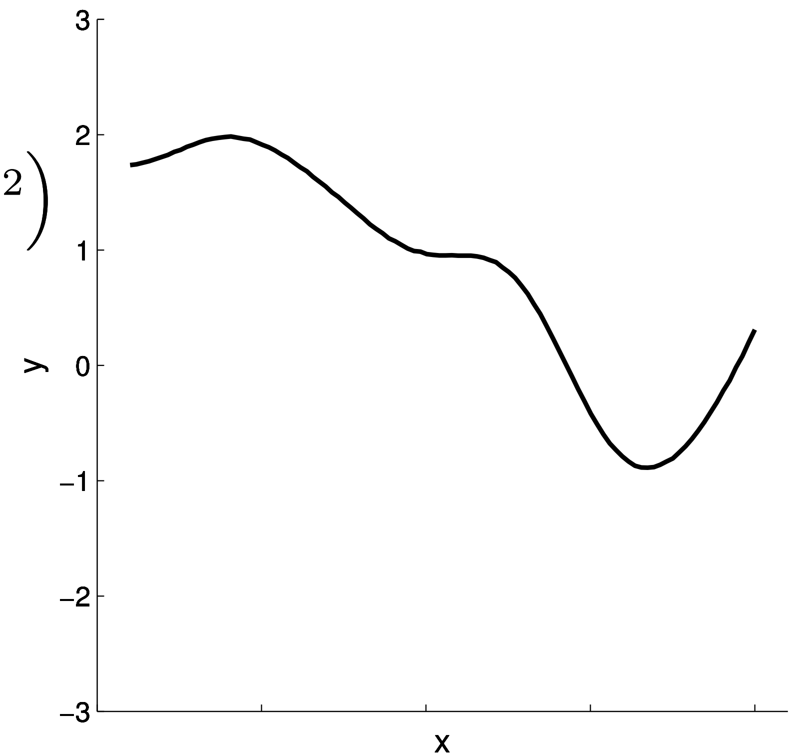
Non-parametric (∞-parametric)

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

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

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

$\Sigma =$ 



Parametric model

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

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

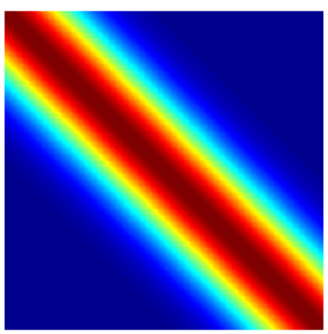
What effect do the hyper-parameters have?

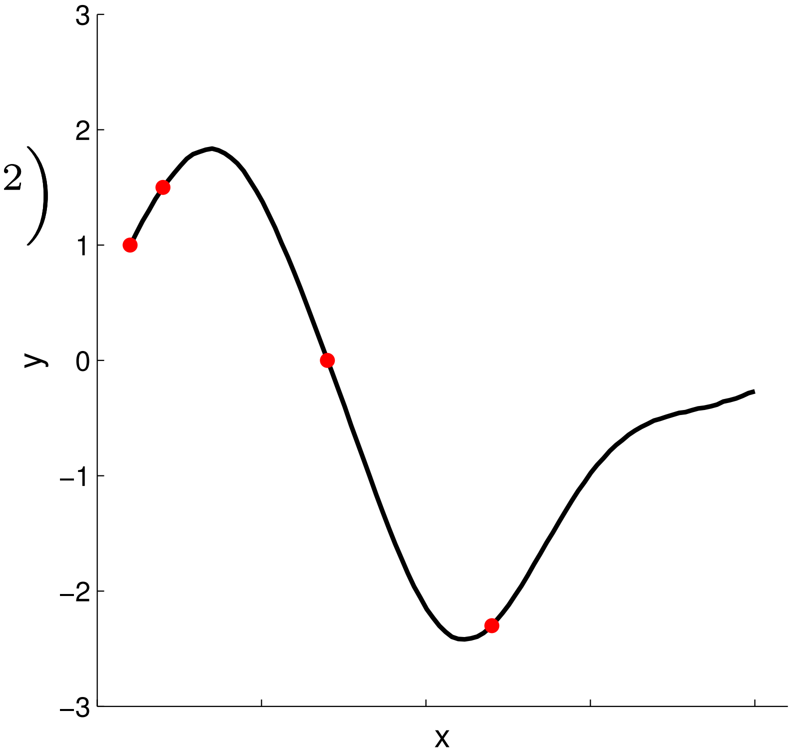
Non-parametric (∞-parametric)

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

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

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

$\Sigma =$ 



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?

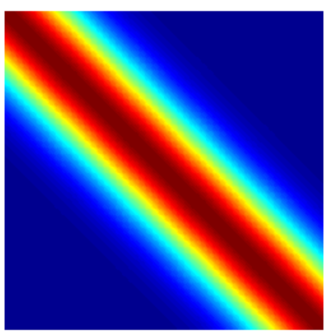
Non-parametric (∞-parametric)

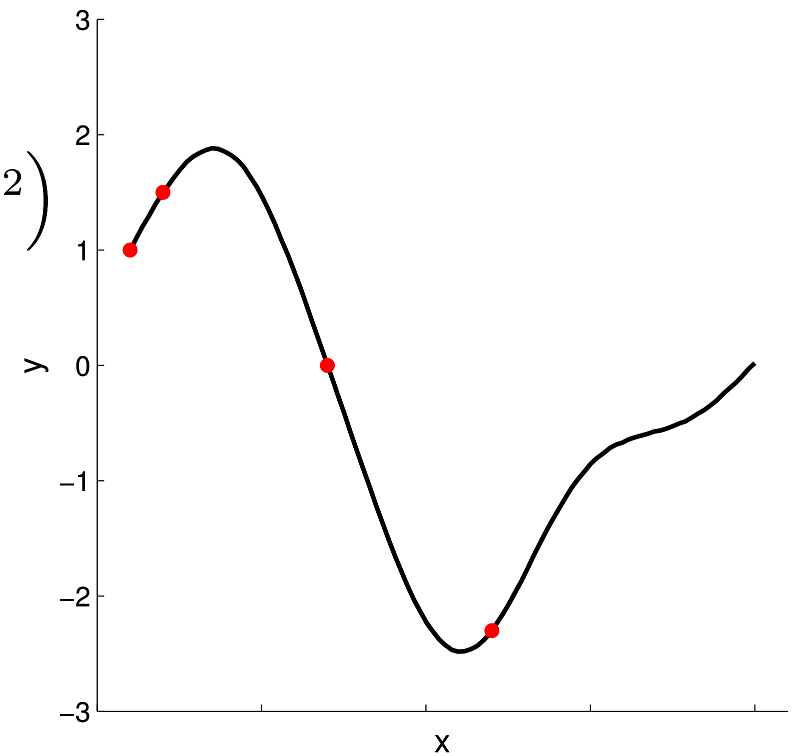
$$p(\mathsf{y}|\theta) = \mathcal{N}(0, \Sigma)$$

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

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

$$\Sigma =$$





Parametric model

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

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

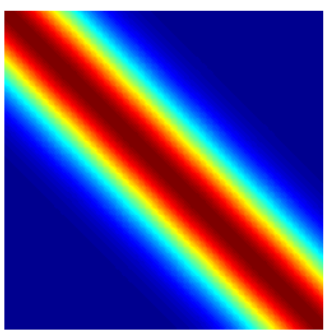
What effect do the hyper-parameters have?

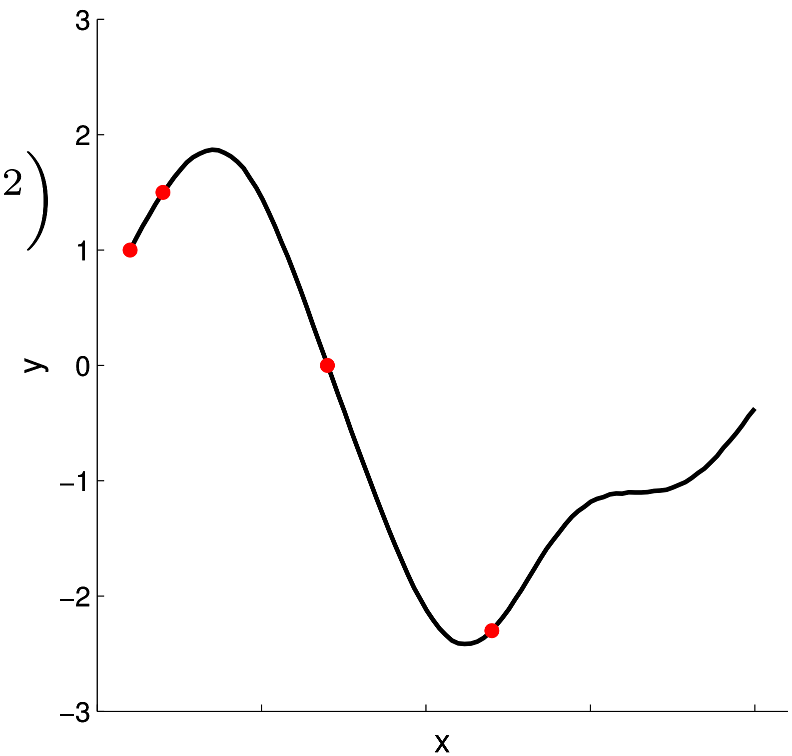
Non-parametric (∞-parametric)

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

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

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

$\Sigma =$ 



Parametric model

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

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

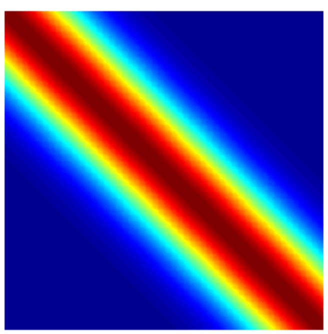
What effect do the hyper-parameters have?

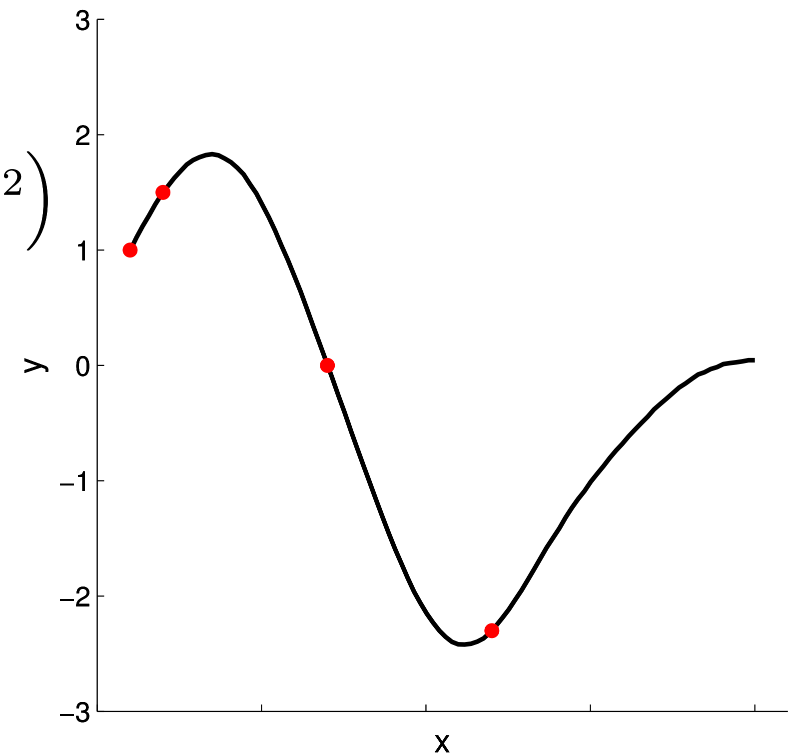
Non-parametric (∞-parametric)

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

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

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

$\Sigma =$ 



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?

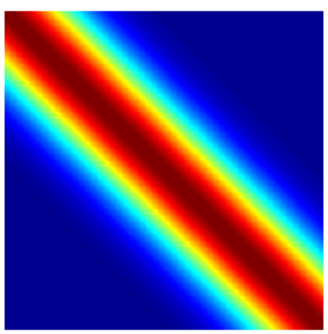
Non-parametric (∞-parametric)

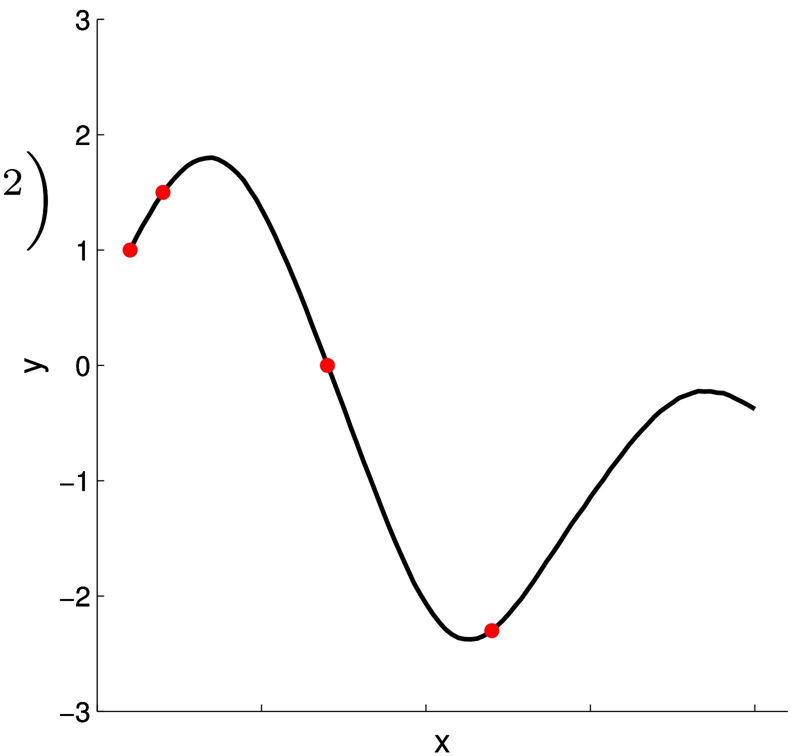
$$p(\mathsf{y}|\theta) = \mathcal{N}(0, \Sigma)$$

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

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

$$\Sigma =$$





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?

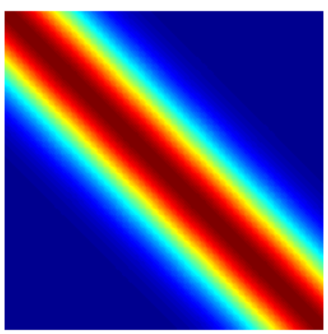
Non-parametric (∞-parametric)

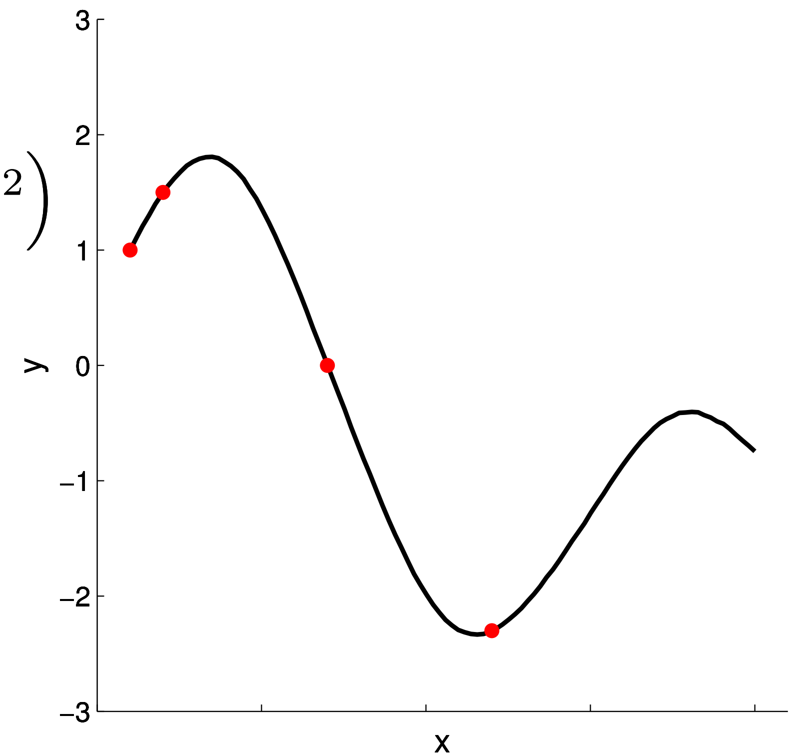
$$p(\mathsf{y}|\theta) = \mathcal{N}(0, \Sigma)$$

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

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

$$\Sigma =$$





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?

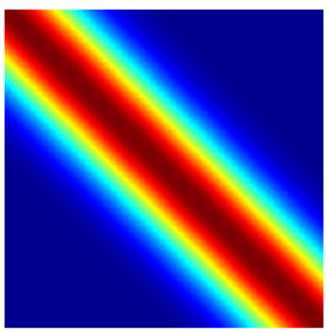
Non-parametric (∞-parametric)

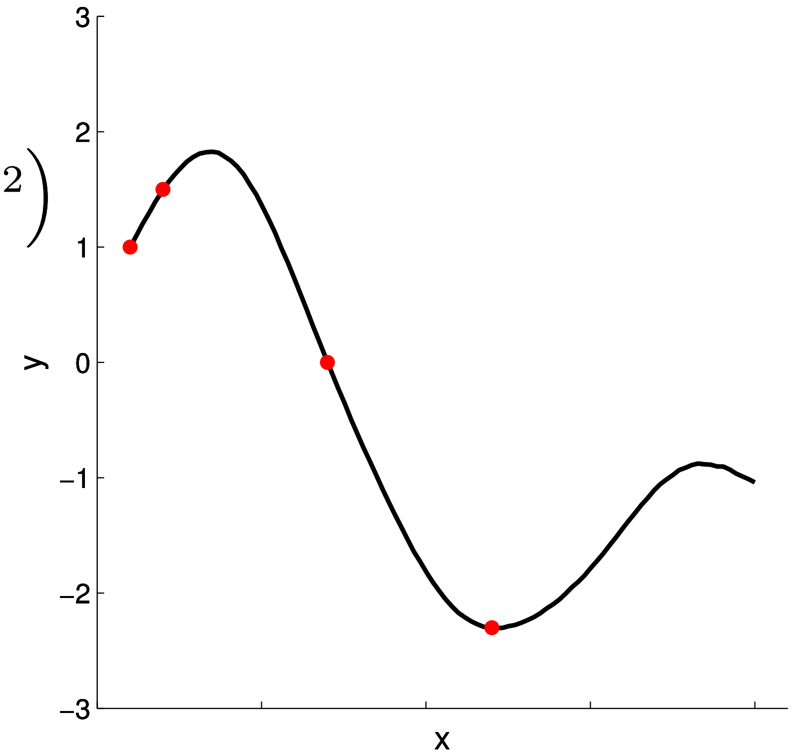
$$p(\mathsf{y}|\theta) = \mathcal{N}(0, \Sigma)$$

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

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

$\Sigma =$





Parametric model

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

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

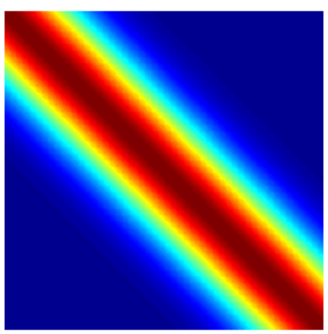
What effect do the hyper-parameters have?

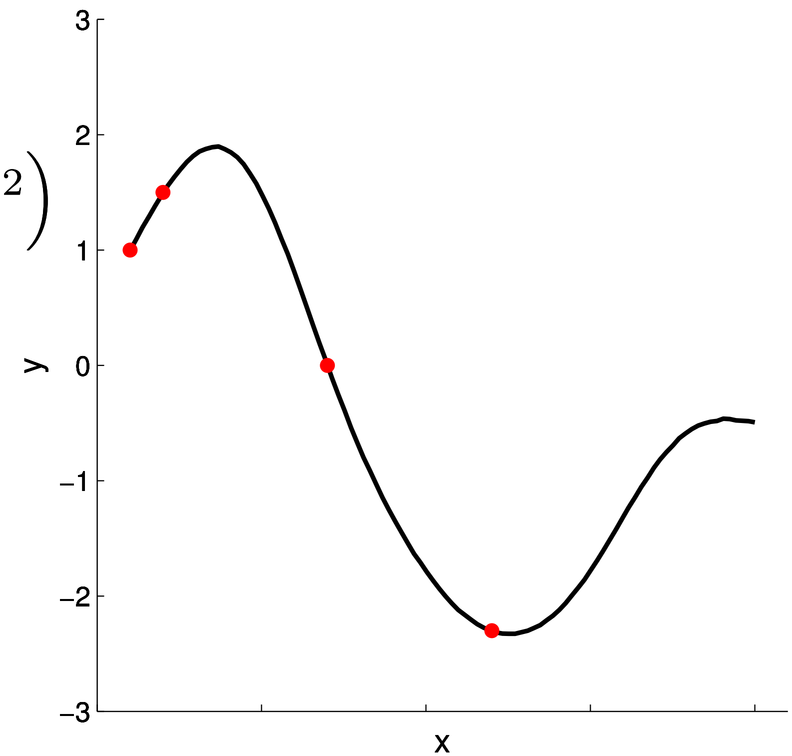
Non-parametric (∞-parametric)

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

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

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

$\Sigma =$ 



Parametric model

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

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

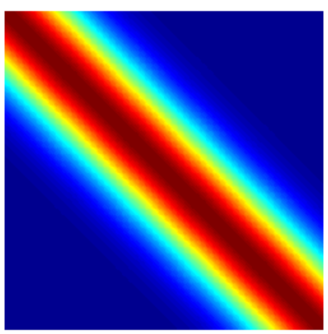
What effect do the hyper-parameters have?

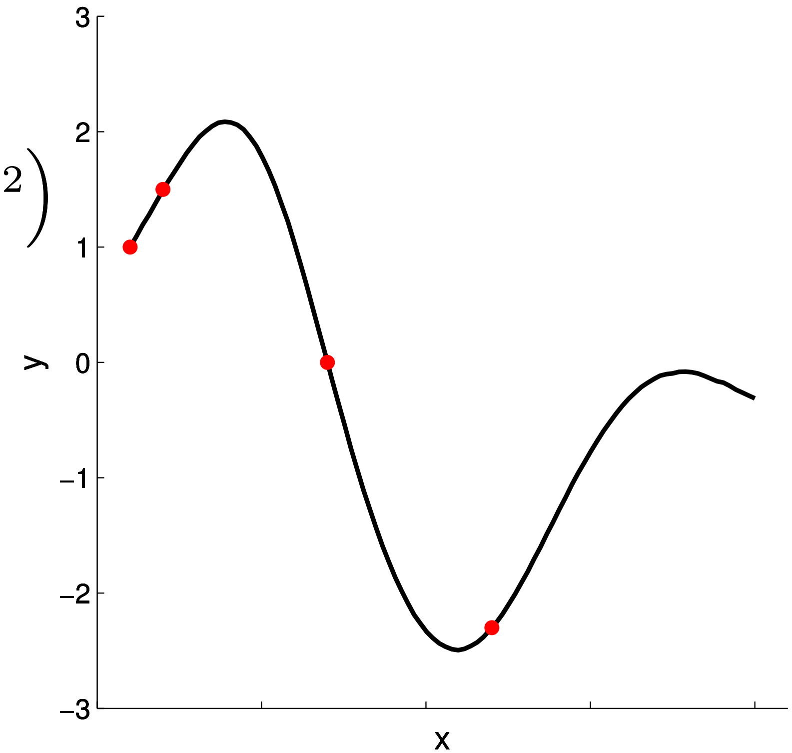
Non-parametric (∞-parametric)

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

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

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

$\Sigma =$ 



Parametric model

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

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

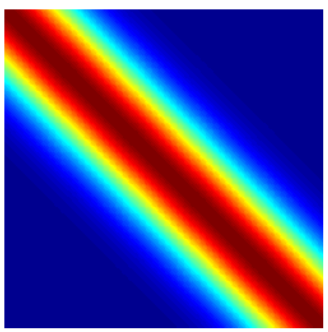
What effect do the hyper-parameters have?

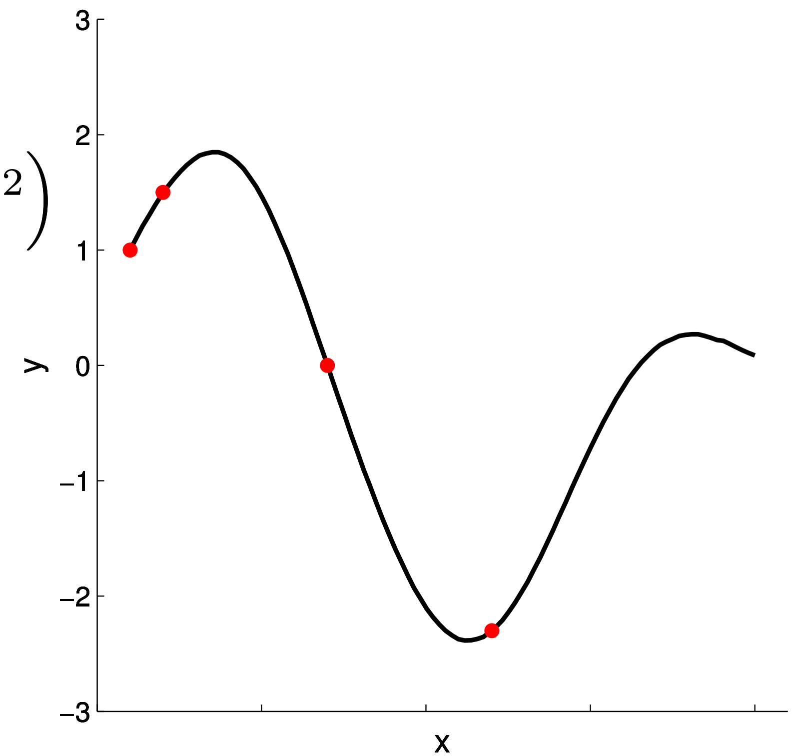
Non-parametric (∞-parametric)

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

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

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

$\Sigma =$ 



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?

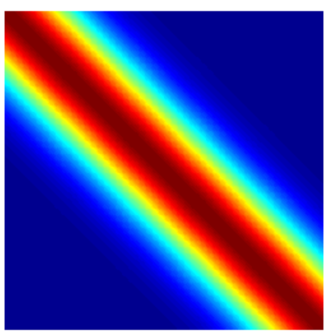
Non-parametric (∞-parametric)

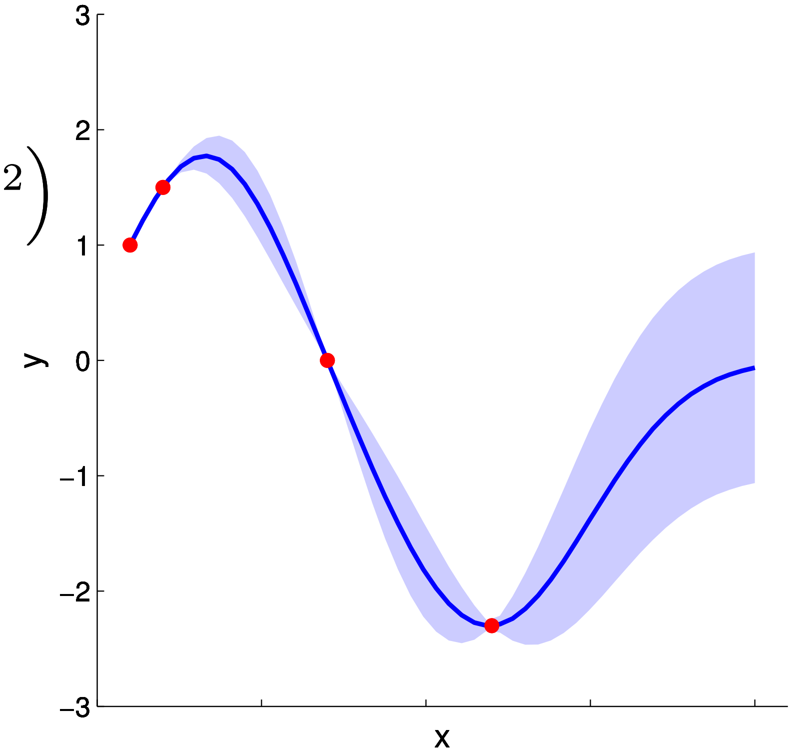
$$p(\mathsf{y}|\theta) = \mathcal{N}(0, \Sigma)$$

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

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

$$\Sigma =$$





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?

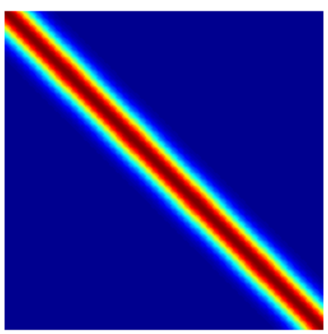
short horizontal length-scale

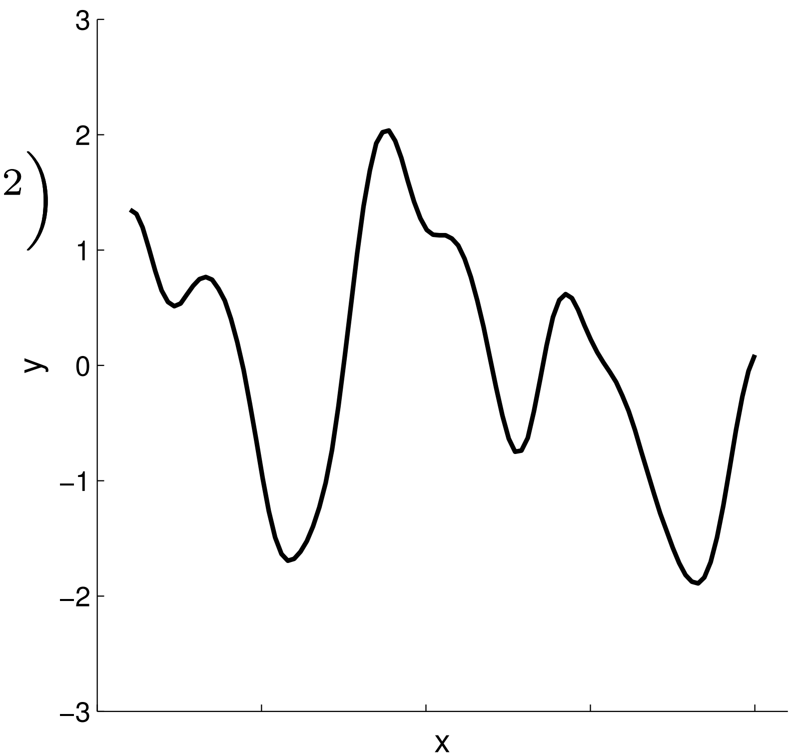
Non-parametric (∞-parametric)

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

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

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

$\Sigma =$ 



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?

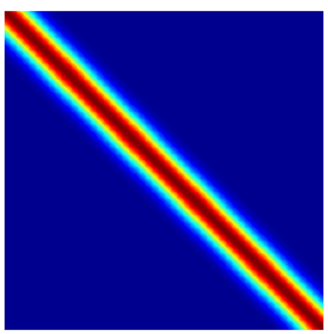
short horizontal length-scale

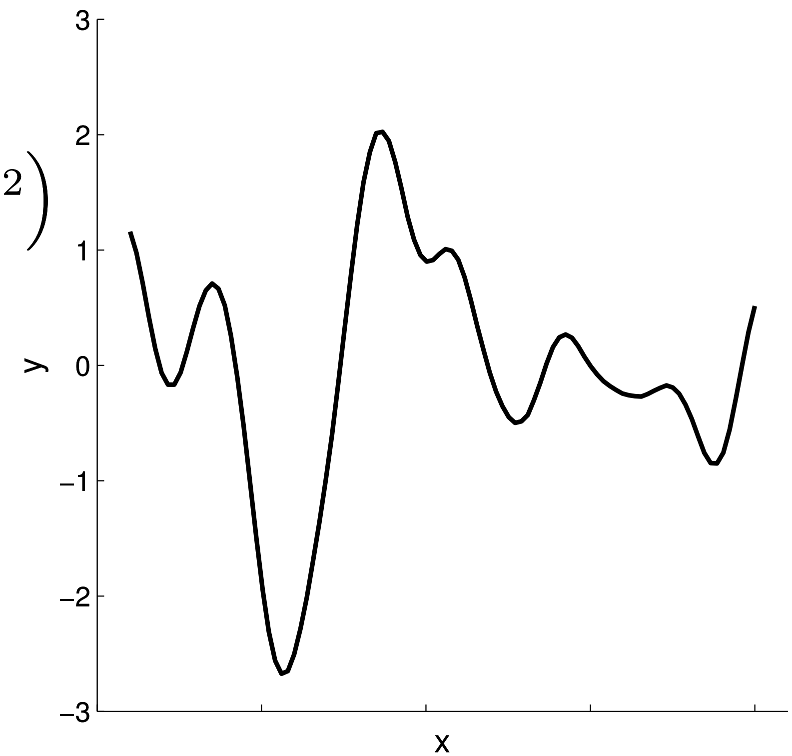
Non-parametric (∞-parametric)

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

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

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

$\Sigma =$ 



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

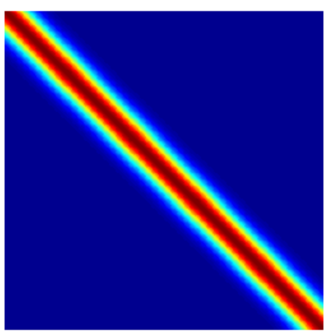
Non-parametric (∞-parametric)

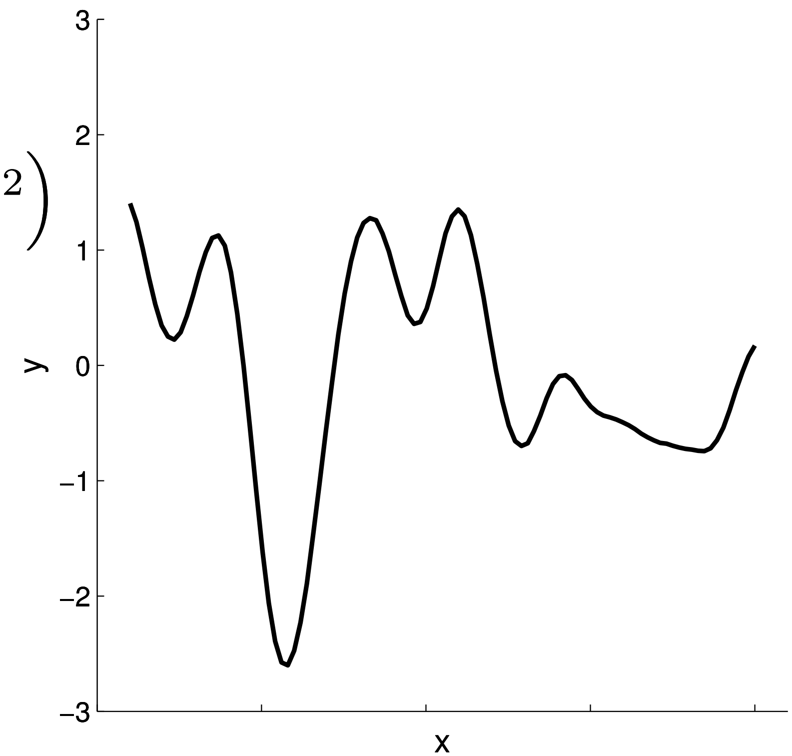
$$p(\mathsf{y}|\theta) = \mathcal{N}(0, \Sigma)$$

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

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

$\Sigma =$





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?

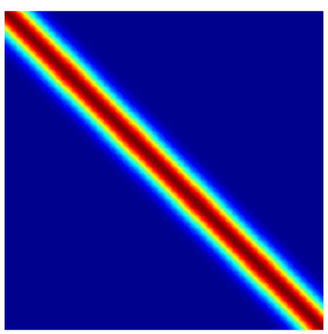
short horizontal length-scale

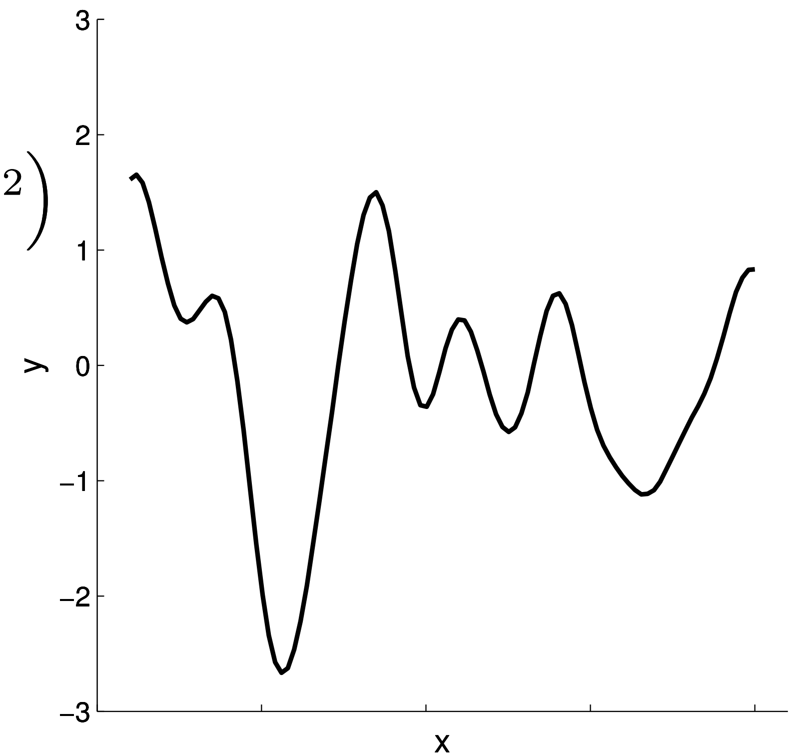
Non-parametric (∞-parametric)

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

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

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

$\Sigma =$ 



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?

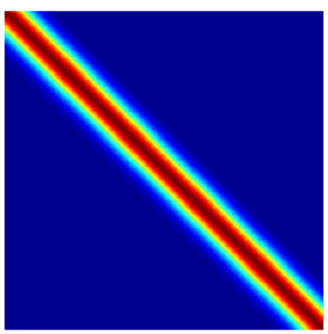
short horizontal length-scale

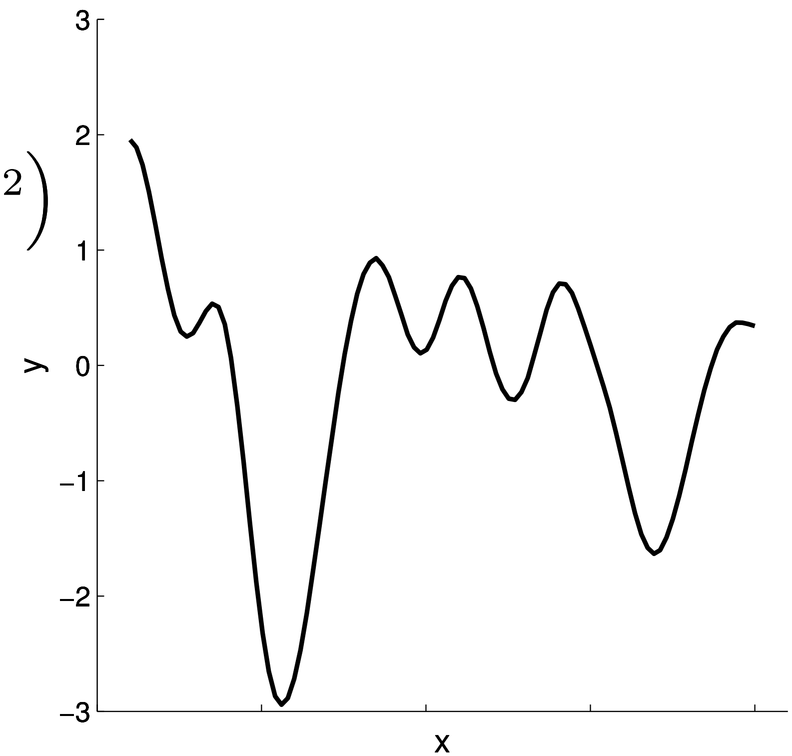
Non-parametric (∞-parametric)

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

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

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

$\Sigma =$ 



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?

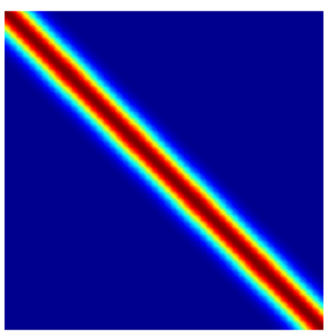
short horizontal length-scale

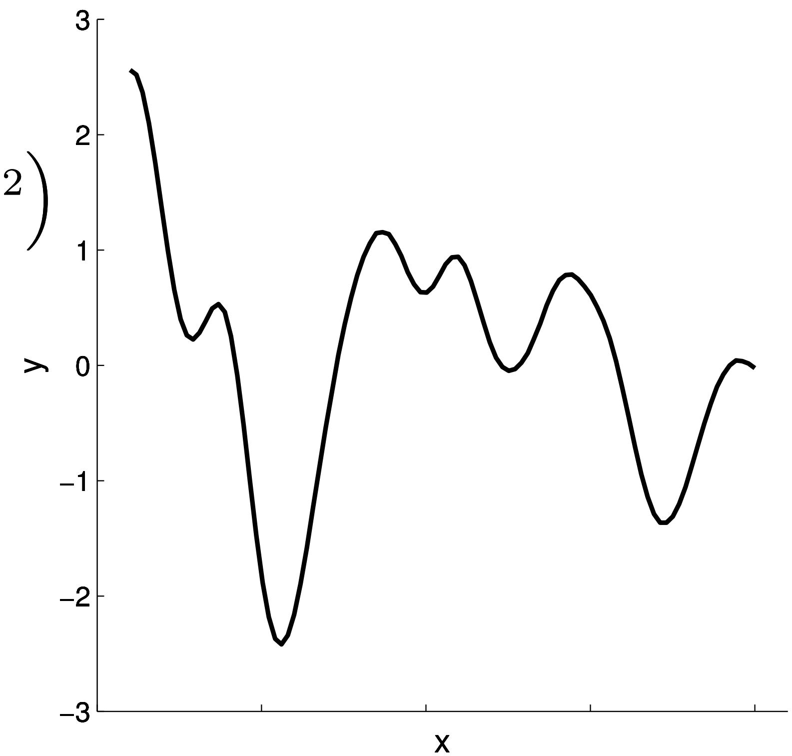
Non-parametric (∞-parametric)

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

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

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

$\Sigma =$ 



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

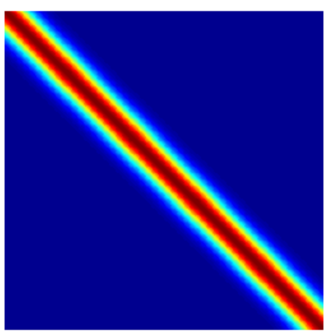
Non-parametric (∞-parametric)

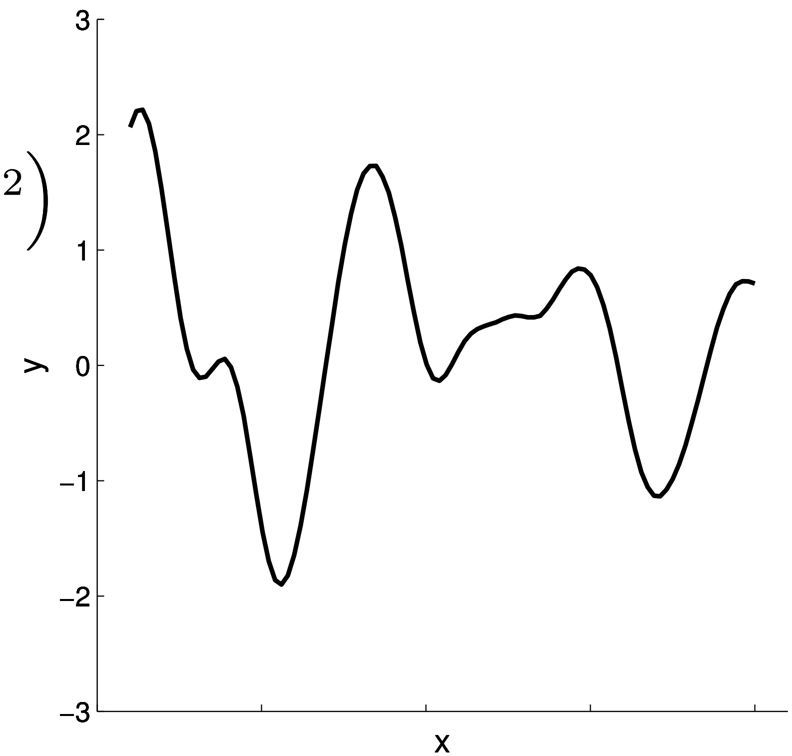
$$p(\mathsf{y}|\theta) = \mathcal{N}(0, \Sigma)$$

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

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

$$\Sigma =$$





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?

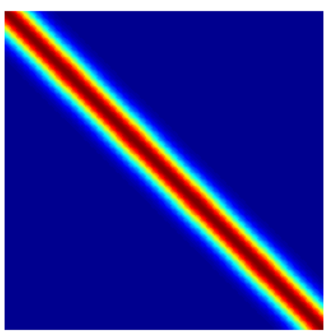
short horizontal length-scale

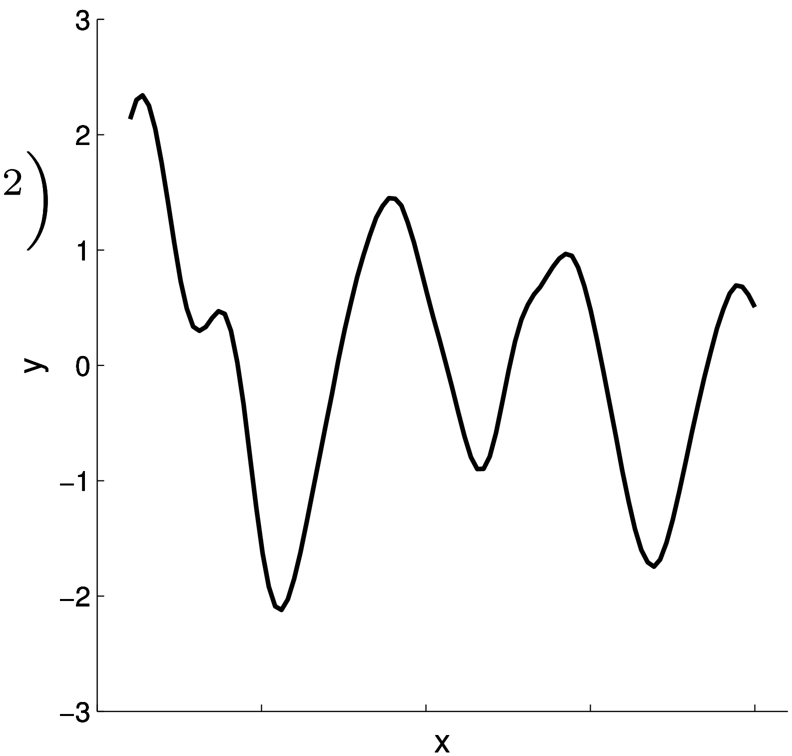
Non-parametric (∞-parametric)

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

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

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

$\Sigma =$ 



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?

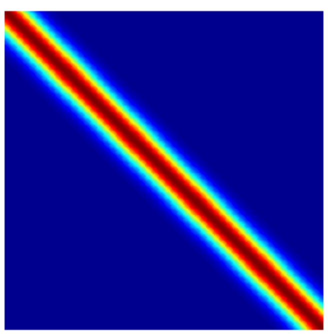
short horizontal length-scale

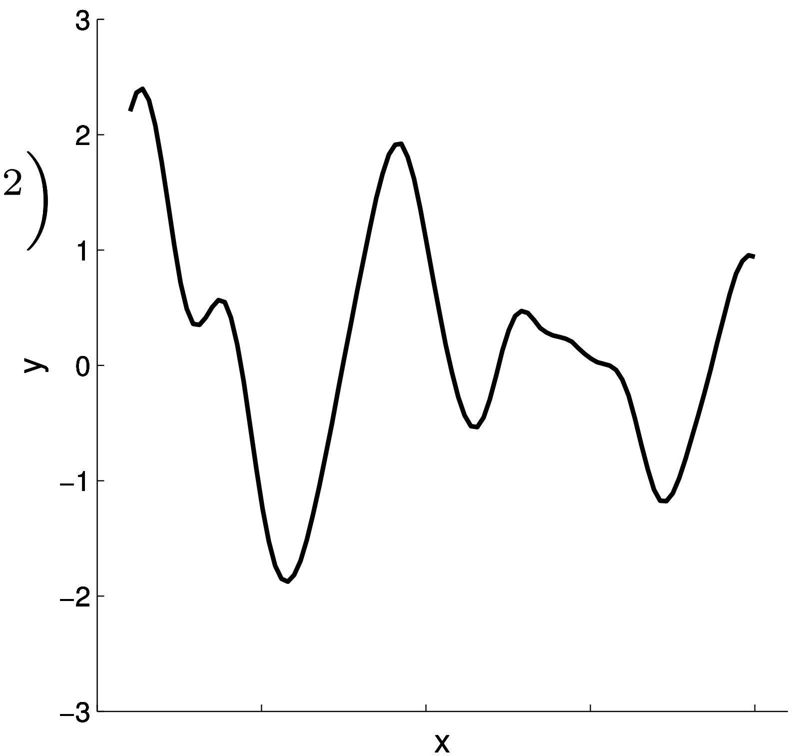
Non-parametric (∞-parametric)

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

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

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

$\Sigma =$ 



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?

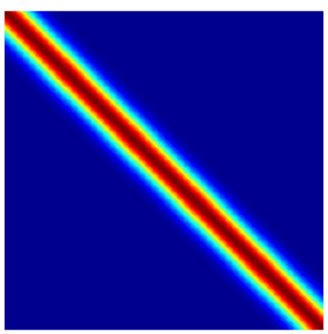
short horizontal length-scale

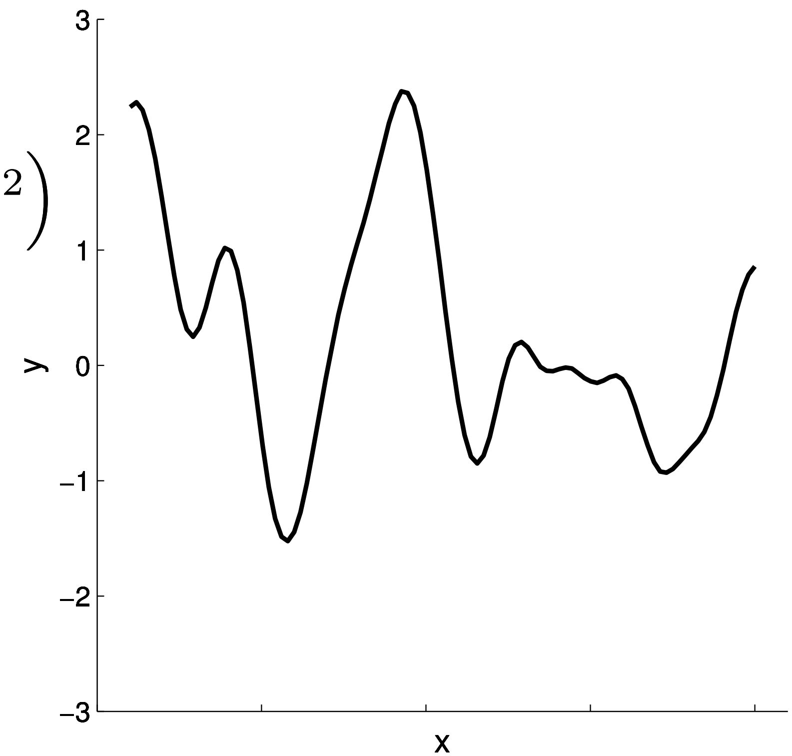
Non-parametric (∞-parametric)

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

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

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

$\Sigma =$ 



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

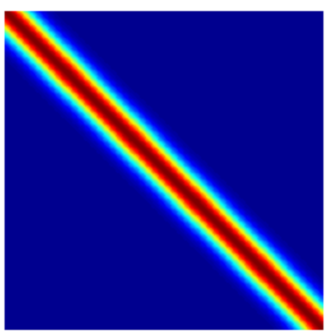
Non-parametric (∞-parametric)

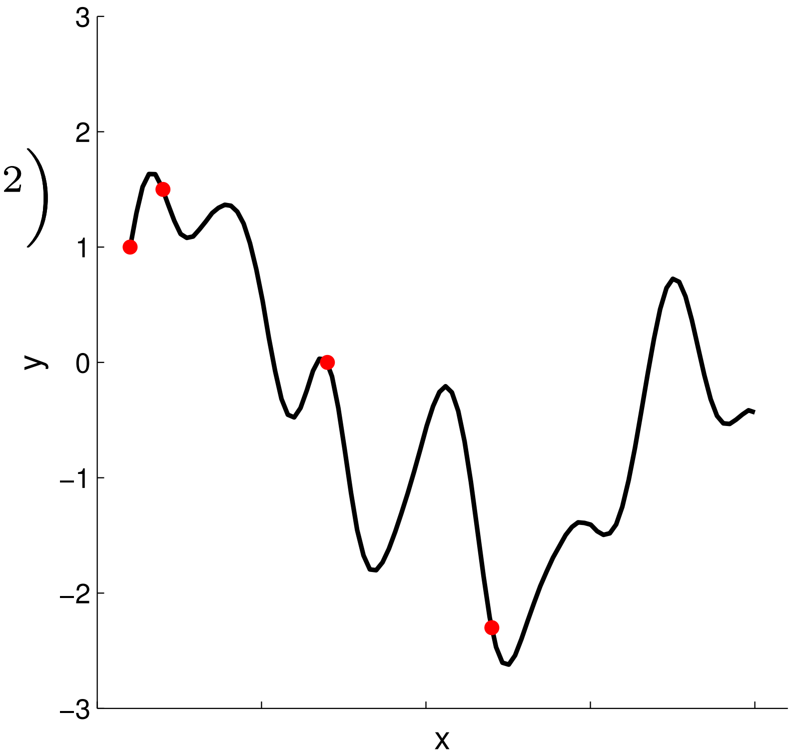
$$p(\mathsf{y}|\theta) = \mathcal{N}(0, \Sigma)$$

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

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

$$\Sigma =$$





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?

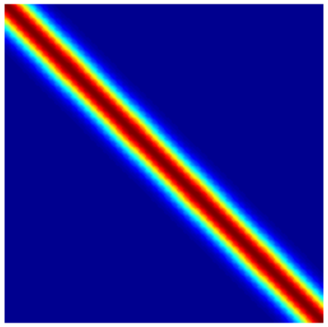
short horizontal length-scale

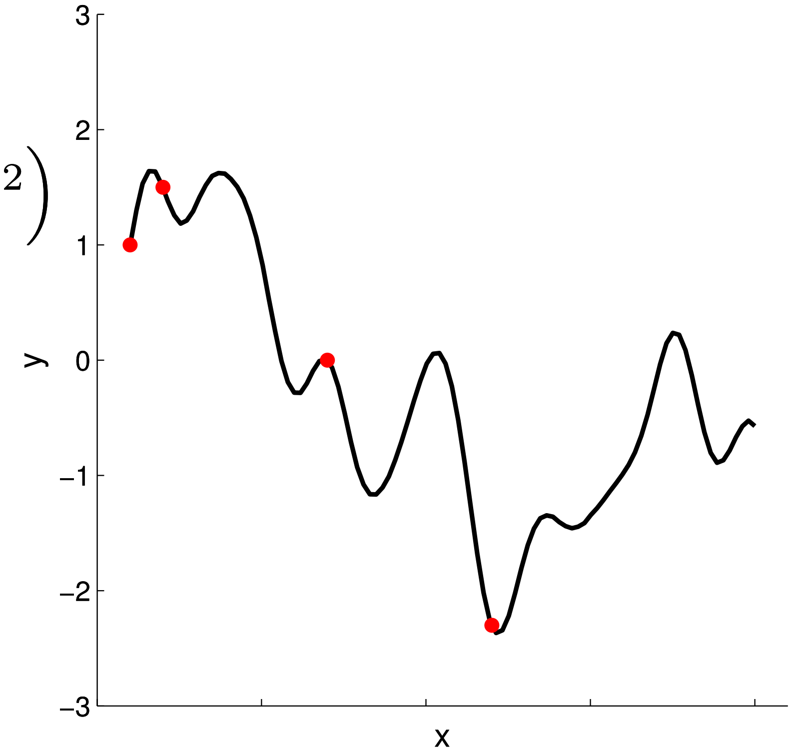
Non-parametric (∞-parametric)

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

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

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

$\Sigma =$ 



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?

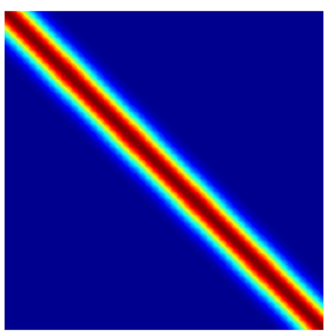
short horizontal length-scale

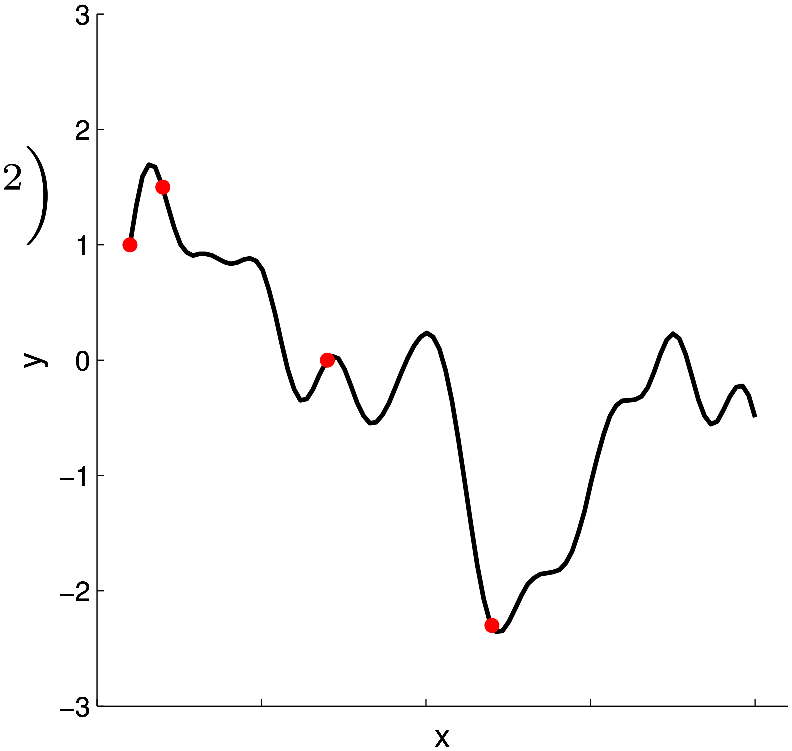
Non-parametric (∞-parametric)

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

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

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

$\Sigma =$ 



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

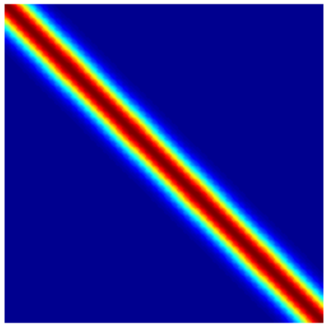
Non-parametric (∞-parametric)

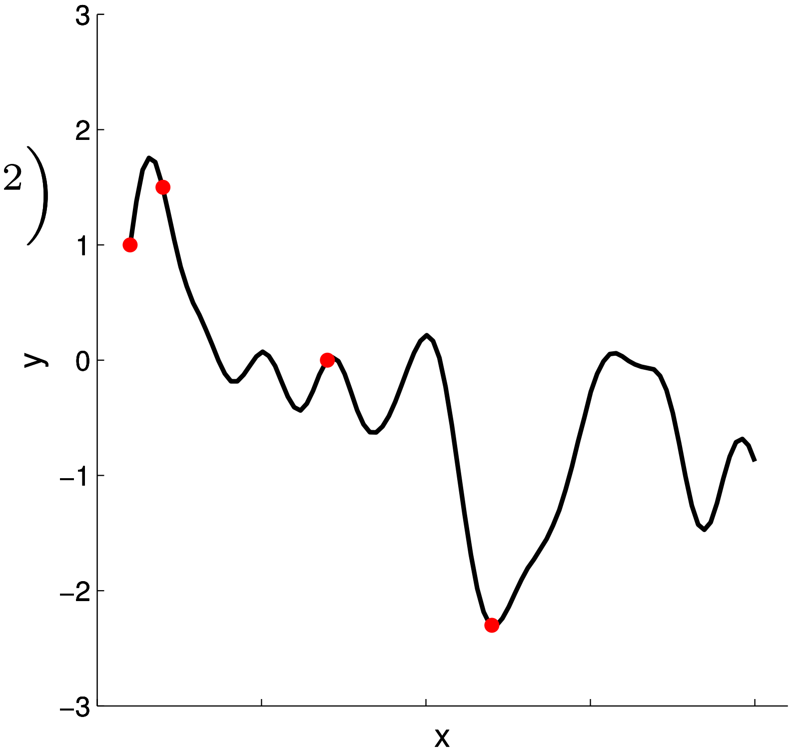
$$p(\mathsf{y}|\theta) = \mathcal{N}(0, \Sigma)$$

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

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

$$\Sigma =$$





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?

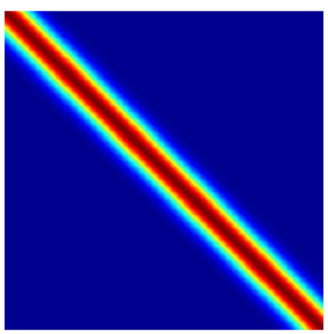
short horizontal length-scale

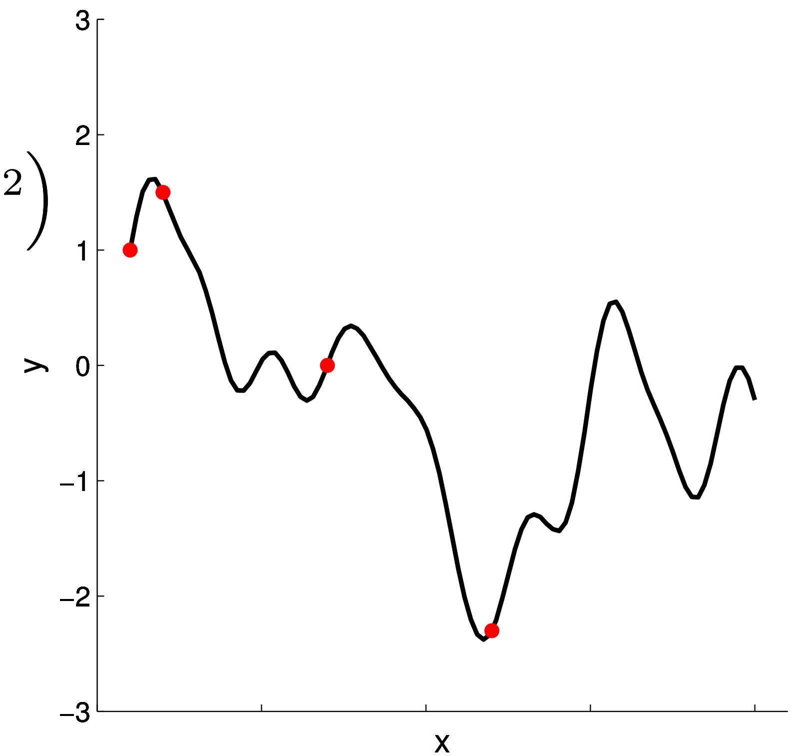
Non-parametric (∞-parametric)

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

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

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

$\Sigma =$ 



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?

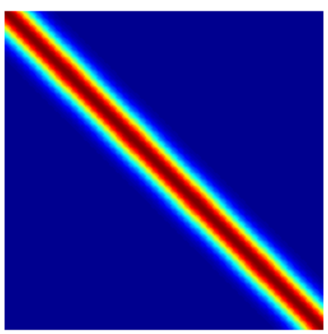
short horizontal length-scale

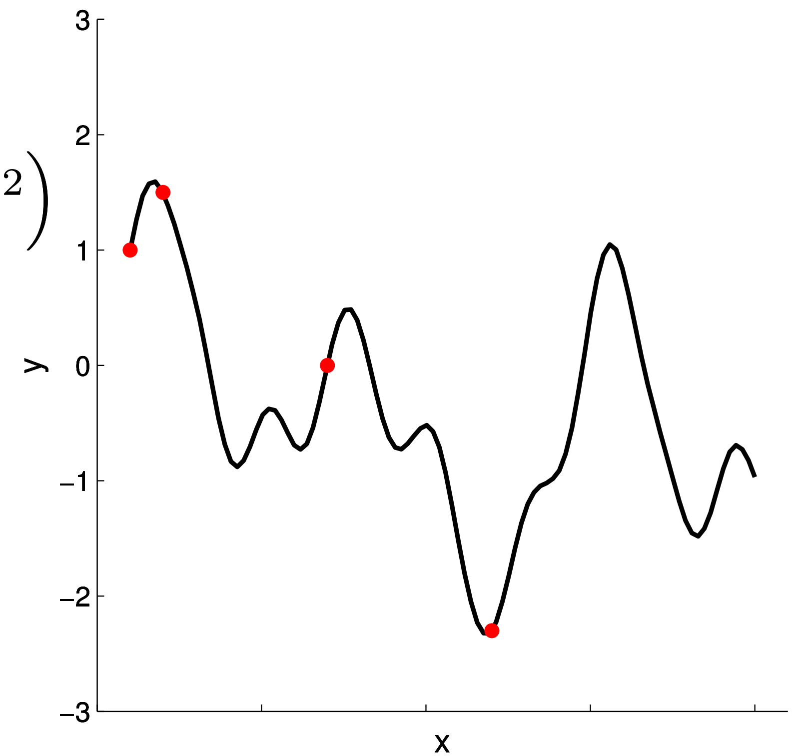
Non-parametric (∞-parametric)

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

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

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

$\Sigma =$ 



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?

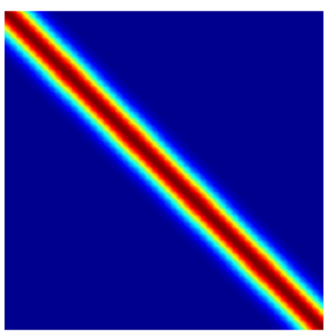
short horizontal length-scale

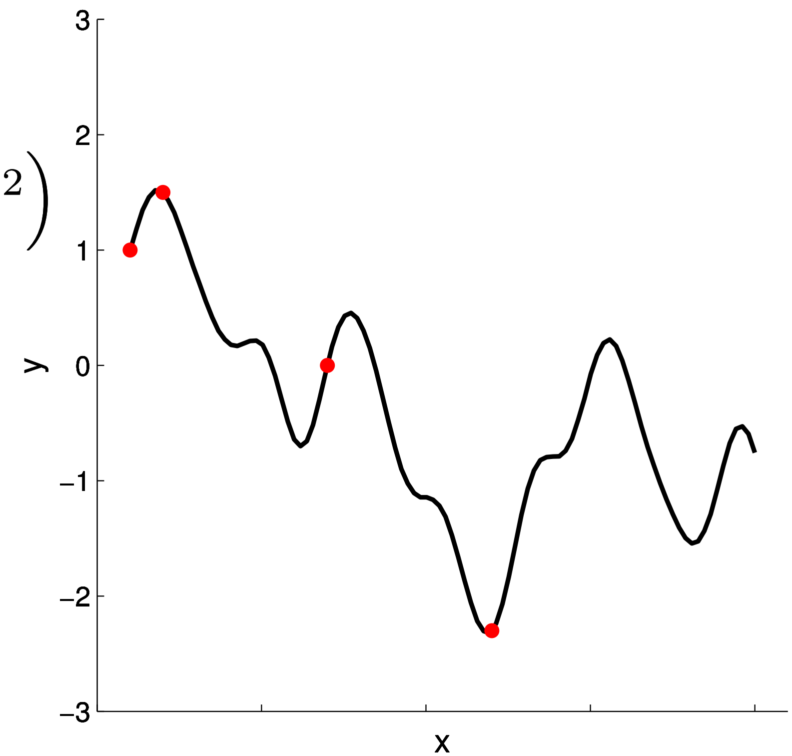
Non-parametric (∞-parametric)

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

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

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

$\Sigma =$ 



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?

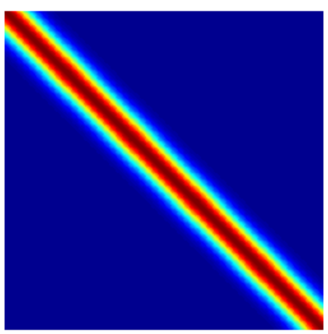
short horizontal length-scale

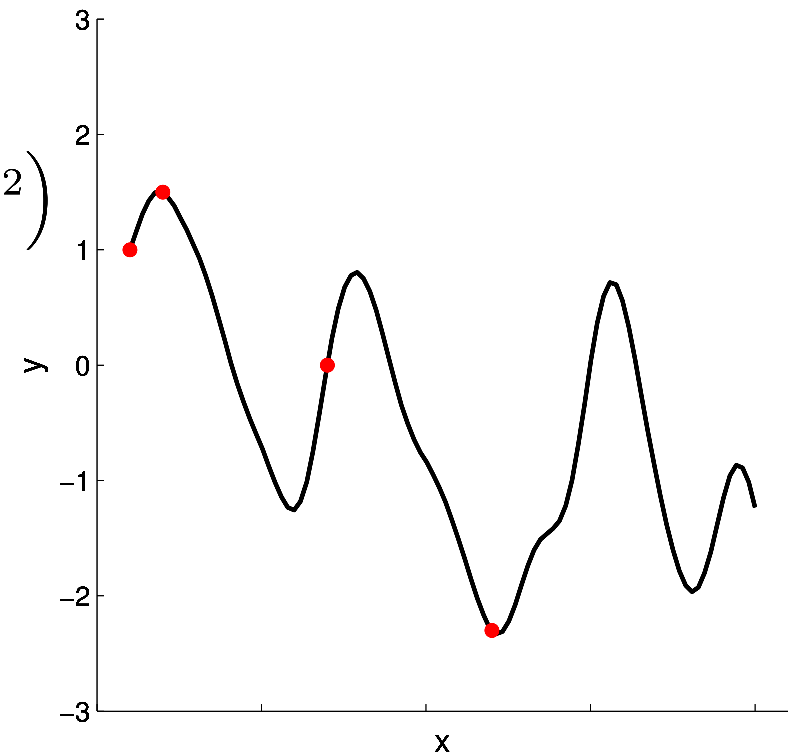
Non-parametric (∞-parametric)

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

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

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

$\Sigma =$ 



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?

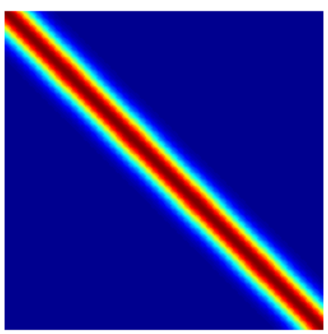
short horizontal length-scale

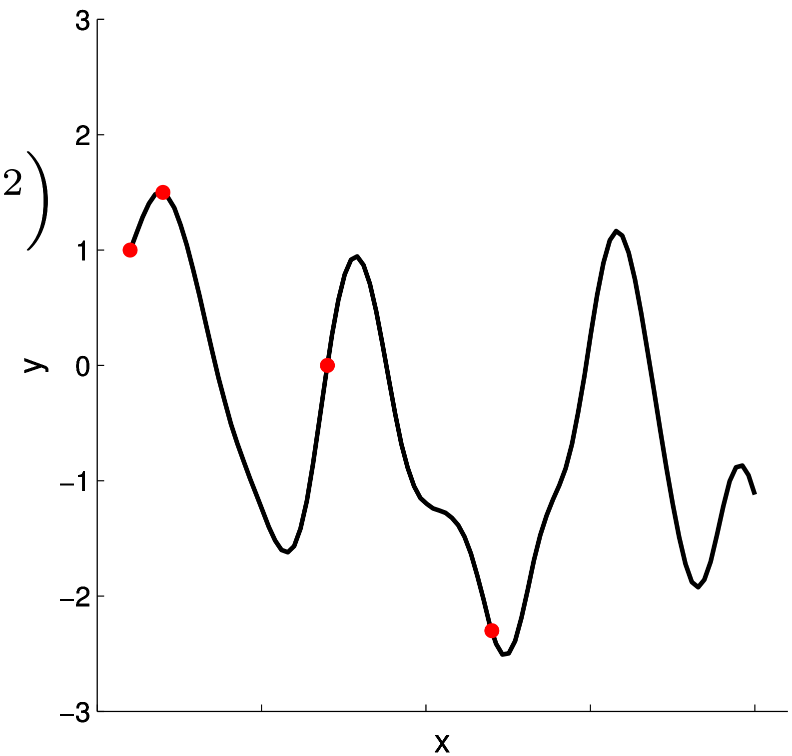
Non-parametric (∞-parametric)

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

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

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

$\Sigma =$ 



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?

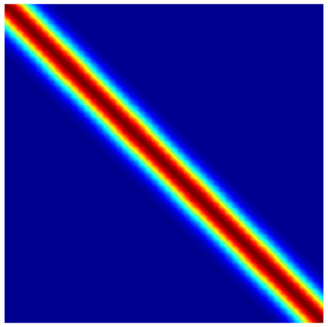
short horizontal length-scale

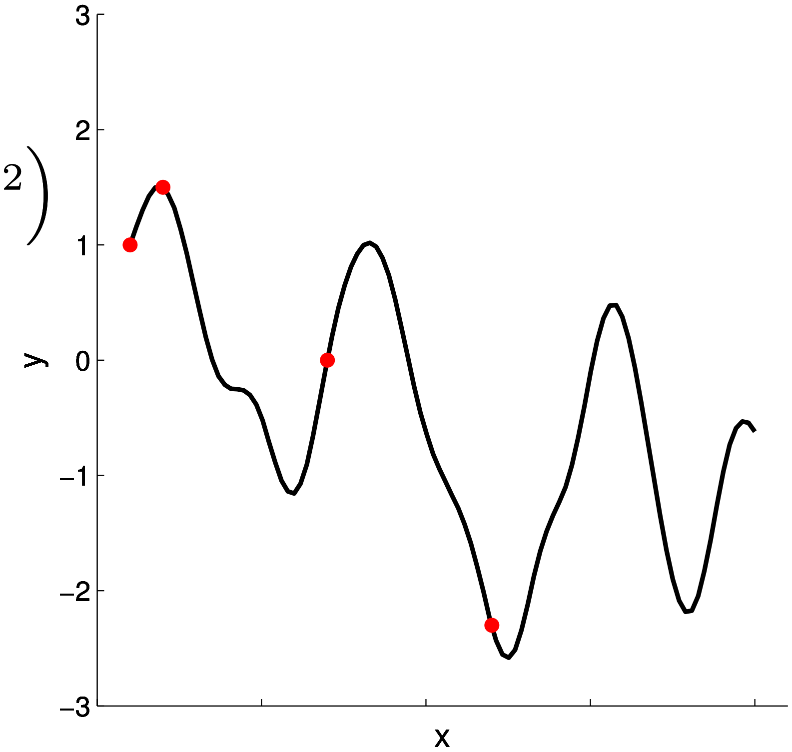
Non-parametric (∞-parametric)

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

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

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

$\Sigma =$ 



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?

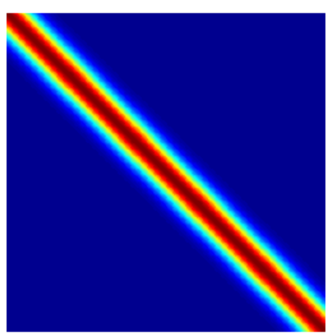
short horizontal length-scale

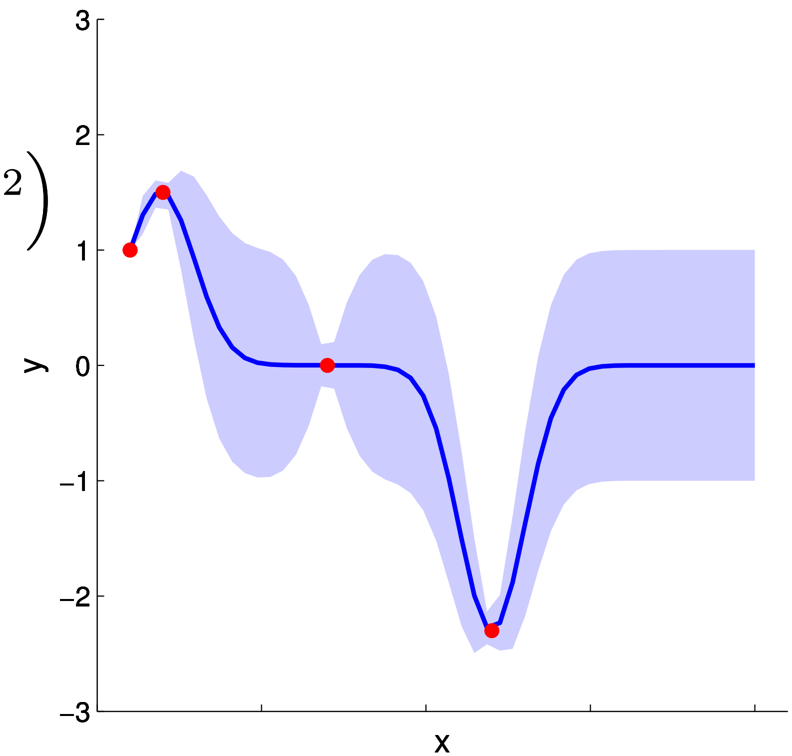
Non-parametric (∞-parametric)

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

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

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

$\Sigma =$ 



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?

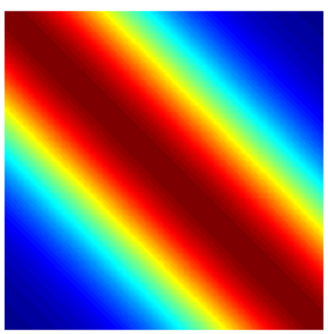
long horizontal length-scale

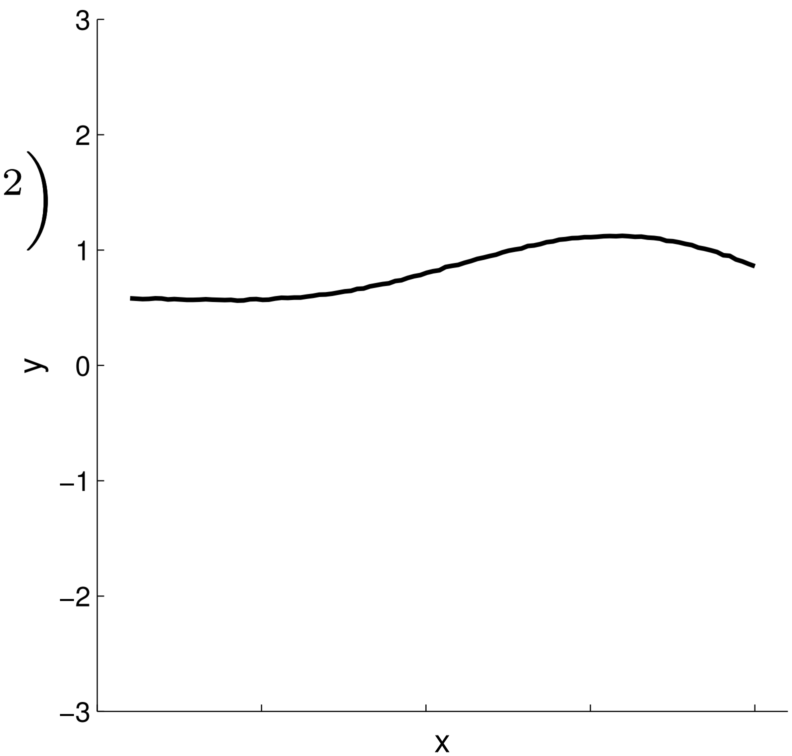
Non-parametric (∞-parametric)

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

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

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

$\Sigma =$ 



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?

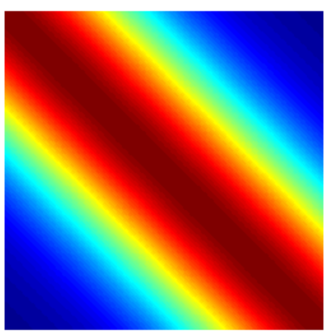
long horizontal length-scale

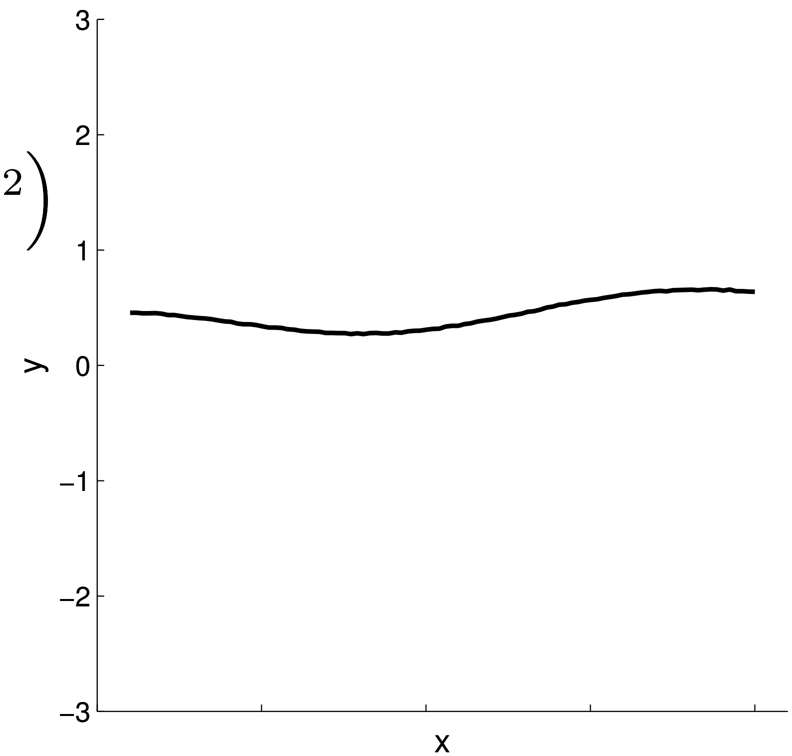
Non-parametric (∞-parametric)

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

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

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

$\Sigma =$ 



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

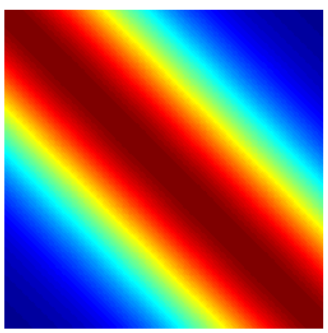
Non-parametric (∞-parametric)

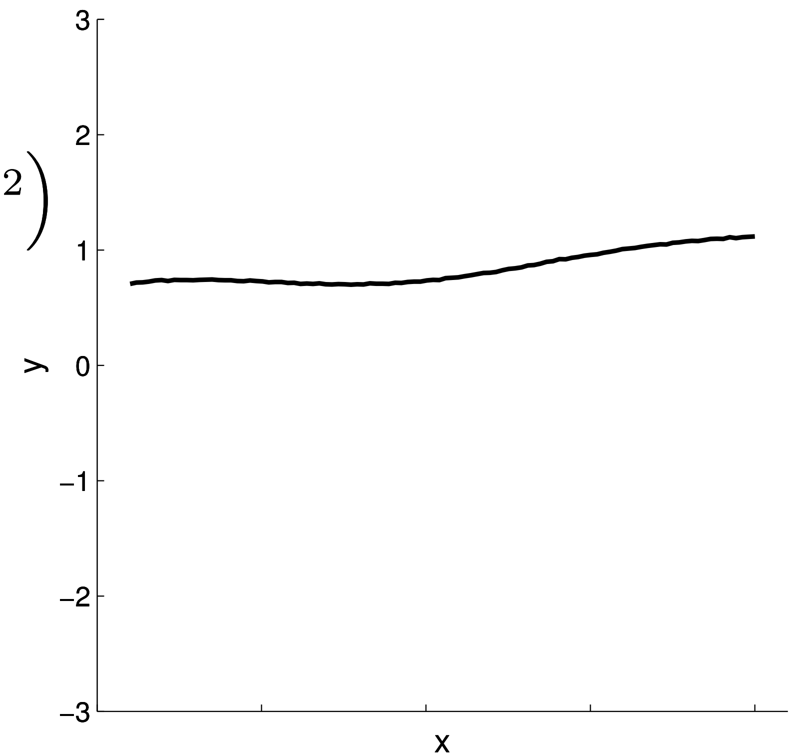
$$p(\mathsf{y}|\theta) = \mathcal{N}(0, \Sigma)$$

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

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

$$\Sigma =$$





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?

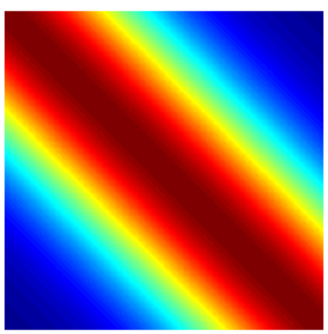
long horizontal length-scale

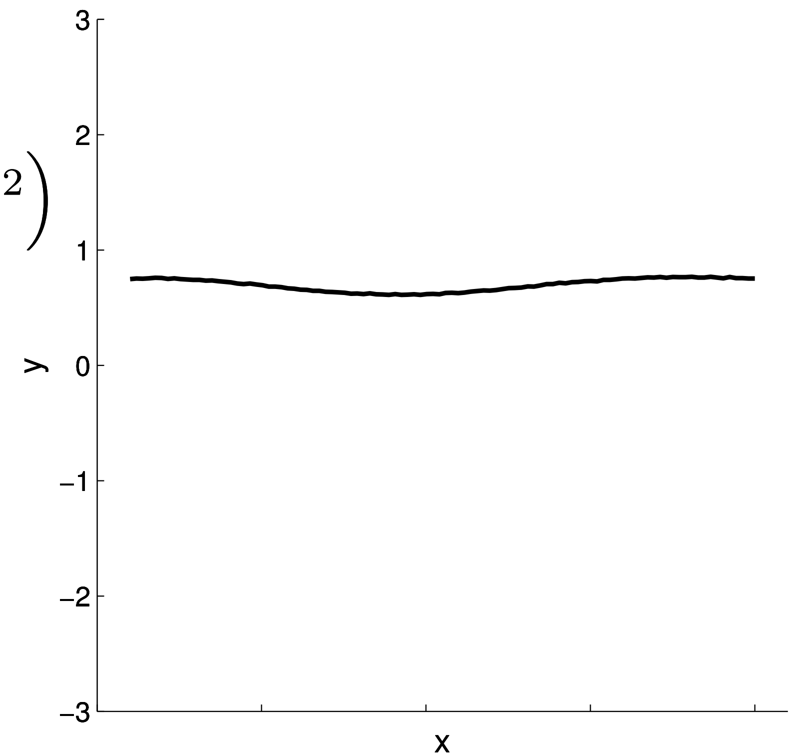
Non-parametric (∞-parametric)

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

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

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

$\Sigma =$ 



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?

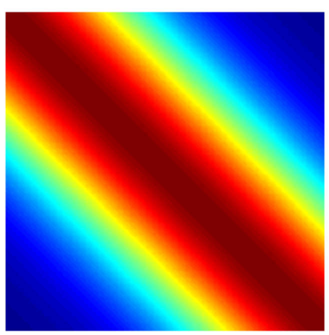
long horizontal length-scale

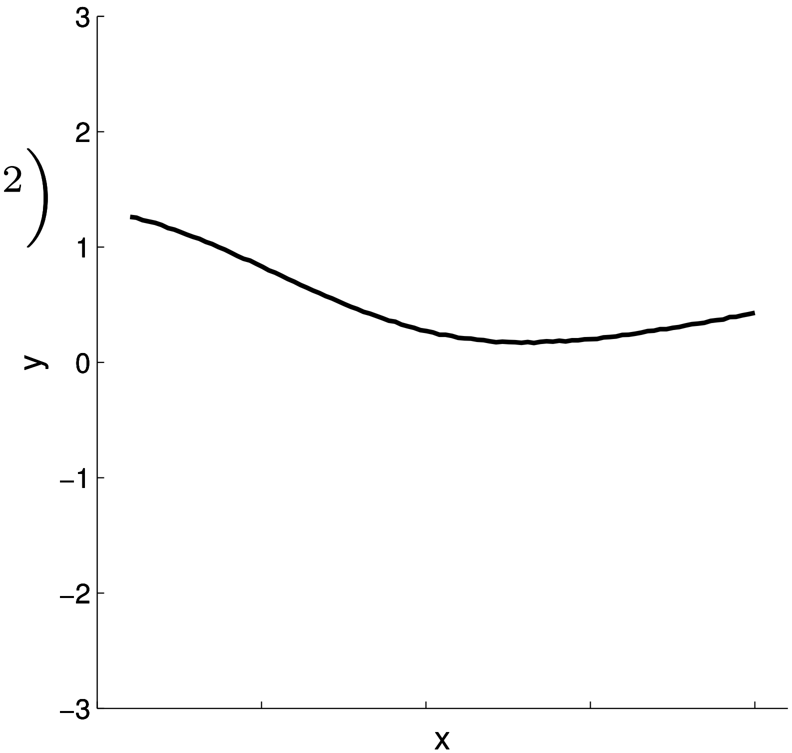
Non-parametric (∞-parametric)

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

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

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

$\Sigma =$ 



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

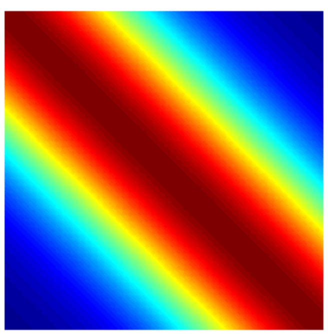
Non-parametric (∞-parametric)

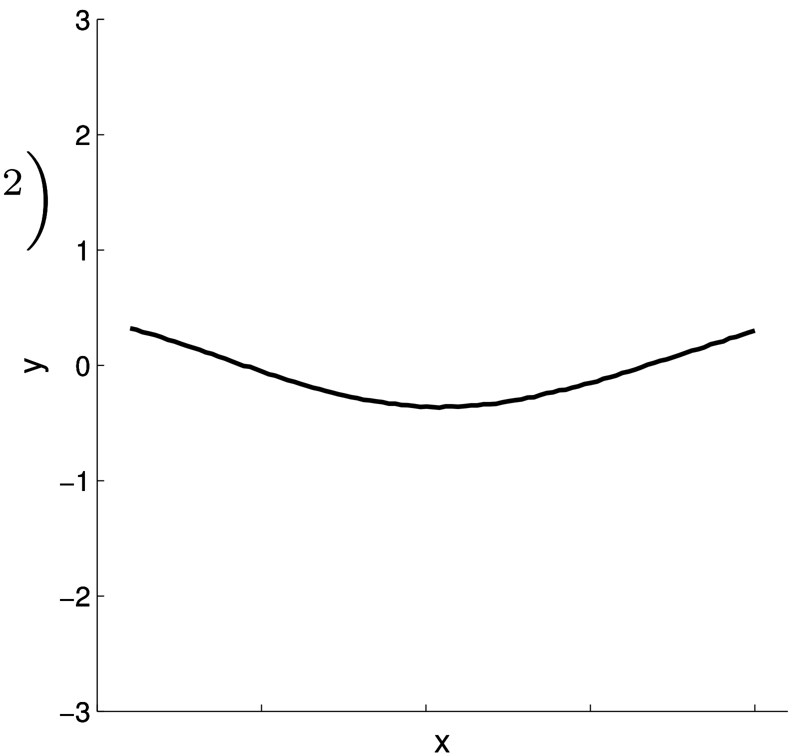
$$p(\mathsf{y}|\theta) = \mathcal{N}(0, \Sigma)$$

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

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

$$\Sigma =$$





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?

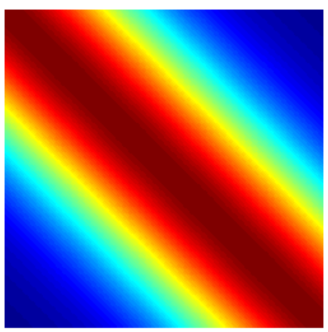
long horizontal length-scale

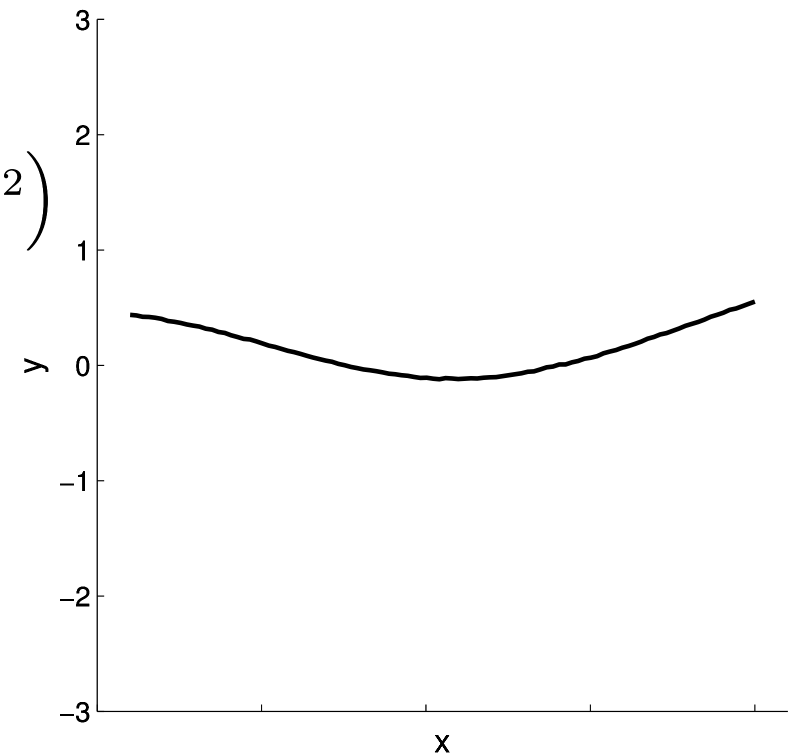
Non-parametric (∞-parametric)

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

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

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

$\Sigma =$ 



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

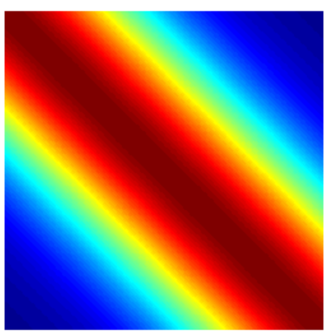
Non-parametric (∞-parametric)

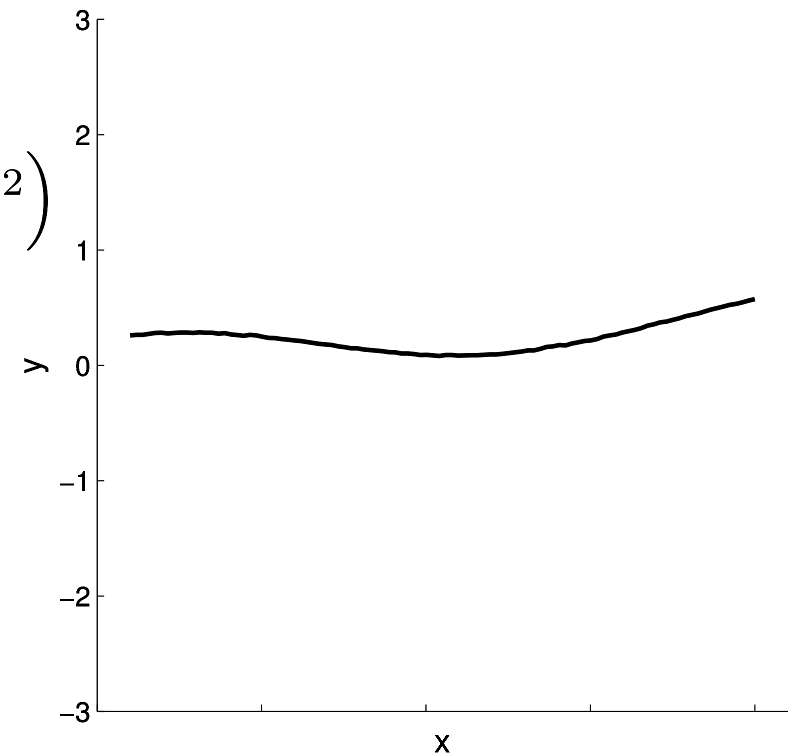
$p(\mathsf{y}|\theta) = \mathcal{N}(0, \Sigma)$

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

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

$\Sigma =$





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?

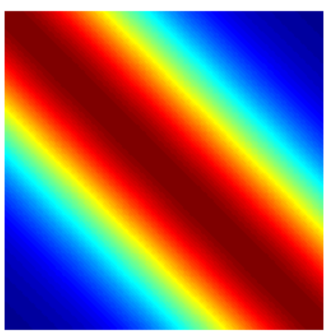
long horizontal length-scale

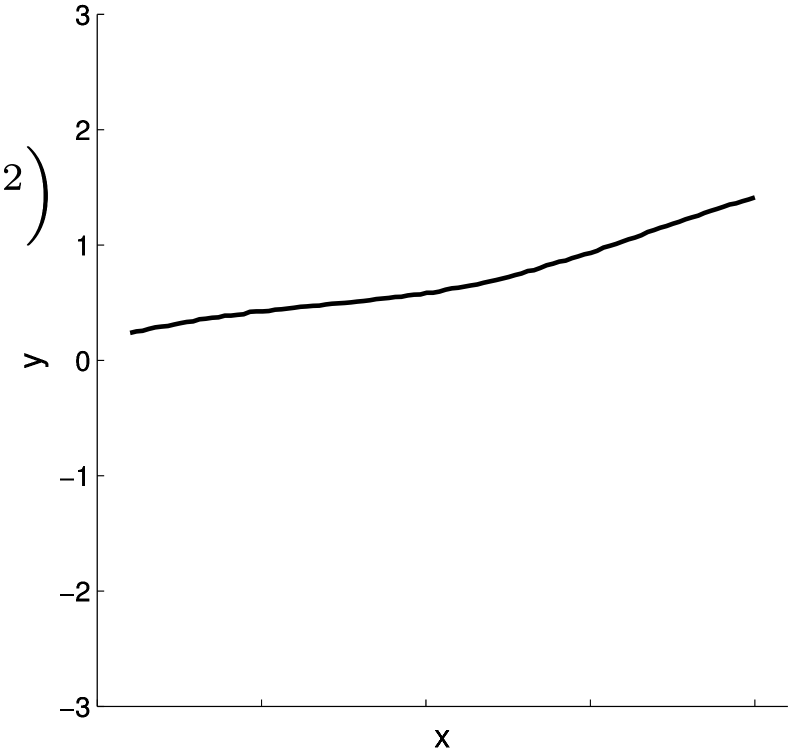
Non-parametric (∞-parametric)

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

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

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

$\Sigma =$ 



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?

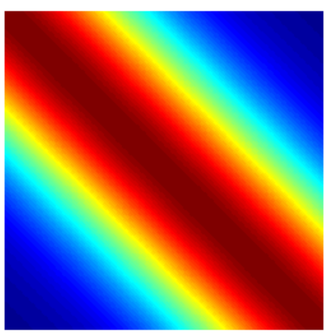
long horizontal length-scale

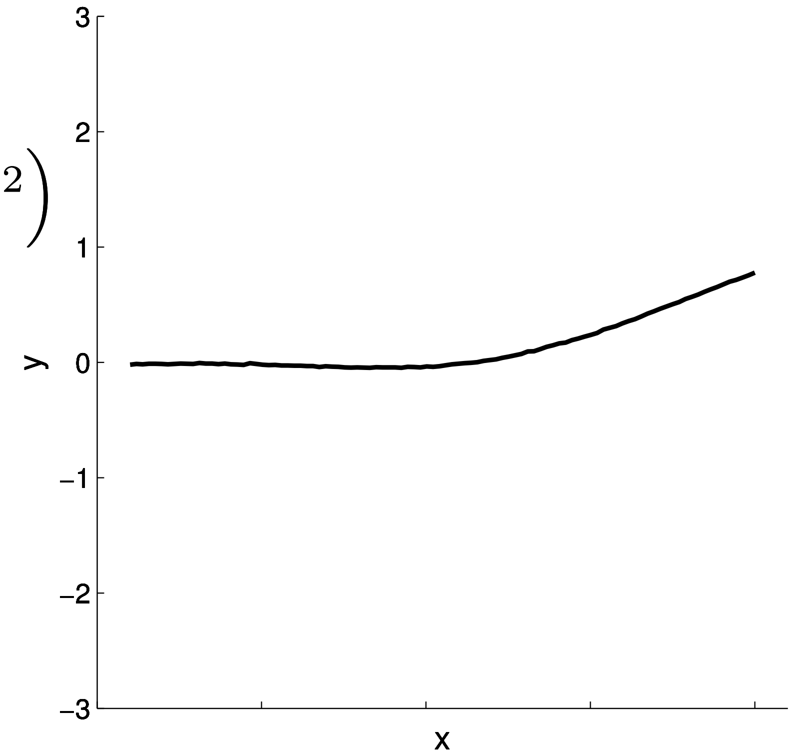
Non-parametric (∞-parametric)

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

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

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

$\Sigma =$ 



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?

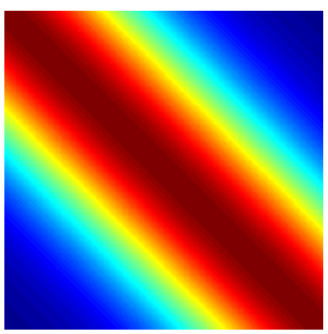
long horizontal length-scale

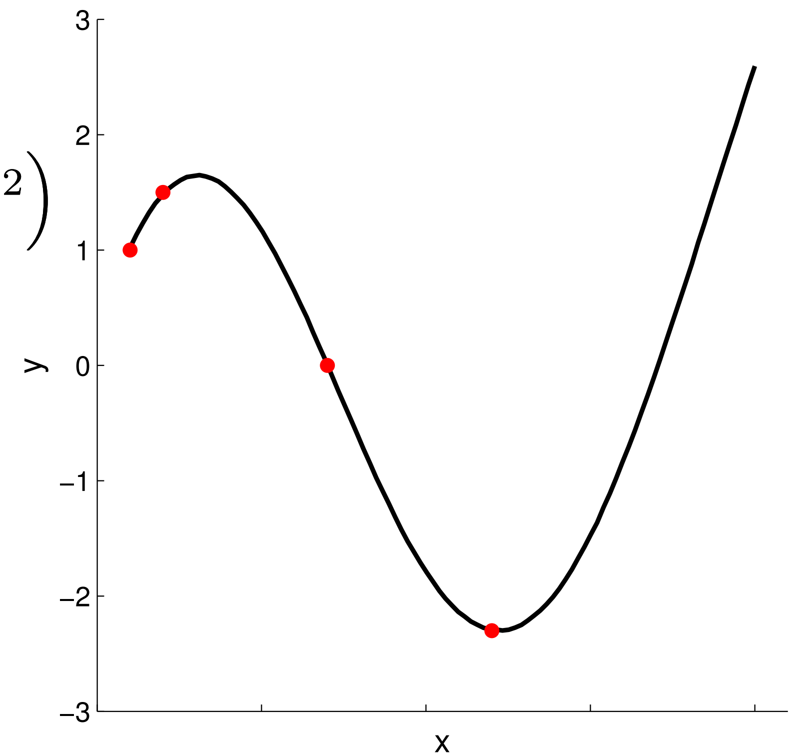
Non-parametric (∞-parametric)

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

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

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

$\Sigma =$ 



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

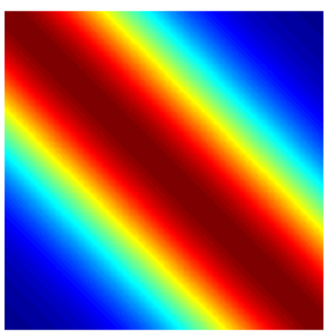
Non-parametric (∞-parametric)

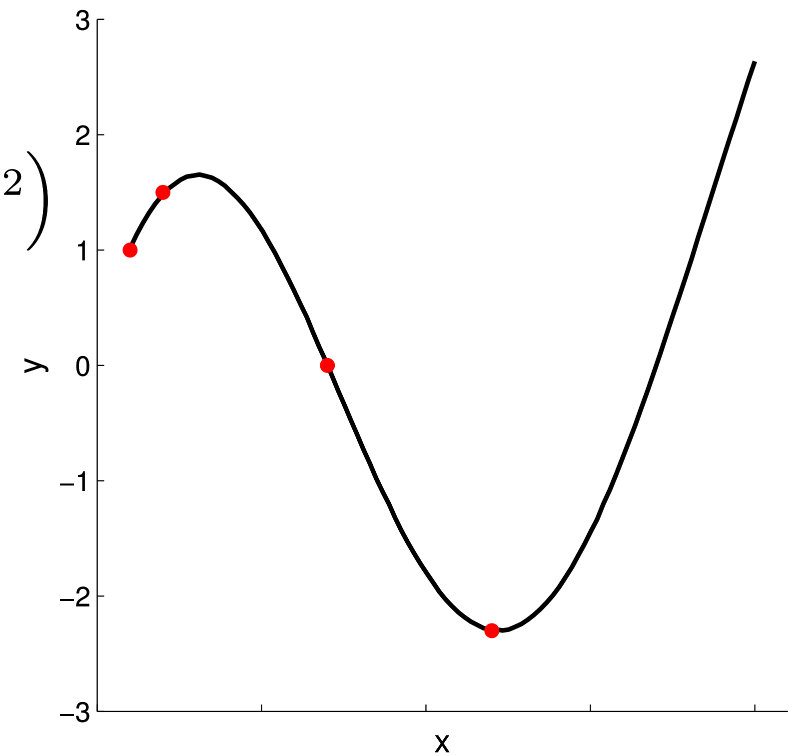
$p(\mathsf{y}|\theta) = \mathcal{N}(0, \Sigma)$

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

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

$\Sigma =$





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

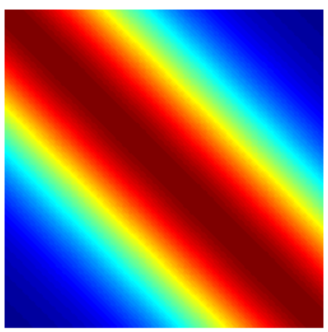
Non-parametric (∞-parametric)

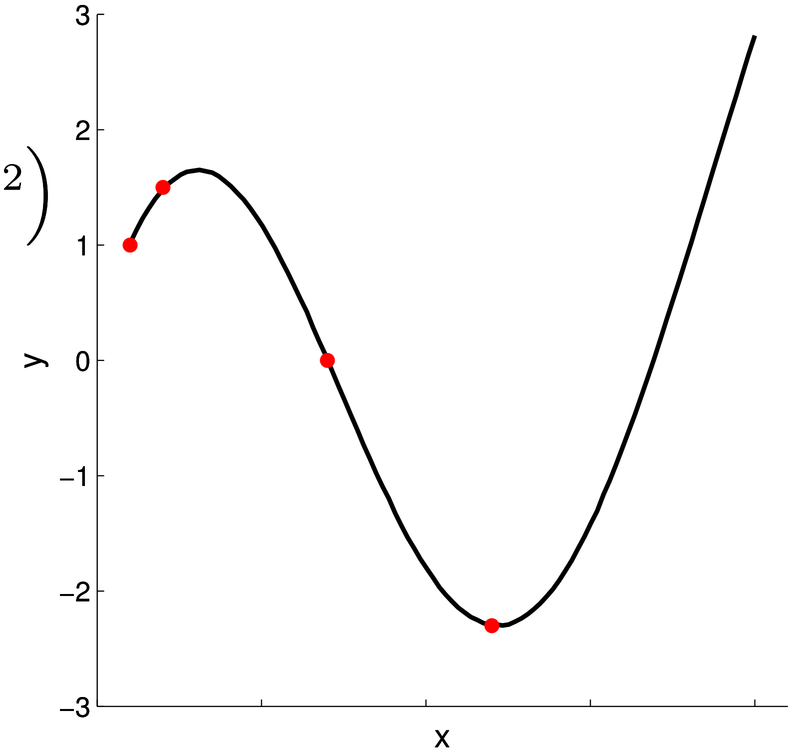
$$p(\mathsf{y}|\theta) = \mathcal{N}(0, \Sigma)$$

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

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

$$\Sigma =$$





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?

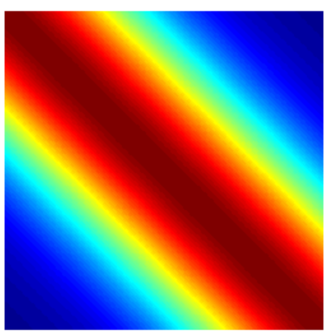
long horizontal length-scale

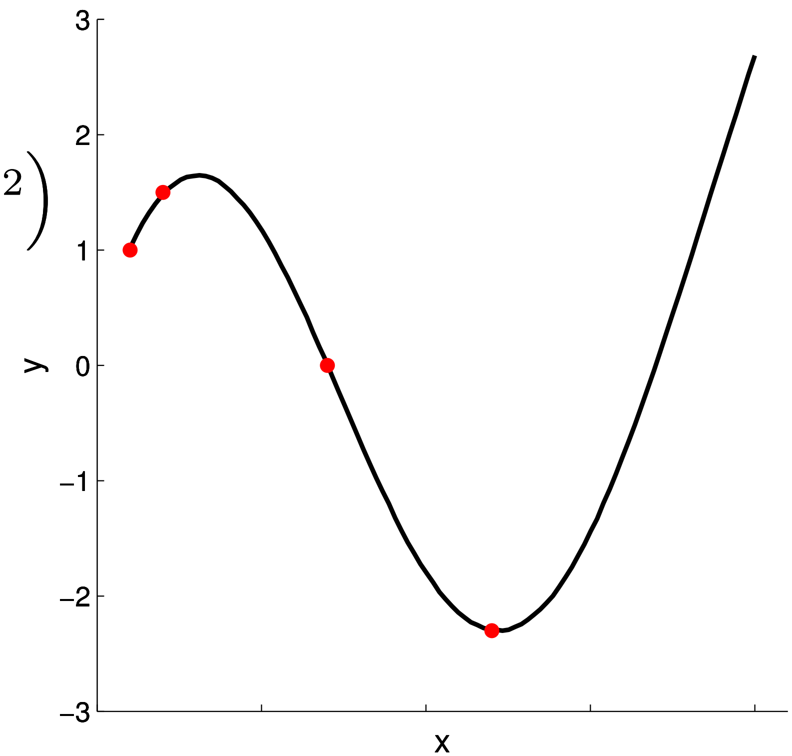
Non-parametric (∞-parametric)

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

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

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

$\Sigma =$ 



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

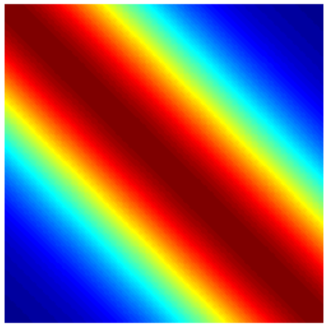
Non-parametric (∞-parametric)

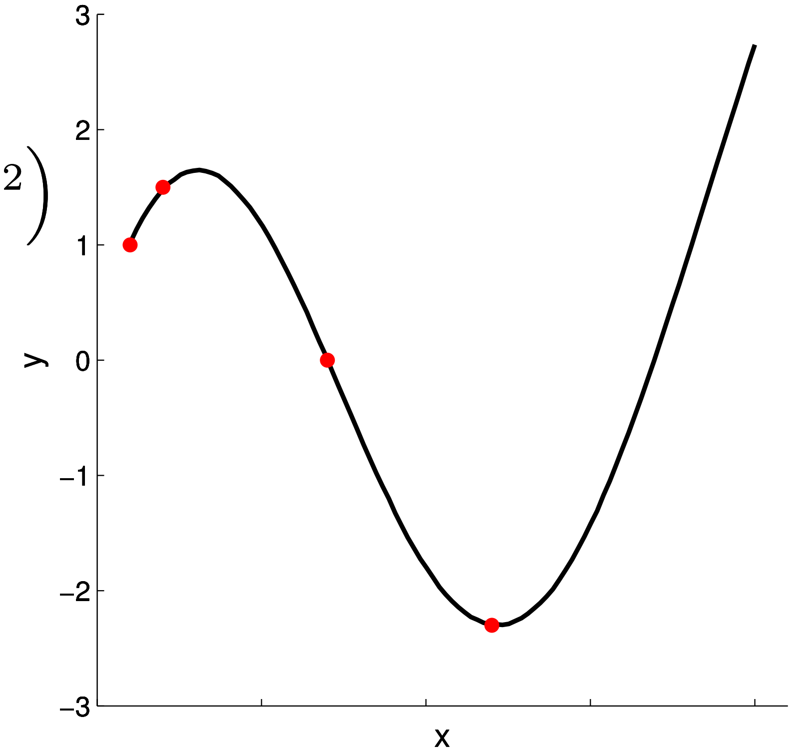
$$p(\mathsf{y}|\theta) = \mathcal{N}(0, \Sigma)$$

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

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

$$\Sigma =$$





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

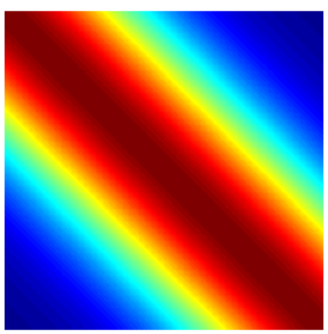
Non-parametric (∞-parametric)

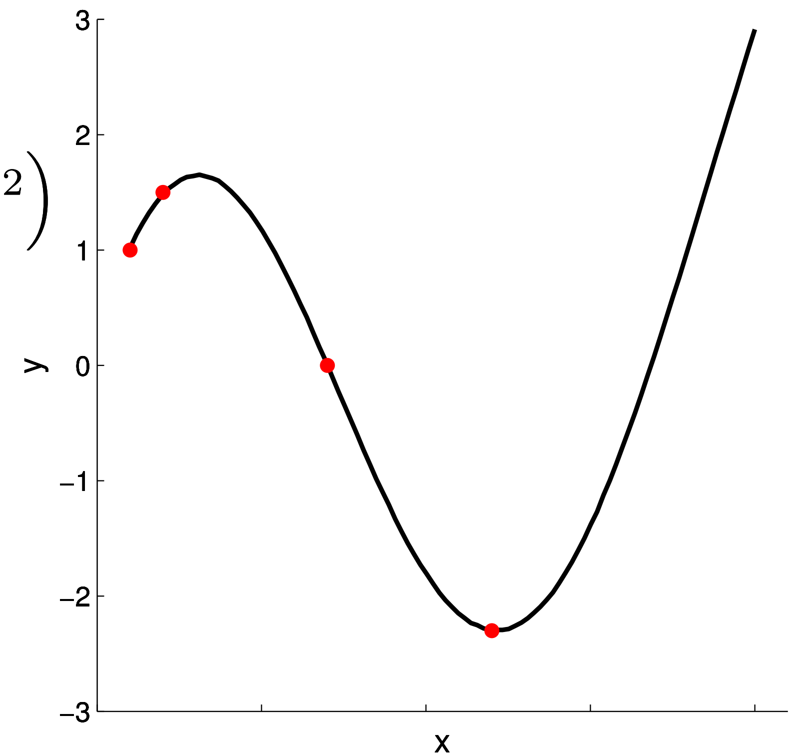
$$p(\mathsf{y}|\theta) = \mathcal{N}(0, \Sigma)$$

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

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

$$\Sigma =$$





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?

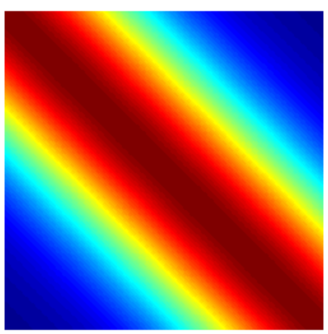
long horizontal length-scale

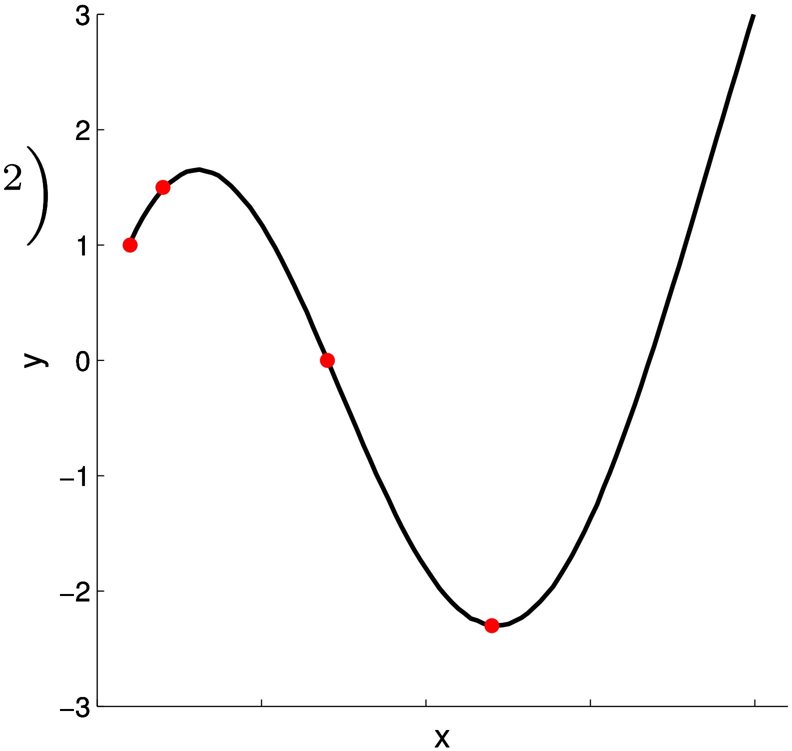
Non-parametric (∞-parametric)

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

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

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

$\Sigma =$ 



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?

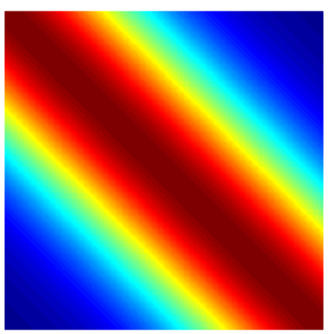
long horizontal length-scale

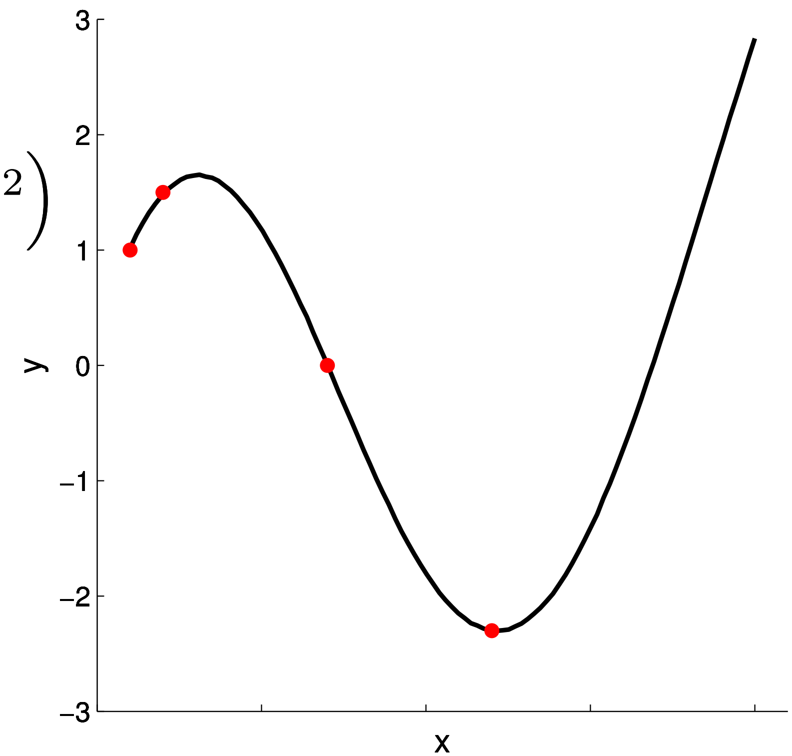
Non-parametric (∞-parametric)

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

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

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

$\Sigma =$ 



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

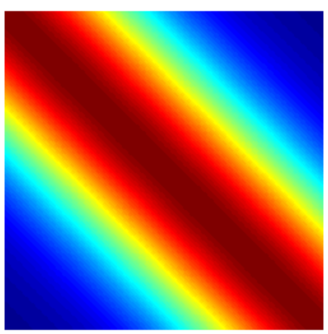
Non-parametric (∞-parametric)

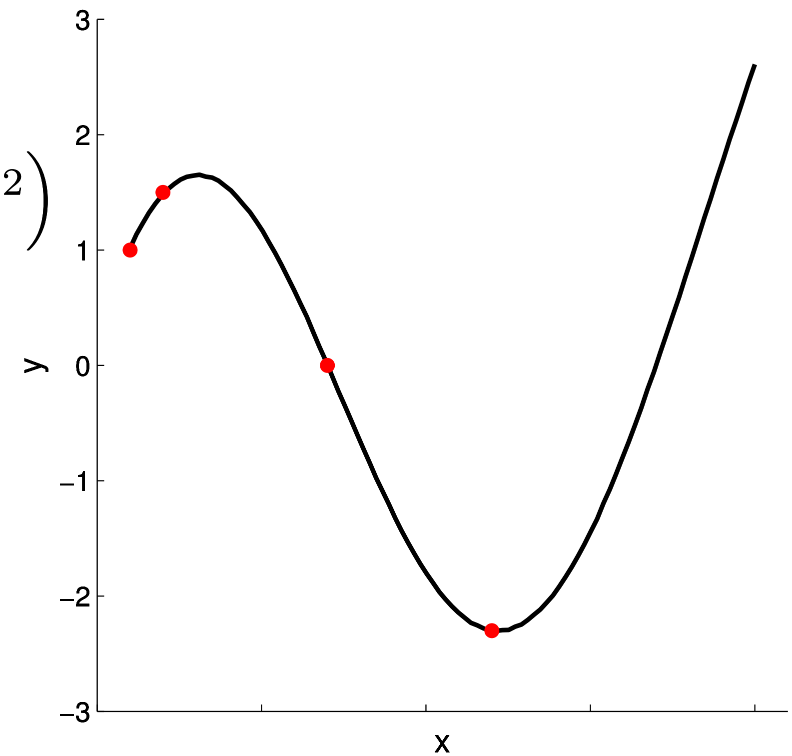
$$p(\mathsf{y}|\theta) = \mathcal{N}(0, \Sigma)$$

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

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

$$\Sigma =$$





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

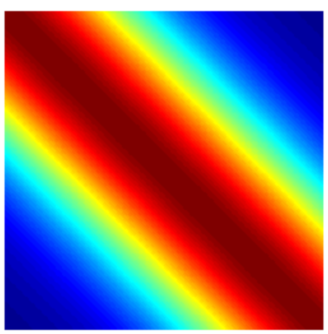
Non-parametric (∞-parametric)

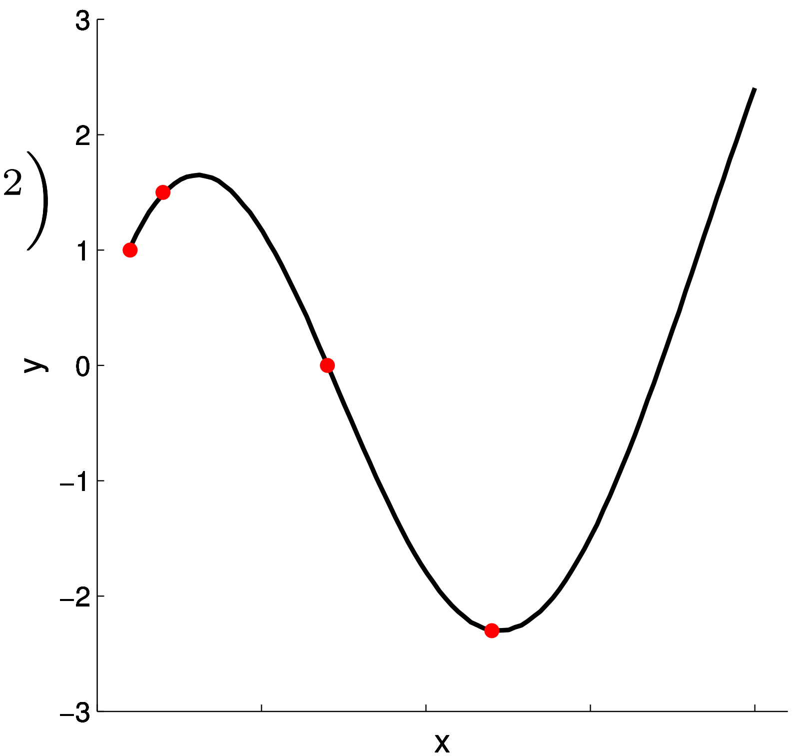
$$p(\mathsf{y}|\theta) = \mathcal{N}(0, \Sigma)$$

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

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

$$\Sigma =$$





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?

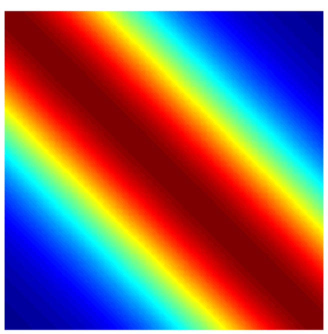
long horizontal length-scale

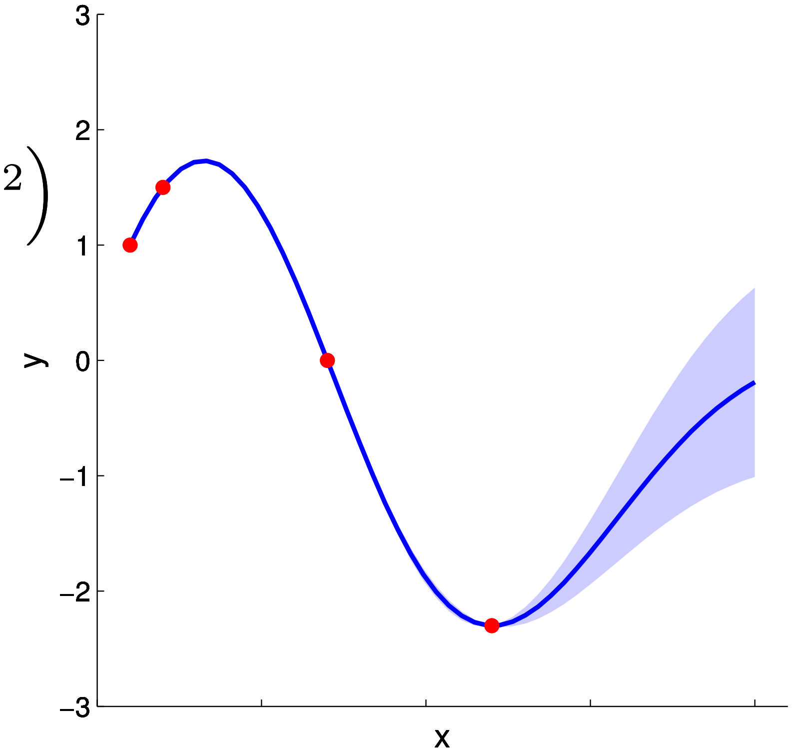
Non-parametric (∞-parametric)

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

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

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

$\Sigma =$ 



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?

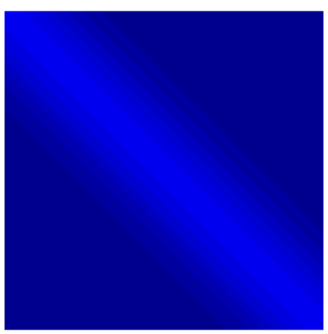
small vertical scale

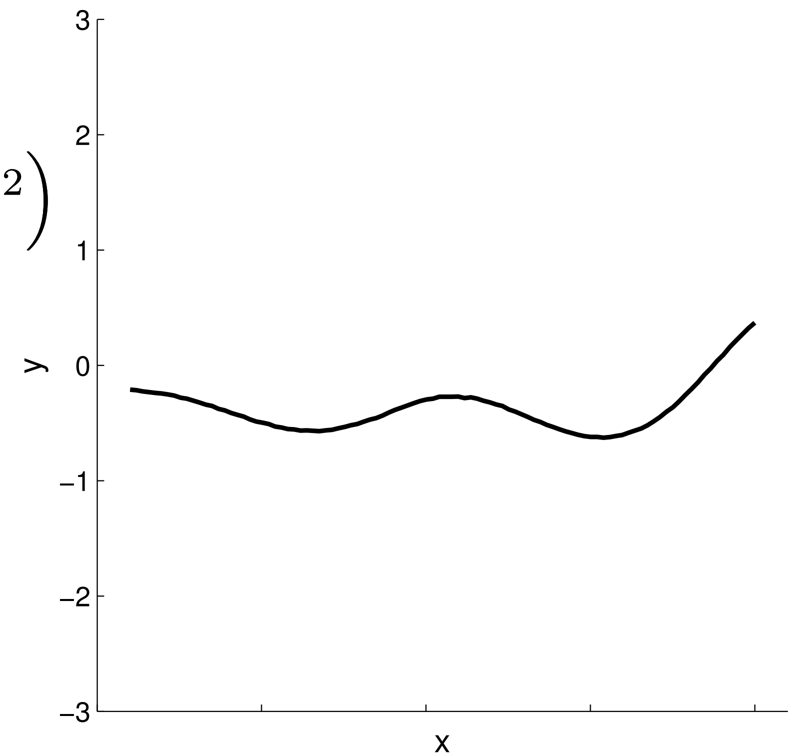
Non-parametric (∞-parametric)

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

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

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

$\Sigma =$ 



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

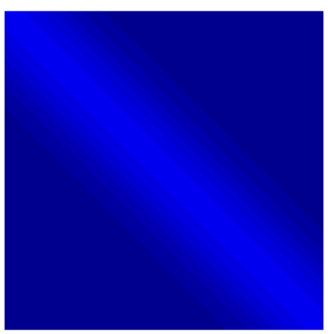
Non-parametric (∞-parametric)

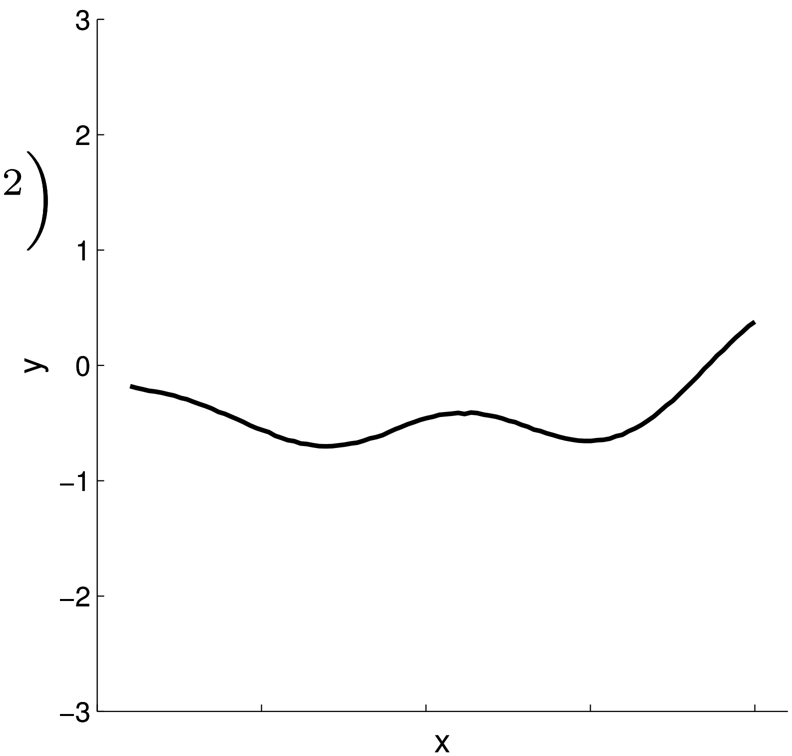
$$p(\mathsf{y}|\theta) = \mathcal{N}(0, \Sigma)$$

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

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

$$\Sigma =$$





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?

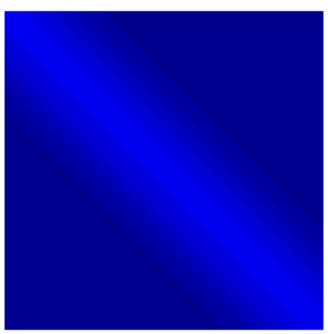
small vertical scale

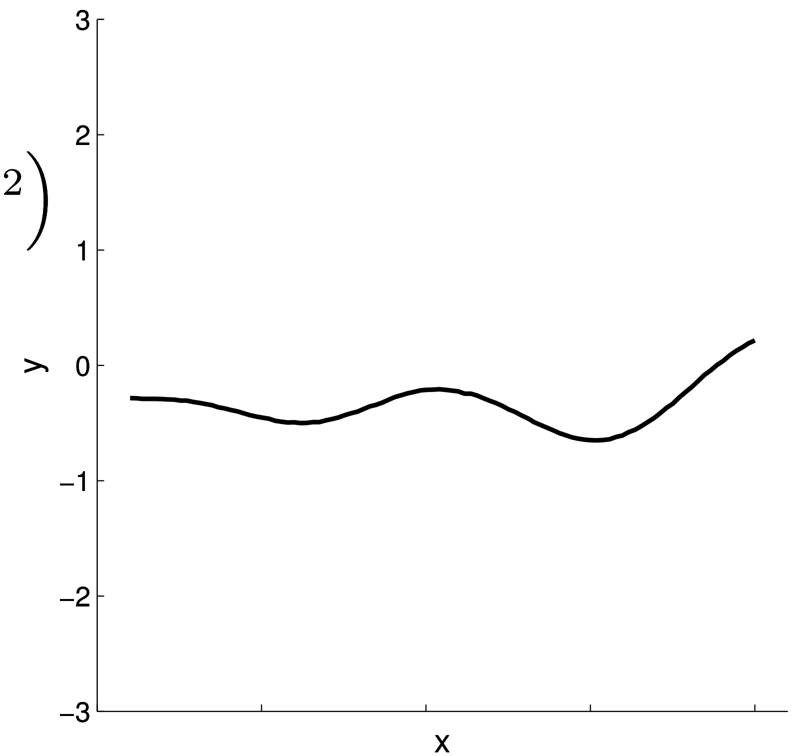
Non-parametric (∞-parametric)

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

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

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

$\Sigma =$ 



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

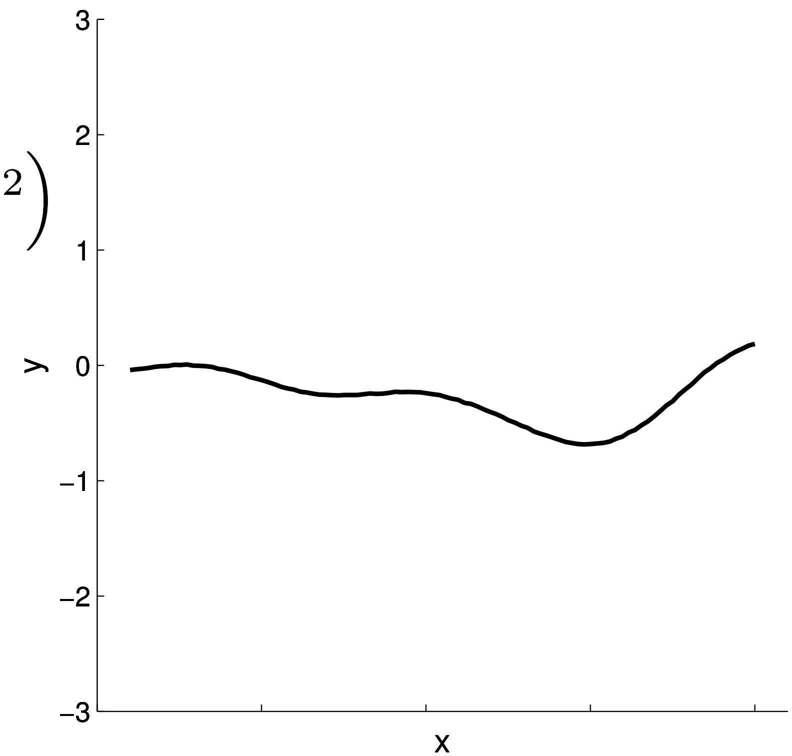
Non-parametric (∞-parametric)

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

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

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

$\Sigma =$ 



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?

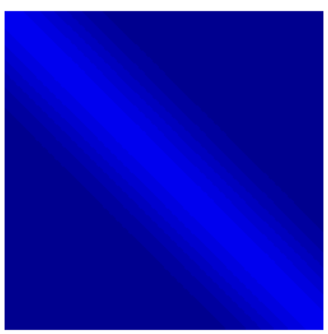
small vertical scale

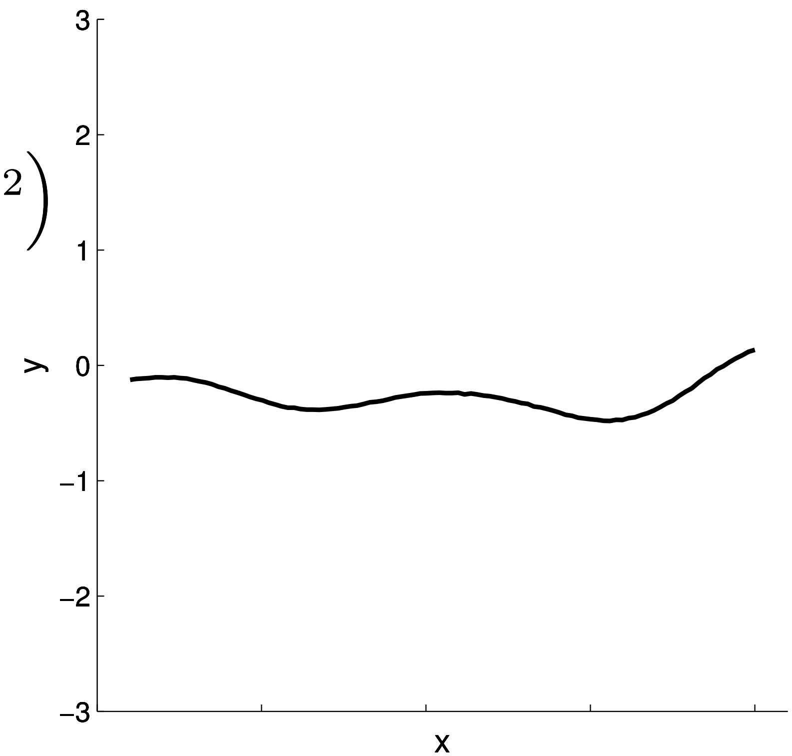
Non-parametric (∞-parametric)

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

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

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

$\Sigma =$ 



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?

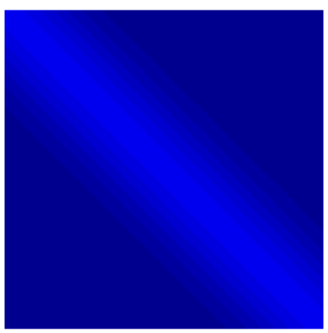
small vertical scale

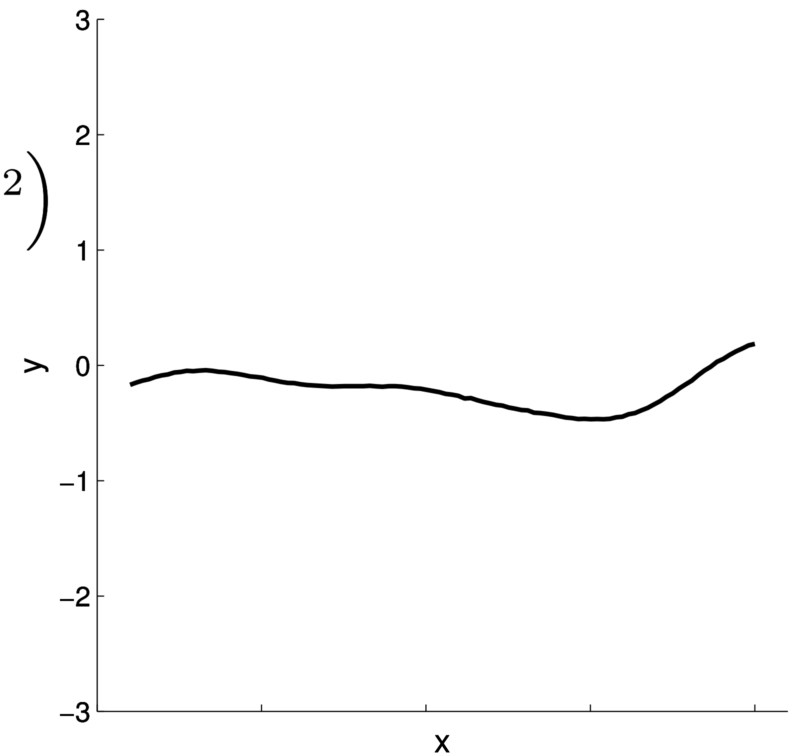
Non-parametric (∞-parametric)

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

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

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

$\Sigma =$ 



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

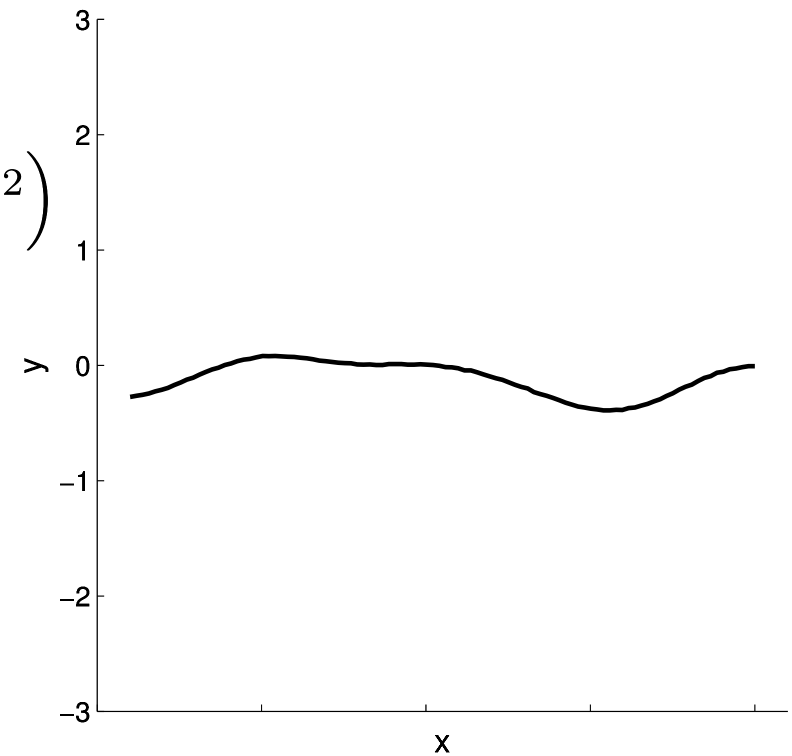
Non-parametric (∞-parametric)

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

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

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

$\Sigma =$ 



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?

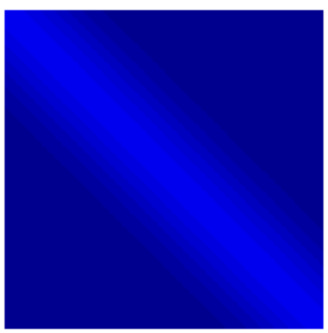
small vertical scale

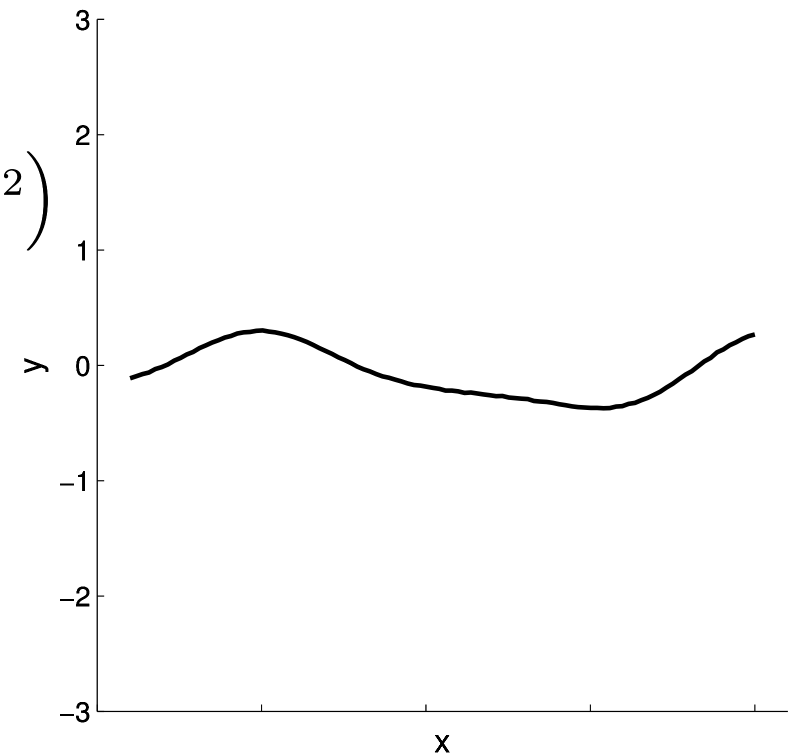
Non-parametric (∞-parametric)

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

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

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

$\Sigma =$ 



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?

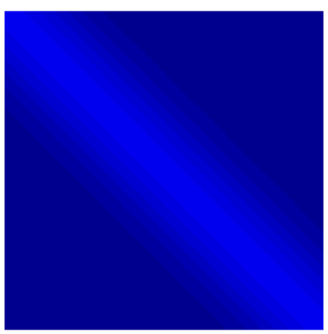
small vertical scale

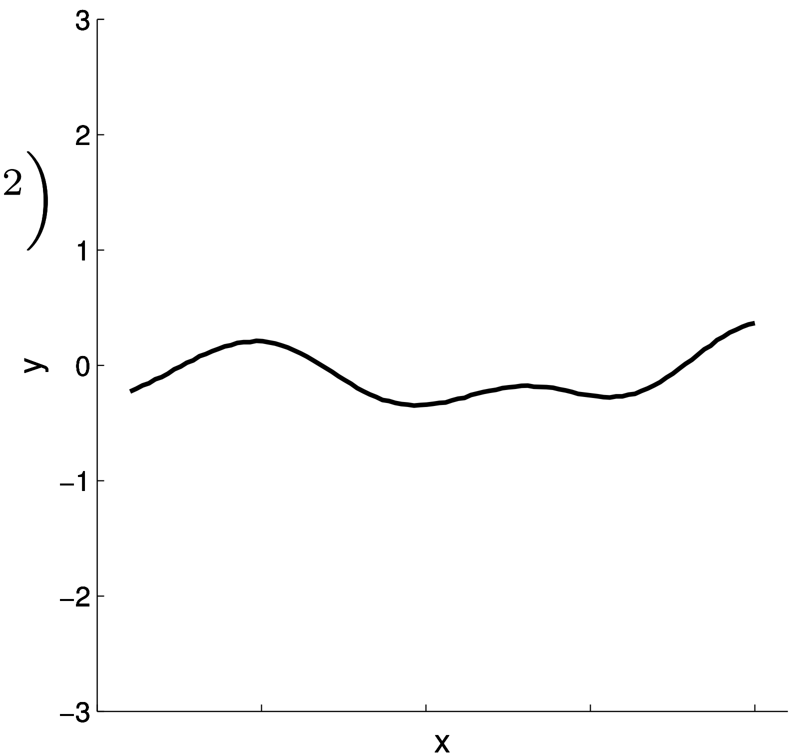
Non-parametric (∞-parametric)

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

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

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

$\Sigma =$ 



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?

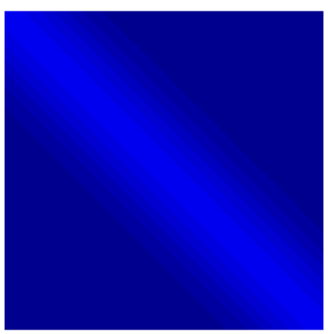
small vertical scale

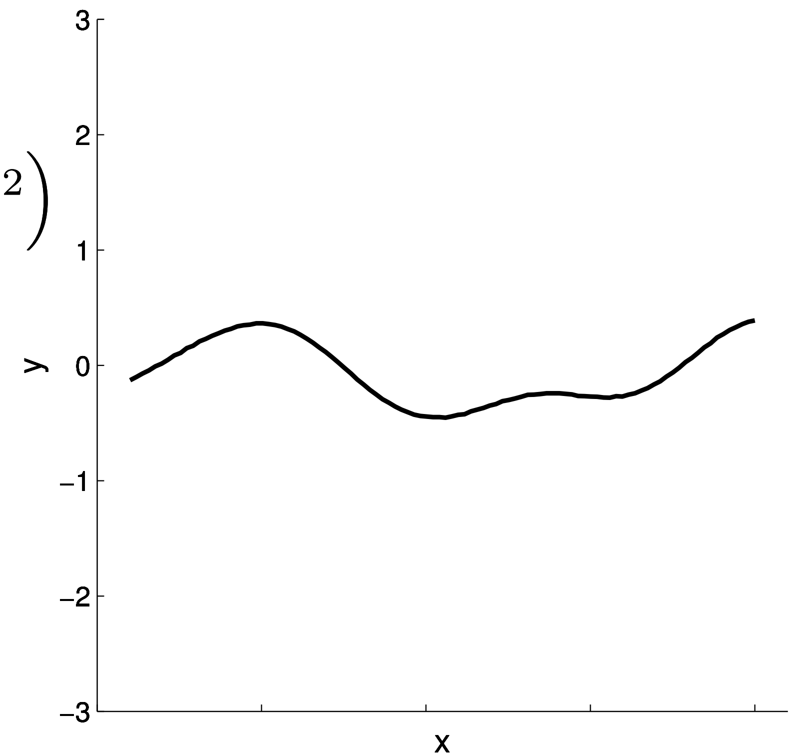
Non-parametric (∞-parametric)

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

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

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

$\Sigma =$ 



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

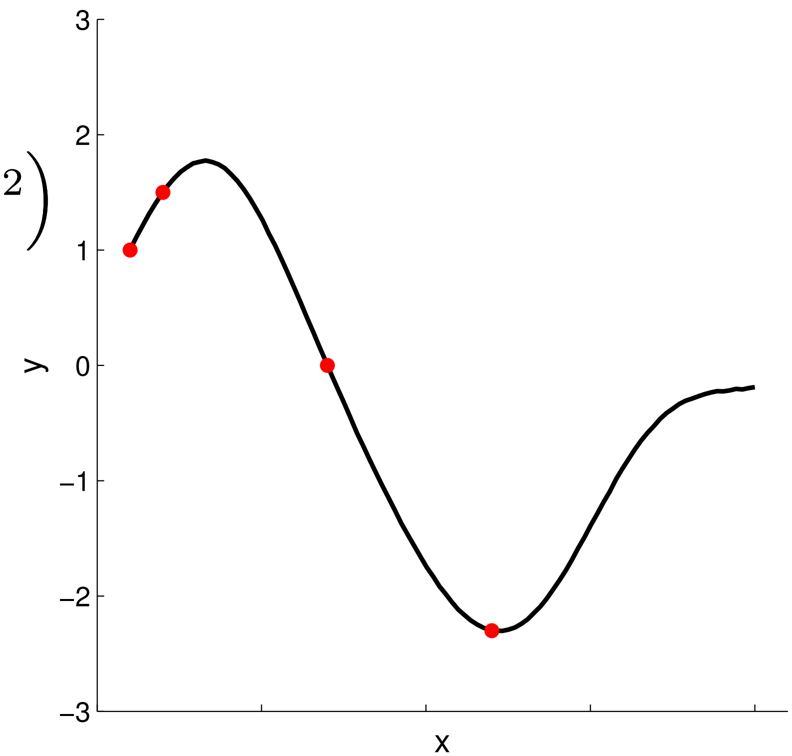
Non-parametric (∞-parametric)

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

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

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

$$\Sigma =$$



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

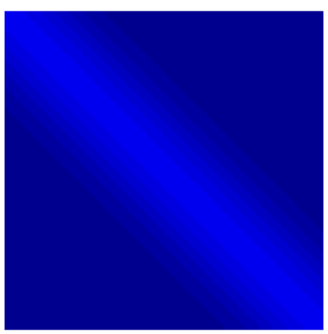
Non-parametric (∞-parametric)

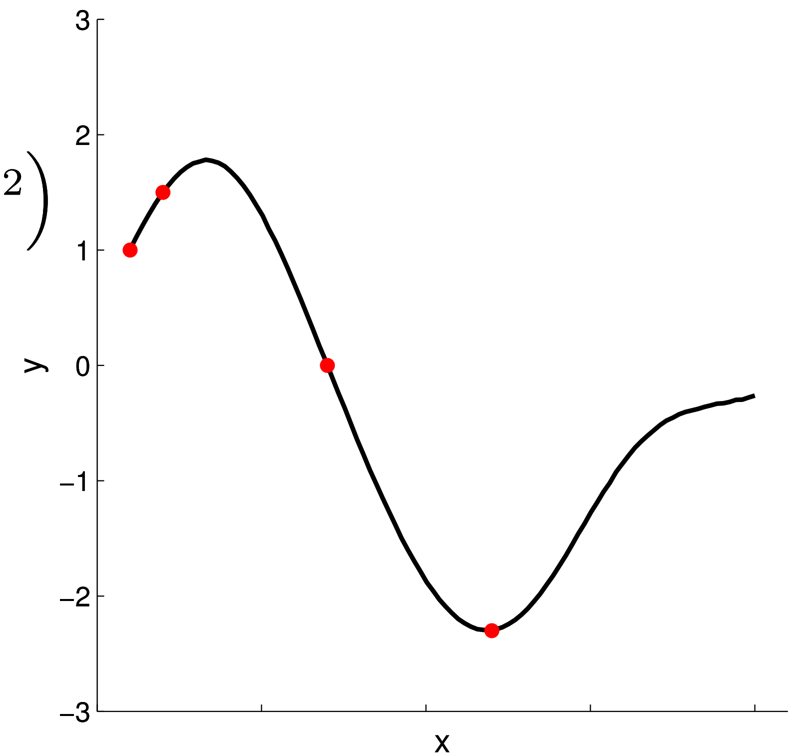
$$p(\mathsf{y}|\theta) = \mathcal{N}(0, \Sigma)$$

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

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

$\Sigma =$





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

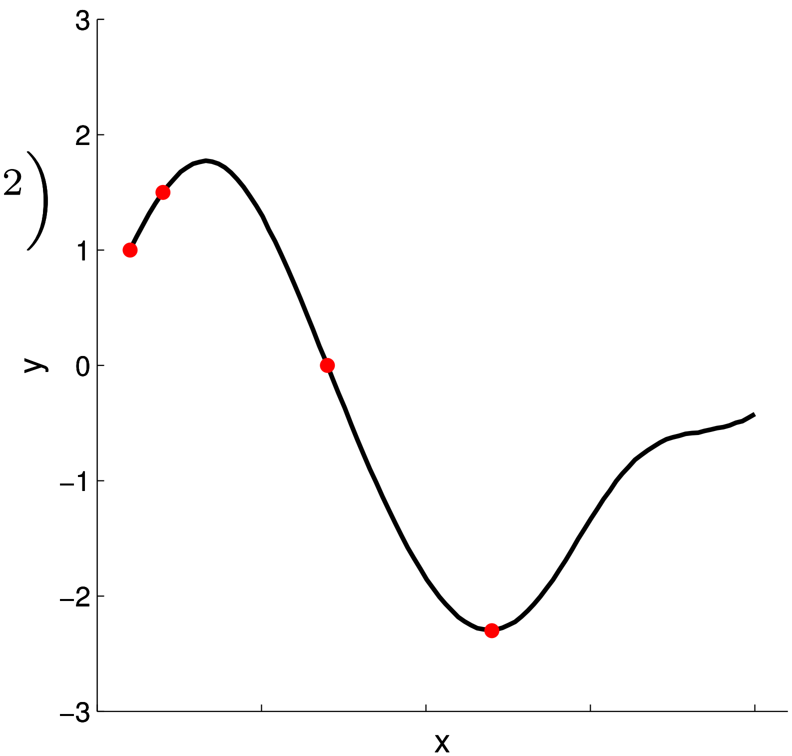
Non-parametric (∞-parametric)

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

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

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

$\Sigma =$ 



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

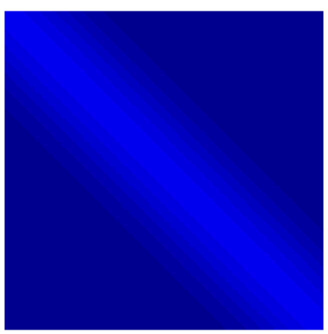
Non-parametric (∞-parametric)

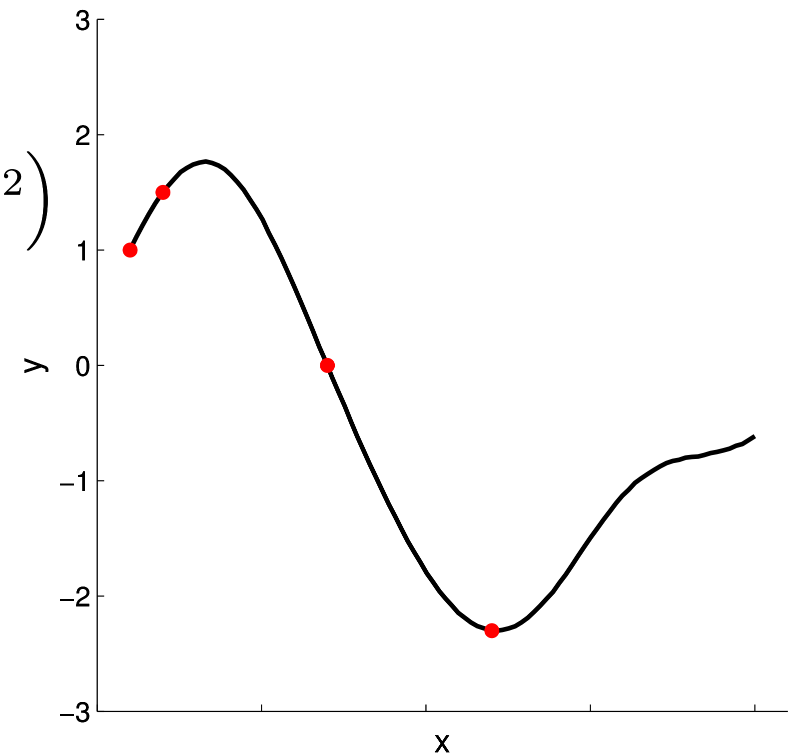
$$p(\mathsf{y}|\theta) = \mathcal{N}(0, \Sigma)$$

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

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

$$\Sigma =$$





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?

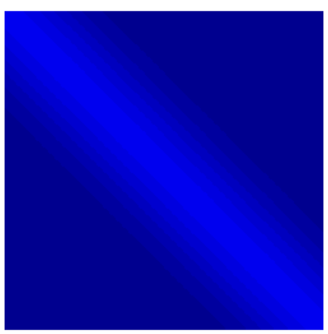
small vertical scale

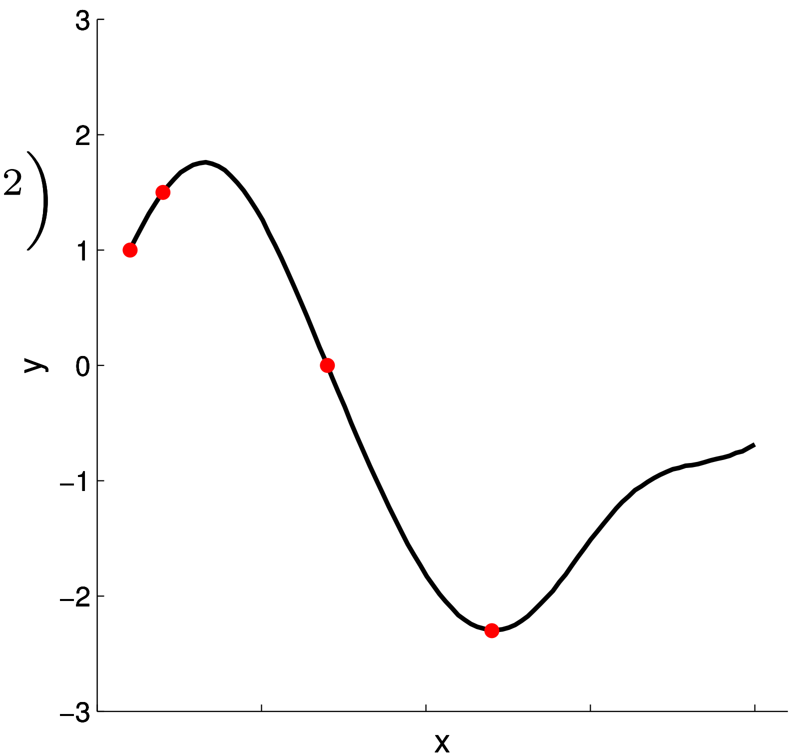
Non-parametric (∞-parametric)

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

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

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

$\Sigma =$ 



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?

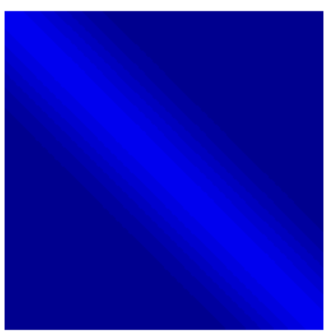
small vertical scale

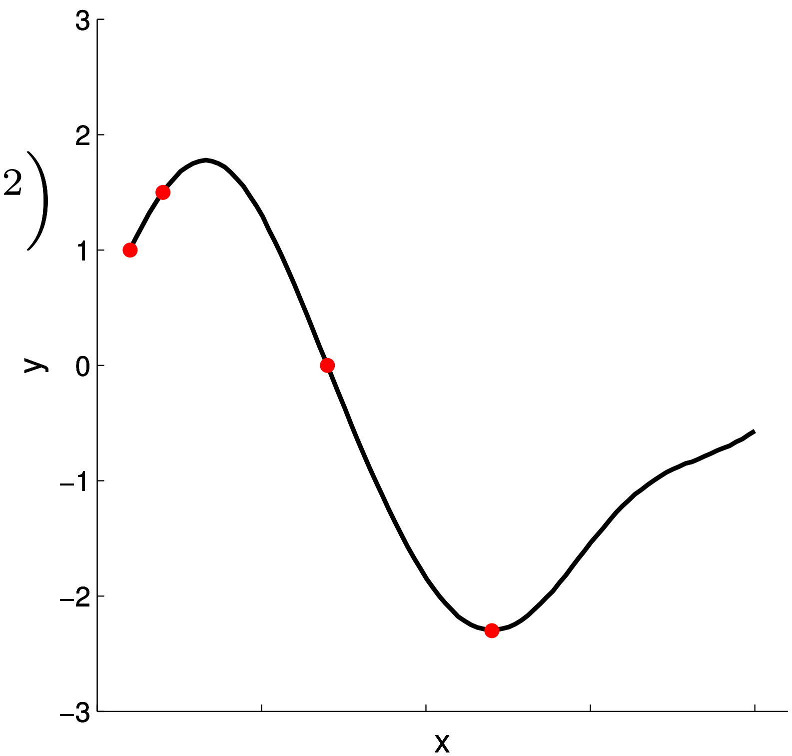
Non-parametric (∞-parametric)

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

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

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

$\Sigma =$ 



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

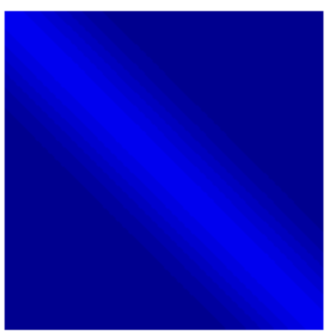
Non-parametric (∞-parametric)

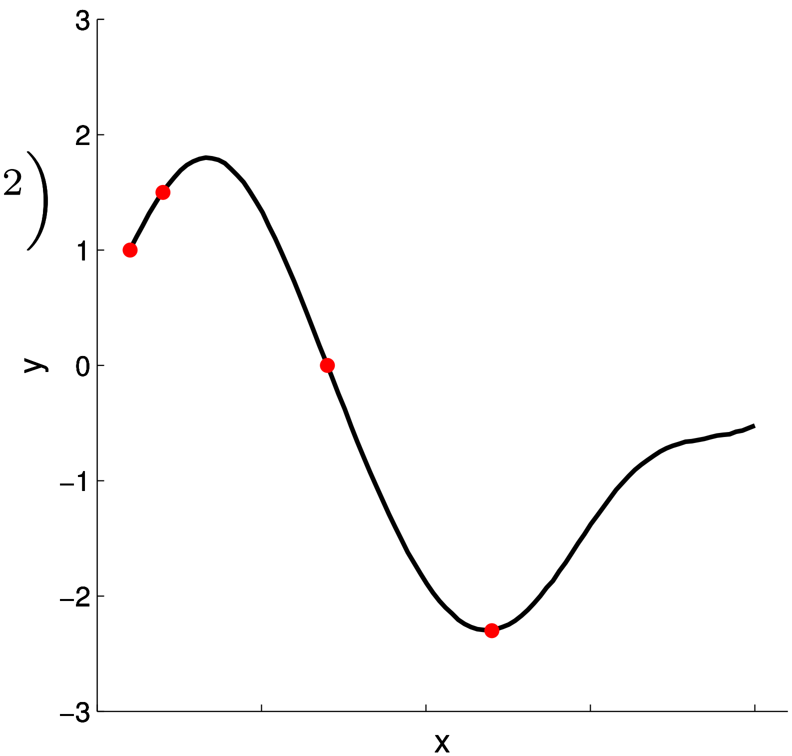
$p(\mathsf{y}|\theta) = \mathcal{N}(0, \Sigma)$

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

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

$\Sigma =$





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

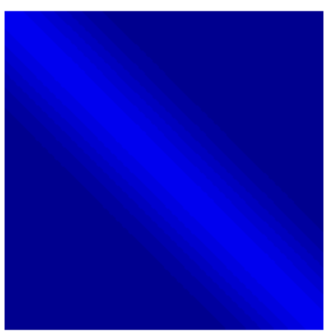
Non-parametric (∞-parametric)

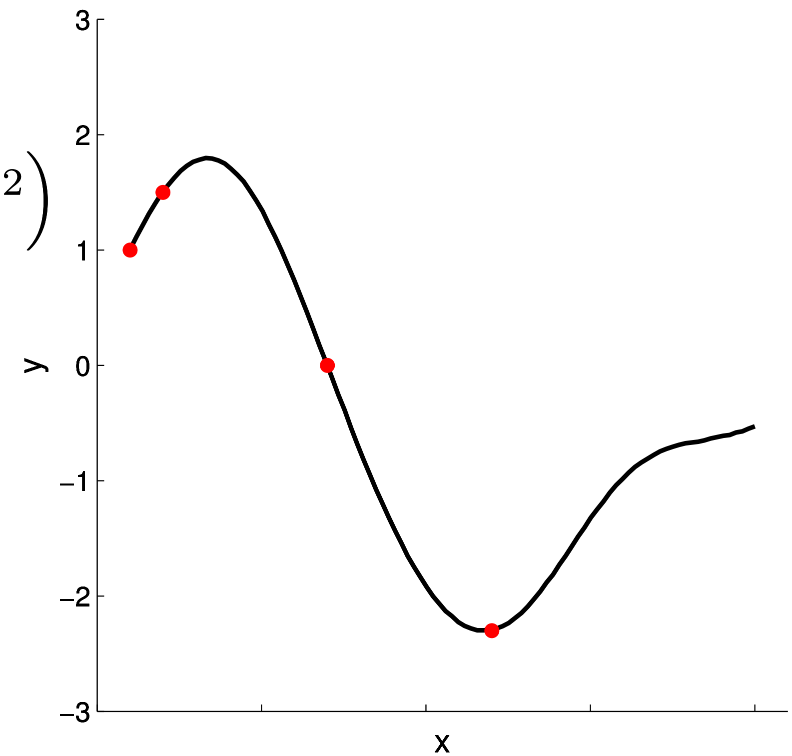
$$p(\mathsf{y}|\theta) = \mathcal{N}(0, \Sigma)$$

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

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

$$\Sigma =$$





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?

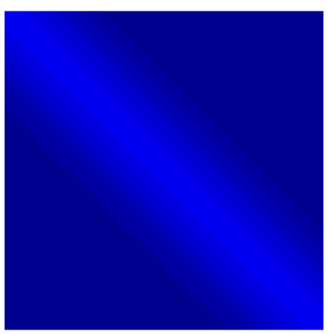
small vertical scale

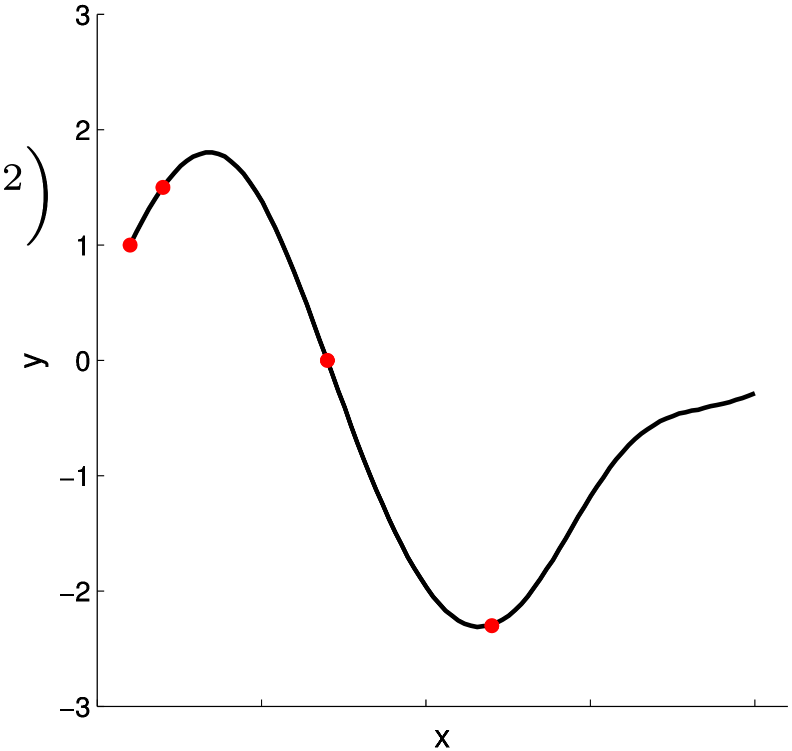
Non-parametric (∞-parametric)

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

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

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

$\Sigma =$ 



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?

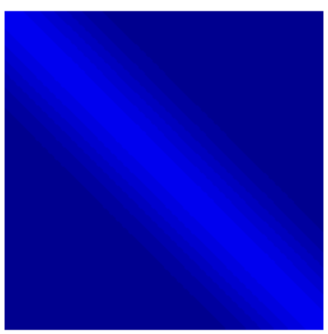
small vertical scale

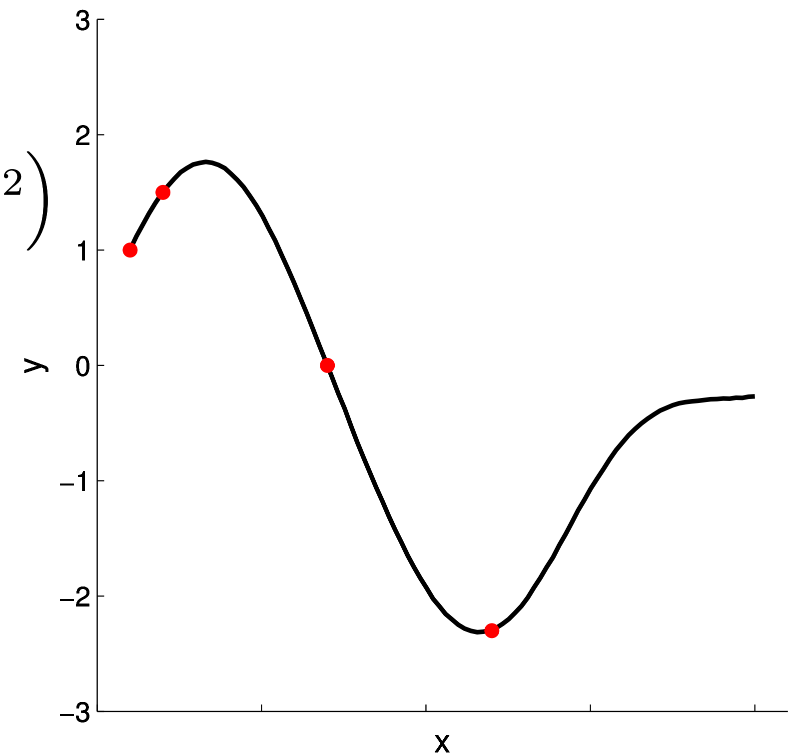
Non-parametric (∞-parametric)

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

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

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

$\Sigma =$ 



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?

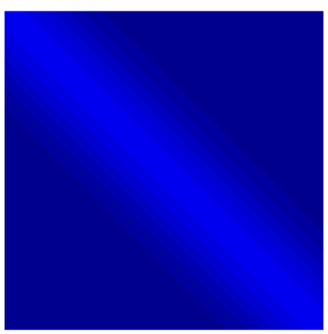
small vertical scale

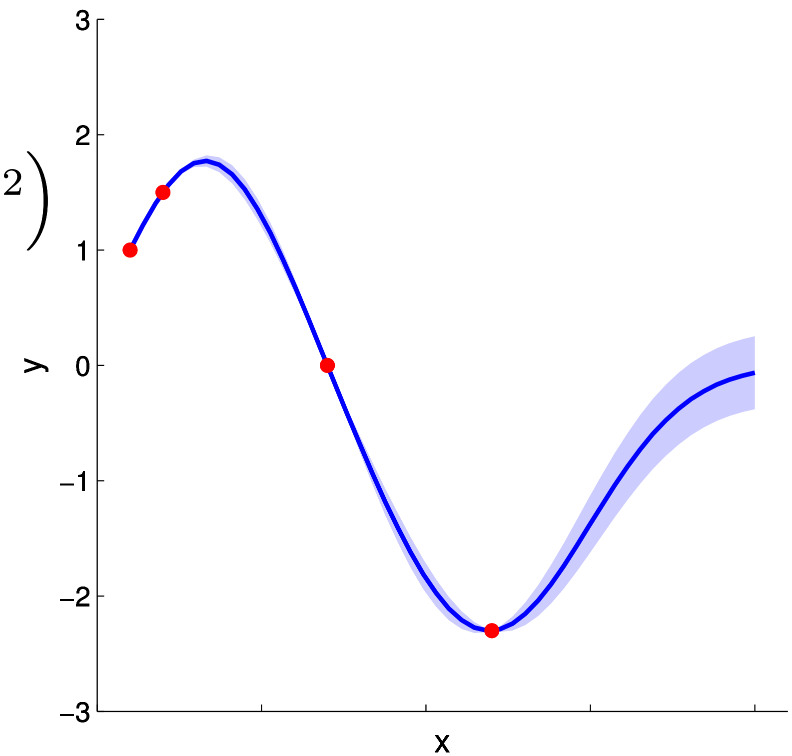
Non-parametric (∞-parametric)

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

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

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

$\Sigma =$ 



Parametric model

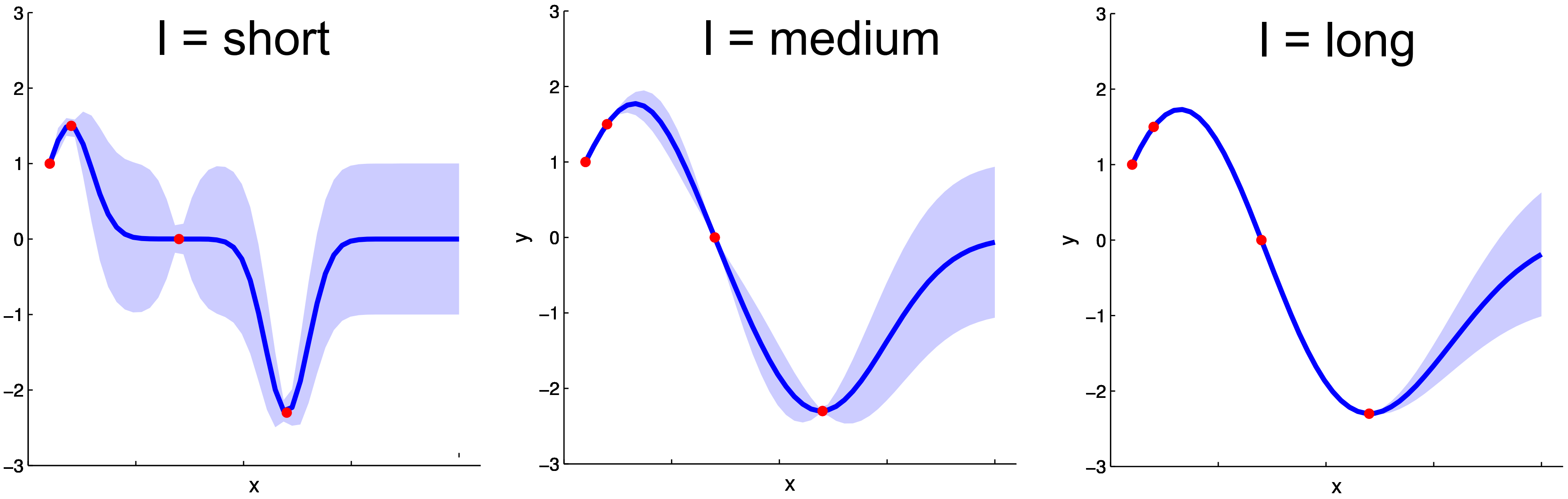
$$\mathsf{y}(\mathsf{x}) = f(\mathsf{x}; \theta) + \sigma_{\mathsf{y}}\epsilon$$

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

What effect do the hyper-parameters have?

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

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



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)

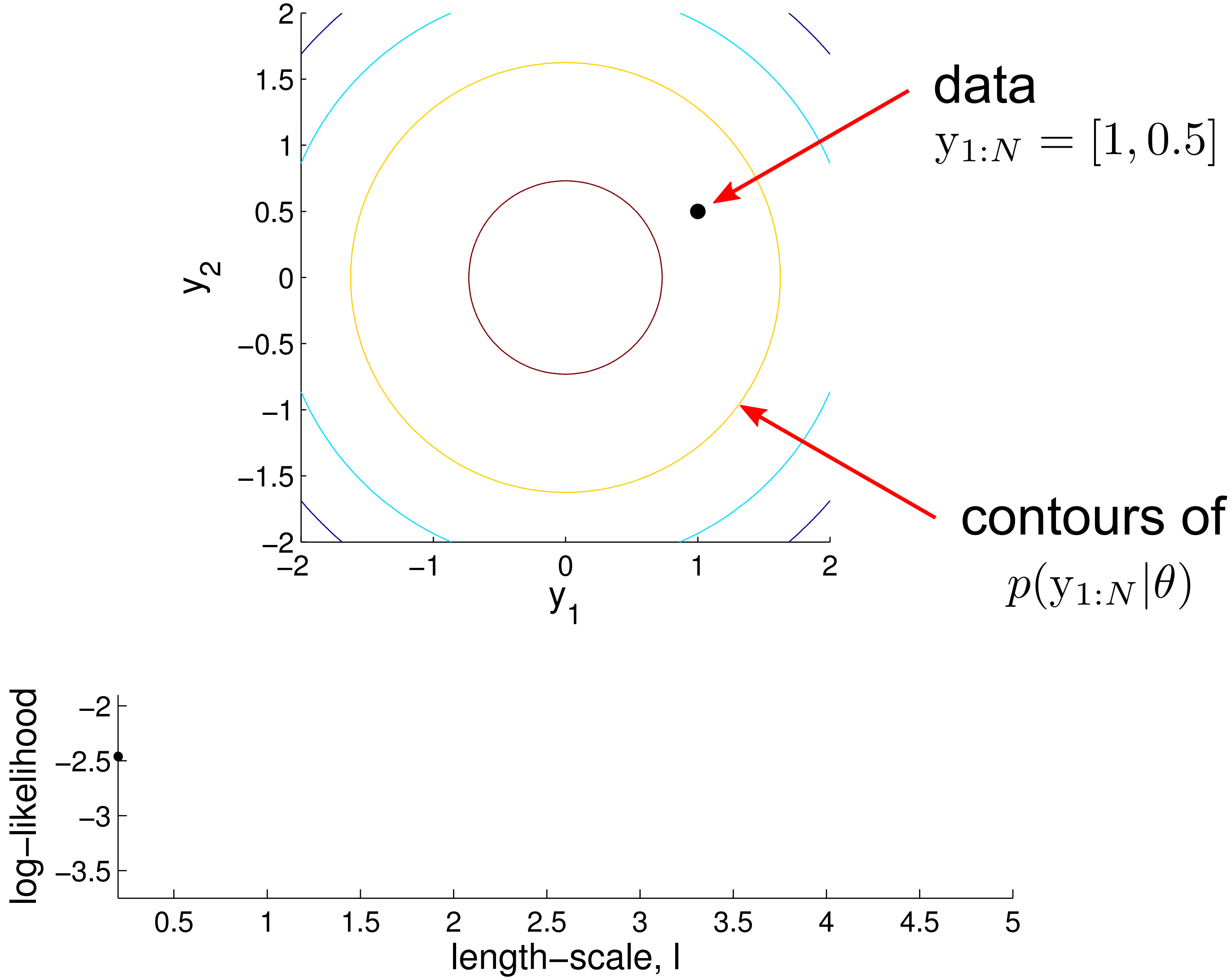
$$p(\theta|\mathbf{y}_{1:N}) = \frac{p(\mathbf{y}_{1:N}|\theta)p(\theta)}{p(\mathbf{y}_{1:N})}$$ (Bayes' Rule)

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

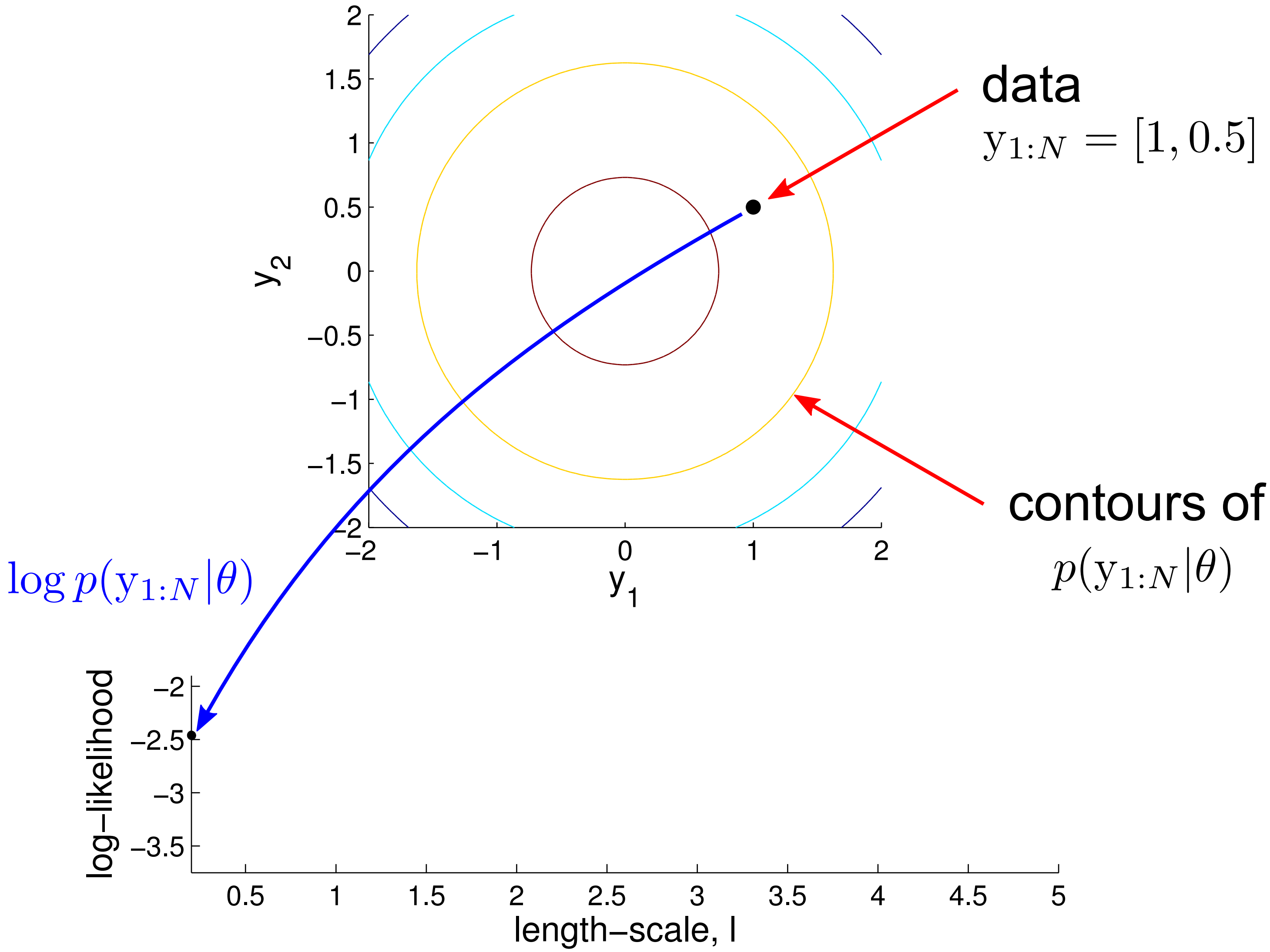
= how well did θ predict the data we observed

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

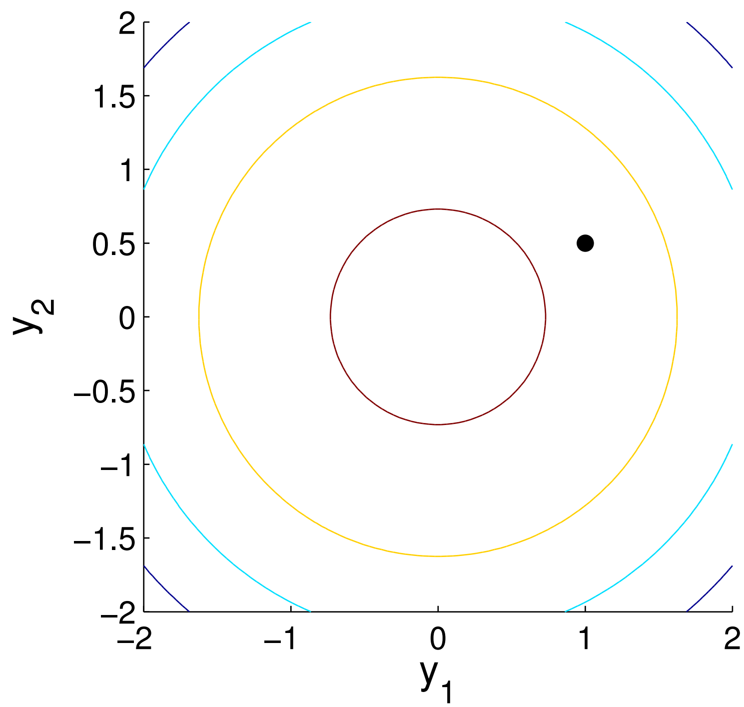
How do we choose the hyper-parameters?

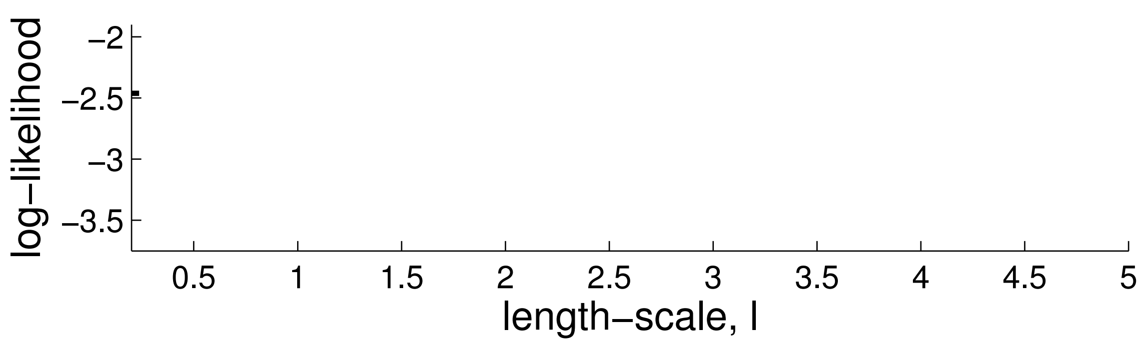


How do we choose the hyper-parameters?

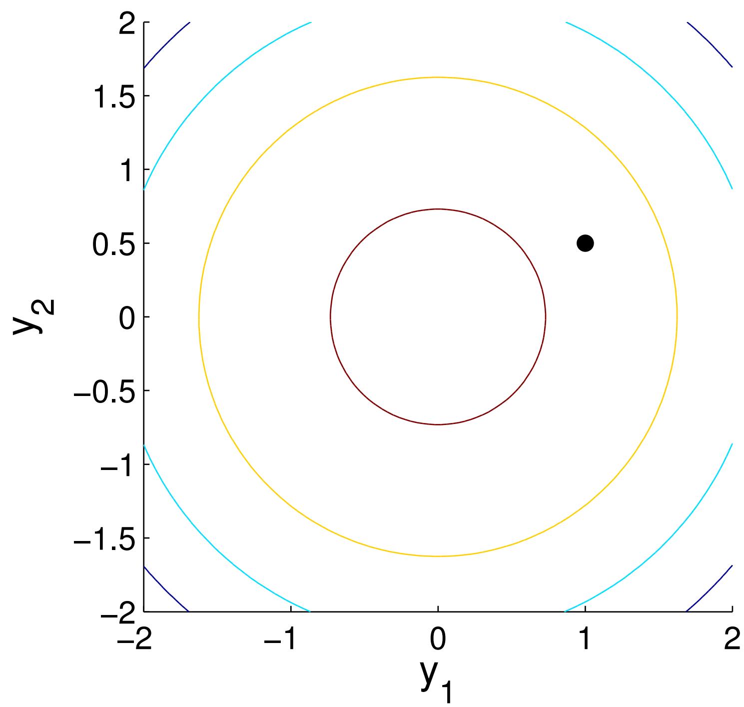


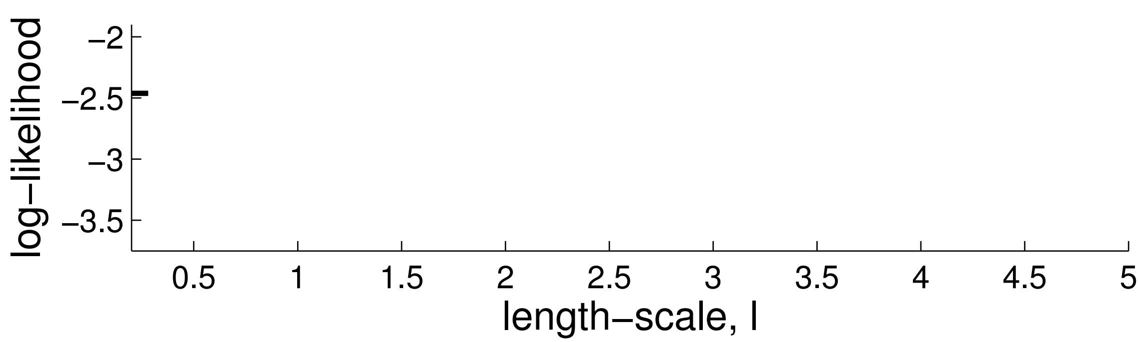
How do we choose the hyper-parameters?



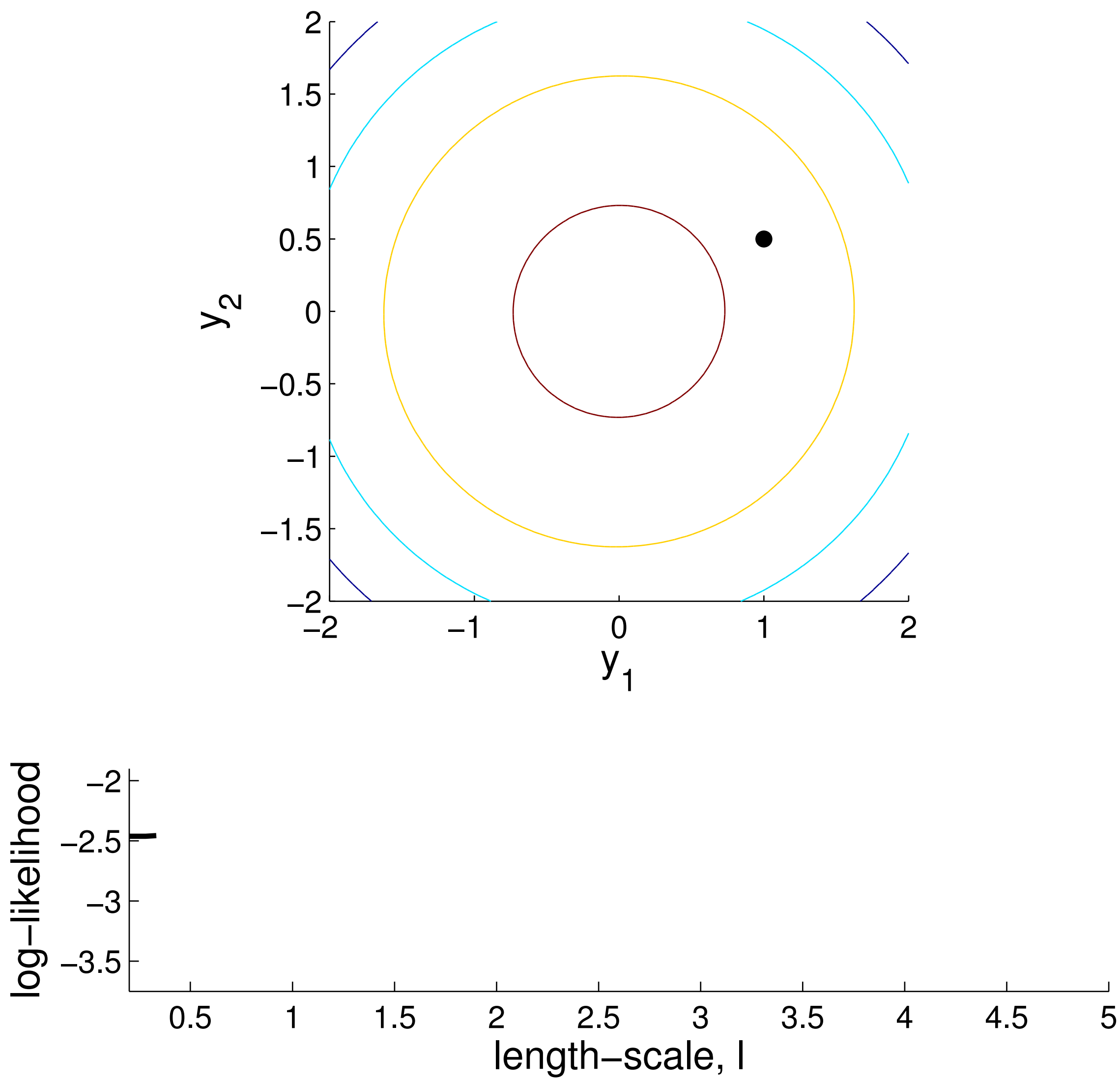


How do we choose the hyper-parameters?

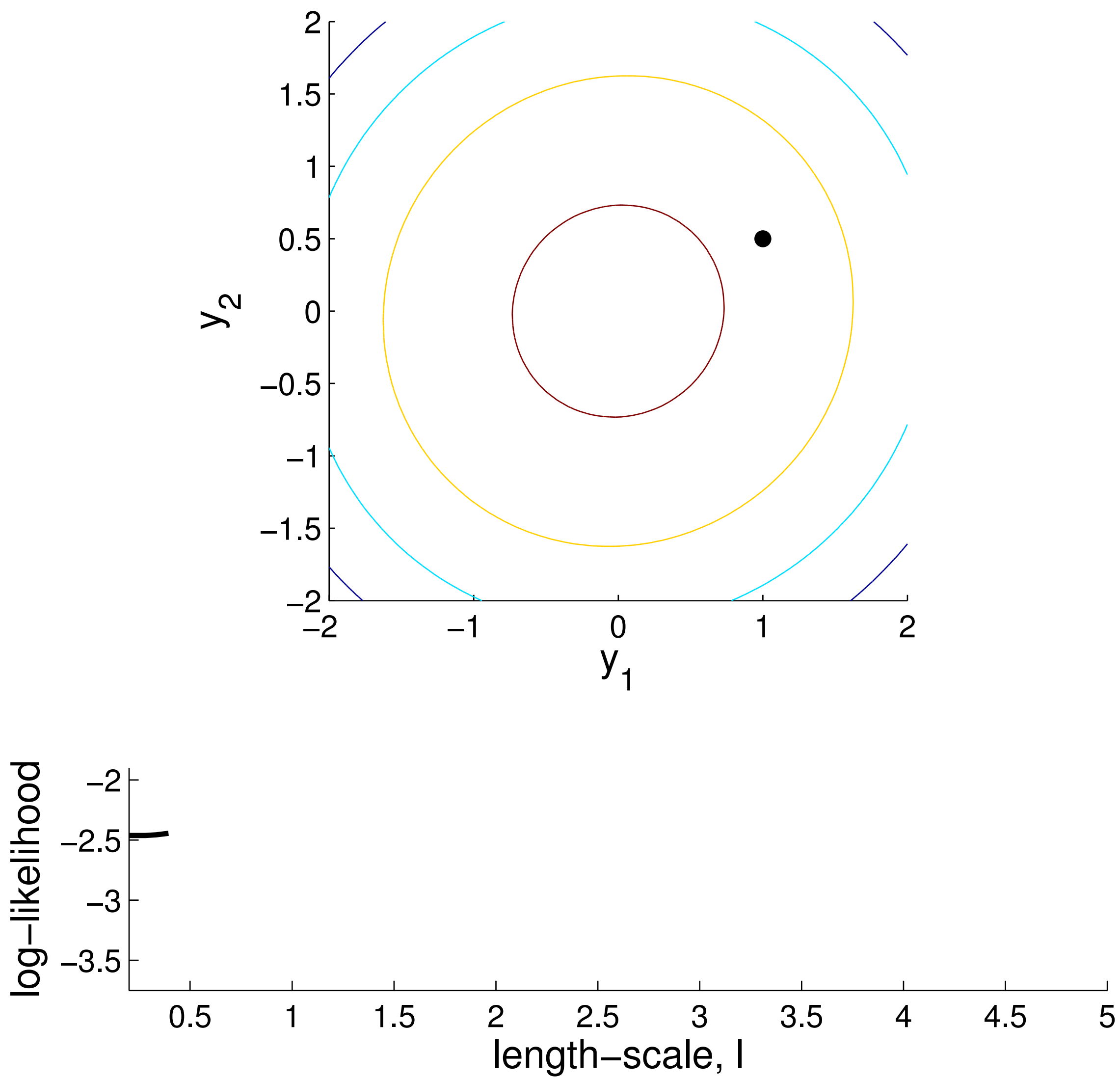




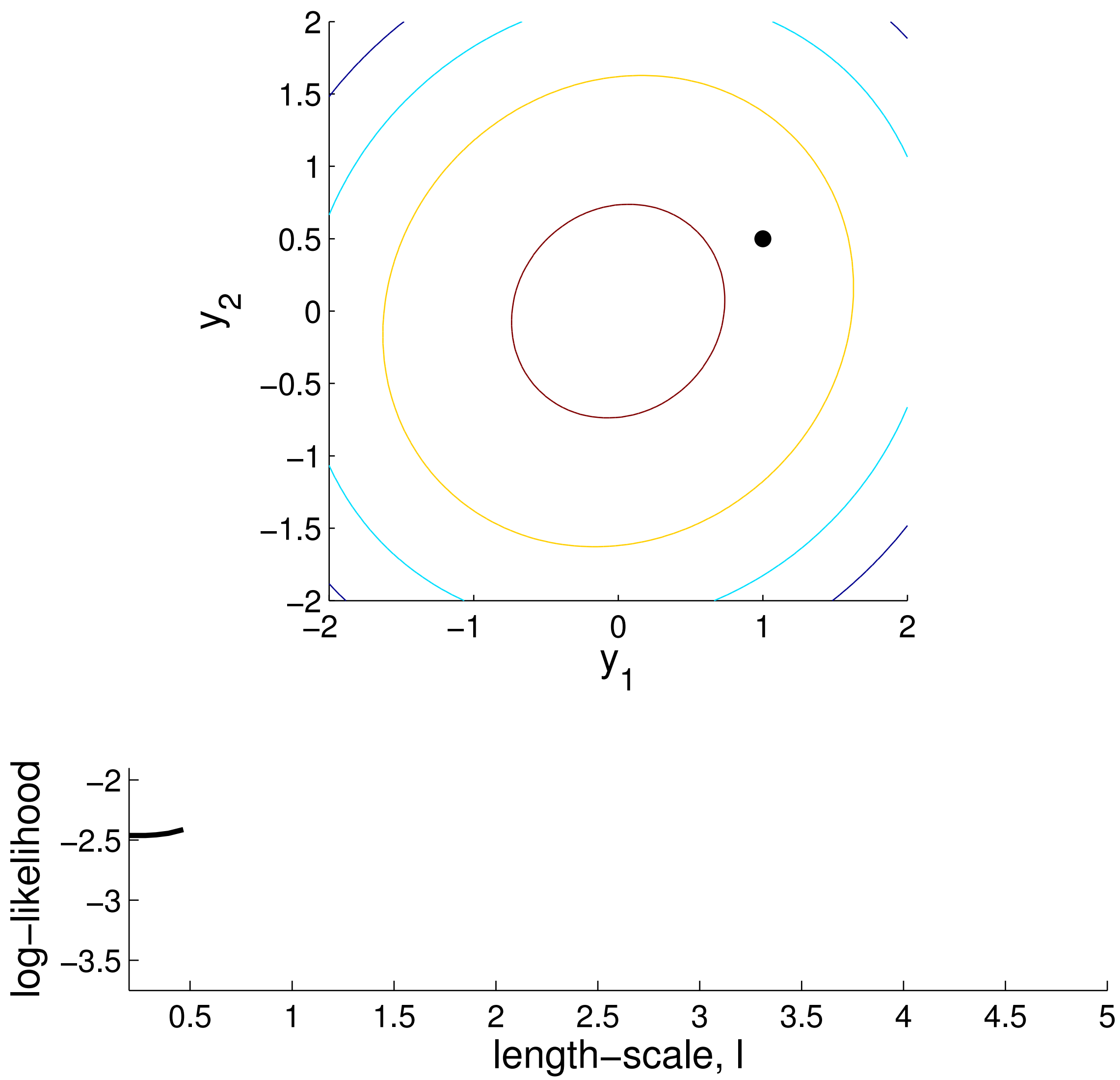
How do we choose the hyper-parameters?



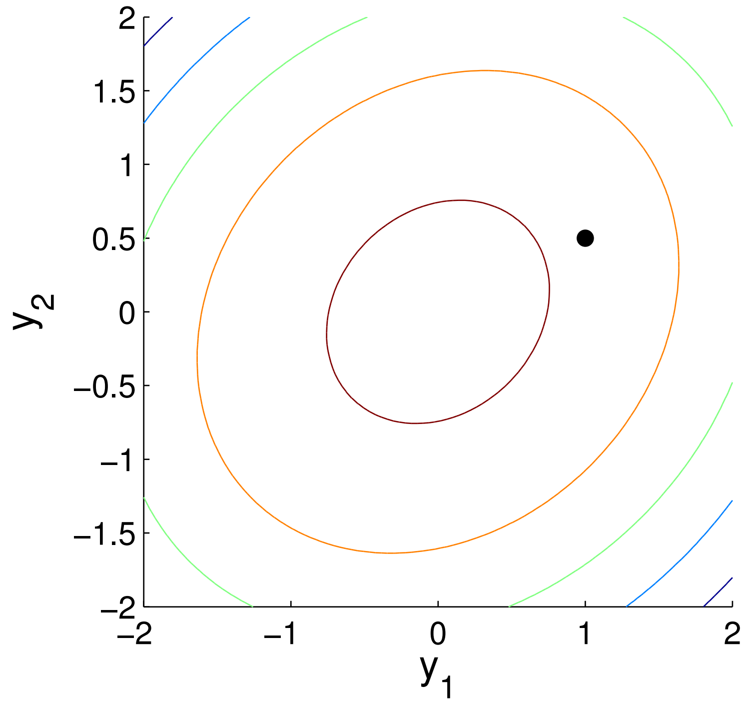
How do we choose the hyper-parameters?

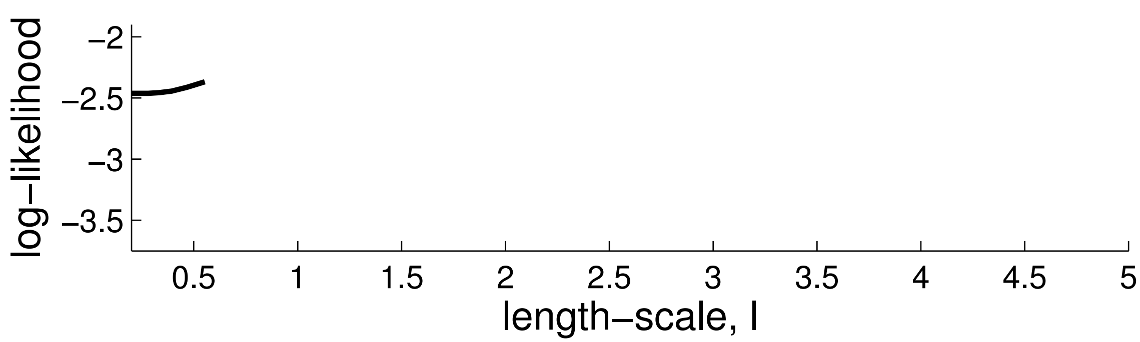


How do we choose the hyper-parameters?

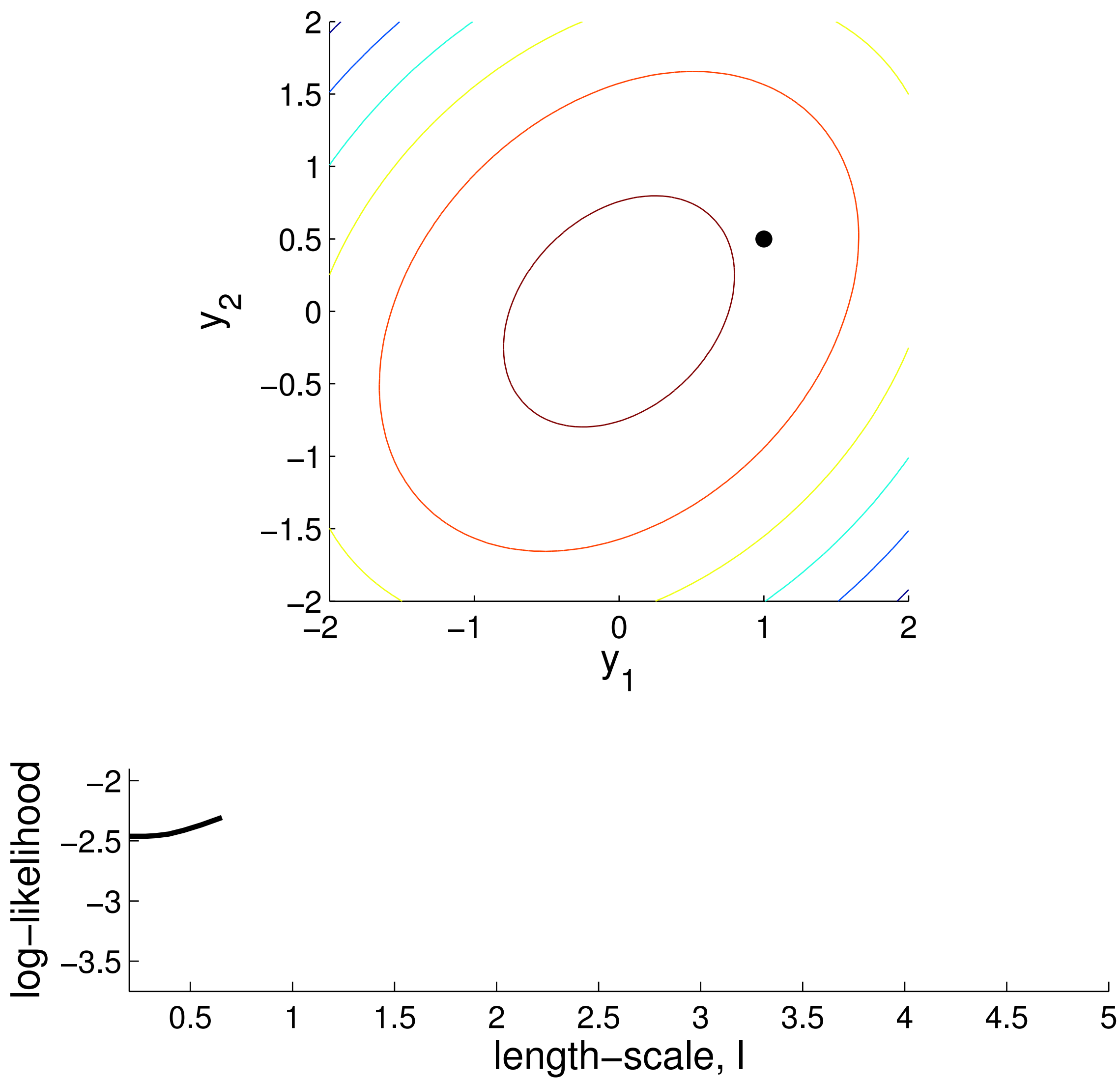


How do we choose the hyper-parameters?

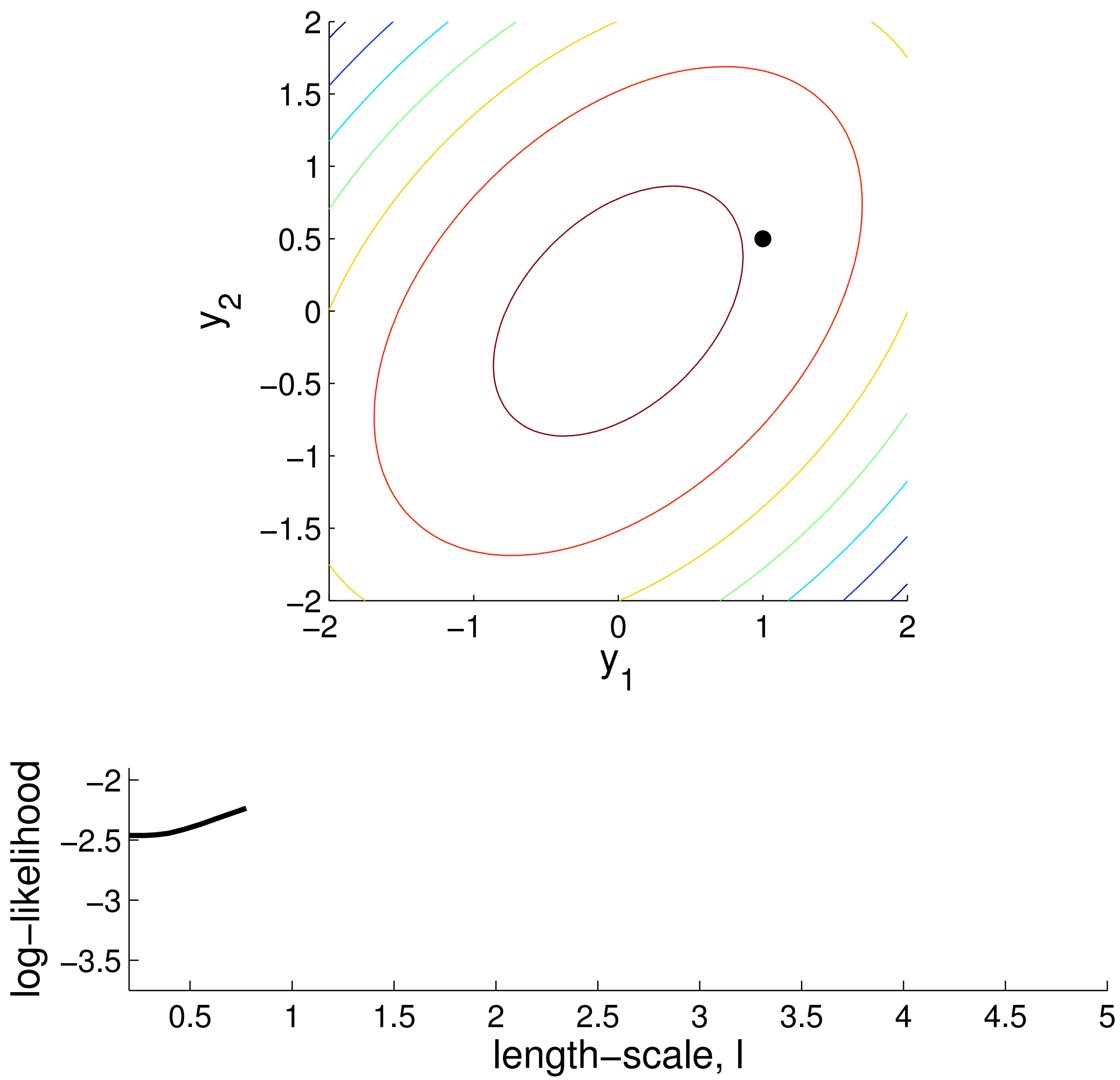




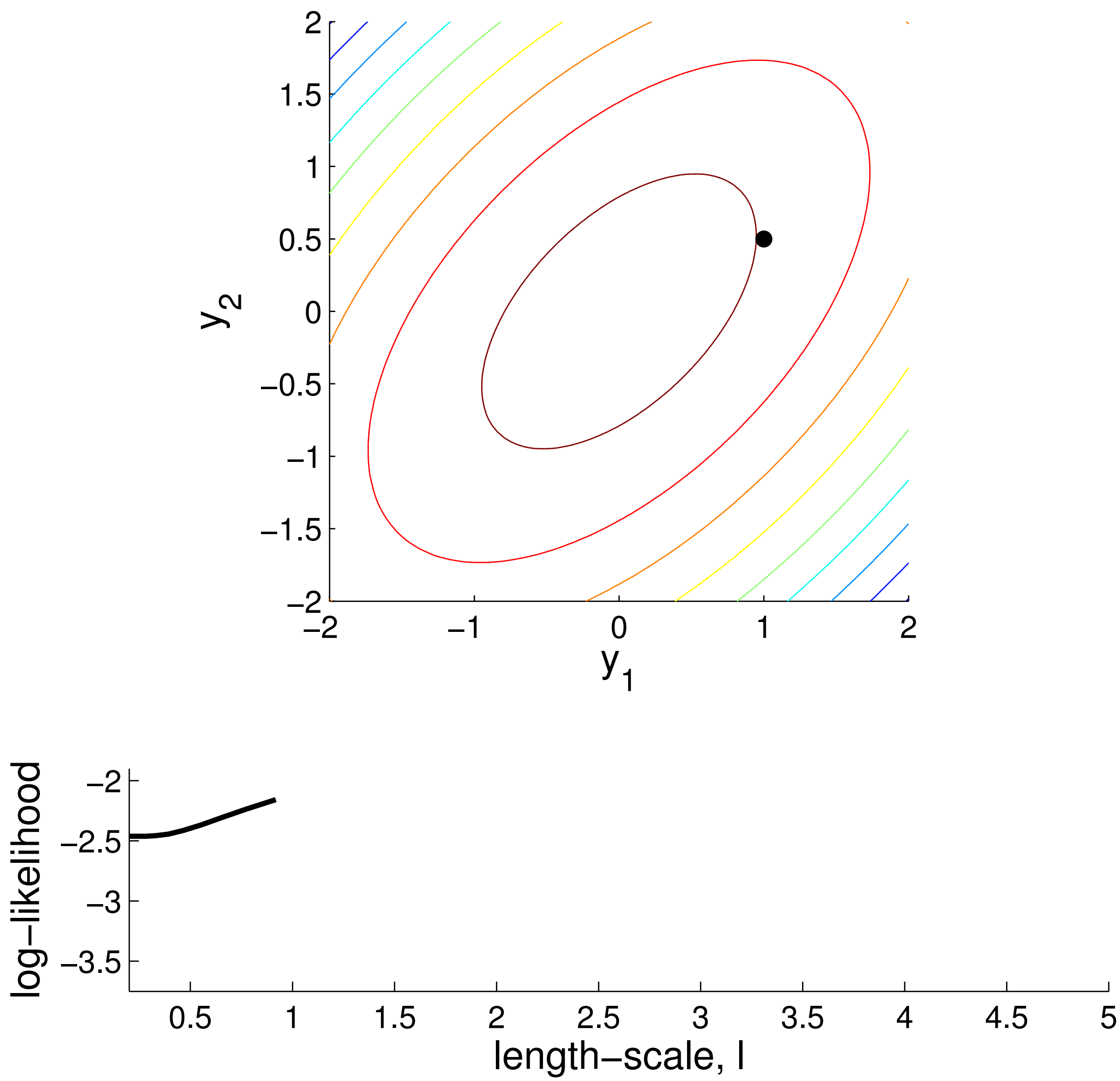
How do we choose the hyper-parameters?



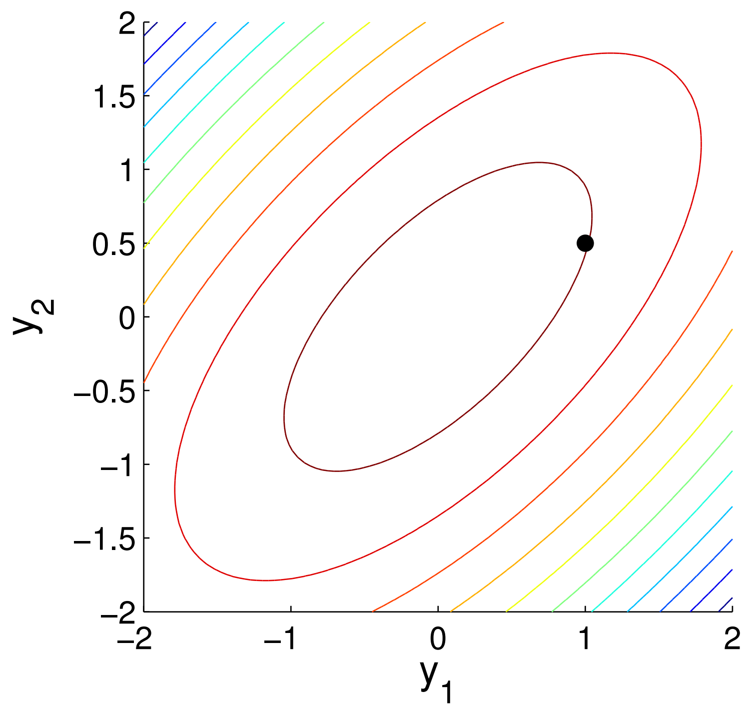
How do we choose the hyper-parameters?

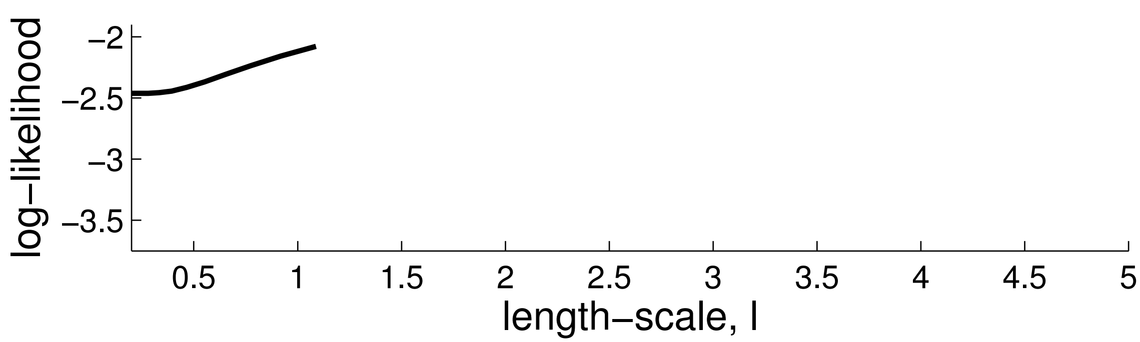


How do we choose the hyper-parameters?

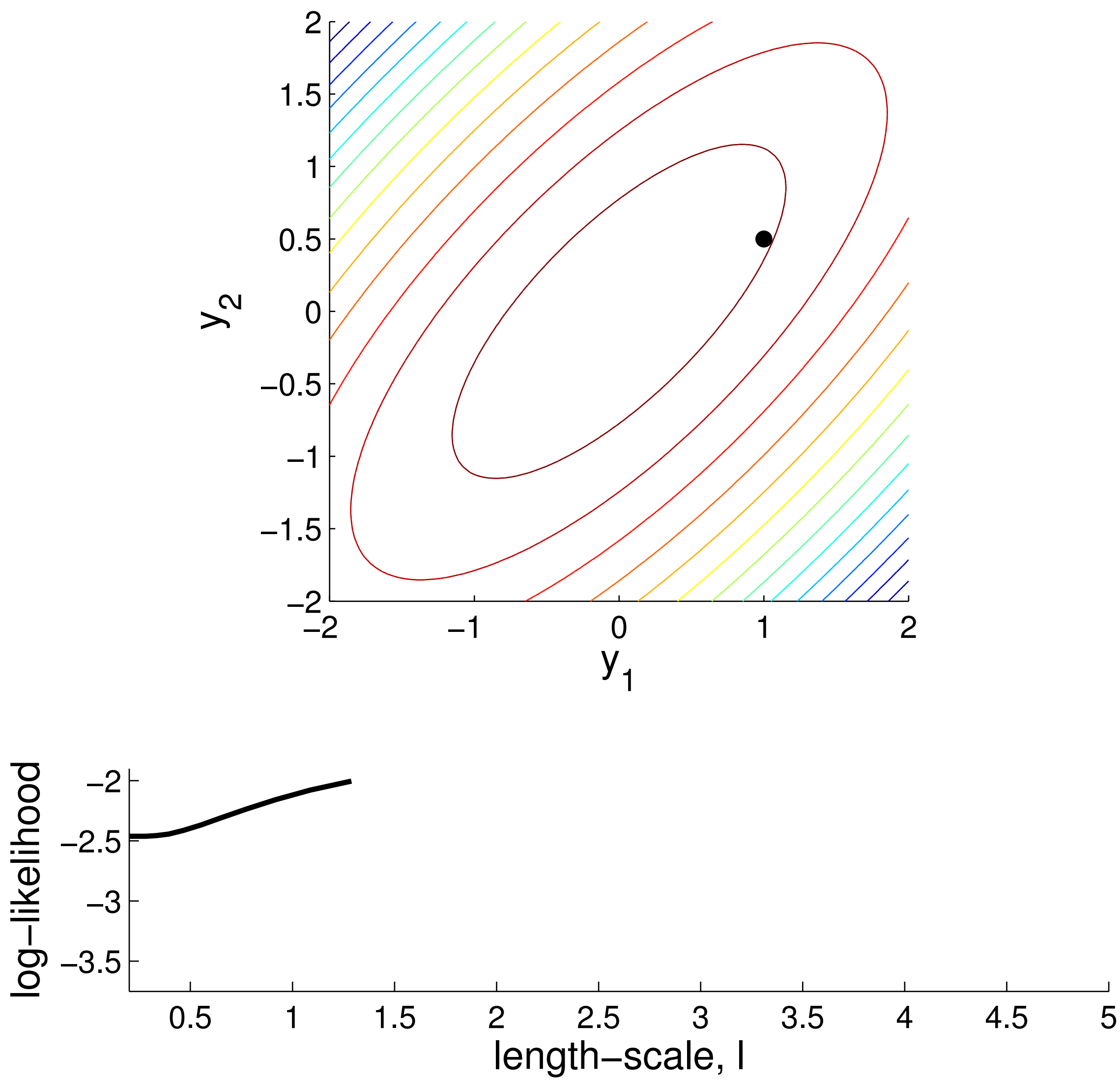


How do we choose the hyper-parameters?

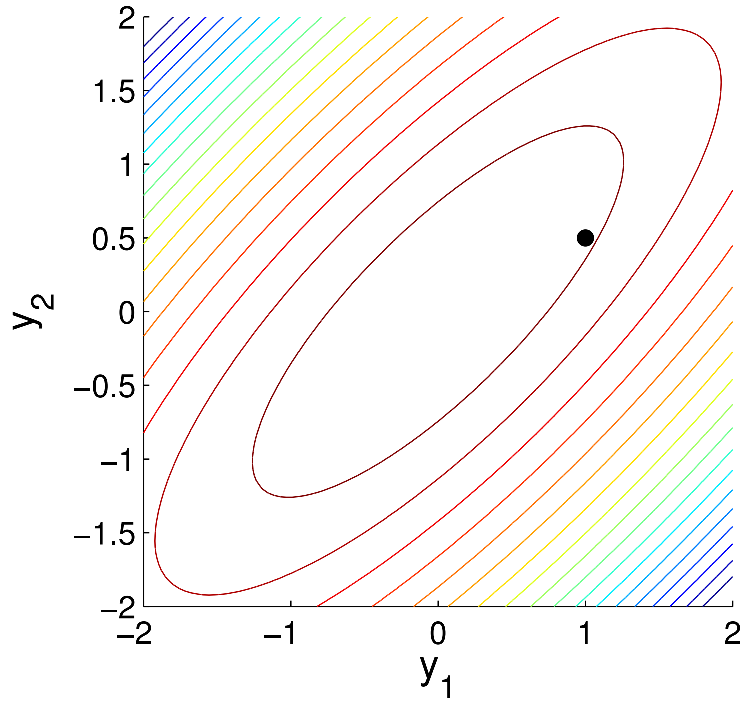


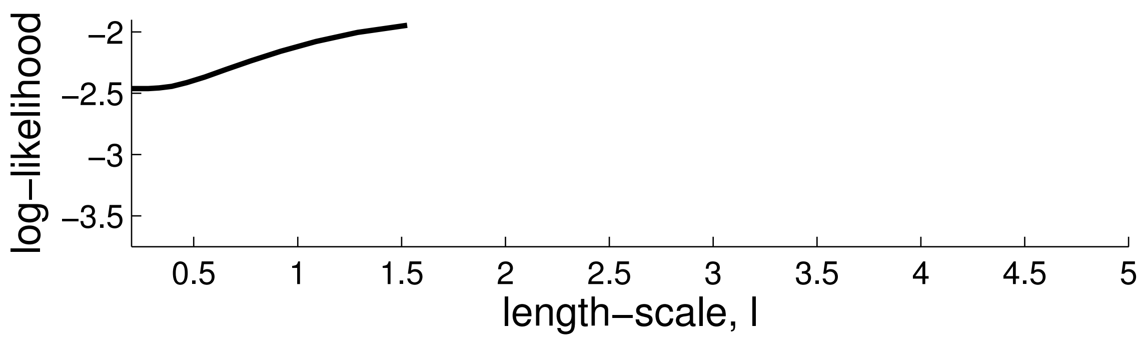


How do we choose the hyper-parameters?

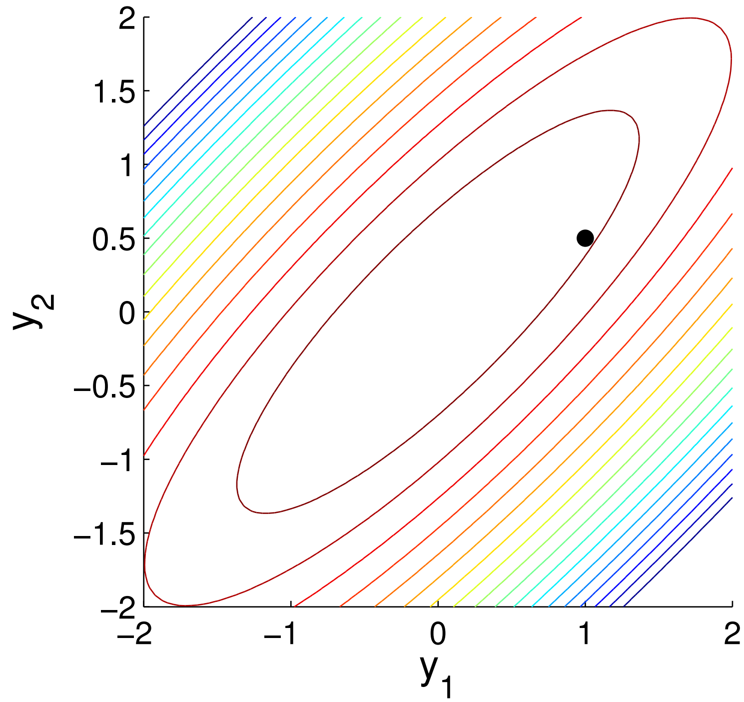


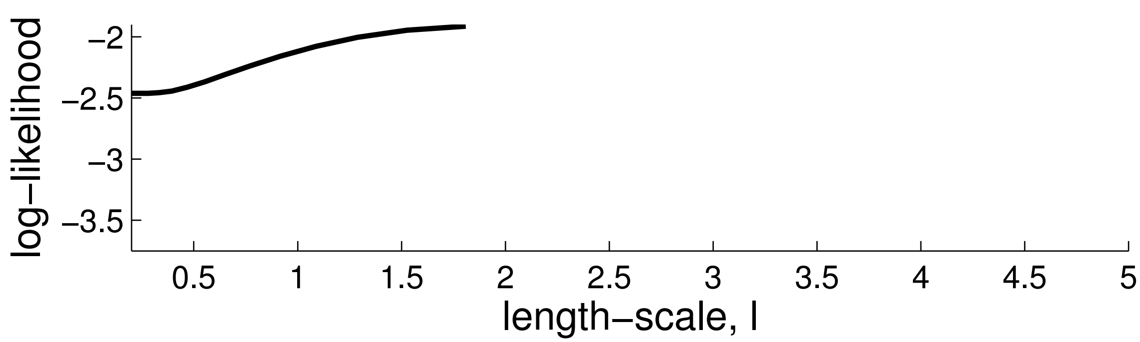
How do we choose the hyper-parameters?



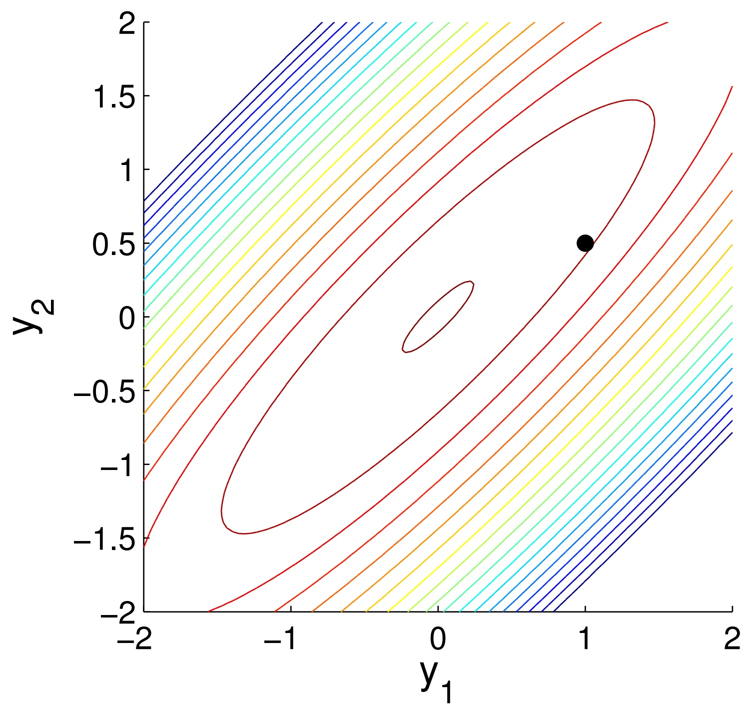


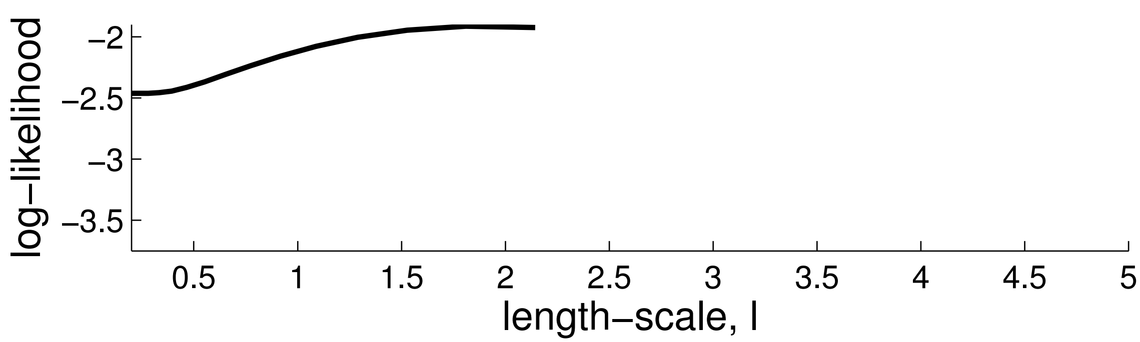
How do we choose the hyper-parameters?



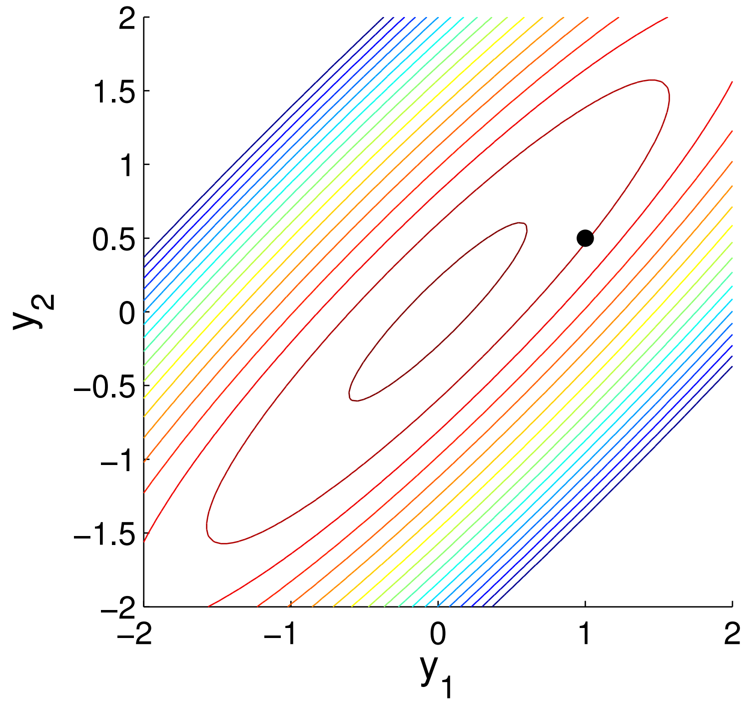


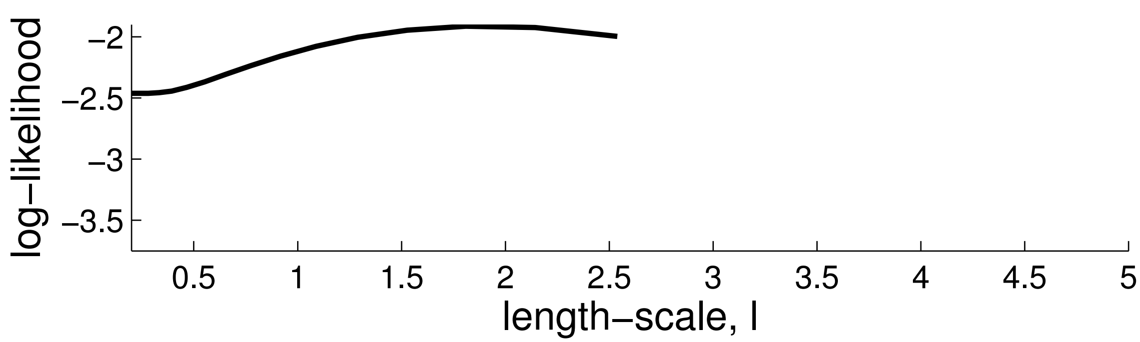
How do we choose the hyper-parameters?



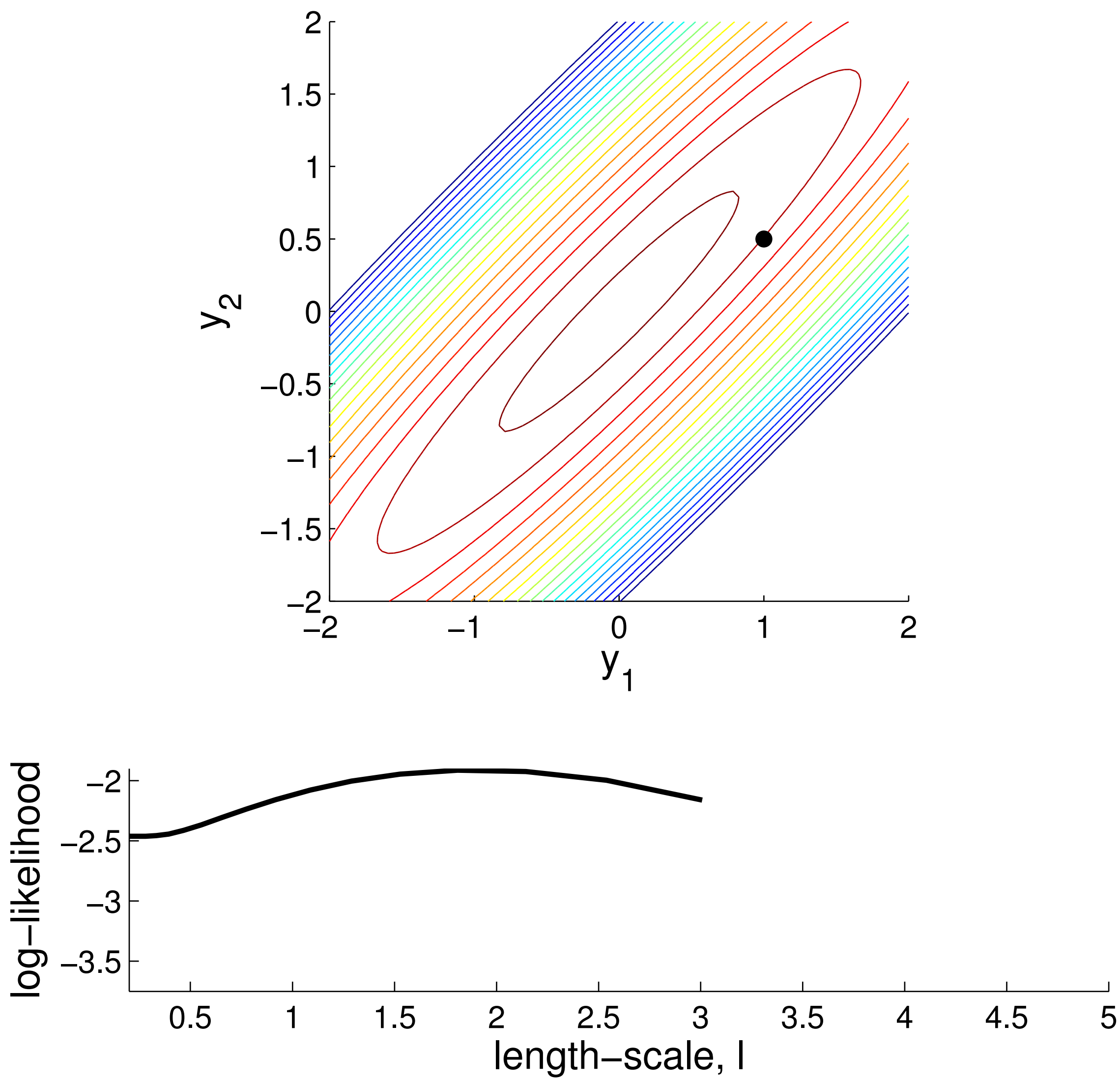


How do we choose the hyper-parameters?

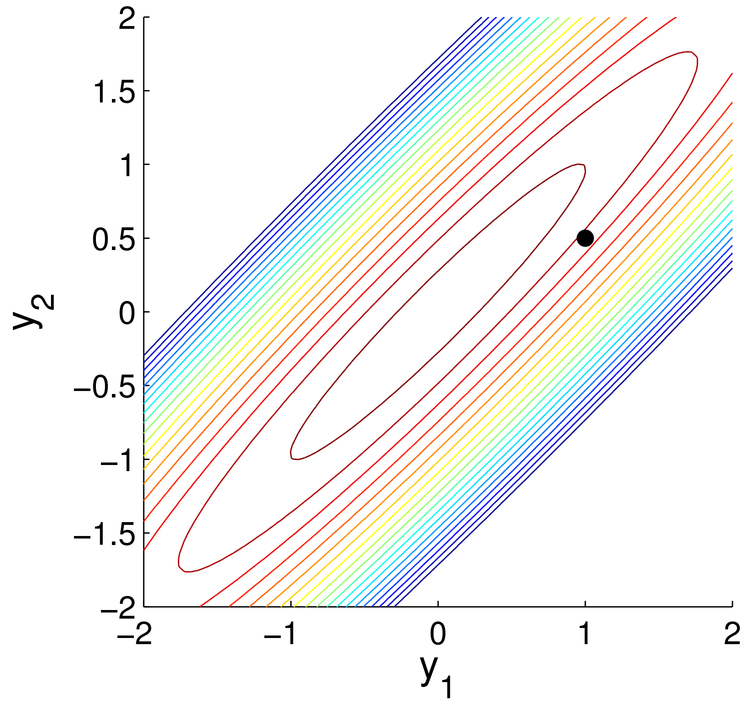


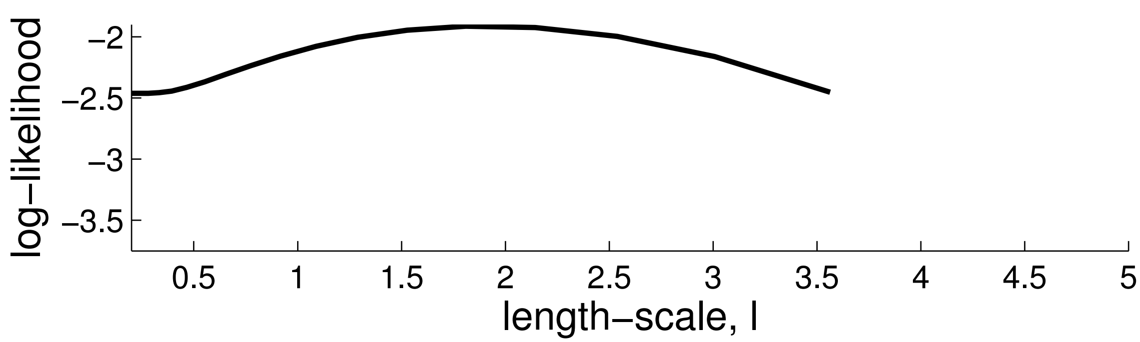


How do we choose the hyper-parameters?

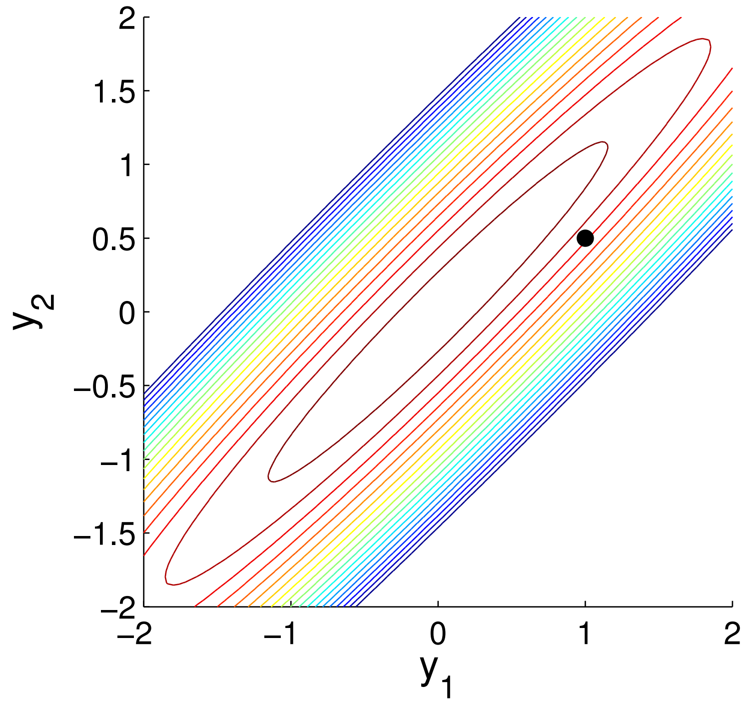


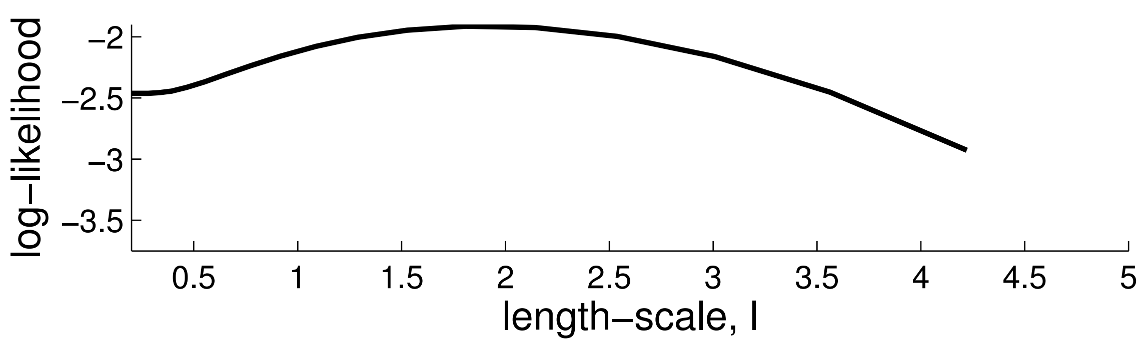
How do we choose the hyper-parameters?



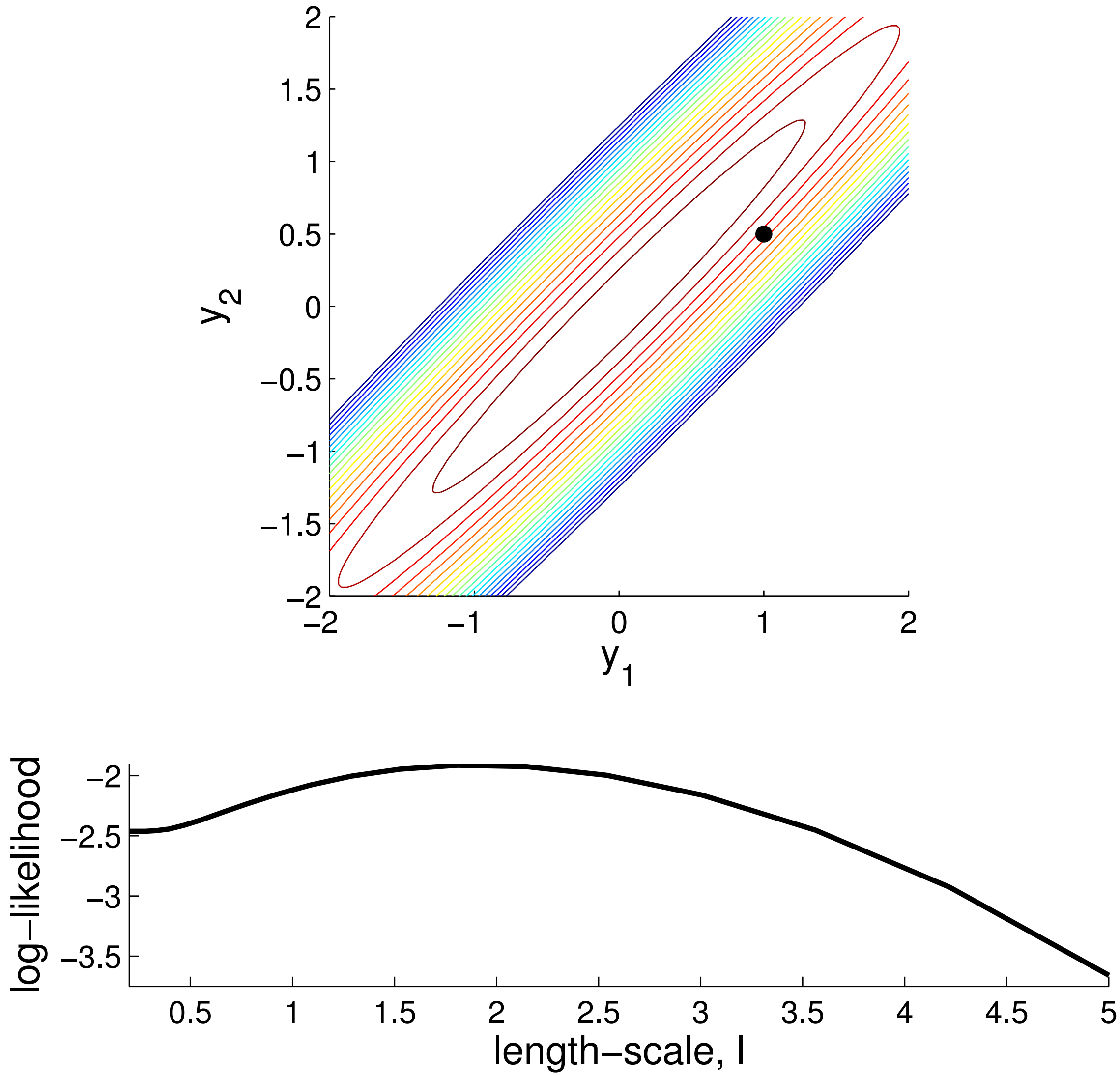


How do we choose the hyper-parameters?

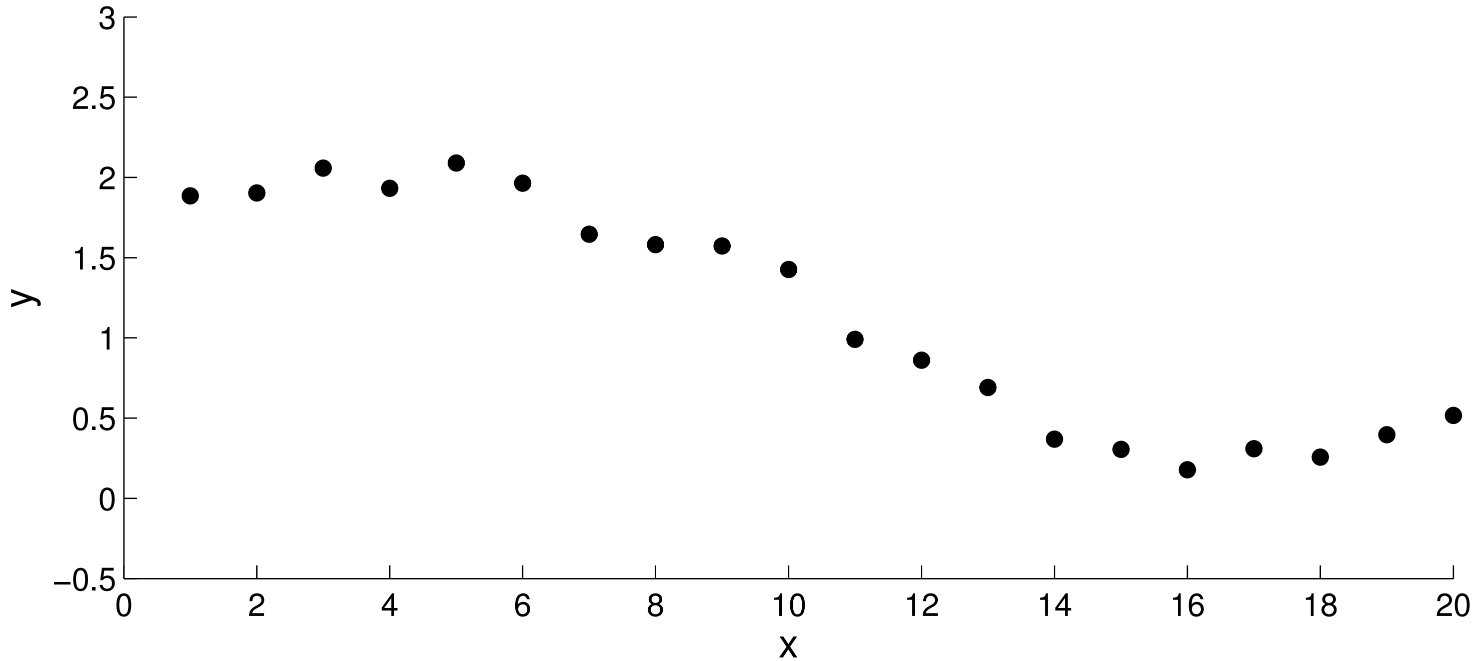




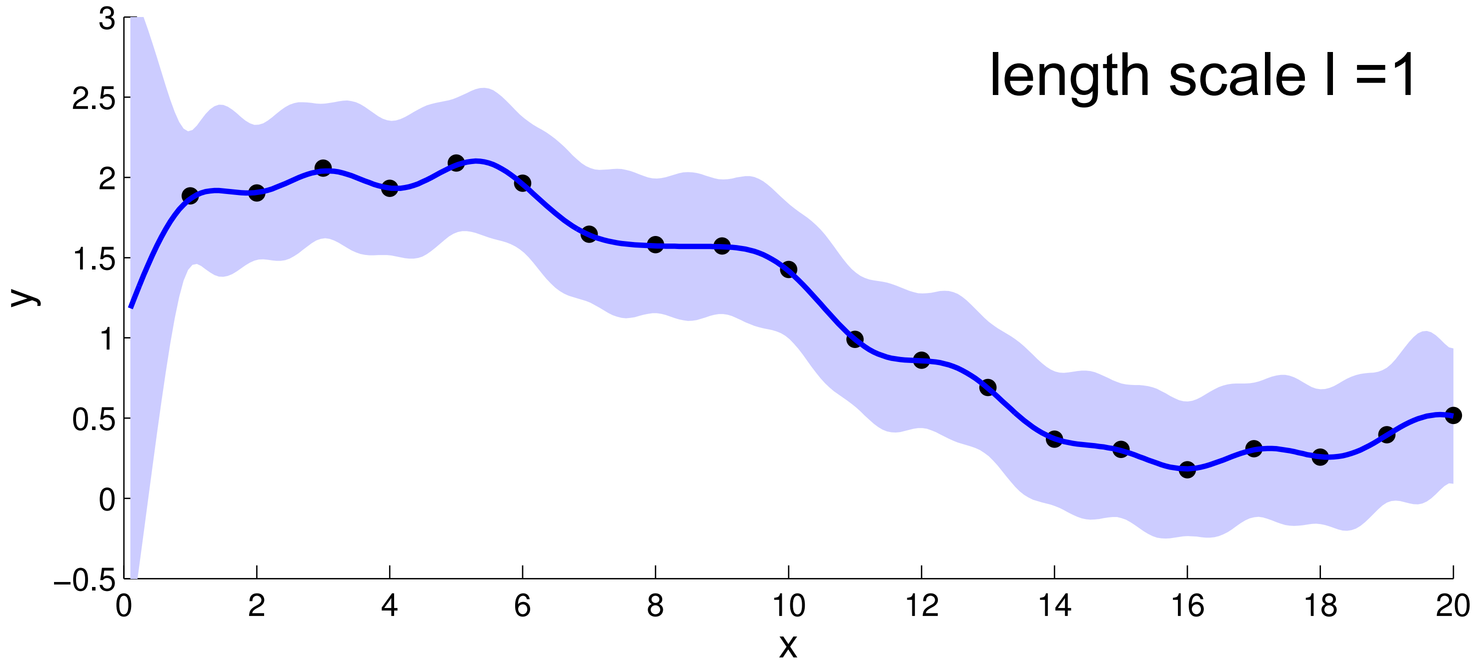
How do we choose the hyper-parameters?



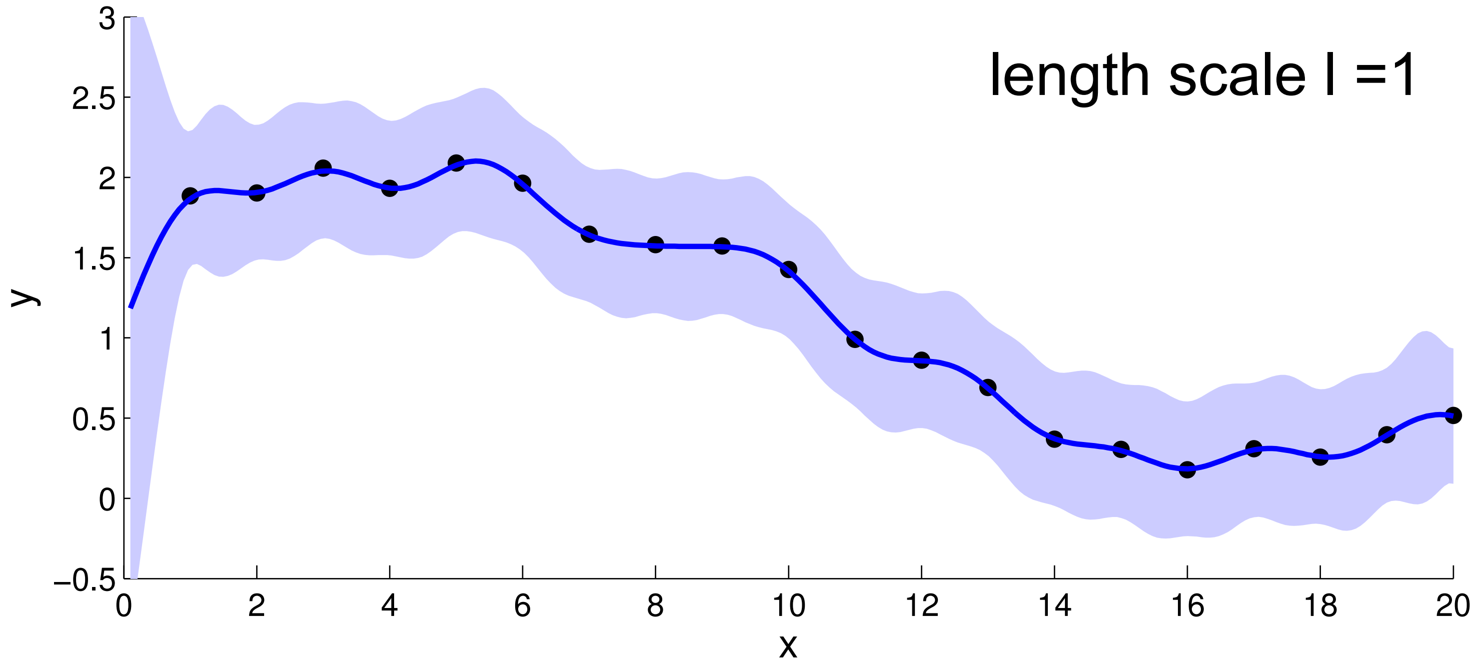
How do we choose the hyper-parameters?

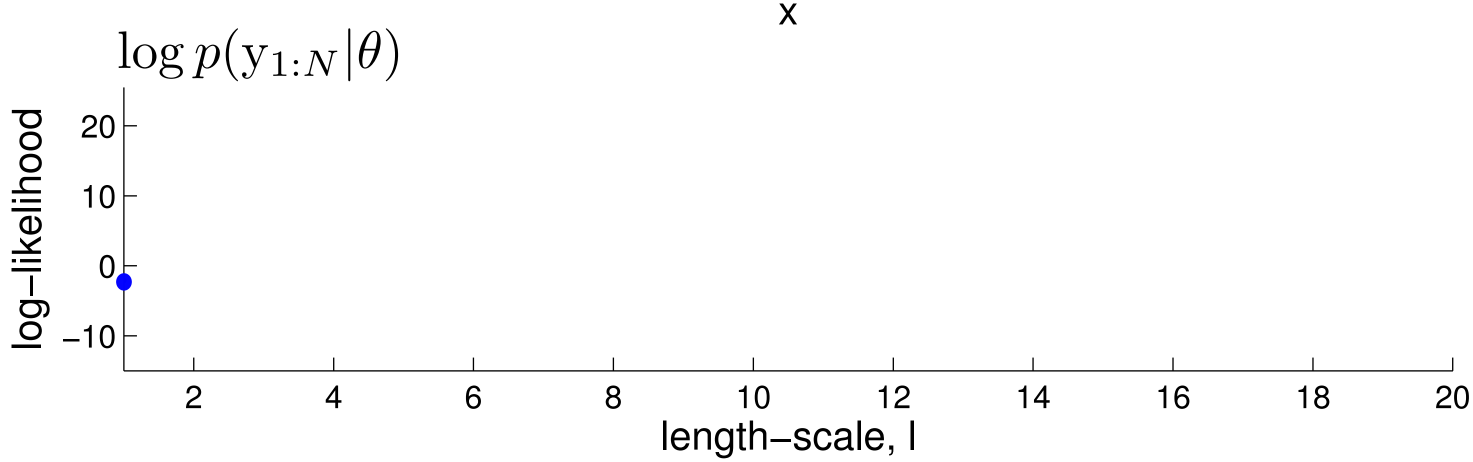


How do we choose the hyper-parameters?

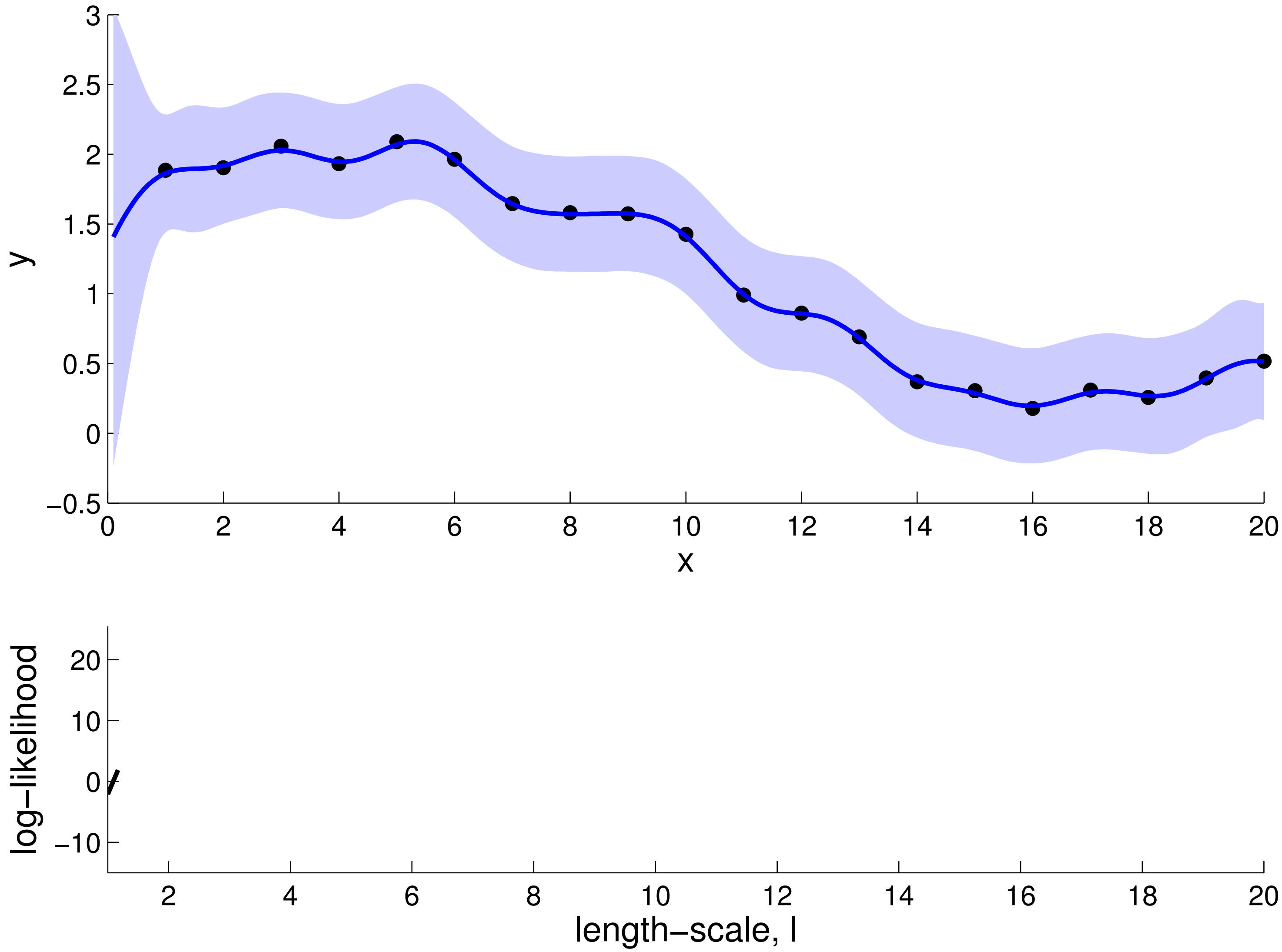


How do we choose the hyper-parameters?

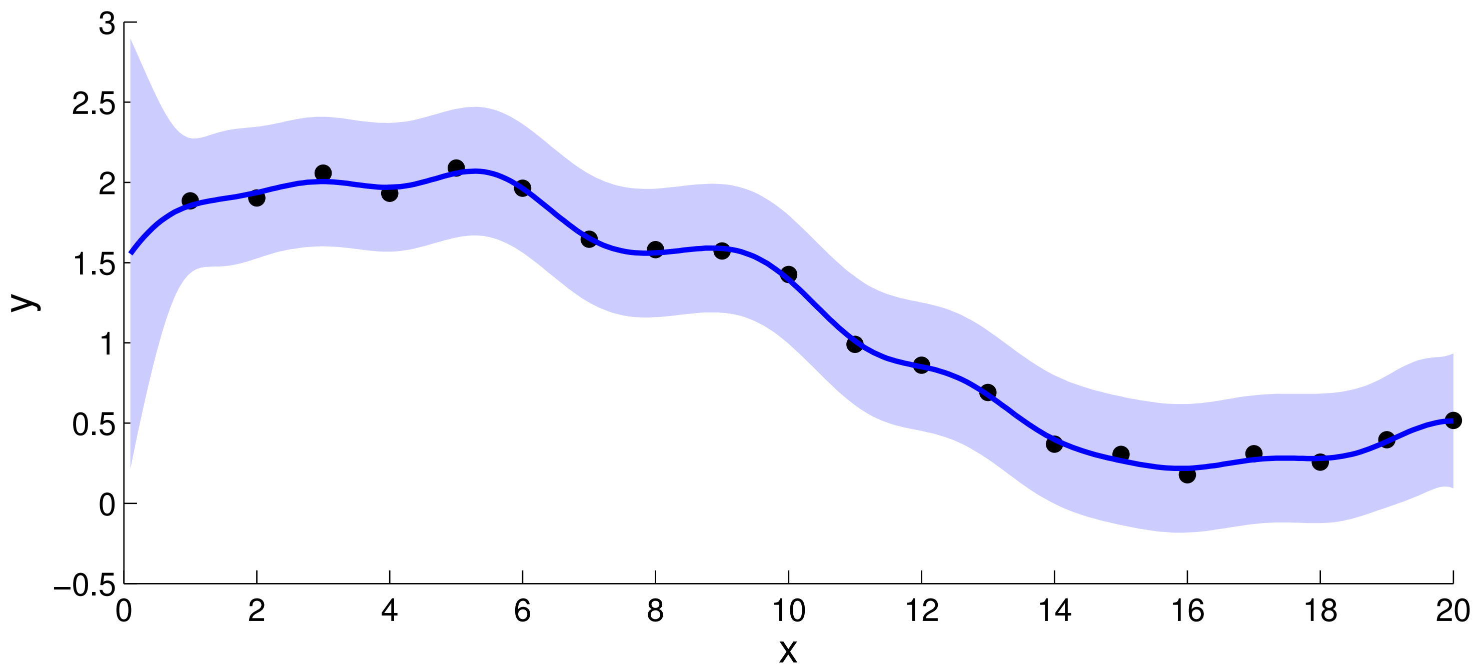


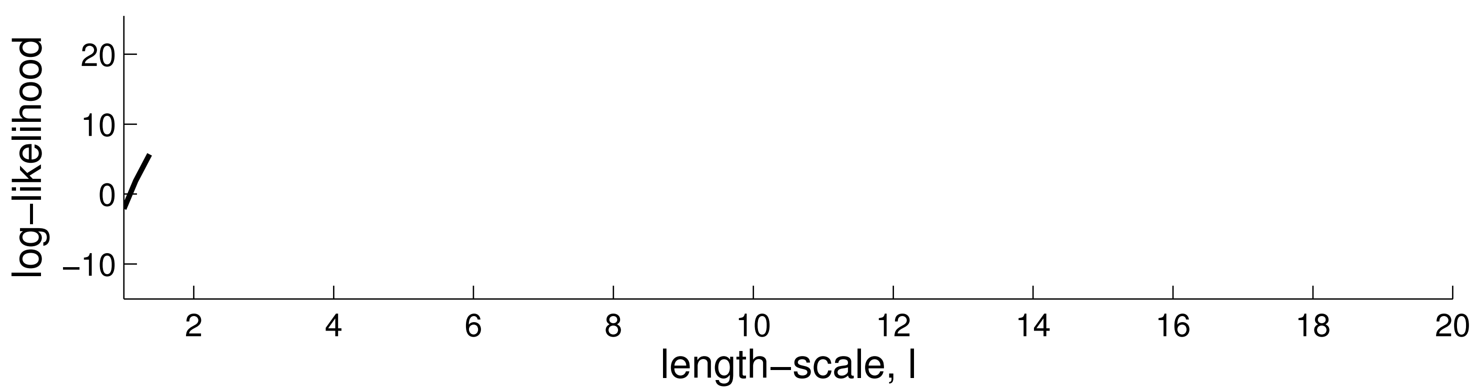


How do we choose the hyper-parameters?

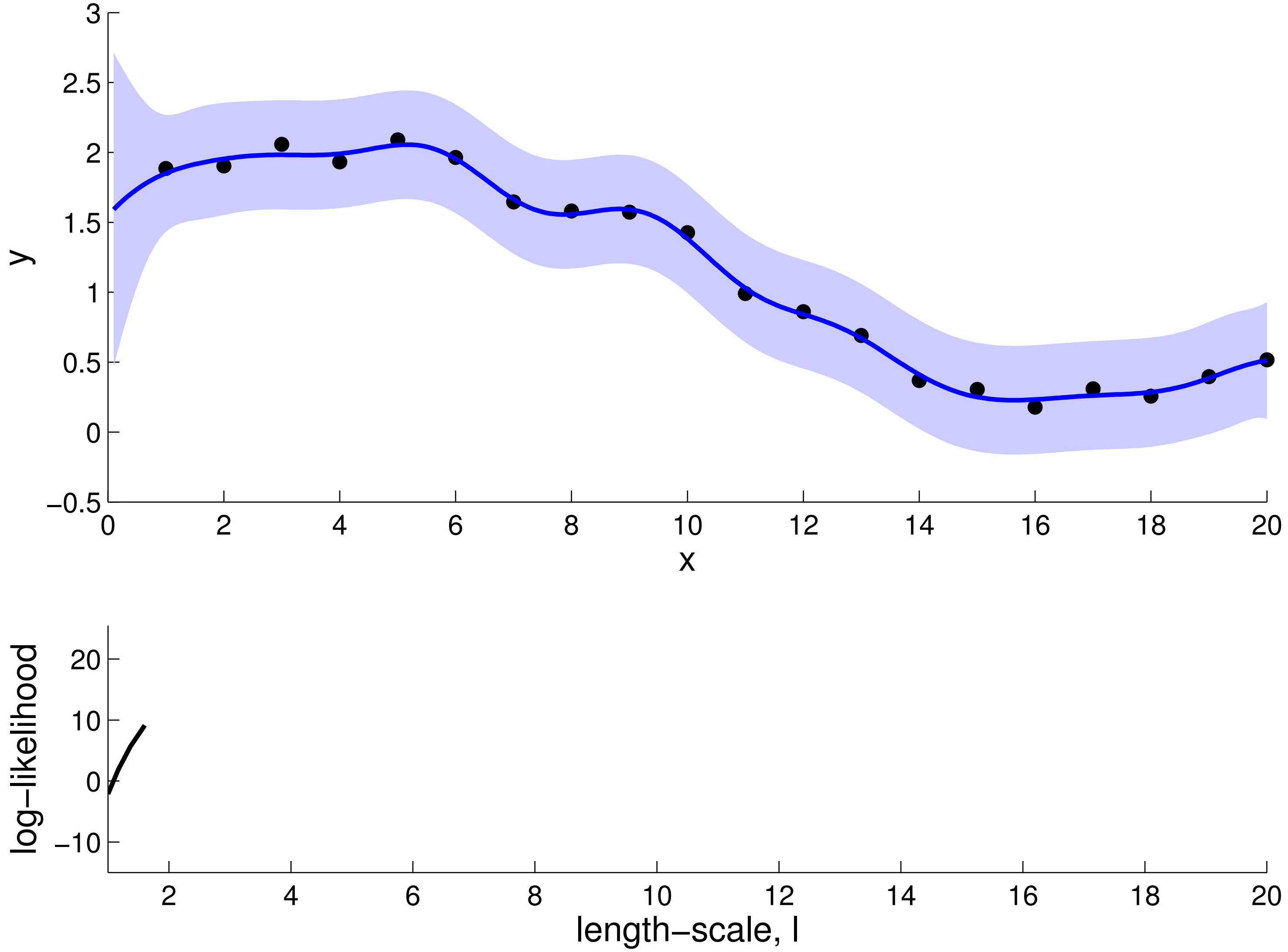


How do we choose the hyper-parameters?

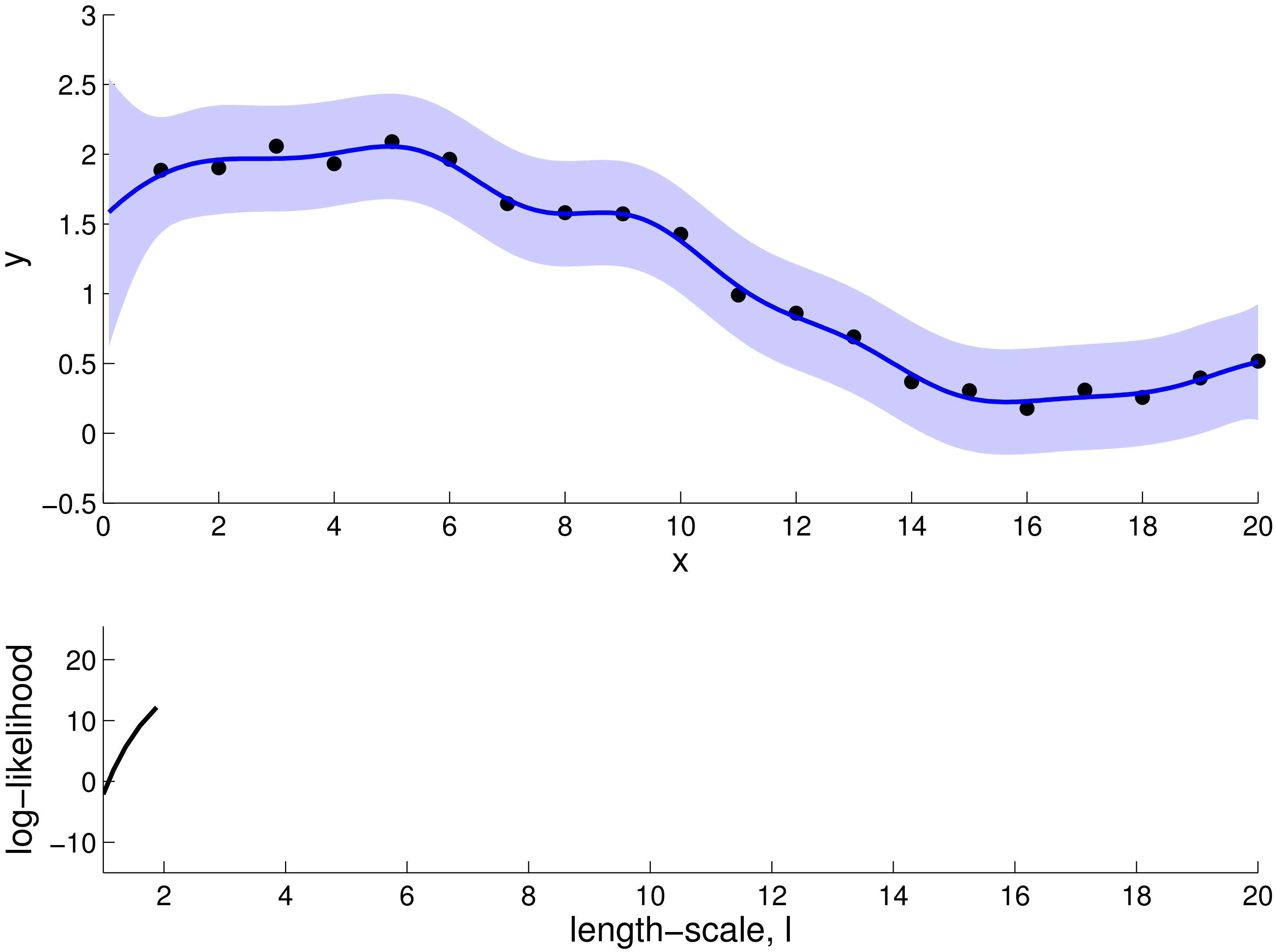




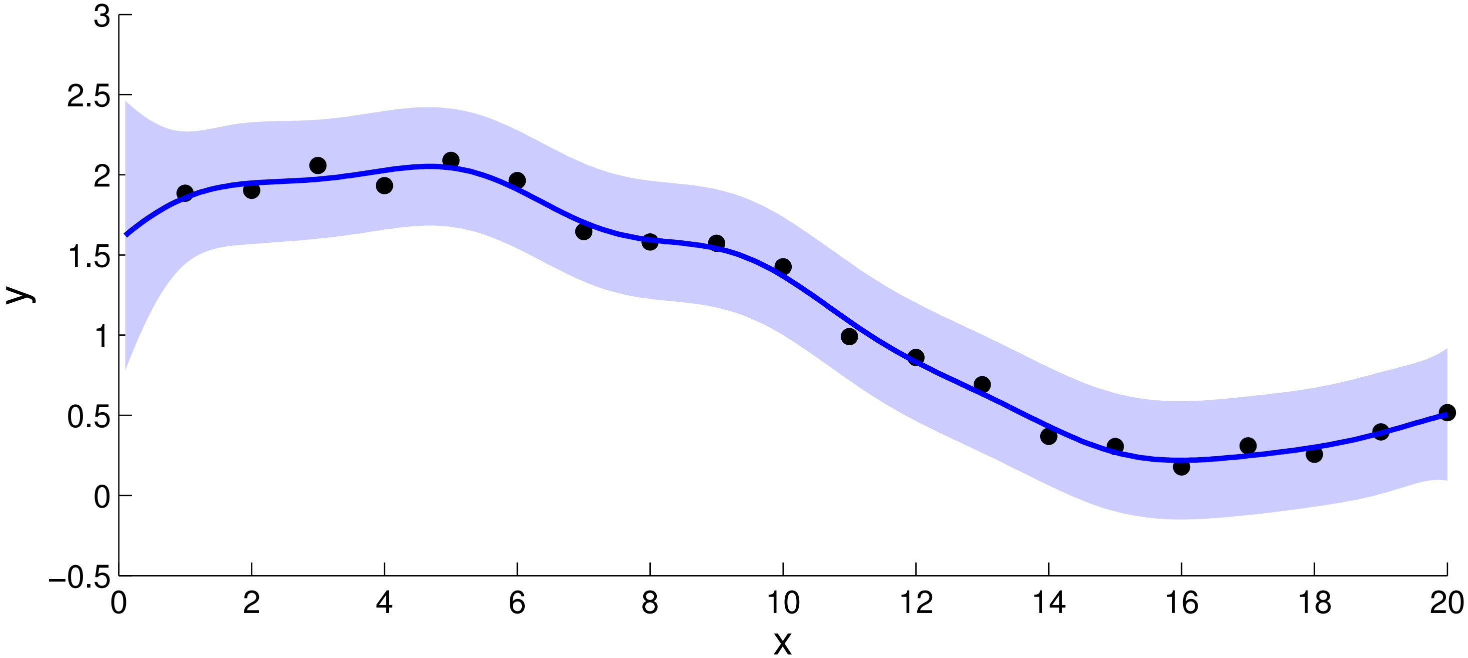
How do we choose the hyper-parameters?

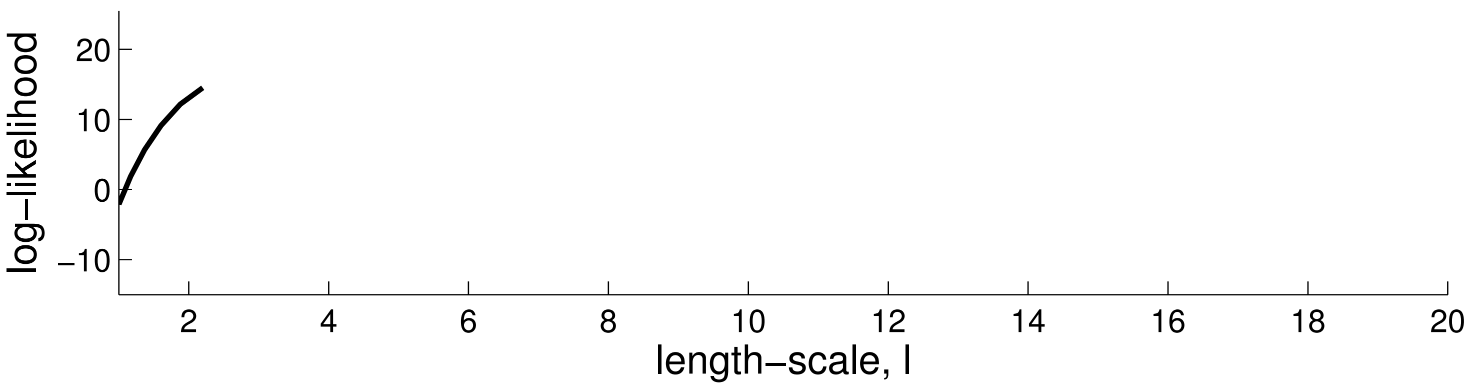


How do we choose the hyper-parameters?

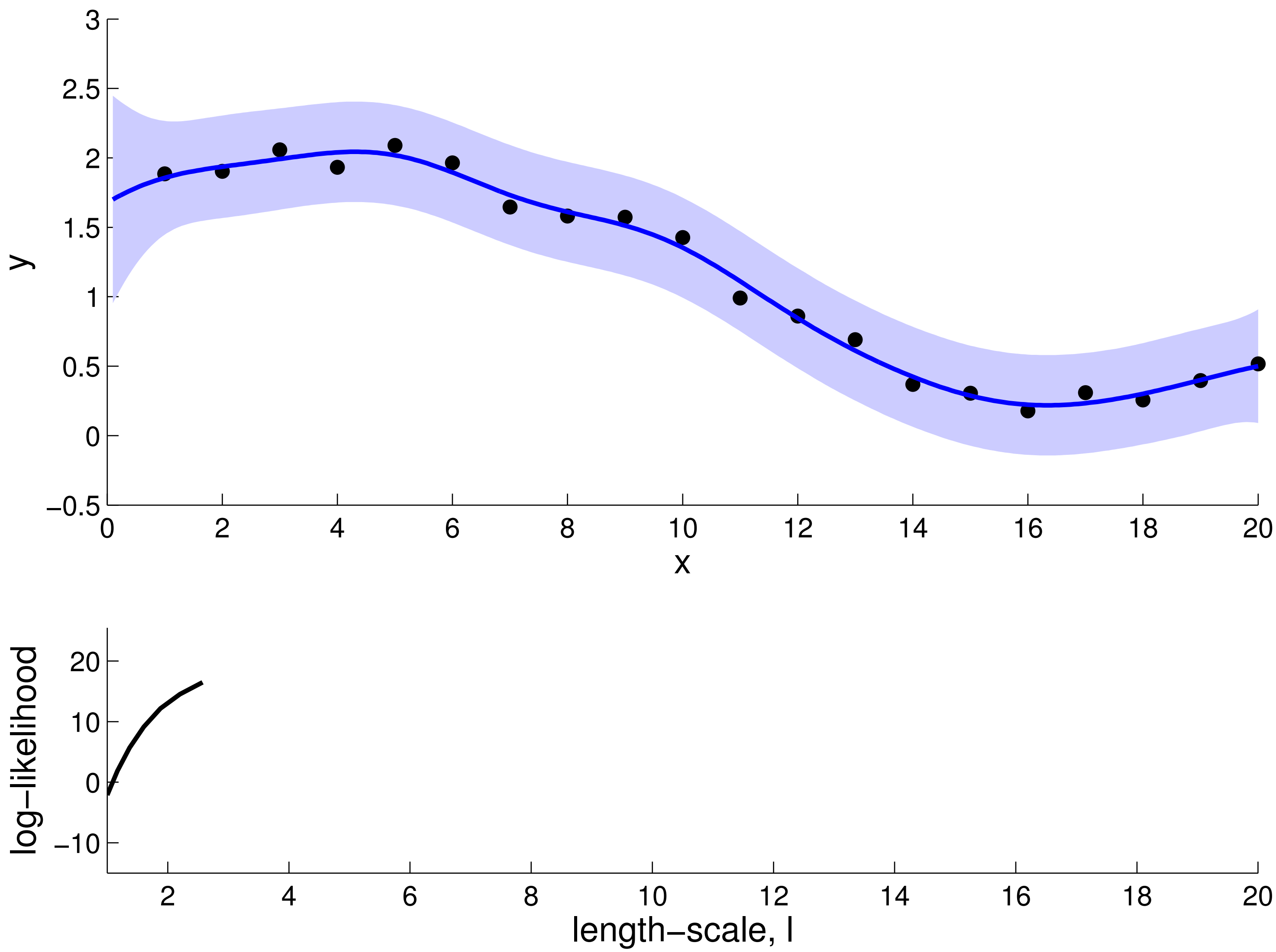


How do we choose the hyper-parameters?

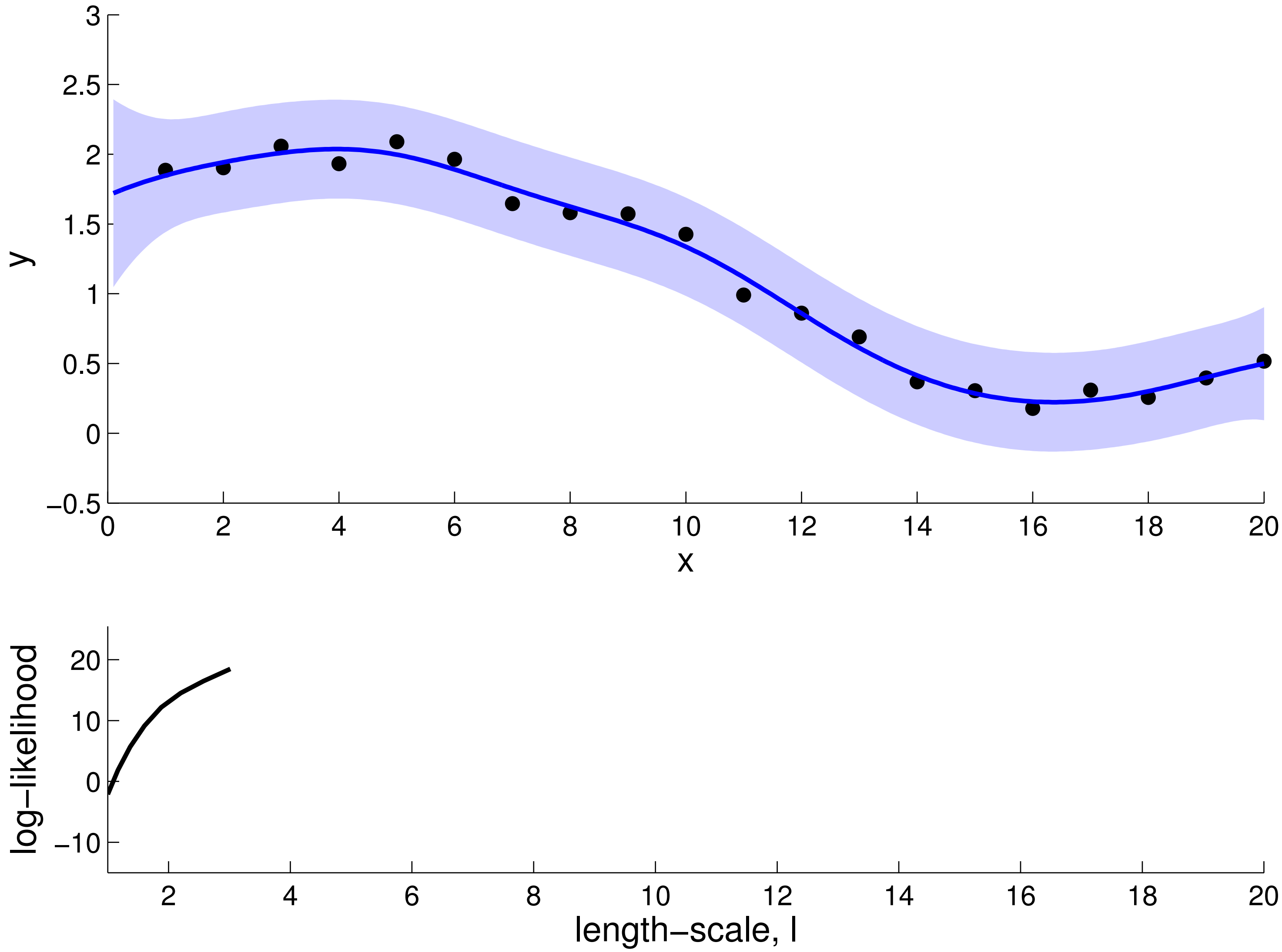




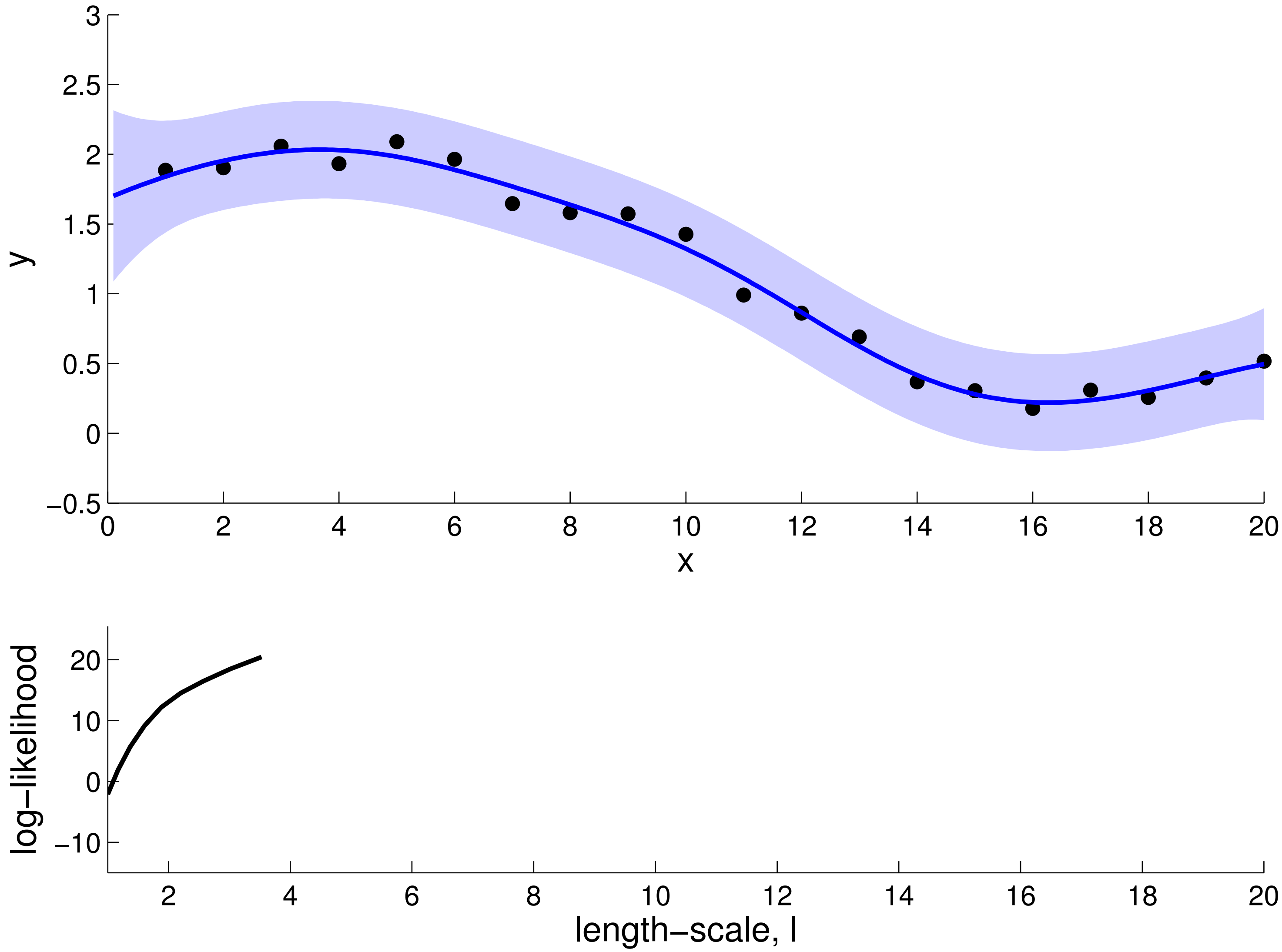
How do we choose the hyper-parameters?



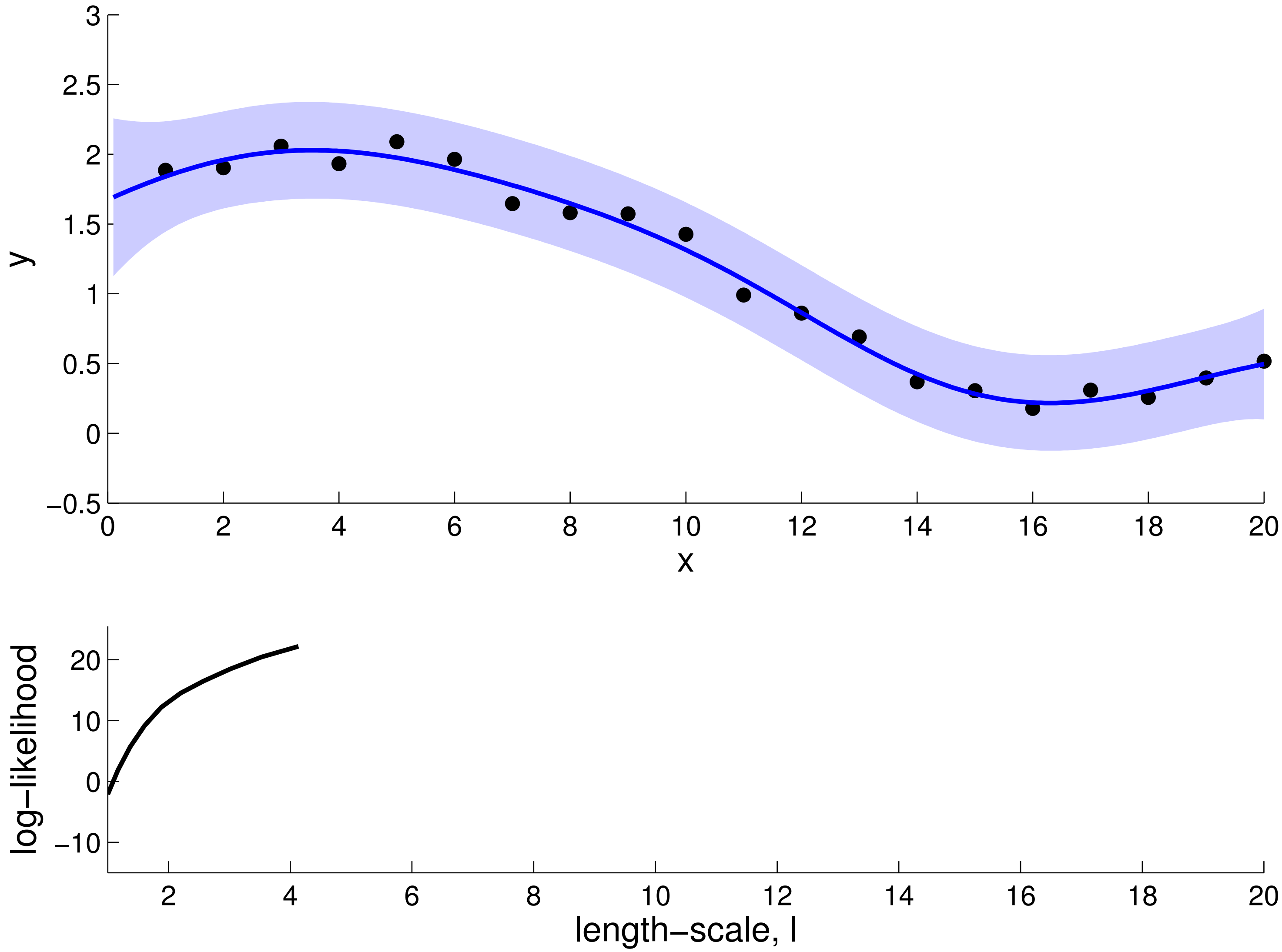
How do we choose the hyper-parameters?



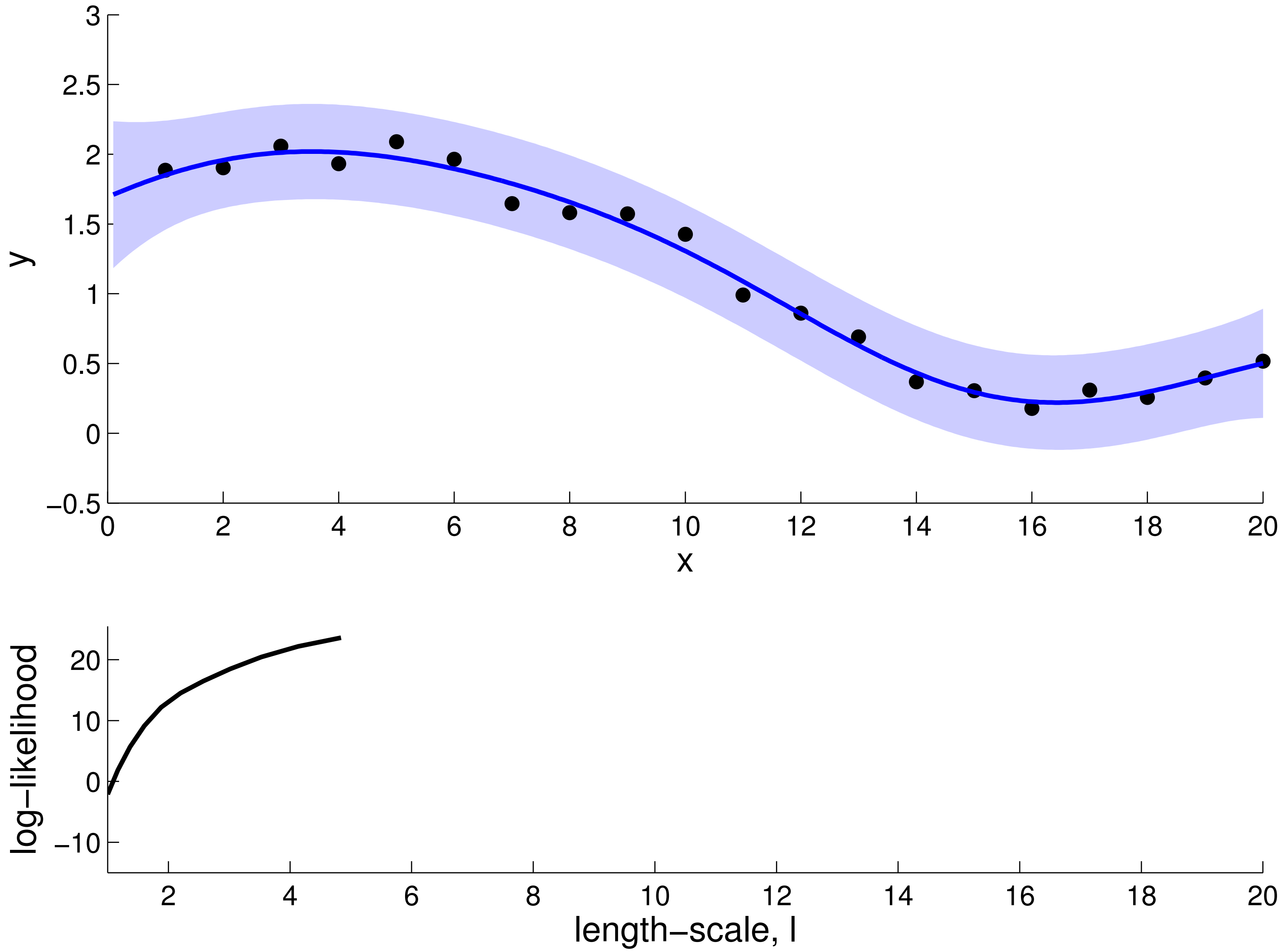
How do we choose the hyper-parameters?



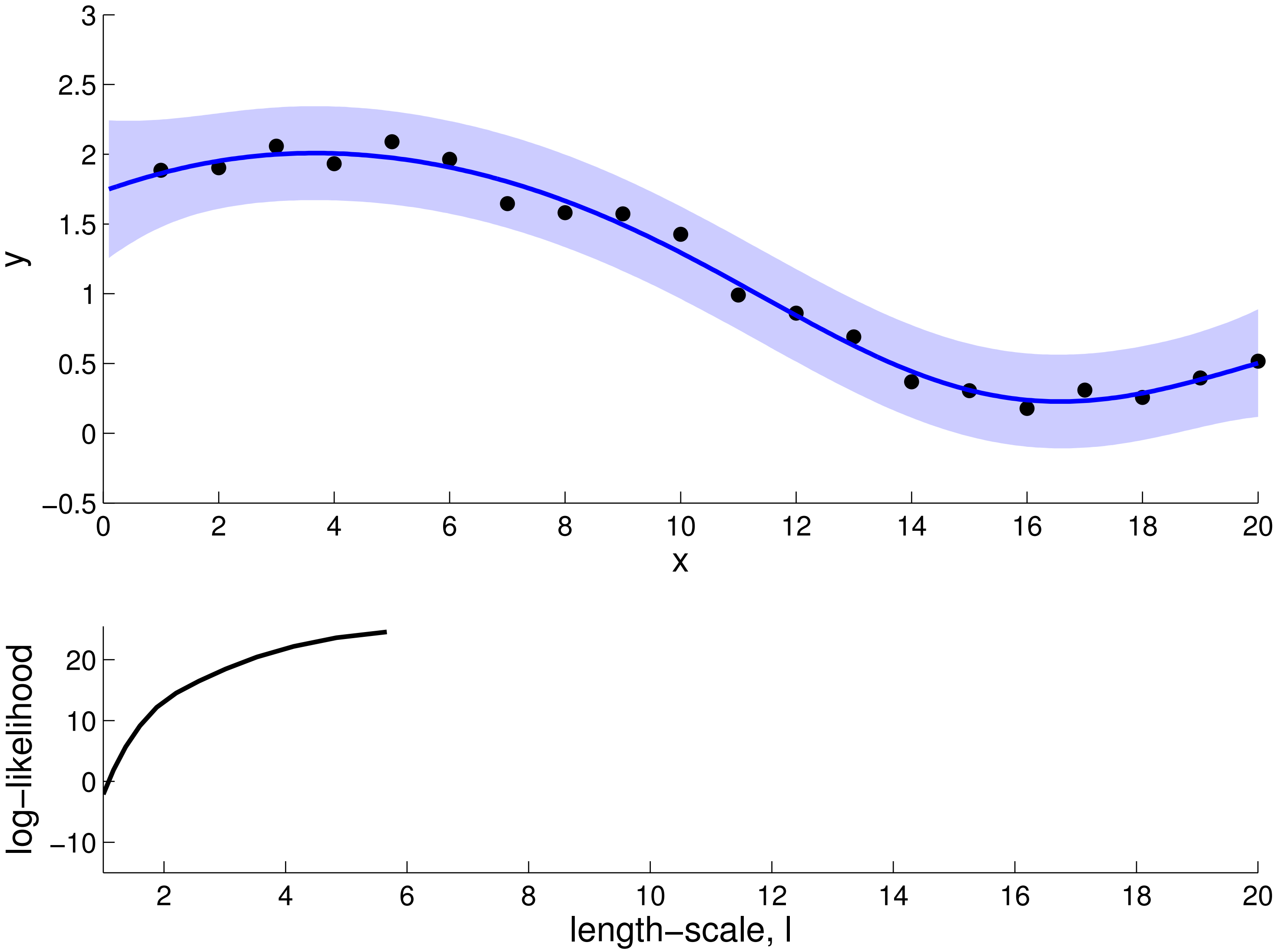
How do we choose the hyper-parameters?



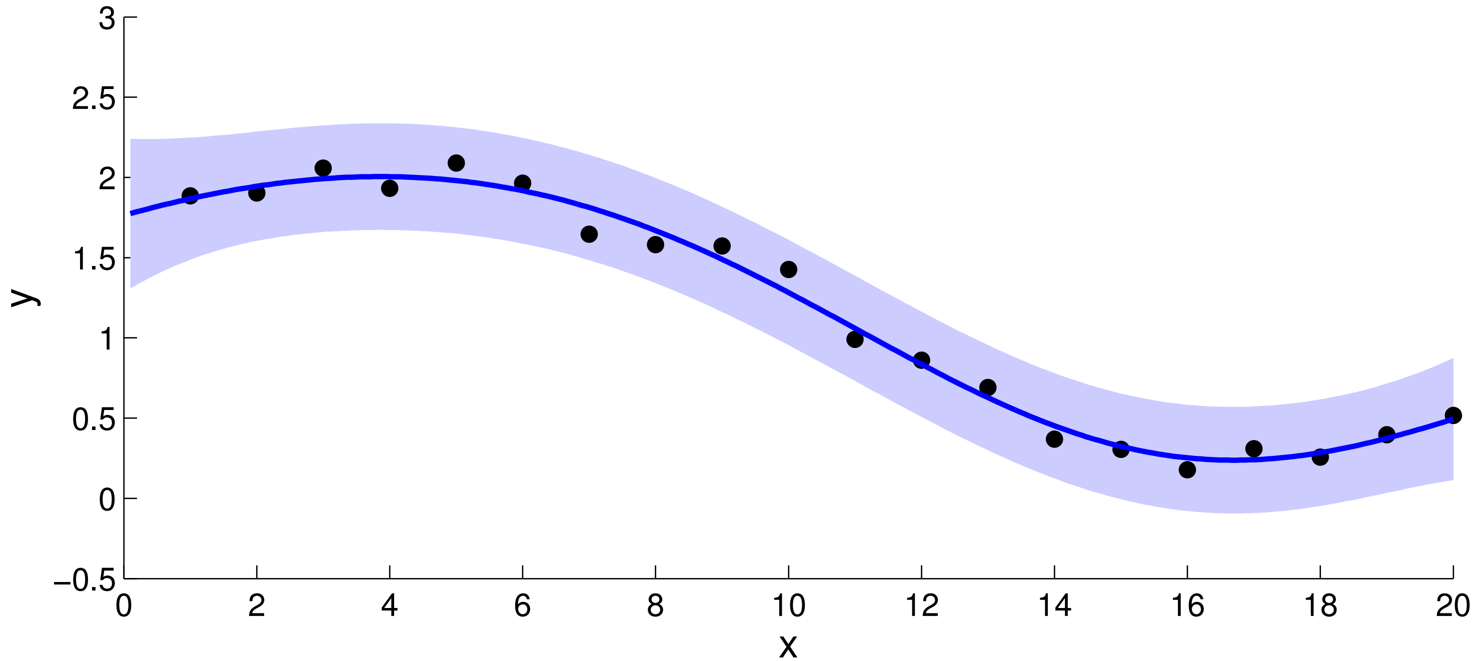
How do we choose the hyper-parameters?

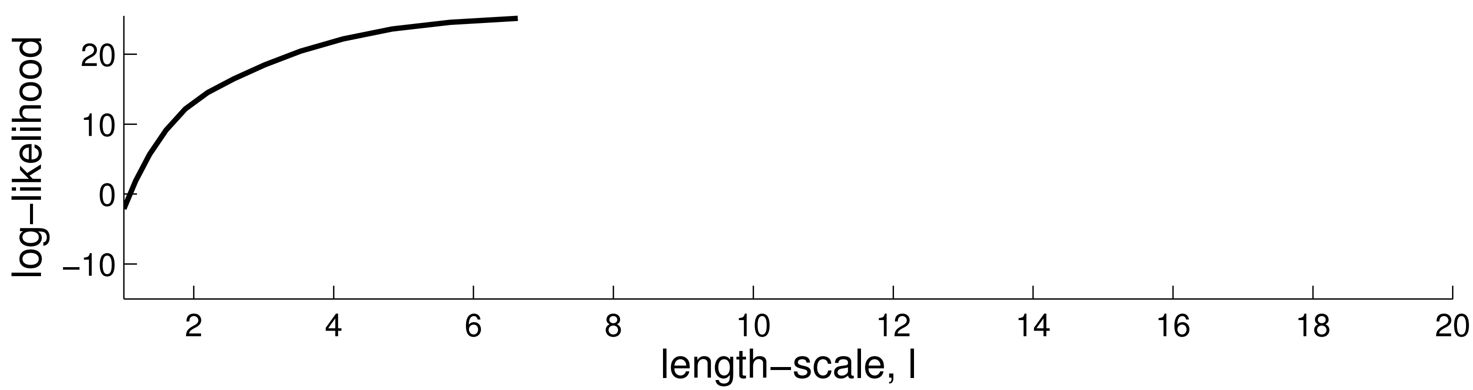


How do we choose the hyper-parameters?

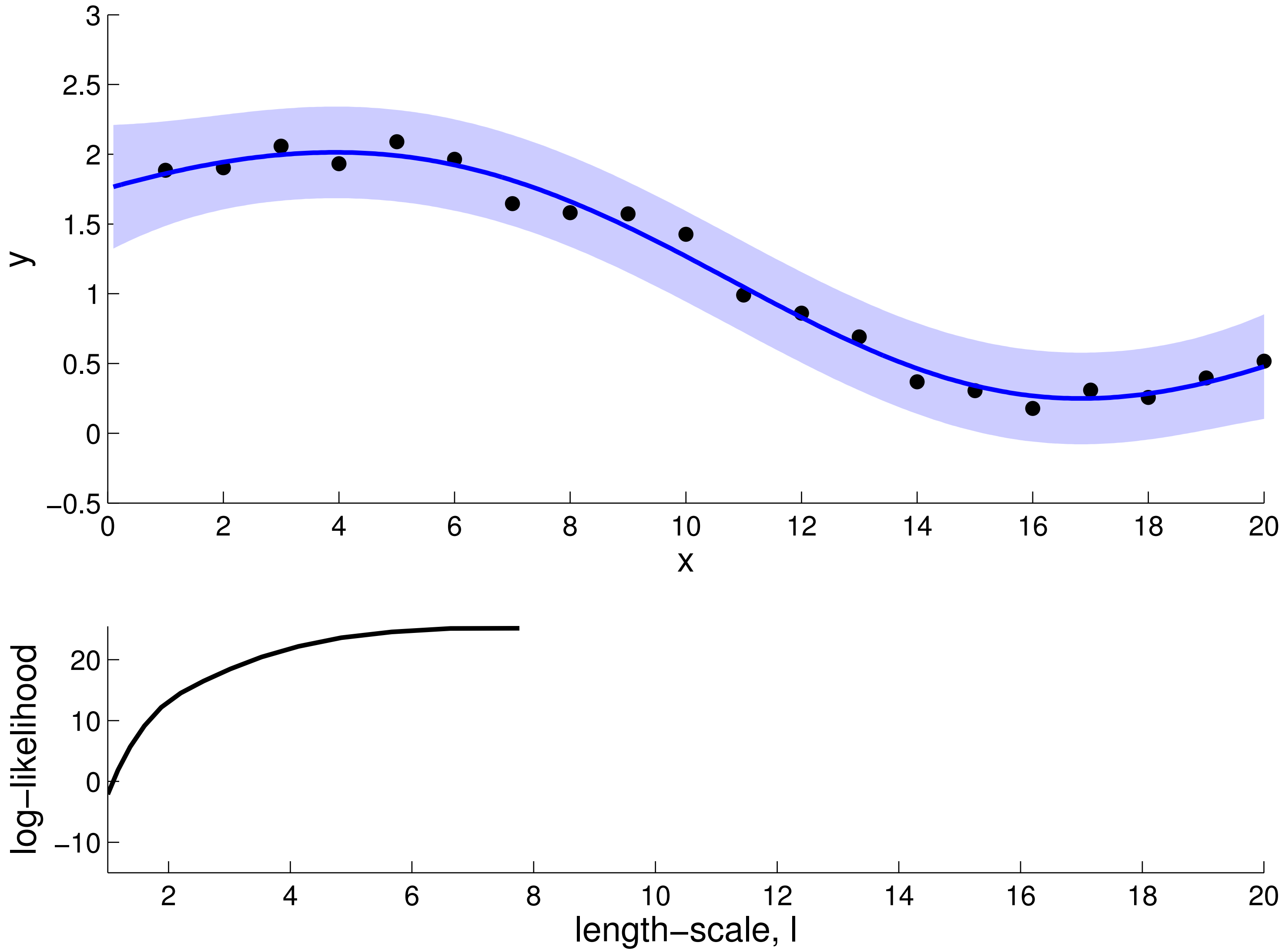


How do we choose the hyper-parameters?

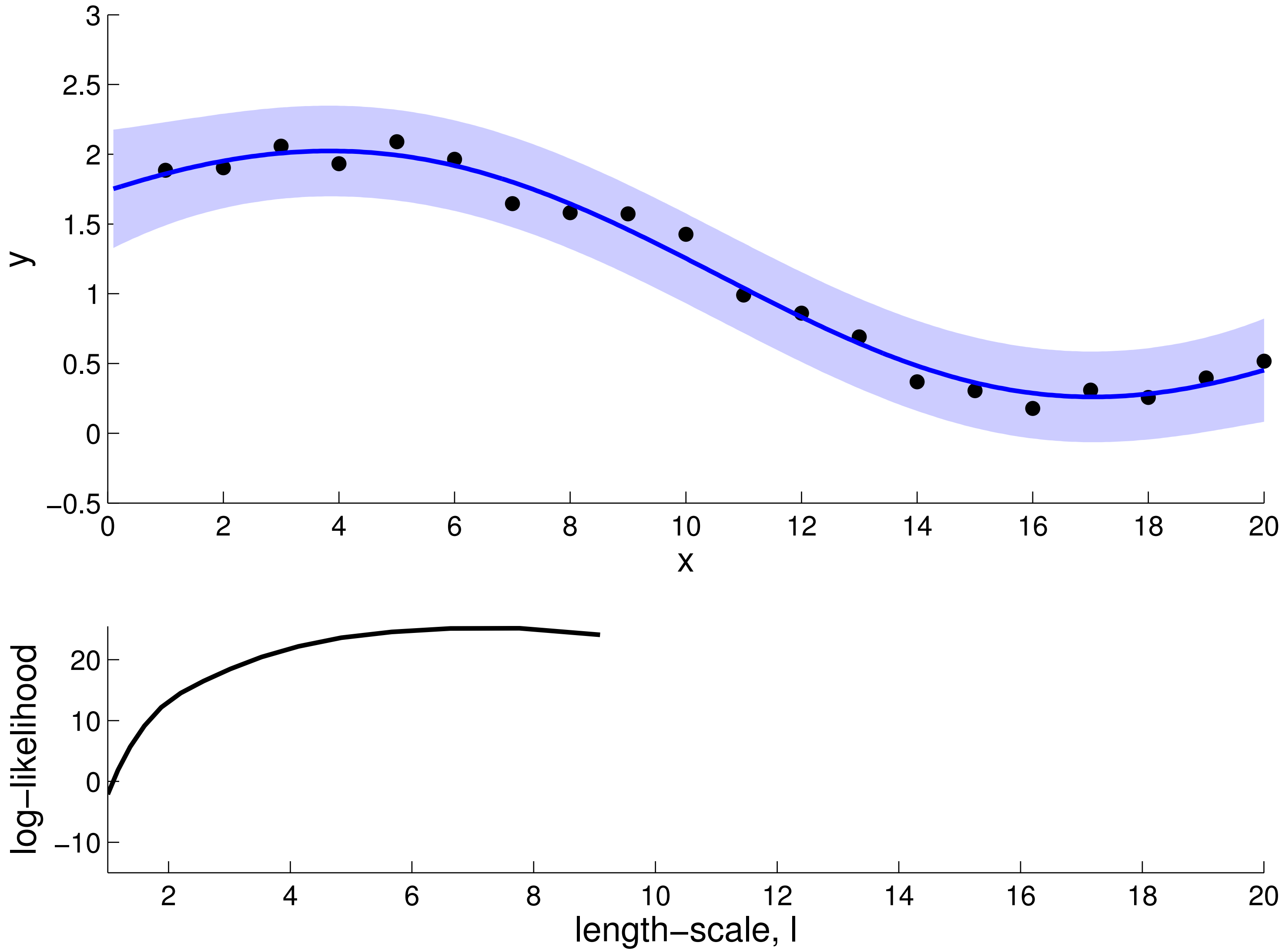




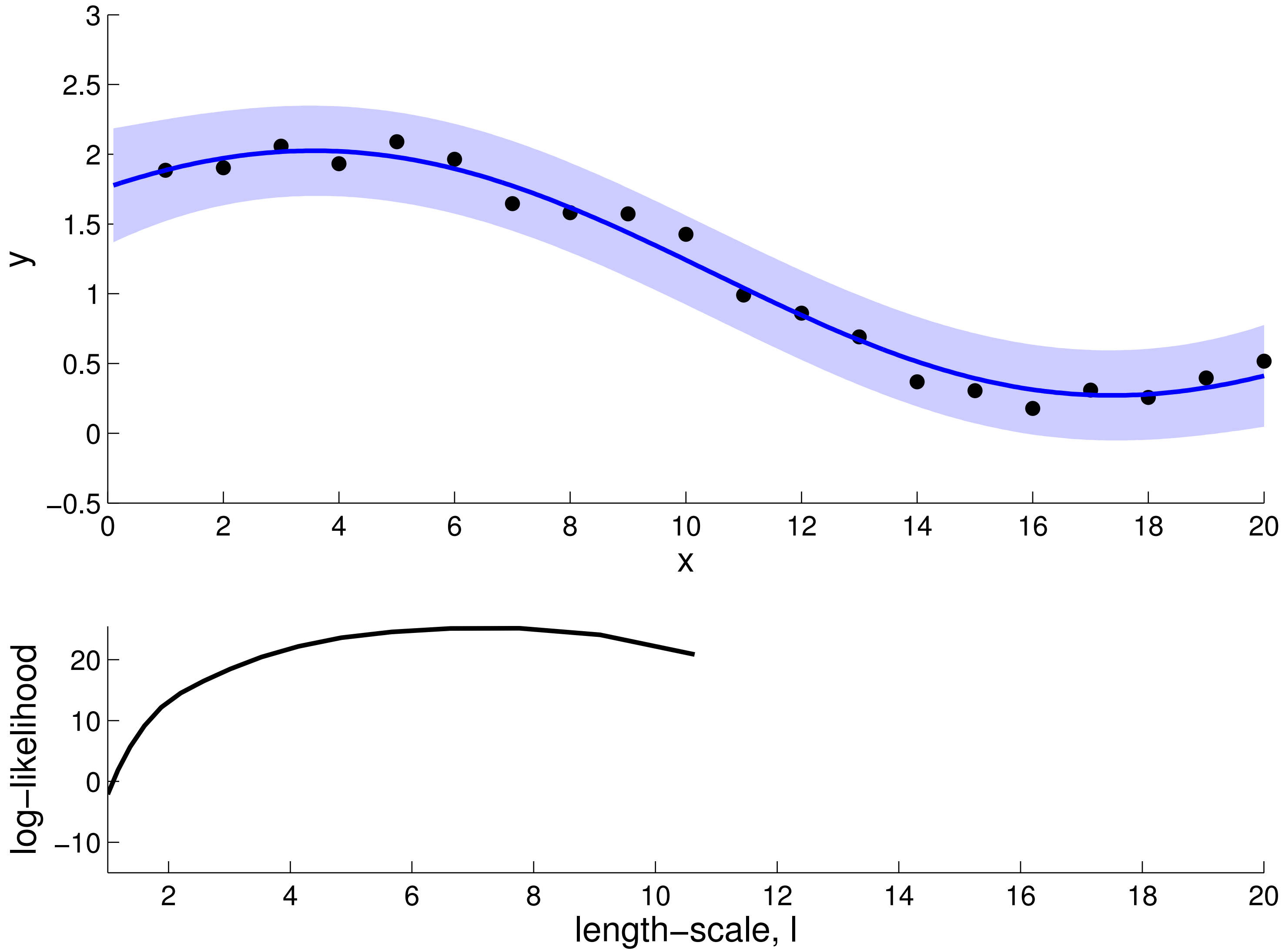
How do we choose the hyper-parameters?



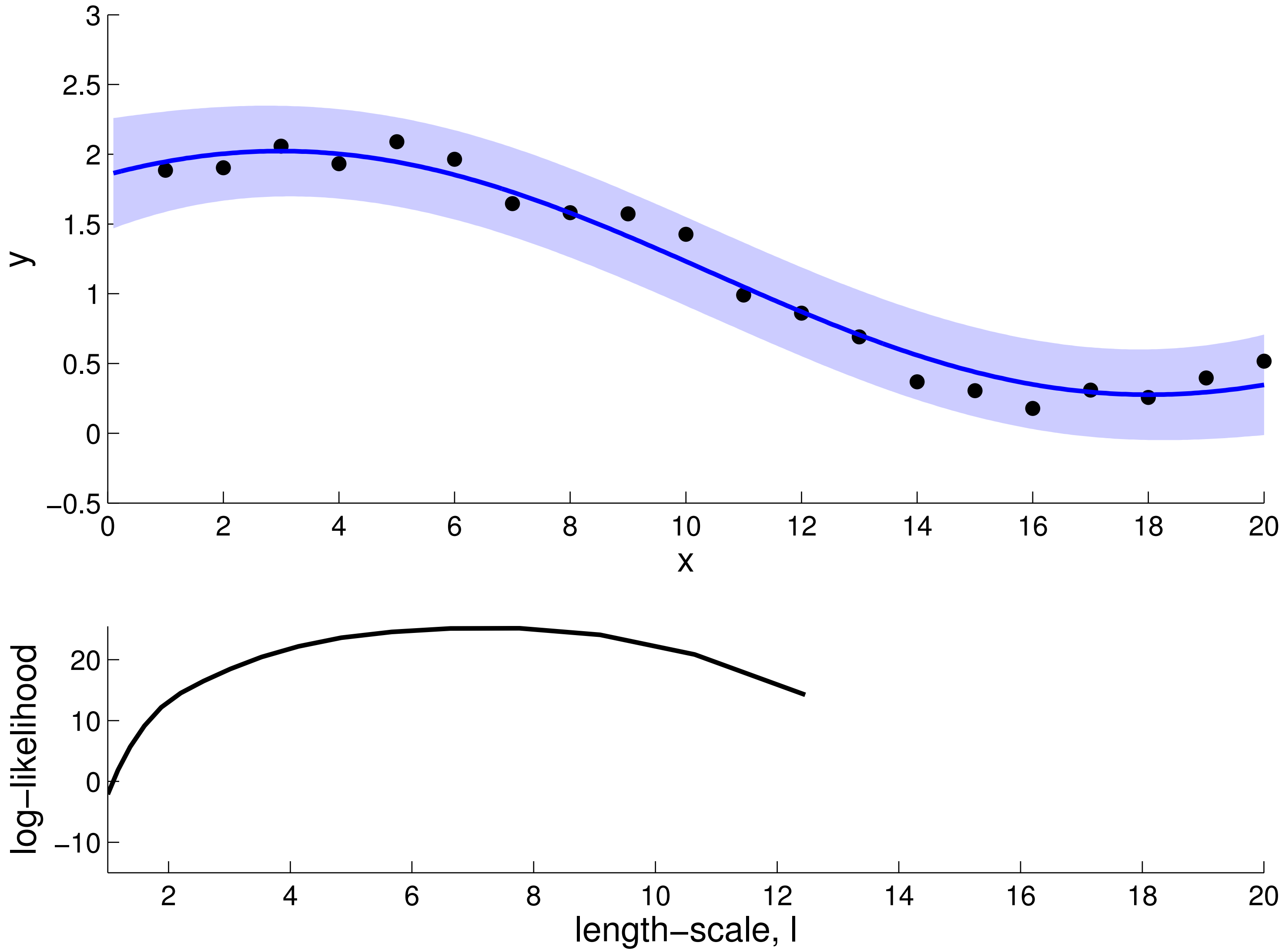
How do we choose the hyper-parameters?



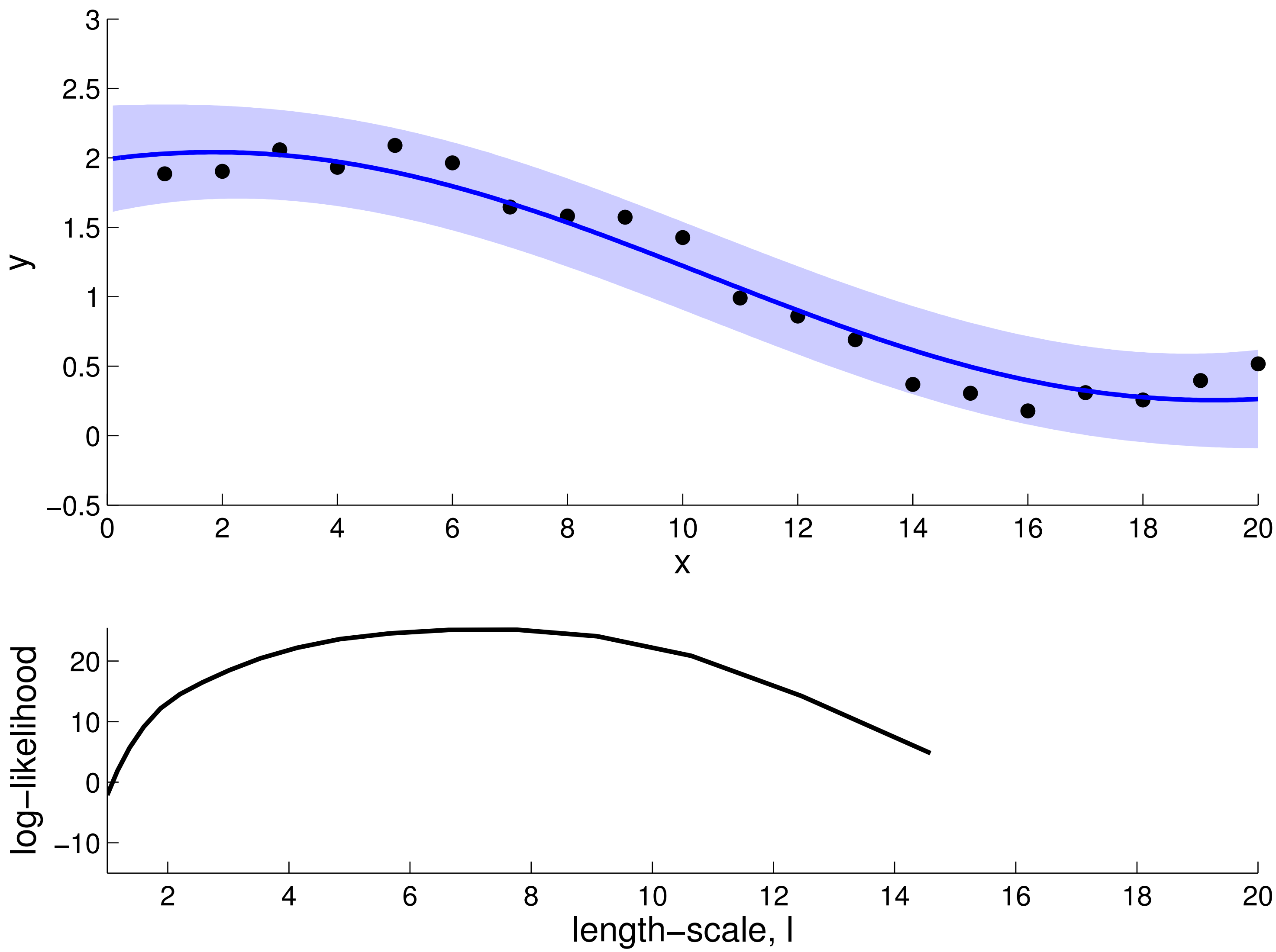
How do we choose the hyper-parameters?



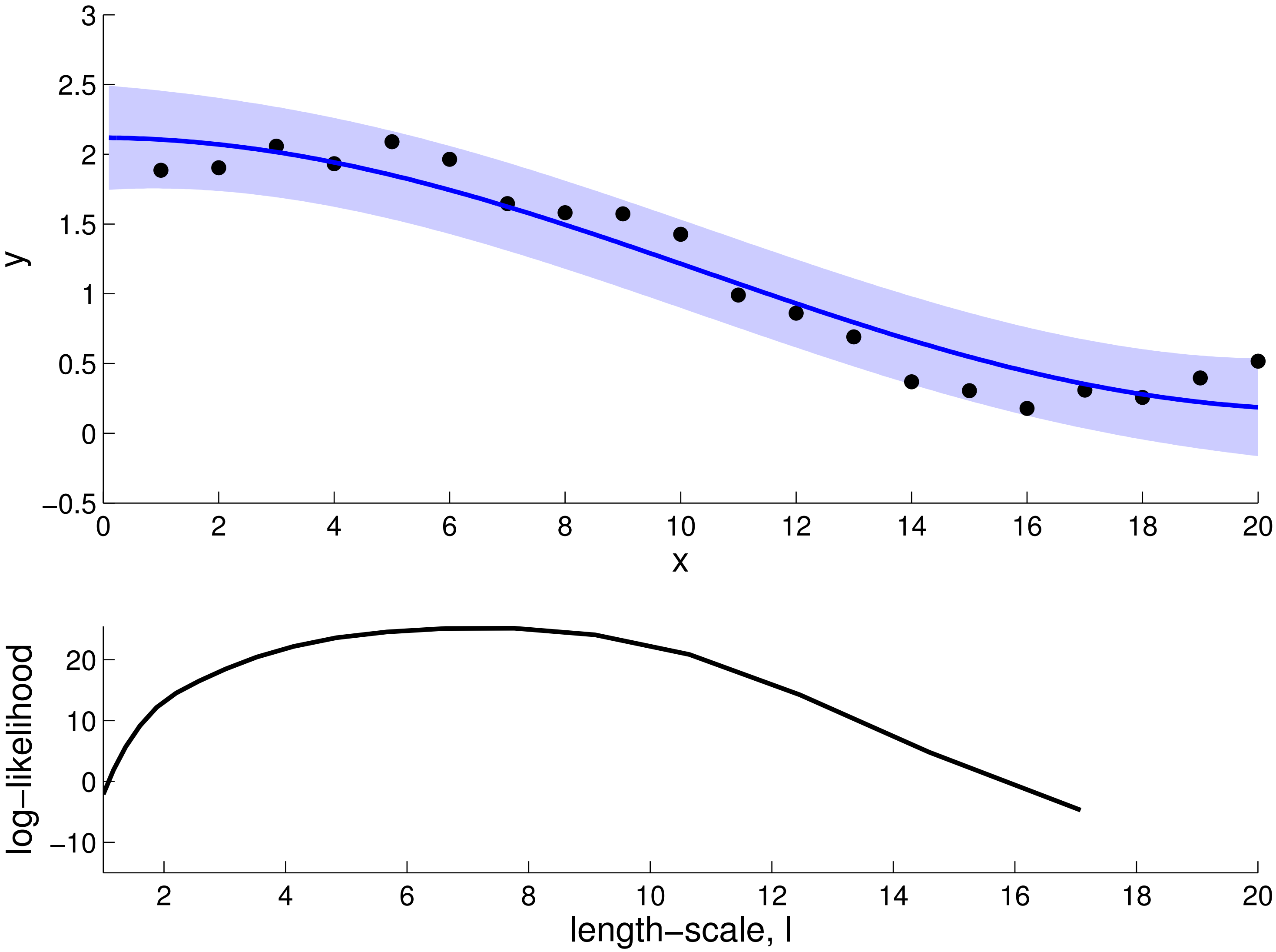
How do we choose the hyper-parameters?



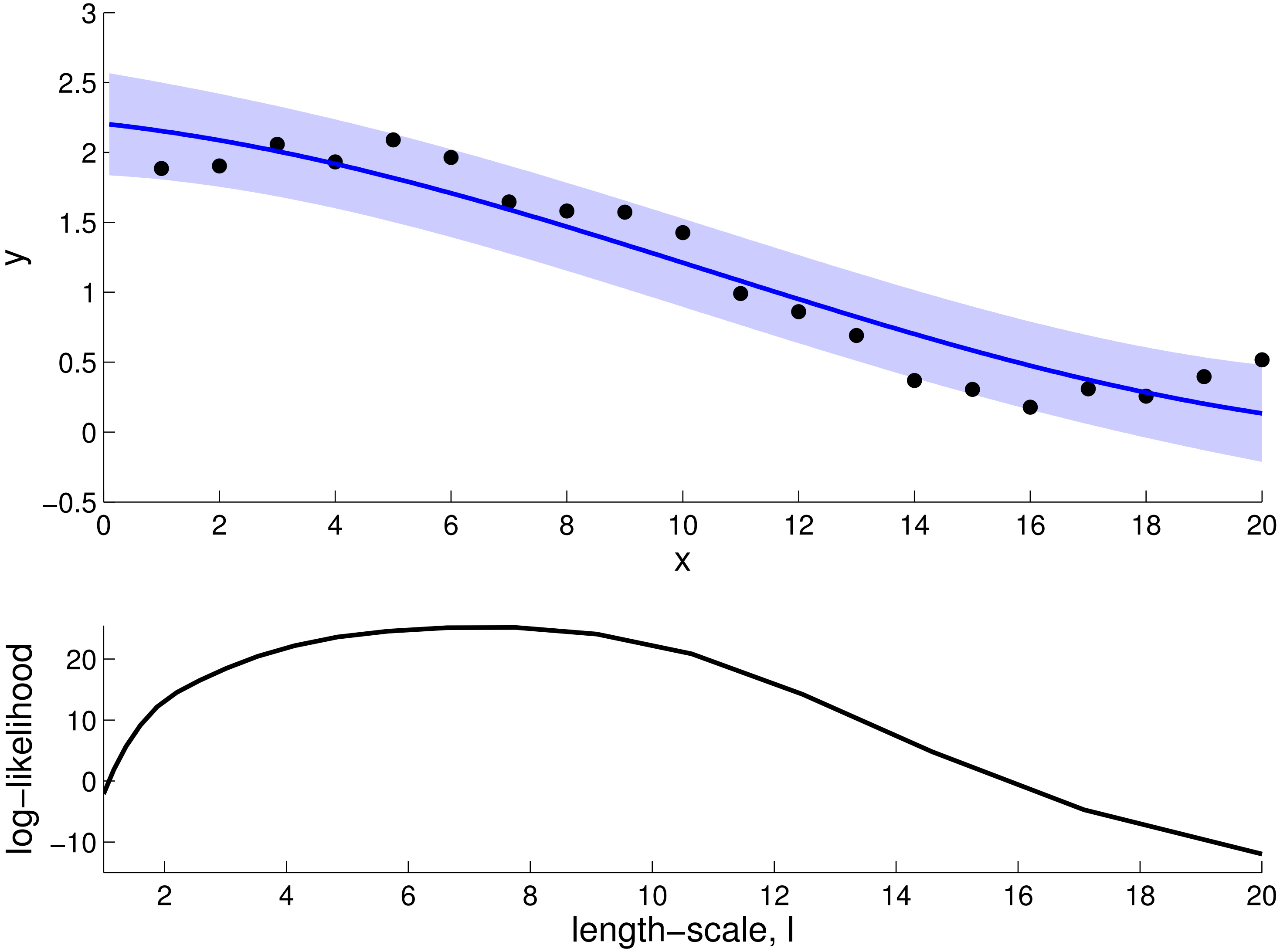
How do we choose the hyper-parameters?



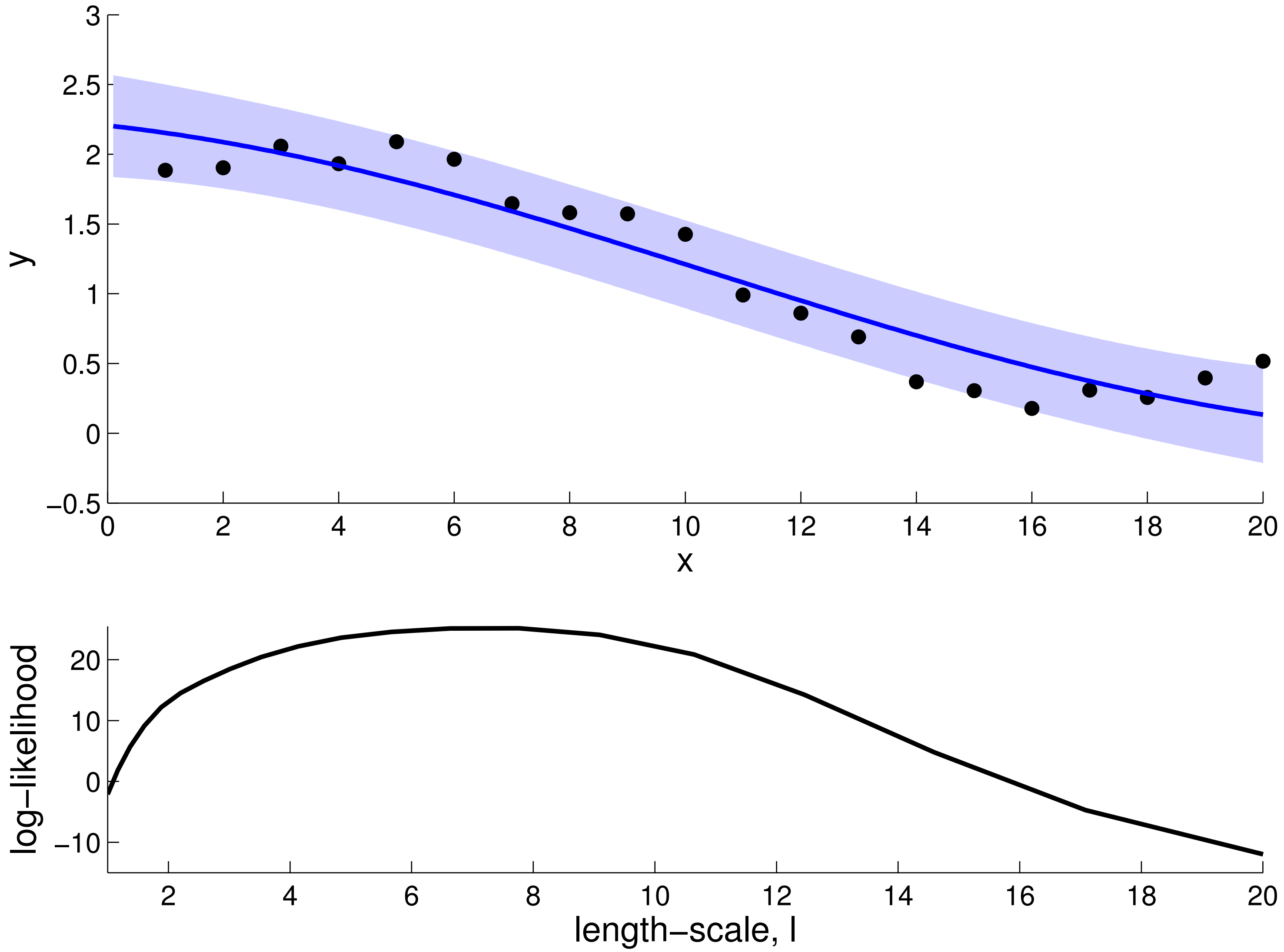
How do we choose the hyper-parameters?



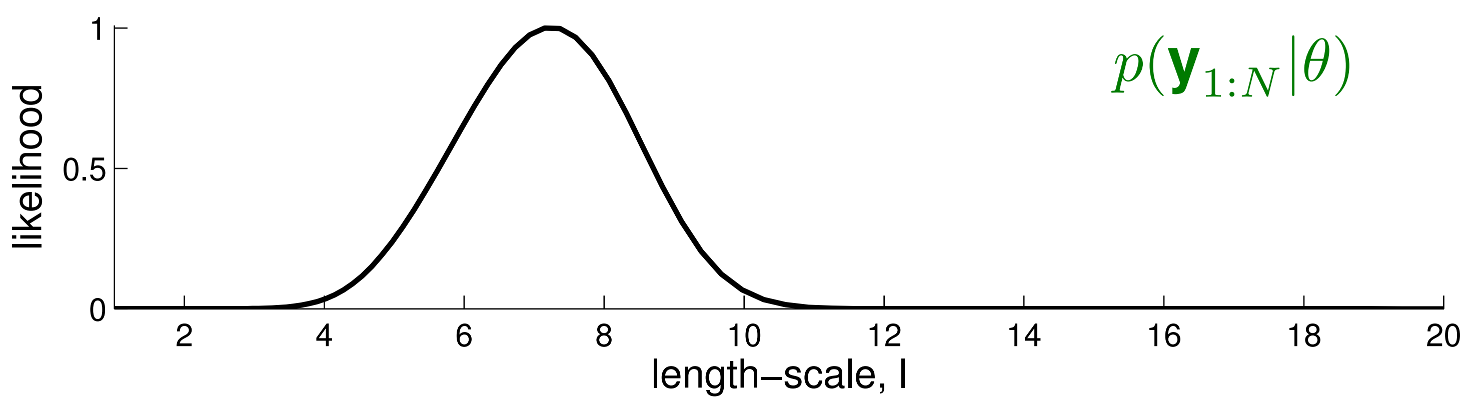
How do we choose the hyper-parameters?



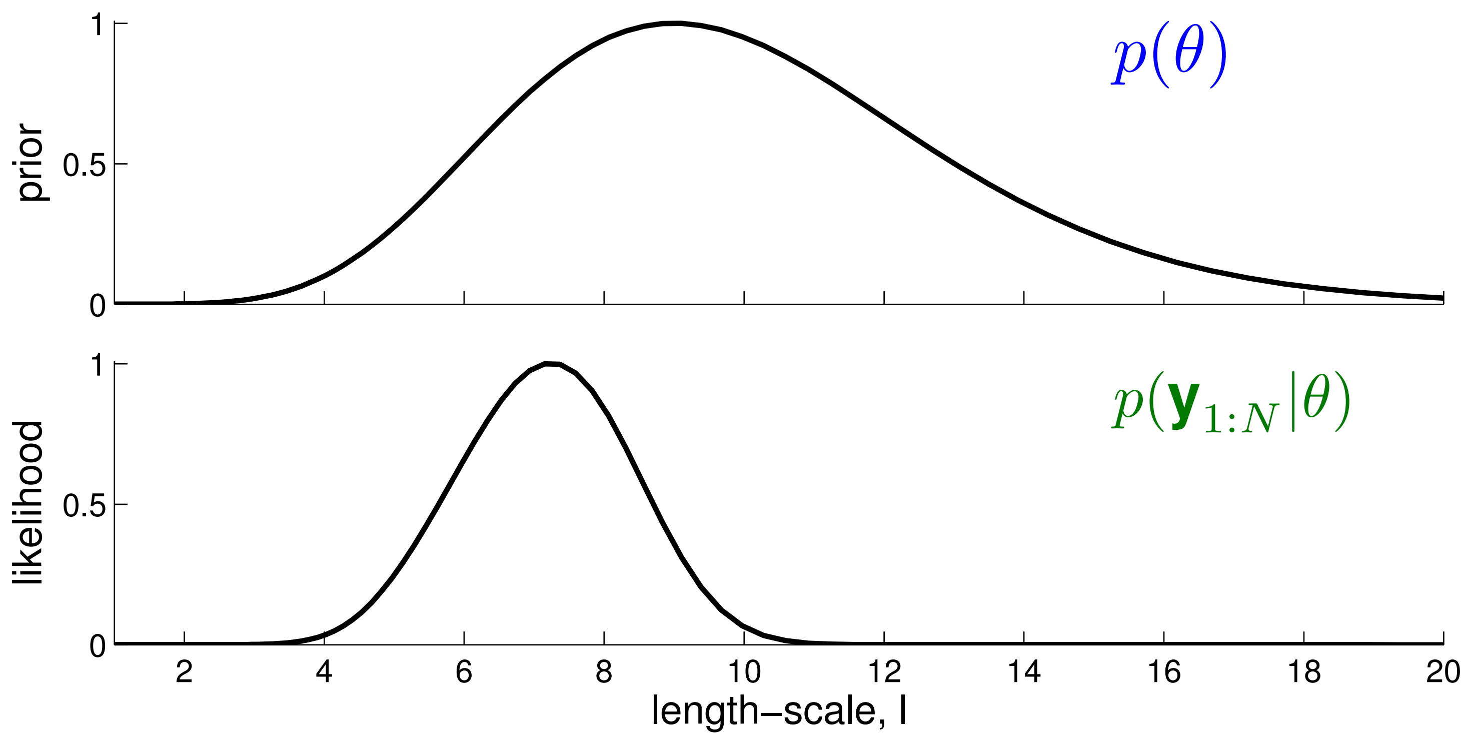
How do we choose the hyper-parameters?



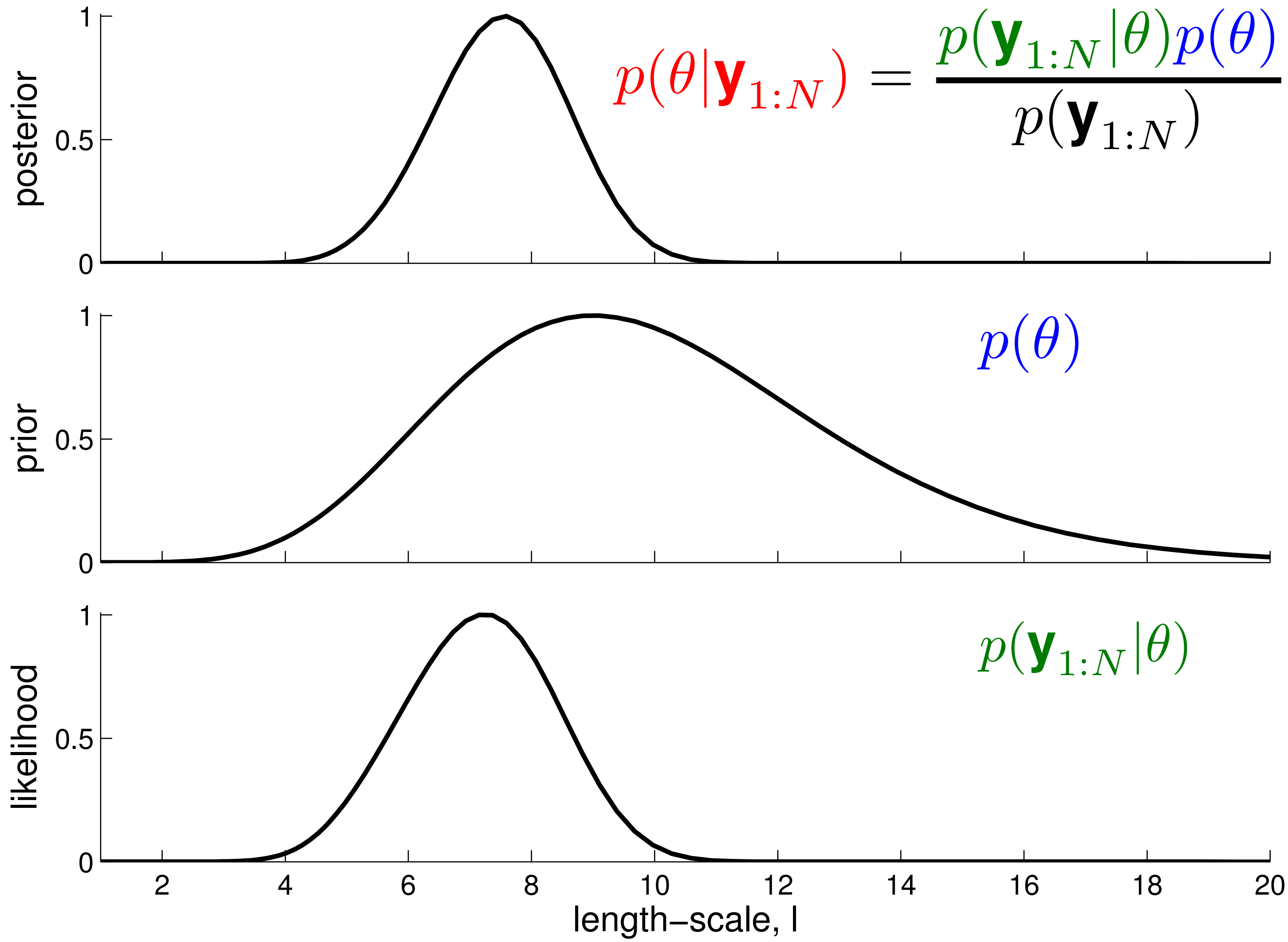
How do we choose the hyper-parameters?



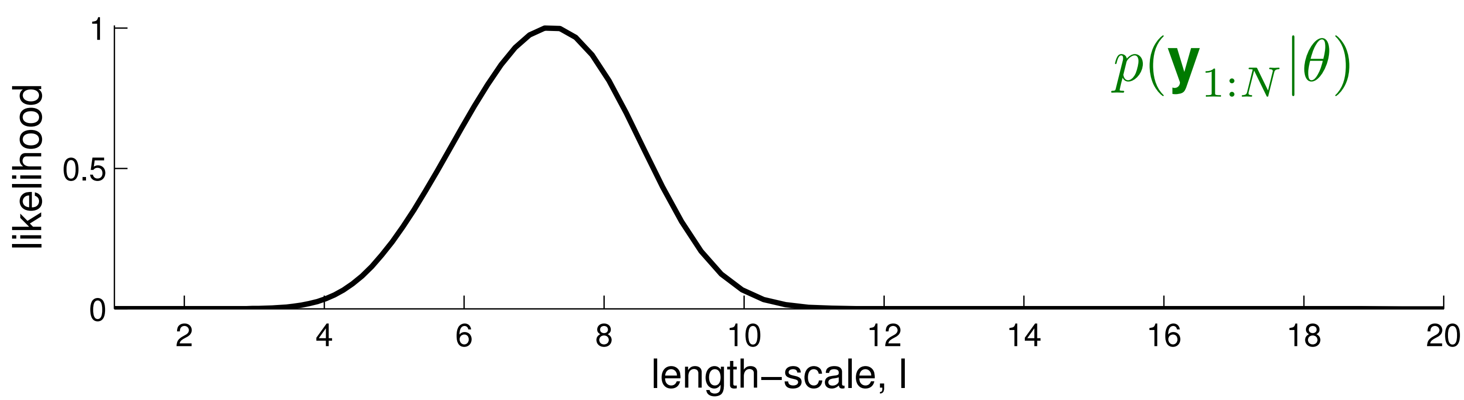
How do we choose the hyper-parameters?



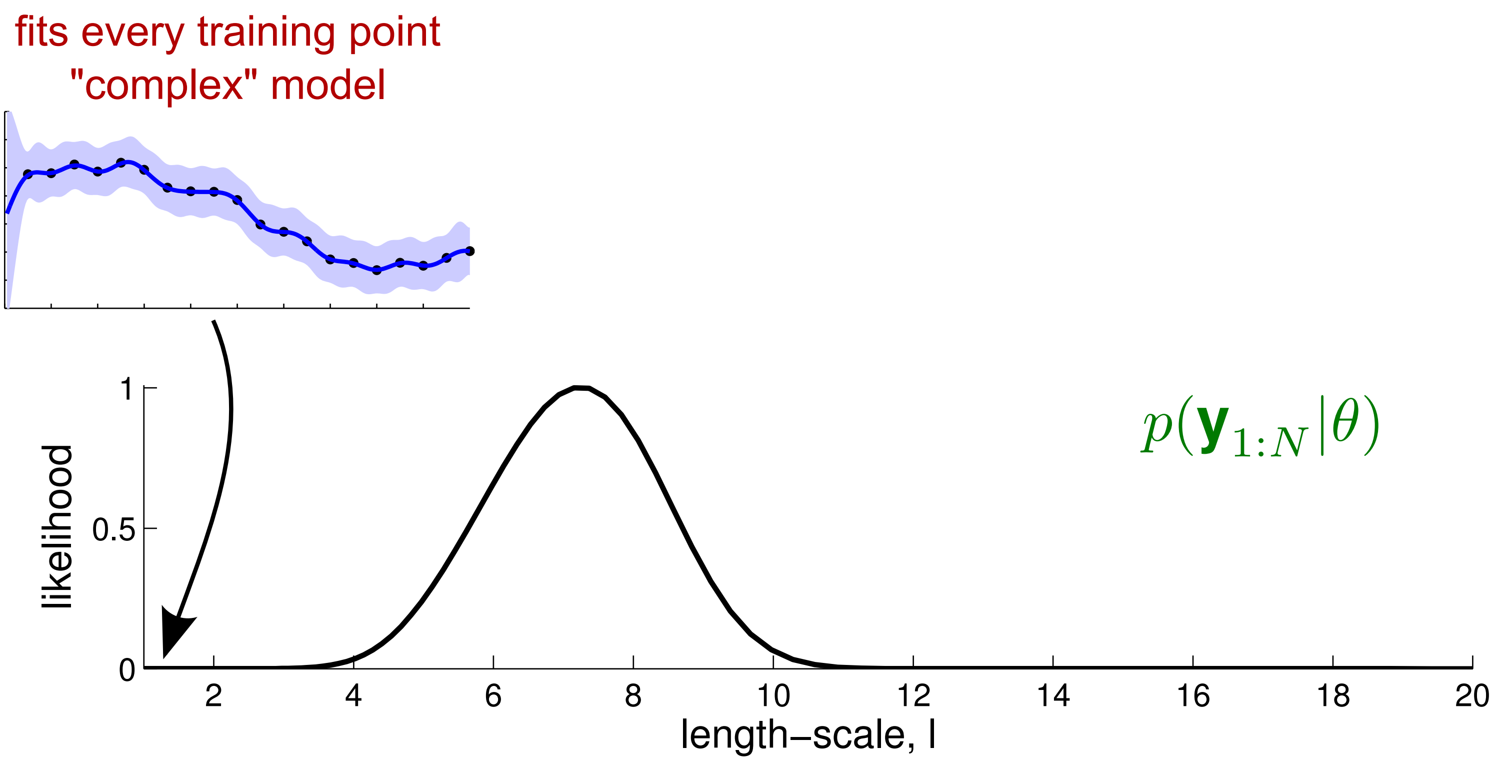
How do we choose the hyper-parameters?



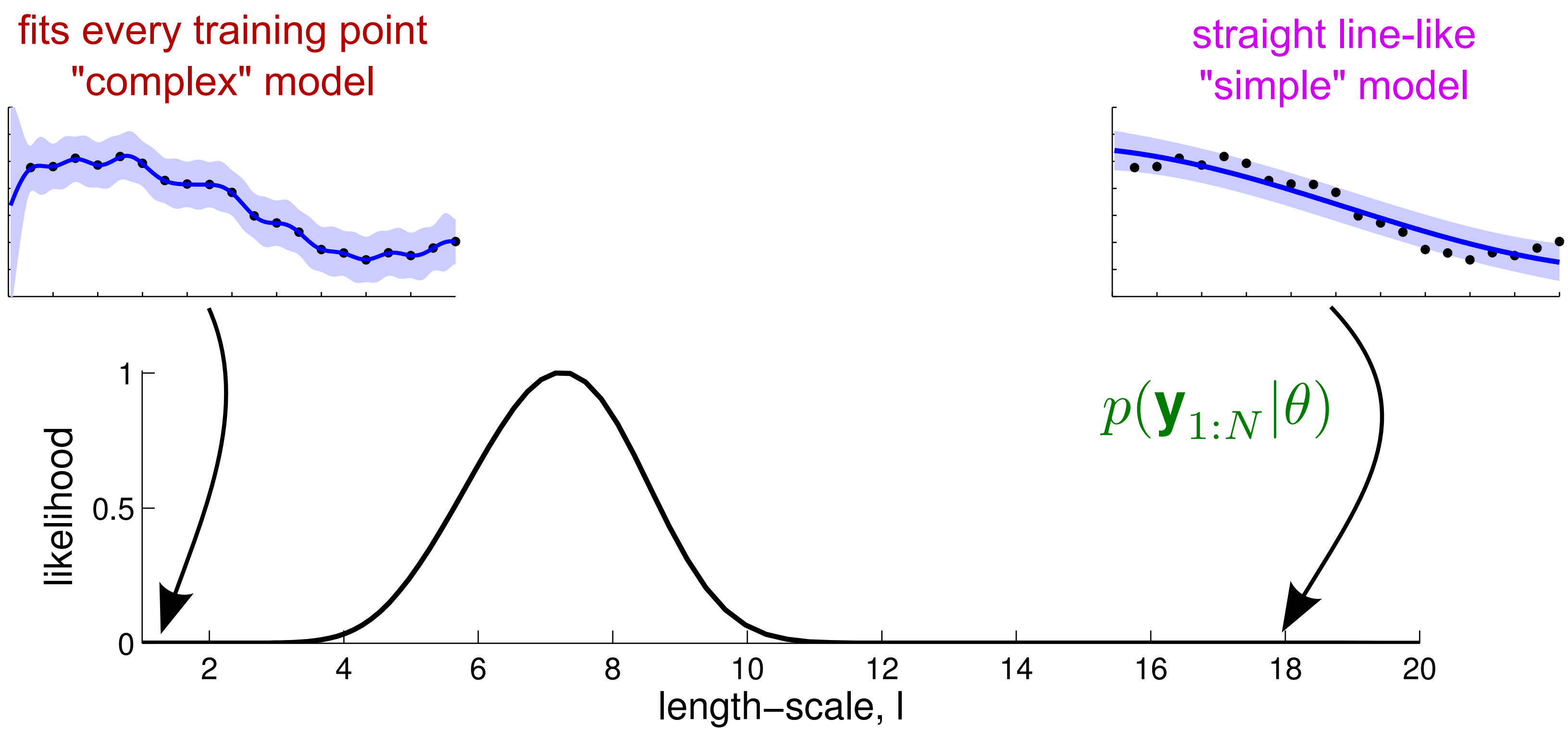
Why does Bayesian inference work?



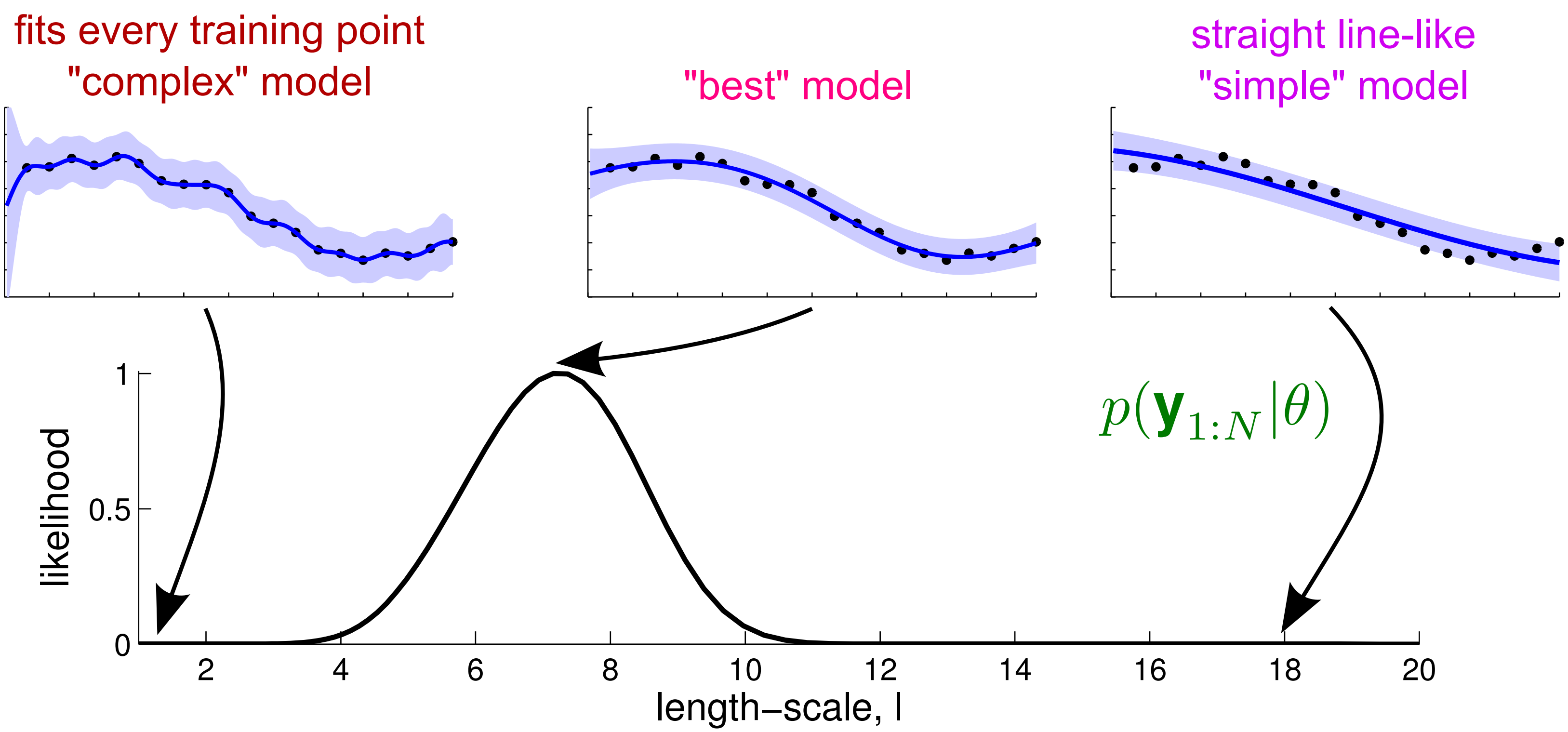
Why does Bayesian inference work?



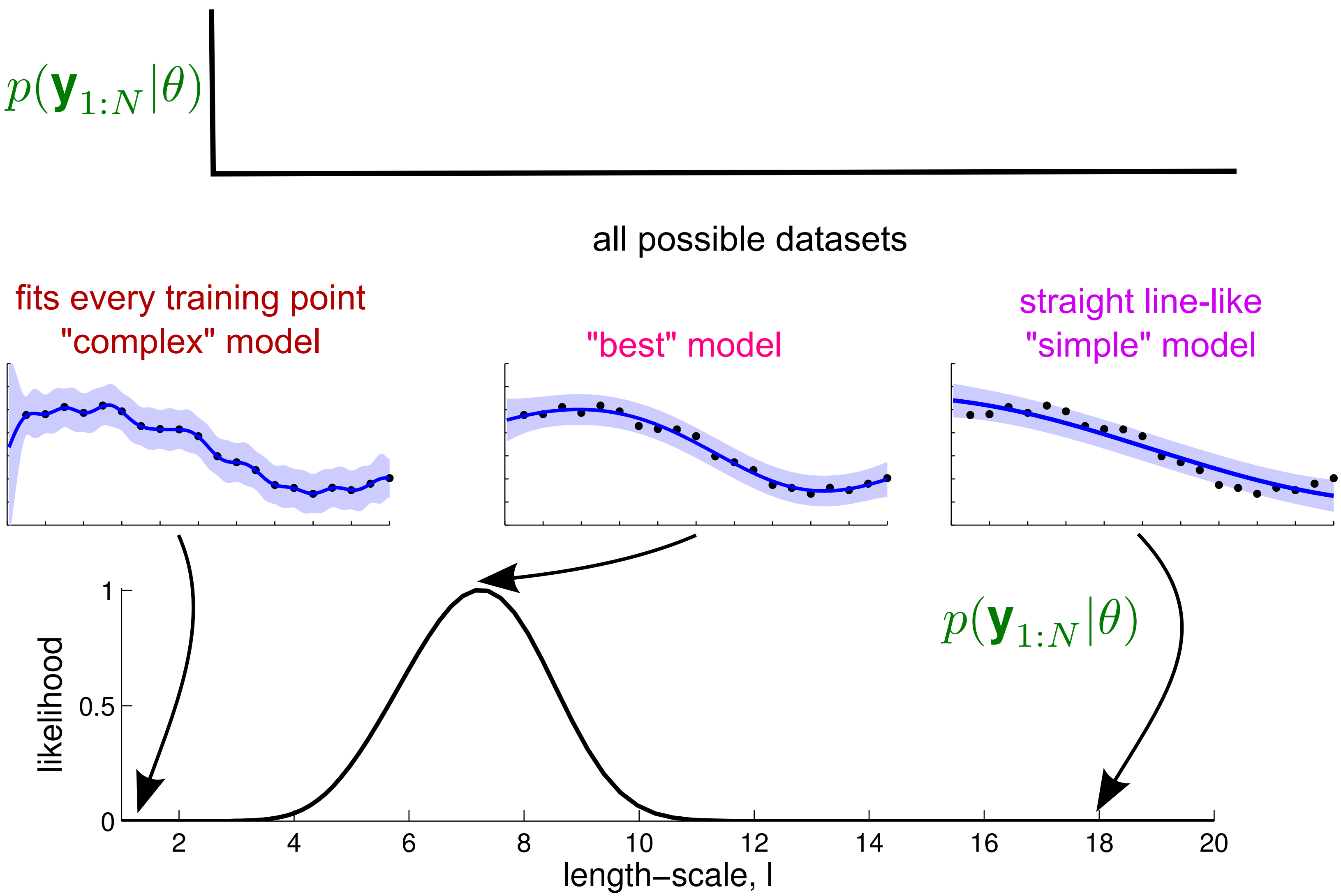
Why does Bayesian inference work?



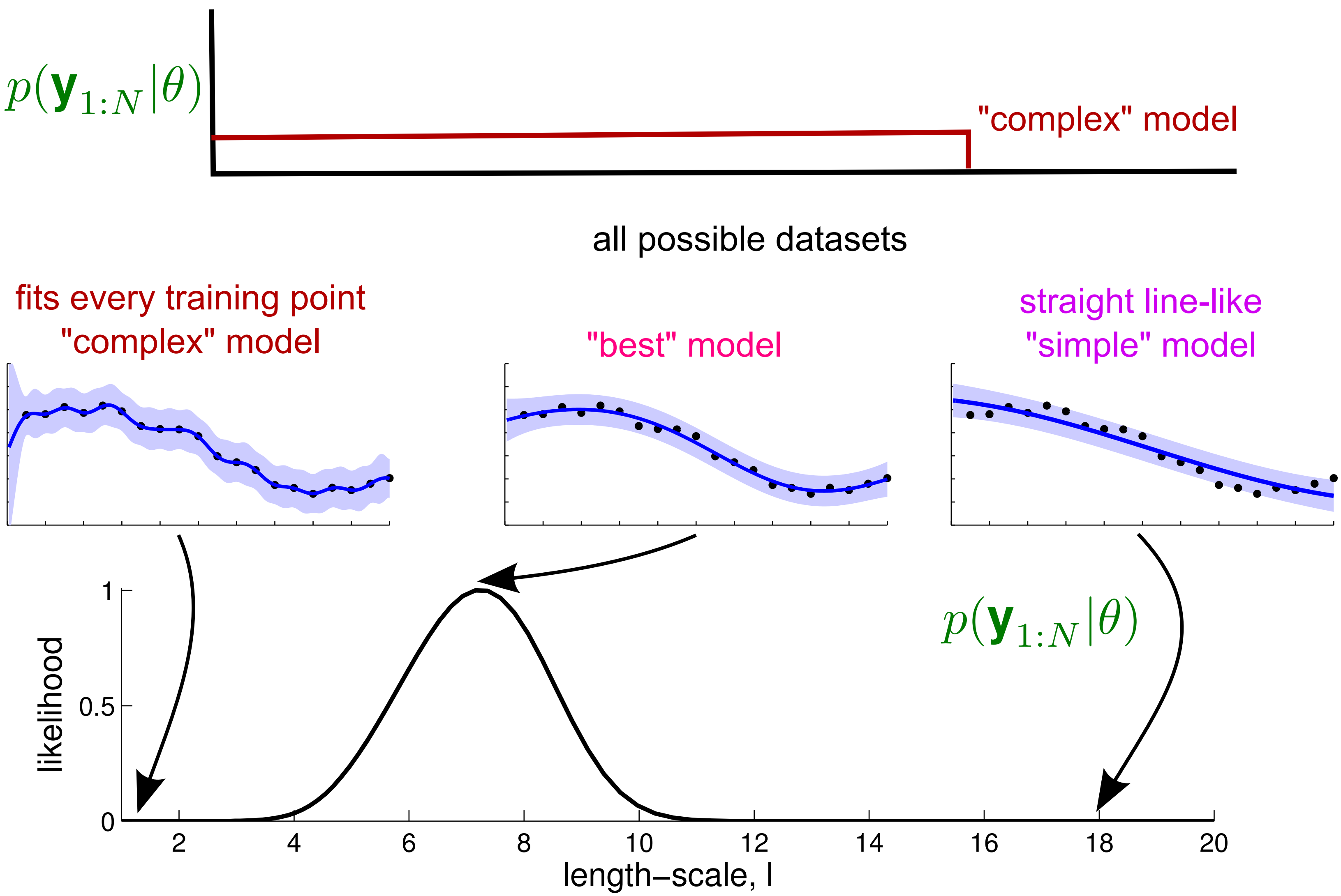
Why does Bayesian inference work?



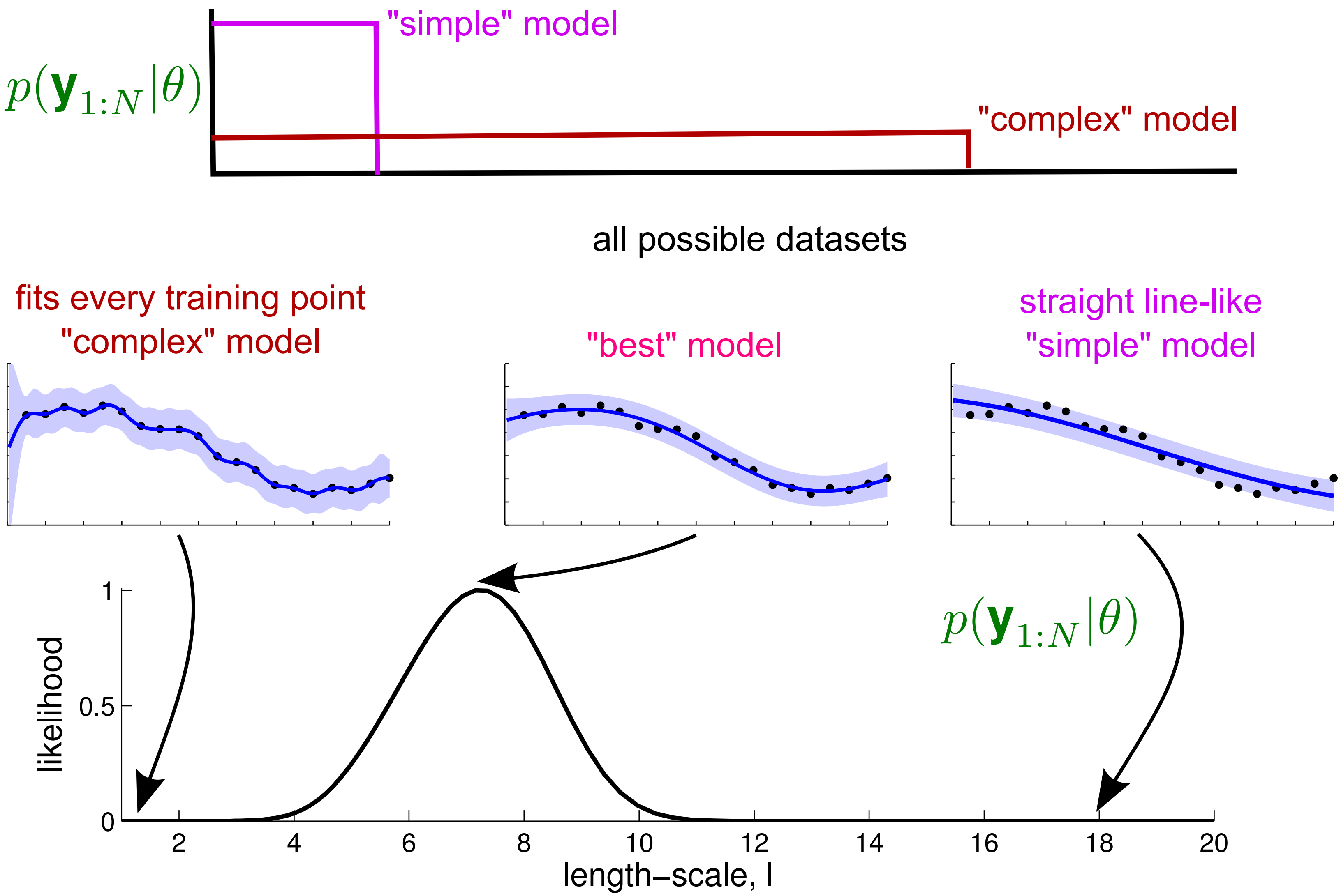
Why does Bayesian inference work?



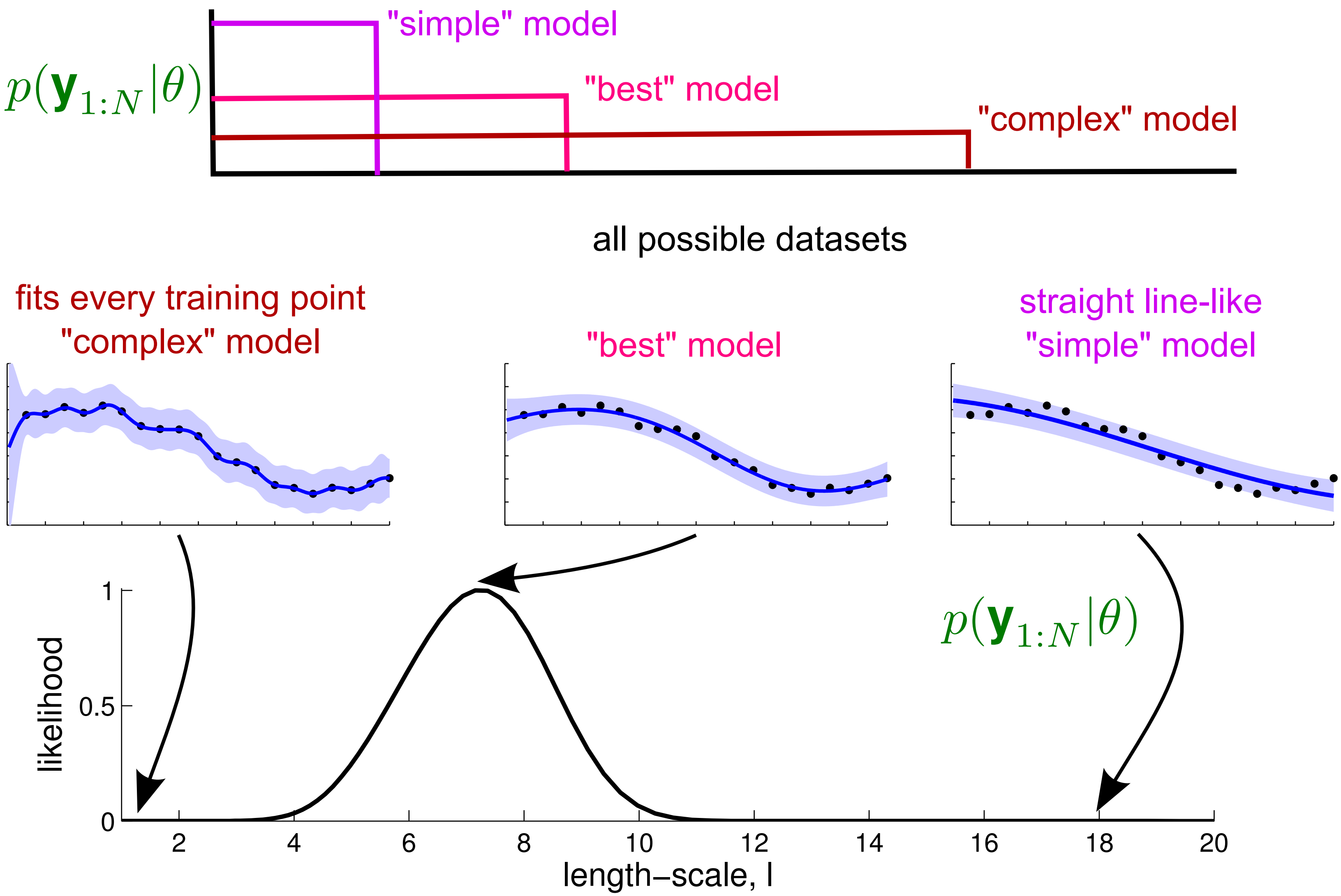
Why does Bayesian inference work?



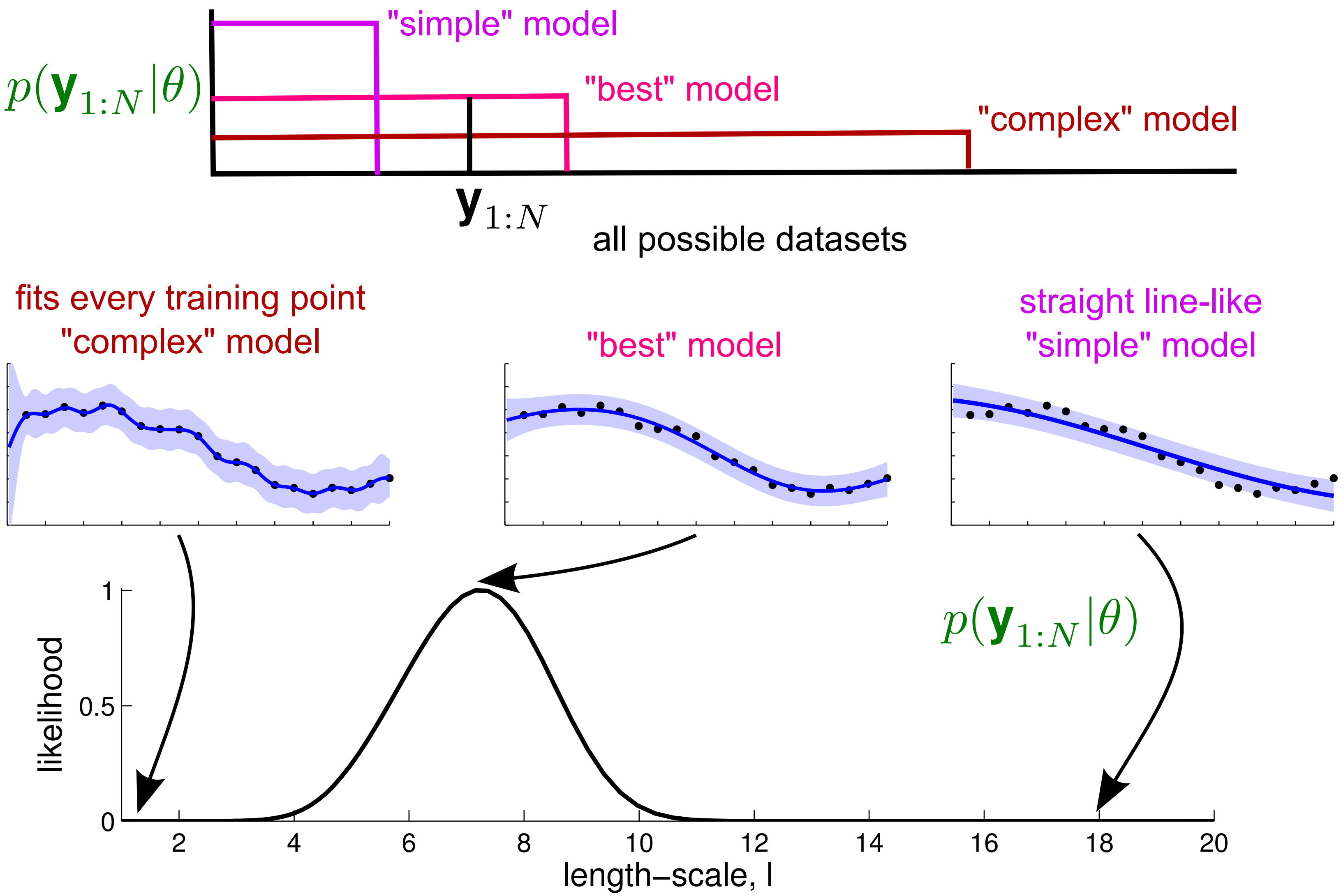
Why does Bayesian inference work?



Why does Bayesian inference work?



Why does Bayesian inference work? Occam's Razor.



How do we make predictions now?

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

- for every setting of the parameters...
- make a prediction for the testing points (as before) $p(\mathbf{y}'|\mathbf{y}_{1:N}, \theta, M)$
- weight the prediction by probability of θ under the posterior $p(\theta|\mathbf{y}_{1:N}, M)$
- sum up for each parameter setting

How do we make predictions now?

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

- for every setting of the parameters...
- make a prediction for the testing points (as before) $p(\mathbf{y}'|\mathbf{y}_{1:N}, \theta, M)$
- weight the prediction by probability of θ under the posterior $p(\theta|\mathbf{y}_{1:N}, M)$
- sum up for each parameter setting

Only need two rules for Bayesian computation: product and sum rules

$$p(\mathsf{x}, \mathsf{y}) = p(\mathsf{x}|\mathsf{y})p(\mathsf{y}) = p(\mathsf{y}|\mathsf{x})p(\mathsf{x}) \qquad\qquad p(\mathsf{x}) = \sum_{\mathsf{y}} p(\mathsf{x}, \mathsf{y})$$

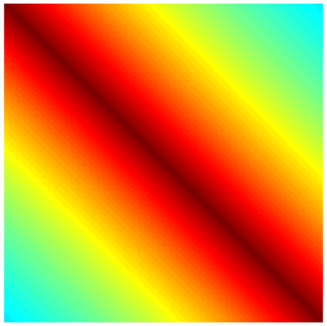
What effect does the form of the covariance function have?

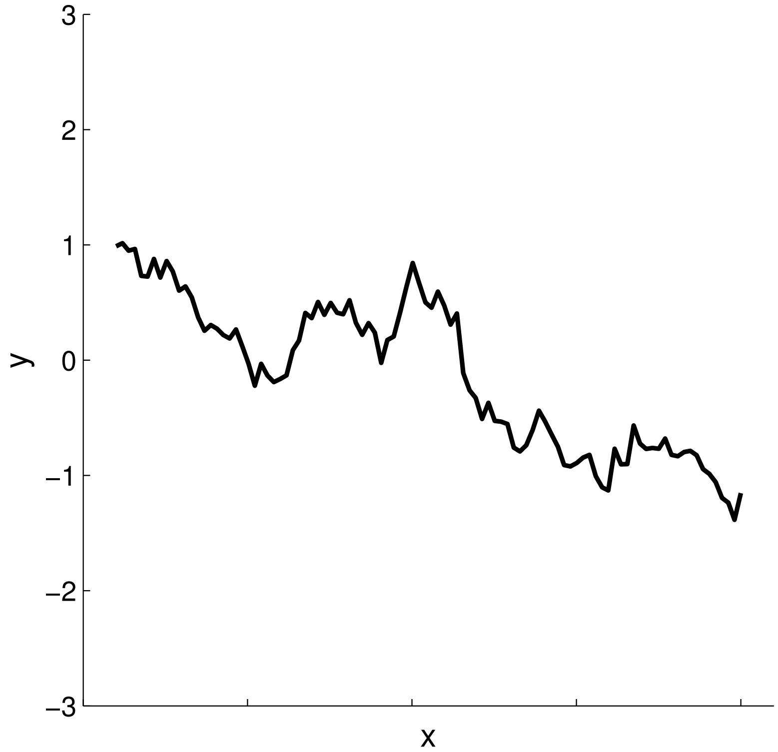
$$K(x_1, x_2) = \sigma^2 \exp\left(-\frac{1}{2l^2}|x_1 - x_2|\right)$$

Laplacian covariance function

Browninan motion

Ornstein-Uhlenbeck

$\Sigma =$ 



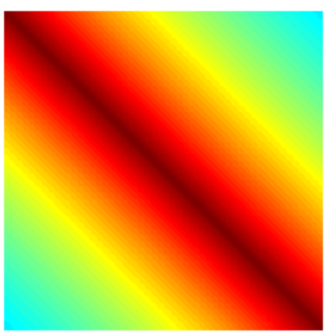
What effect does the form of the covariance function have?

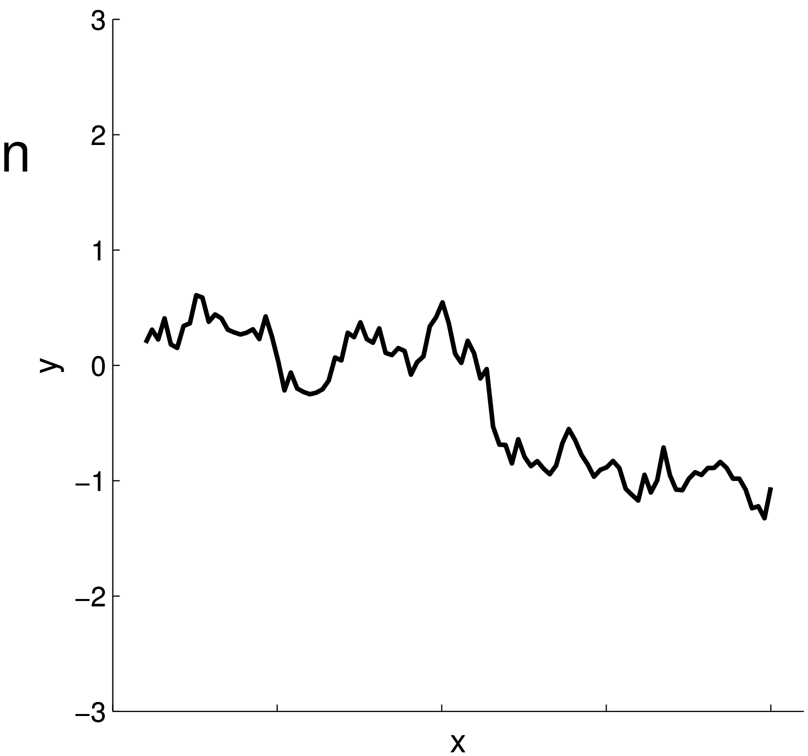
$$K(x_1, x_2) = \sigma^2 \exp\left(-\frac{1}{2l^2}|x_1 - x_2|\right)$$

Laplacian covariance function

Browninan motion

Ornstein-Uhlenbeck

$\Sigma =$ 



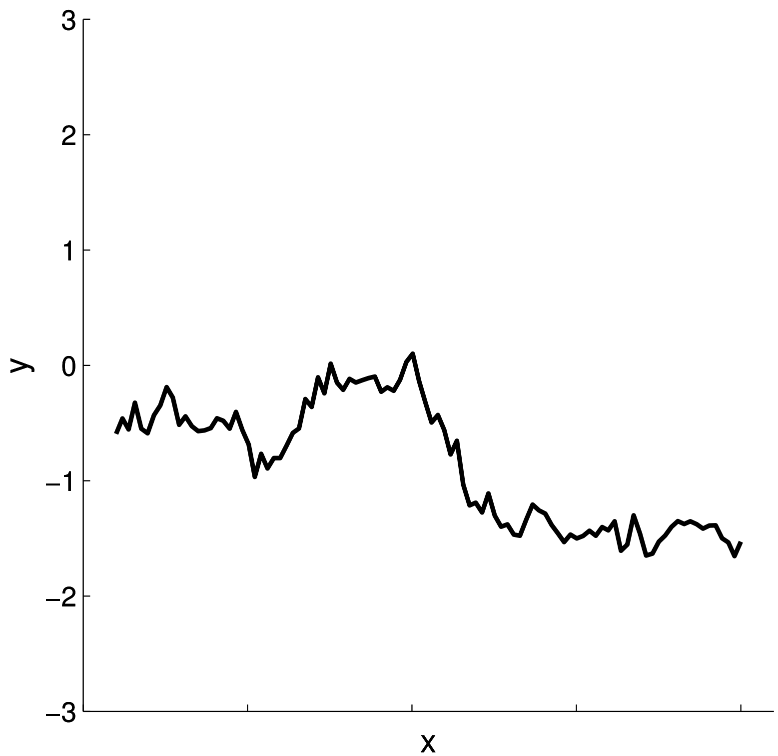
What effect does the form of the covariance function have?

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

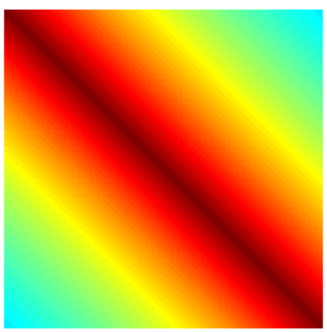
Laplacian covariance function

Browninan motion

Ornstein-Uhlenbeck



$\Sigma =$



What effect does the form of the covariance function have?

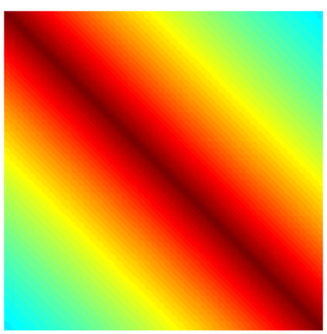
$$K(x_1, x_2) = \sigma^2 \exp\left(-\frac{1}{2l^2}|x_1 - x_2|\right)$$

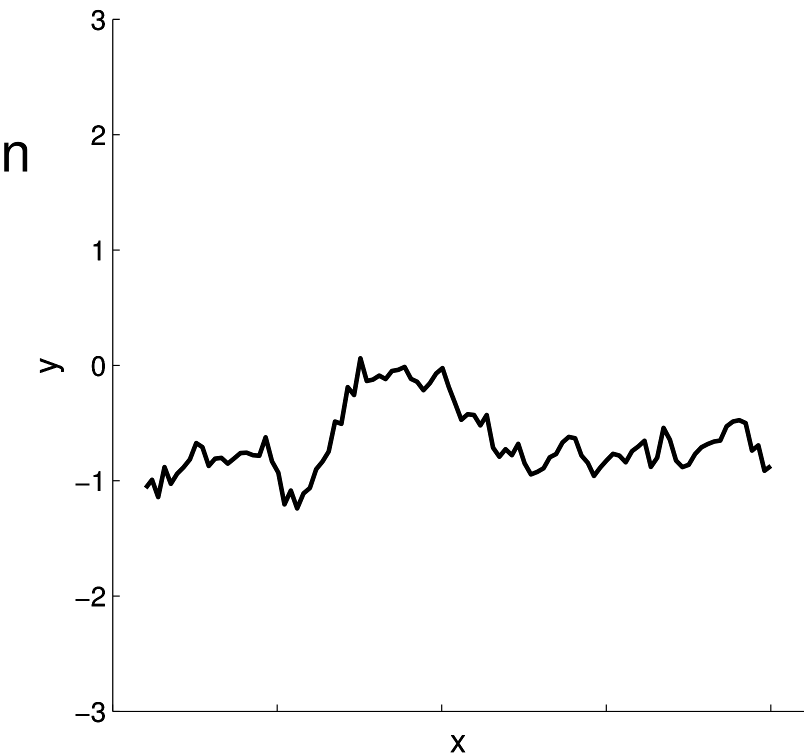
Laplacian covariance function

Browninan motion

Ornstein-Uhlenbeck

$\Sigma =$





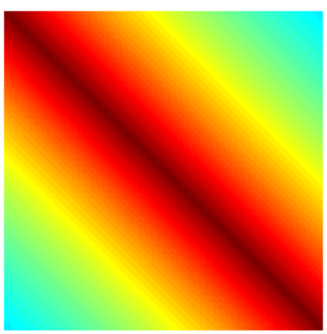
What effect does the form of the covariance function have?

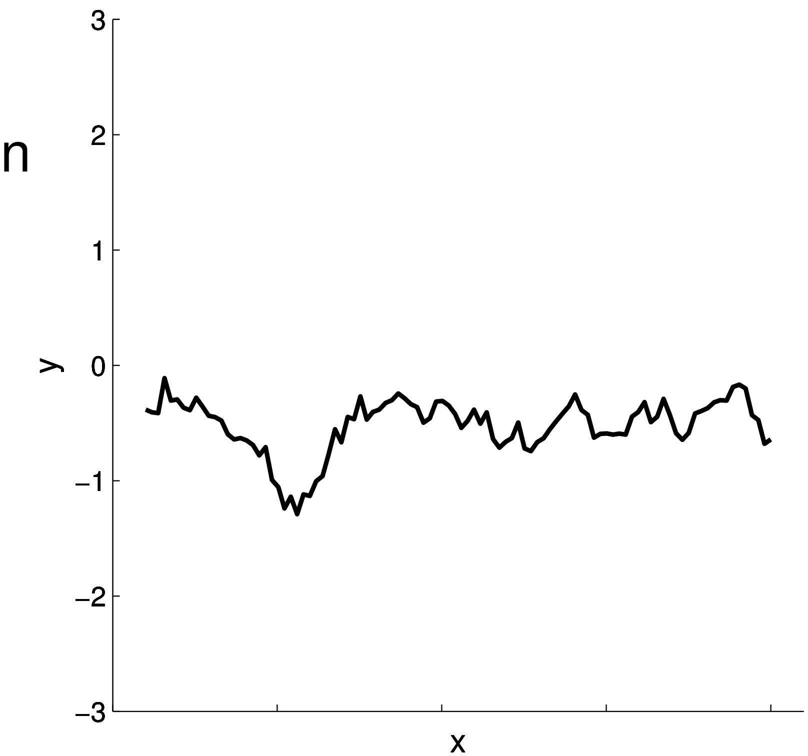
$$K(x_1, x_2) = \sigma^2 \exp\left(-\frac{1}{2l^2}|x_1 - x_2|\right)$$

Laplacian covariance function

Browninan motion

Ornstein-Uhlenbeck

$\Sigma =$ 



What effect does the form of the covariance function have?

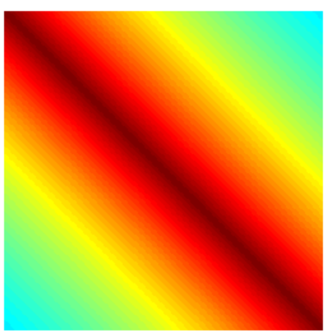
$$K(x_1, x_2) = \sigma^2 \exp\left(-\frac{1}{2l^2}|x_1 - x_2|\right)$$

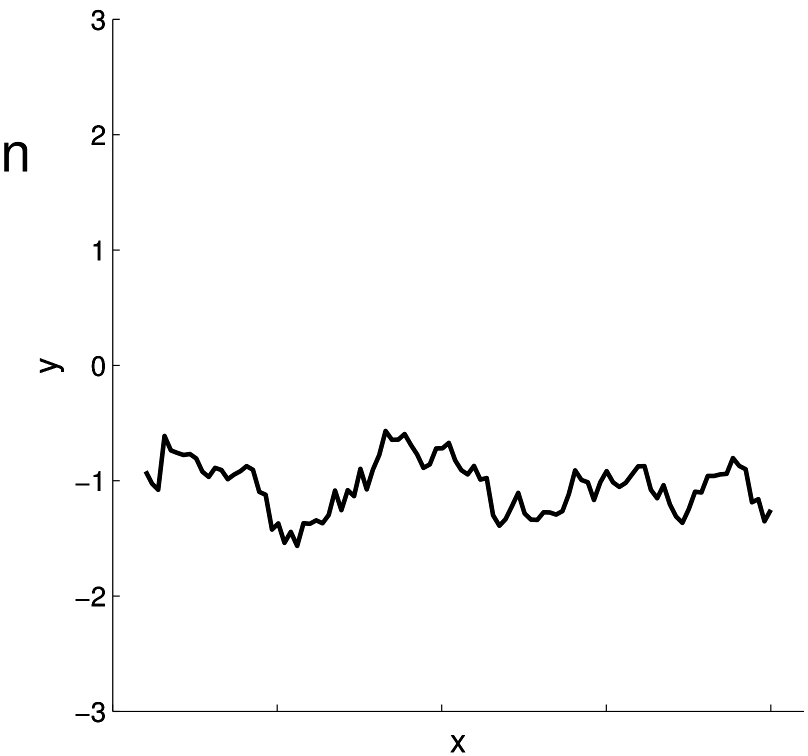
Laplacian covariance function

Browninan motion

Ornstein-Uhlenbeck

$\Sigma =$





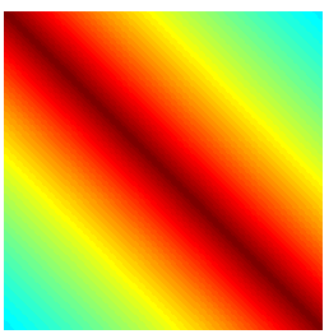
What effect does the form of the covariance function have?

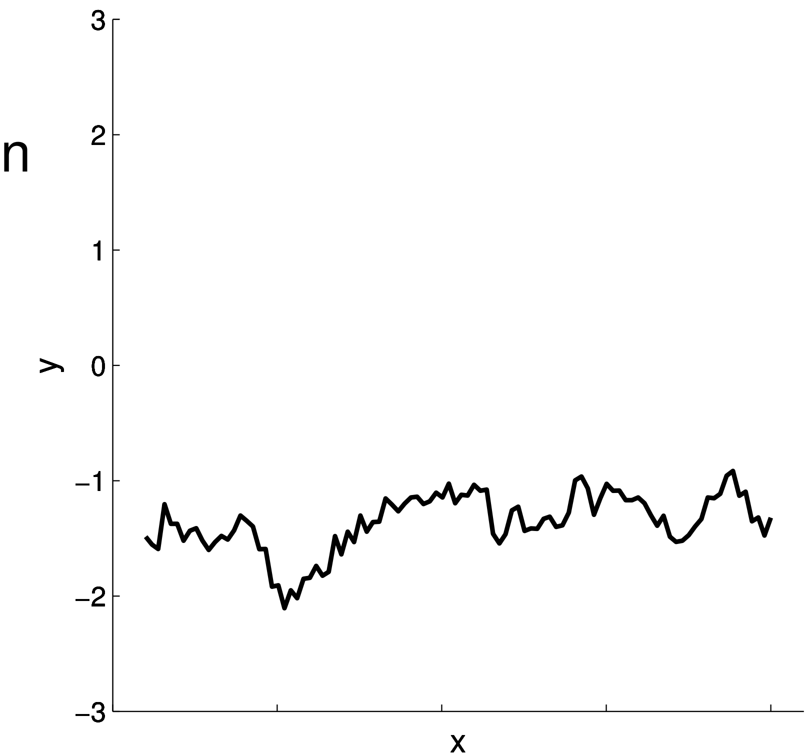
$$K(x_1, x_2) = \sigma^2 \exp\left(-\frac{1}{2l^2}|x_1 - x_2|\right)$$

Laplacian covariance function

Browninan motion

Ornstein-Uhlenbeck

$\Sigma =$ 



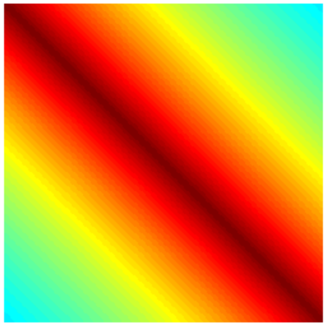
What effect does the form of the covariance function have?

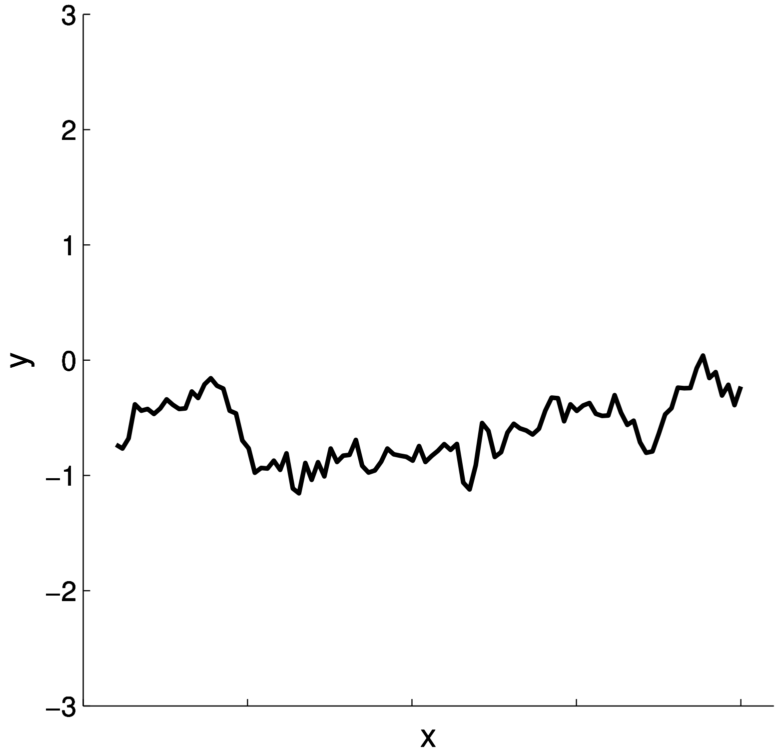
$$K(x_1, x_2) = \sigma^2 \exp\left(-\frac{1}{2l^2}|x_1 - x_2|\right)$$

Laplacian covariance function

Browninan motion

Ornstein-Uhlenbeck

$\Sigma =$ 



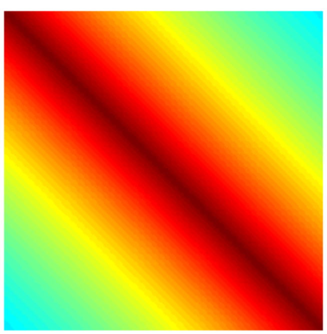
What effect does the form of the covariance function have?

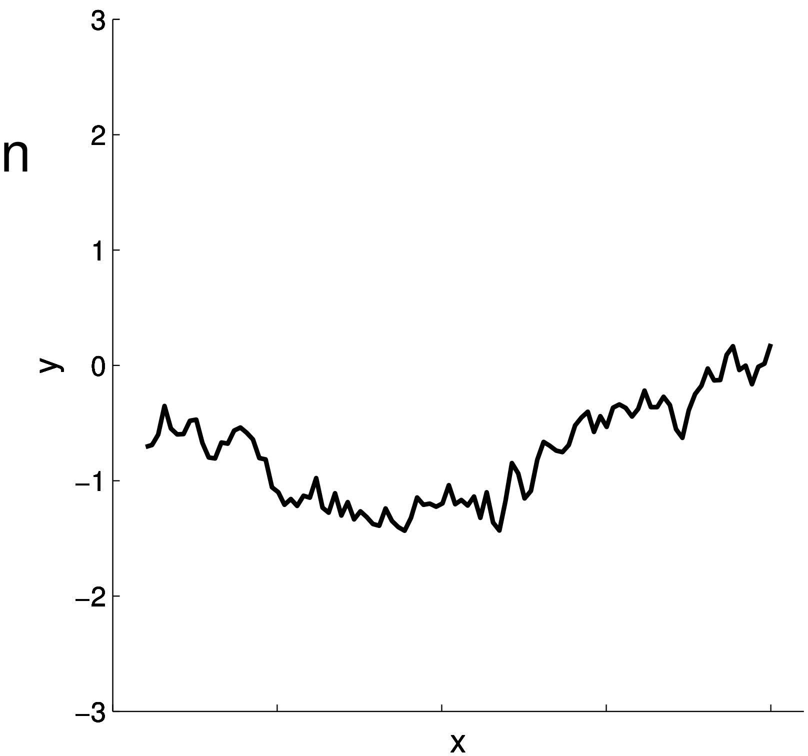
$$K(x_1, x_2) = \sigma^2 \exp\left(-\frac{1}{2l^2}|x_1 - x_2|\right)$$

Laplacian covariance function

Browninan motion

Ornstein-Uhlenbeck

$\Sigma =$ 



What effect does the form of the covariance function have?

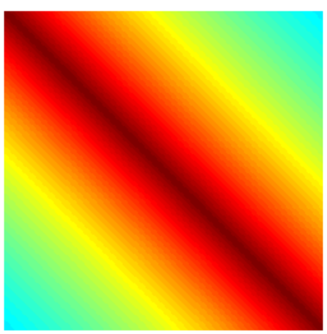
$$K(x_1, x_2) = \sigma^2 \exp\left(-\frac{1}{2l^2}|x_1 - x_2|\right)$$

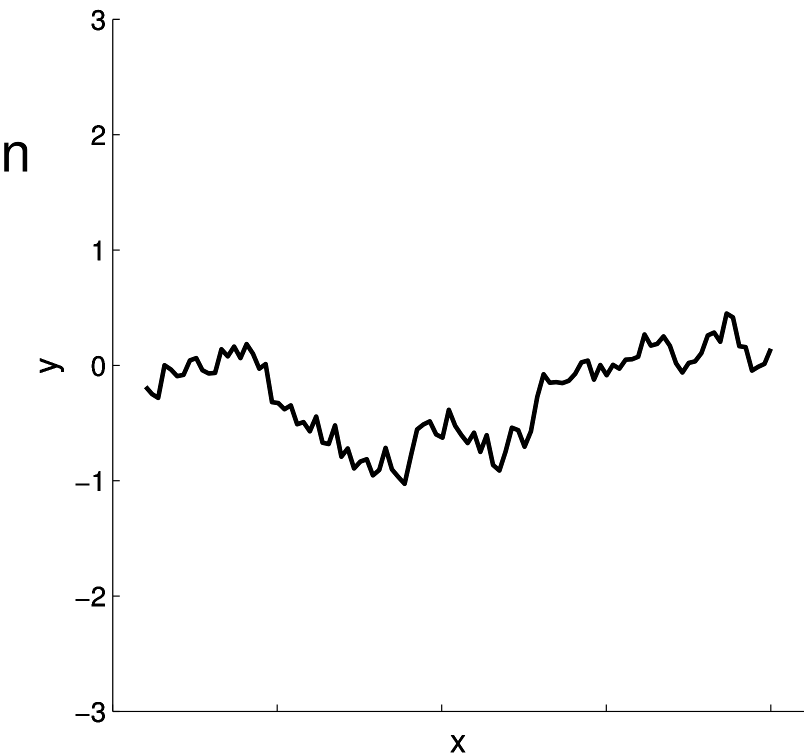
Laplacian covariance function

Browninan motion

Ornstein-Uhlenbeck

$\Sigma =$





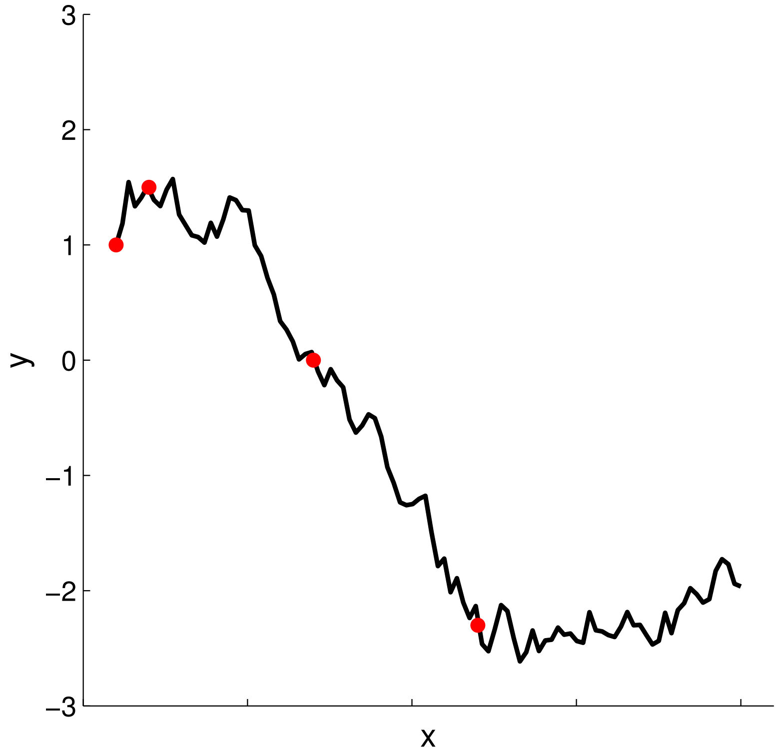
What effect does the form of the covariance function have?

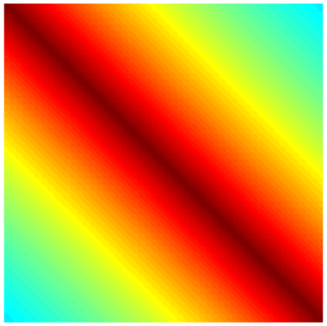
$$\mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) = \sigma^2 \exp\left(-\frac{1}{2\mathsf{l}^2}|\mathsf{x}_1 - \mathsf{x}_2|\right)$$

Laplacian covariance function

Browninan motion

Ornstein-Uhlenbeck



$\Sigma =$ 

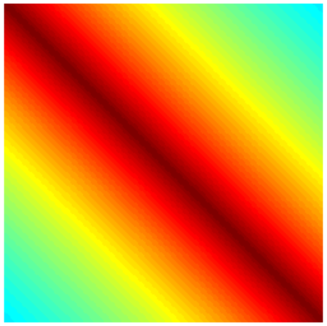
What effect does the form of the covariance function have?

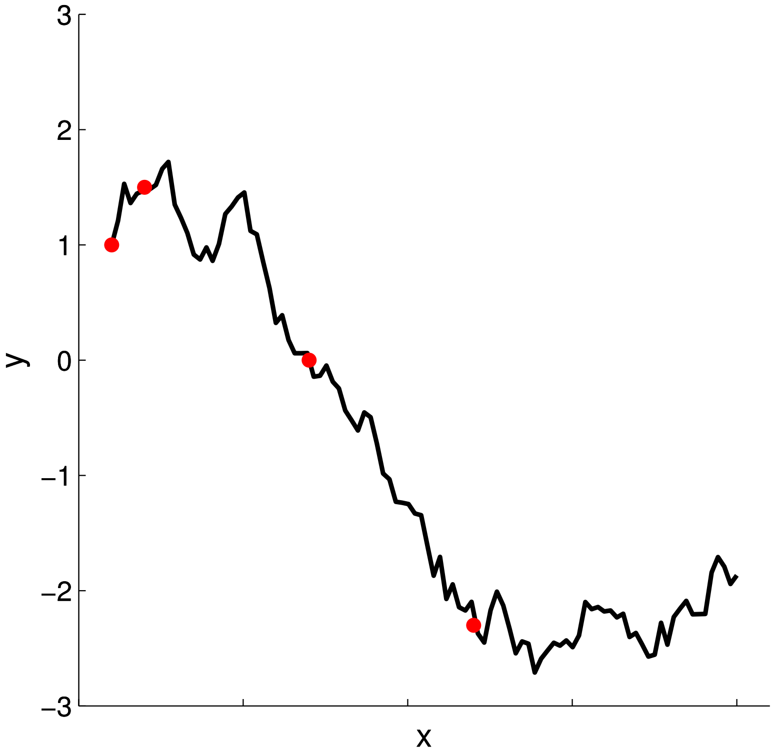
$$K(x_1, x_2) = \sigma^2 \exp\left(-\frac{1}{2l^2}|x_1 - x_2|\right)$$

Laplacian covariance function

Browninan motion

Ornstein-Uhlenbeck

$\Sigma =$ 



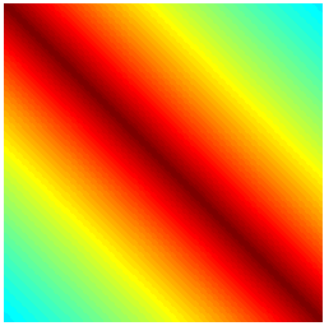
What effect does the form of the covariance function have?

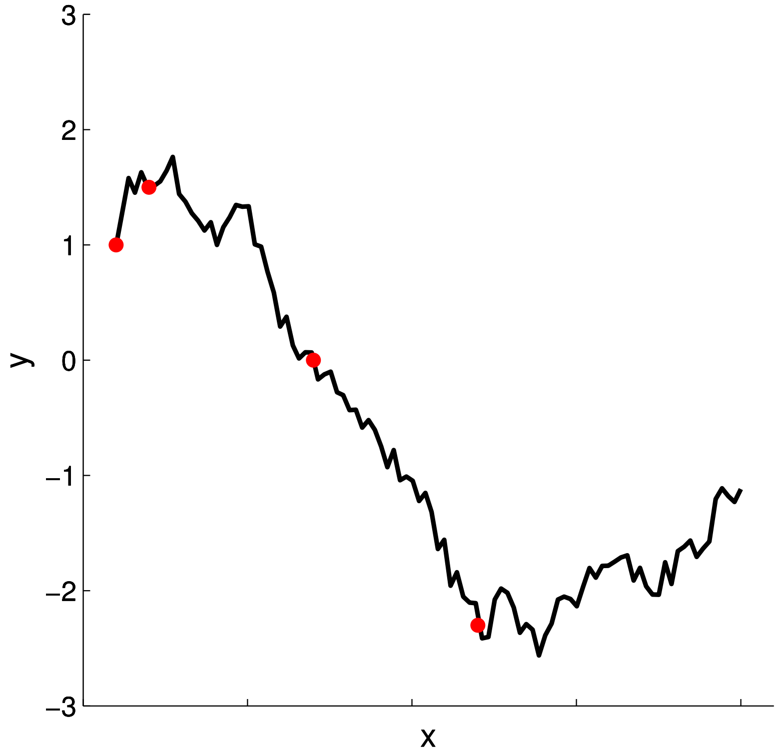
$$K(x_1, x_2) = \sigma^2 \exp\left(-\frac{1}{2l^2}|x_1 - x_2|\right)$$

Laplacian covariance function

Browninan motion

Ornstein-Uhlenbeck

$\Sigma =$ 



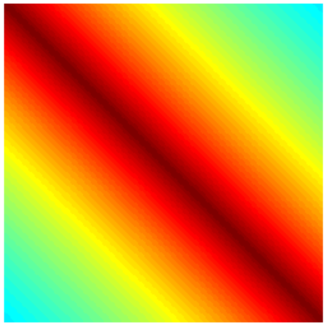
What effect does the form of the covariance function have?

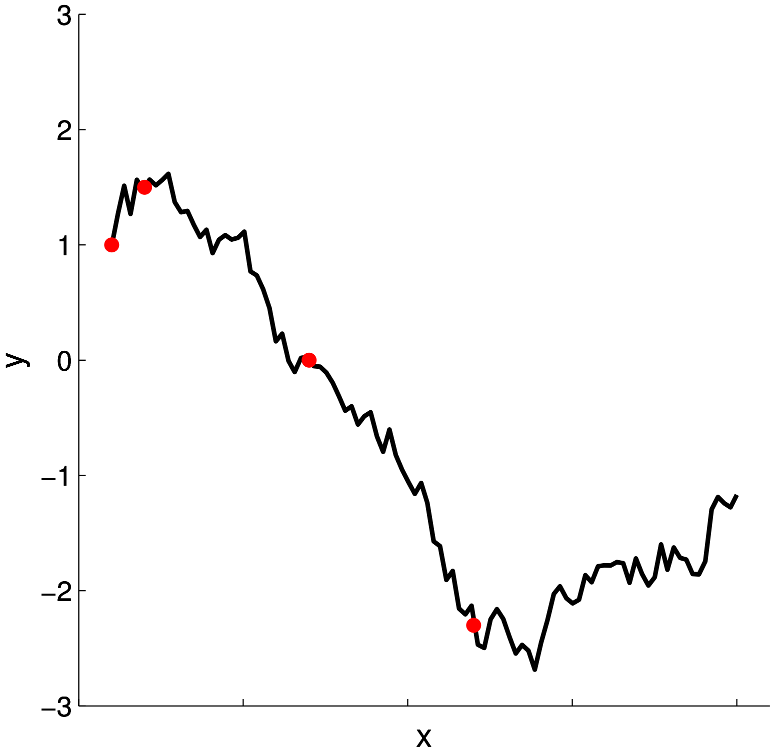
$$\mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) = \sigma^2 \exp\left(-\frac{1}{2\mathsf{l}^2}|\mathsf{x}_1 - \mathsf{x}_2|\right)$$

Laplacian covariance function

Browninan motion

Ornstein-Uhlenbeck

$\Sigma =$ 



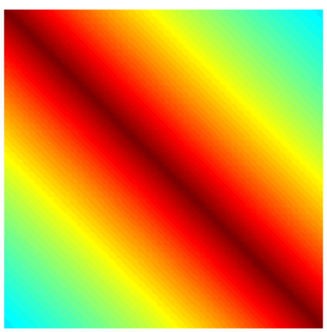
What effect does the form of the covariance function have?

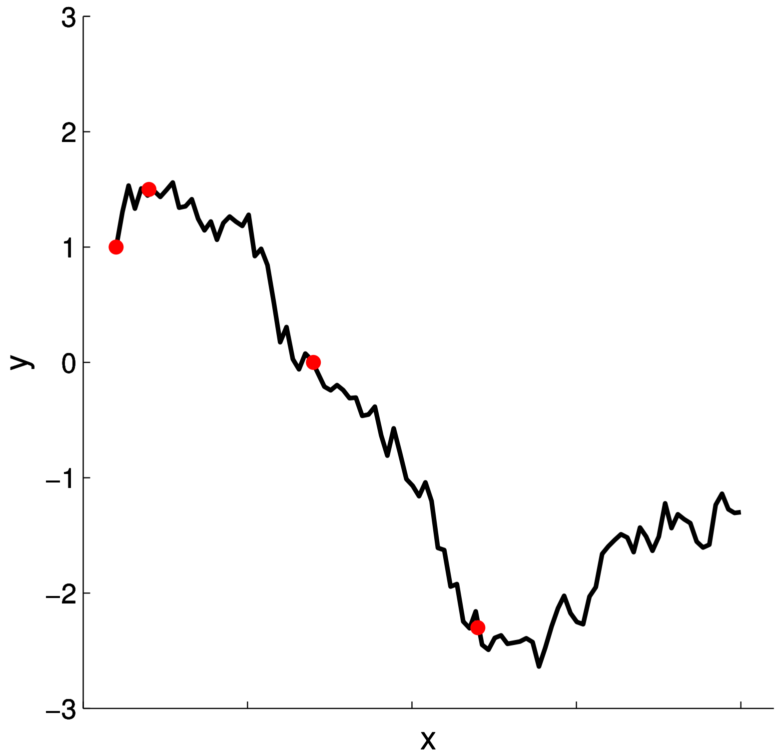
$$K(x_1, x_2) = \sigma^2 \exp\left(-\frac{1}{2l^2}|x_1 - x_2|\right)$$

Laplacian covariance function

Browninan motion

Ornstein-Uhlenbeck

$\Sigma =$ 



What effect does the form of the covariance function have?

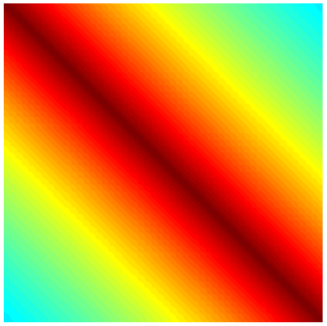
$$\mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) = \sigma^2 \exp\left(-\frac{1}{2\mathsf{l}^2}|\mathsf{x}_1 - \mathsf{x}_2|\right)$$

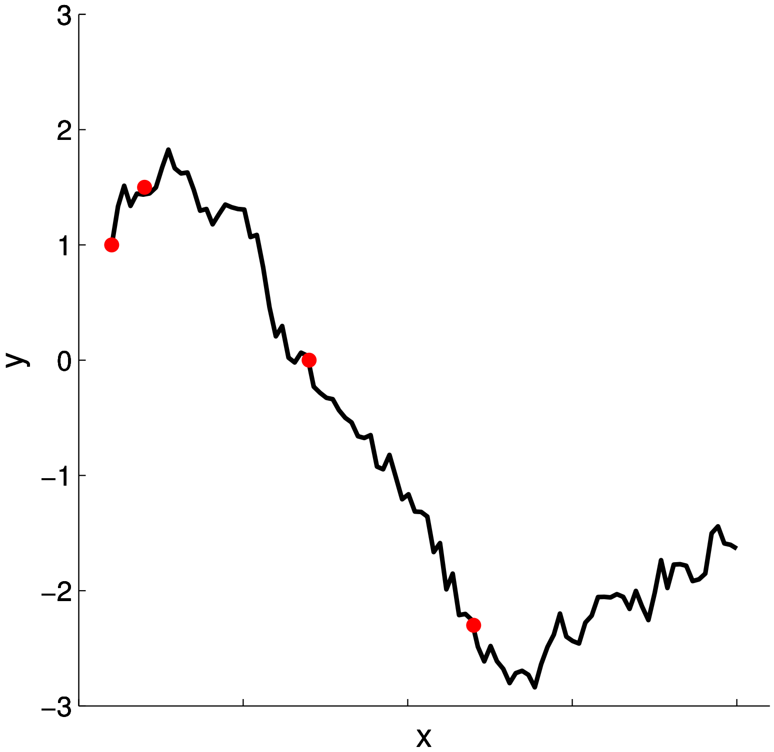
Laplacian covariance function

Browninan motion

Ornstein-Uhlenbeck

$\Sigma =$





What effect does the form of the covariance function have?

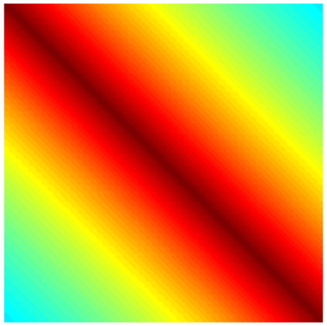
$$K(x_1, x_2) = \sigma^2 \exp\left(-\frac{1}{2l^2}|x_1 - x_2|\right)$$

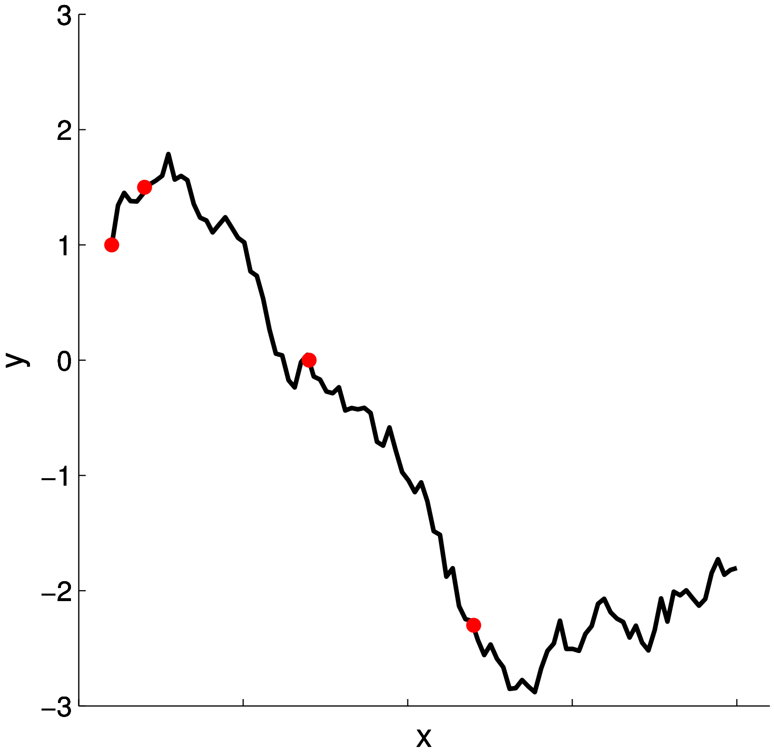
Laplacian covariance function

Browninan motion

Ornstein-Uhlenbeck

$\Sigma =$





What effect does the form of the covariance function have?

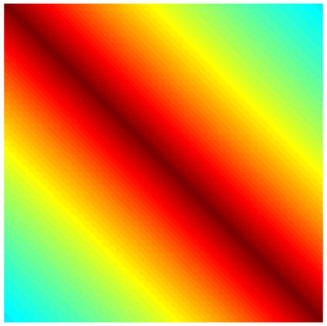
$$K(x_1, x_2) = \sigma^2 \exp\left(-\frac{1}{2l^2}|x_1 - x_2|\right)$$

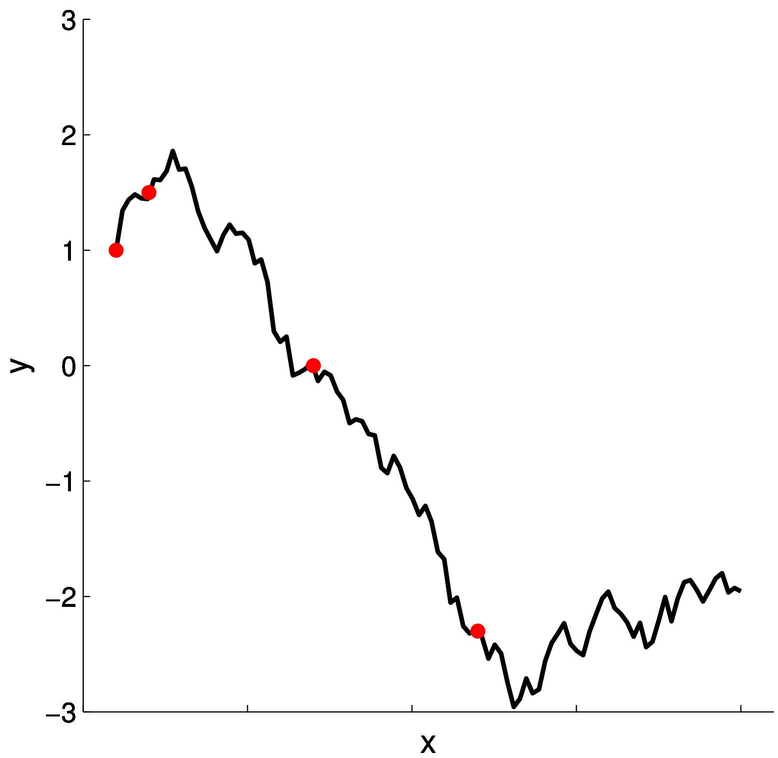
Laplacian covariance function

Browninan motion

Ornstein-Uhlenbeck

$\Sigma =$





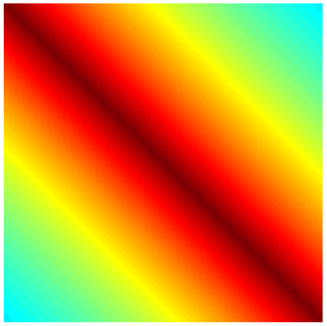
What effect does the form of the covariance function have?

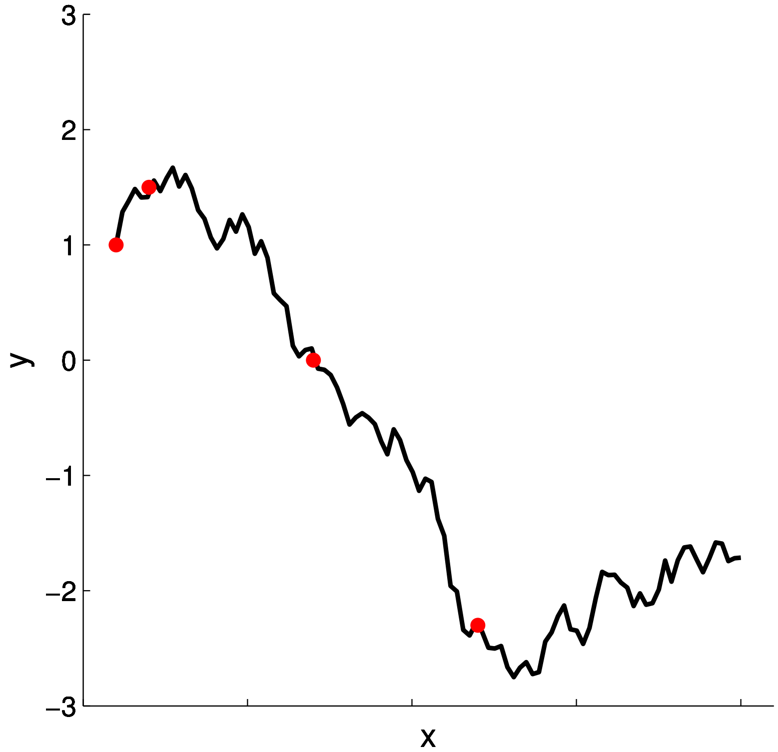
$$K(x_1, x_2) = \sigma^2 \exp\left(-\frac{1}{2l^2}|x_1 - x_2|\right)$$

Laplacian covariance function

Browninan motion

Ornstein-Uhlenbeck

$\Sigma =$ 



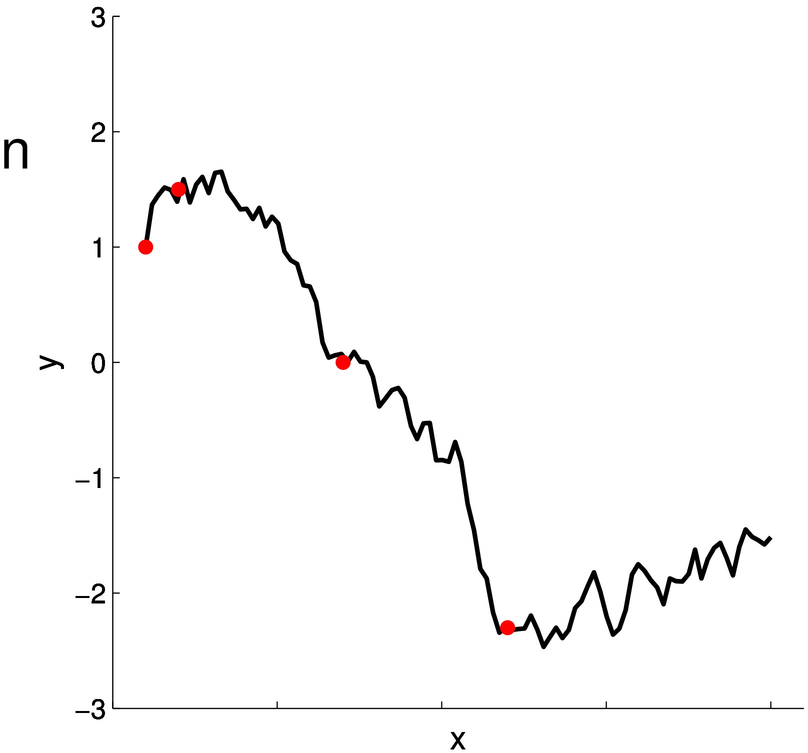
What effect does the form of the covariance function have?

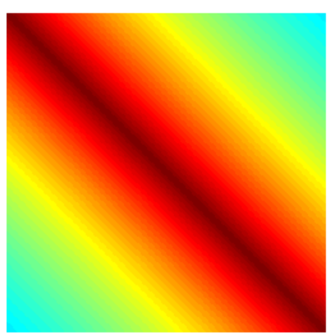
$$\mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) = \sigma^2 \exp\left(-\frac{1}{2\mathsf{l}^2}|\mathsf{x}_1 - \mathsf{x}_2|\right)$$

Laplacian covariance function

Browninan motion

Ornstein-Uhlenbeck



$\Sigma =$ 

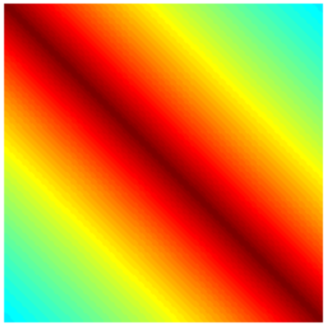
What effect does the form of the covariance function have?

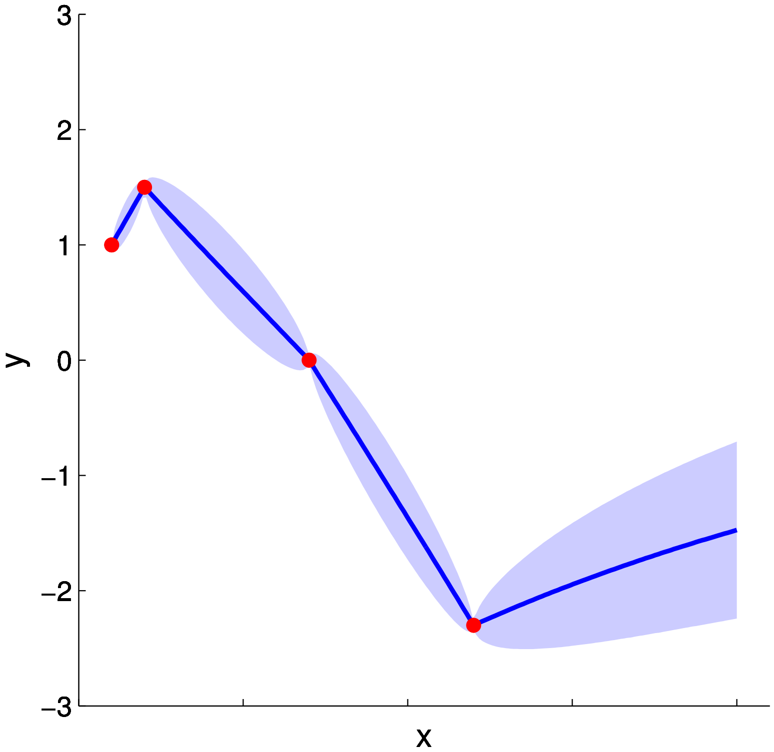
$$K(x_1, x_2) = \sigma^2 \exp\left(-\frac{1}{2l^2}|x_1 - x_2|\right)$$

Laplacian covariance function

Browninan motion

Ornstein-Uhlenbeck

$\Sigma =$ 

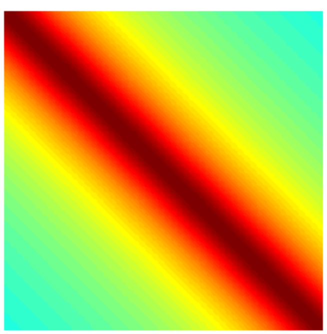


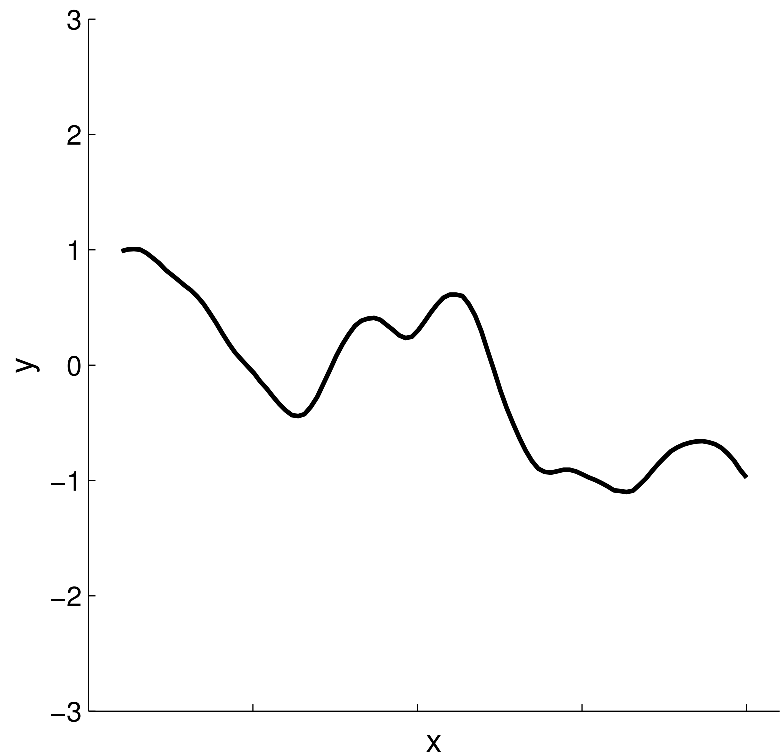
What effect does the form of the covariance function have?

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

Rational Quadratic

$\Sigma =$



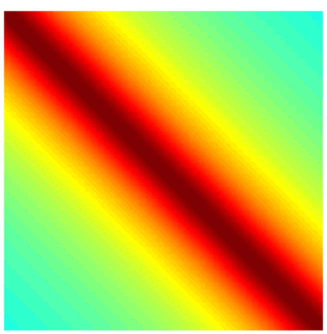


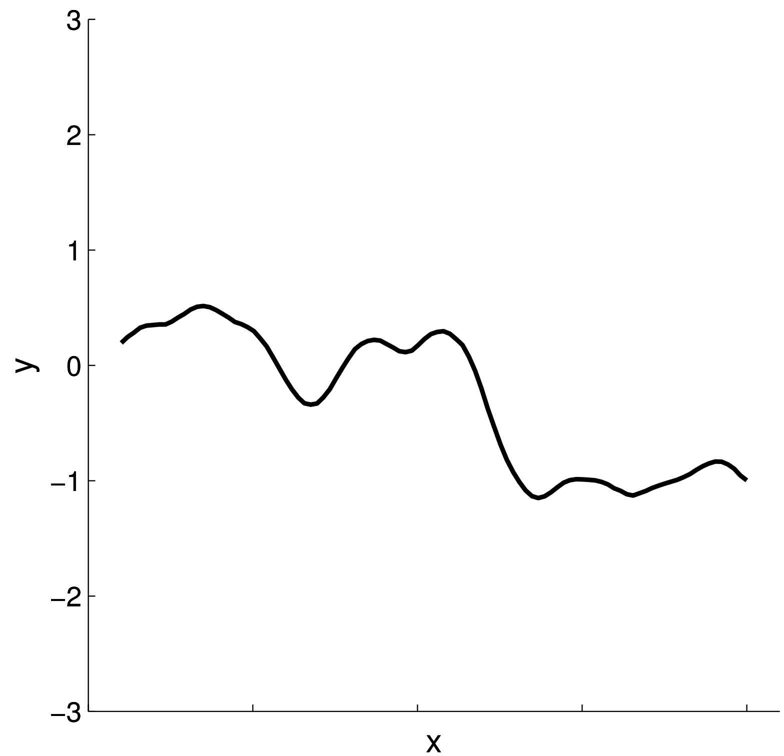
What effect does the form of the covariance function have?

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

Rational Quadratic

$\Sigma =$

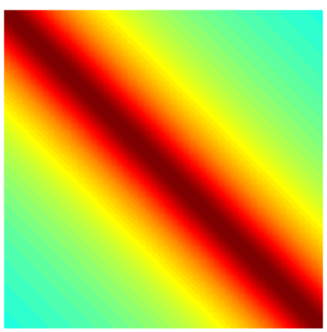


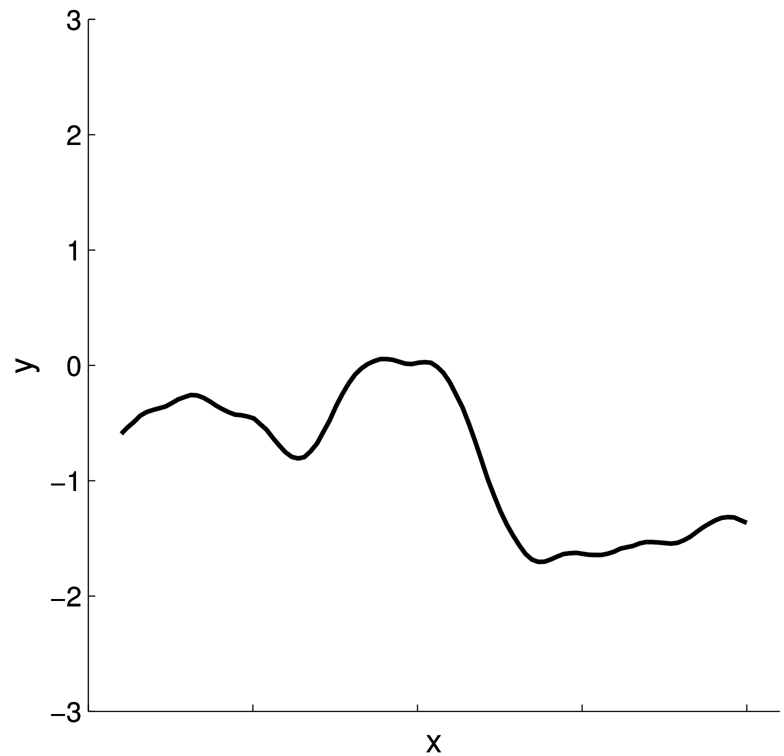


What effect does the form of the covariance function have?

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

Rational Quadratic

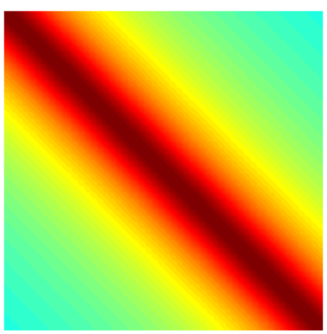
$\Sigma =$ 

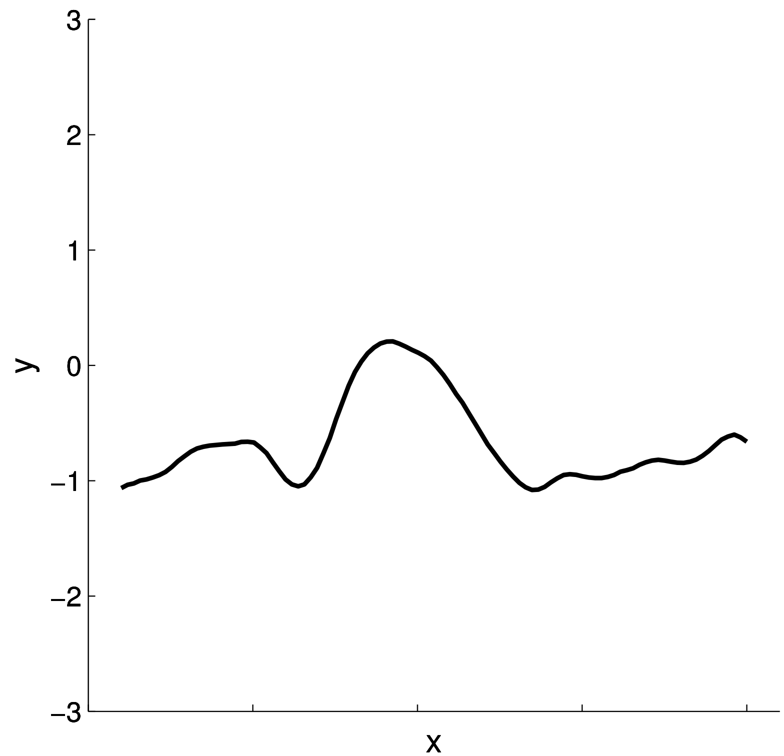


What effect does the form of the covariance function have?

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

Rational Quadratic

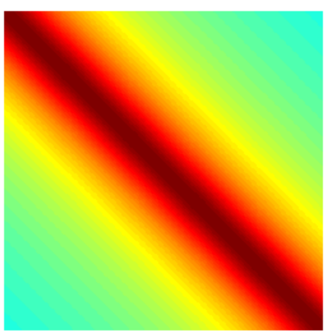
$\Sigma =$ 

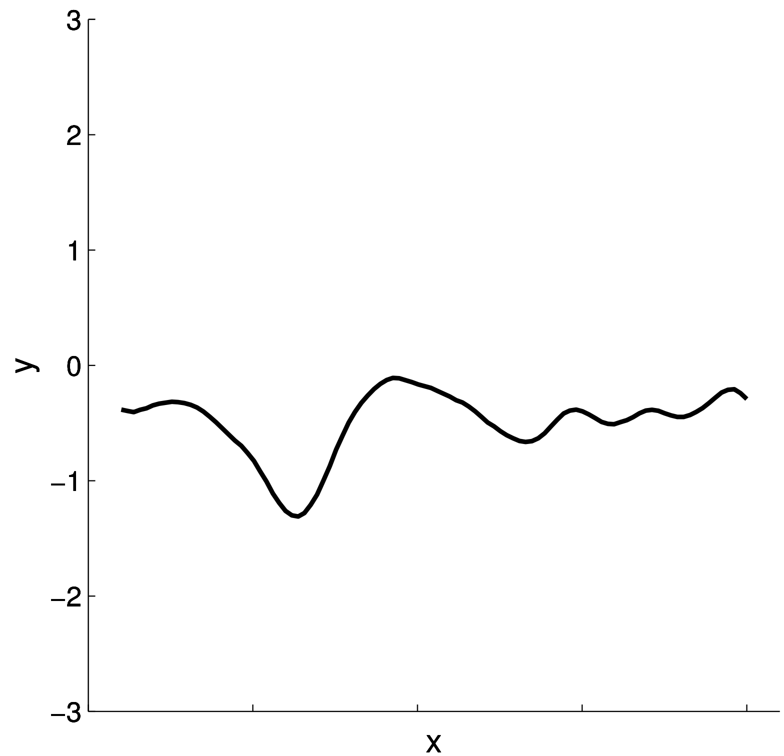


What effect does the form of the covariance function have?

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

Rational Quadratic

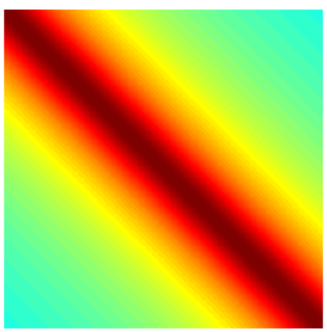
$\Sigma =$ 

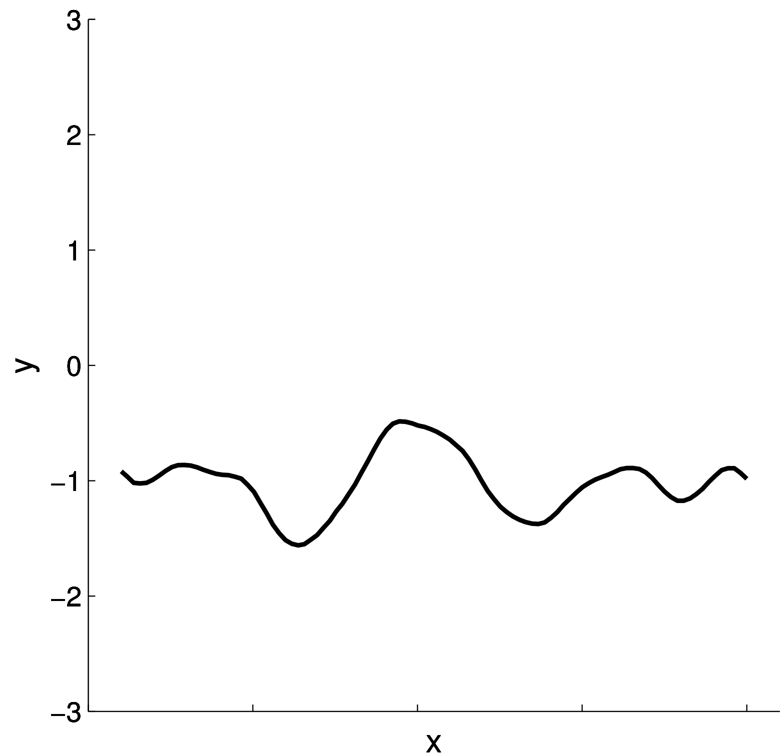


What effect does the form of the covariance function have?

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

Rational Quadratic

$\Sigma =$ 

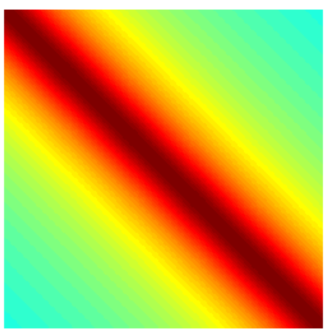


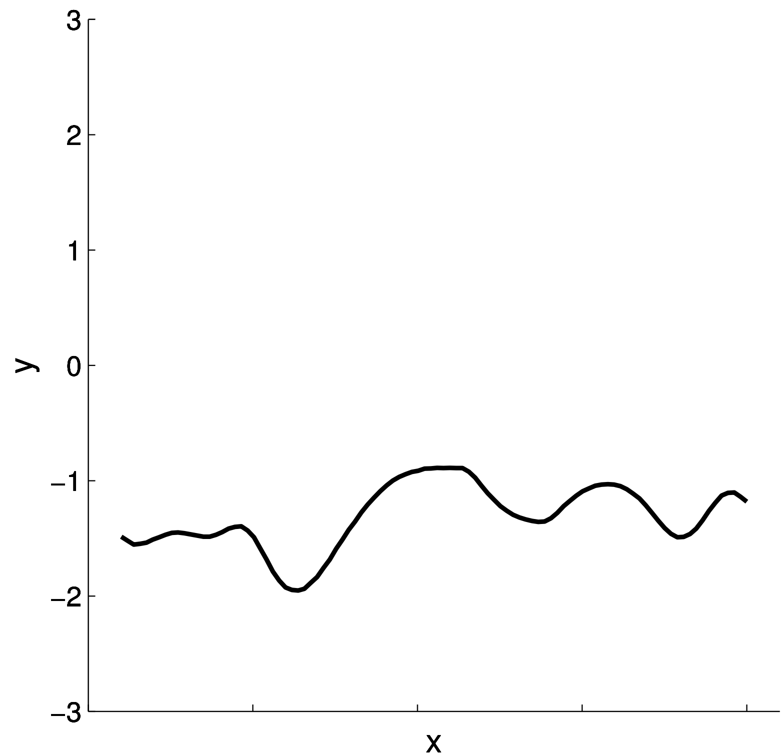
What effect does the form of the covariance function have?

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

Rational Quadratic

$\Sigma =$



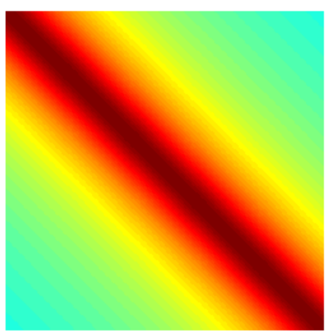


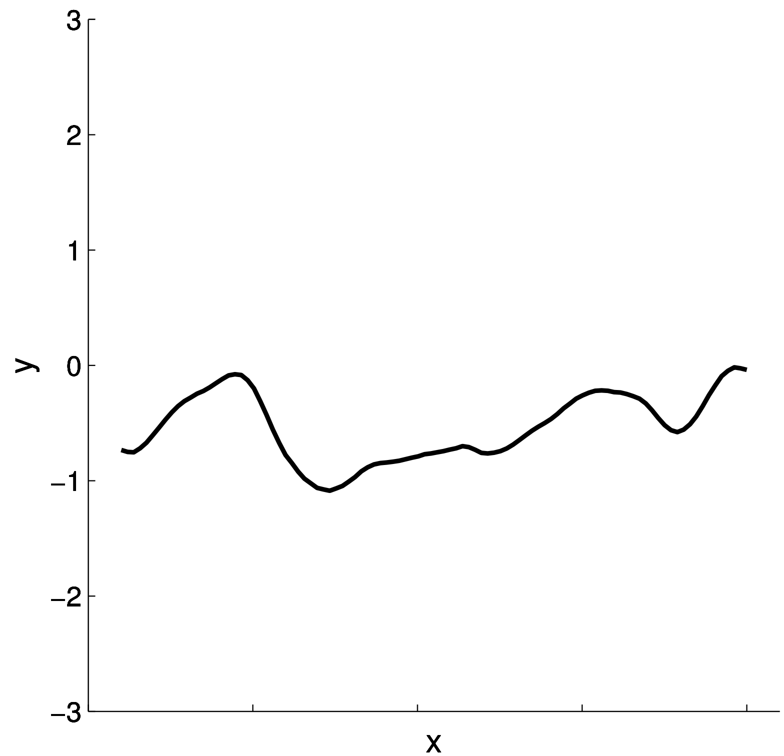
What effect does the form of the covariance function have?

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

Rational Quadratic

$\Sigma =$

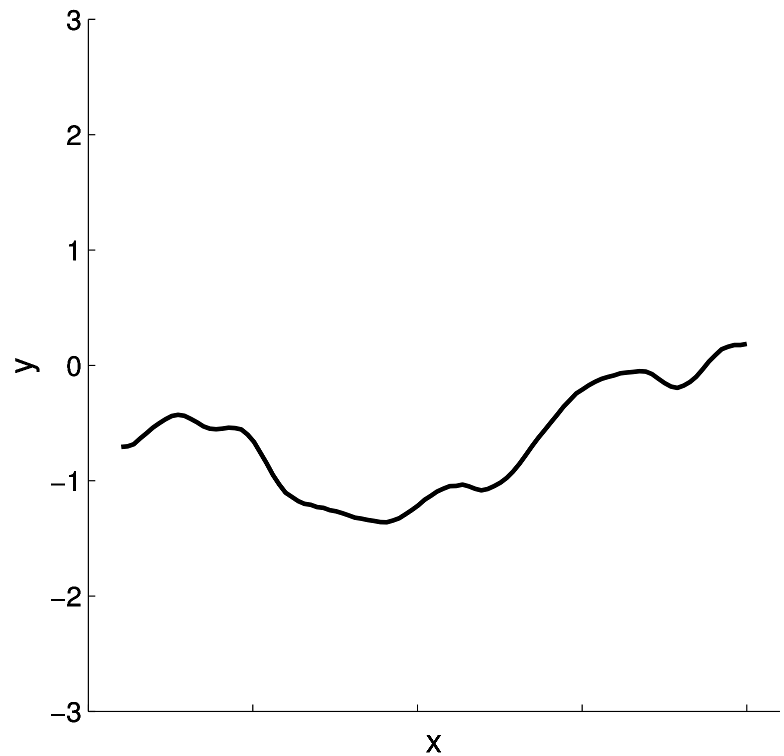




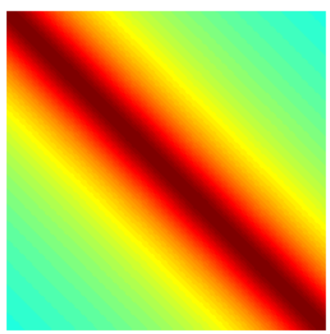
What effect does the form of the covariance function have?

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

Rational Quadratic



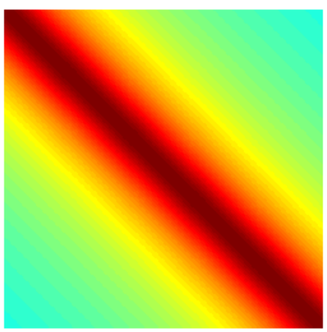
$\Sigma =$

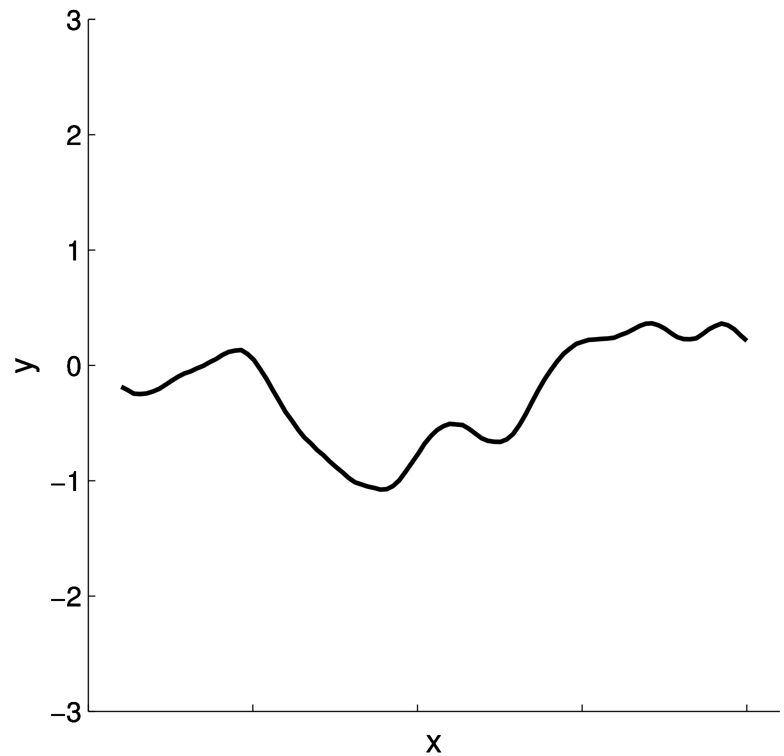


What effect does the form of the covariance function have?

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

Rational Quadratic

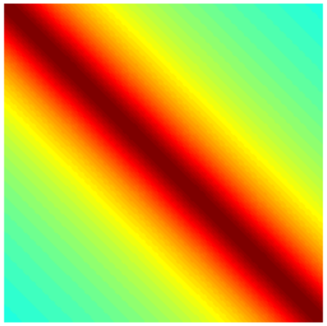
$\Sigma =$ 

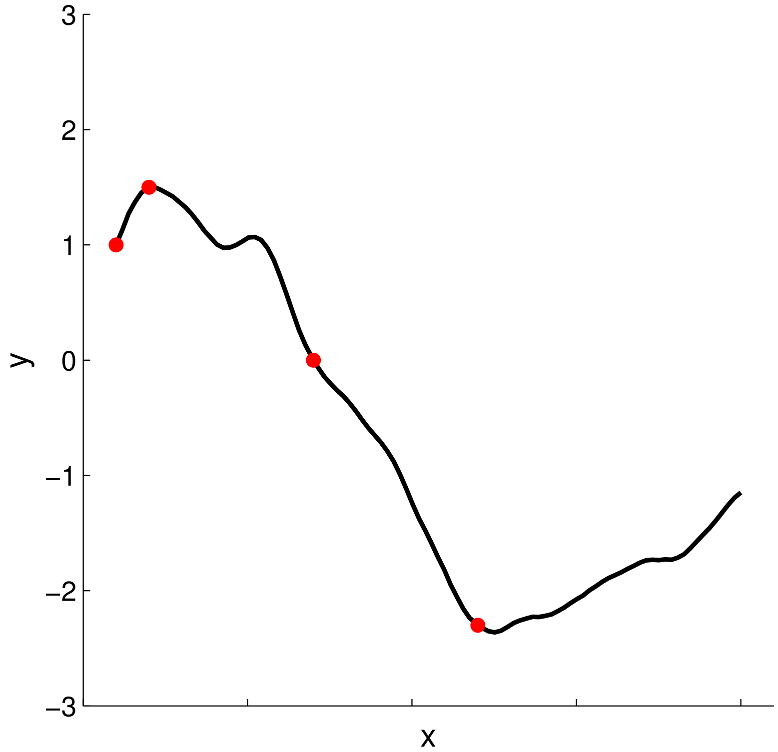


What effect does the form of the covariance function have?

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

Rational Quadratic

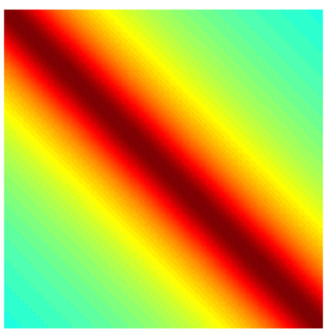
$\Sigma =$ 

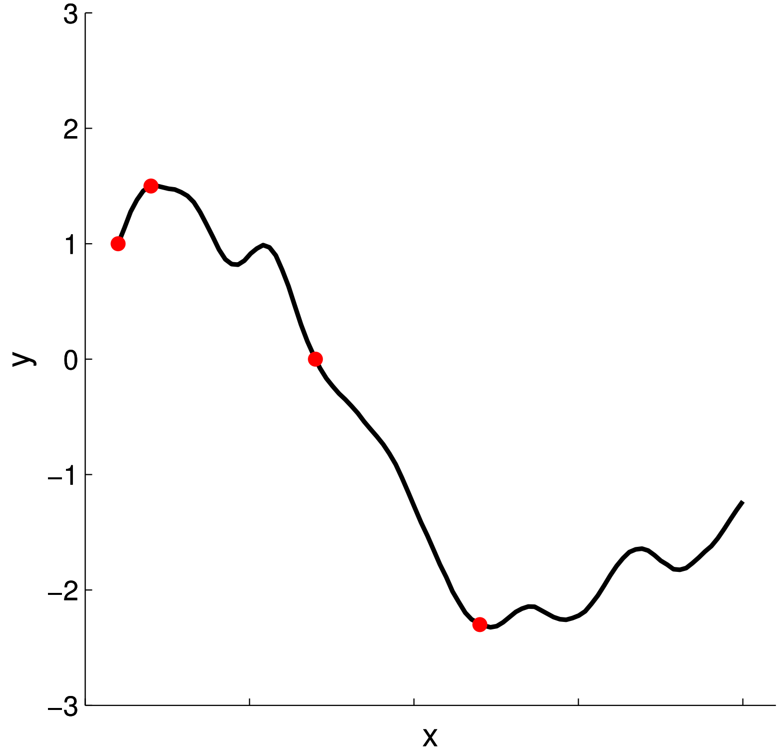


What effect does the form of the covariance function have?

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

Rational Quadratic

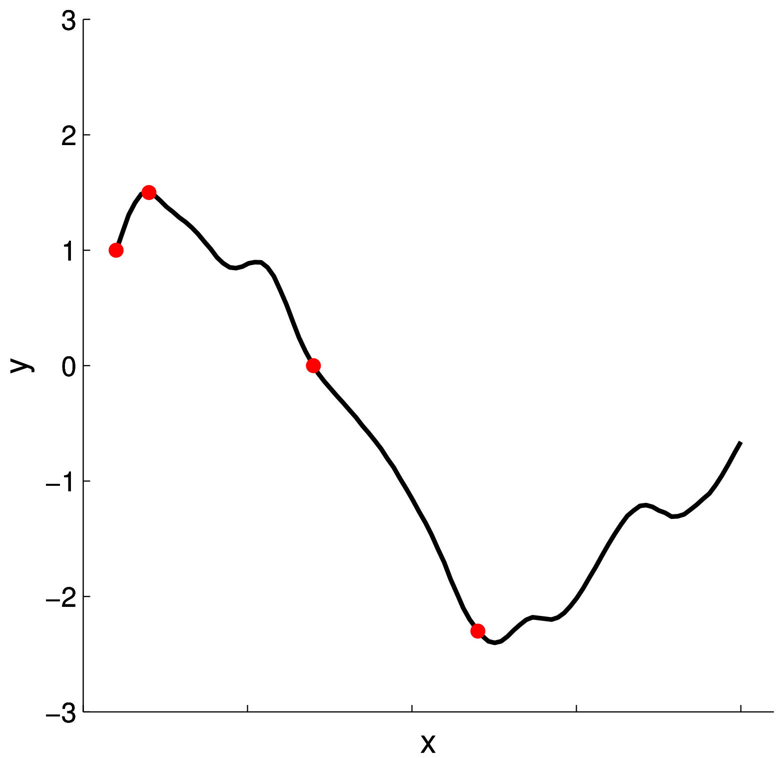
$\Sigma =$ 



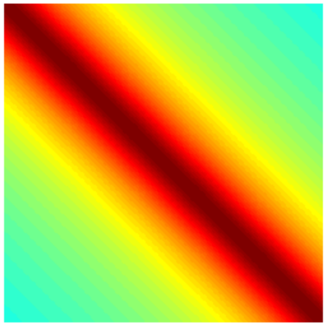
What effect does the form of the covariance function have?

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

Rational Quadratic



$\Sigma =$

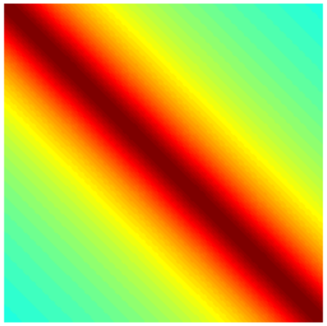


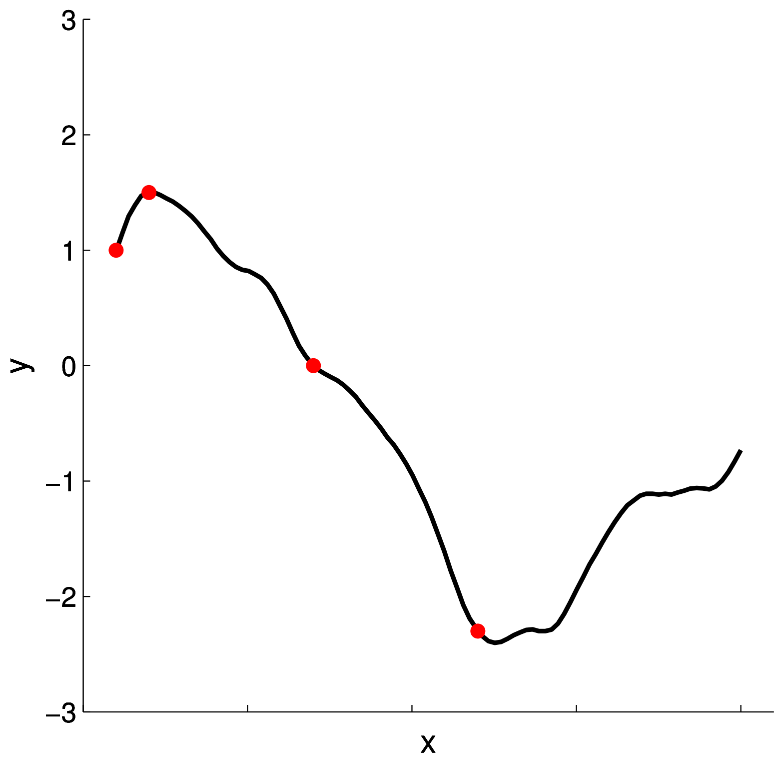
What effect does the form of the covariance function have?

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

Rational Quadratic

$\Sigma =$

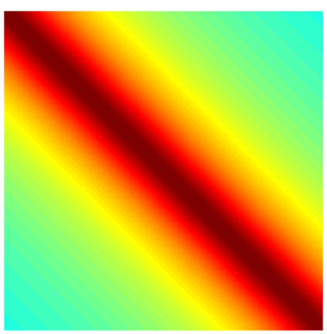


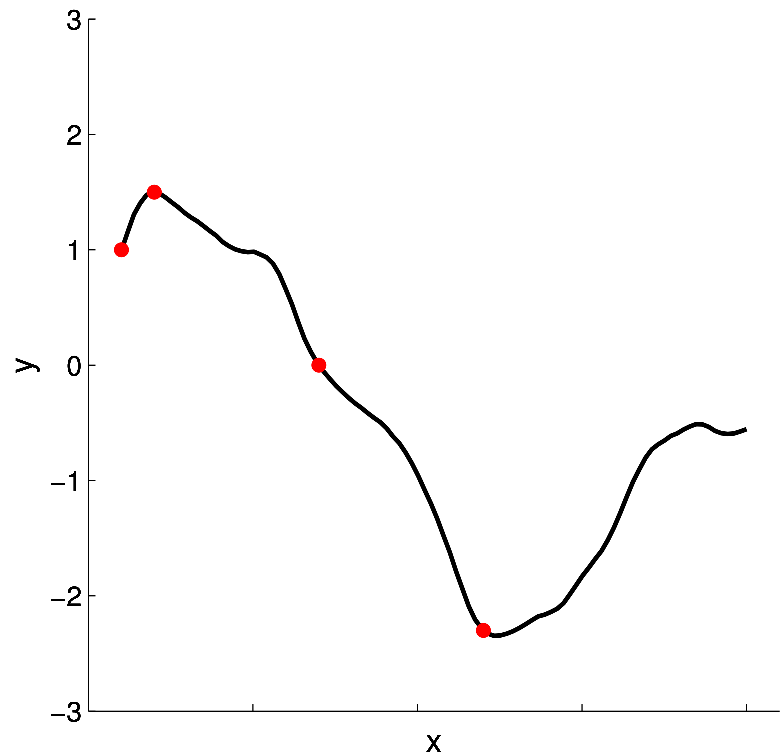


What effect does the form of the covariance function have?

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

Rational Quadratic

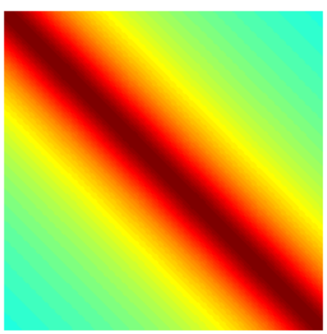
$\Sigma =$ 

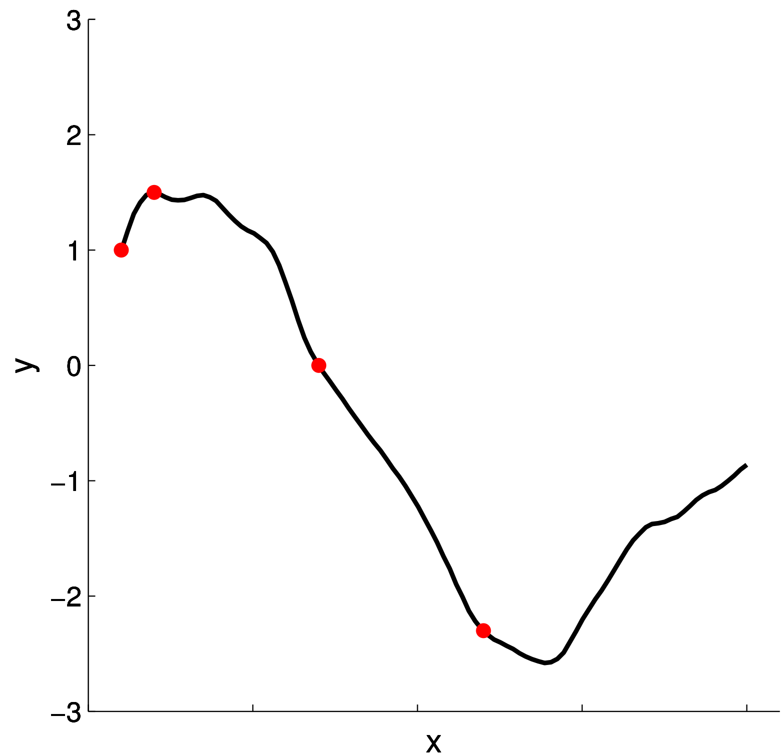


What effect does the form of the covariance function have?

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

Rational Quadratic

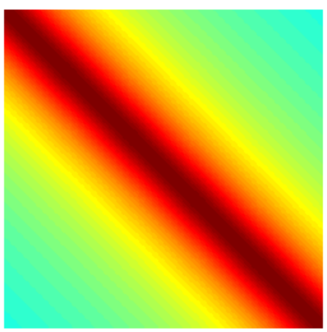
$\Sigma =$ 

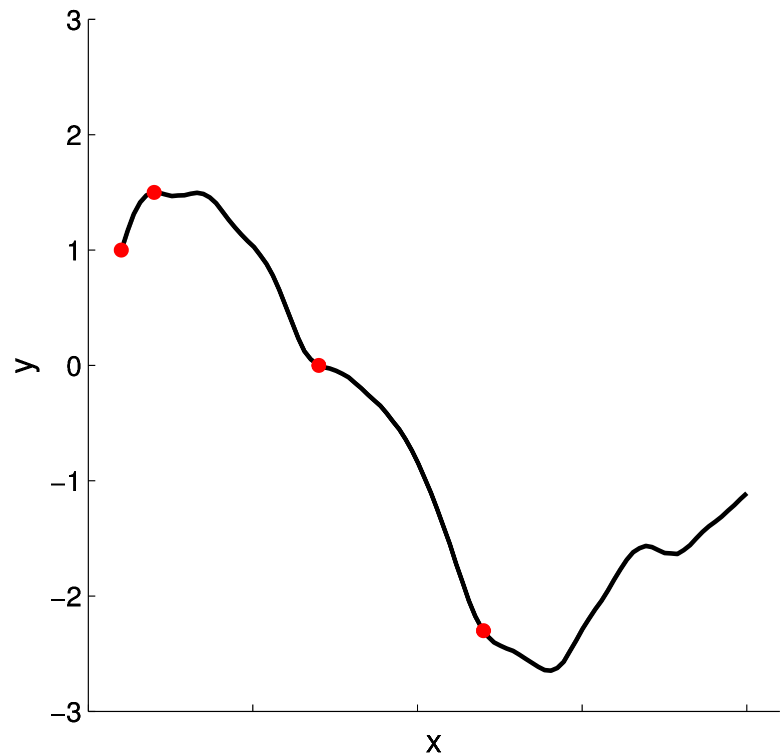


What effect does the form of the covariance function have?

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

Rational Quadratic

$\Sigma =$ 

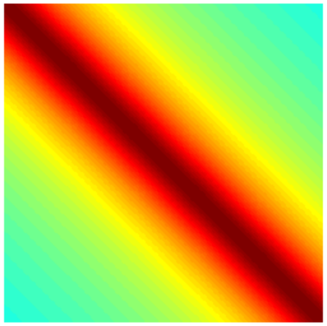


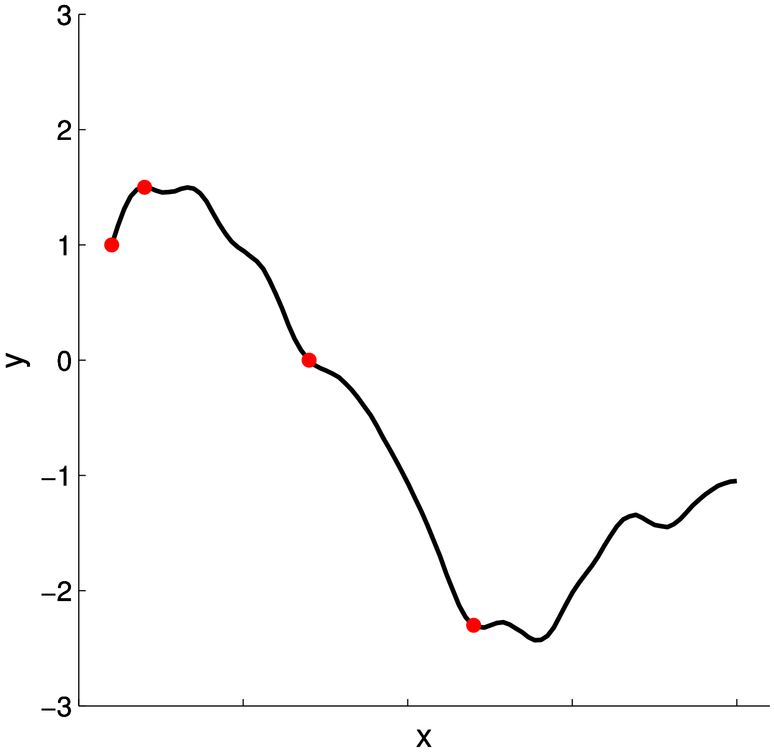
What effect does the form of the covariance function have?

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

Rational Quadratic

$\Sigma =$

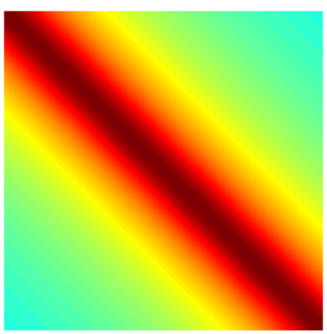


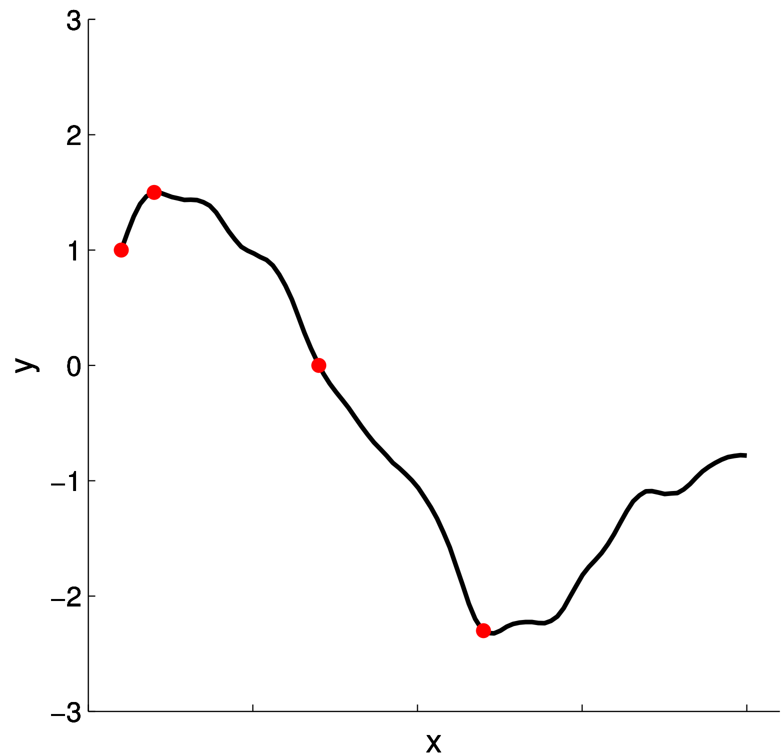


What effect does the form of the covariance function have?

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

Rational Quadratic

$\Sigma =$ 

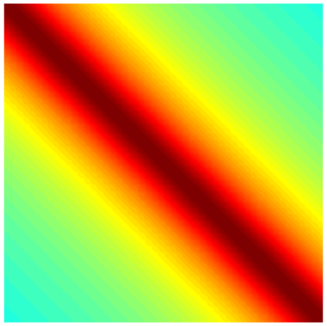


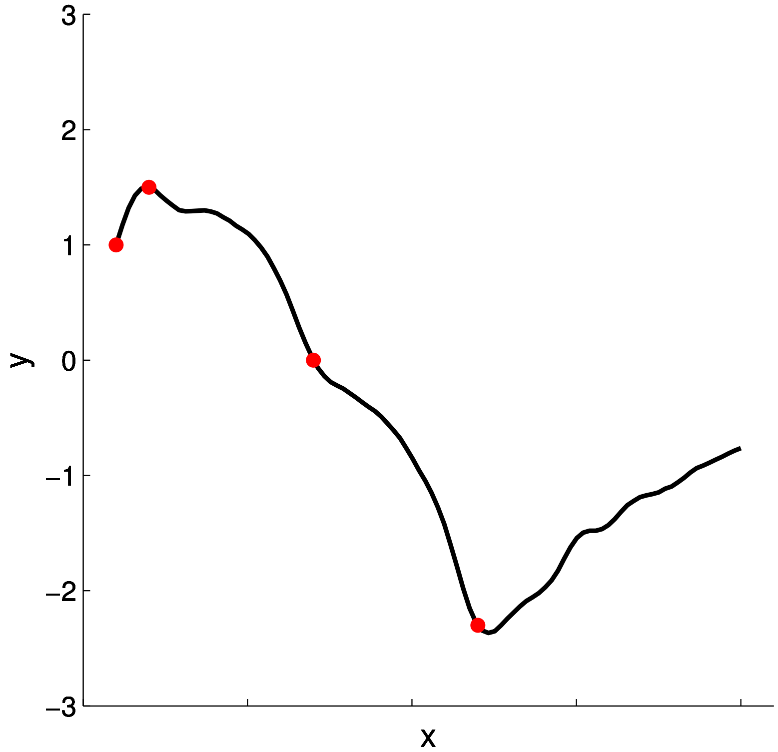
What effect does the form of the covariance function have?

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

Rational Quadratic

$\Sigma =$

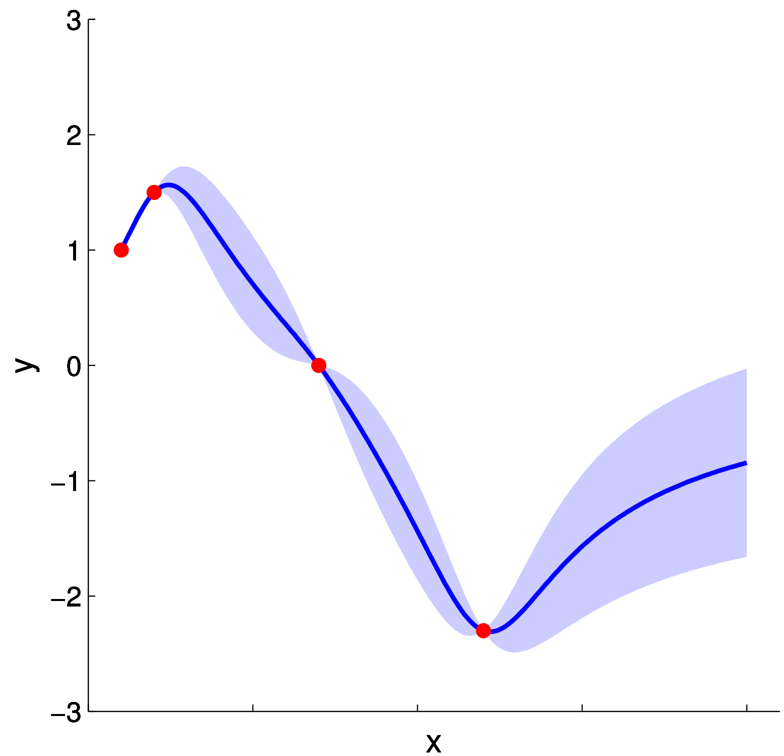




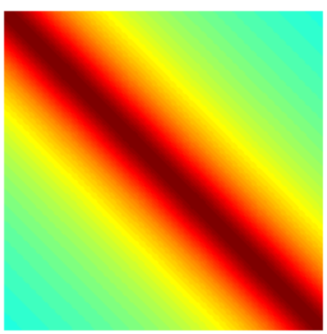
What effect does the form of the covariance function have?

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

Rational Quadratic



$\Sigma =$

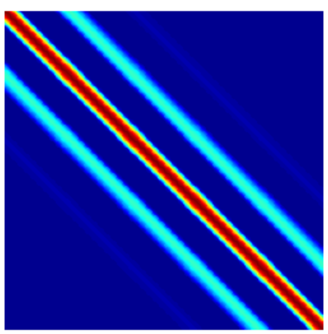


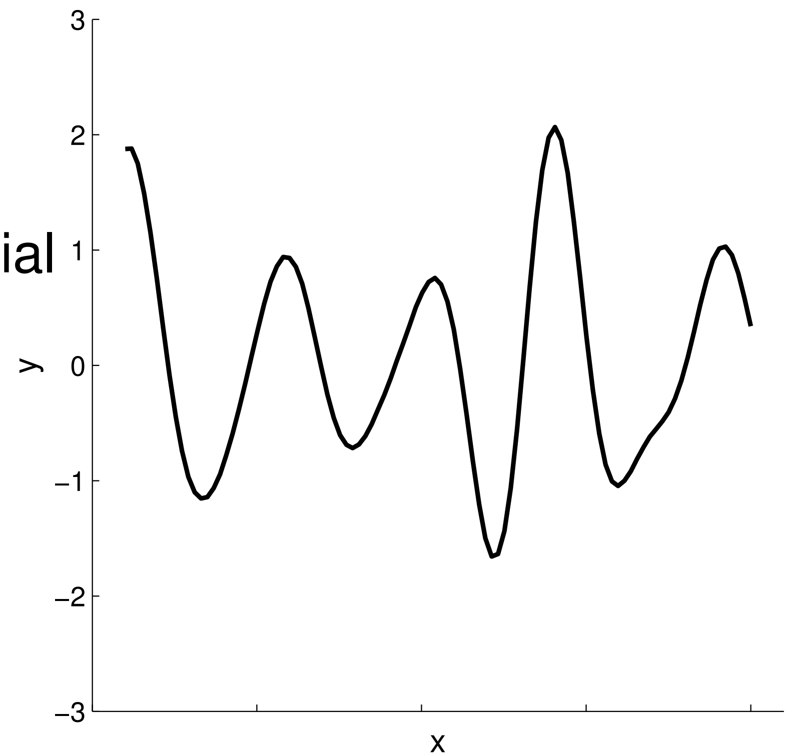
What effect does the form of the covariance function have?

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

Periodic

sinusoid $\times$ squared exponential

$\Sigma =$ 

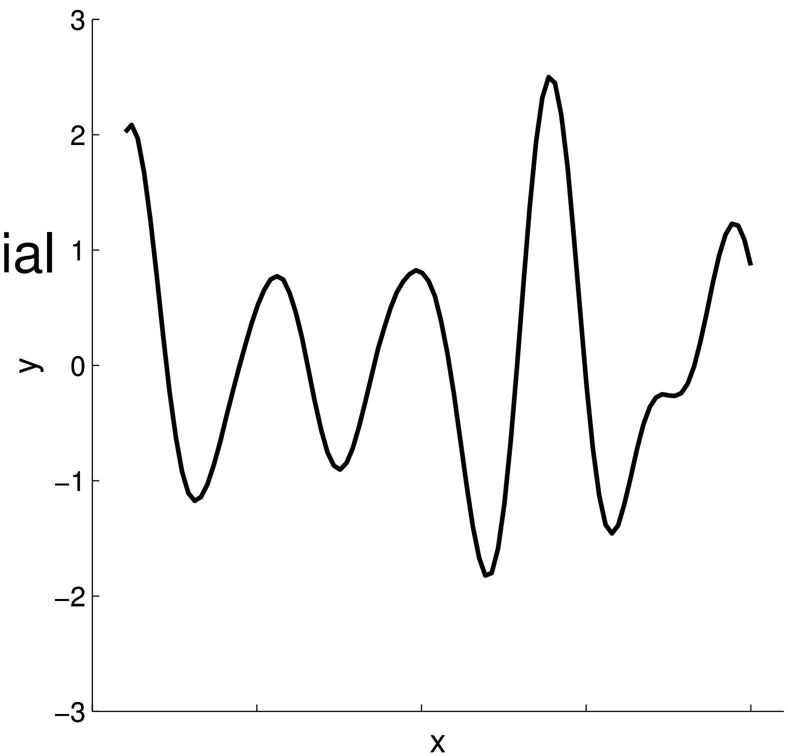


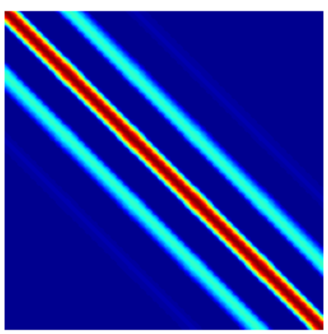
What effect does the form of the covariance function have?

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

Periodic

sinusoid $\times$ squared exponential



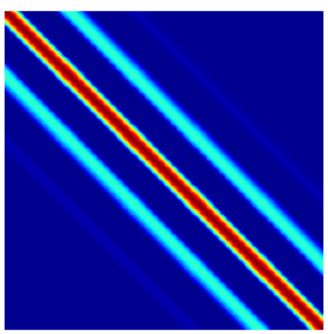
$\Sigma =$ 

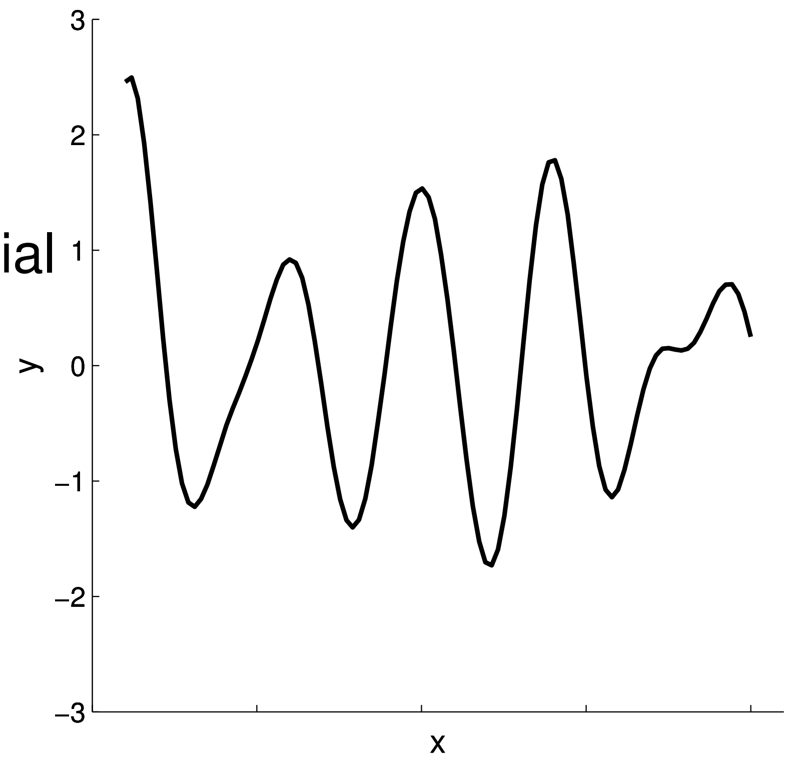
What effect does the form of the covariance function have?

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

Periodic

sinusoid $\times$ squared exponential

$\Sigma =$ 

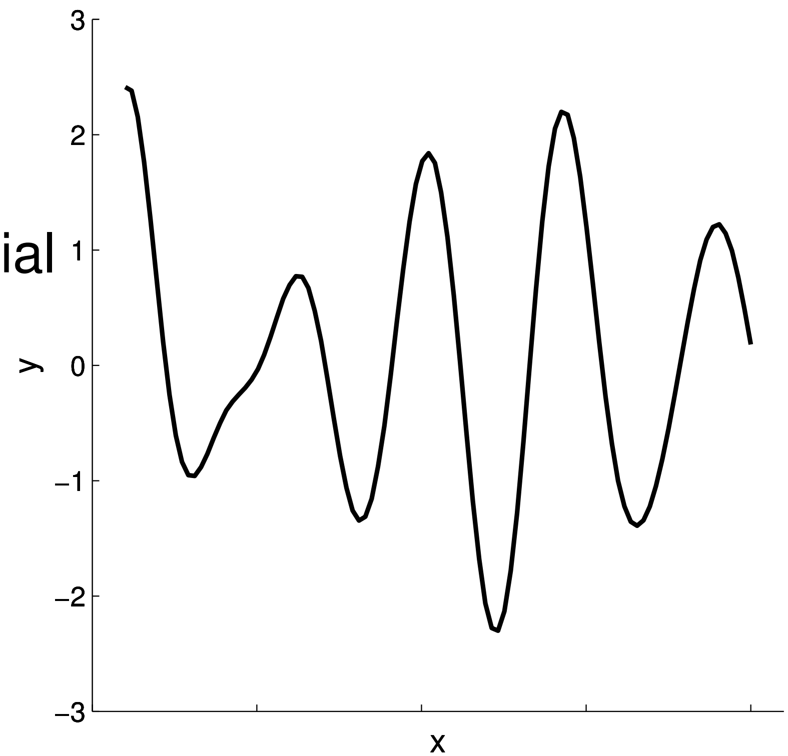


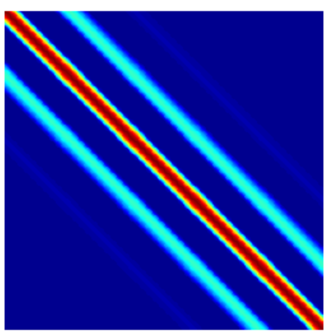
What effect does the form of the covariance function have?

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

Periodic

sinusoid $\times$ squared exponential



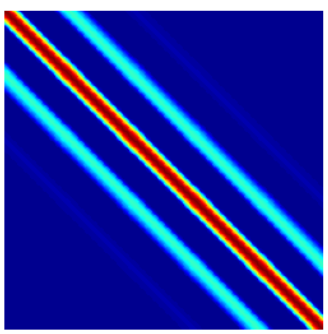
$\Sigma =$ 

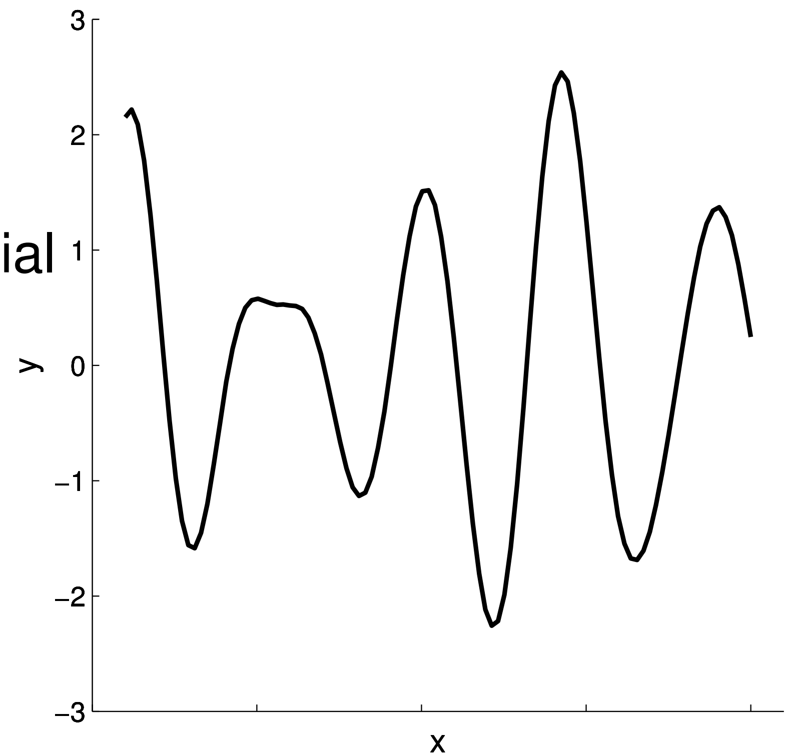
What effect does the form of the covariance function have?

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

Periodic

sinusoid $\times$ squared exponential

$\Sigma =$ 

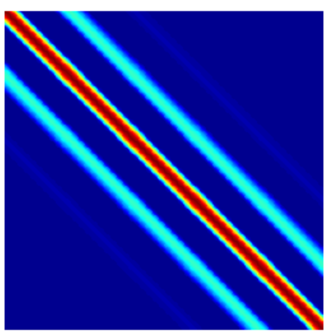


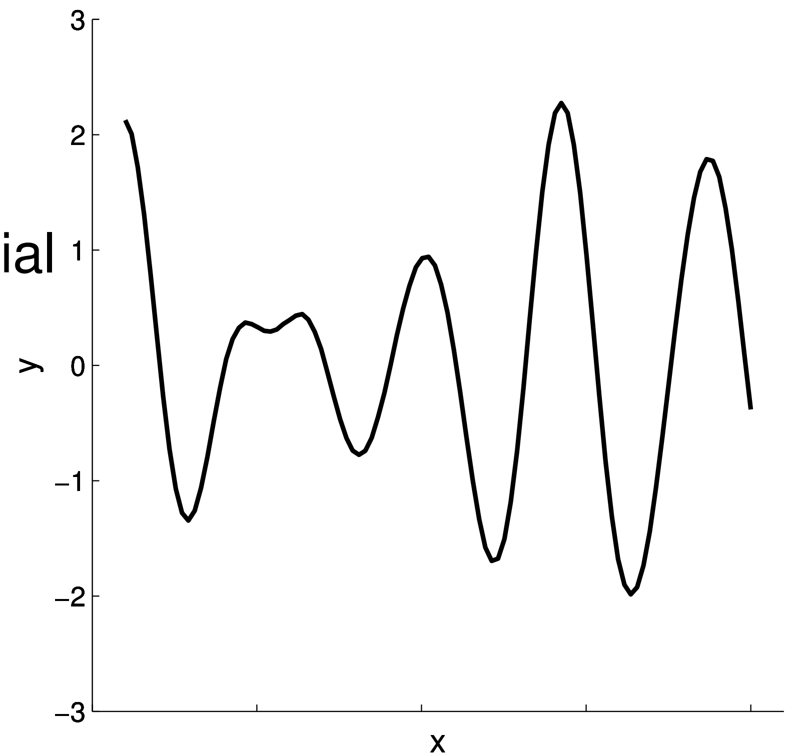
What effect does the form of the covariance function have?

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

Periodic

sinusoid $\times$ squared exponential

$\Sigma =$ 

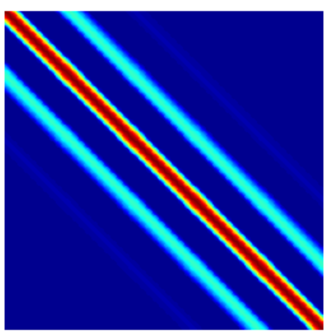


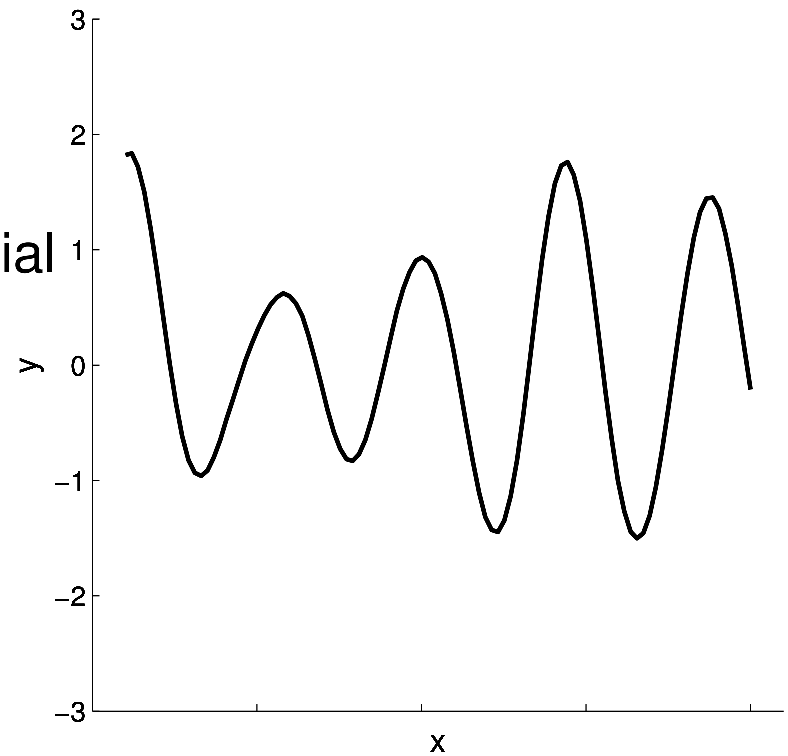
What effect does the form of the covariance function have?

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

Periodic

sinusoid $\times$ squared exponential

$\Sigma =$ 

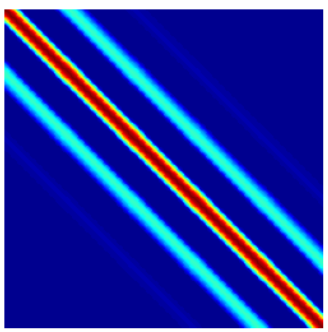


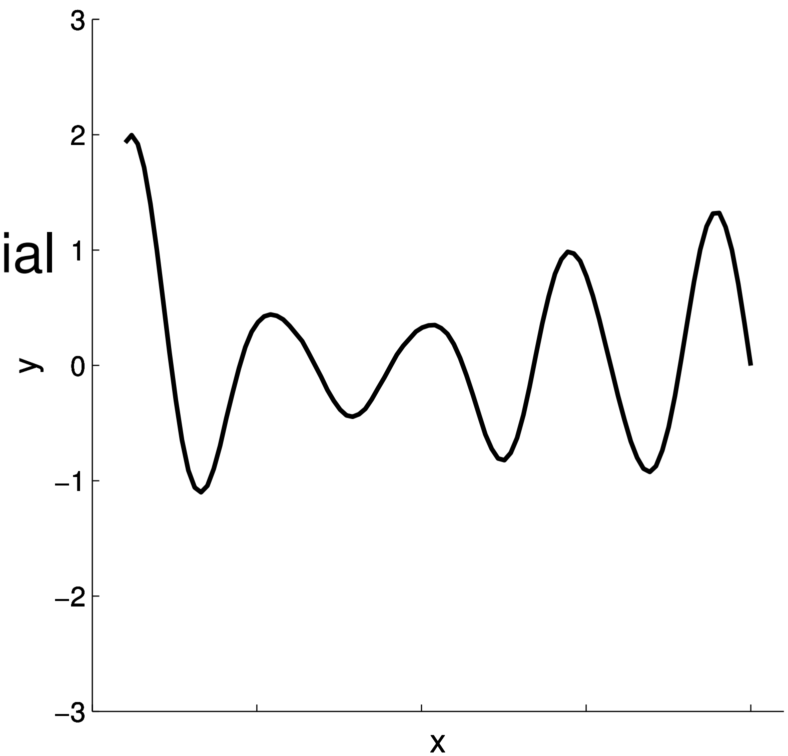
What effect does the form of the covariance function have?

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

Periodic

sinusoid $\times$ squared exponential

$\Sigma =$ 

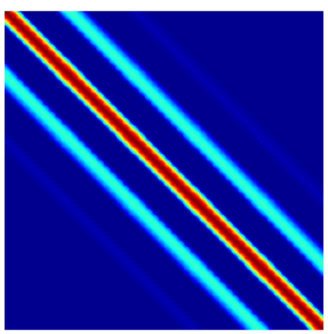


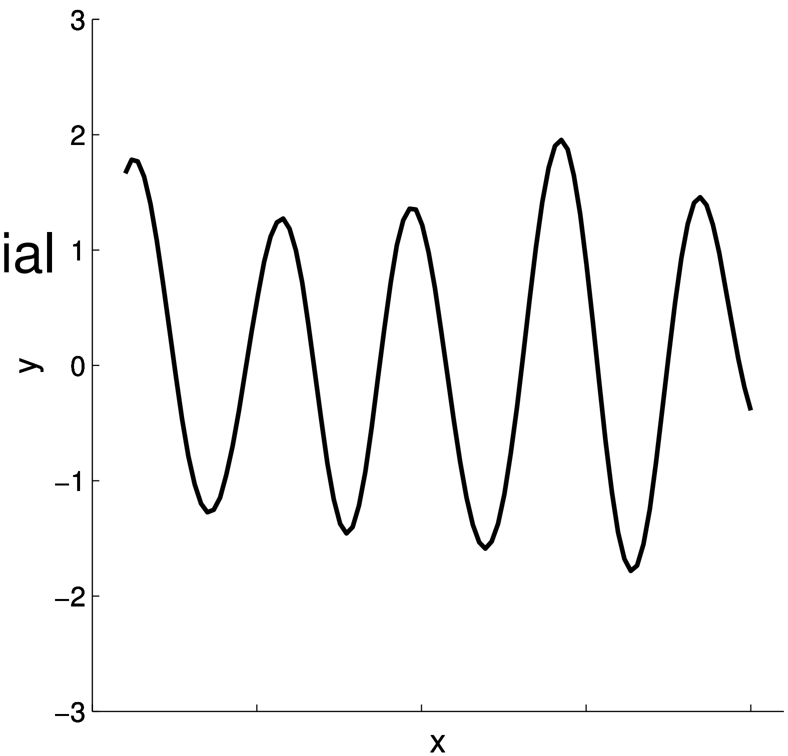
What effect does the form of the covariance function have?

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

Periodic

sinusoid $\times$ squared exponential

$\Sigma =$ 

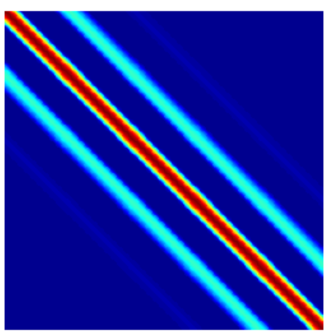


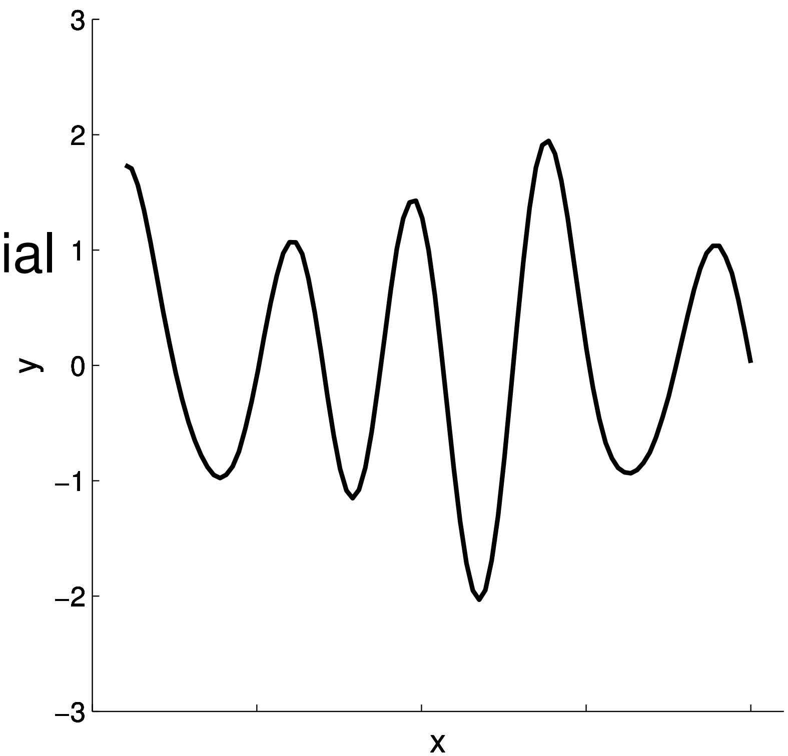
What effect does the form of the covariance function have?

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

Periodic

sinusoid $\times$ squared exponential

$\Sigma =$ 

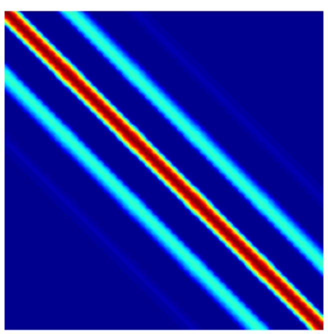


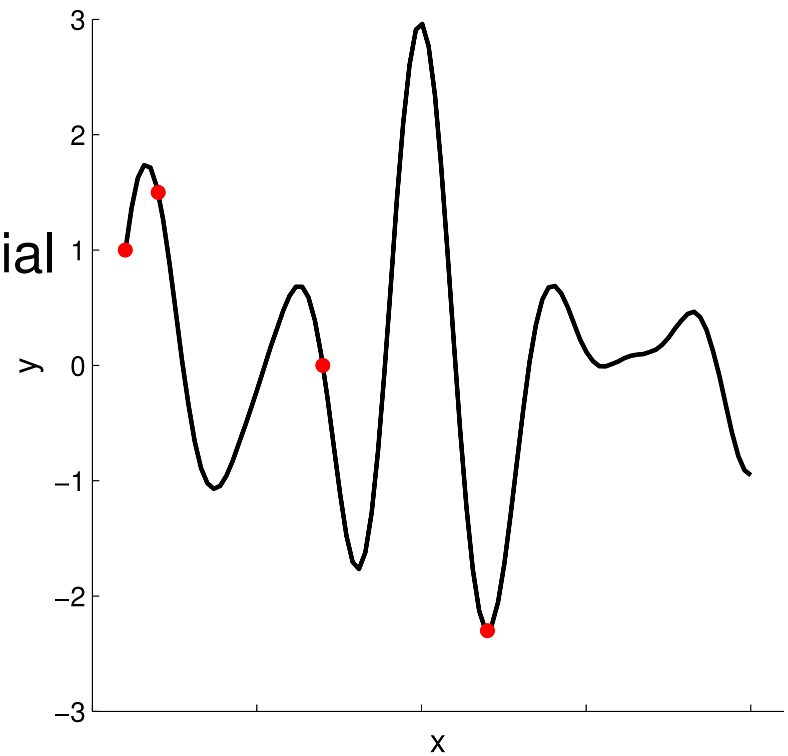
What effect does the form of the covariance function have?

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

Periodic

sinusoid $\times$ squared exponential

$\Sigma =$ 

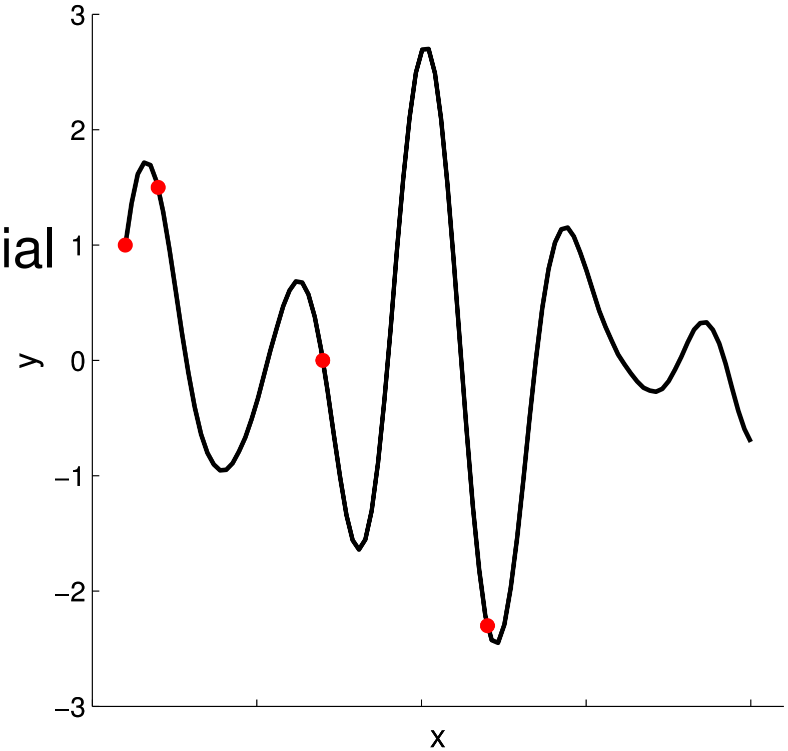


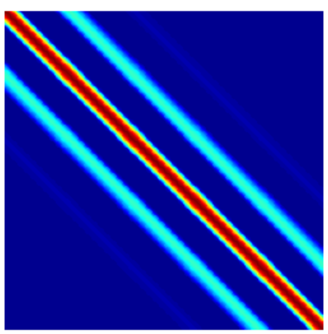
What effect does the form of the covariance function have?

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

Periodic

sinusoid $\times$ squared exponential



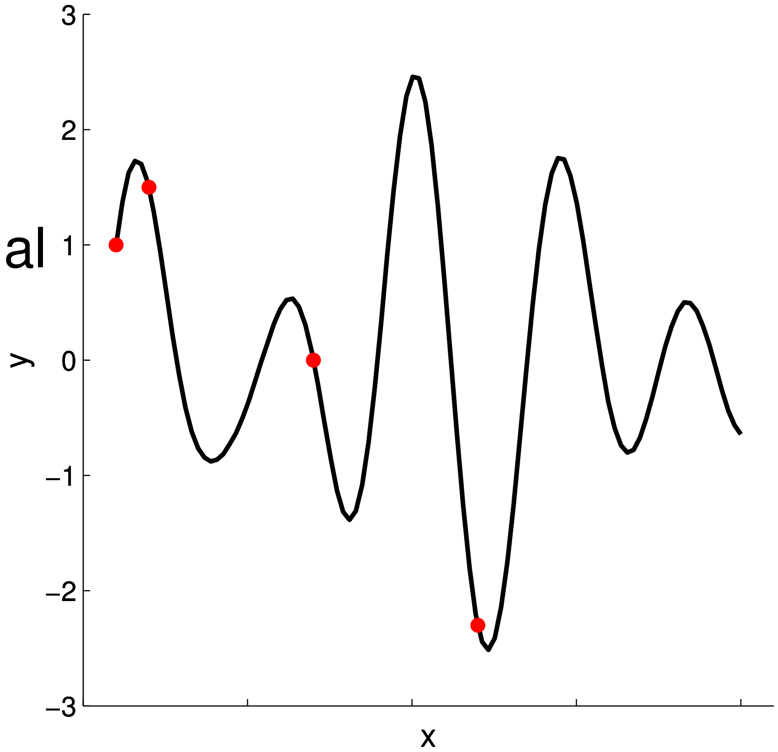
$\Sigma =$ 

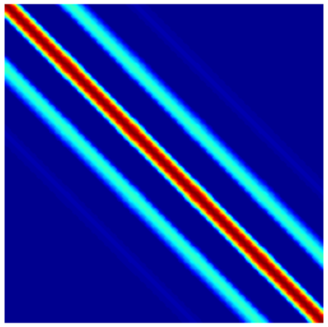
What effect does the form of the covariance function have?

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

Periodic

sinusoid $\times$ squared exponential



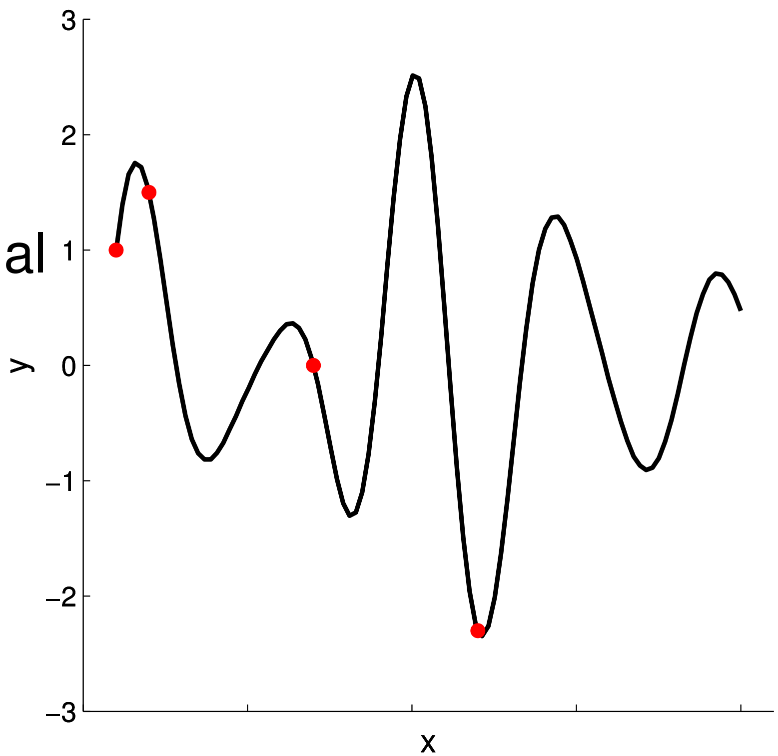
$\Sigma =$ 

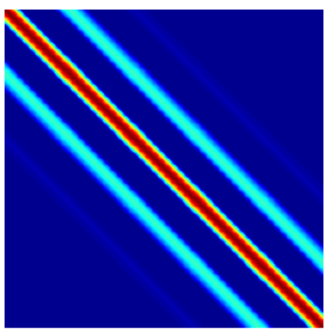
What effect does the form of the covariance function have?

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

Periodic

sinusoid $\times$ squared exponential



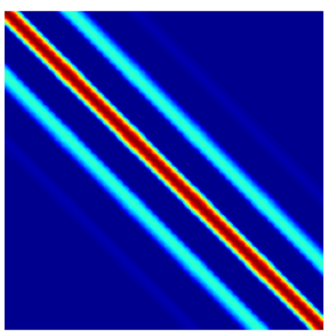
$\Sigma =$ 

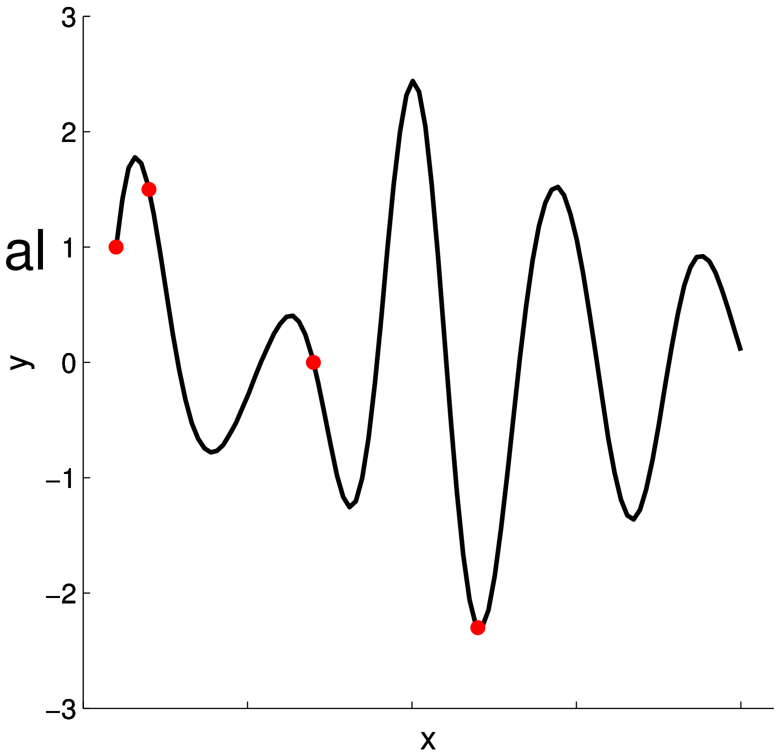
What effect does the form of the covariance function have?

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

Periodic

sinusoid $\times$ squared exponential

$\Sigma =$ 

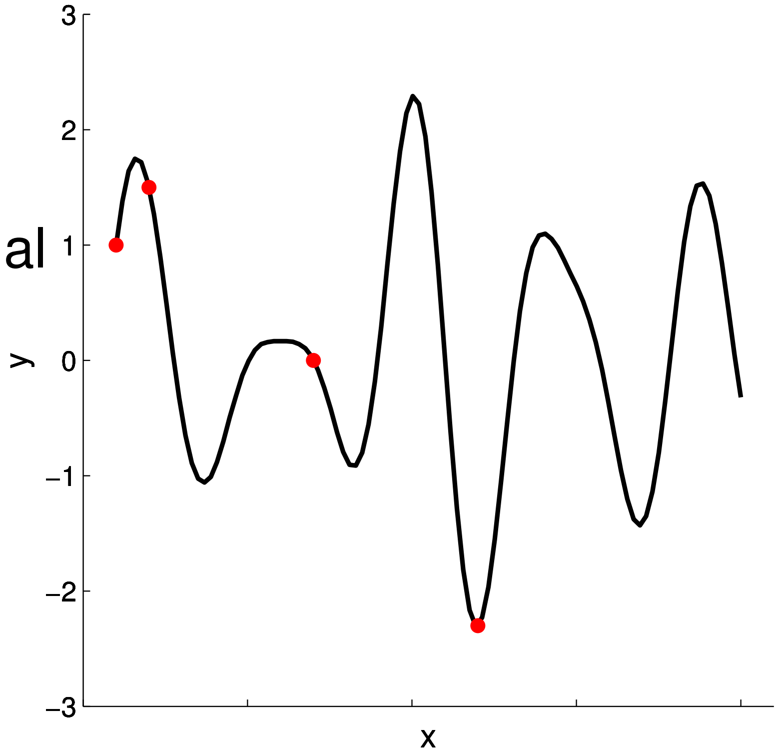


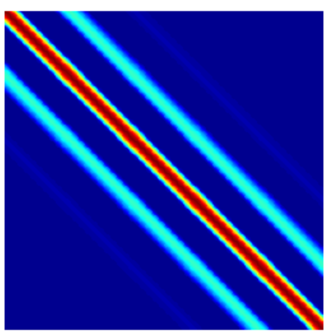
What effect does the form of the covariance function have?

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

Periodic

sinusoid $\times$ squared exponential



$\Sigma =$ 

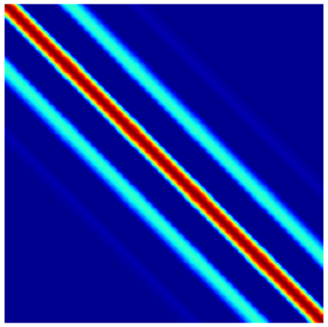
What effect does the form of the covariance function have?

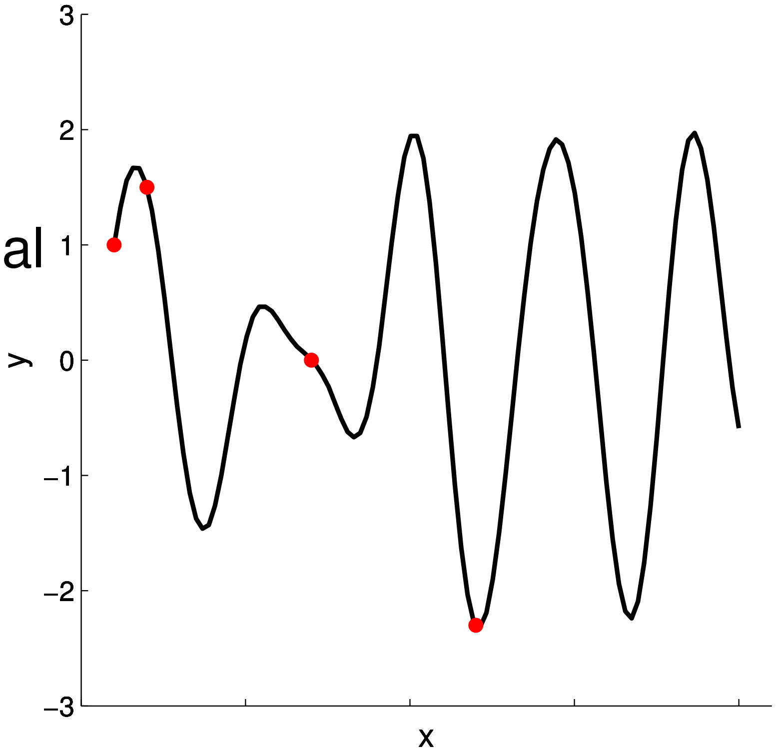
$$K(x_1, x_2) = \sigma^2 \cos(\omega(x_1 - x_2)) \exp\left(-\frac{1}{2l^2}(x_1 - x_2)^2\right)$$

Periodic

sinusoid $\times$ squared exponential

$\Sigma =$



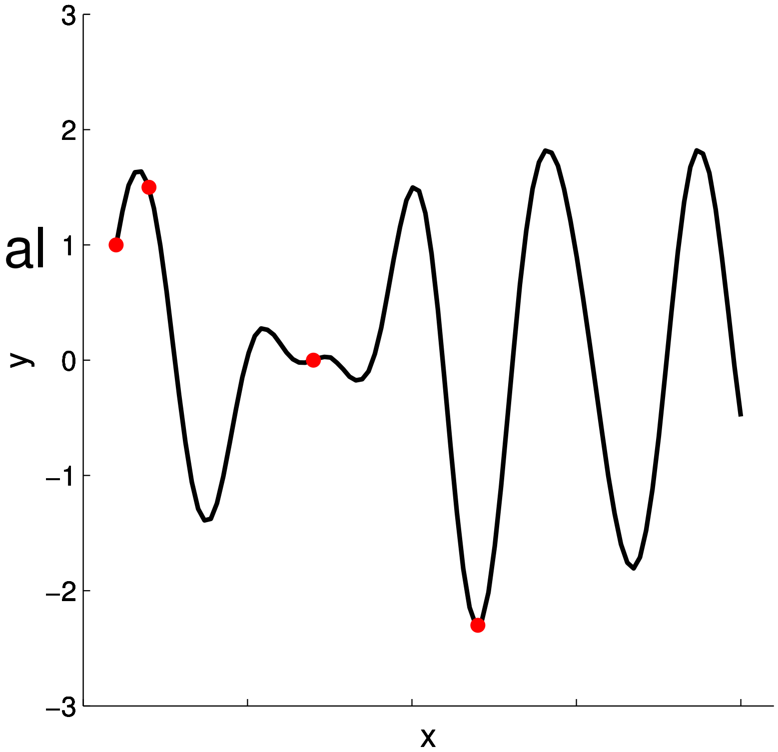


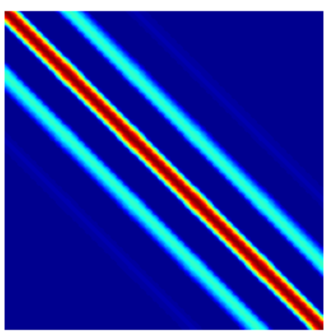
What effect does the form of the covariance function have?

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

Periodic

sinusoid $\times$ squared exponential



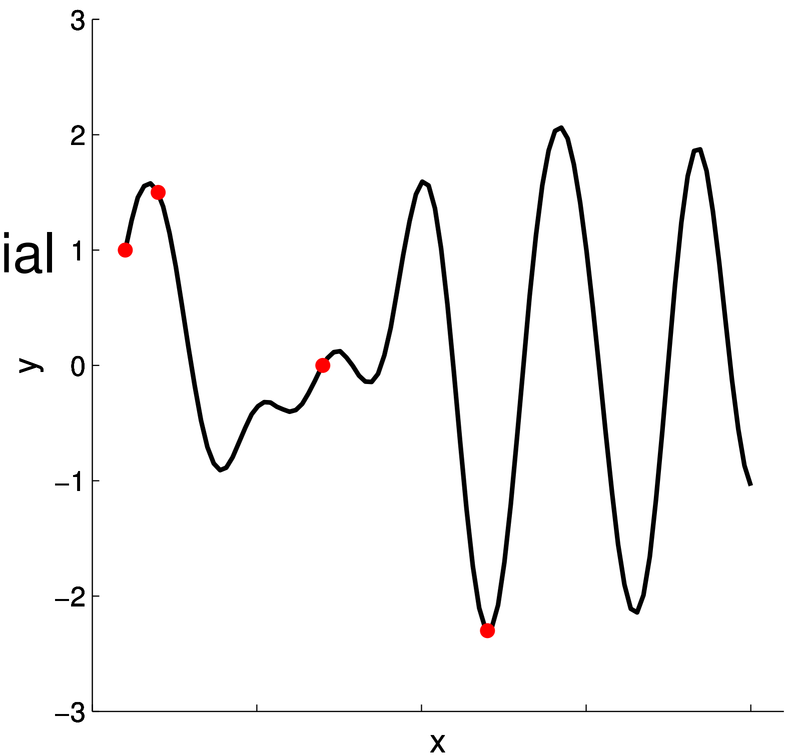
$\Sigma =$ 

What effect does the form of the covariance function have?

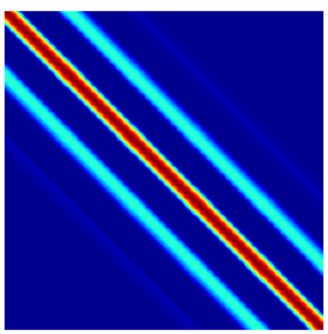
$$\mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) = \sigma^2 \cos(\omega(\mathsf{x}_1 - \mathsf{x}_2)) \exp\left(-\frac{1}{2\mathsf{l}^2}(\mathsf{x}_1 - \mathsf{x}_2)^2\right)$$

Periodic

sinusoid $\times$ squared exponential



$\Sigma =$

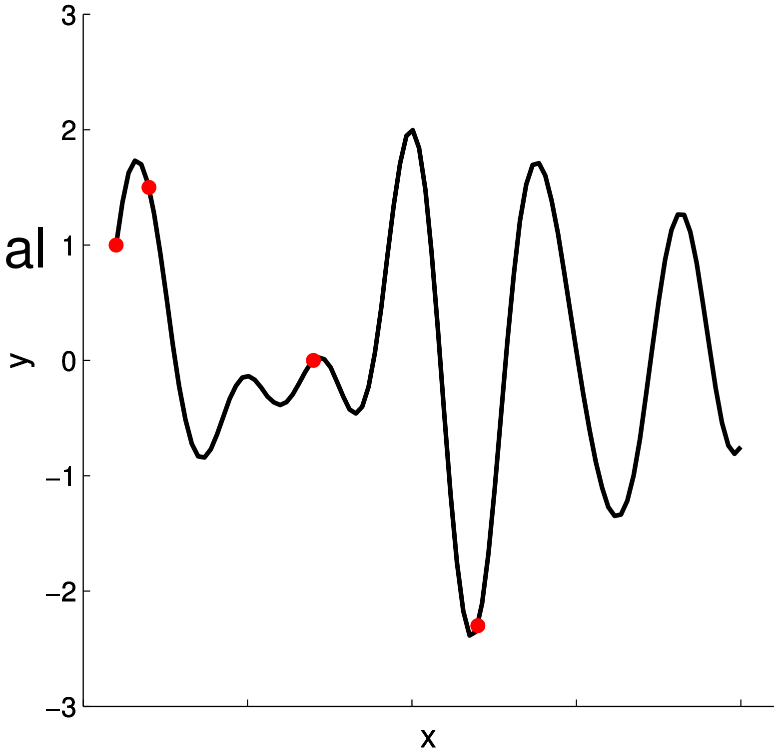


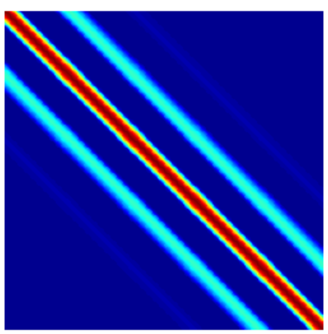
What effect does the form of the covariance function have?

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

Periodic

sinusoid $\times$ squared exponential



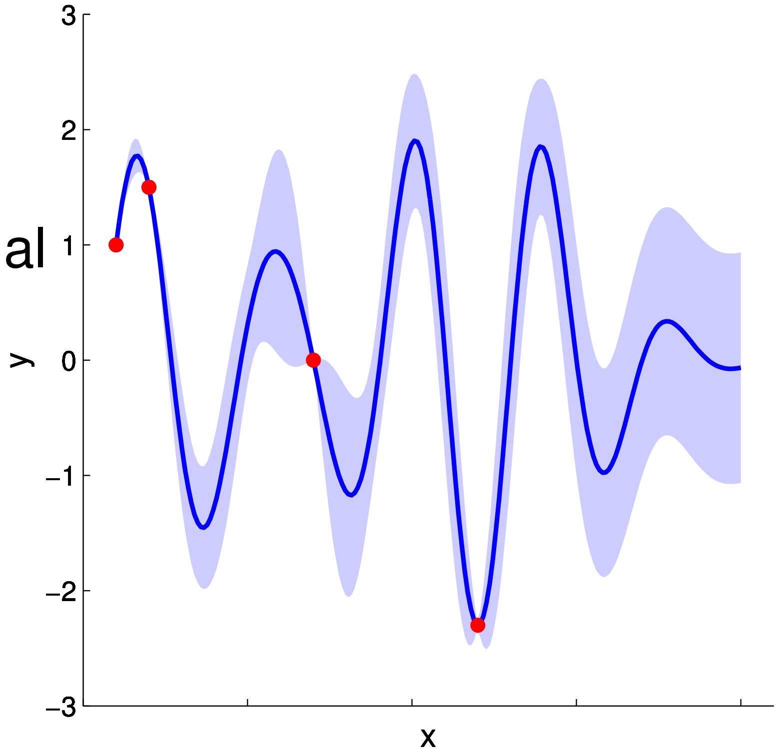
$\Sigma =$ 

What effect does the form of the covariance function have?

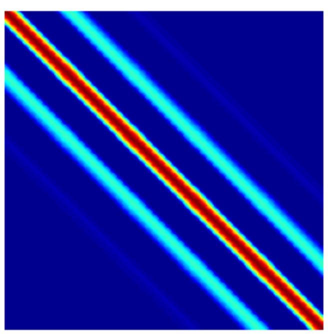
$$K(x_1, x_2) = \sigma^2 \cos(\omega(x_1 - x_2)) \exp\left(-\frac{1}{2l^2}(x_1 - x_2)^2\right)$$

Periodic

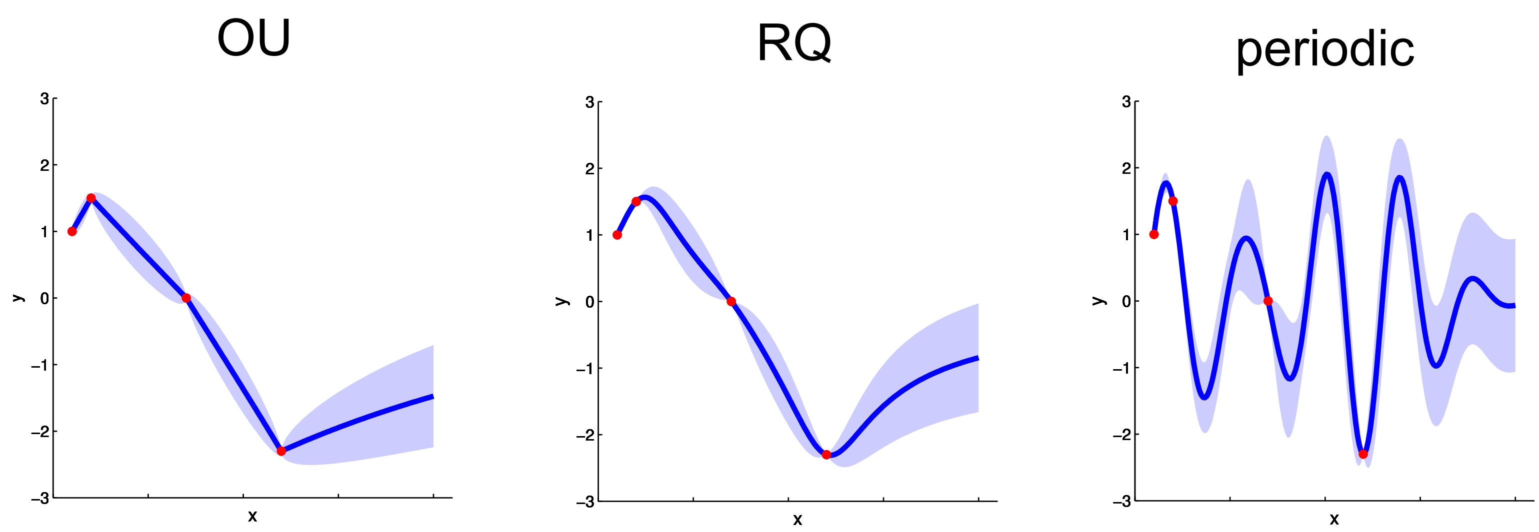
sinusoid $\times$ squared exponential



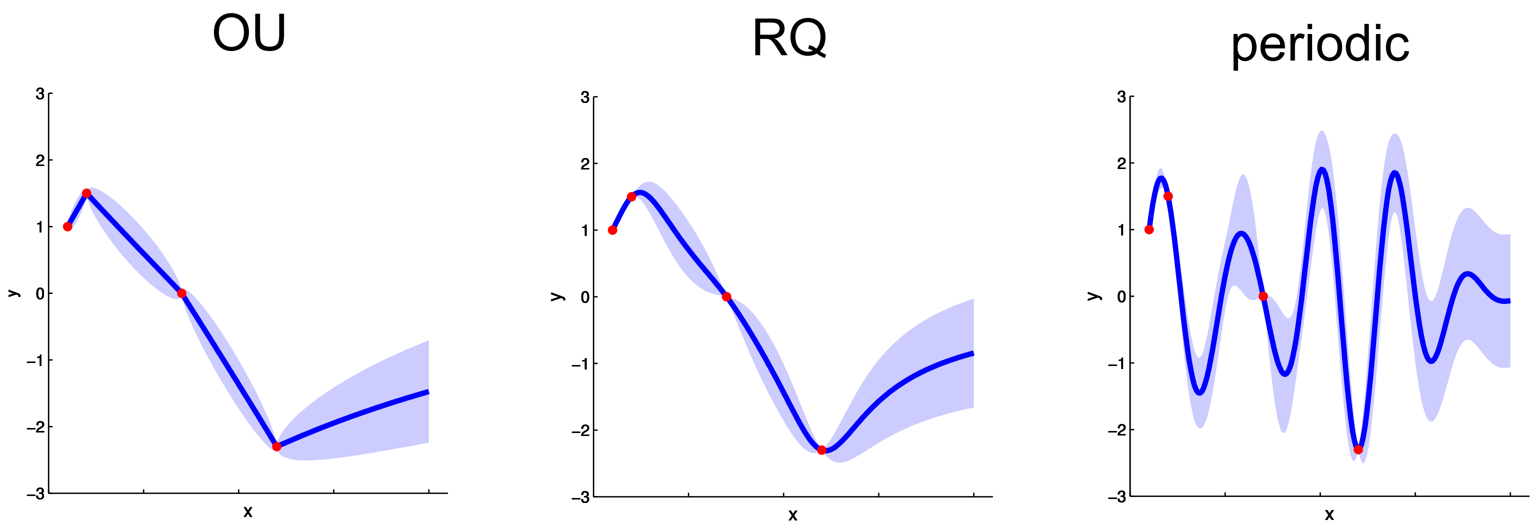
$\Sigma =$



The covariance function has a large effect



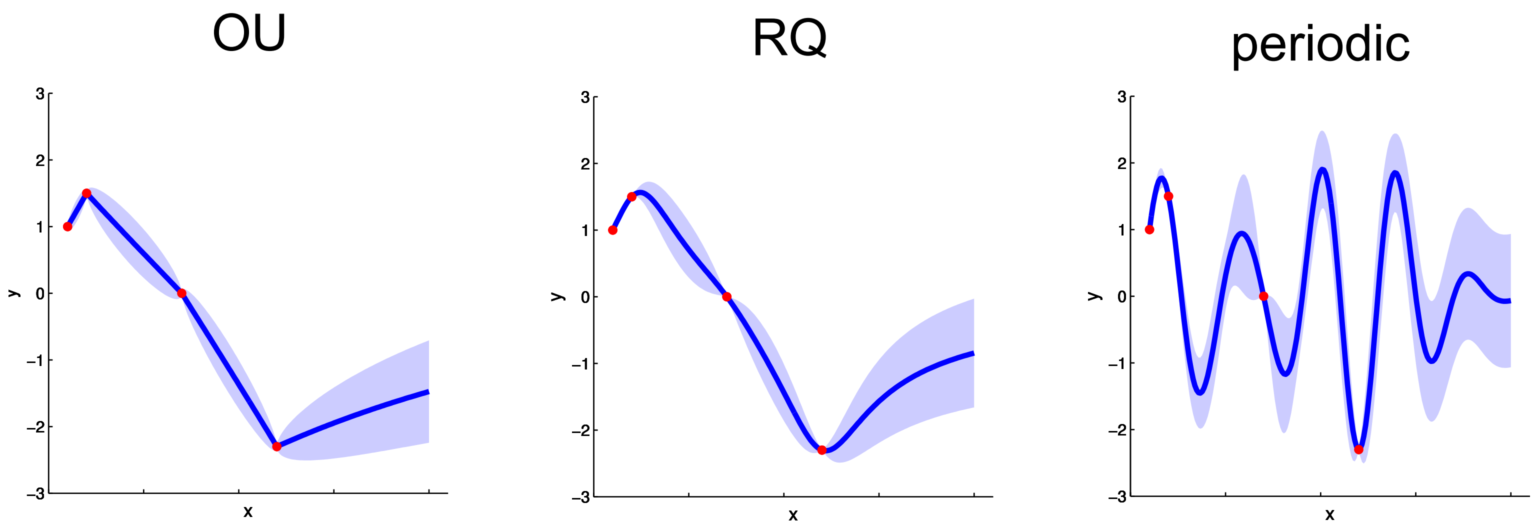
The covariance function has a large effect



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

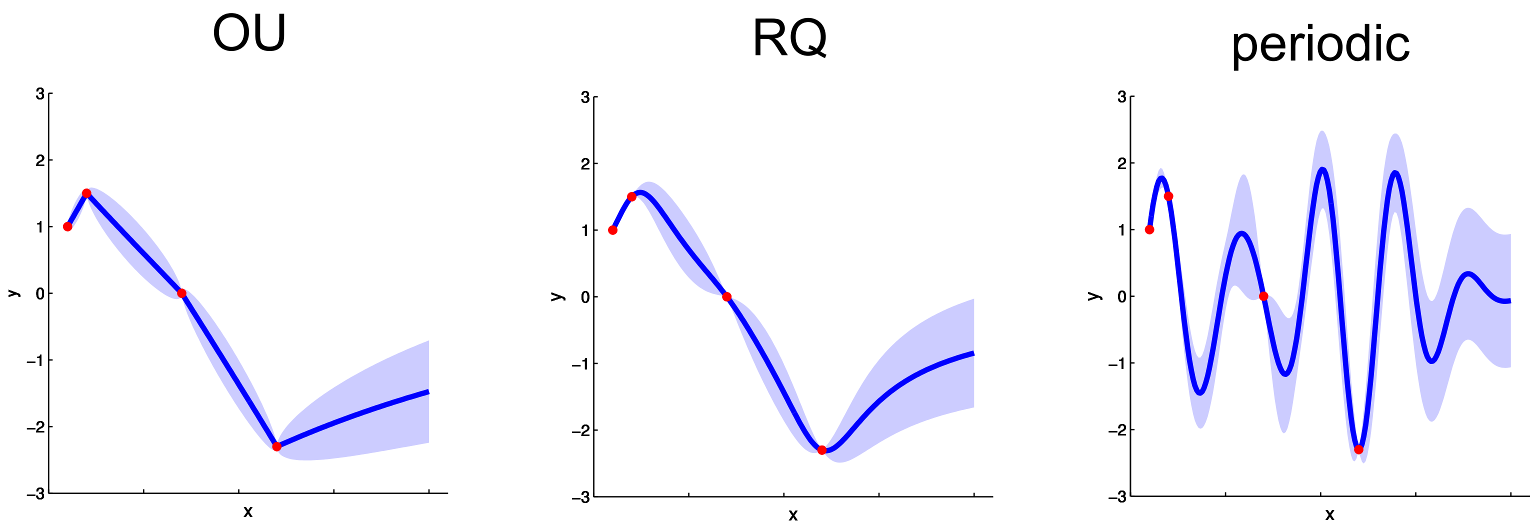


Bayesian model comparison:

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

prior over models

The covariance function has a large effect



Bayesian model comparison:

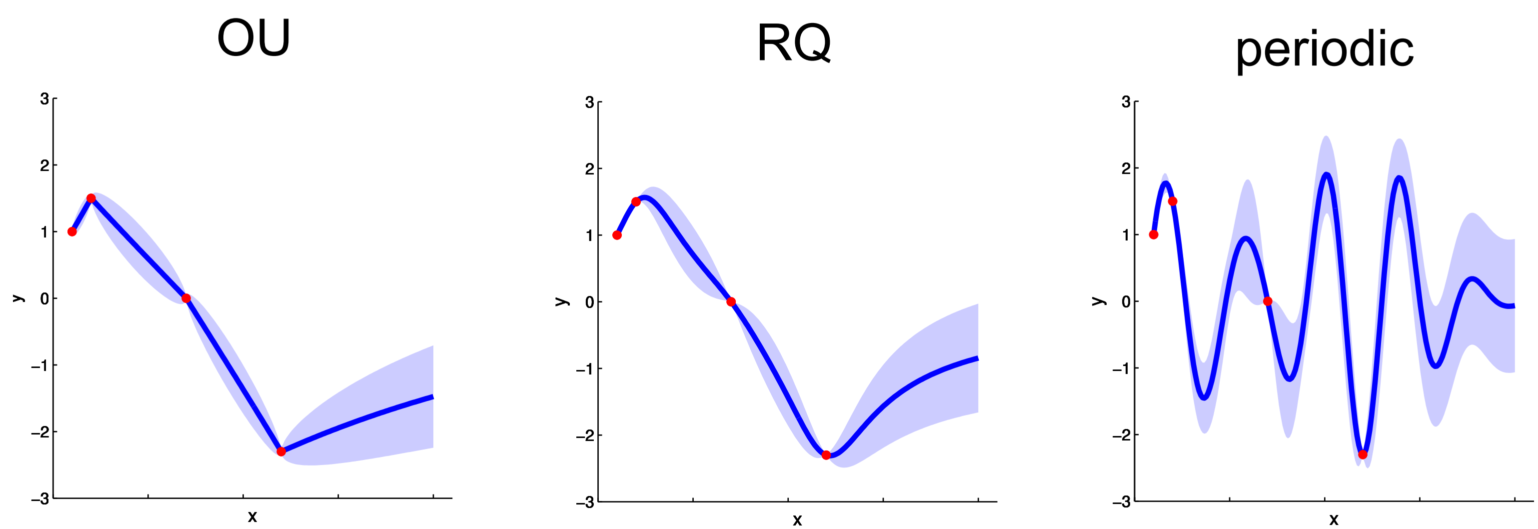
prior over models

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

marginal likelihood

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

The covariance function has a large effect



Bayesian model comparison:

prior over models

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

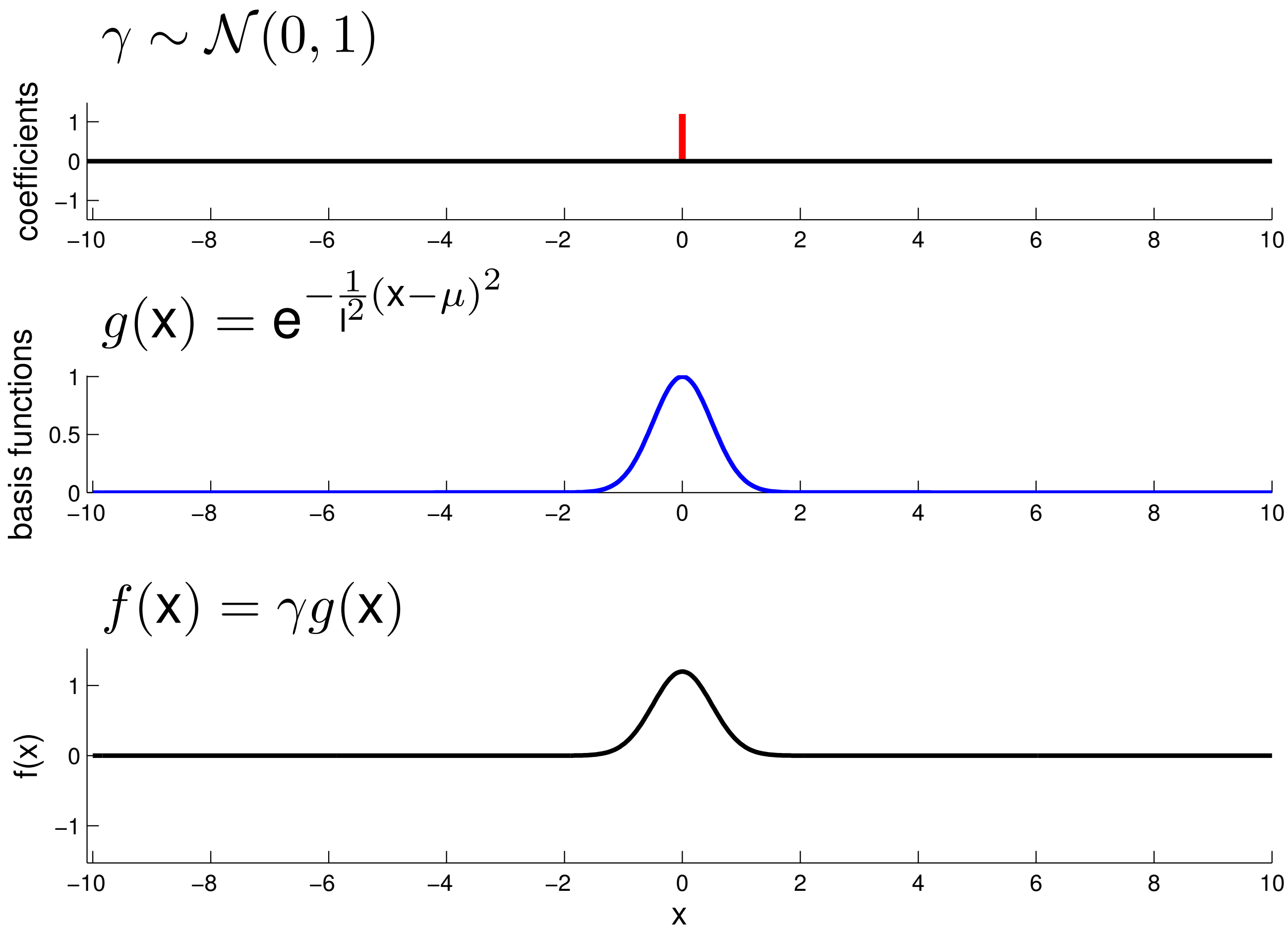
marginal likelihood

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

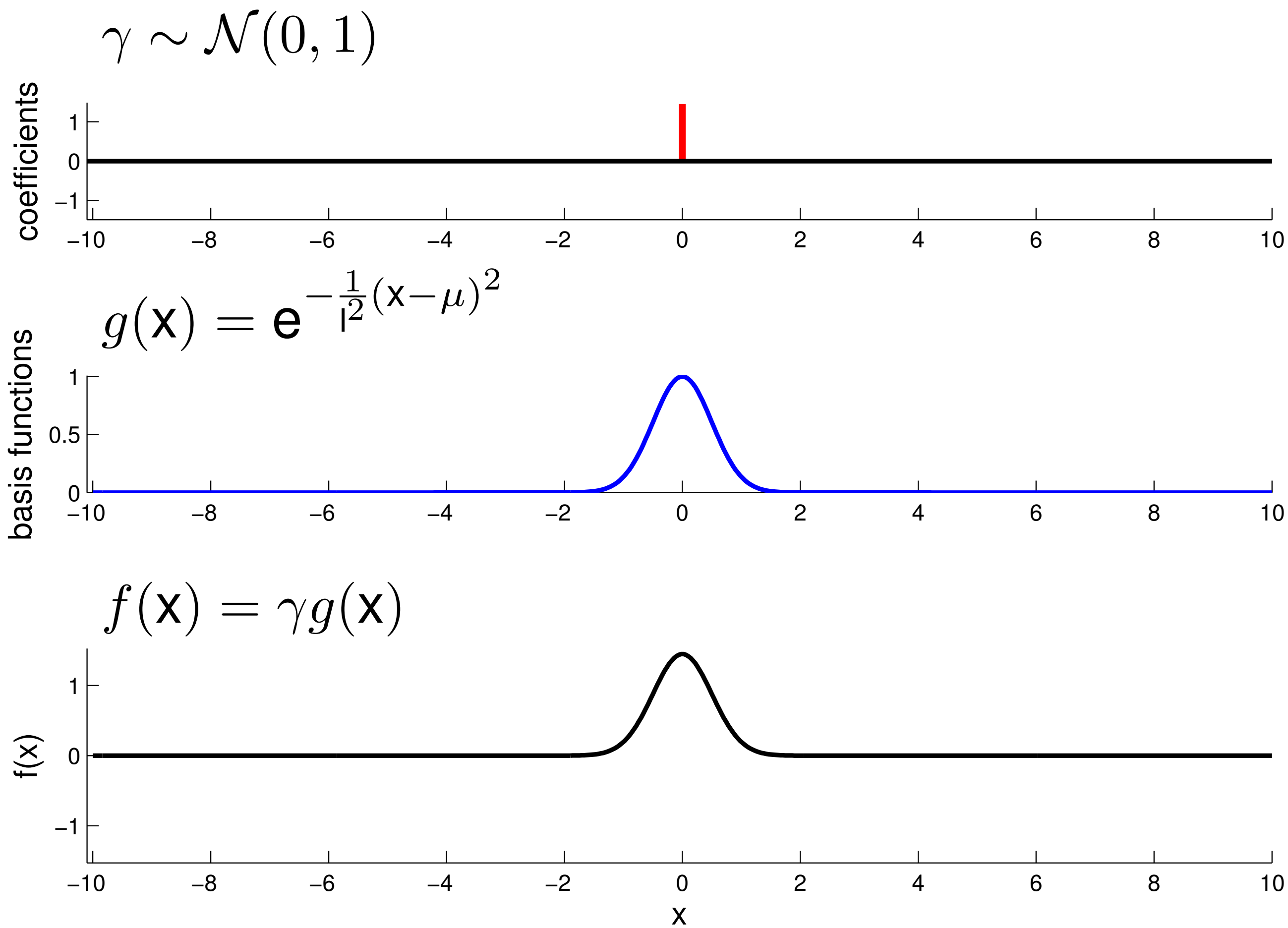
Health warnings: Hard to compute (need approximations)

Often results are very sensitive to the priors $p(\theta|M)$

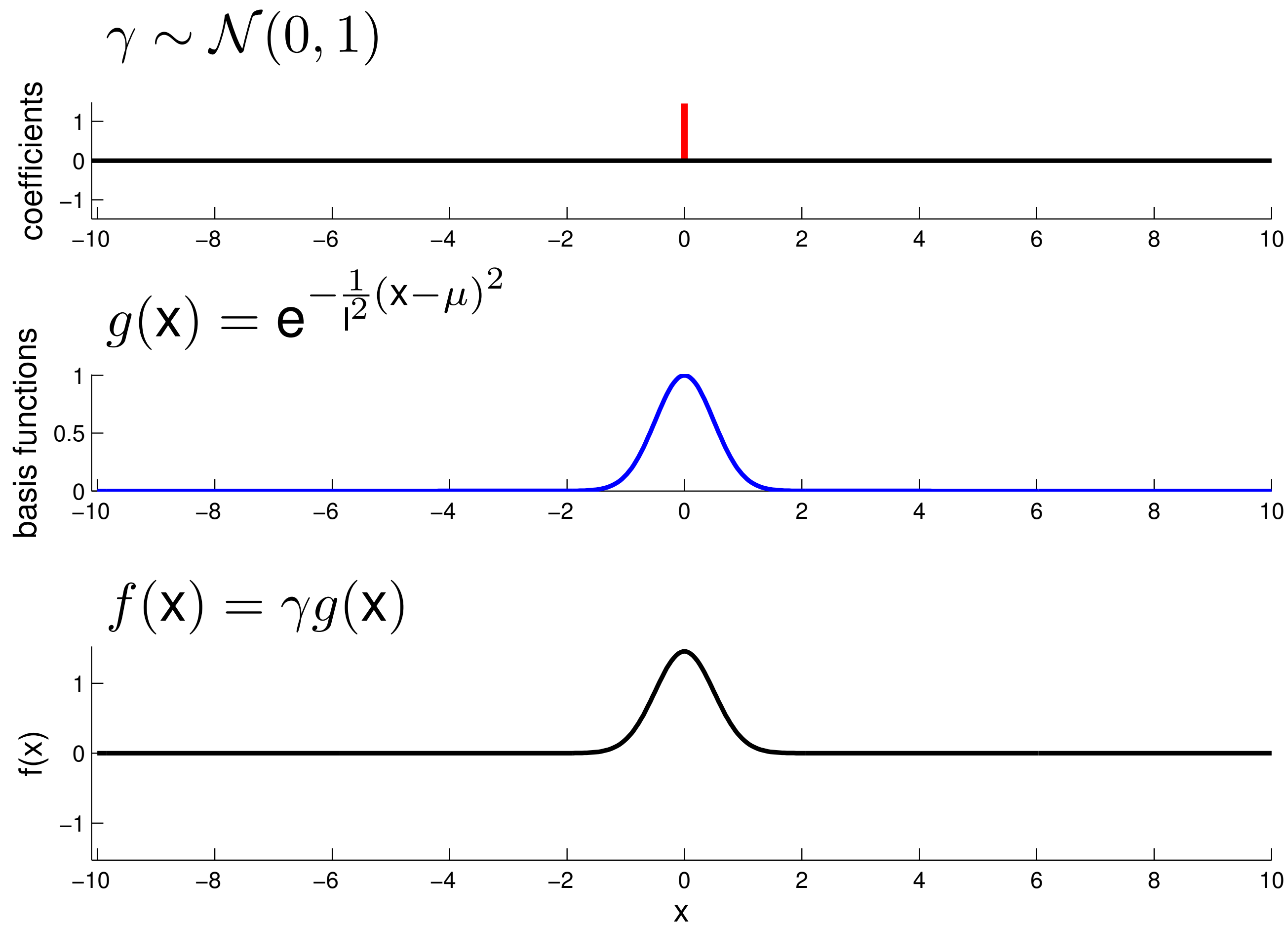
Basis function view of Gaussian processes



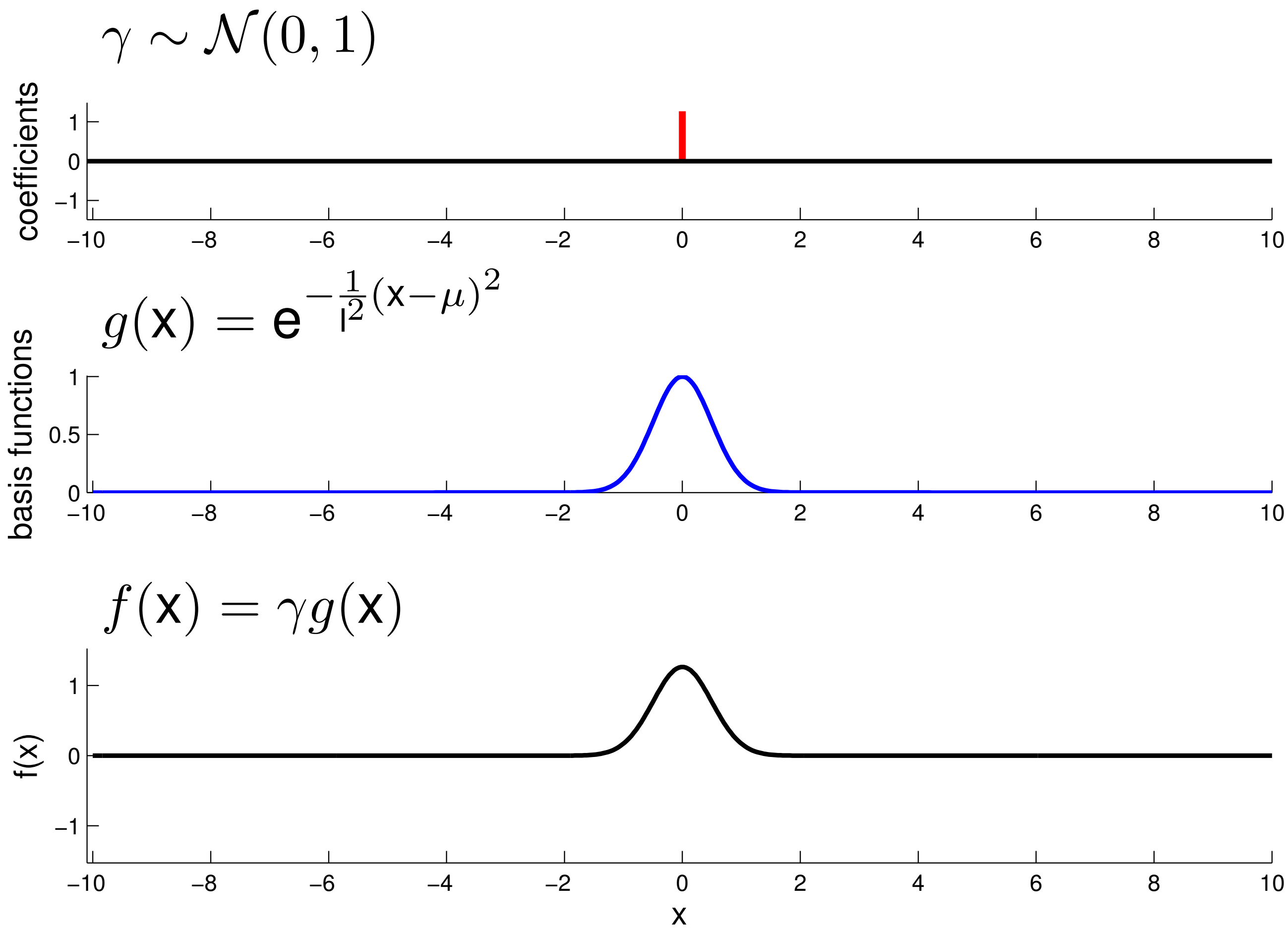
Basis function view of Gaussian processes



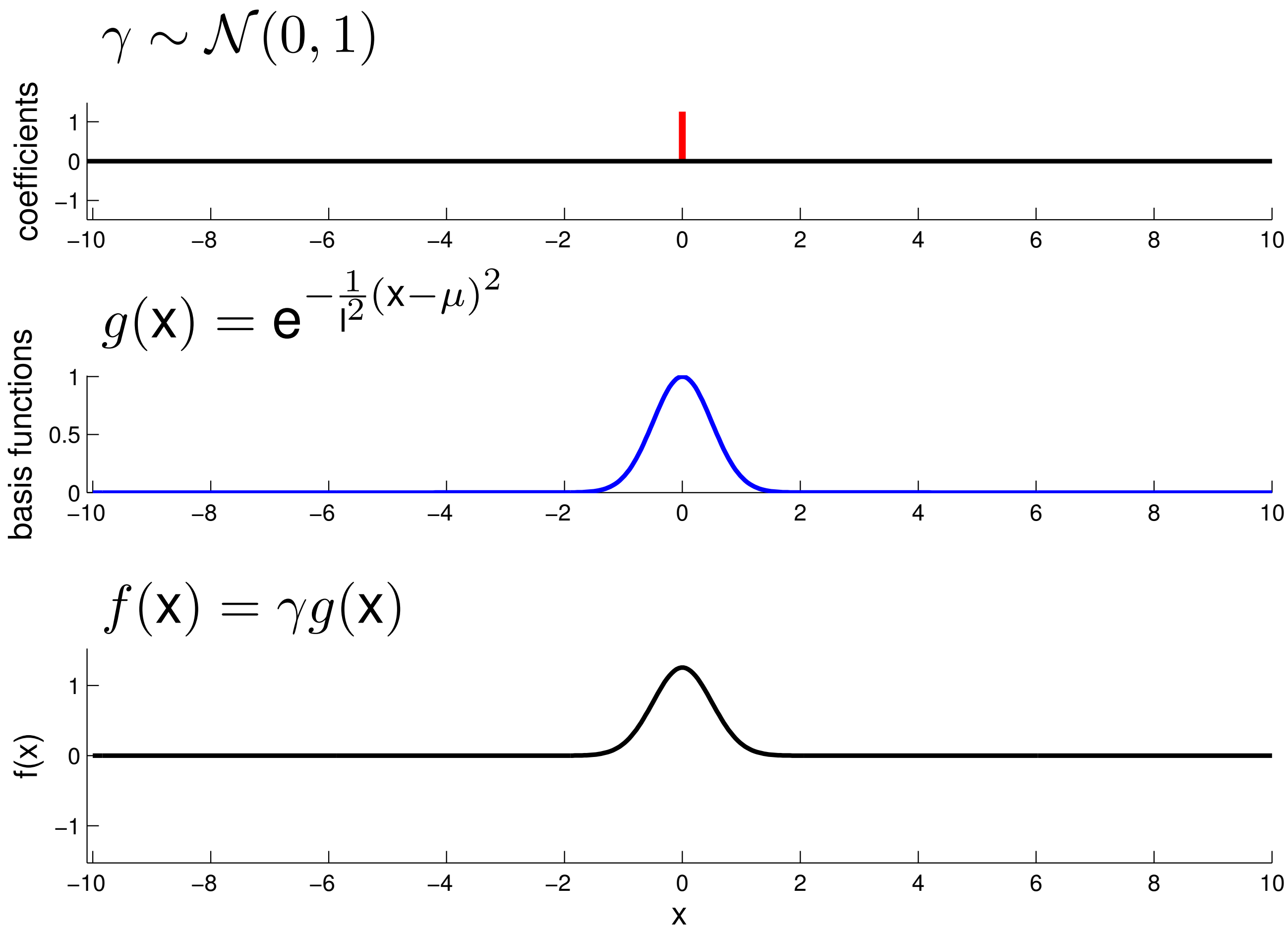
Basis function view of Gaussian processes



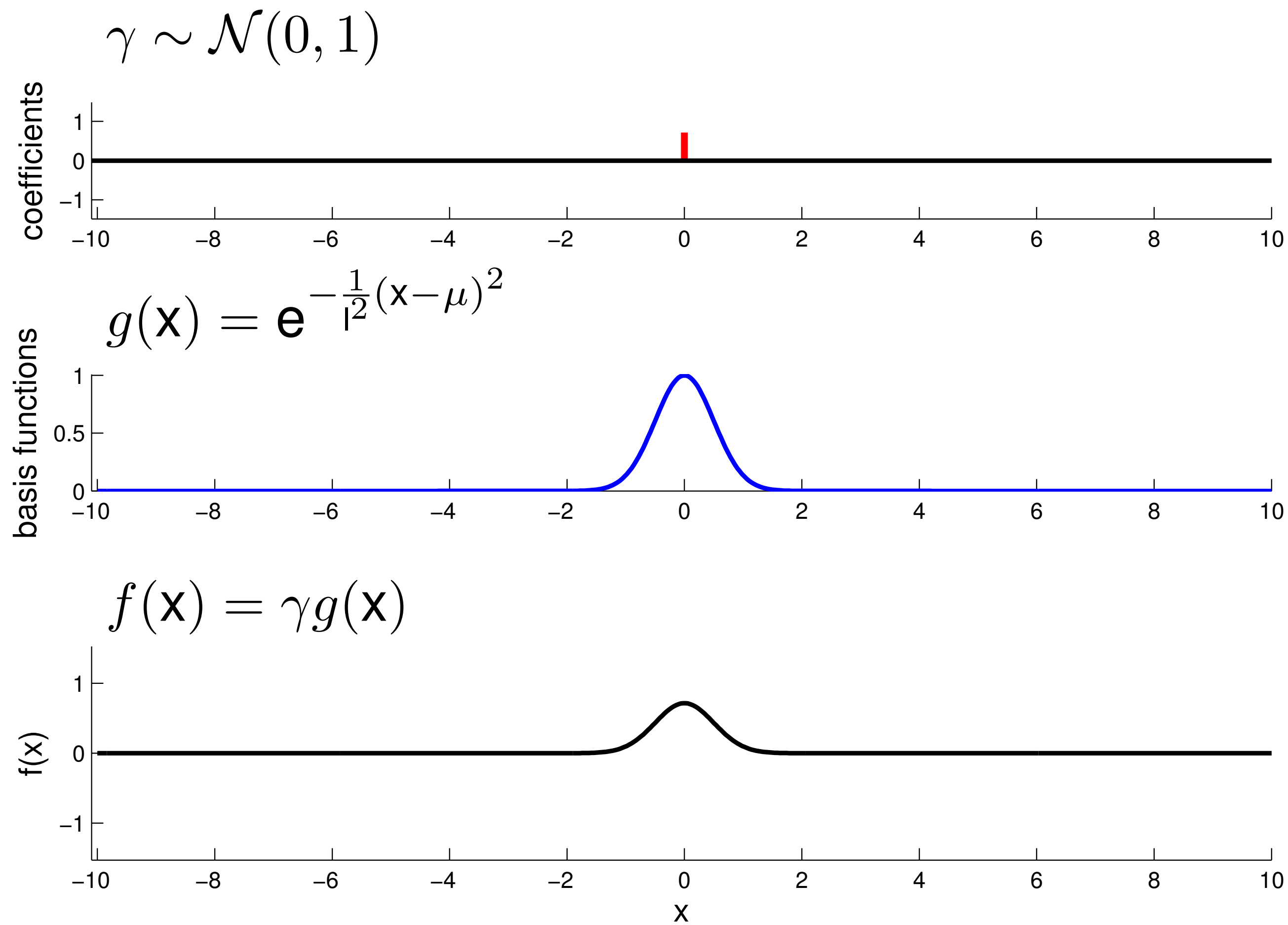
Basis function view of Gaussian processes



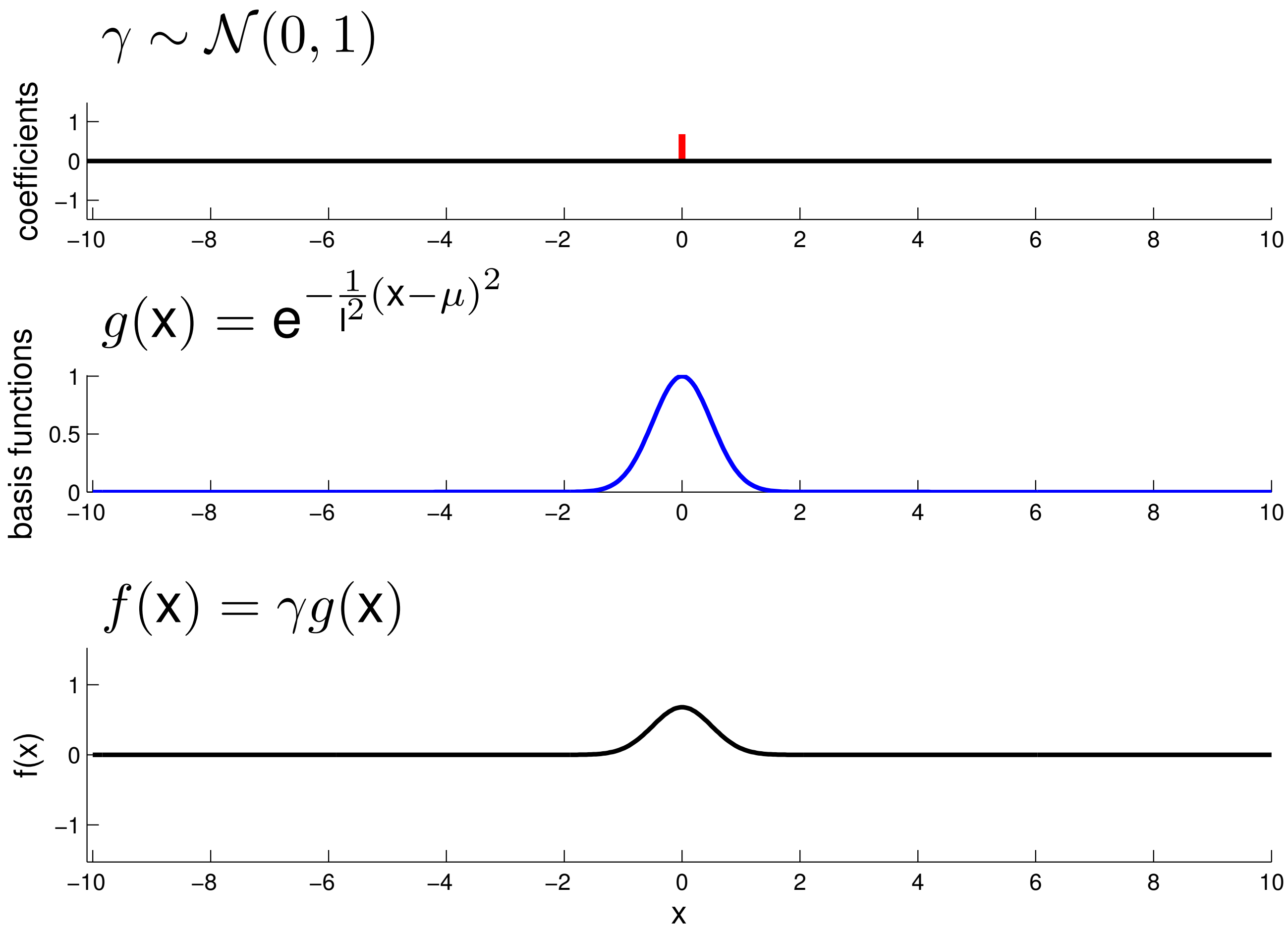
Basis function view of Gaussian processes



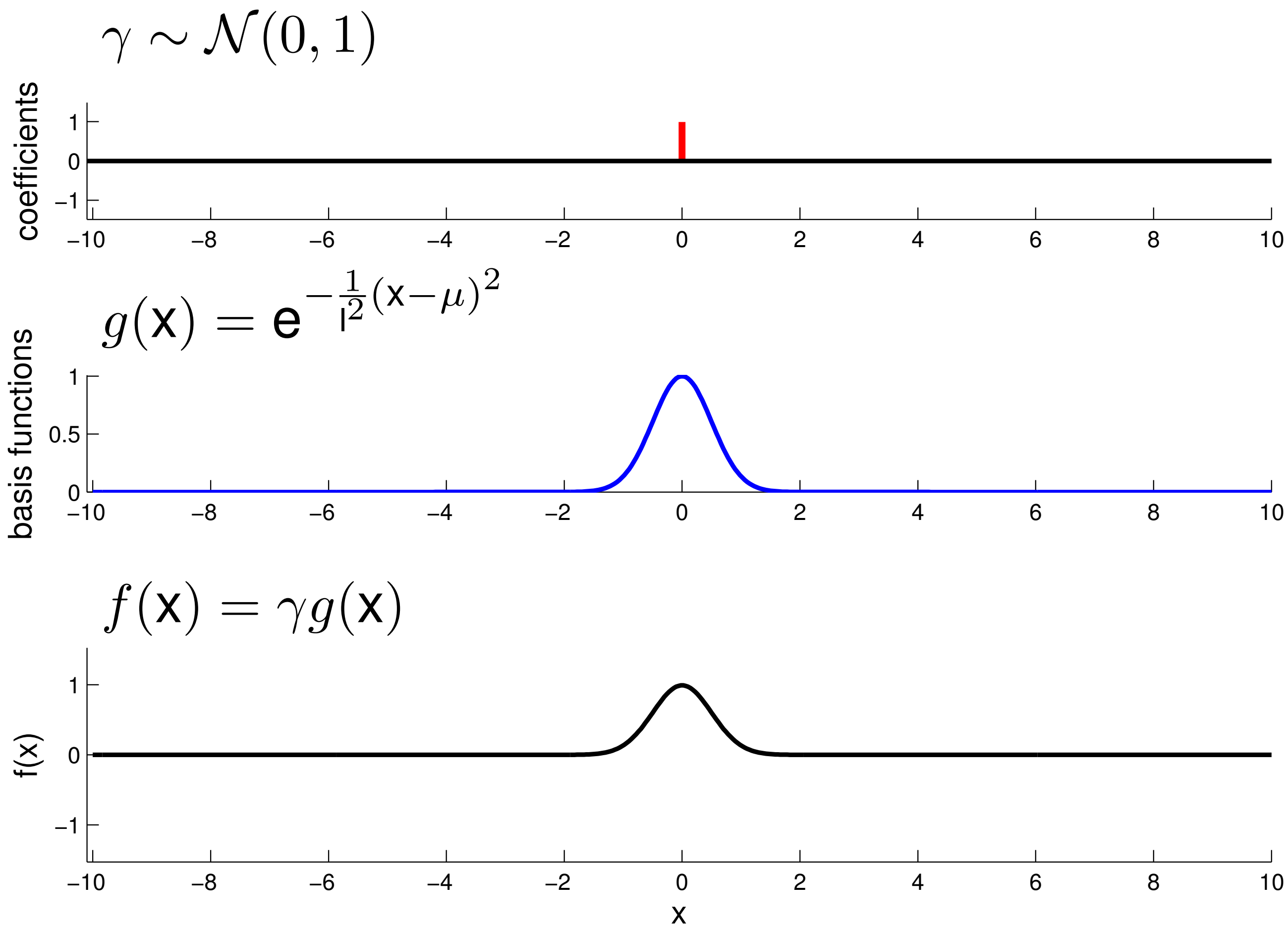
Basis function view of Gaussian processes



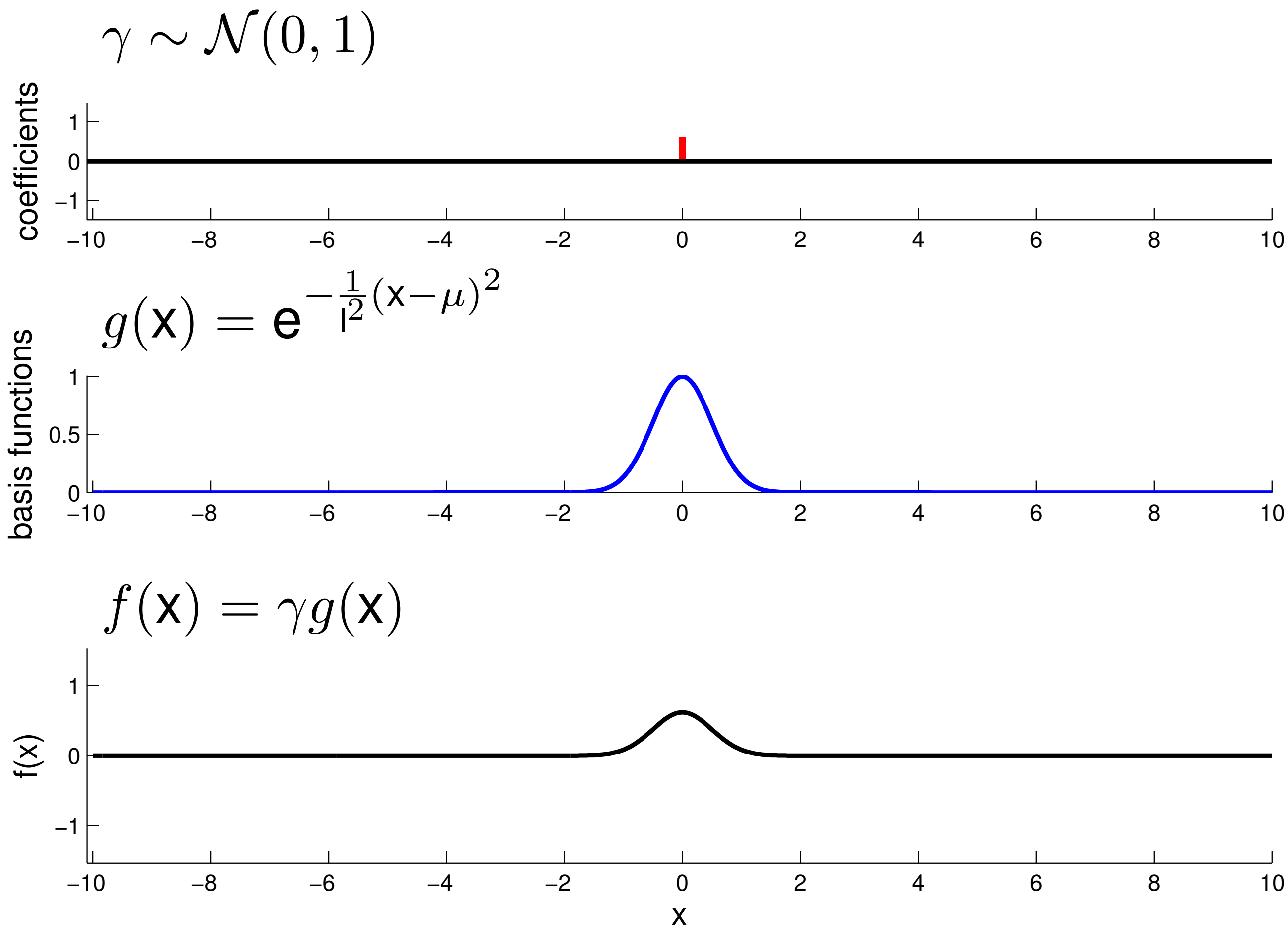
Basis function view of Gaussian processes



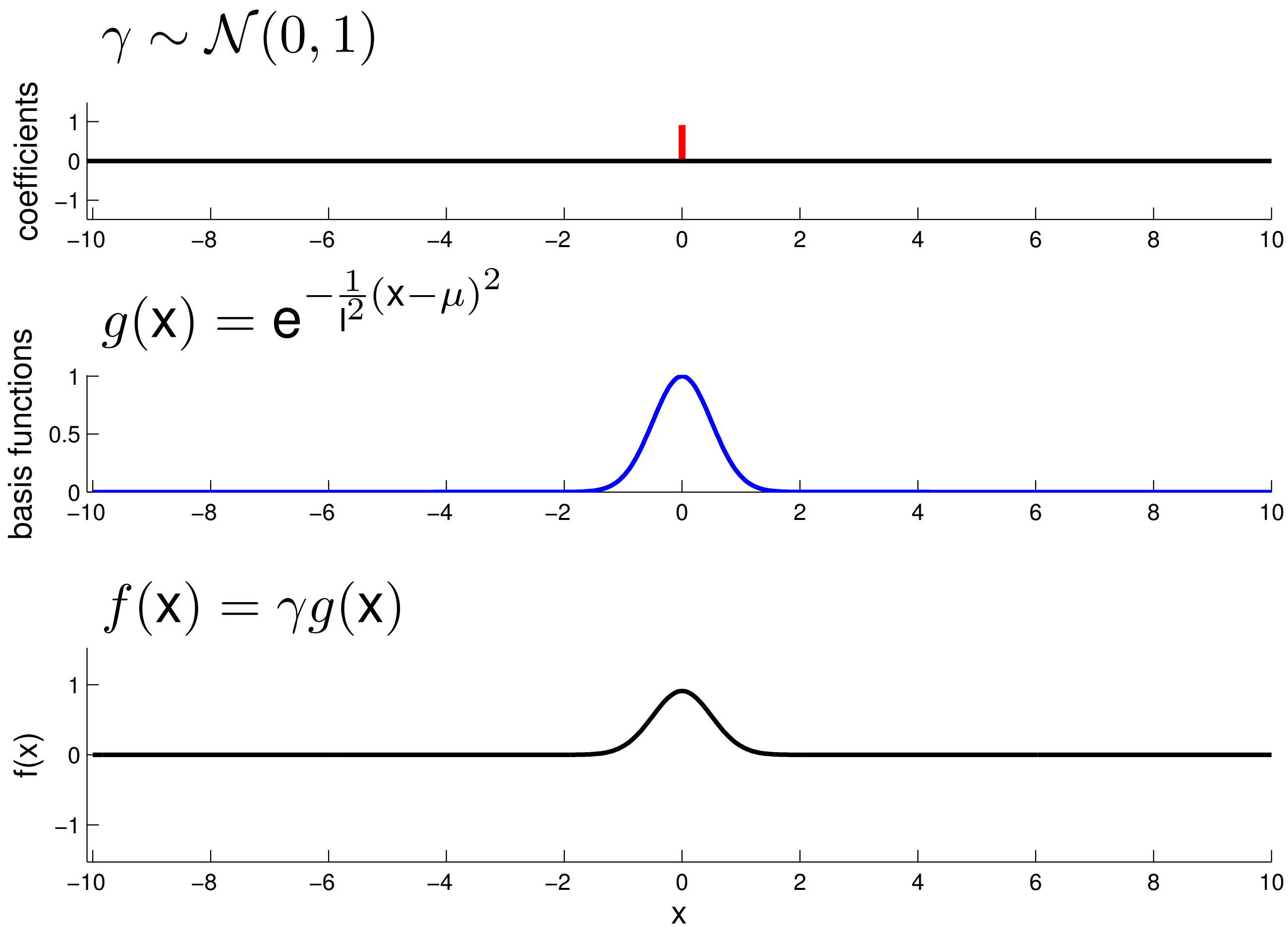
Basis function view of Gaussian processes



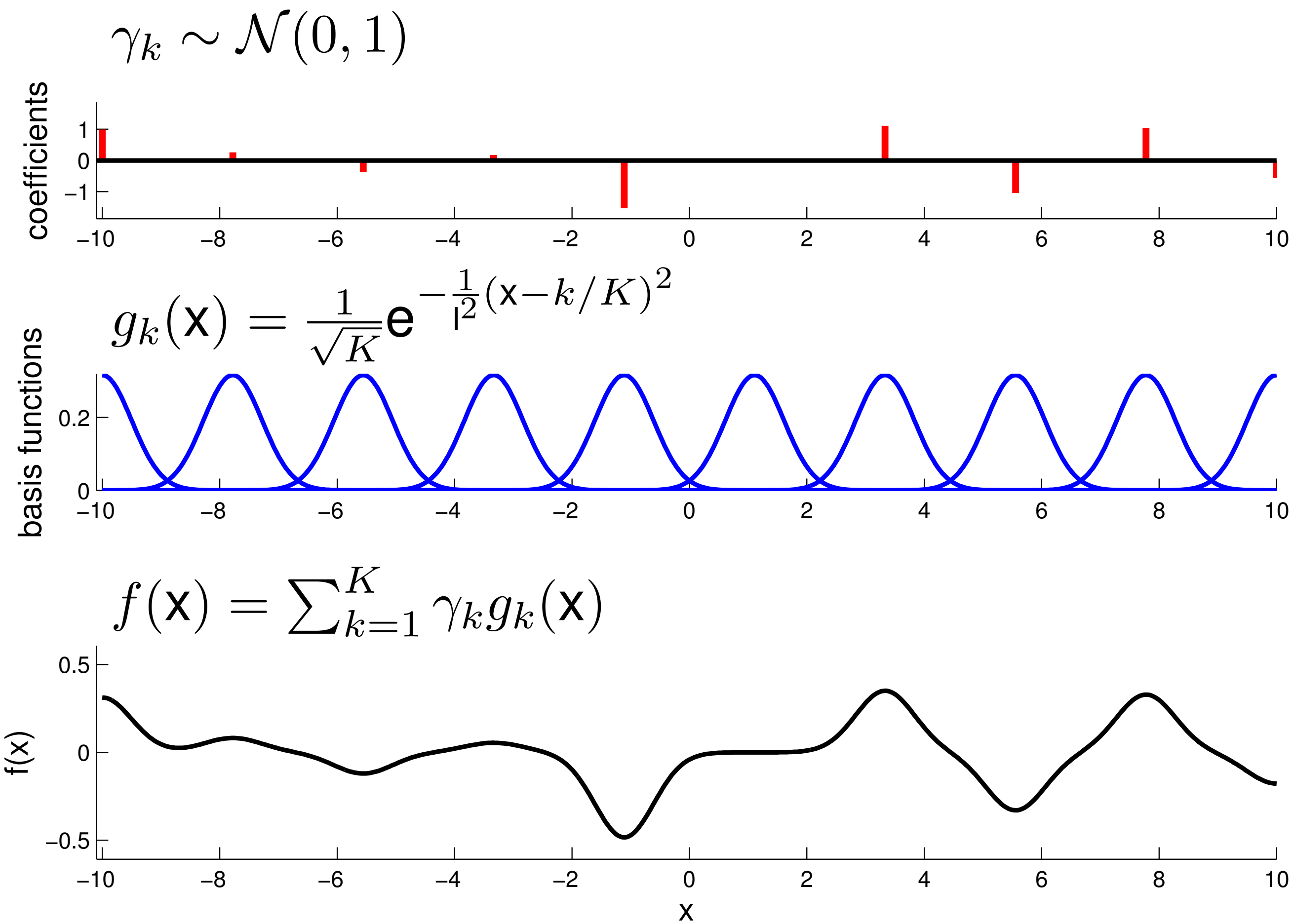
Basis function view of Gaussian processes



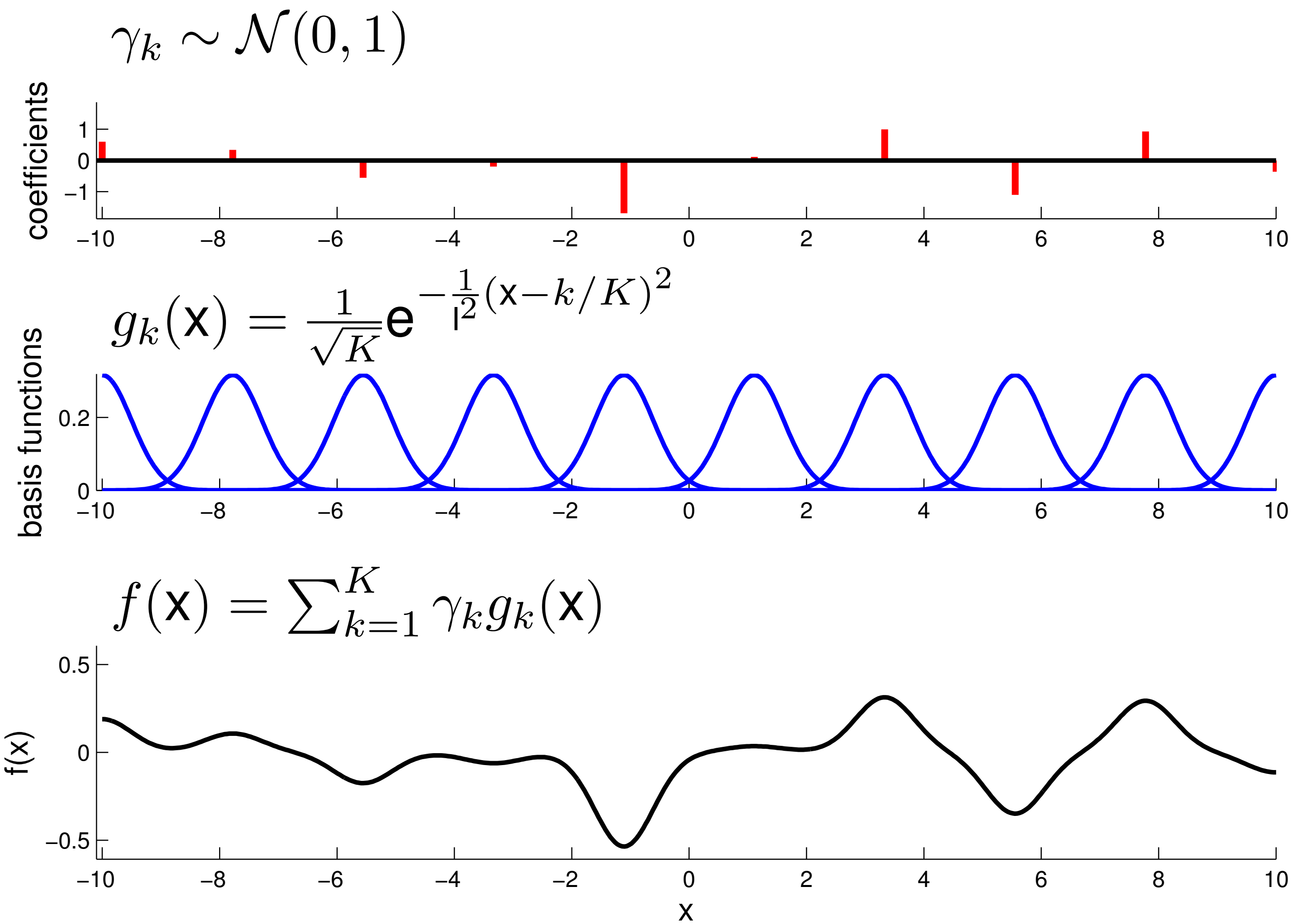
Basis function view of Gaussian processes



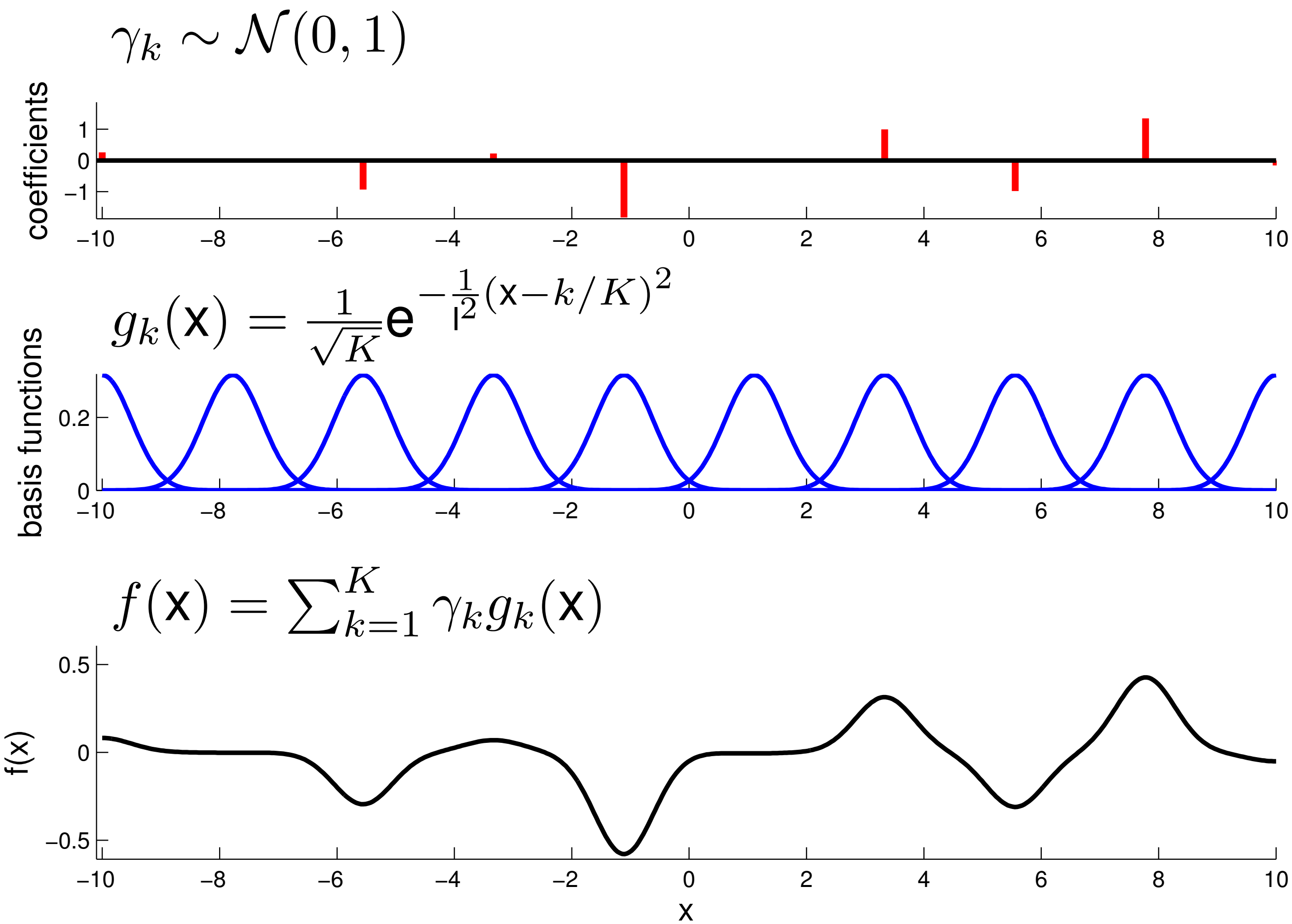
Basis function view of Gaussian processes



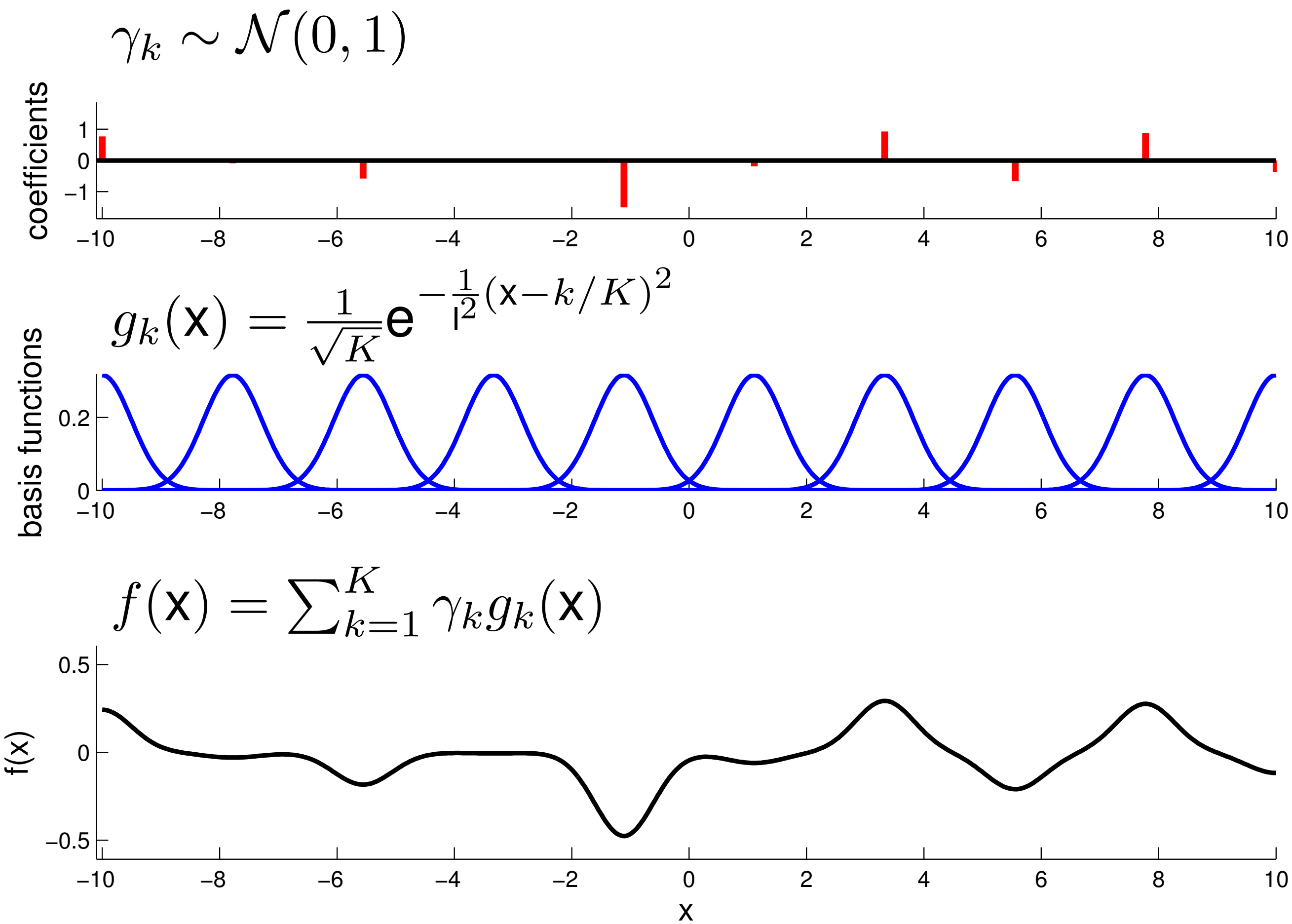
Basis function view of Gaussian processes



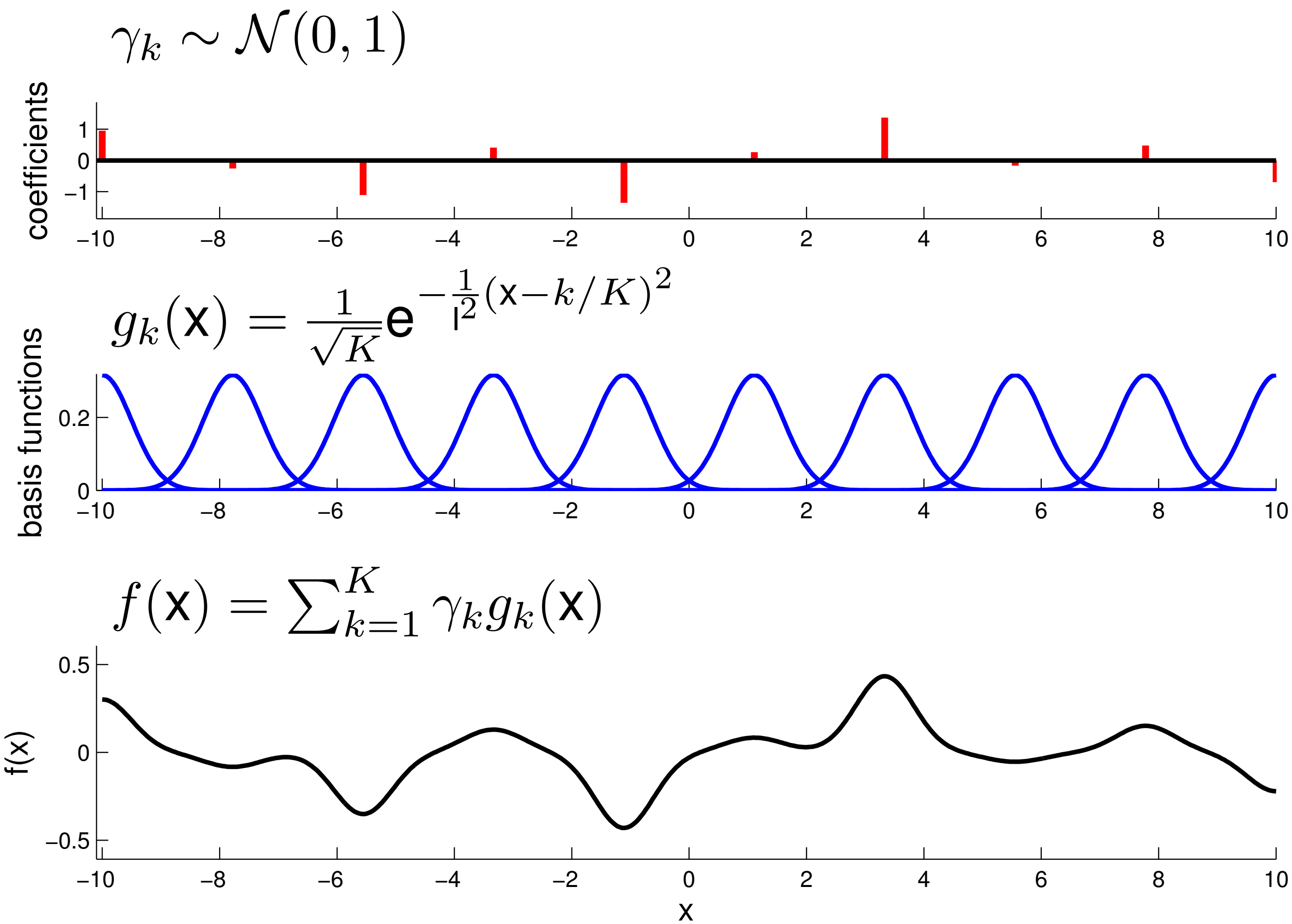
Basis function view of Gaussian processes



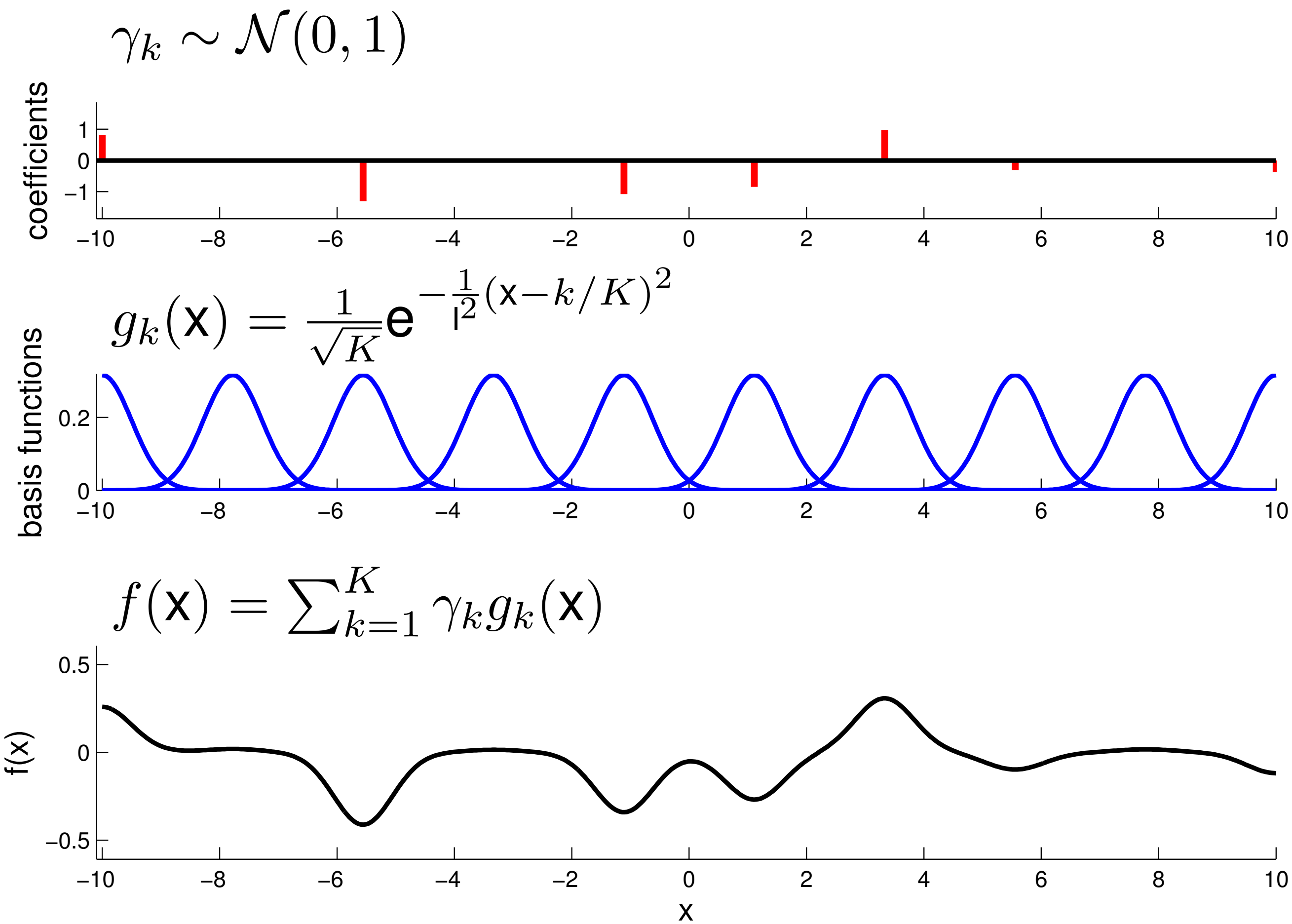
Basis function view of Gaussian processes



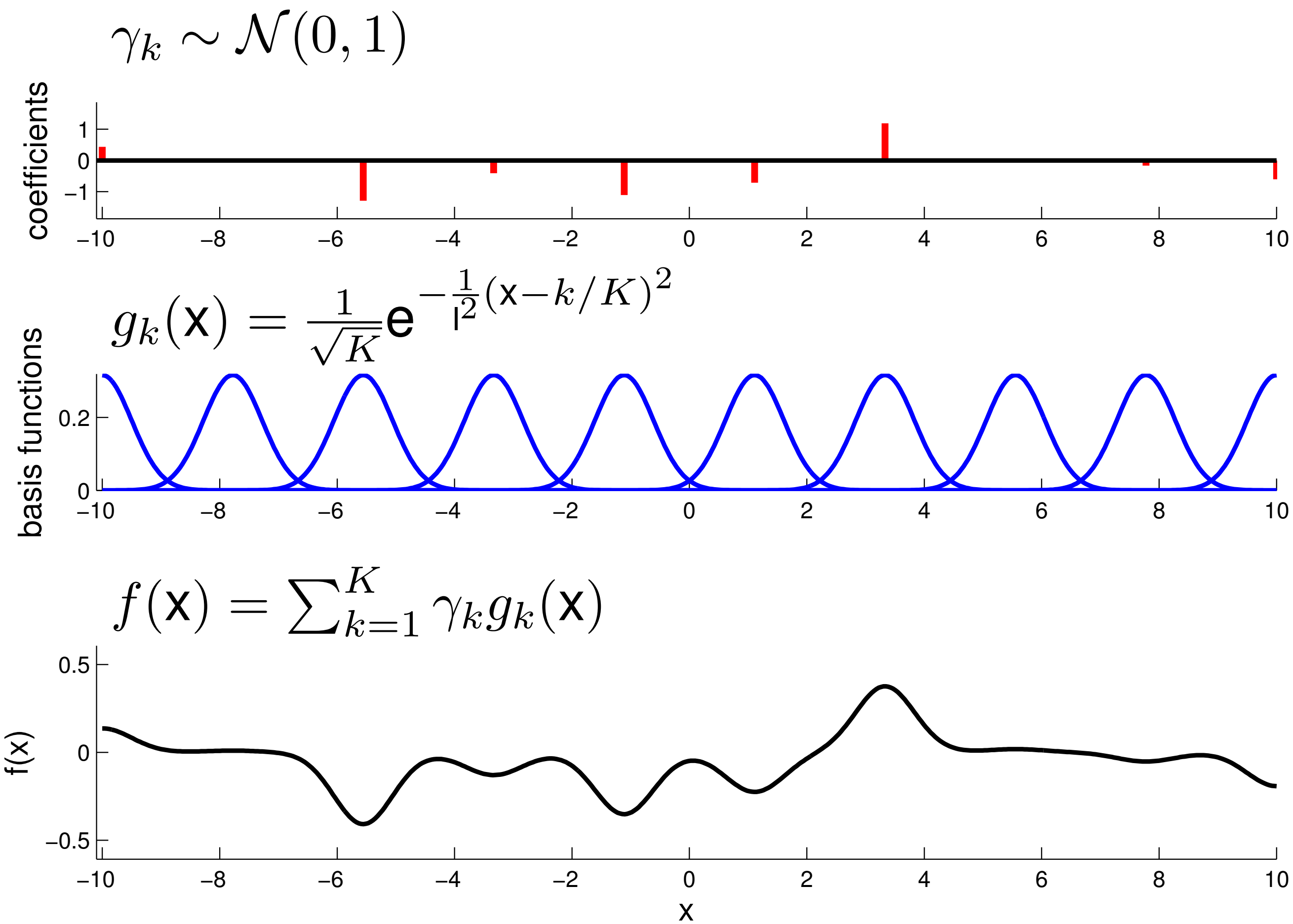
Basis function view of Gaussian processes



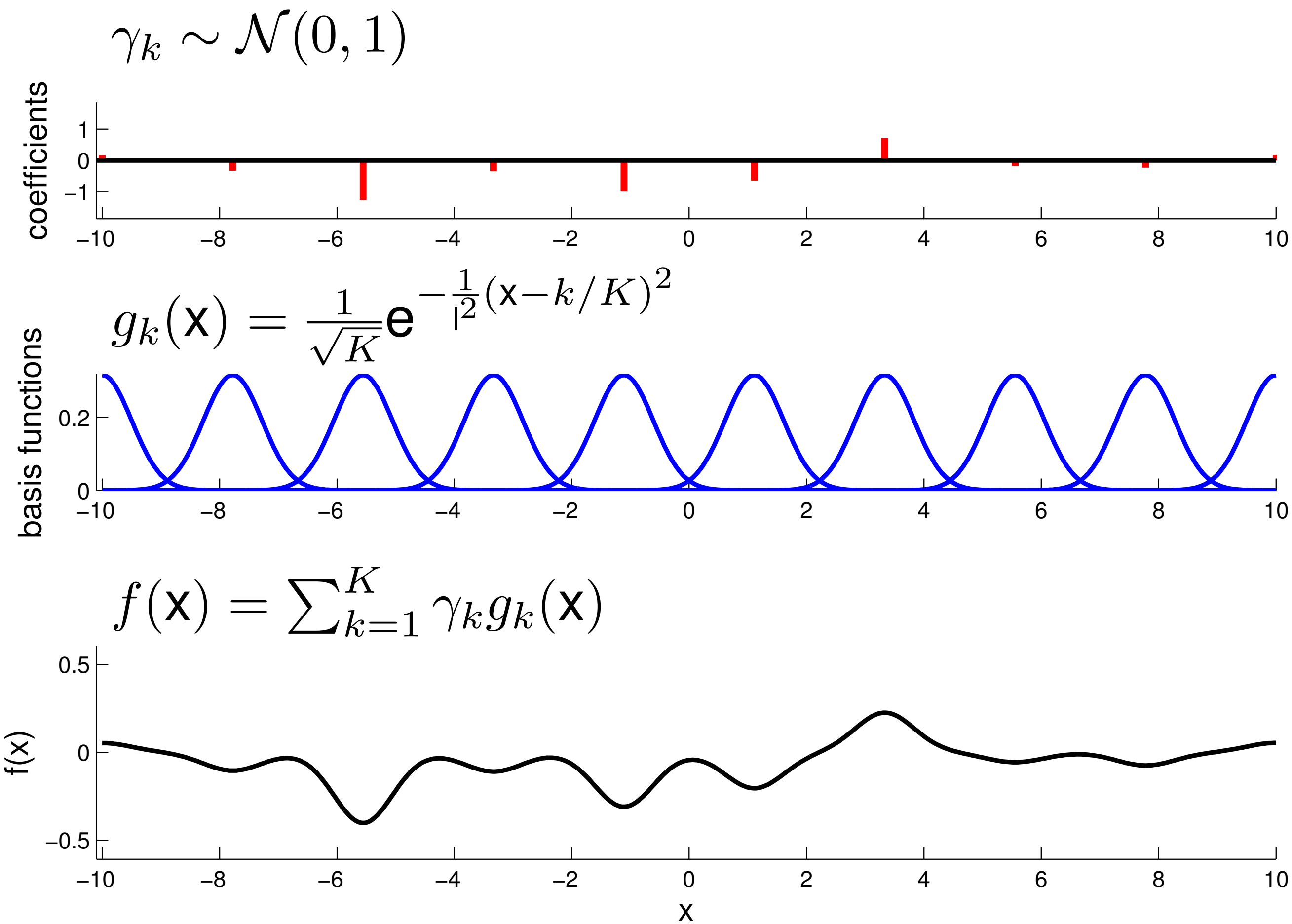
Basis function view of Gaussian processes



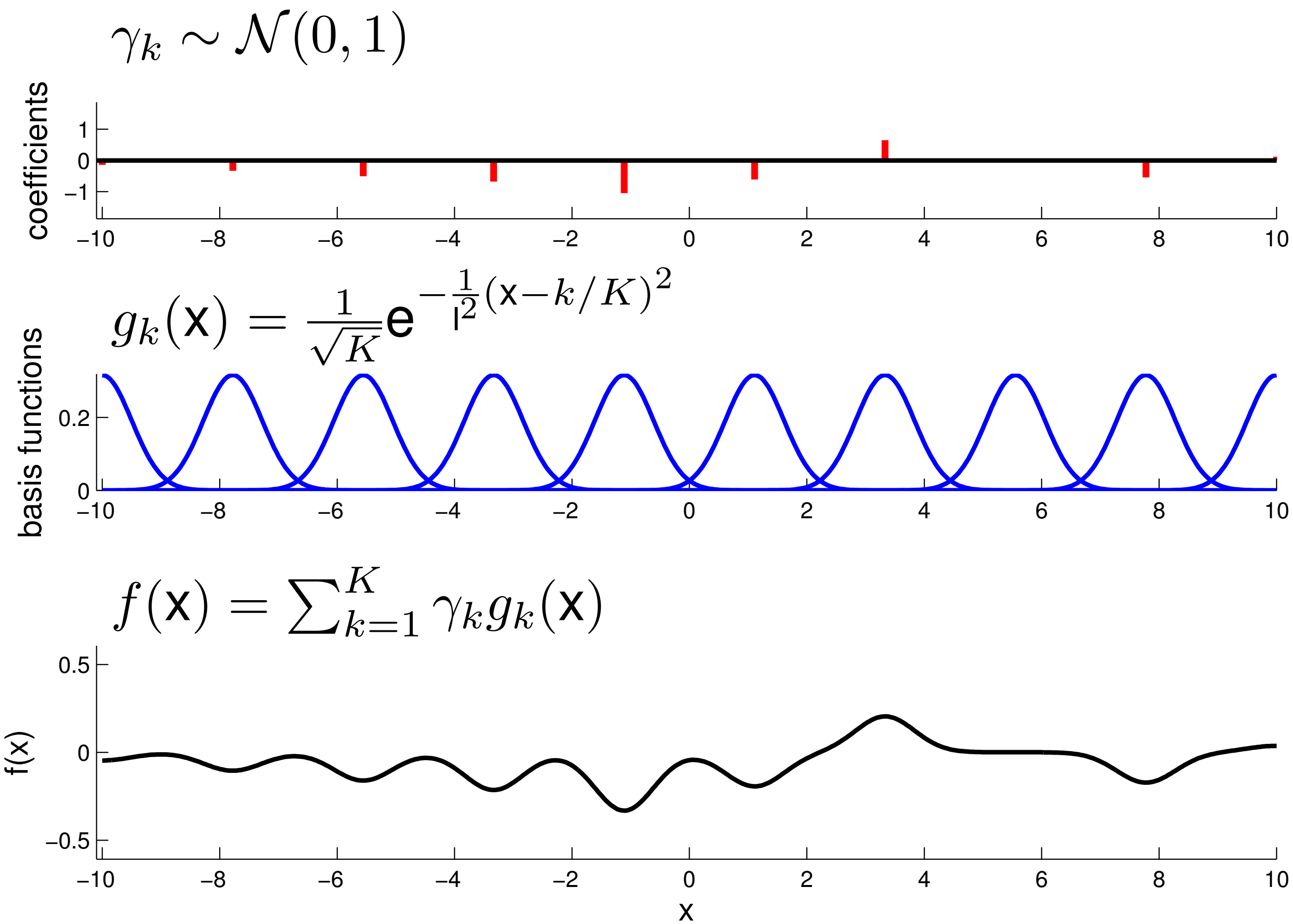
Basis function view of Gaussian processes



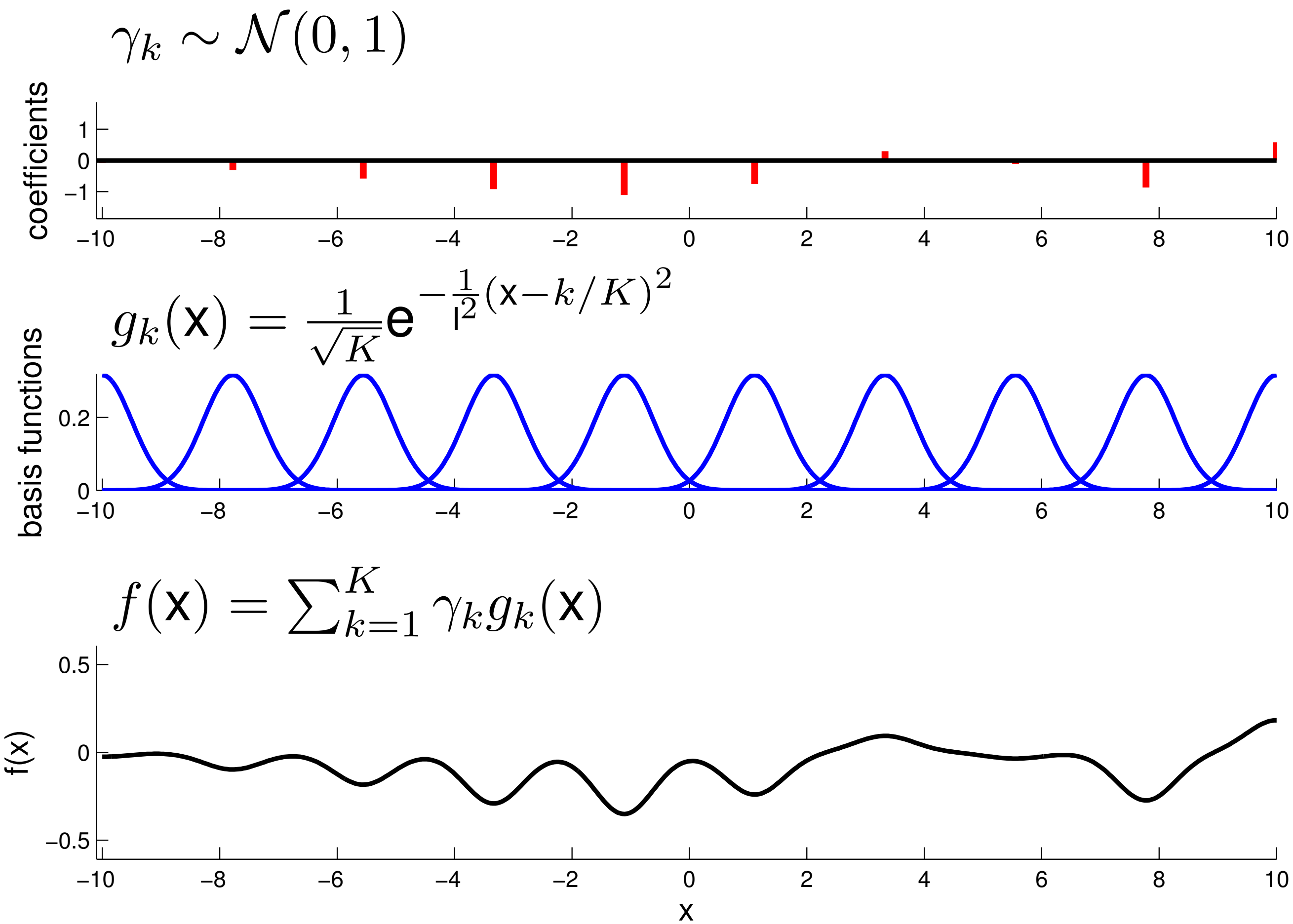
Basis function view of Gaussian processes



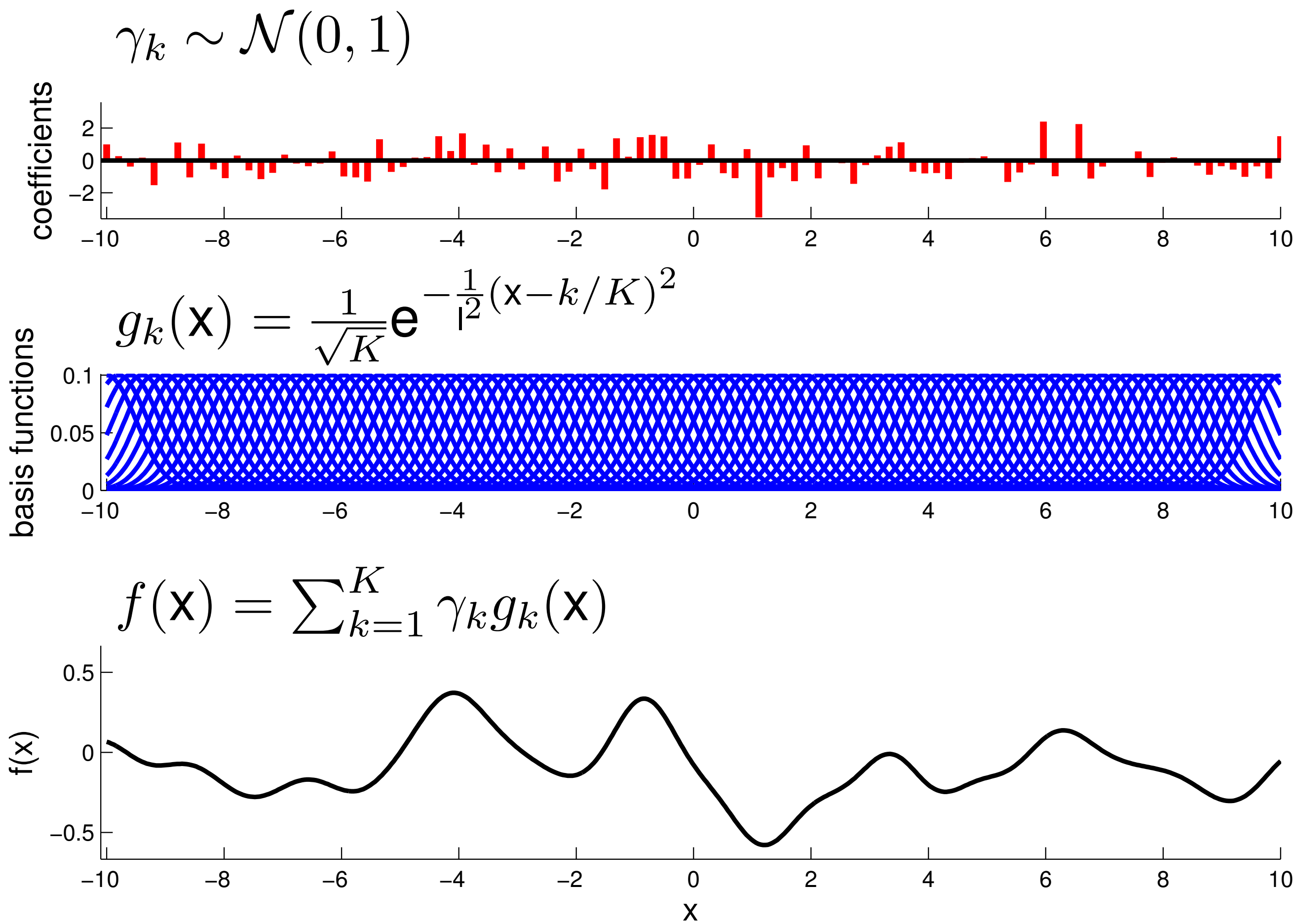
Basis function view of Gaussian processes



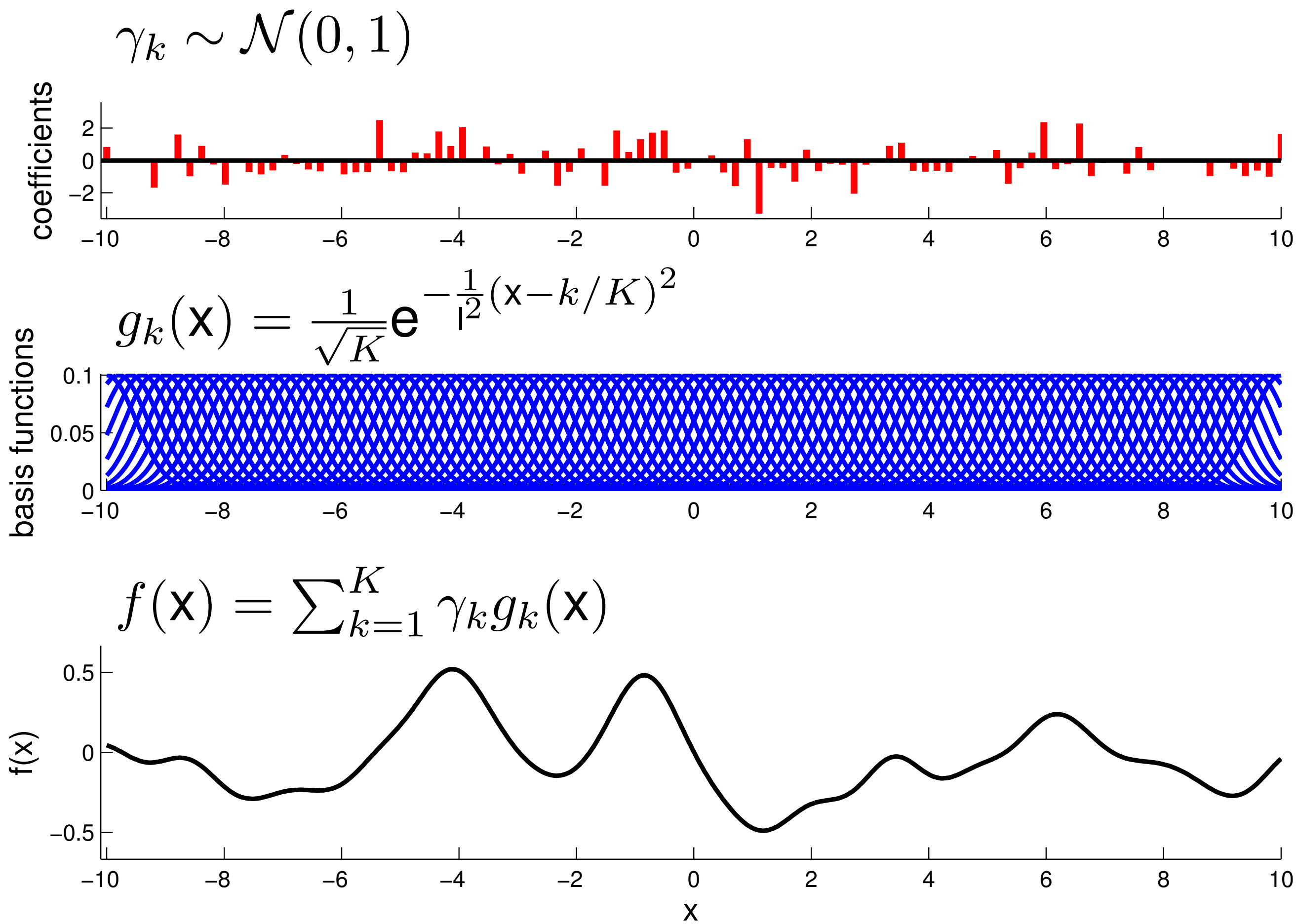
Basis function view of Gaussian processes



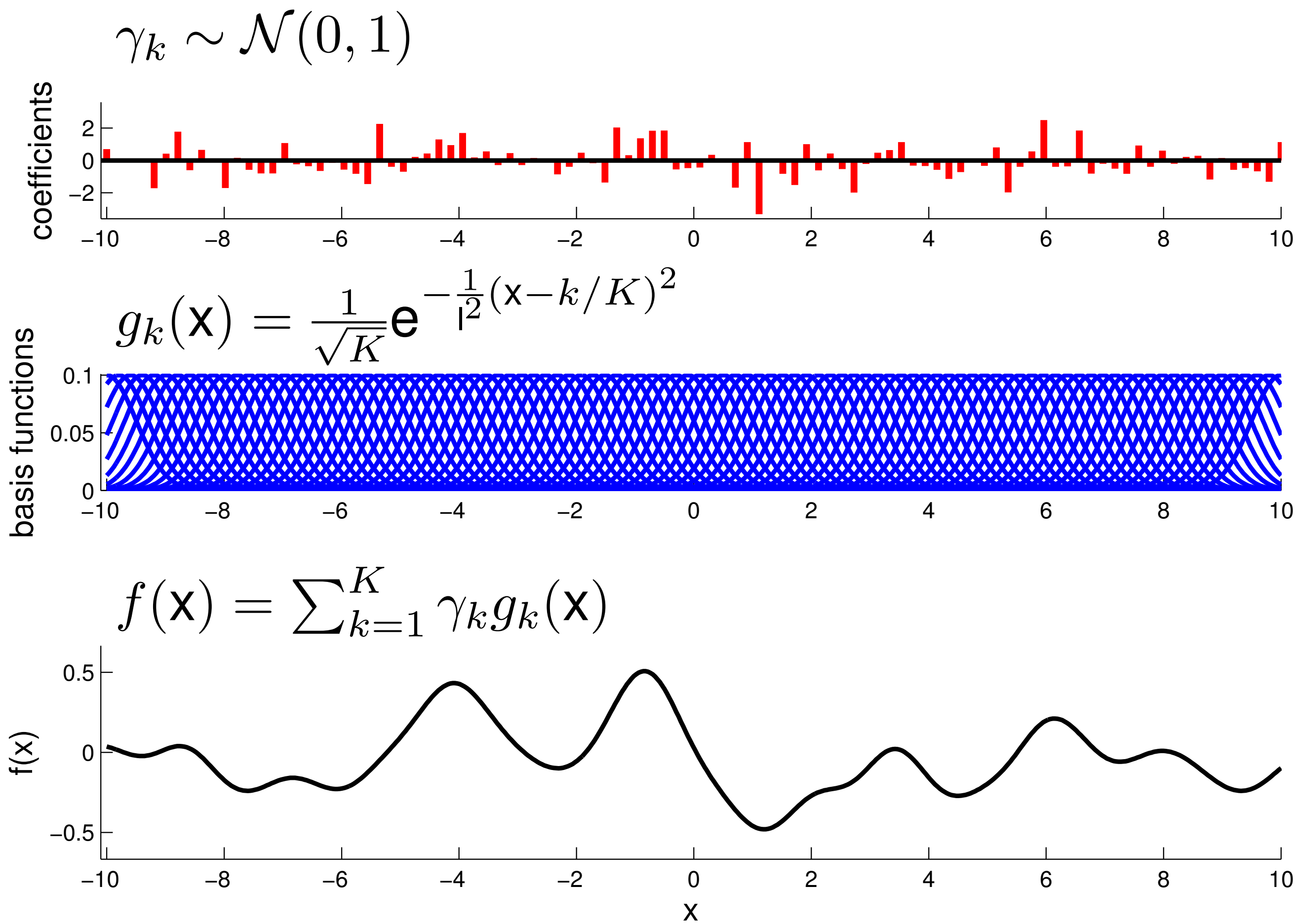
Basis function view of Gaussian processes



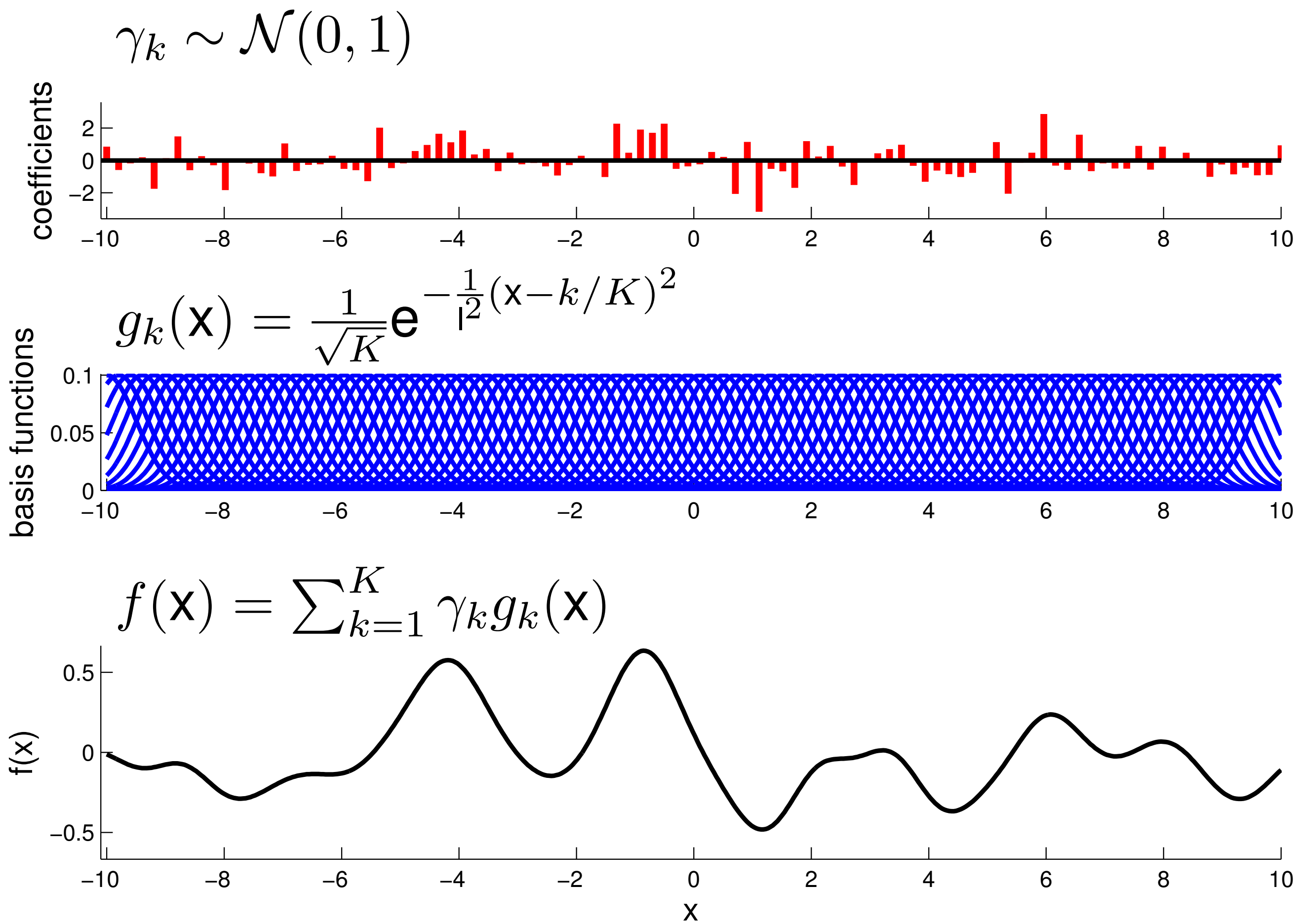
Basis function view of Gaussian processes



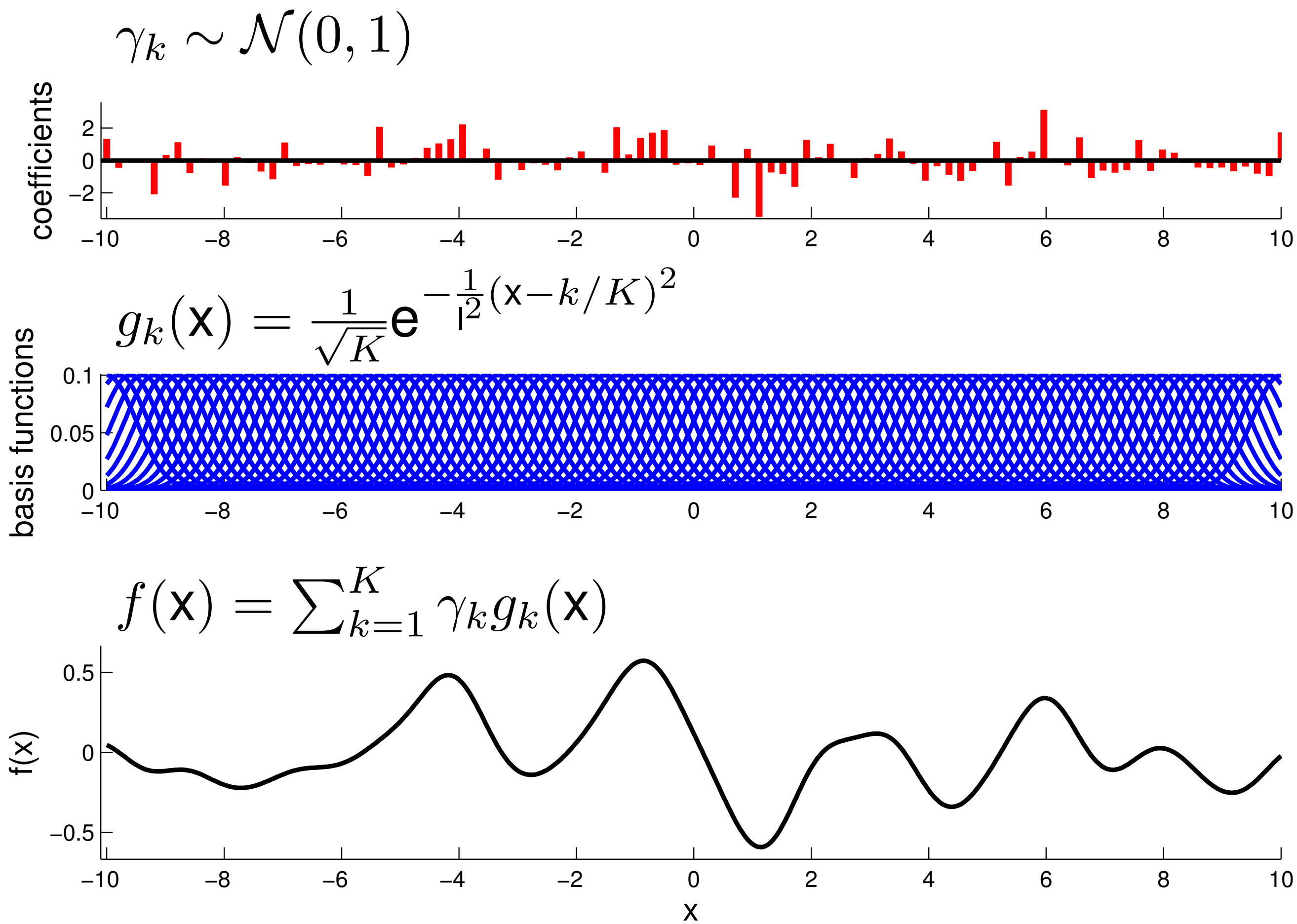
Basis function view of Gaussian processes



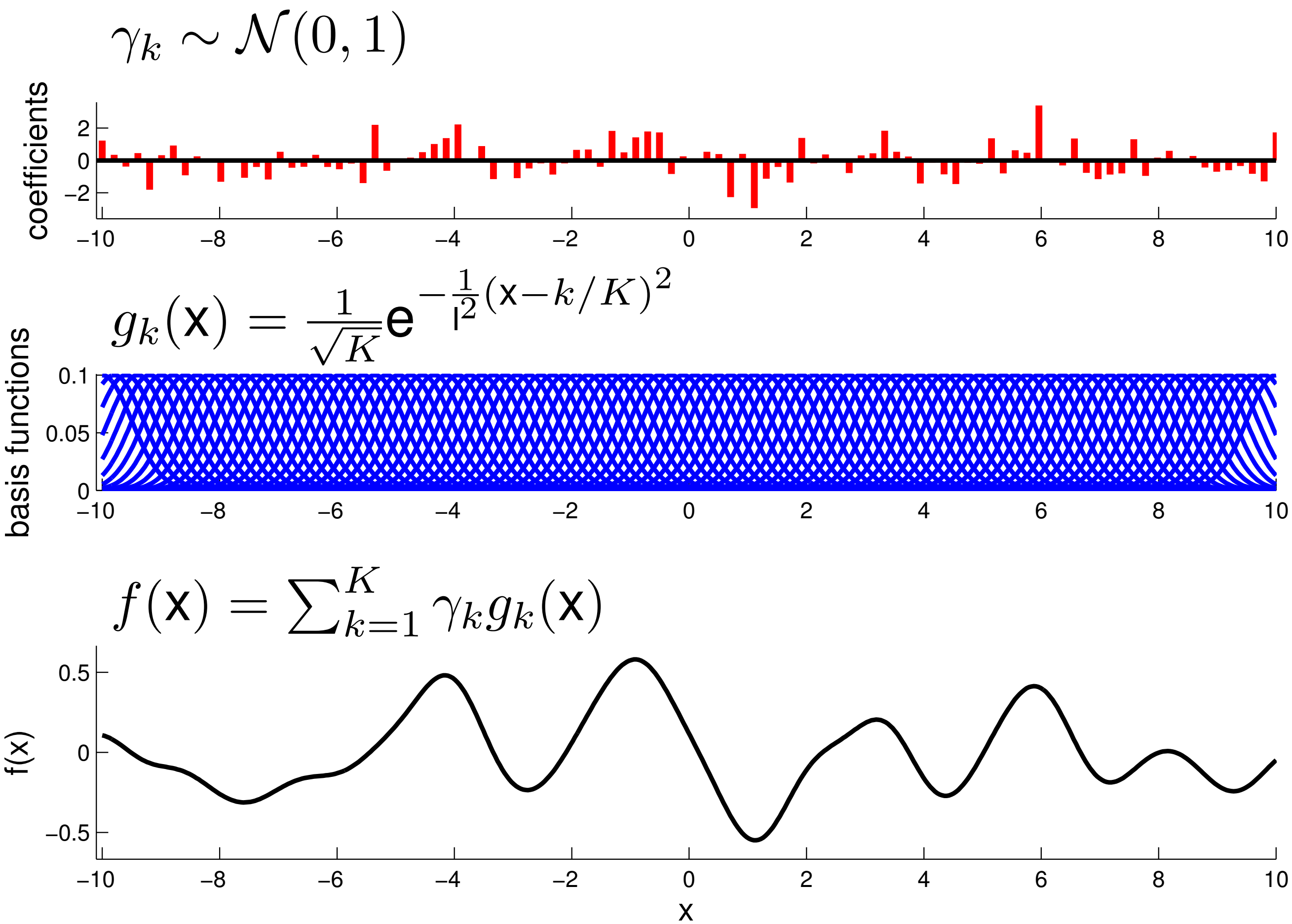
Basis function view of Gaussian processes



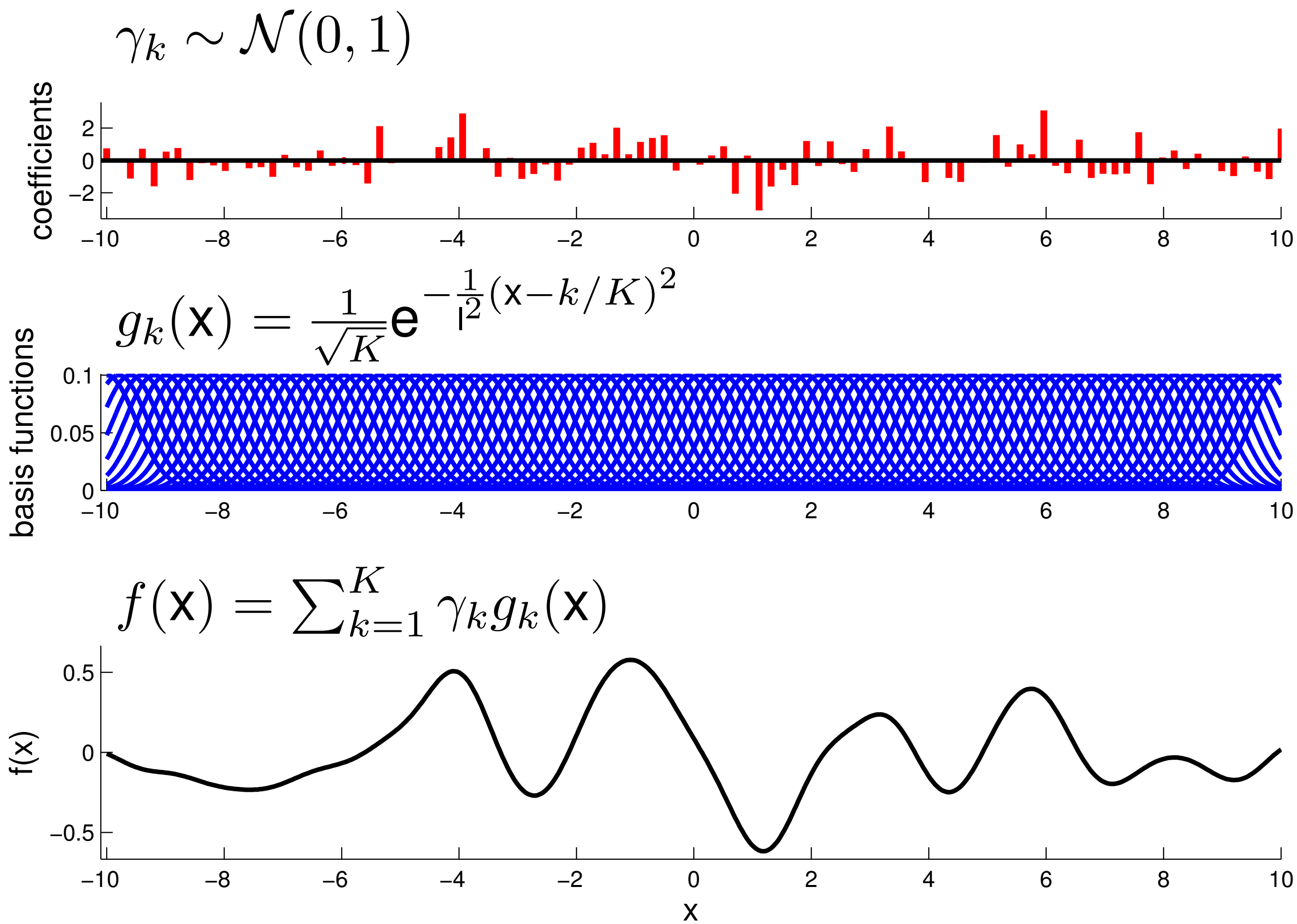
Basis function view of Gaussian processes



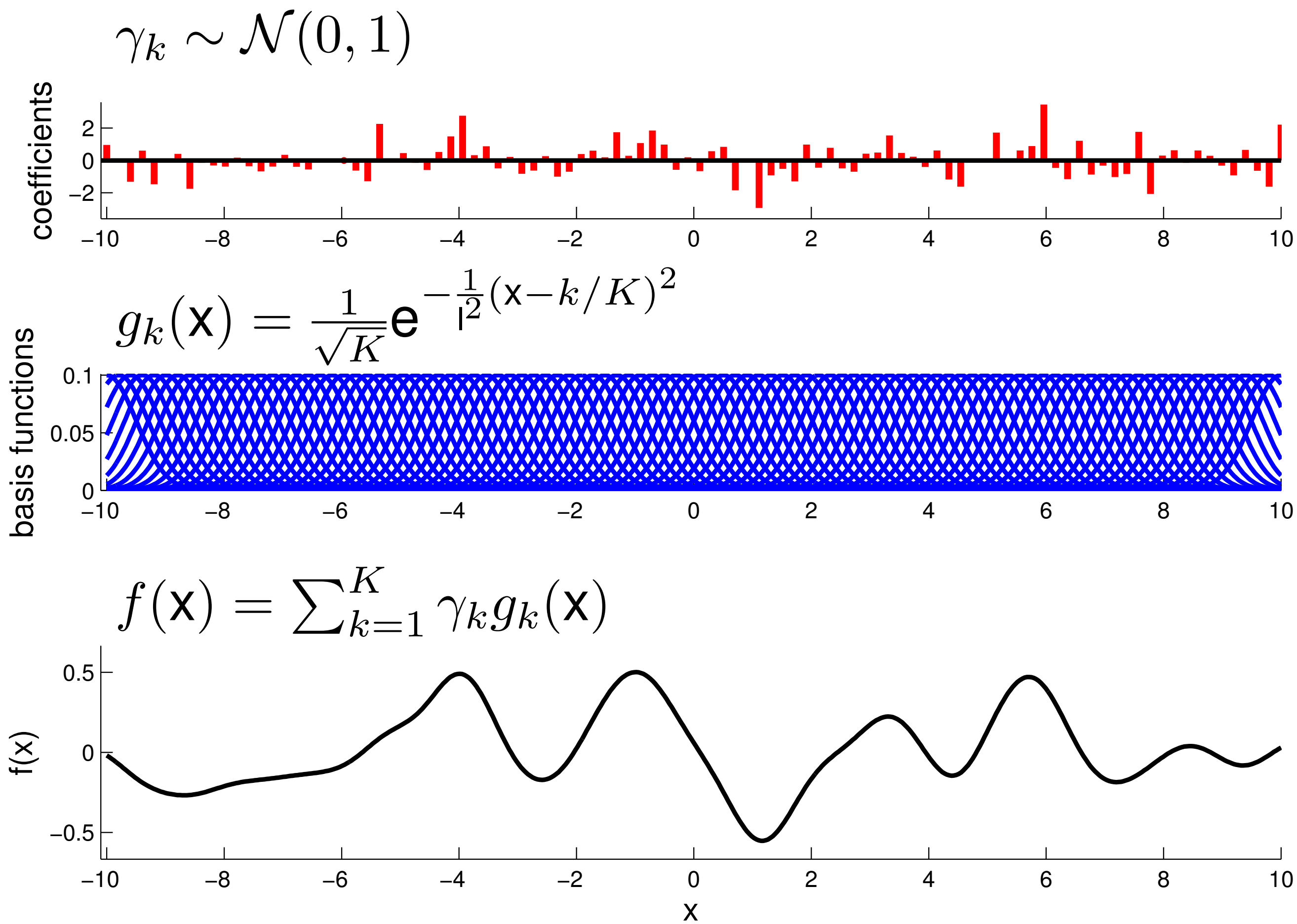
Basis function view of Gaussian processes



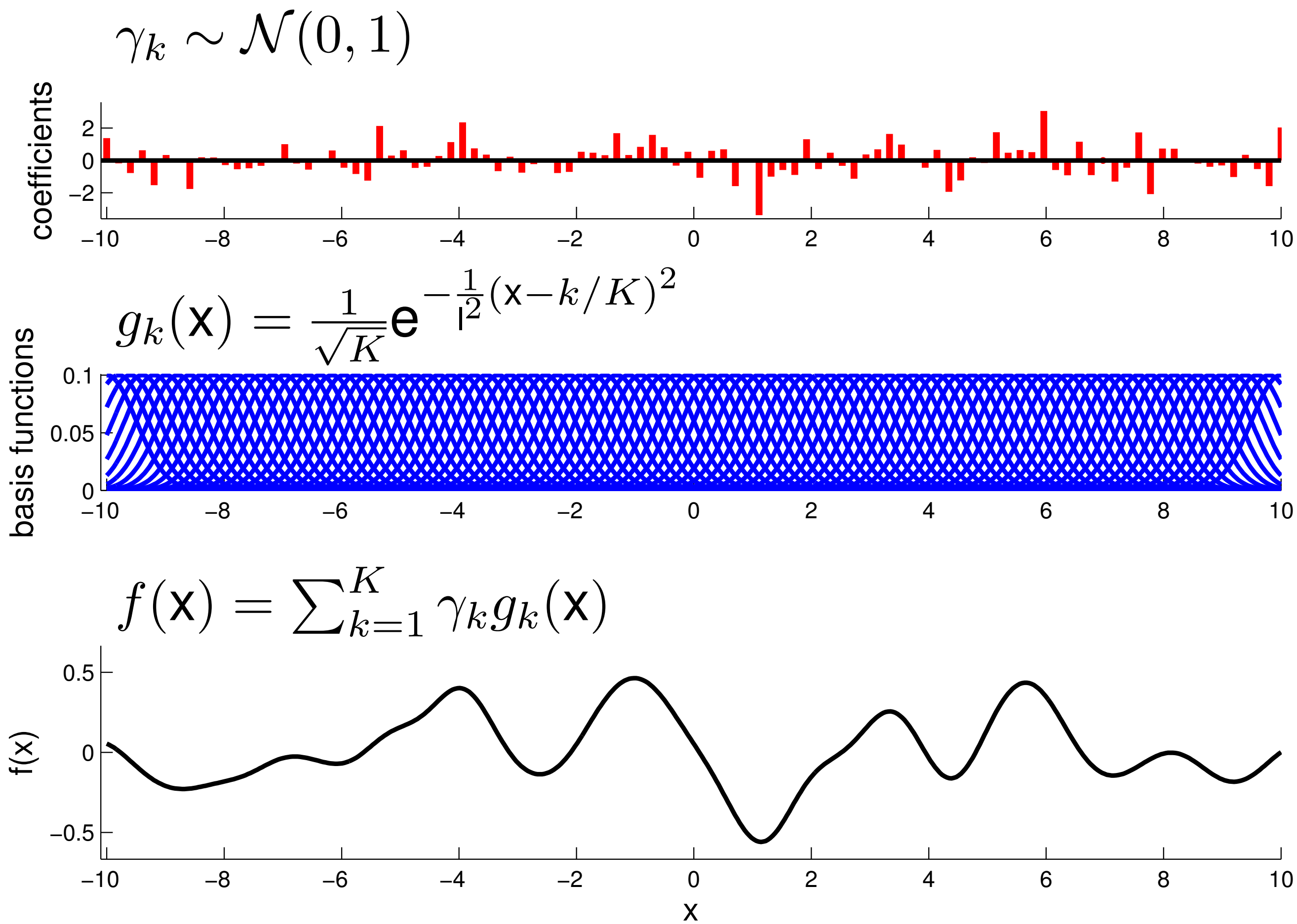
Basis function view of Gaussian processes



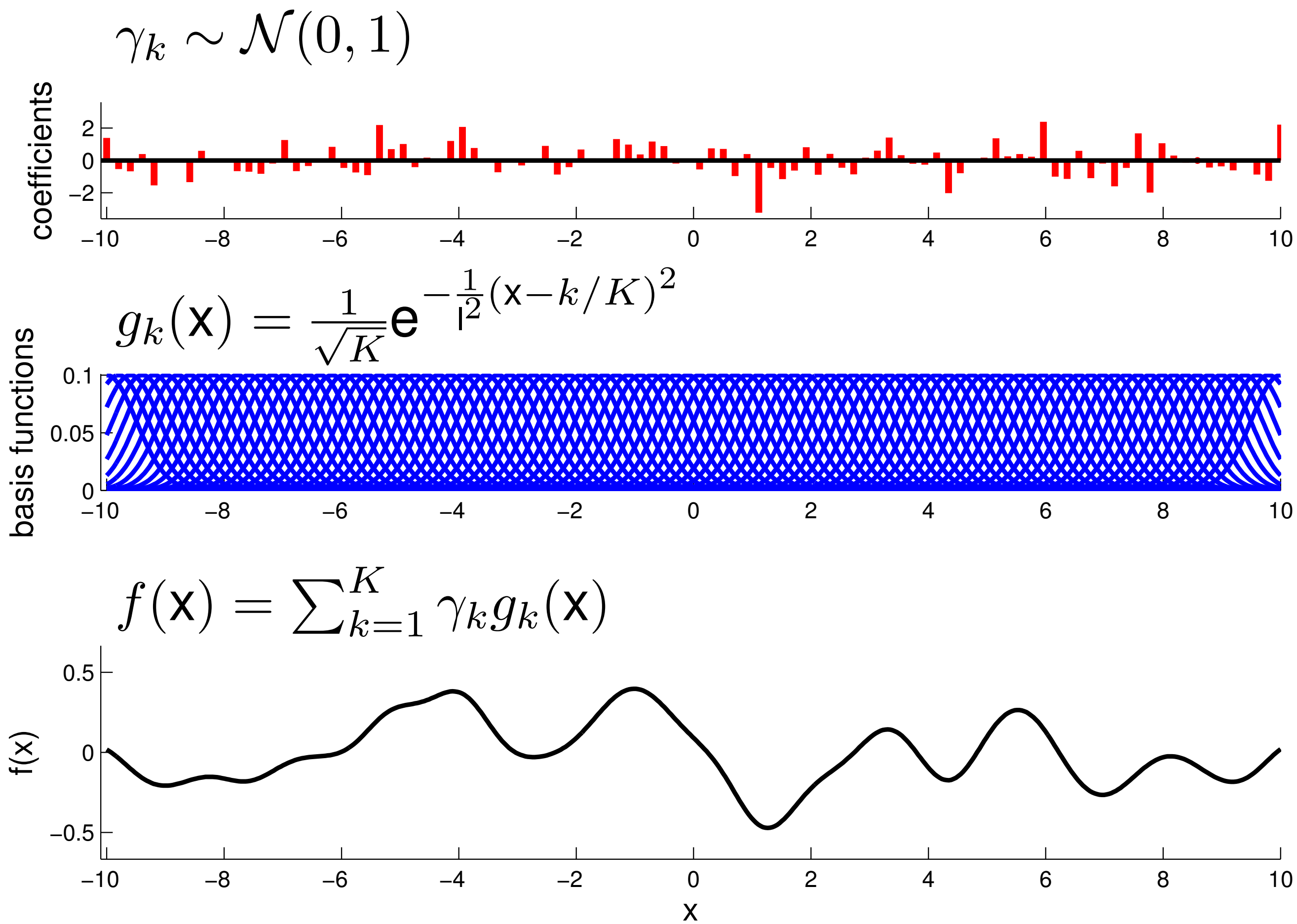
Basis function view of Gaussian processes



Basis function view of Gaussian processes



Basis function view of Gaussian processes



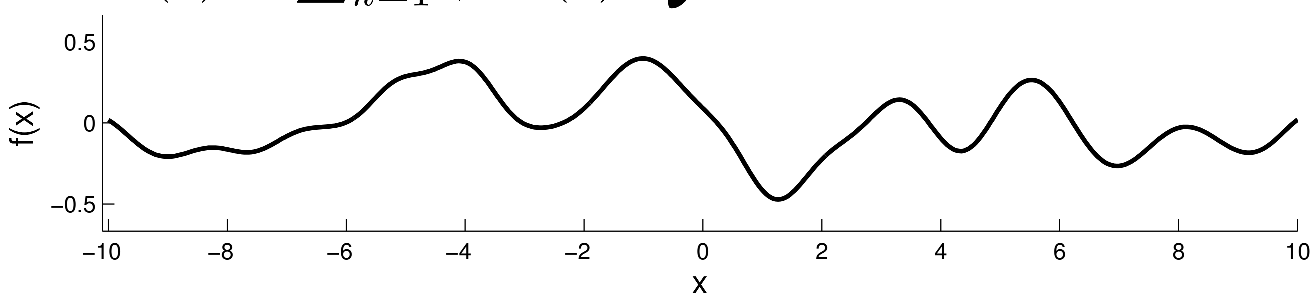
Basis function view of Gaussian processes

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

$$g_k(\mathsf{x}) = \frac{1}{\sqrt{K}} \mathsf{e}^{-\frac{1}{\mathsf{l}^2}(\mathsf{x}-k/K)^2}$$

$$f(\mathsf{x}) = \sum_{k=1}^{K} \gamma_k g_k(\mathsf{x})$$

$$m(\mathsf{x}) = \langle f(\mathsf{x}) \rangle$$



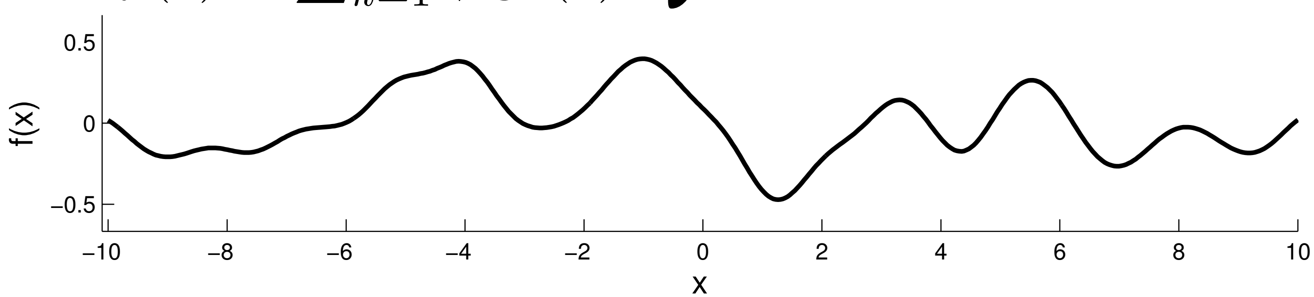
Basis function view of Gaussian processes

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

$$g_k(\mathsf{x}) = \frac{1}{\sqrt{K}} \mathsf{e}^{-\frac{1}{\mathsf{l}^2}(\mathsf{x}-k/K)^2}$$

$$f(\mathsf{x}) = \sum_{k=1}^{K} \gamma_k g_k(\mathsf{x})$$

$$m(\mathsf{x}) = \langle \sum_k \gamma_k g_k(\mathsf{x}) \rangle$$



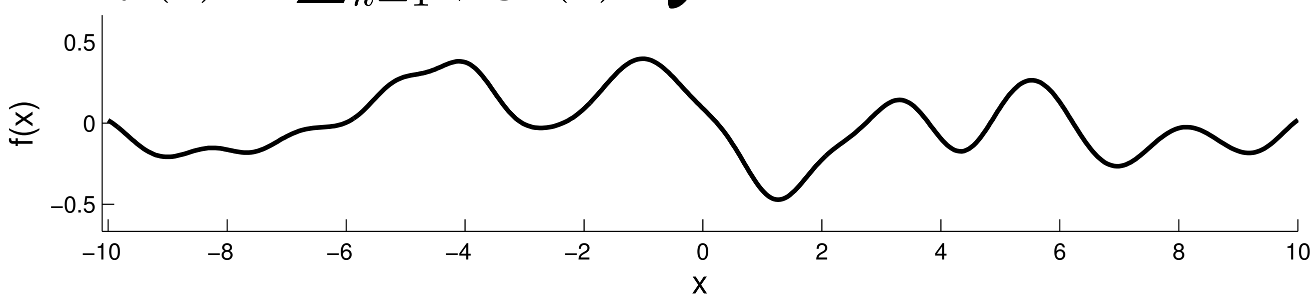
Basis function view of Gaussian processes

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

$$g_k(\mathrm{x}) = \frac{1}{\sqrt{K}} \mathrm{e}^{-\frac{1}{\mathrm{l}^2}(\mathrm{x}-k/K)^2}$$

$$f(\mathrm{x}) = \sum_{k=1}^{K} \gamma_k g_k(\mathrm{x})$$

$$m(\mathrm{x}) = \sum_k \langle \gamma_k \rangle g_k(\mathrm{x})$$



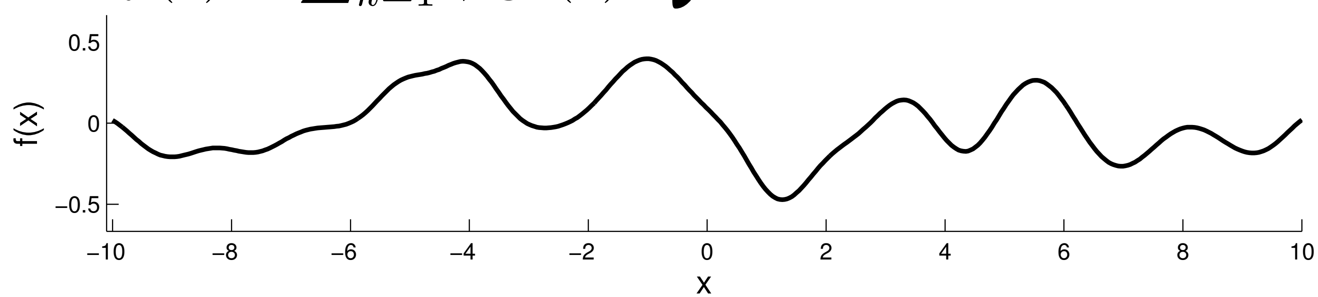
Basis function view of Gaussian processes

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

$$g_k(\mathsf{x}) = \frac{1}{\sqrt{K}} \mathsf{e}^{-\frac{1}{\mathsf{l}^2}(\mathsf{x}-k/K)^2}$$

$$f(\mathsf{x}) = \sum_{k=1}^{K} \gamma_k g_k(\mathsf{x})$$

$$m(\mathsf{x}) = 0$$



Basis function view of Gaussian processes

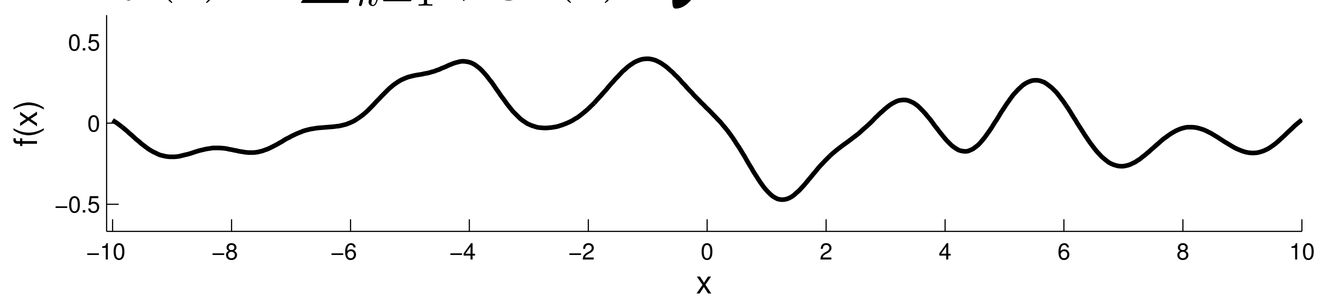
$$\gamma_k \sim \mathcal{N}(0,1)$$

$$g_k(\mathsf{x}) = \frac{1}{\sqrt{K}} \mathsf{e}^{-\frac{1}{\mathsf{l}^2}(\mathsf{x}-k/K)^2}$$

$$f(\mathsf{x}) = \sum_{k=1}^{K} \gamma_k g_k(\mathsf{x})$$

$$m(\mathsf{x}) = 0$$

$$\mathsf{K}(\mathsf{x},\mathsf{x}') = \langle f(\mathsf{x}) f(\mathsf{x}') \rangle$$



Basis function view of Gaussian processes

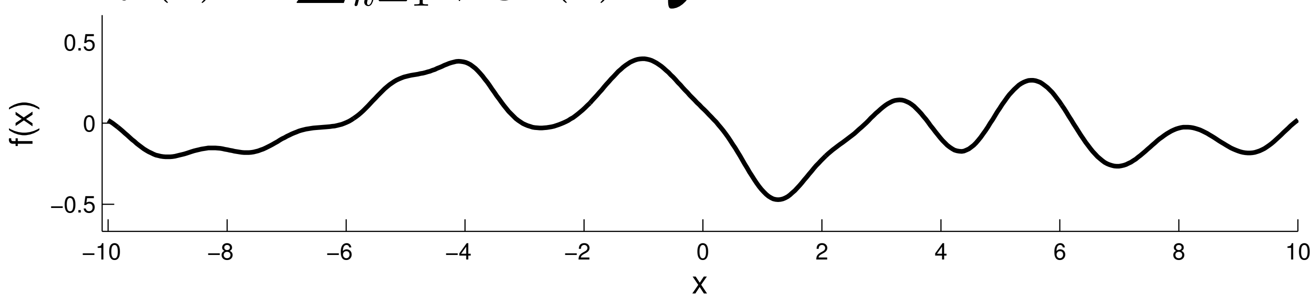
$$\gamma_k \sim \mathcal{N}(0,1)$$

$$g_k(\mathsf{x}) = \frac{1}{\sqrt{K}} \mathsf{e}^{-\frac{1}{\mathsf{l}^2}(\mathsf{x}-k/K)^2}$$

$$f(\mathsf{x}) = \sum_{k=1}^{K} \gamma_k g_k(\mathsf{x})$$

$$m(\mathsf{x}) = 0$$

$$\mathsf{K}(\mathsf{x},\mathsf{x}') = \langle \sum_k \gamma_k g_k(\mathsf{x}) \sum_{k'} \gamma_{k'} g_{k'}(\mathsf{x}') \rangle$$



Basis function view of Gaussian processes

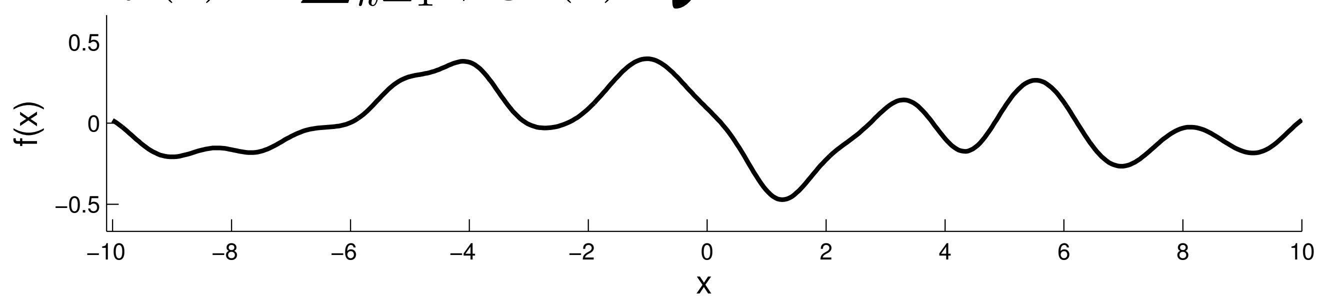
$$\gamma_k \sim \mathcal{N}(0,1)$$

$$g_k(\mathsf{x}) = \frac{1}{\sqrt{K}} \mathsf{e}^{-\frac{1}{\mathsf{l}^2}(\mathsf{x}-k/K)^2}$$

$$f(\mathsf{x}) = \sum_{k=1}^{K} \gamma_k g_k(\mathsf{x})$$

$$m(\mathsf{x}) = 0$$

$$\mathsf{K}(\mathsf{x},\mathsf{x}') = \sum_{k,k'} \langle \gamma_k \gamma_{k'} \rangle g_k(\mathsf{x}) g_{k'}(\mathsf{x}')$$



Basis function view of Gaussian processes

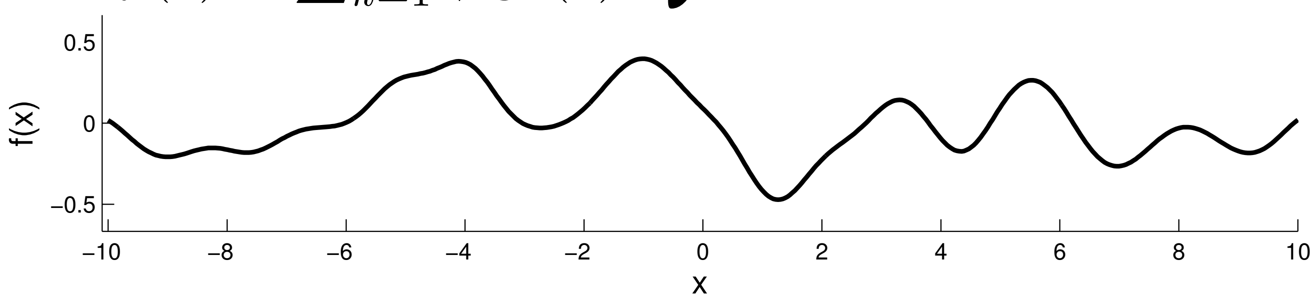
$$\gamma_k \sim \mathcal{N}(0,1)$$

$$g_k(\mathsf{x}) = \frac{1}{\sqrt{K}} \mathsf{e}^{-\frac{1}{\mathsf{l}^2}(\mathsf{x}-k/K)^2}$$

$$f(\mathsf{x}) = \sum_{k=1}^{K} \gamma_k g_k(\mathsf{x})$$

$$m(\mathsf{x}) = 0$$

$$\mathsf{K}(\mathsf{x},\mathsf{x}') = \sum_{k,k'} \delta_{k,k'} g_k(\mathsf{x}) g_{k'}(\mathsf{x}')$$



Basis function view of Gaussian processes

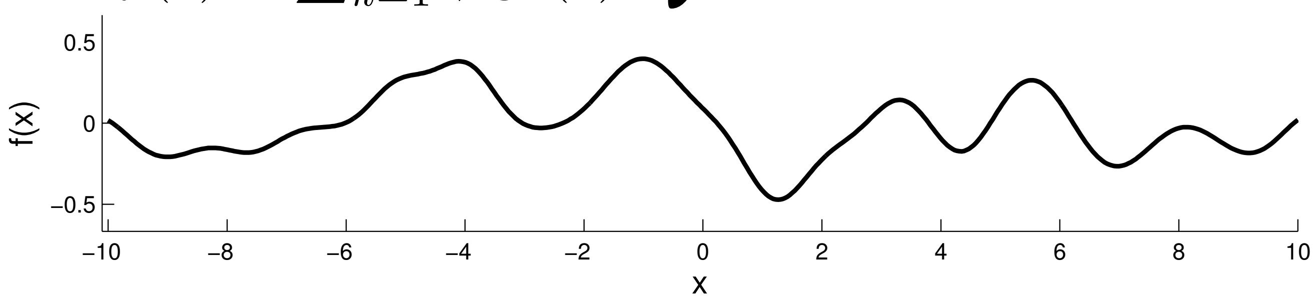
$$\gamma_k \sim \mathcal{N}(0,1)$$

$$g_k(\mathsf{x}) = \frac{1}{\sqrt{K}} \mathsf{e}^{-\frac{1}{\mathsf{l}^2}(\mathsf{x}-k/K)^2}$$

$$f(\mathsf{x}) = \sum_{k=1}^{K} \gamma_k g_k(\mathsf{x})$$

$$m(\mathsf{x}) = 0$$

$$\mathsf{K}(\mathsf{x},\mathsf{x}') = \sum_k g_k(\mathsf{x}) g_k(\mathsf{x}')$$



Basis function view of Gaussian processes

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

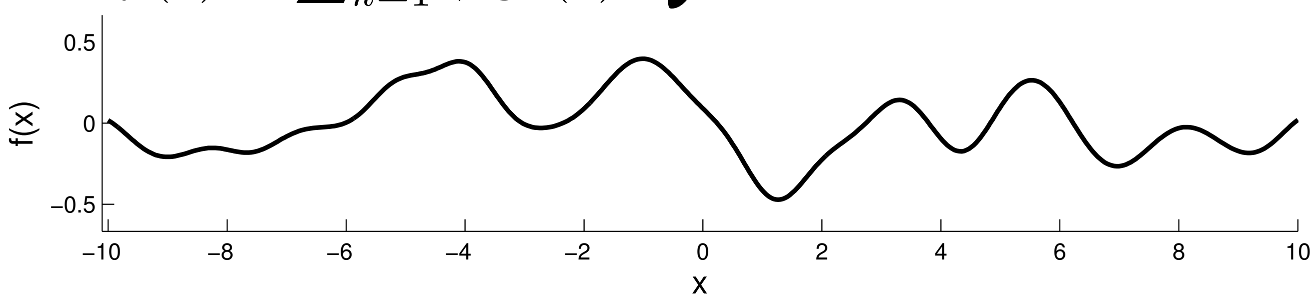
$$g_k(\mathsf{x}) = \frac{1}{\sqrt{K}} \mathsf{e}^{-\frac{1}{\mathsf{l}^2}(\mathsf{x}-k/K)^2}$$

$$f(\mathsf{x}) = \sum_{k=1}^{K} \gamma_k g_k(\mathsf{x})$$

$$m(\mathsf{x}) = 0$$

$$\mathsf{K}(\mathsf{x},\mathsf{x}') = \sum_k g_k(\mathsf{x}) g_k(\mathsf{x}')$$

$$= \frac{1}{K} \sum_k \mathsf{e}^{-\frac{1}{\mathsf{l}^2}(\mathsf{x}-k/K)^2 - \frac{1}{\mathsf{l}^2}(\mathsf{x}'-k/K)^2}$$



Basis function view of Gaussian processes

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

$$g_k(\mathsf{x}) = \frac{1}{\sqrt{K}} \mathsf{e}^{-\frac{1}{\mathsf{l}^2}(\mathsf{x}-k/K)^2}$$

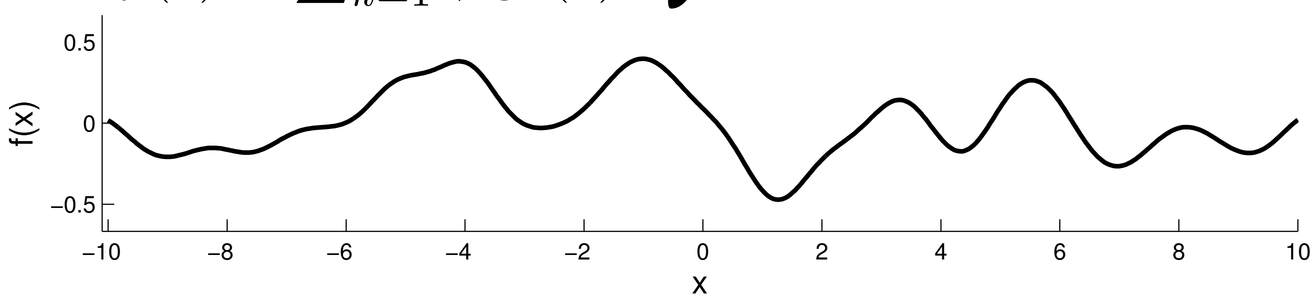
$$f(\mathsf{x}) = \sum_{k=1}^{K} \gamma_k g_k(\mathsf{x})$$

$$m(\mathsf{x}) = 0$$

$$\mathsf{K}(\mathsf{x},\mathsf{x}') = \sum_k g_k(\mathsf{x}) g_k(\mathsf{x}')$$

$$= \frac{1}{K} \sum_k \mathsf{e}^{-\frac{1}{\mathsf{l}^2}(\mathsf{x}-k/K)^2 - \frac{1}{\mathsf{l}^2}(\mathsf{x}'-k/K)^2}$$

$$\underset{K \to \infty}{\longrightarrow} \quad \int du \;\; \mathsf{e}^{-\frac{1}{\mathsf{l}^2}(\mathsf{x}-u)^2 - \frac{1}{\mathsf{l}^2}(\mathsf{x}'-u)^2}$$



Basis function view of Gaussian processes

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

$$g_k(\mathsf{x}) = \frac{1}{\sqrt{K}} \mathsf{e}^{-\frac{1}{\mathsf{l}^2}(\mathsf{x}-k/K)^2}$$

$$f(\mathsf{x}) = \sum_{k=1}^{K} \gamma_k g_k(\mathsf{x})$$

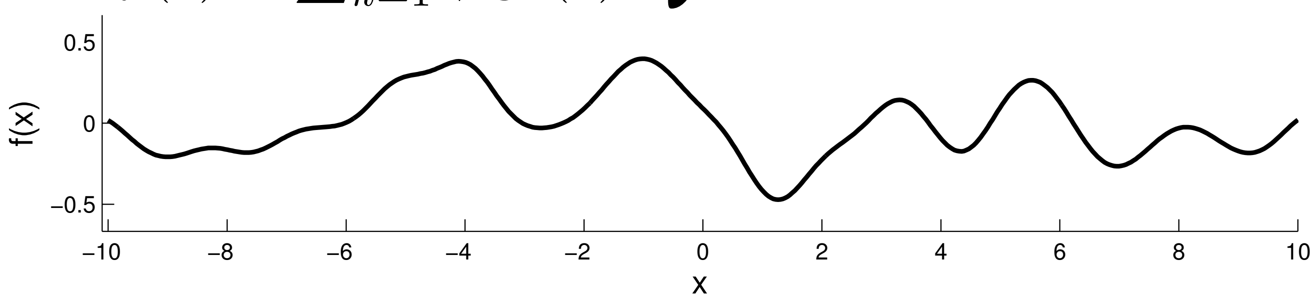
$$m(\mathsf{x}) = 0$$

$$\mathsf{K}(\mathsf{x},\mathsf{x}') = \sum_k g_k(\mathsf{x}) g_k(\mathsf{x}')$$

$$= \frac{1}{K} \sum_k \mathsf{e}^{-\frac{1}{\mathsf{l}^2}(\mathsf{x}-k/K)^2 - \frac{1}{\mathsf{l}^2}(\mathsf{x}'-k/K)^2}$$

$$\underset{K \to \infty}{\longrightarrow} \int du \;\; \mathsf{e}^{-\frac{1}{\mathsf{l}^2}(\mathsf{x}-u)^2 - \frac{1}{\mathsf{l}^2}(\mathsf{x}'-u)^2}$$

$$\propto \mathsf{e}^{-\frac{1}{2\mathsf{l}^2}(\mathsf{x}-\mathsf{x}')^2}$$



Basis function view of Gaussian processes

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

$$g_k(\mathsf{x}) = \frac{1}{\sqrt{K}} \mathsf{e}^{-\frac{1}{\mathsf{l}^2}(\mathsf{x}-k/K)^2}$$

$$f(\mathsf{x}) = \sum_{k=1}^{K} \gamma_k g_k(\mathsf{x})$$

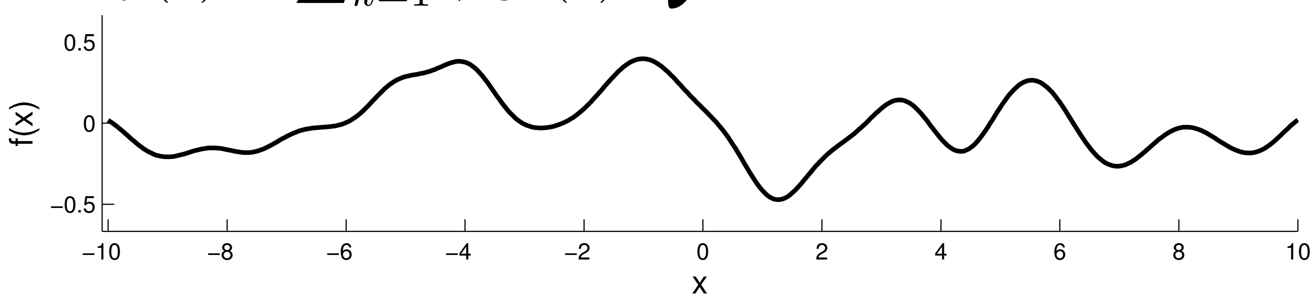
$$m(\mathsf{x}) = 0$$

$$\mathsf{K}(\mathsf{x},\mathsf{x}') = \sum_k g_k(\mathsf{x}) g_k(\mathsf{x}')$$

$$= \frac{1}{K} \sum_k \mathsf{e}^{-\frac{1}{\mathsf{l}^2}(\mathsf{x}-k/K)^2 - \frac{1}{\mathsf{l}^2}(\mathsf{x}'-k/K)^2}$$

$$\underset{K\to\infty}{\longrightarrow} \int du \;\; \mathsf{e}^{-\frac{1}{\mathsf{l}^2}(\mathsf{x}-u)^2 - \frac{1}{\mathsf{l}^2}(\mathsf{x}'-u)^2}$$

$$\propto \mathsf{e}^{-\frac{1}{2\mathsf{l}^2}(\mathsf{x}-\mathsf{x}')^2}$$

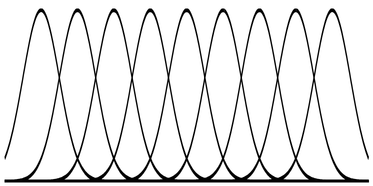


Gaussian processes $\equiv$ models with ∞ parameters

Basis function view of Gaussian Processes

$$\mathsf{y} = \sum_k \gamma_k g_k(\mathsf{x}) \qquad \gamma_k \sim \mathcal{N}(0, \Gamma_k)$$

basis function



$g_k(\mathsf{x})$

$$\exp\left(-\frac{1}{l}(\mathsf{x} - \mu_k)^2\right)$$

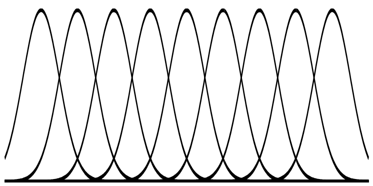
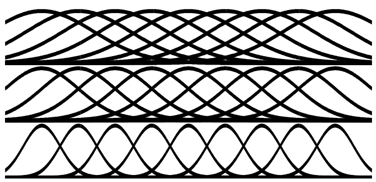
$\mathsf{K}(\mathsf{x}, \mathsf{x}')$

$$\exp\left(-\frac{1}{2l}(\mathsf{x} - \mathsf{x}')^2\right)$$

squared exponential

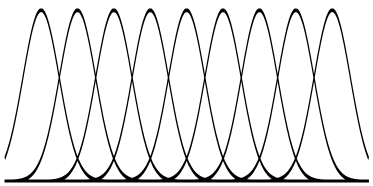
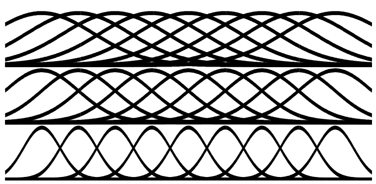
Basis function view of Gaussian Processes

$$\mathsf{y} = \sum_k \gamma_k g_k(\mathsf{x}) \qquad \gamma_k \sim \mathcal{N}(0, \Gamma_k)$$

basis function	$g_k(\mathsf{x})$	$\mathsf{K}(\mathsf{x}, \mathsf{x}')$	
	$\exp\left(-\frac{1}{l}(\mathsf{x} - \mu_k)^2\right)$	$\exp\left(-\frac{1}{2l}(\mathsf{x} - \mathsf{x}')^2\right)$	squared exponential
	$\exp\left(-\frac{1}{l_k}(\mathsf{x} - \mu_k)^2\right)$	$\left(1 + \frac{1}{2\alpha \mathsf{l}^2}\lvert\mathsf{x} - \mathsf{x}'\rvert\right)^{-\alpha}$	rational quadratic $\Gamma_k = \mathsf{invGam}(\mathsf{l})$

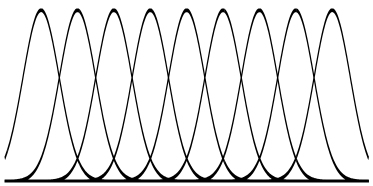
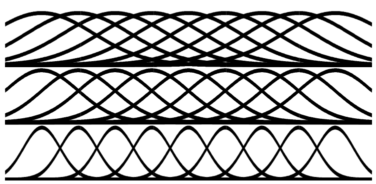
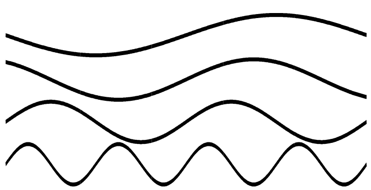
Basis function view of Gaussian Processes

$$\mathsf{y} = \sum_k \gamma_k g_k(\mathsf{x}) \qquad \gamma_k \sim \mathcal{N}(0, \Gamma_k)$$

basis function	$g_k(\mathsf{x})$	$\mathsf{K}(\mathsf{x}, \mathsf{x}')$	
	$\exp\left(-\frac{1}{l}(\mathsf{x} - \mu_k)^2\right)$	$\exp\left(-\frac{1}{2l}(\mathsf{x} - \mathsf{x}')^2\right)$	squared exponential
	$\exp\left(-\frac{1}{l_k}(\mathsf{x} - \mu_k)^2\right)$	mixture of SEs: $\left(1 + \frac{1}{2\alpha \mathsf{l}^2}\lvert\mathsf{x} - \mathsf{x}'\rvert\right)^{-\alpha}$	rational quadratic $\Gamma_k = \mathsf{invGam(l)}$

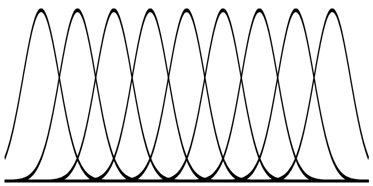
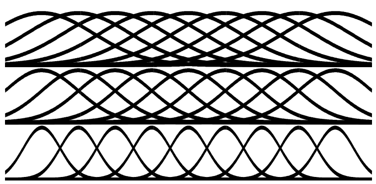
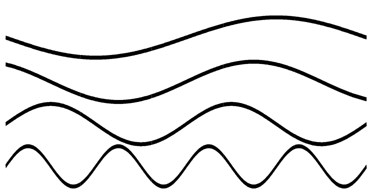
Basis function view of Gaussian Processes

$$\mathsf{y} = \sum_k \gamma_k g_k(\mathsf{x}) \qquad \gamma_k \sim \mathcal{N}(0, \Gamma_k)$$

basis function	$g_k(\mathsf{x})$	$\mathsf{K}(\mathsf{x}, \mathsf{x}')$	
	$\exp\left(-\frac{1}{l}(\mathsf{x} - \mu_k)^2\right)$	$\exp\left(-\frac{1}{2l}(\mathsf{x} - \mathsf{x}')^2\right)$	squared exponential
	$\exp\left(-\frac{1}{l_k}(\mathsf{x} - \mu_k)^2\right)$	mixture of SEs: $\left(1 + \frac{1}{2\alpha \mathsf{l}^2}\lvert\mathsf{x} - \mathsf{x}'\rvert\right)^{-\alpha}$	rational quadratic $\Gamma_k = \text{invGam}(\mathsf{l})$
	$\sin(\omega_k \mathsf{x})$ & $\cos(\omega_k \mathsf{x})$	$\sum_k \Gamma_k \cos(\omega_k(\mathsf{x} - \mathsf{x}'))$	any stationary covariance (Fourier basis)

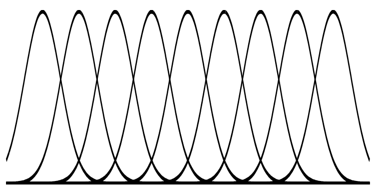
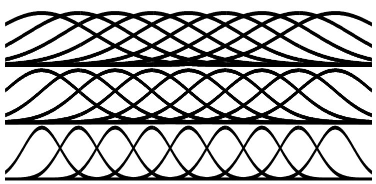
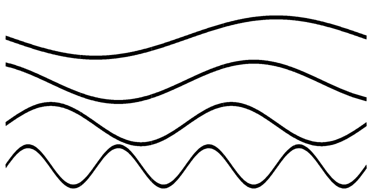
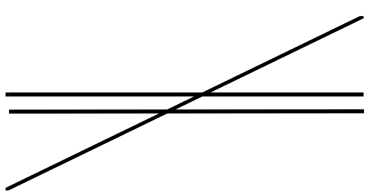
Basis function view of Gaussian Processes

$$\mathsf{y} = \sum_k \gamma_k g_k(\mathsf{x}) \qquad \gamma_k \sim \mathcal{N}(0, \Gamma_k)$$

basis function	$g_k(\mathsf{x})$	$\mathsf{K}(\mathsf{x},\mathsf{x}')$	
	$\exp\left(-\frac{1}{l}(\mathsf{x}-\mu_k)^2\right)$	$\exp\left(-\frac{1}{2l}(\mathsf{x}-\mathsf{x}')^2\right)$	squared exponential
	$\exp\left(-\frac{1}{l_k}(\mathsf{x}-\mu_k)^2\right)$	mixture of SEs $\left(1+\frac{1}{2\alpha \mathsf{l}^2}\lvert\mathsf{x}-\mathsf{x}'\rvert\right)^{-\alpha}$	rational quadratic $\Gamma_k = \text{invGam}(\mathsf{l})$
	data = Fourier series with Gaussian coefficients $\sin(\omega_k \mathsf{x})$ & $\cos(\omega_k \mathsf{x})$	covariance even function $\sum_k \Gamma_k \cos(\omega_k(\mathsf{x}-\mathsf{x}'))$	any stationary covariance (Fourier basis)

Basis function view of Gaussian Processes

$$\mathsf{y} = \sum_k \gamma_k g_k(\mathsf{x}) \qquad \gamma_k \sim \mathcal{N}(0, \Gamma_k)$$

basis function	$g_k(\mathsf{x})$	$\mathsf{K}(\mathsf{x}, \mathsf{x}')$	
	$\exp\left(-\frac{1}{l}(\mathsf{x} - \mu_k)^2\right)$	$\exp\left(-\frac{1}{2l}(\mathsf{x} - \mathsf{x}')^2\right)$	squared exponential
	$\exp\left(-\frac{1}{l_k}(\mathsf{x} - \mu_k)^2\right)$	mixture of SEs ↓ $\left(1 + \frac{1}{2\alpha \mathsf{l}^2}\lvert\mathsf{x} - \mathsf{x}'\rvert\right)^{-\alpha}$	rational quadratic $\Gamma_k = \text{invGam}(\mathsf{l})$
	data = Fourier series with Gaussian coefficients $\sin(\omega_k \mathsf{x})$ & $\cos(\omega_k \mathsf{x})$	covariance even function ↓ $\sum_k \Gamma_k \cos(\omega_k(\mathsf{x} - \mathsf{x}'))$	any stationary covariance (Fourier basis)
	x and 1	$\Gamma_1 \mathsf{x}\mathsf{x}' + \Gamma_2$	linear regression

Basis function view of Gaussian Processes

Basis function view useful because:

- **connects to classical approaches**

- **basis function generative model is useful theoretically**
  - Q: are draws from a SE continuous and differentiable?
  - A: they are (almost surely, almost everywhere) as the basis functions are

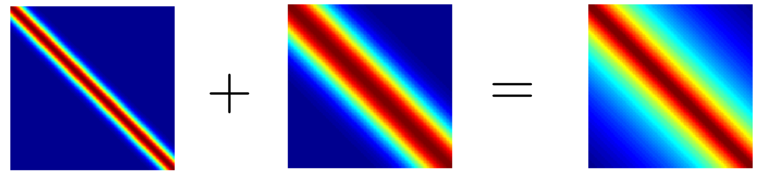
- **thinking about the basis function generative model is useful practically**
  - Q: how could I construct a periodic covariance?
  - A: use periodic basis functions with Gaussian weights

Making new covariance functions from old

(positive) linear combinations
of covariance functions

Making new covariance functions from old

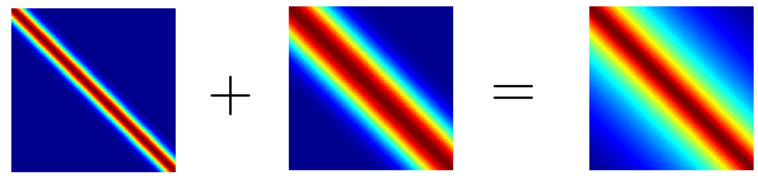
(positive) linear combinations
of covariance functions



e.g. scale mixture of SE = rational quadratic

Making new covariance functions from old

(positive) linear combinations of covariance functions

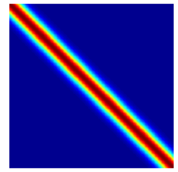
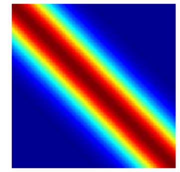
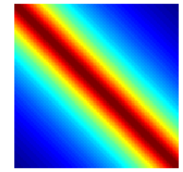


e.g. scale mixture of SE = rational quadratic

multiplication of covariance functions

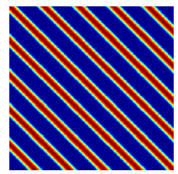
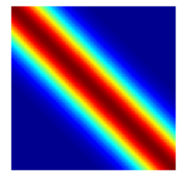
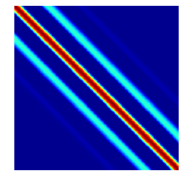
Making new covariance functions from old

(positive) linear combinations of covariance functions

 $+$  $=$ 

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

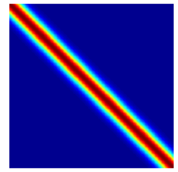
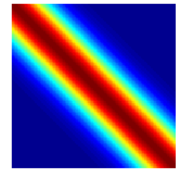
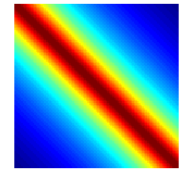
multiplication of covariance functions

 $\times$  $=$ 

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

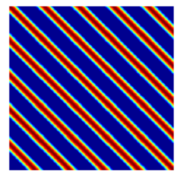
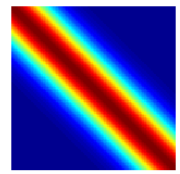
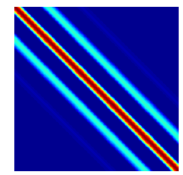
Making new covariance functions from old

(positive) linear combinations of covariance functions

 +  = 

e.g. scale mixture of SE = rational quadratic

multiplication of covariance functions

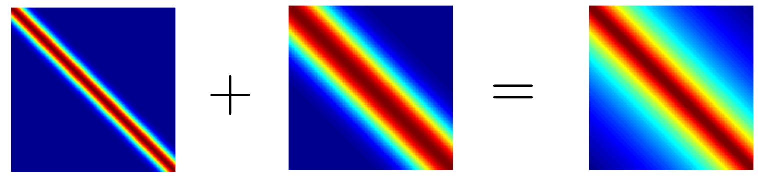
 $\times$  = 

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

derivative of GP = GP

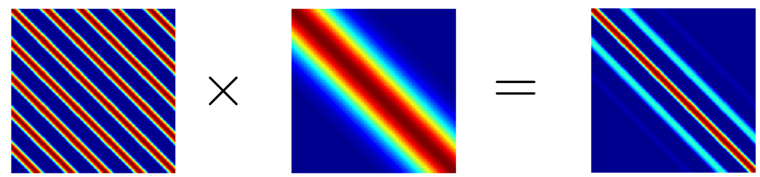
Making new covariance functions from old

(positive) linear combinations of covariance functions



e.g. scale mixture of SE = rational quadratic

multiplication of covariance functions

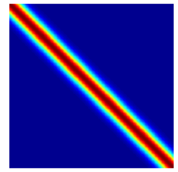
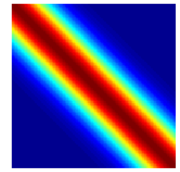
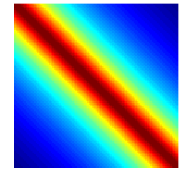


e.g. periodic × SE = $\cos(\omega \Delta x) \exp(-\frac{1}{2l^2} \Delta x^2)$

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

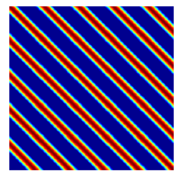
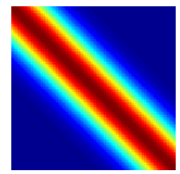
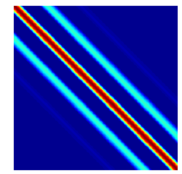
Making new covariance functions from old

(positive) linear combinations of covariance functions

 $+$  $=$ 

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

multiplication of covariance functions

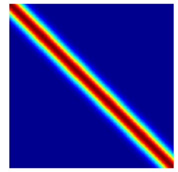
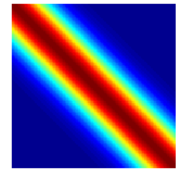
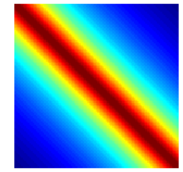
 $\times$  $=$ 

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

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

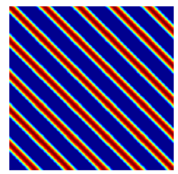
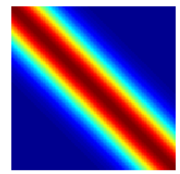
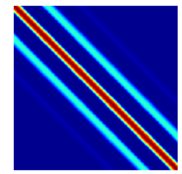
Making new covariance functions from old

(positive) linear combinations of covariance functions

 $+$  $=$ 

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

multiplication of covariance functions

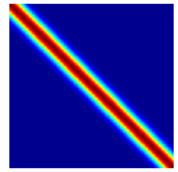
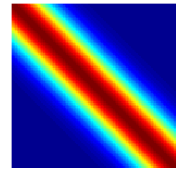
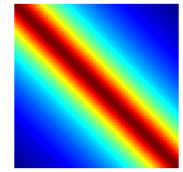
 $\times$  $=$ 

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

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

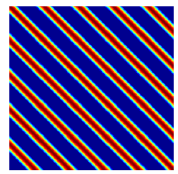
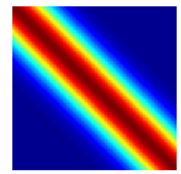
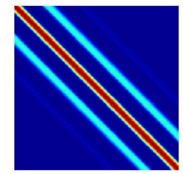
Making new covariance functions from old

(positive) linear combinations of covariance functions

 $+$  $=$ 

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

multiplication of covariance functions

 $\times$  $=$ 

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

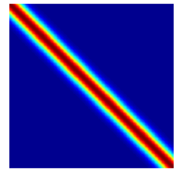
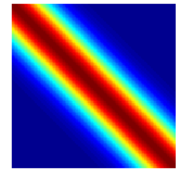
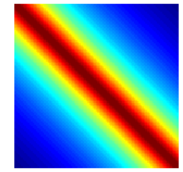
derivative of GP = GP

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

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

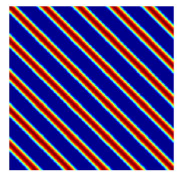
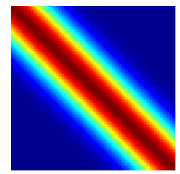
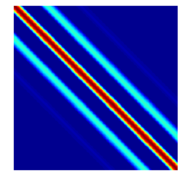
Making new covariance functions from old

(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

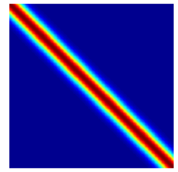
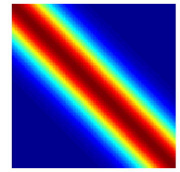
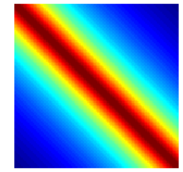
$$\frac{\mathsf{d}}{\mathsf{dx}} \mathsf{y}(\mathsf{x}) = \frac{\mathsf{d}}{\mathsf{dx}} \sum_{k=1}^{\infty} \gamma_k \mathsf{g}_k(\mathsf{x}) = \sum_{k=1}^{\infty} \gamma_k \frac{\mathsf{d}}{\mathsf{dx}} \mathsf{g}_k(\mathsf{x})$$

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

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

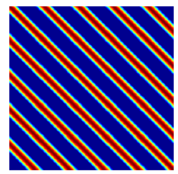
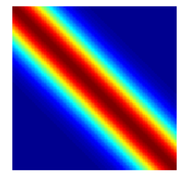
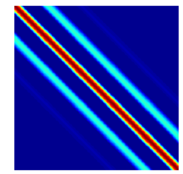
Making new covariance functions from old

(positive) linear combinations of covariance functions

 $+$  $=$ 

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

multiplication of covariance functions

 $\times$  $=$ 

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

derivative of GP = GP

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

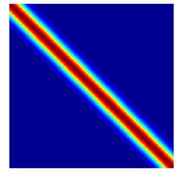
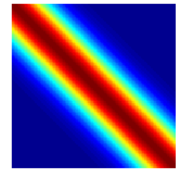
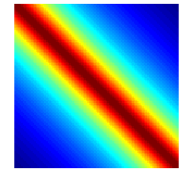
new basis: $\mathrm{g}'_k(\mathrm{x}) = \frac{\mathrm{d}}{\mathrm{dx}}\mathrm{g}_k(\mathrm{x})$

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

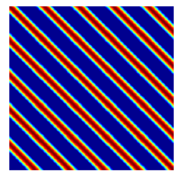
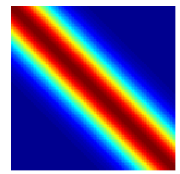
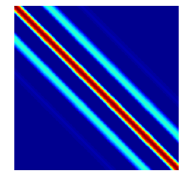
Making new covariance functions from old

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

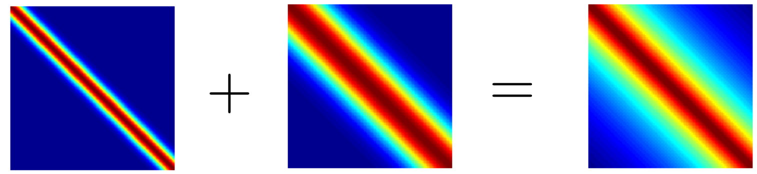
new covariance: $\mathsf{K}'(\mathsf{x},\mathsf{x}') = \frac{\mathsf{d}}{\mathsf{dx}}\frac{\mathsf{d}}{\mathsf{dx}'}\mathsf{K}(\mathsf{x},\mathsf{x}')$

integral of GP = GP

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

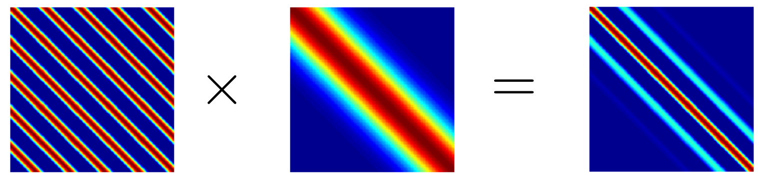
Making new covariance functions from old

(positive) linear combinations of covariance functions



e.g. scale mixture of SE = rational quadratic

multiplication of covariance functions



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

derivative of GP = GP

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

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

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

integral of GP = GP

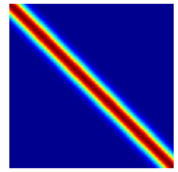
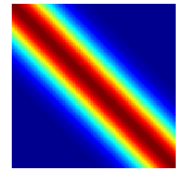
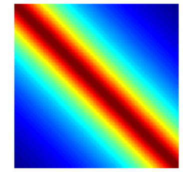
$$\int \mathrm{dx}\, y(x) = \sum_{k=1}^{\infty} \gamma_k \int \mathrm{dx}\, g_k(x)$$

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

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

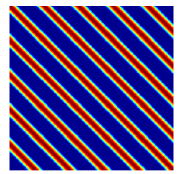
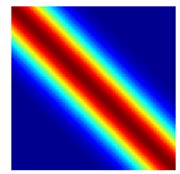
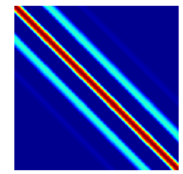
Making new covariance functions from old

(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

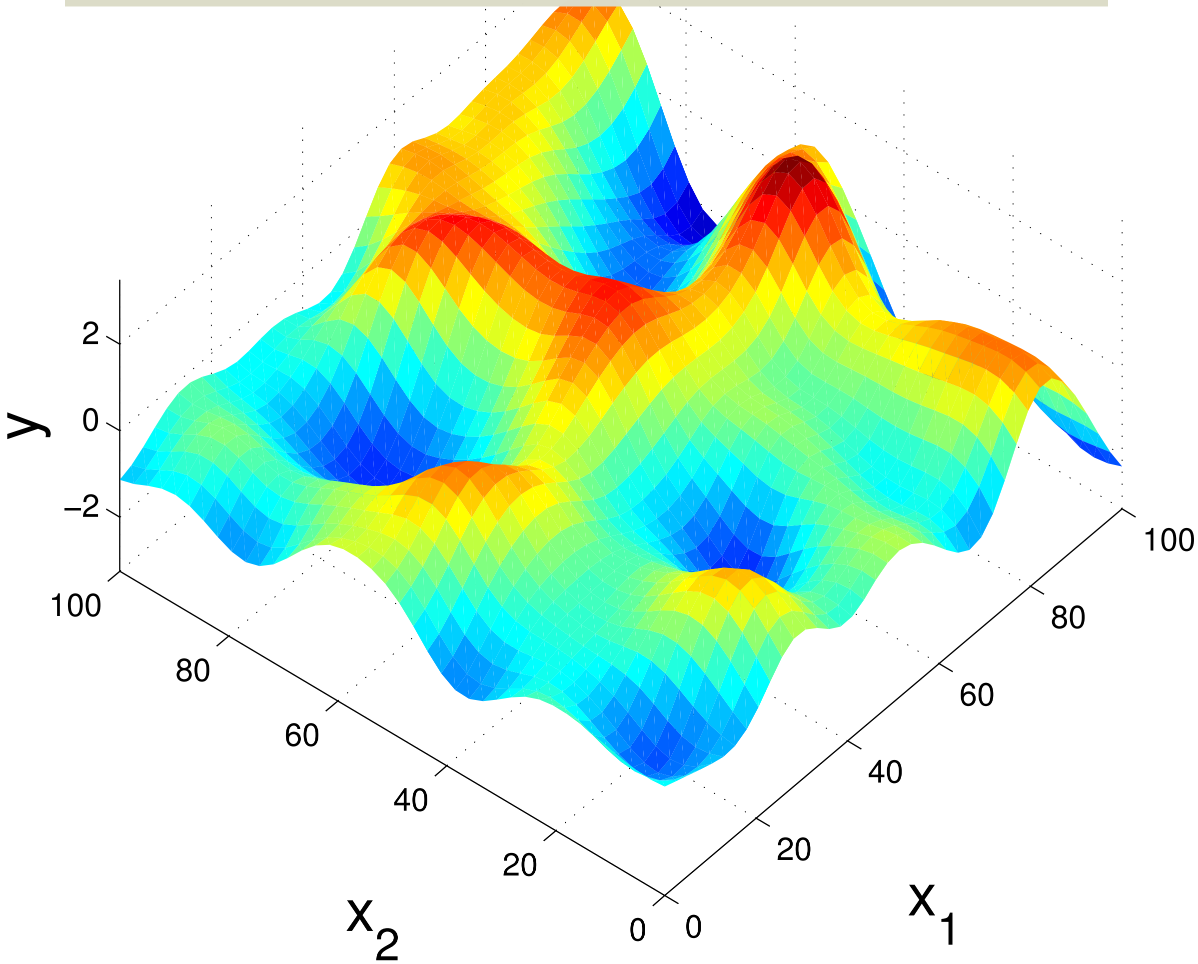
filtering a GP = GP

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

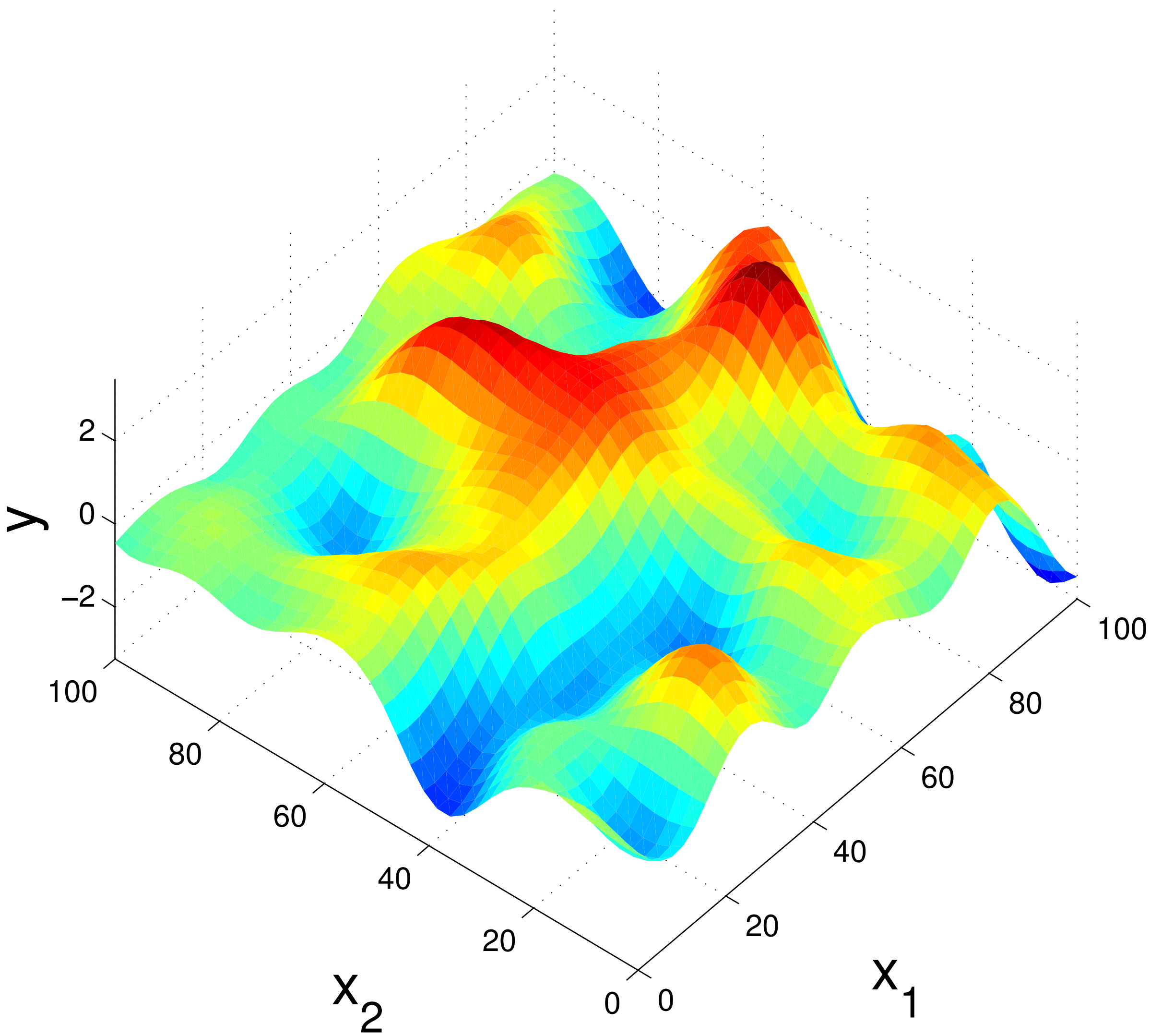
$\mathrm{K}'(\mathrm{x}, \mathrm{x}') = \mathrm{V}(\mathrm{x}) \otimes \mathrm{K}(\mathrm{x}, \mathrm{x}') \otimes \mathrm{V}(\mathrm{x}')$

Higher dimensional input spaces

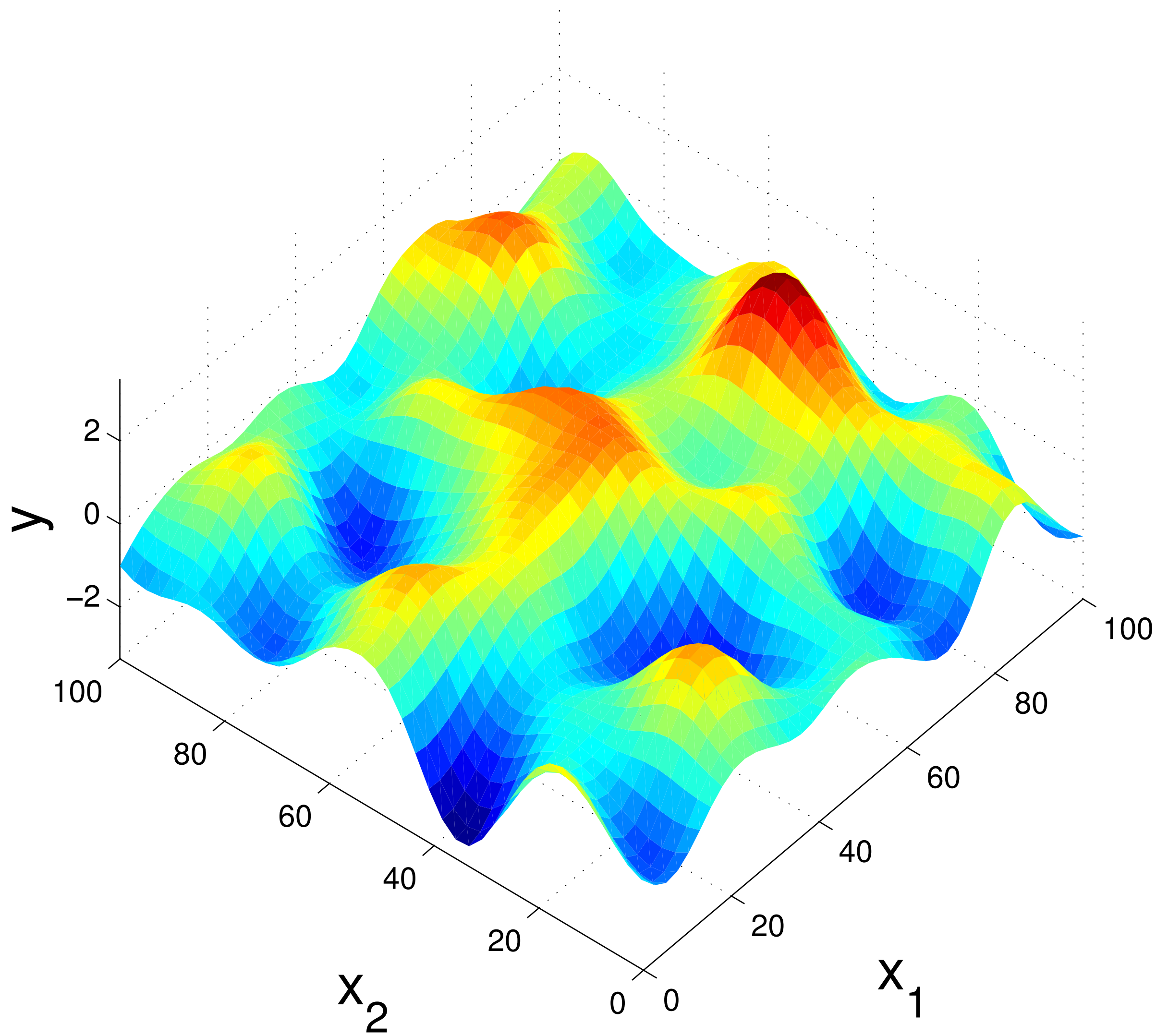
$$K(\mathbf{x},\mathbf{x}') = \sigma^2 \exp\left(-\frac{1}{2l_1^2}(\mathsf{x}_1 - \mathsf{x}_1')^2 - \frac{1}{2l_2^2}(\mathsf{x}_2 - \mathsf{x}_2')^2\right)$$



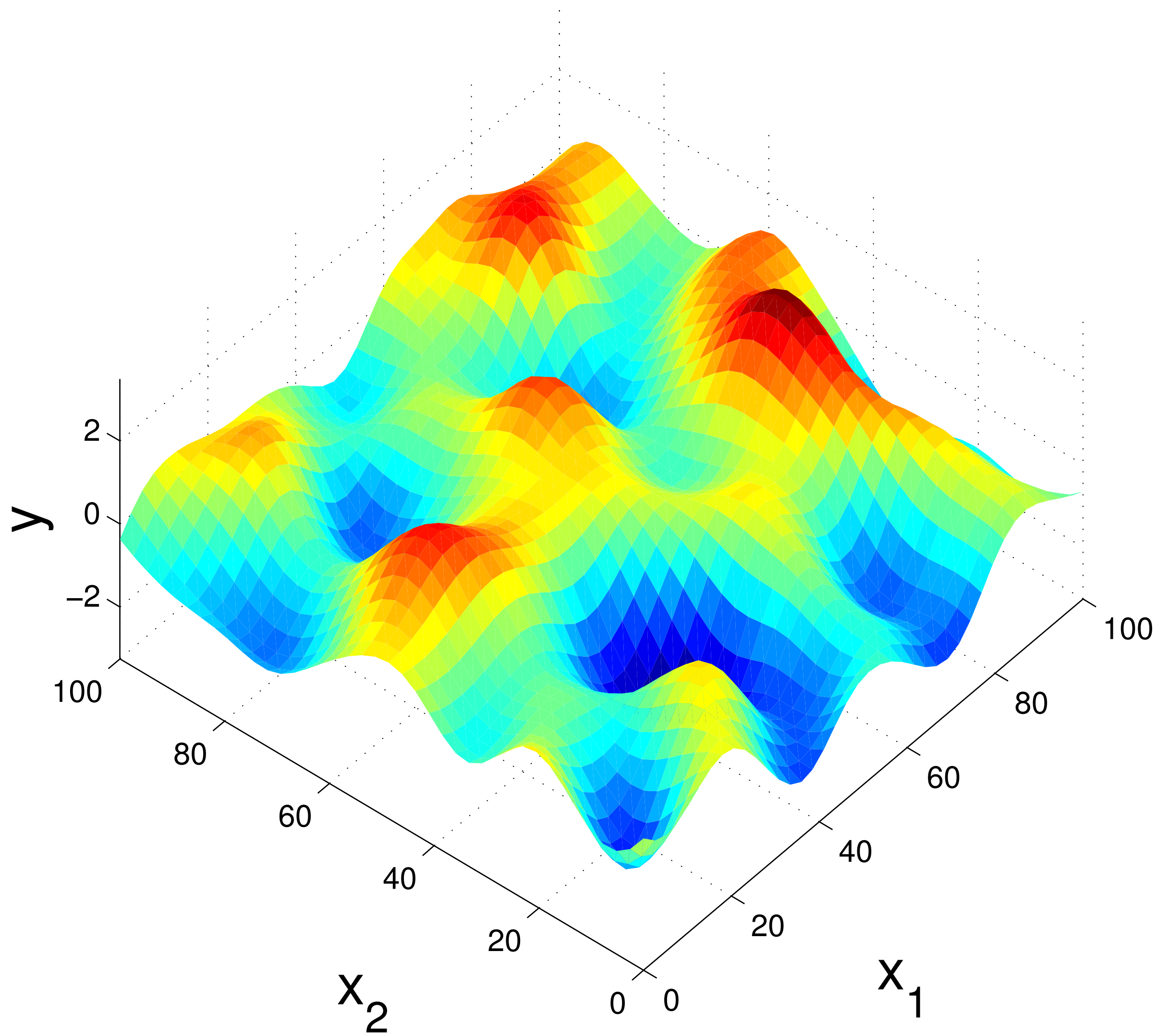
Higher dimensional input spaces



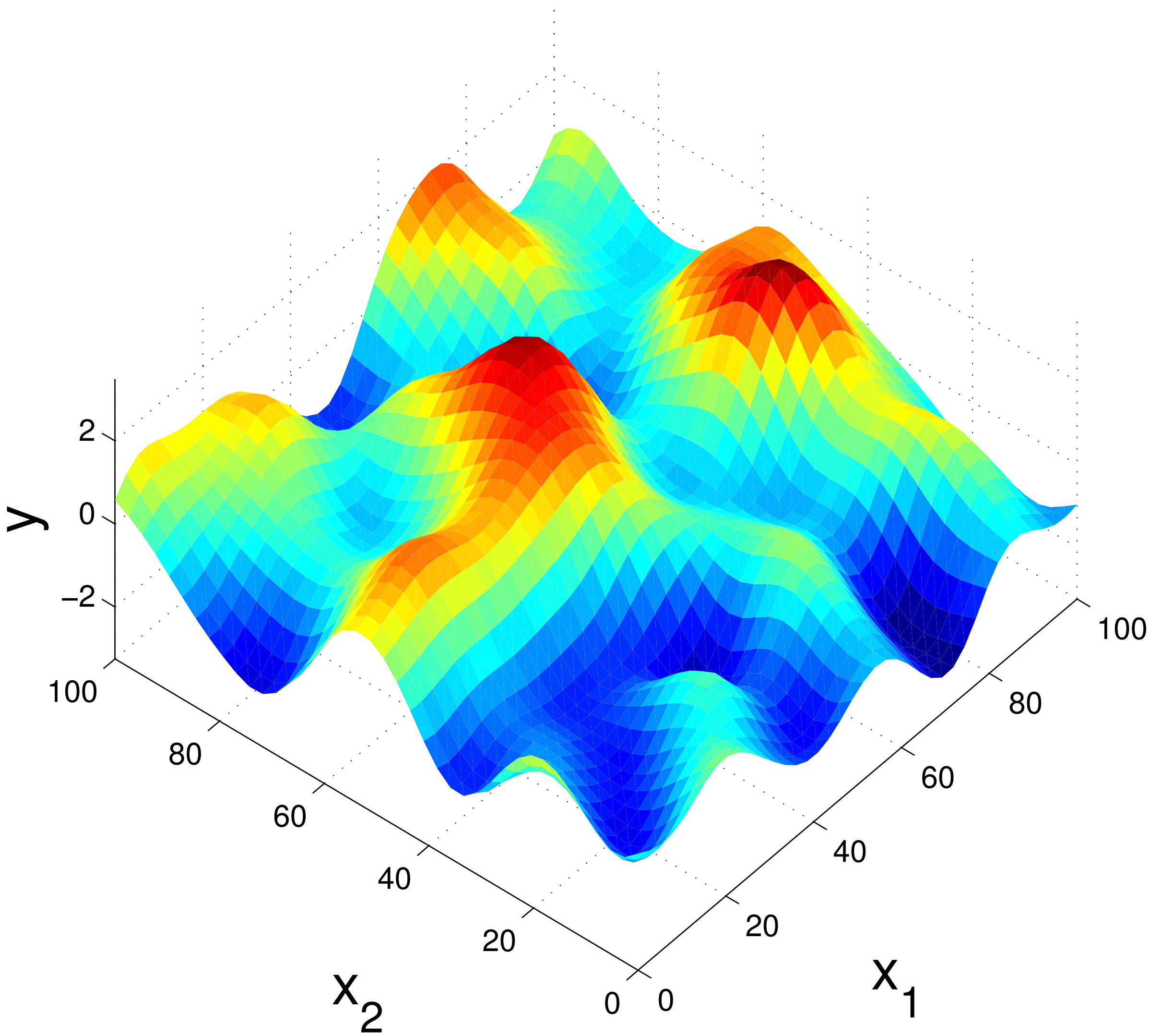
Higher dimensional input spaces



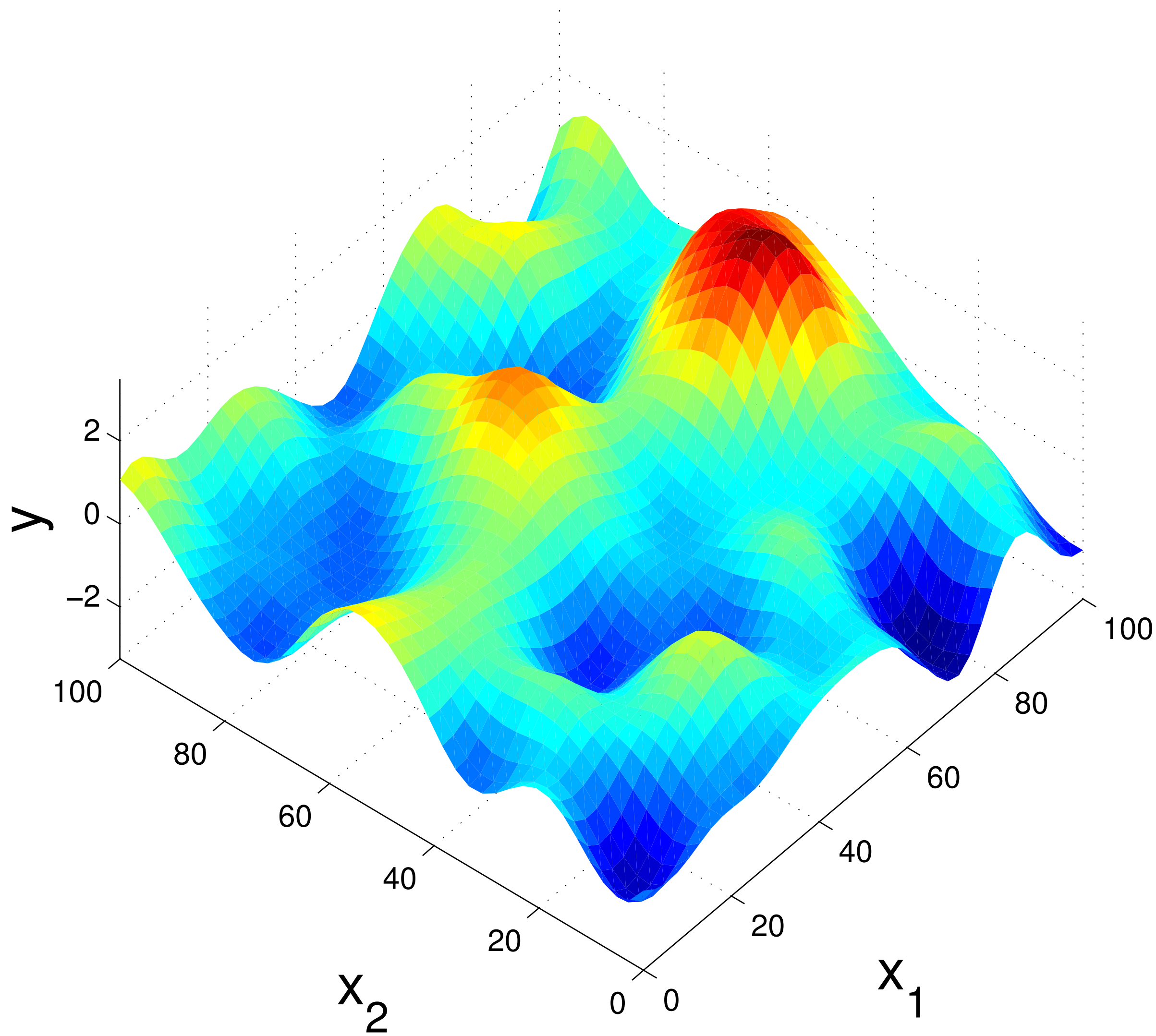
Higher dimensional input spaces



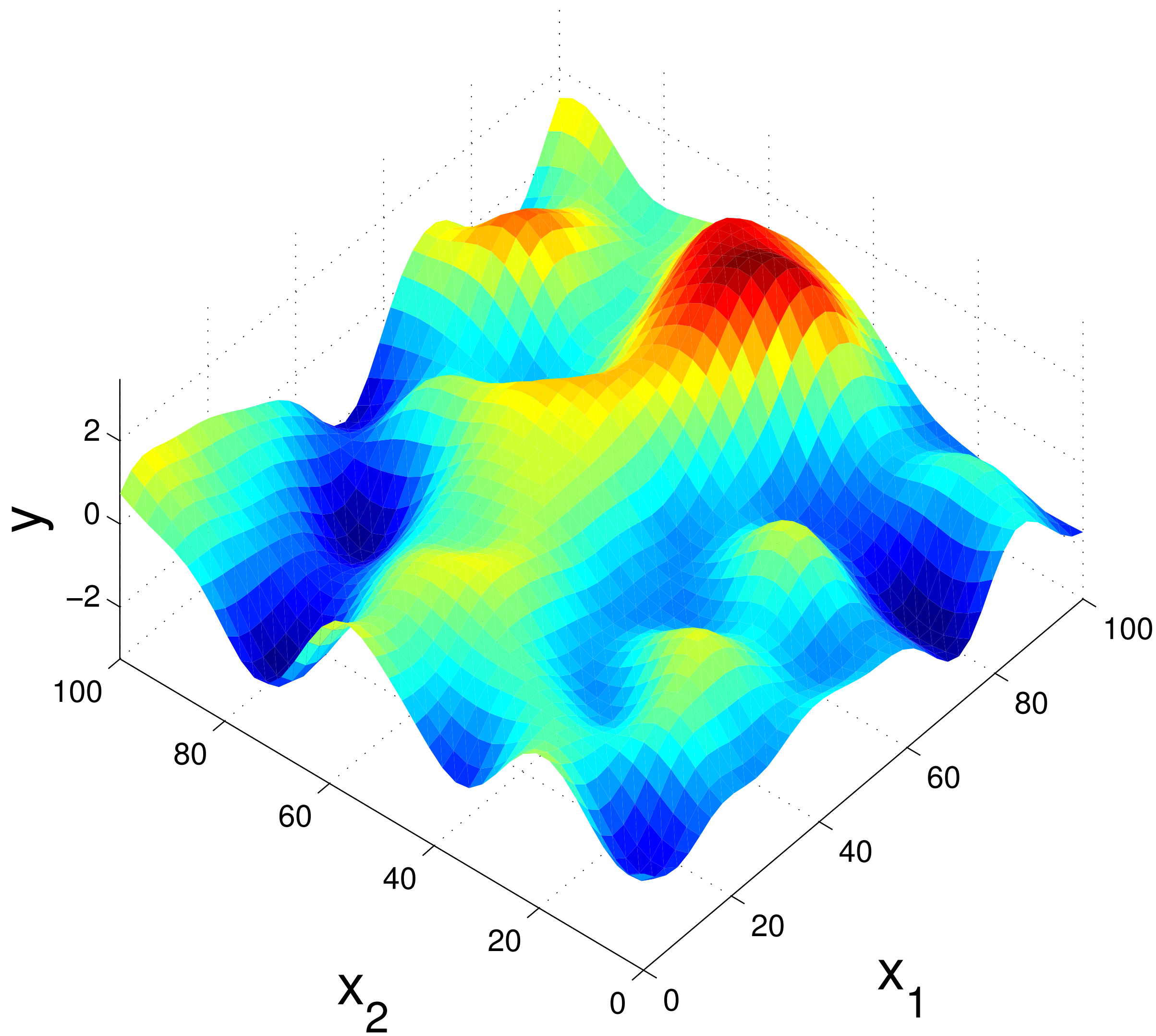
Higher dimensional input spaces



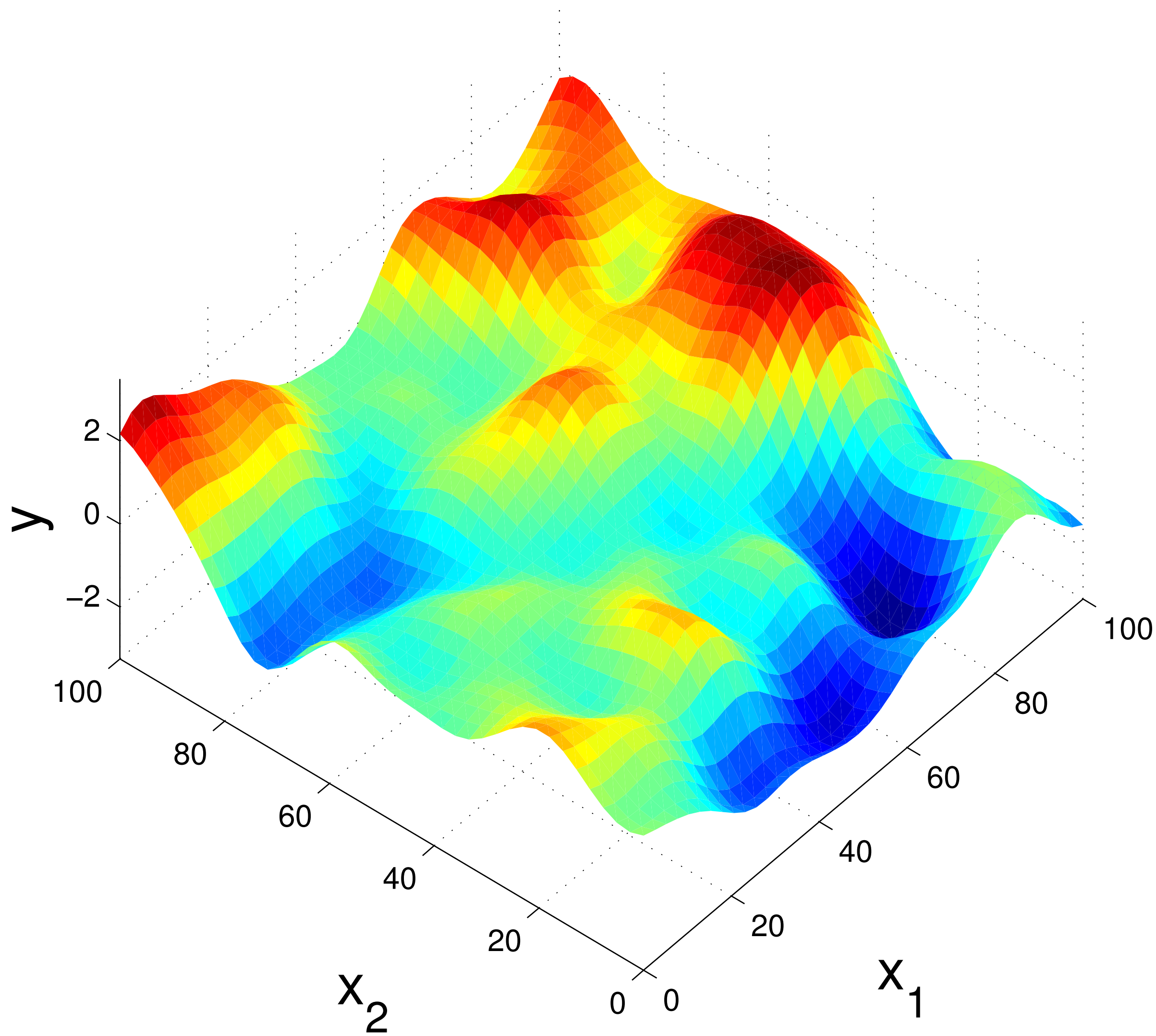
Higher dimensional input spaces



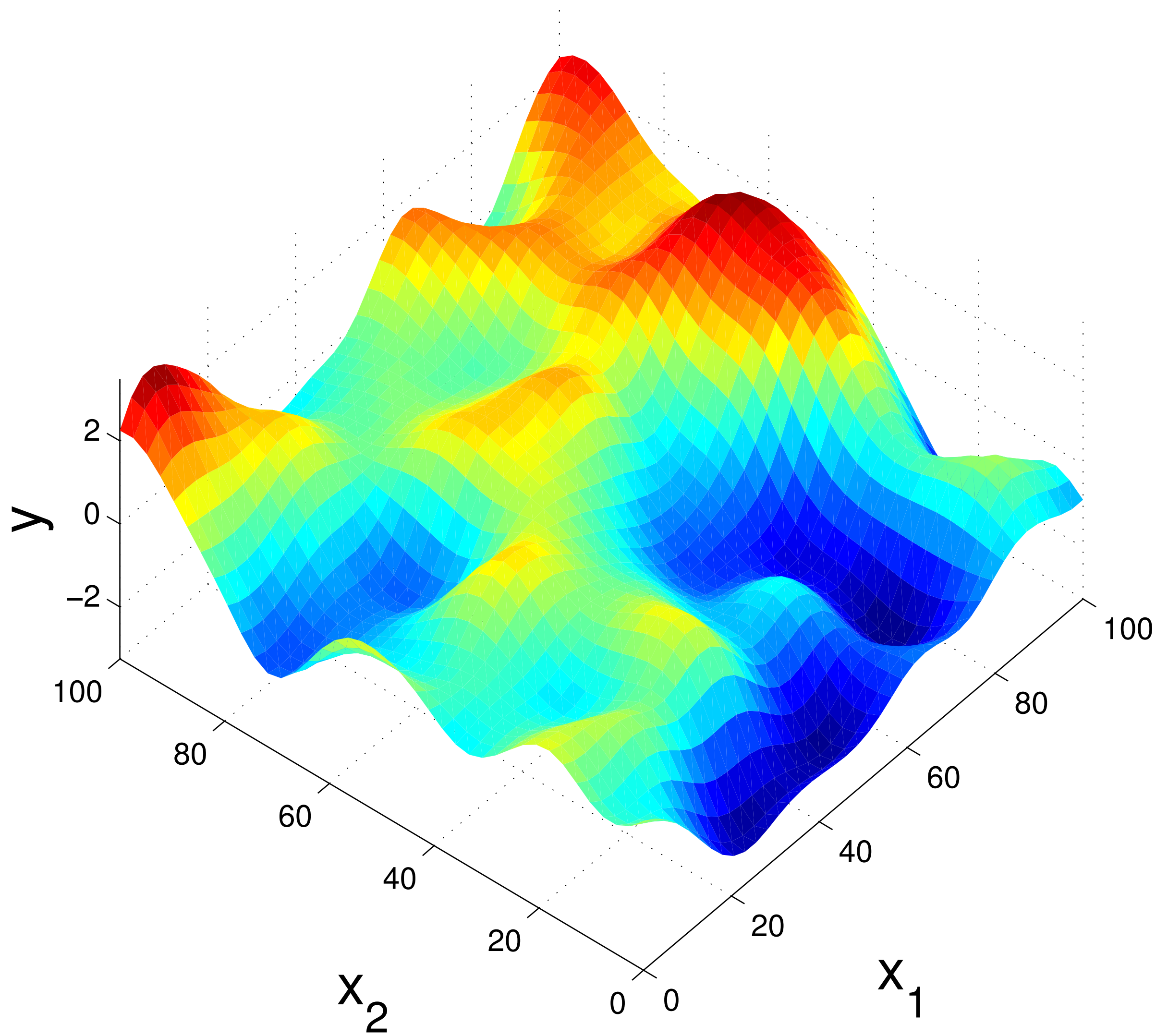
Higher dimensional input spaces



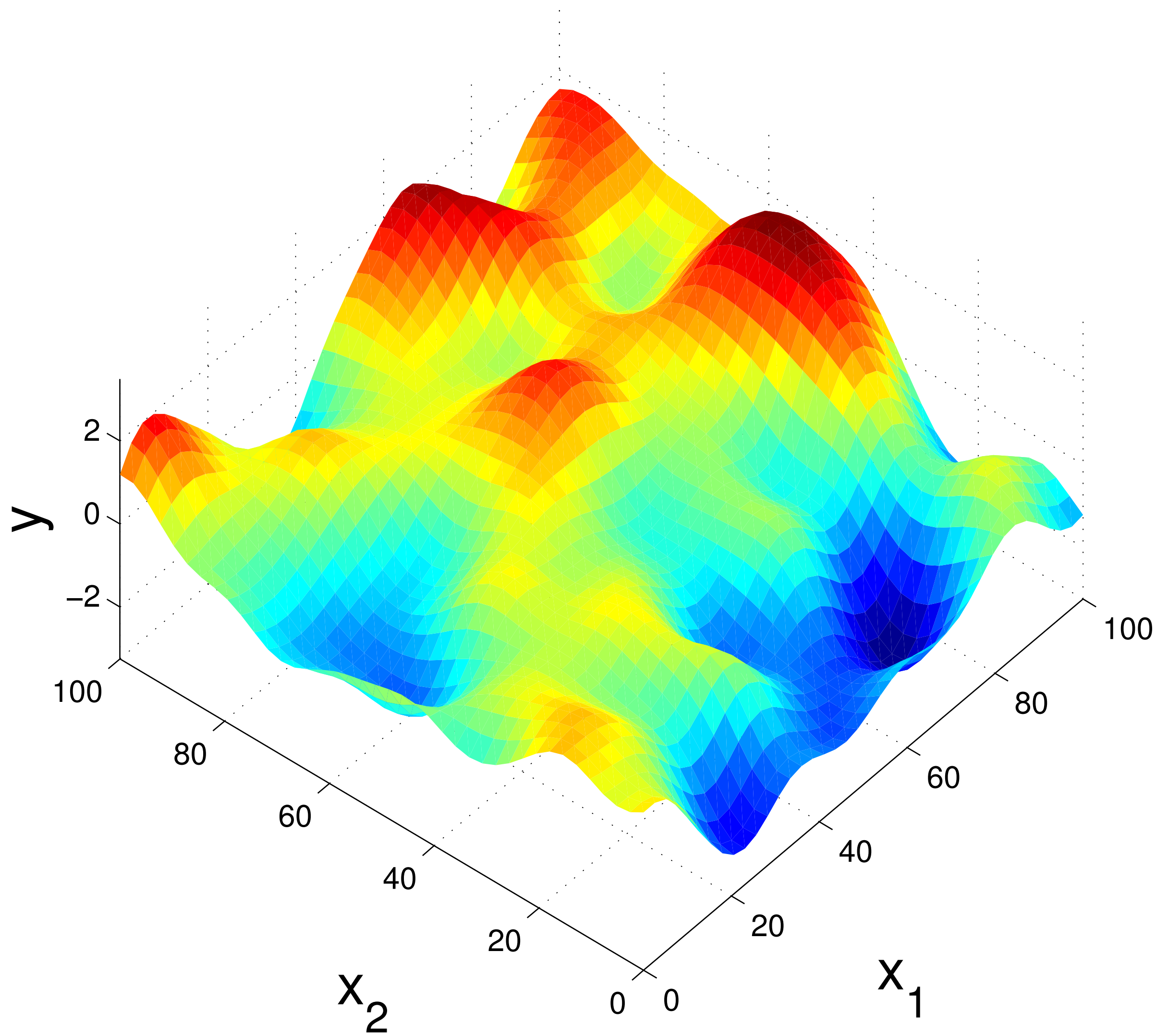
Higher dimensional input spaces



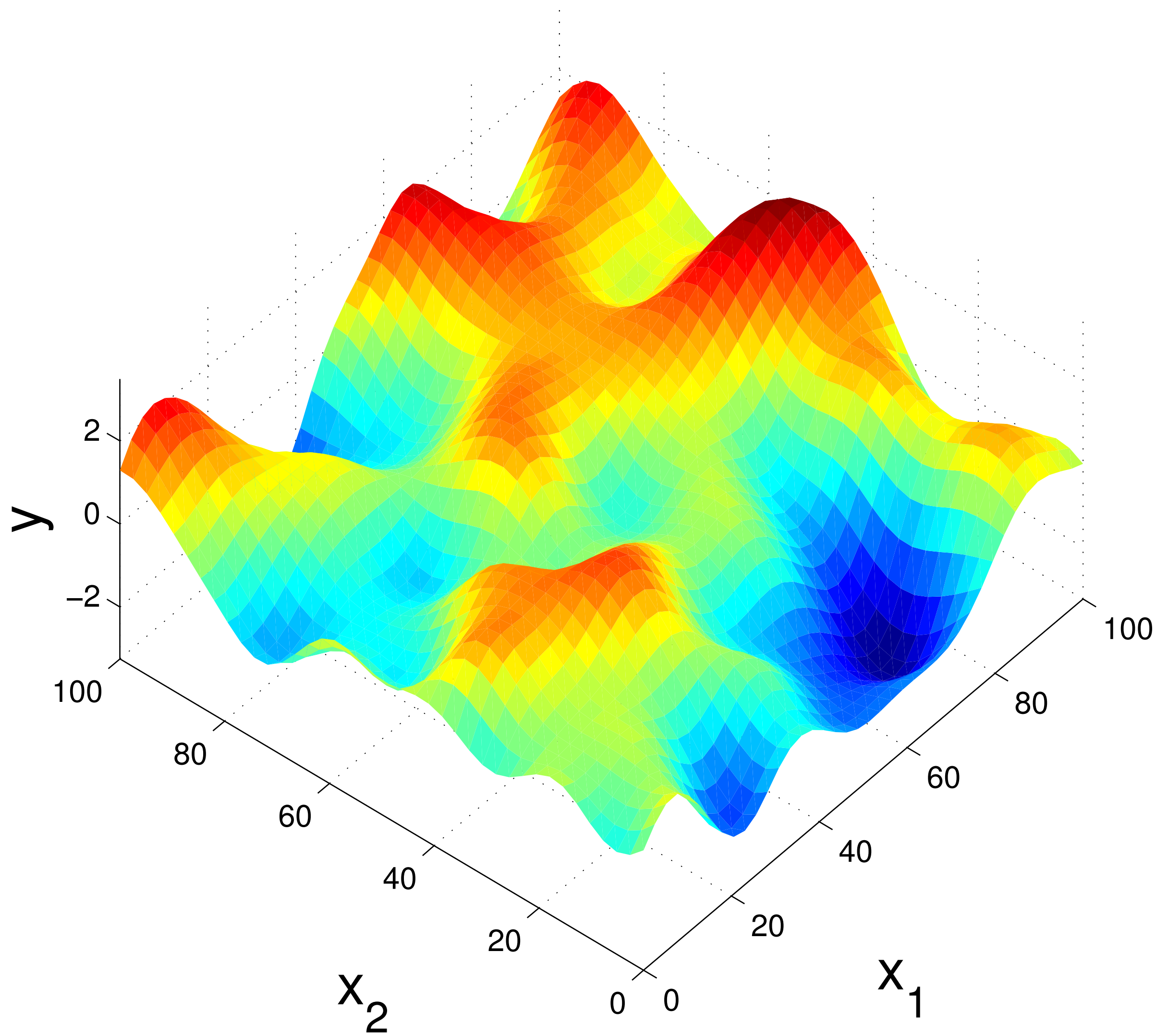
Higher dimensional input spaces



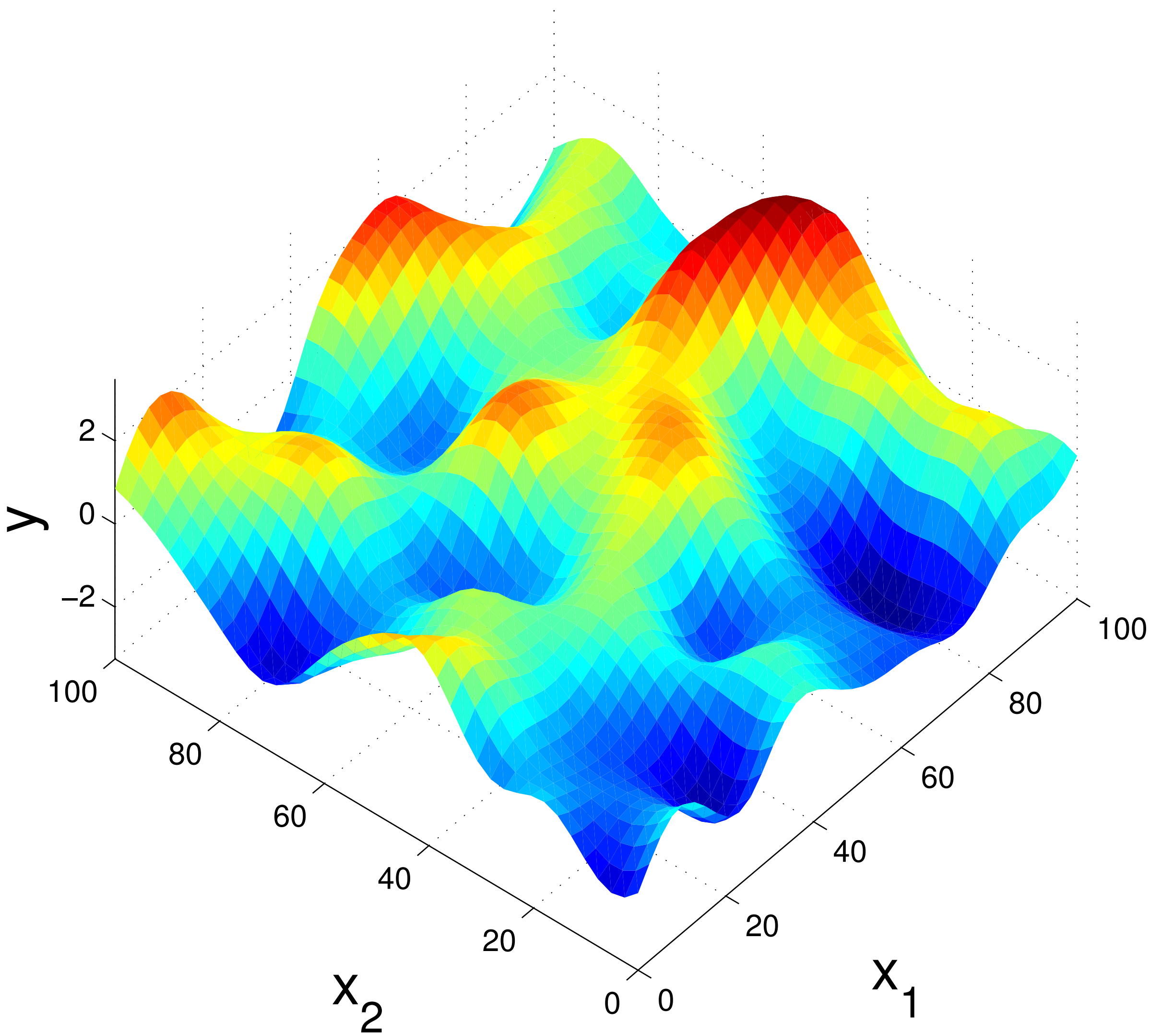
Higher dimensional input spaces



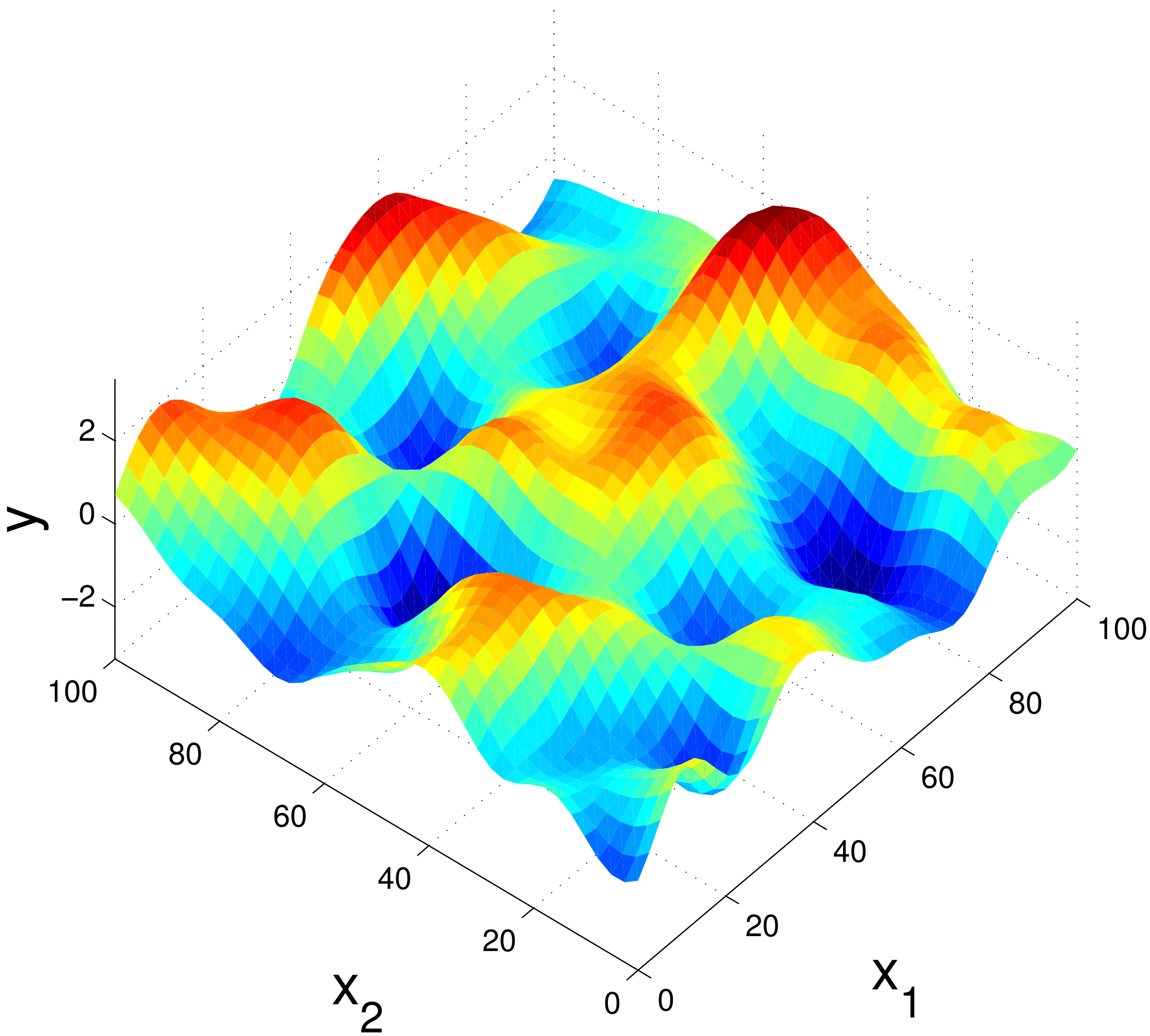
Higher dimensional input spaces



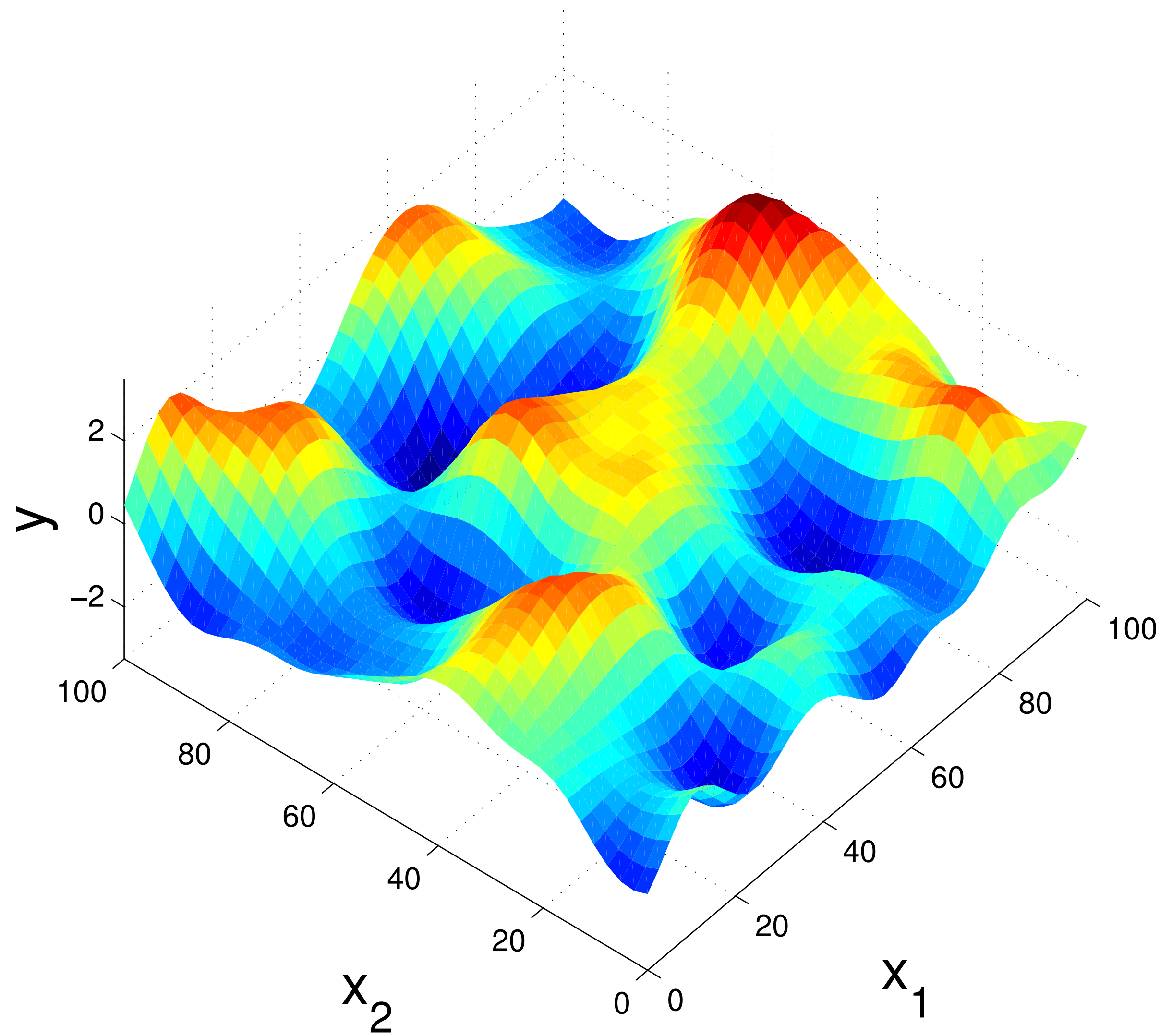
Higher dimensional input spaces



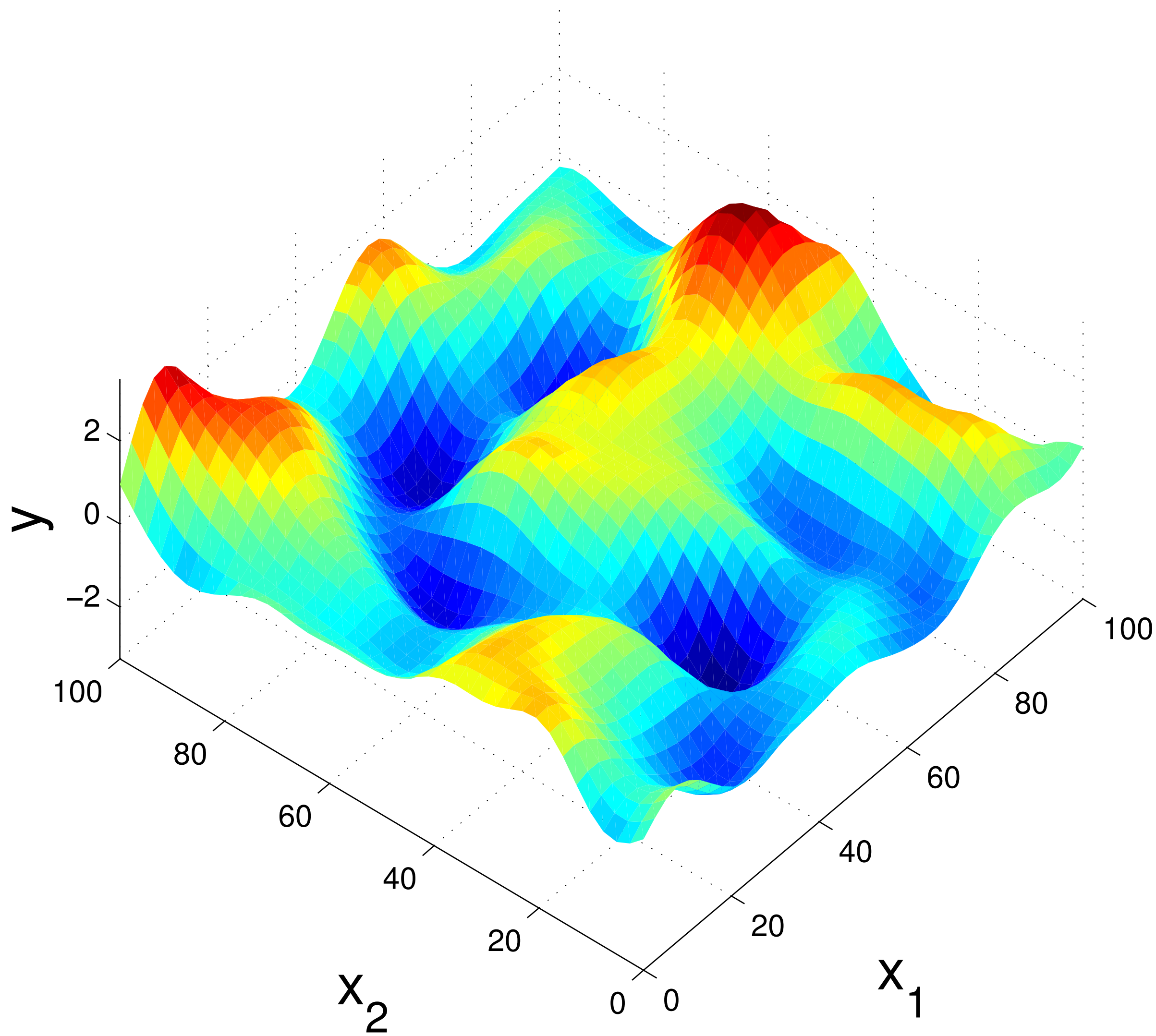
Higher dimensional input spaces



Higher dimensional input spaces



Higher dimensional input spaces



Computational cost

- prediction task
  - train on N points
  - test on M points

- prediction equations

$$\mu_M = \mathsf{K}_{MN}\mathsf{K}_{NN}^{-1}\mathsf{y}_N$$

$$\Sigma_{MM} = \mathsf{K}_{MM} - \mathsf{K}_{MN}\mathsf{K}_{NN}^{-1}\mathsf{K}_{NM}$$

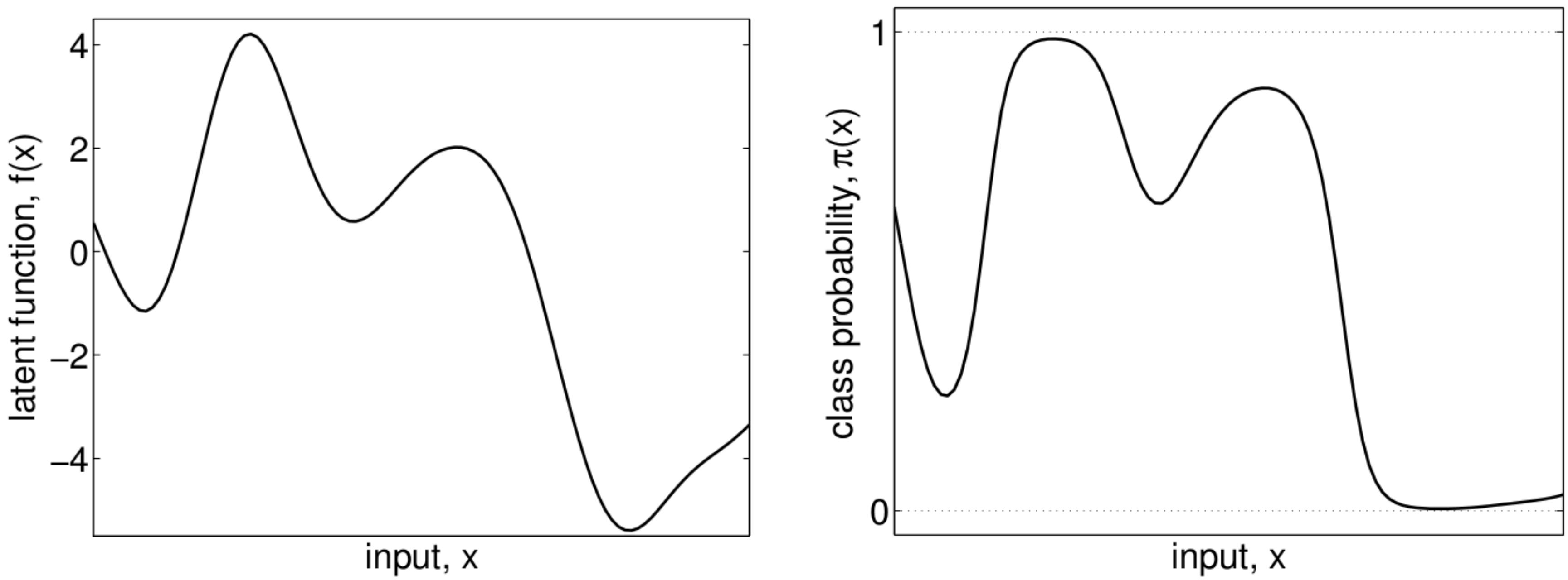
- Full cost $\mathcal{O}((N+M)^3)$ just variances $\mathcal{O}(N^2M)$

- Without special structure, computation is limited to $\mathcal{O}(1000)$ variables

$\Rightarrow$ Computational cost is a major limitation of GPs

Beyond regression: classification

Idea: points near each other in the input space tend to have similar labels



Class probability related to latent function:

$$\pi(\mathsf{x}) = p(\mathsf{y} = 1 | f(\mathsf{x})) = \Phi(f(\mathsf{x}))$$

Logistic link function a typical choice: $\Phi(f) = \frac{1}{1+\exp(-f)}$

Likelihood non-Gaussian $\Rightarrow$ prediction analytically intractable $\Rightarrow$ require approximations

Beyond regression

GPs useful when ever a prior over functions is required

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

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/`