Carnegie Mellon

School of Computer Science

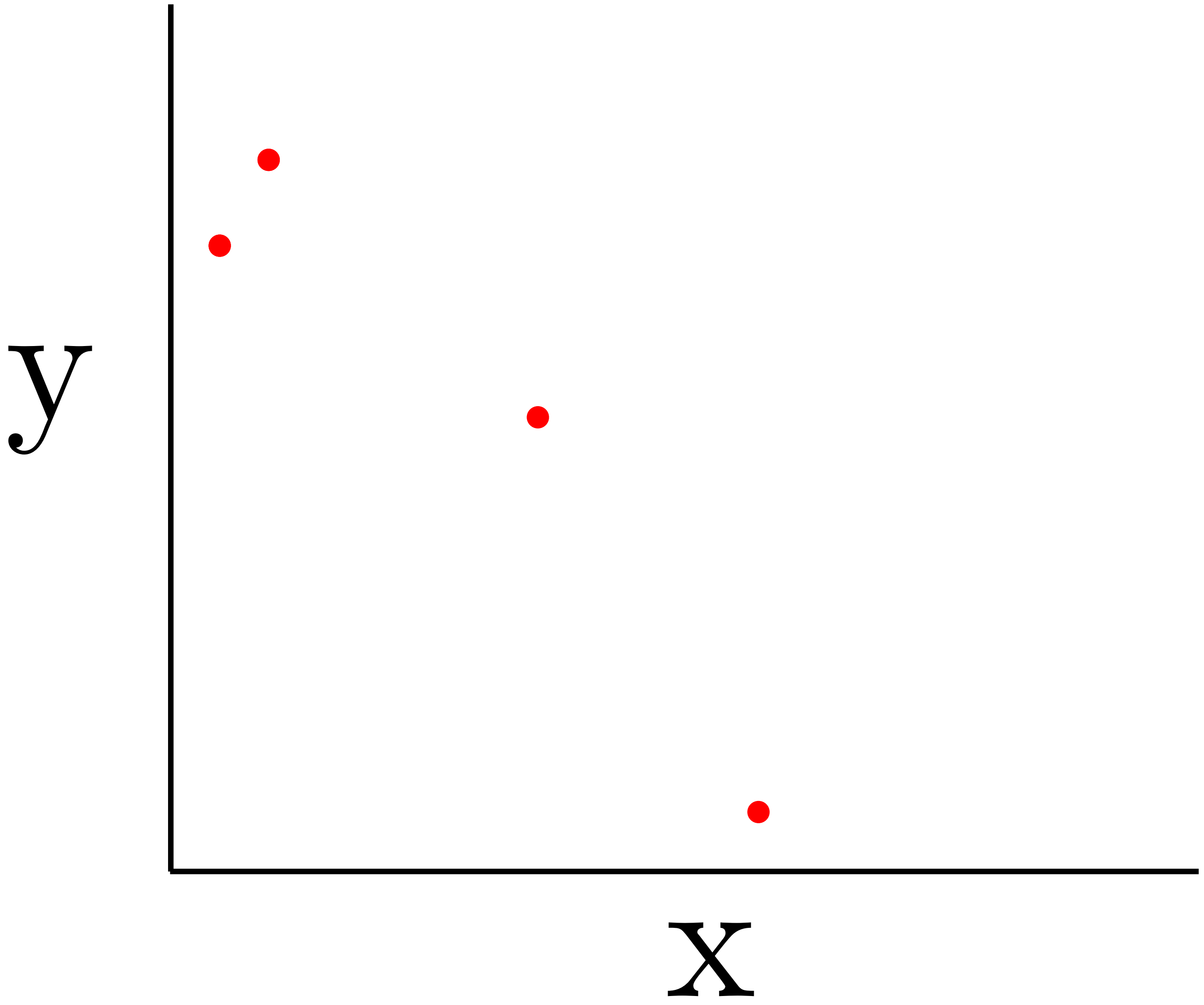
Deep Reinforcement Learning and Control

Bayesian Optimization- Gaussian Processes

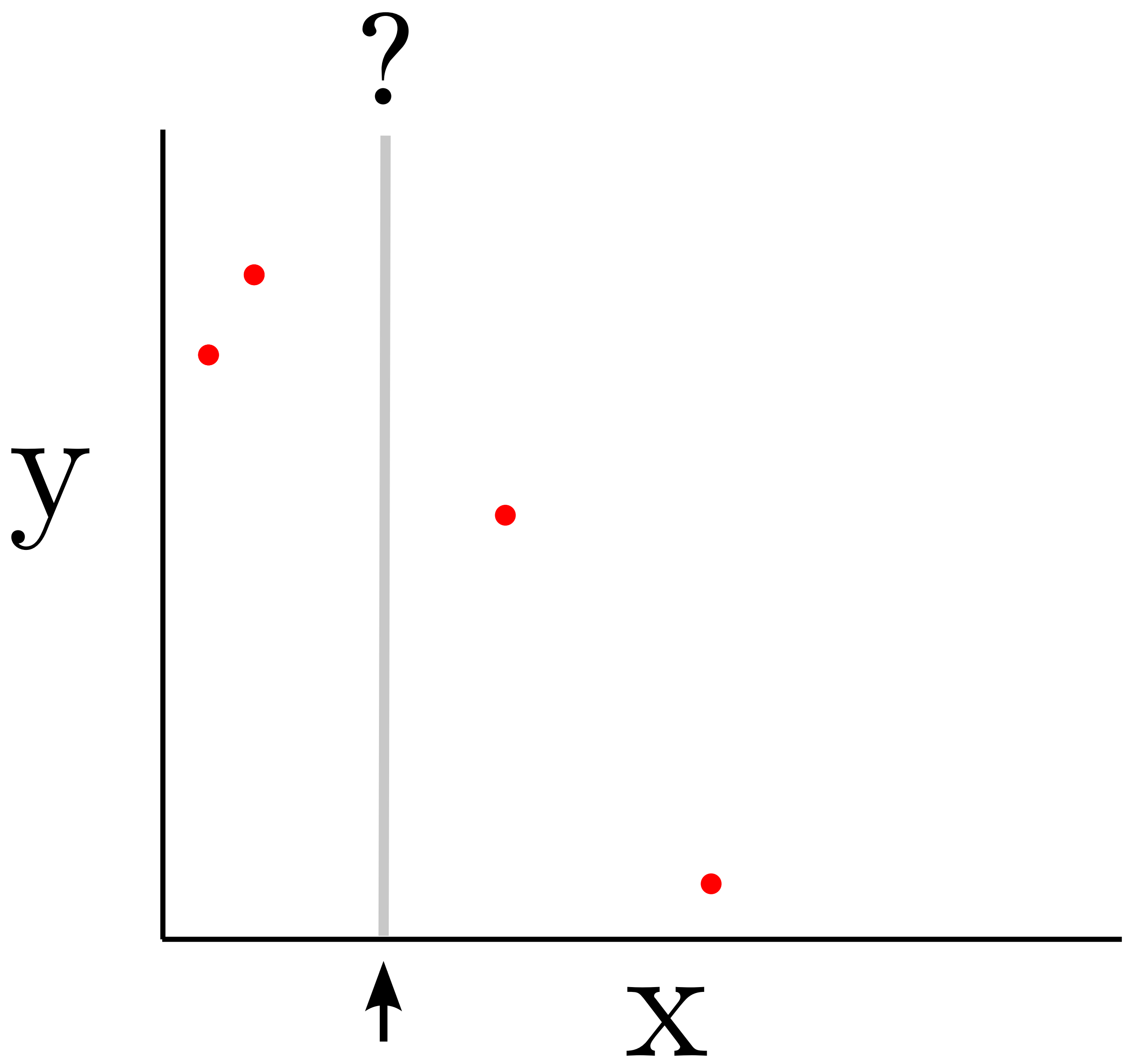
CMU 10-403

Katerina Fragkiadaki

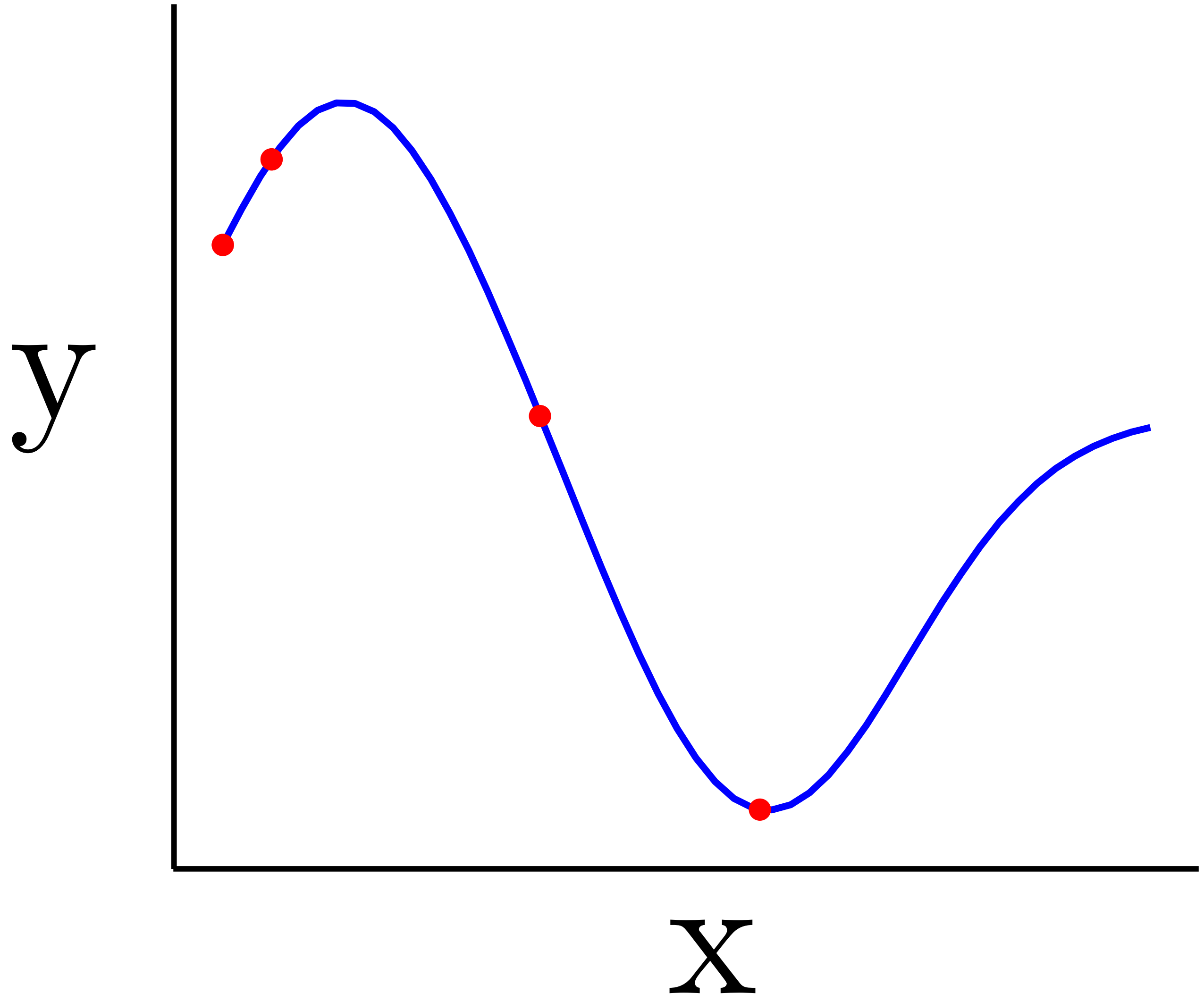
Motivation: non-linear regression



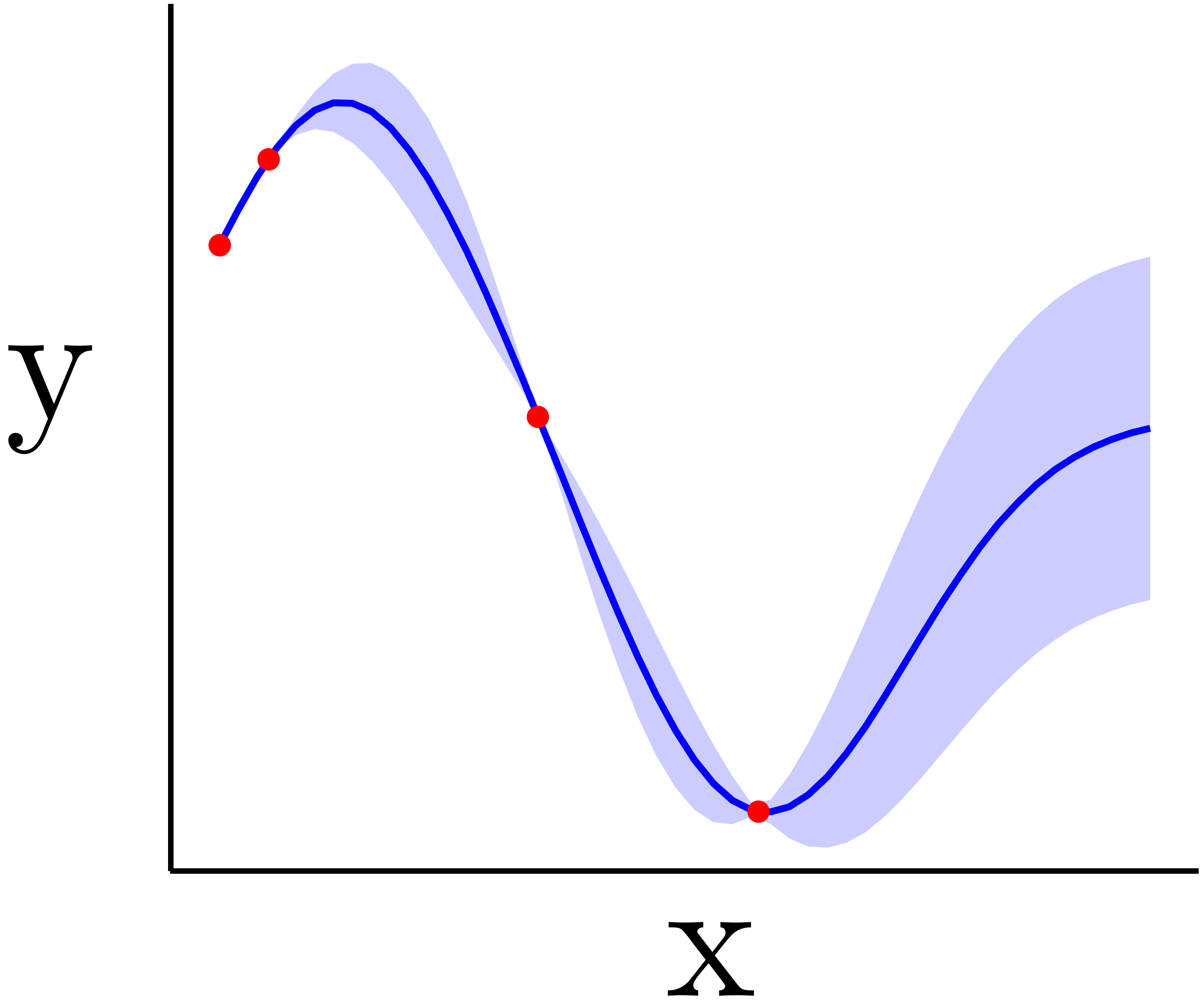
Motivation: non-linear regression



Motivation: non-linear regression

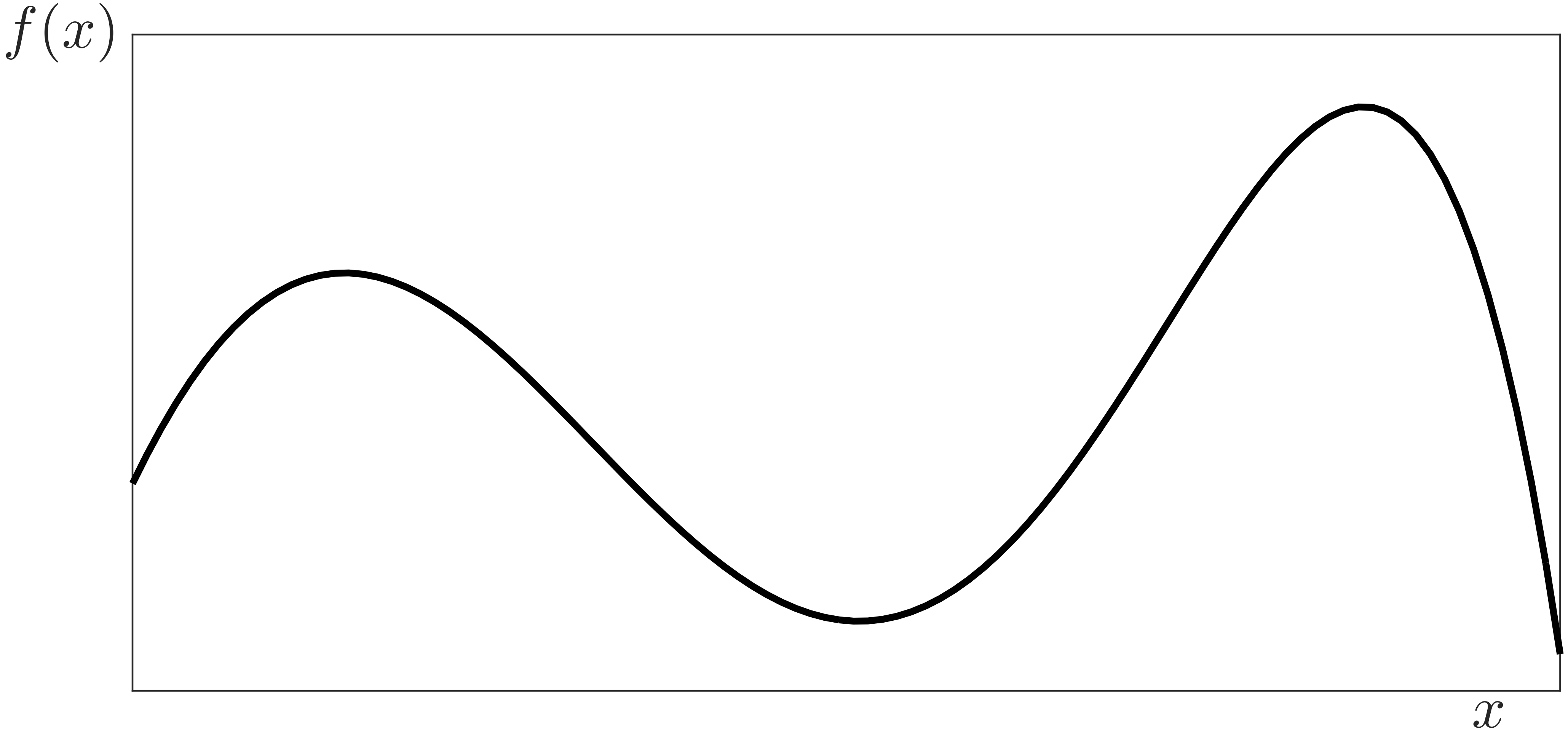


Motivation: non-linear regression



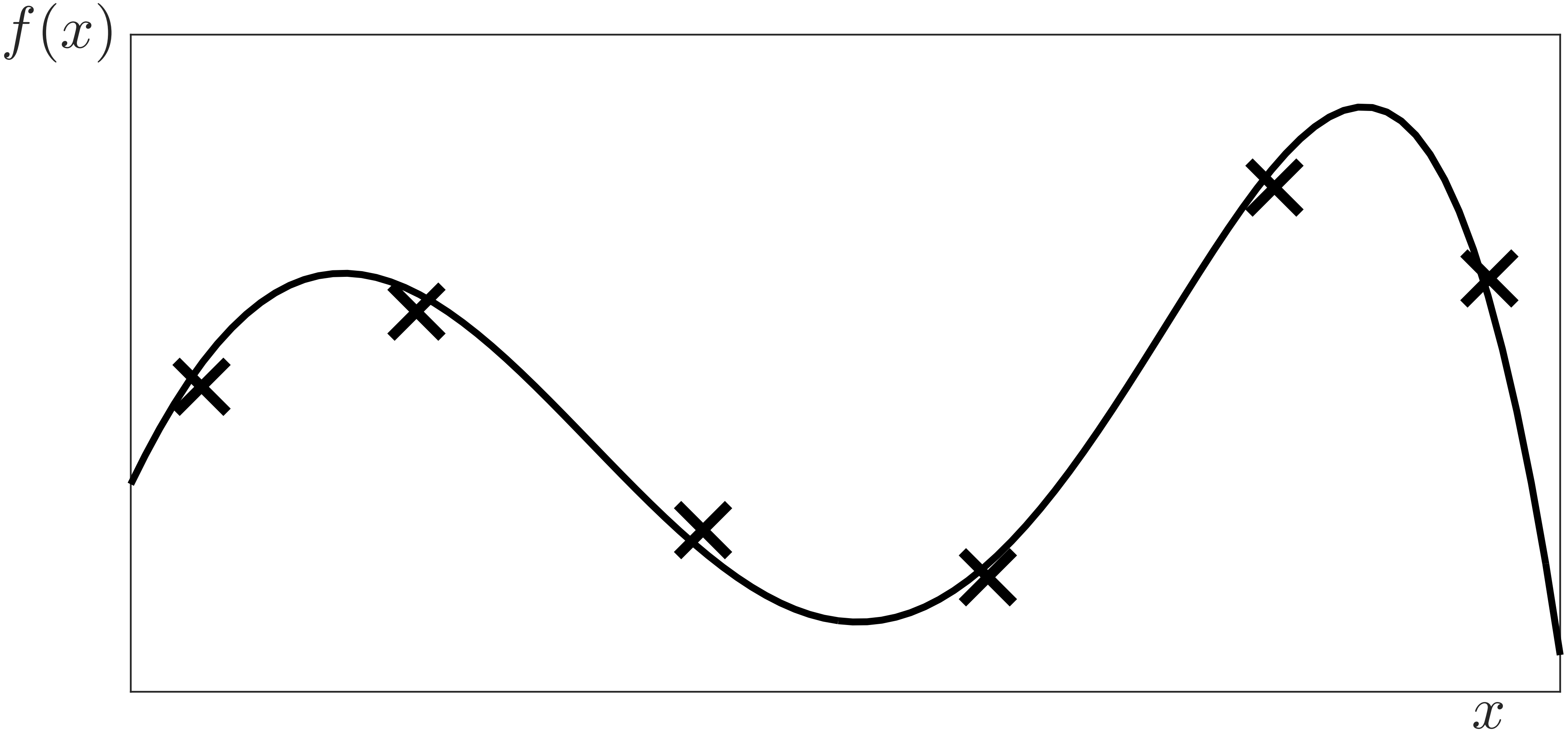
Bandit/Black-box Optimisation

$f : \mathcal{X} \to \mathbb{R}$ is an expensive black-box function, accessible only via noisy evaluations.



Bandit/Black-box Optimisation

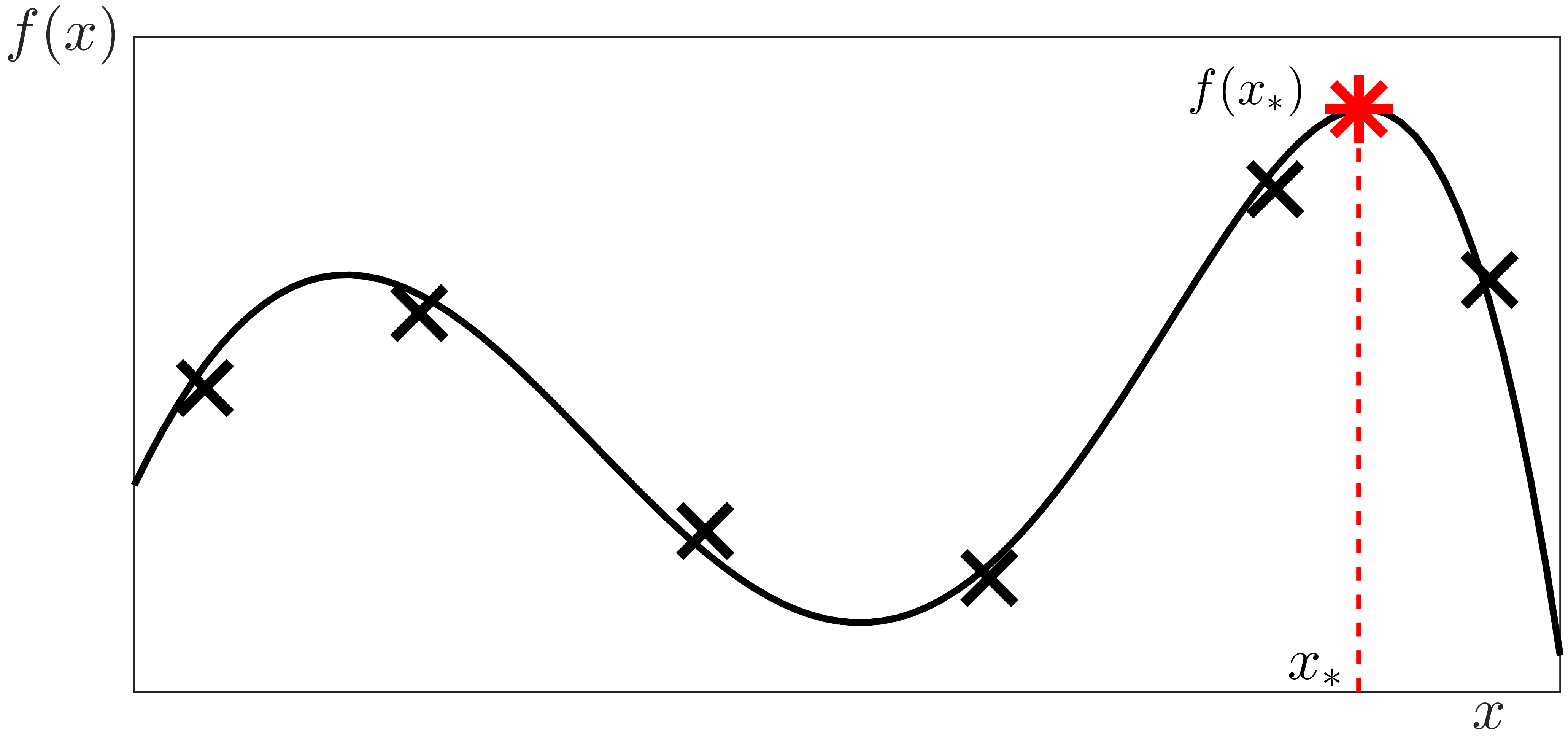
$f : \mathcal{X} \to \mathbb{R}$ is an expensive black-box function, accessible only via noisy evaluations.



Bandit/Black-box Optimisation

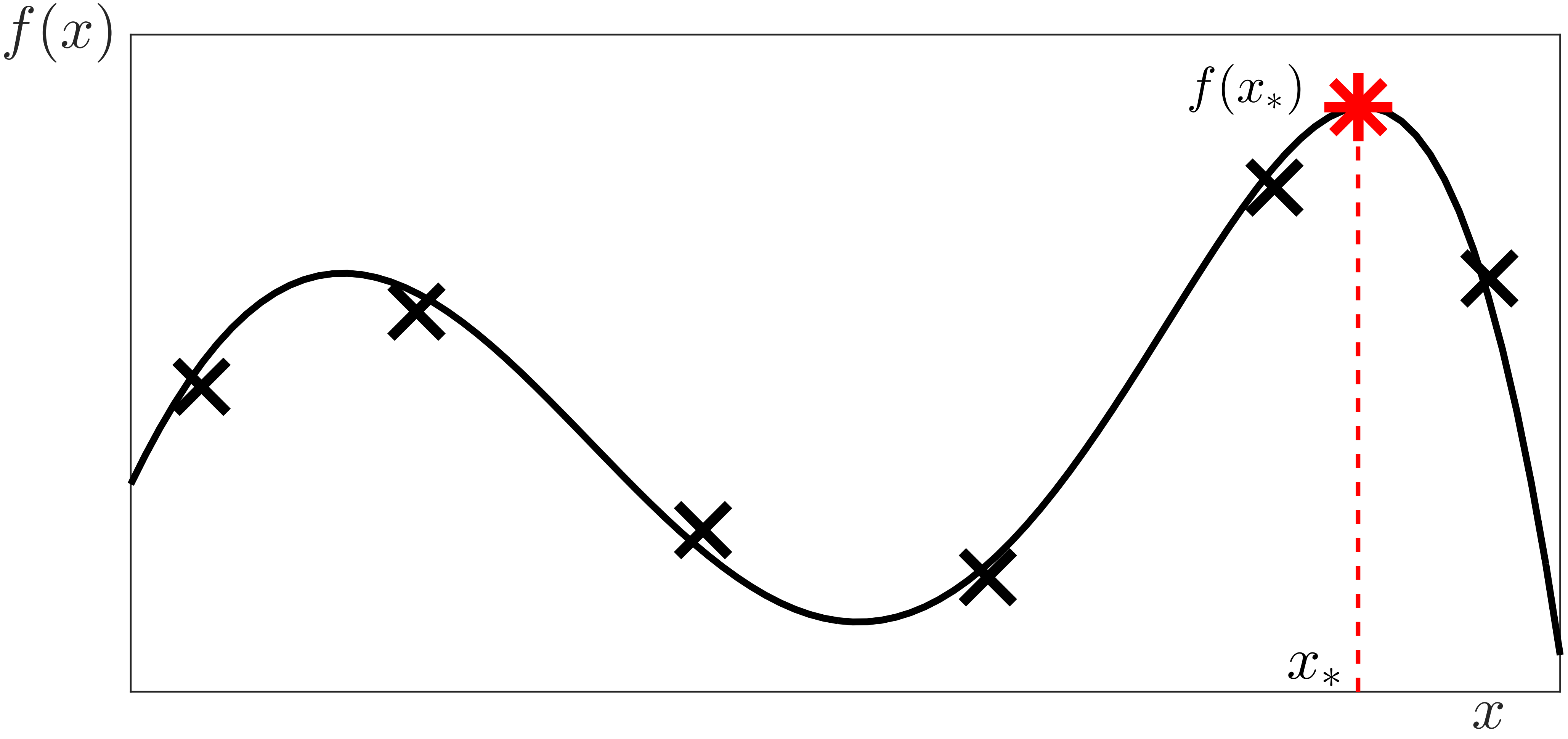
$f : \mathcal{X} \to \mathbb{R}$ is an expensive black-box function, accessible only via noisy evaluations.

Let $x_\star = \operatorname{argmax}_x f(x)$.



Bandit/Black-box Optimisation

$f : \mathcal{X} \to \mathbb{R}$ is an expensive black-box function, accessible only via noisy evaluations.
Let $x_\star = \operatorname{argmax}_x f(x)$.



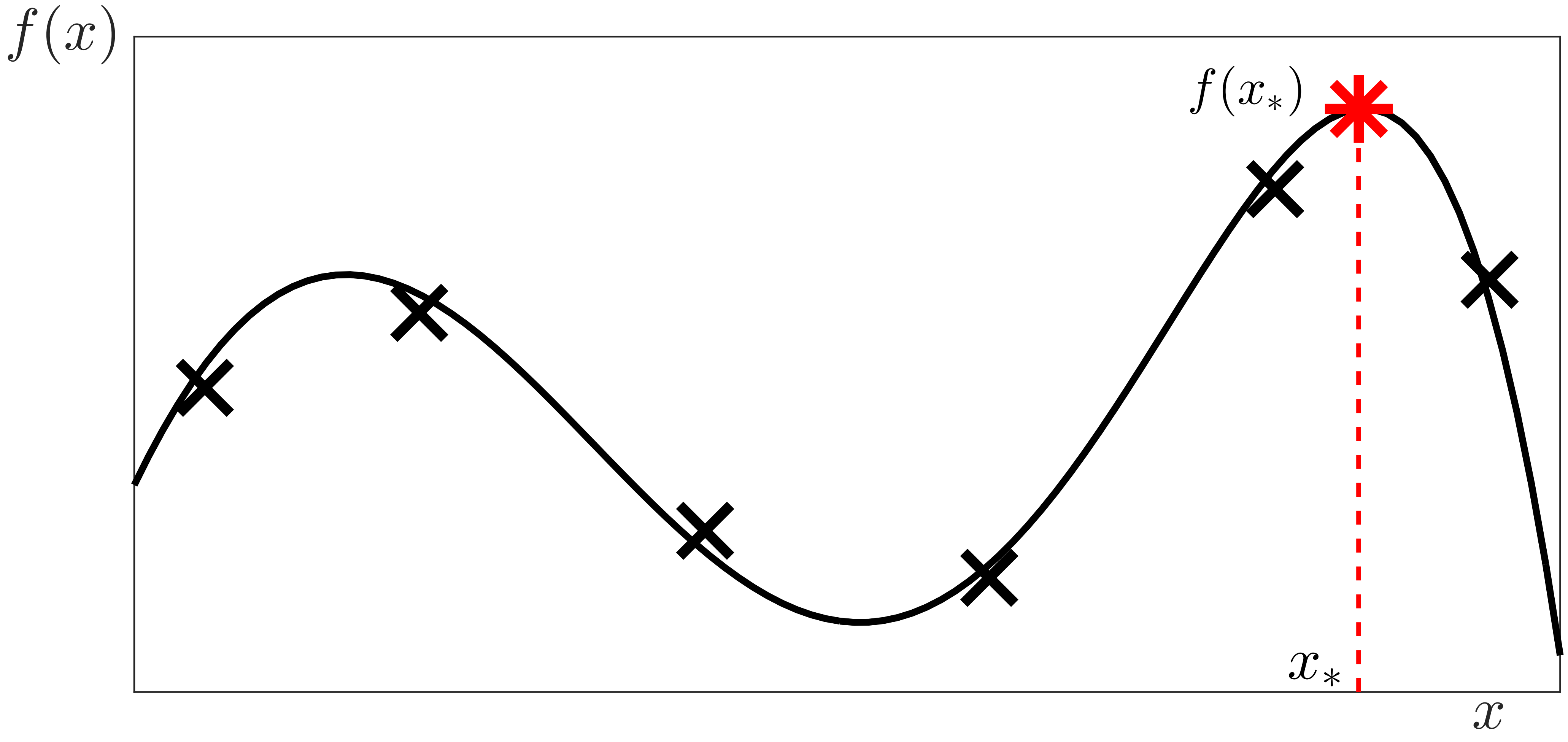
Simple Regret after n evaluations

$$\mathrm{SR}(n) = f(x_\star) \; - \max_{t=1,\ldots,n} f(x_t).$$

Bandit/Black-box Optimisation

$f : \mathcal{X} \to \mathbb{R}$ is an expensive black-box function, accessible only via noisy evaluations.

Let $x_\star = \mathrm{argmax}_x \, f(x)$.

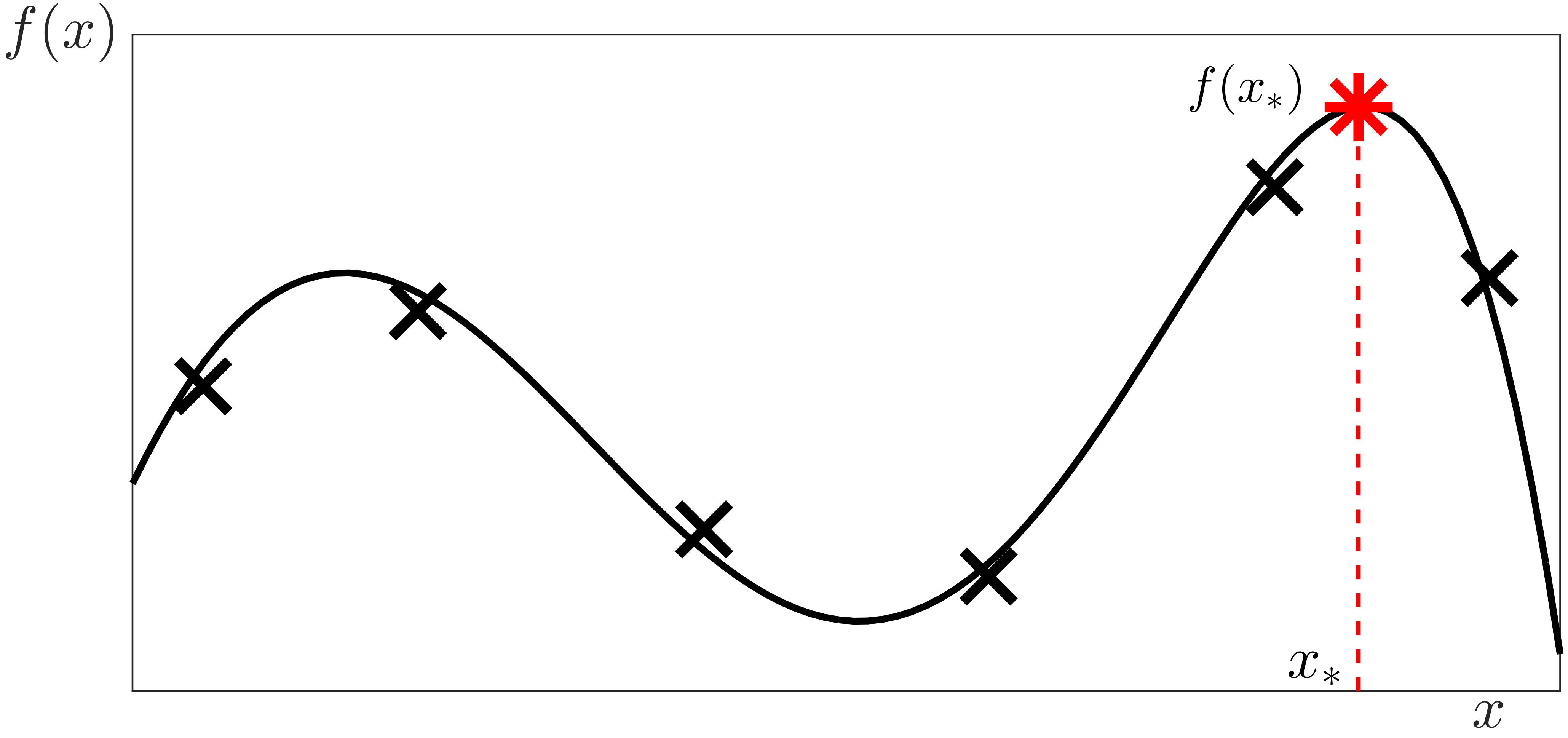


Cumulative Regret after n evaluations

$$\mathrm{CR}(n) = \sum_{t=1}^{n} \Big(f(x_\star) \; - f(x_t) \Big)$$

Bandit/Black-box Optimisation

$f : \mathcal{X} \to \mathbb{R}$ is an expensive black-box function, accessible only via noisy evaluations.
Let $x_\star = \mathrm{argmax}_x f(x)$.

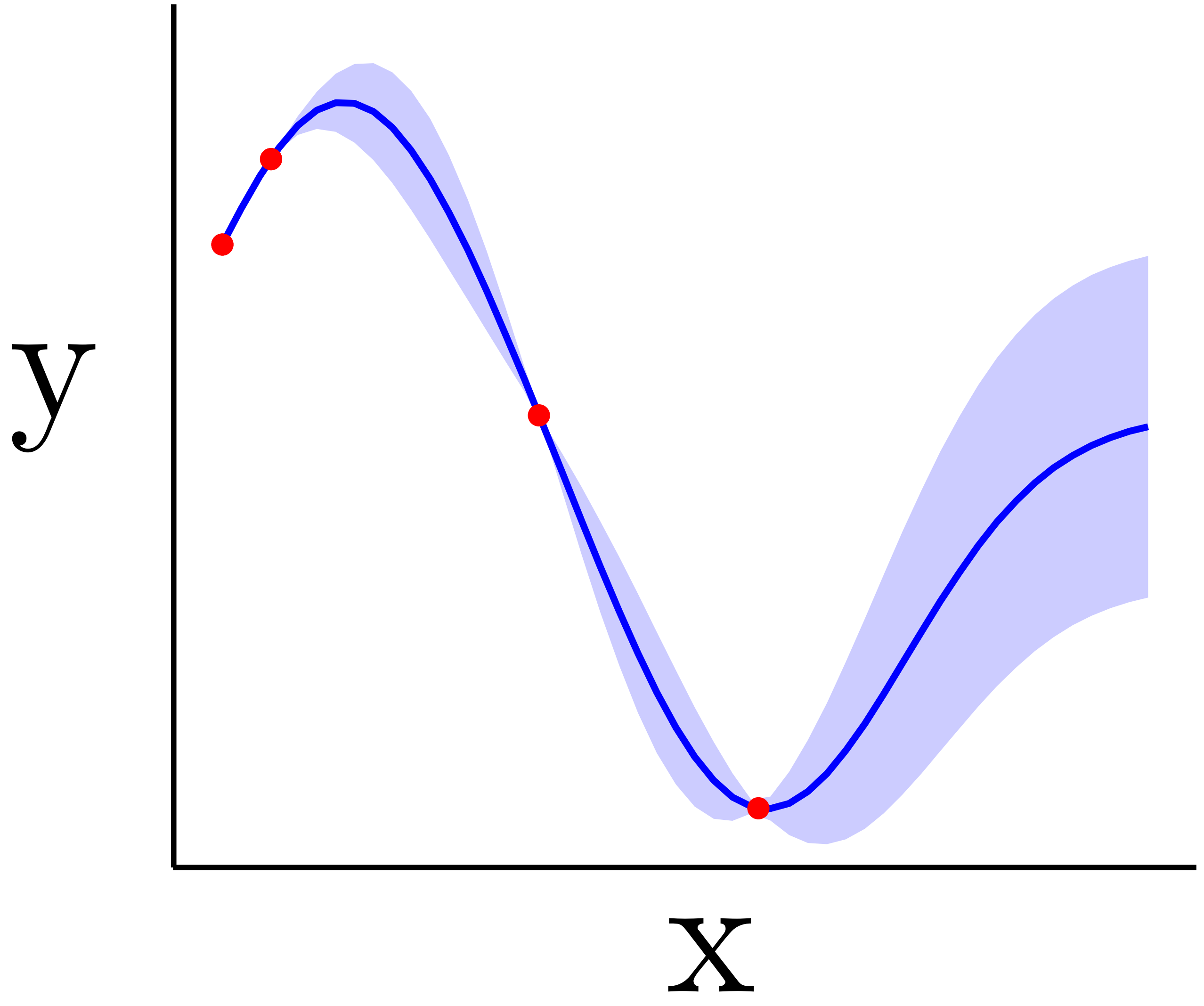


Simple Regret after n evaluations

$$\mathrm{SR}(n) = f(x_\star) - \max_{t=1,\ldots,n} f(x_t).$$

Motivation: non-linear regression

Can we do this with a plain old Gaussian?



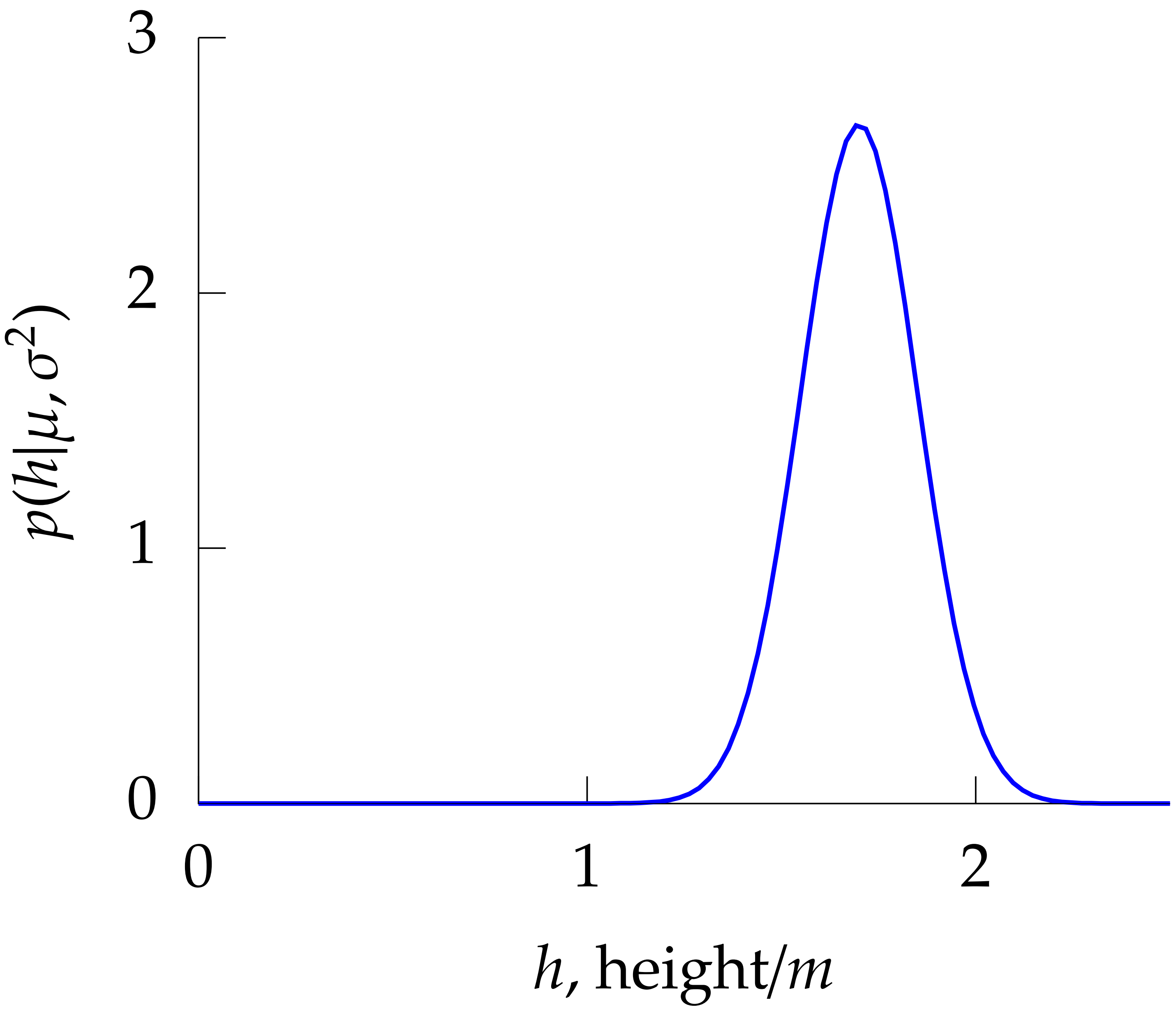
The Gaussian Density

- Perhaps the most common probability density.

$$
\begin{aligned}
p(y|\mu, \sigma^2) &= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y-\mu)^2}{2\sigma^2}\right) \\
&\triangleq \mathcal{N}\left(y|\mu, \sigma^2\right)
\end{aligned}
$$

- The Gaussian density.

Gaussian Density



The Gaussian PDF with $\mu = 1.7$ and variance $\sigma^2 = 0.0225$. Mean shown as red line. It could represent the heights of a population of students.

Gaussian Density

$$\mathcal{N}\left(y|\mu,\sigma^2\right) = \frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{(y-\mu)^2}{2\sigma^2}\right)$$

σ^2 is the variance of the density and μ is the mean.

Two Important Gaussian Properties

Sum of Gaussians

- Sum of Gaussian variables is also Gaussian.

$$y_i \sim \mathcal{N}\left(\mu_i, \sigma_i^2\right)$$

Two Important Gaussian Properties

Sum of Gaussians

- Sum of Gaussian variables is also Gaussian.

$$y_i \sim \mathcal{N}\left(\mu_i, \sigma_i^2\right)$$

And the sum is distributed as

$$\sum_{i=1}^{n} y_i \sim \mathcal{N}\left(\sum_{i=1}^{n} \mu_i, \sum_{i=1}^{n} \sigma_i^2\right)$$

Two Important Gaussian Properties

Sum of Gaussians

- Sum of Gaussian variables is also Gaussian.

  $$y_i \sim \mathcal{N}\left(\mu_i, \sigma_i^2\right)$$

  And the sum is distributed as

  $$\sum_{i=1}^{n} y_i \sim \mathcal{N}\left(\sum_{i=1}^{n} \mu_i, \sum_{i=1}^{n} \sigma_i^2\right)$$

  (*Aside*: As sum increases, sum of non-Gaussian, finite variance variables is also Gaussian [central limit theorem].)

Two Important Gaussian Properties

Sum of Gaussians

- Sum of Gaussian variables is also Gaussian.

$$y_i \sim \mathcal{N}\left(\mu_i, \sigma_i^2\right)$$

And the sum is distributed as

$$\sum_{i=1}^{n} y_i \sim \mathcal{N}\left(\sum_{i=1}^{n} \mu_i, \sum_{i=1}^{n} \sigma_i^2\right)$$

(*Aside*: As sum increases, sum of non-Gaussian, finite variance variables is also Gaussian [central limit theorem].)

Two Important Gaussian Properties

Scaling a Gaussian

- Scaling a Gaussian leads to a Gaussian.

Two Important Gaussian Properties

Scaling a Gaussian

- Scaling a Gaussian leads to a Gaussian.

$$y \sim \mathcal{N}\left(\mu, \sigma^2\right)$$

Two Important Gaussian Properties

Scaling a Gaussian

- Scaling a Gaussian leads to a Gaussian.

$$y \sim \mathcal{N}\left(\mu, \sigma^2\right)$$

And the scaled density is distributed as

$$wy \sim \mathcal{N}\left(w\mu, w^2\sigma^2\right)$$

Multivariate Consequence

- If

$$\mathbf{x} \sim \mathcal{N}\left(\mu, \boldsymbol{\Sigma}\right)$$

Multivariate Consequence

- If

$$\mathbf{x} \sim \mathcal{N}\left(\mu, \mathbf{\Sigma}\right)$$

- And

$$\mathbf{y} = \mathbf{W}\mathbf{x}$$

Multivariate Consequence

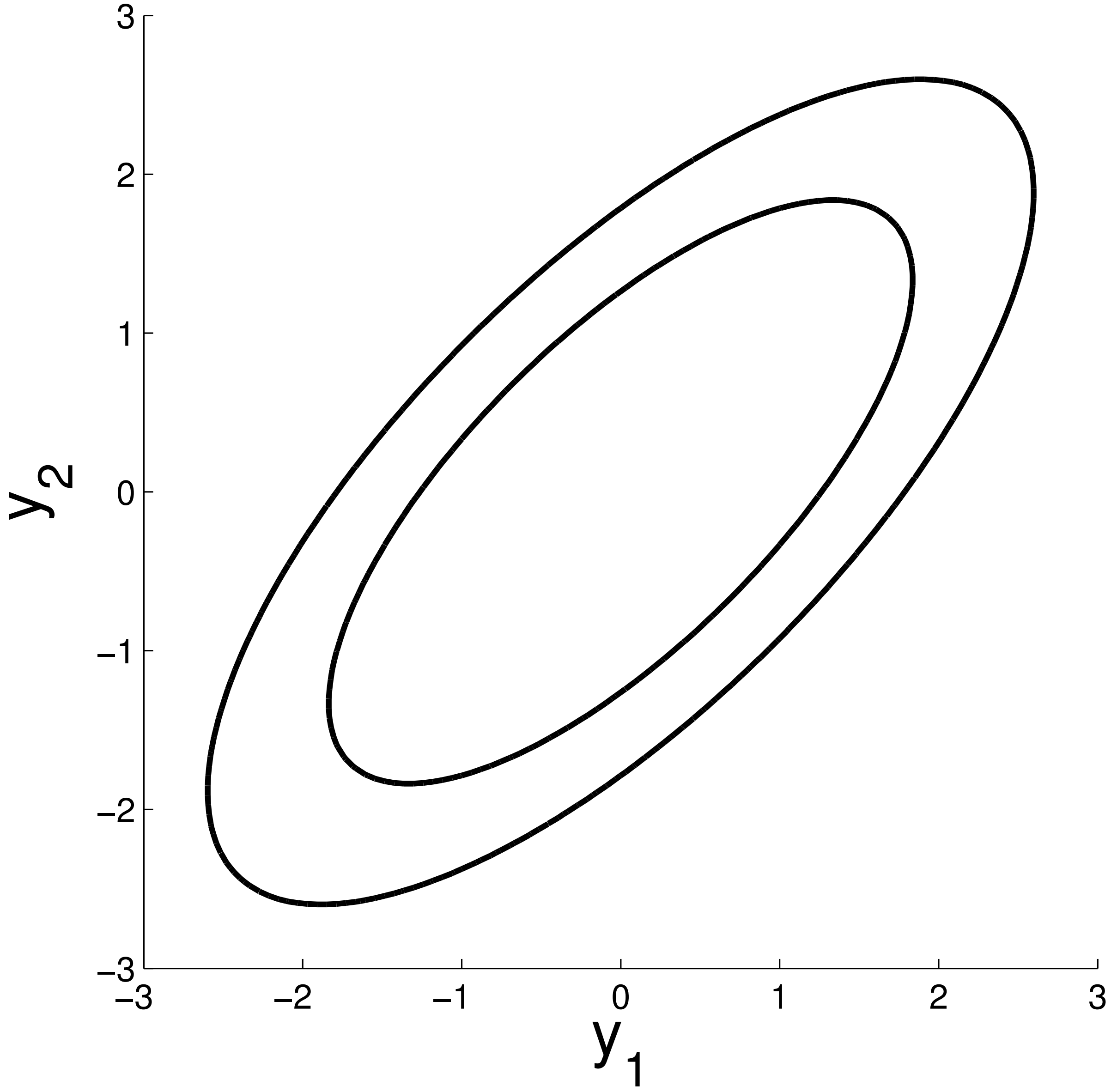
- If
$$\mathbf{x} \sim \mathcal{N}\left(\boldsymbol{\mu}, \boldsymbol{\Sigma}\right)$$

- And
$$\mathbf{y} = \mathbf{W}\mathbf{x}$$

- Then
$$\mathbf{y} \sim \mathcal{N}\left(\mathbf{W}\boldsymbol{\mu}, \mathbf{W}\boldsymbol{\Sigma}\mathbf{W}^{\top}\right)$$

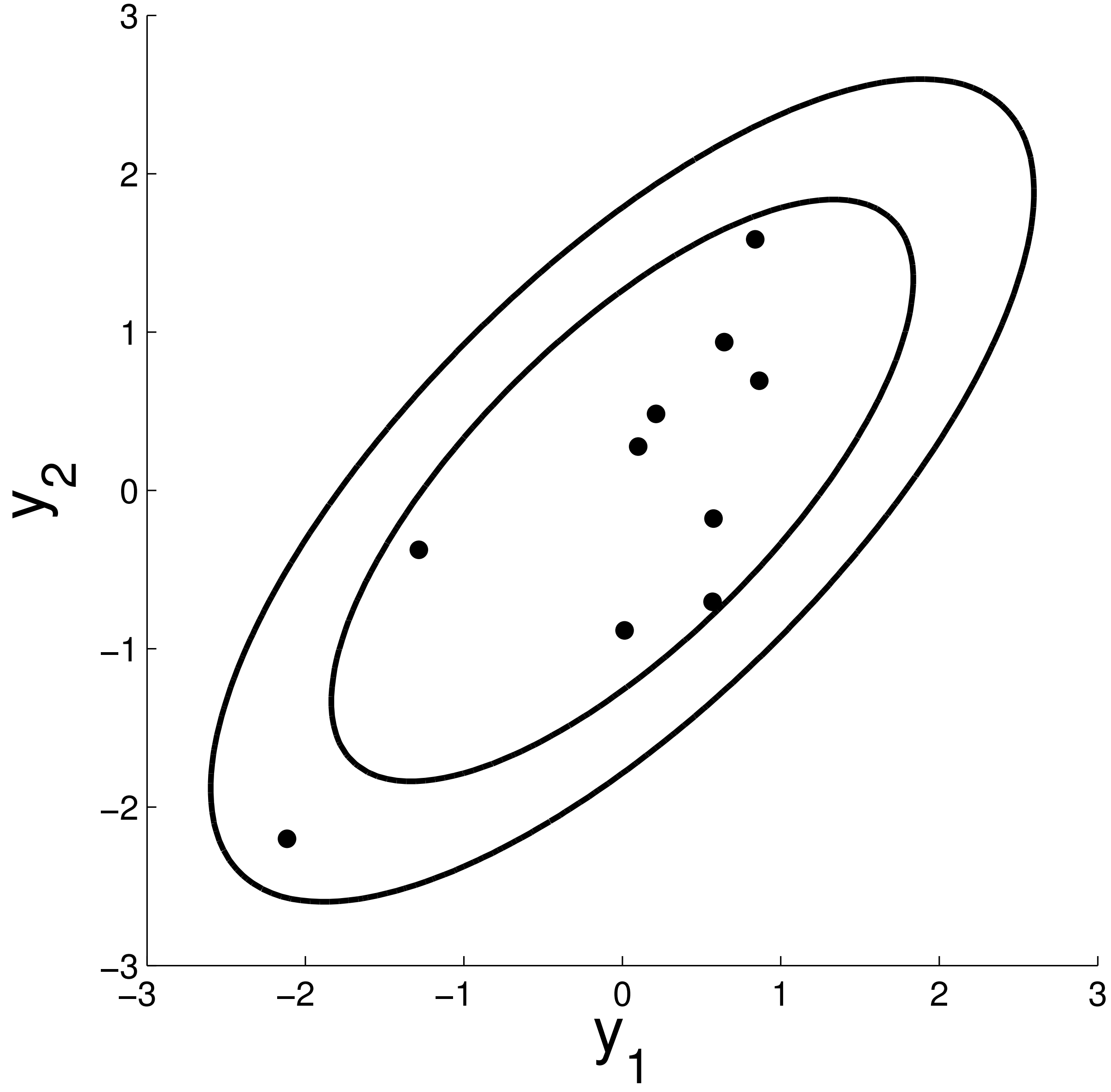
Gaussian distribution

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



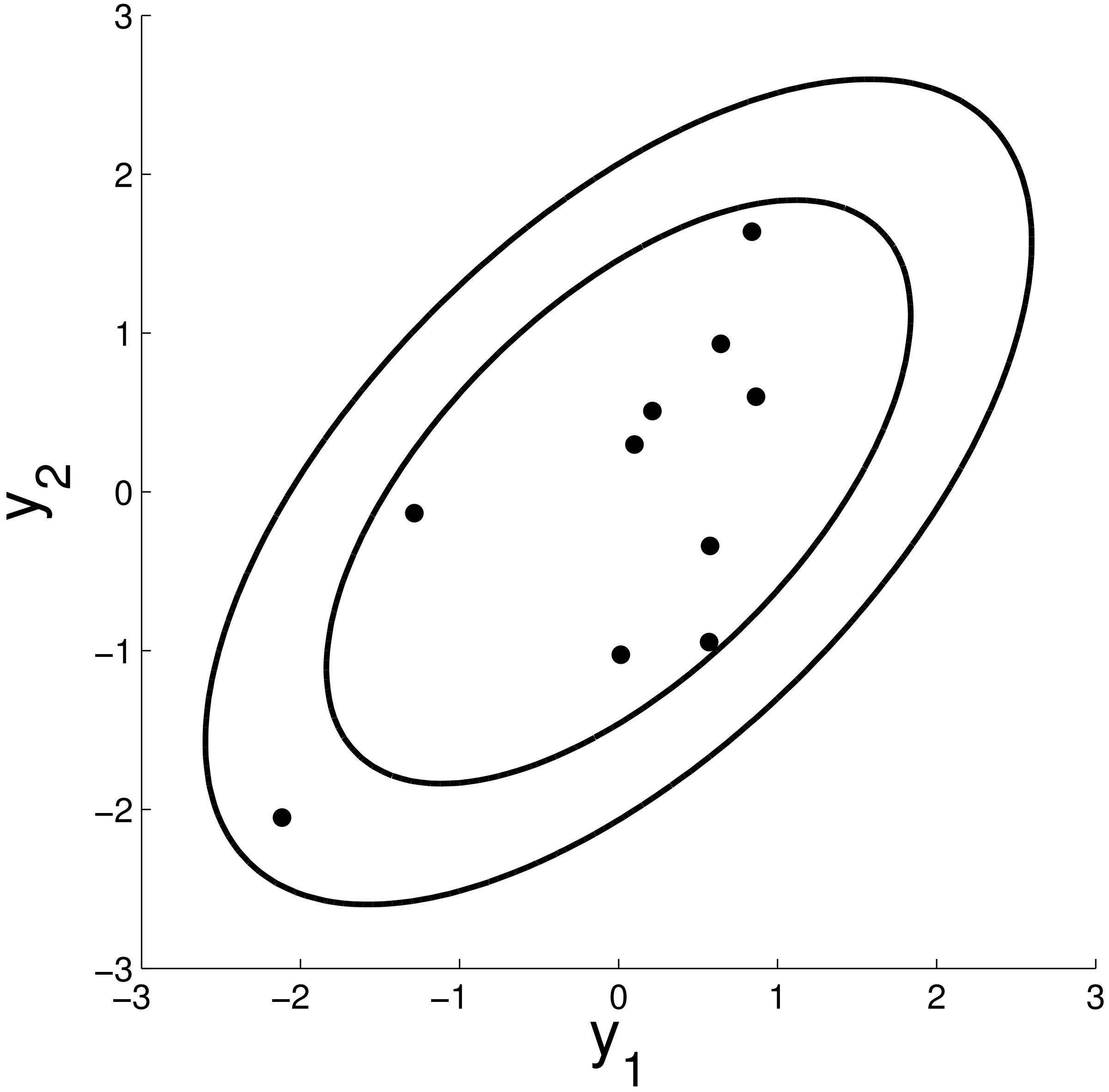
Gaussian distribution

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



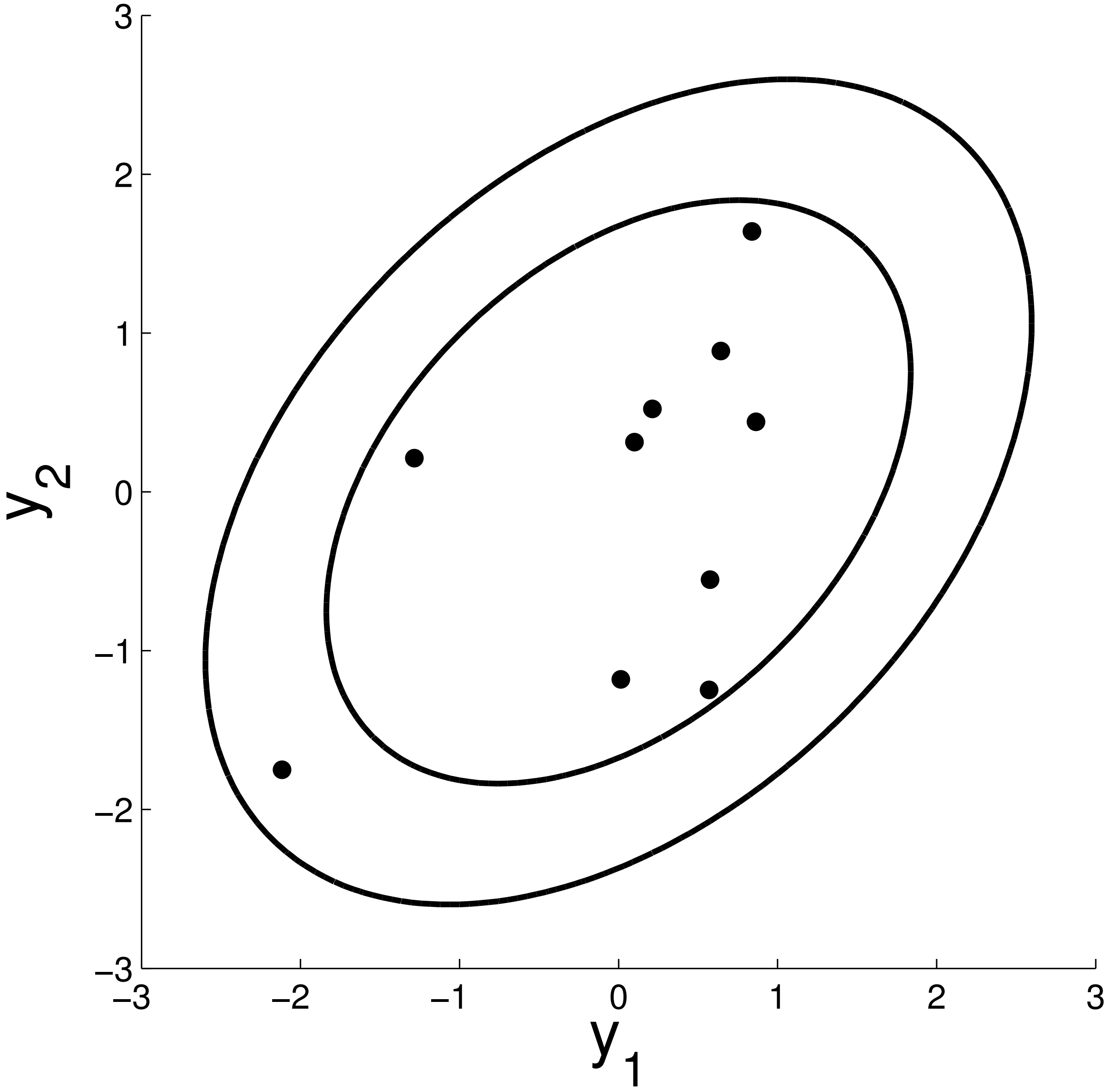
Gaussian distribution

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



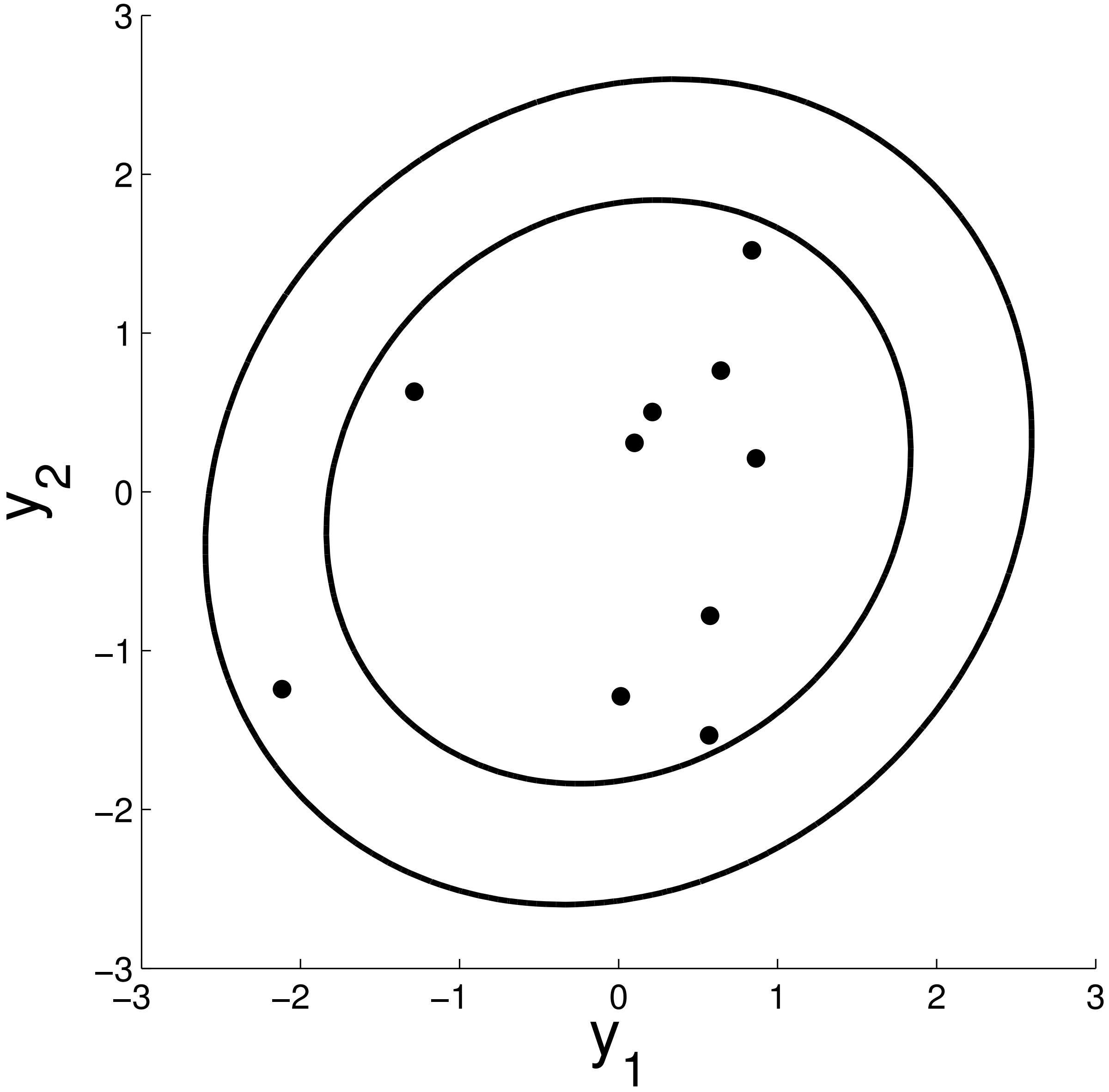
Gaussian distribution

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



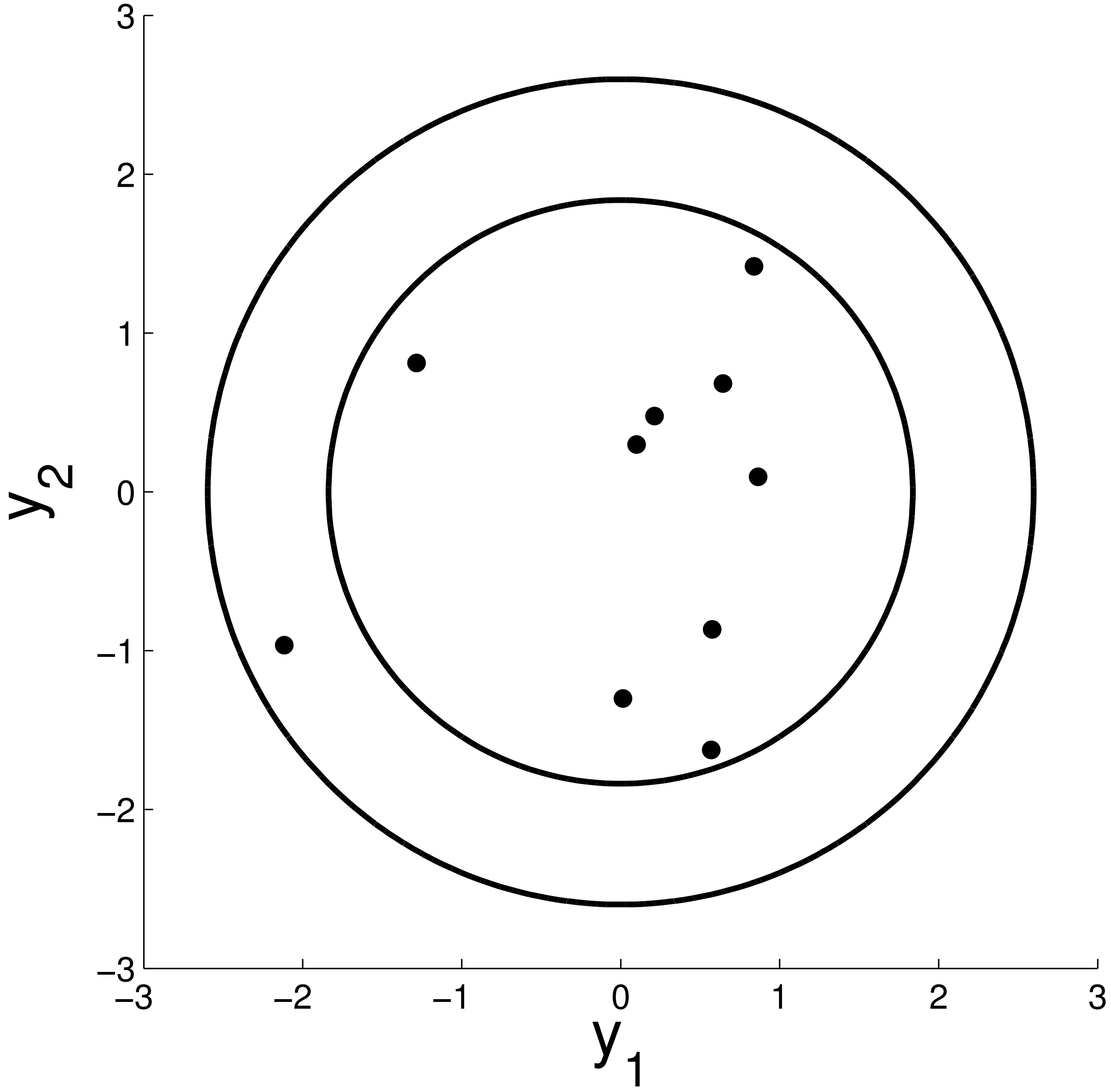
Gaussian distribution

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



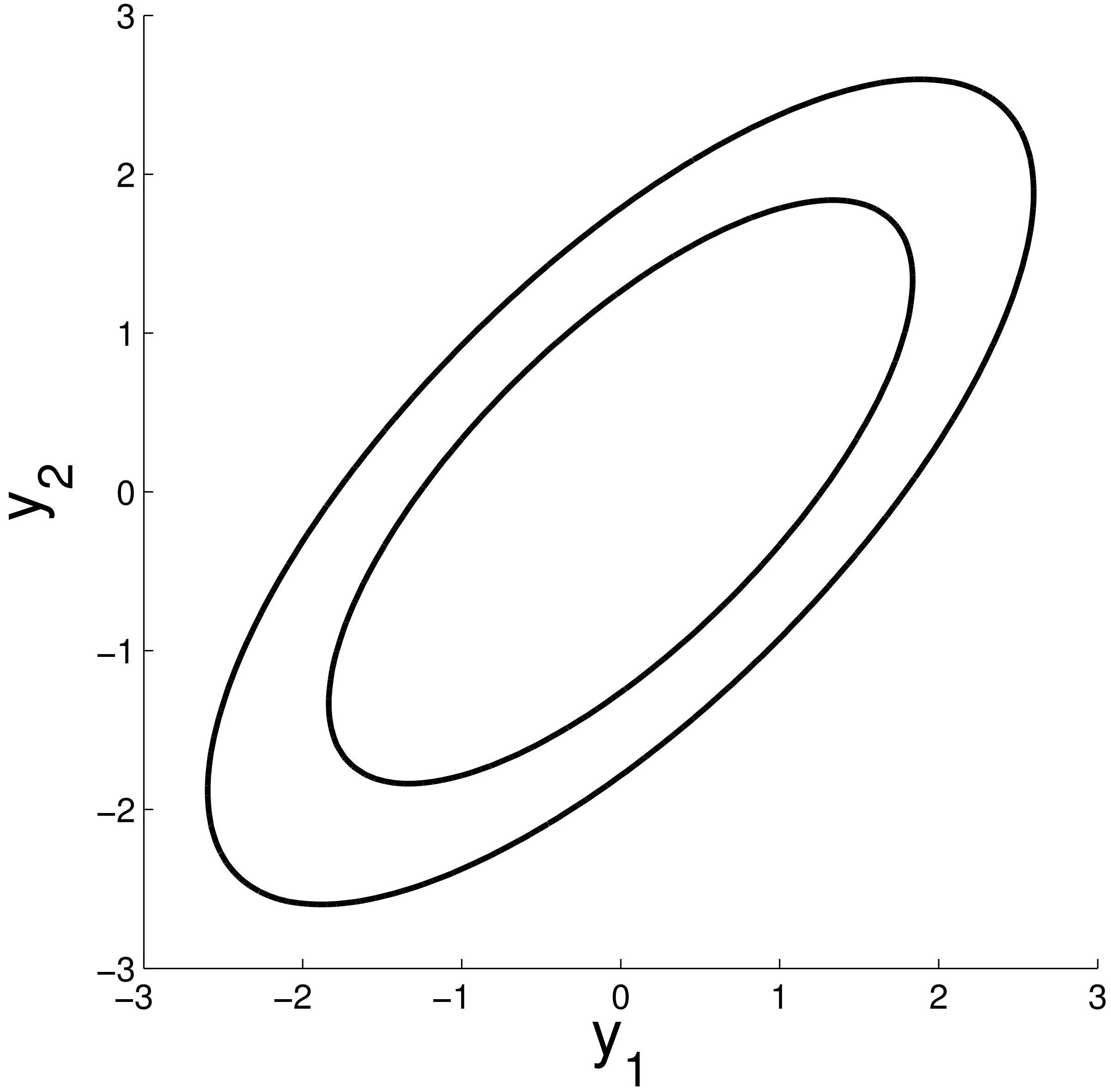
Gaussian distribution

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



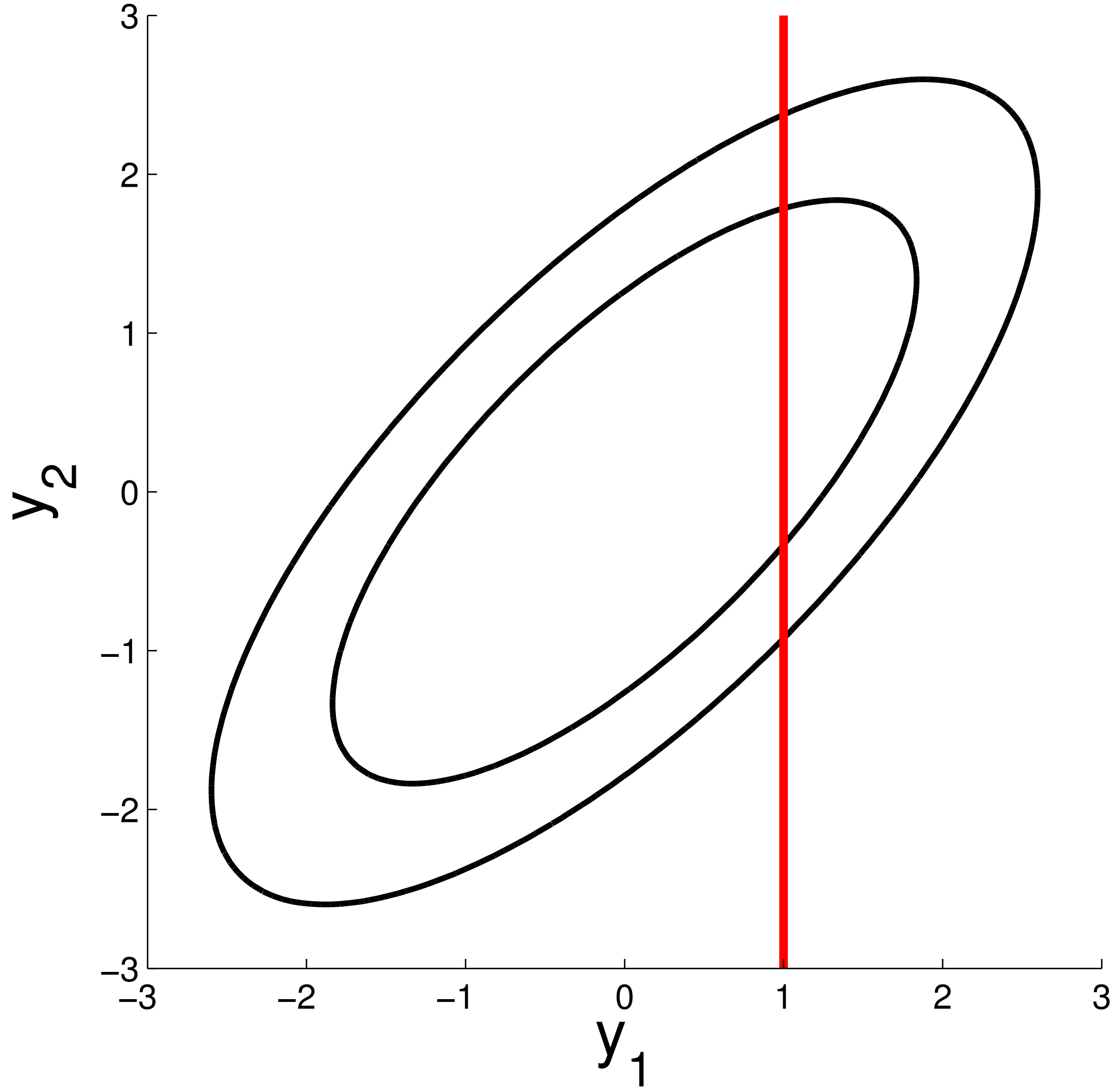
Gaussian distribution

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



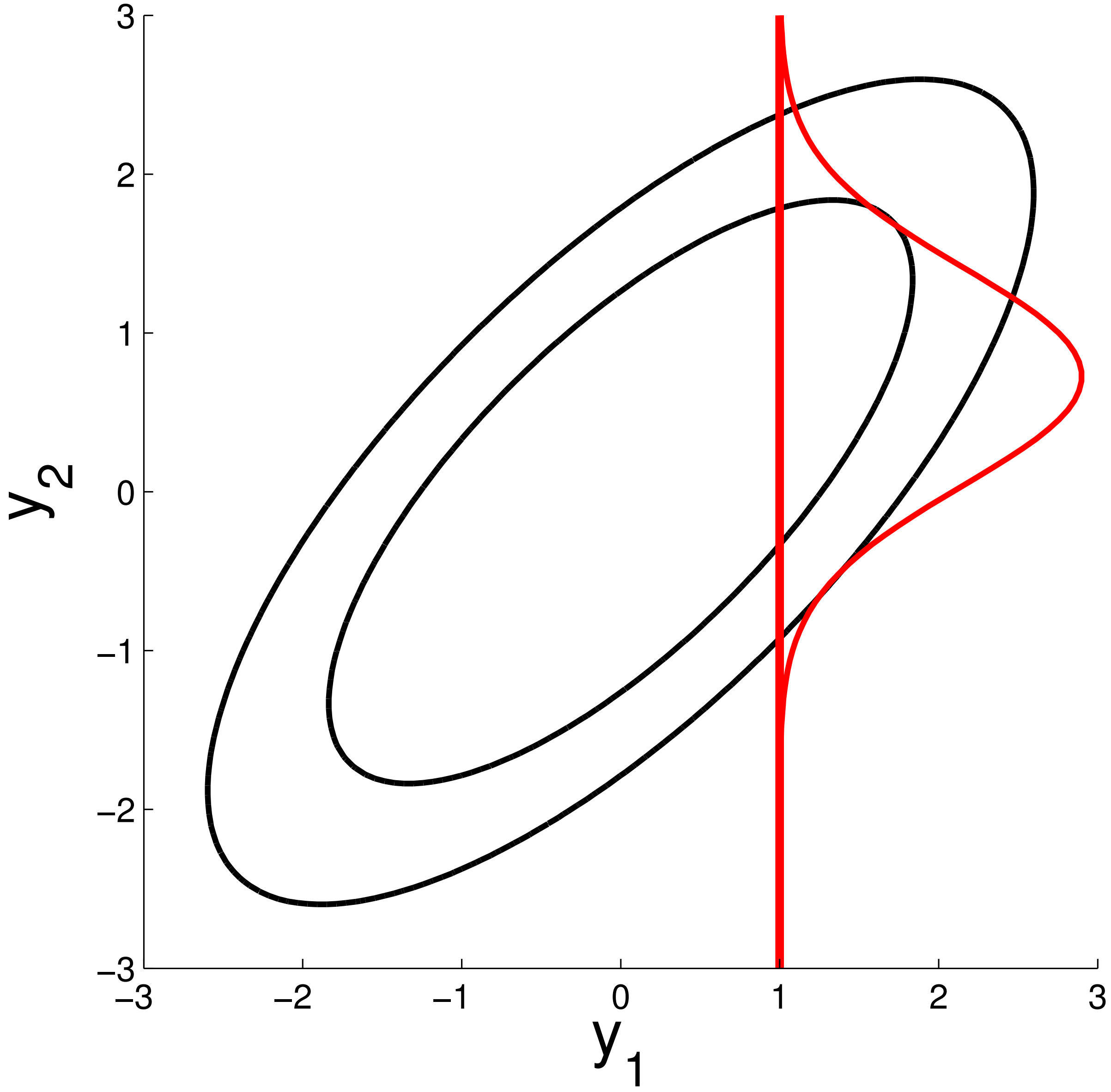
Gaussian distribution - Conditioning

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



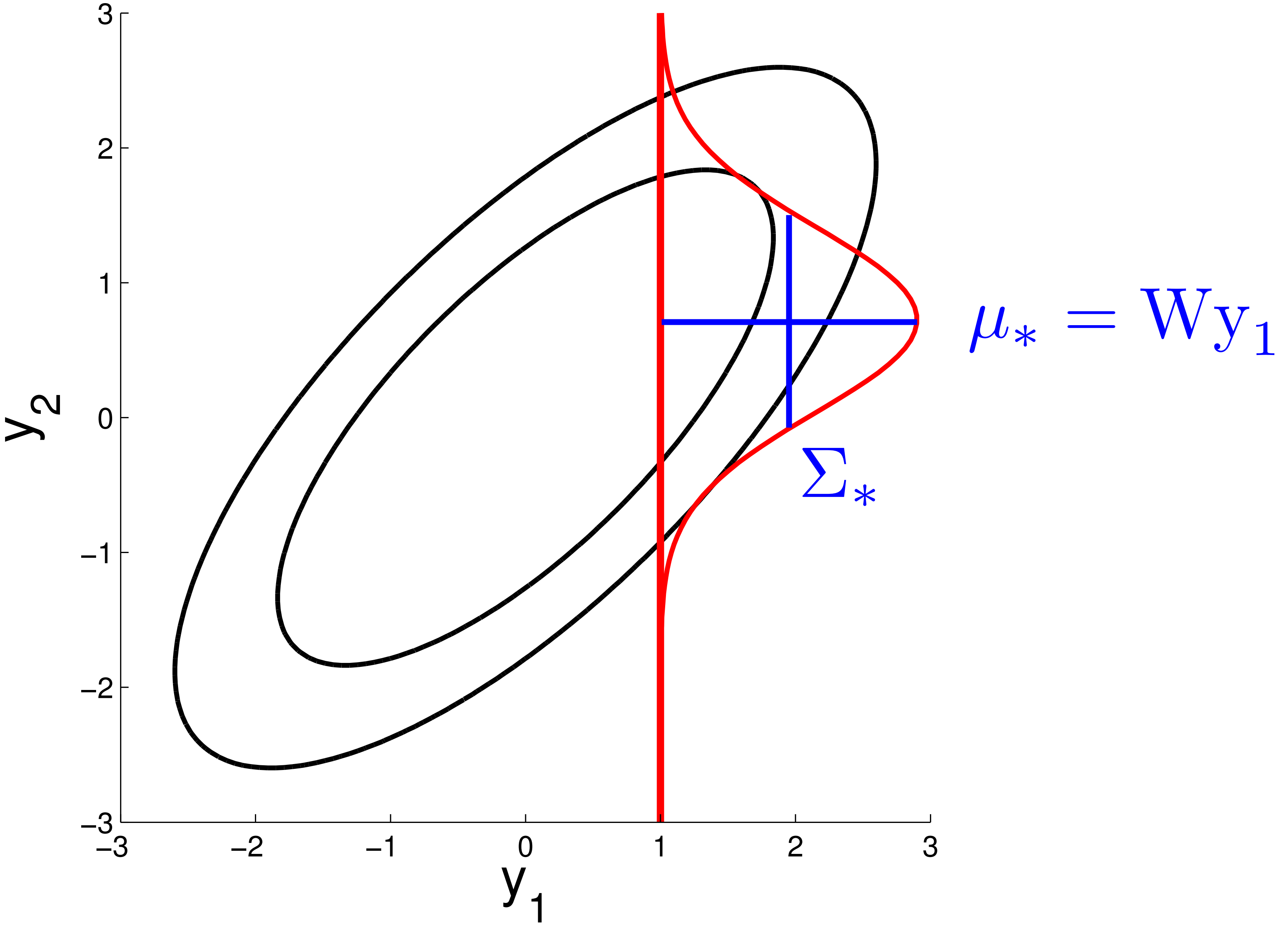
Gaussian distribution - Conditioning

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



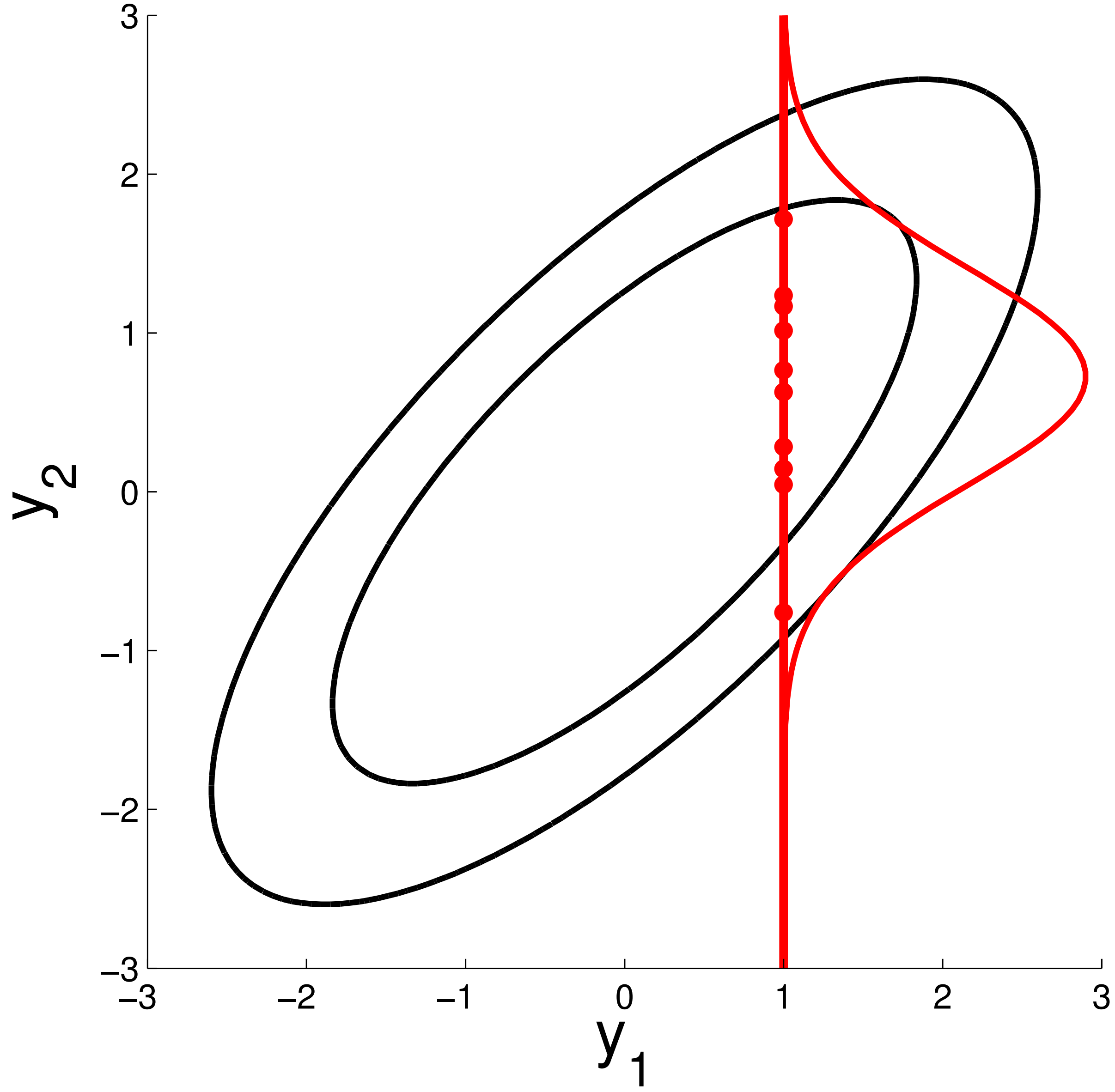
Gaussian distribution - Conditioning

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



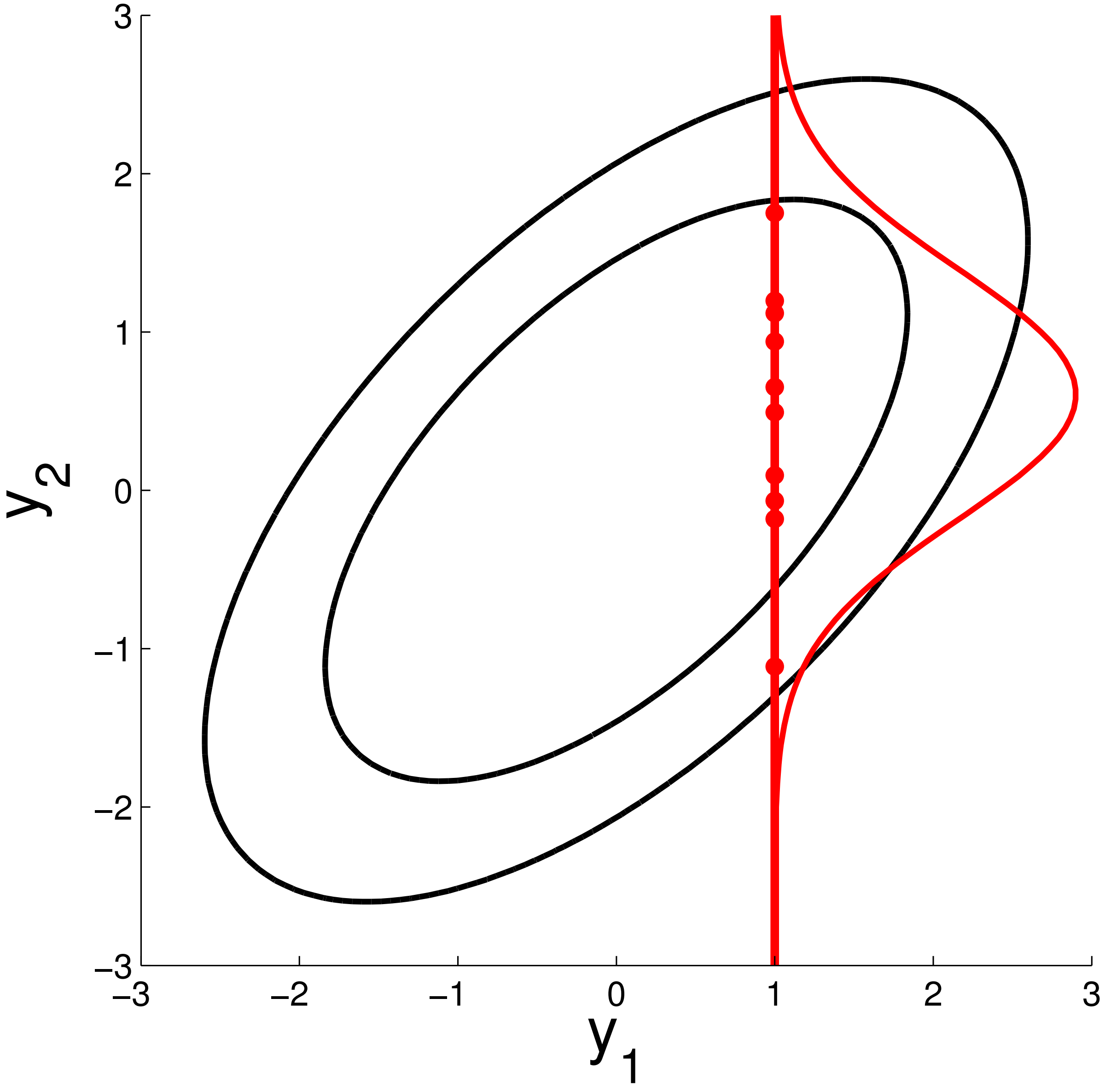
Gaussian distribution - Conditioning

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



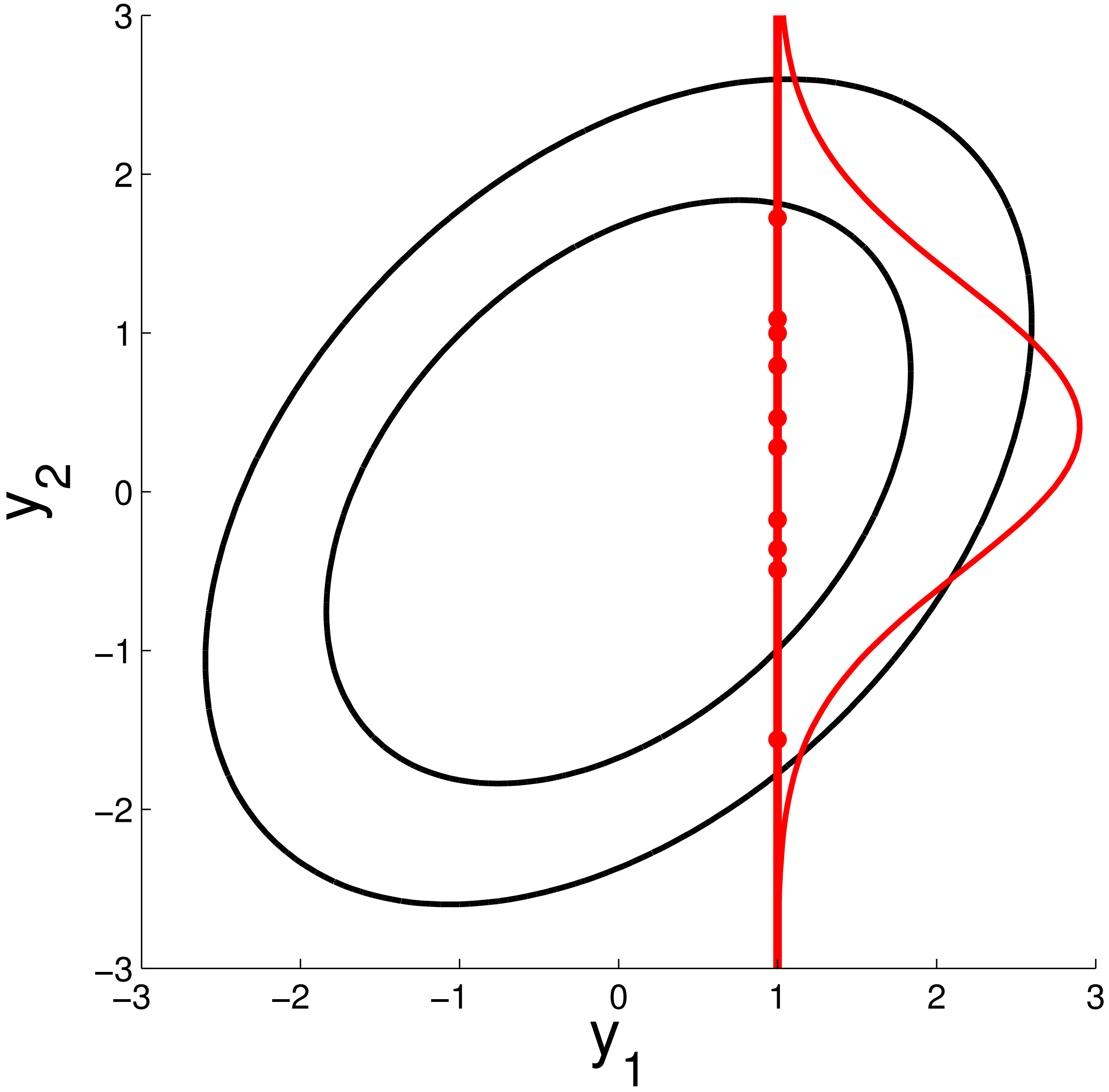
Gaussian distribution - Conditioning

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



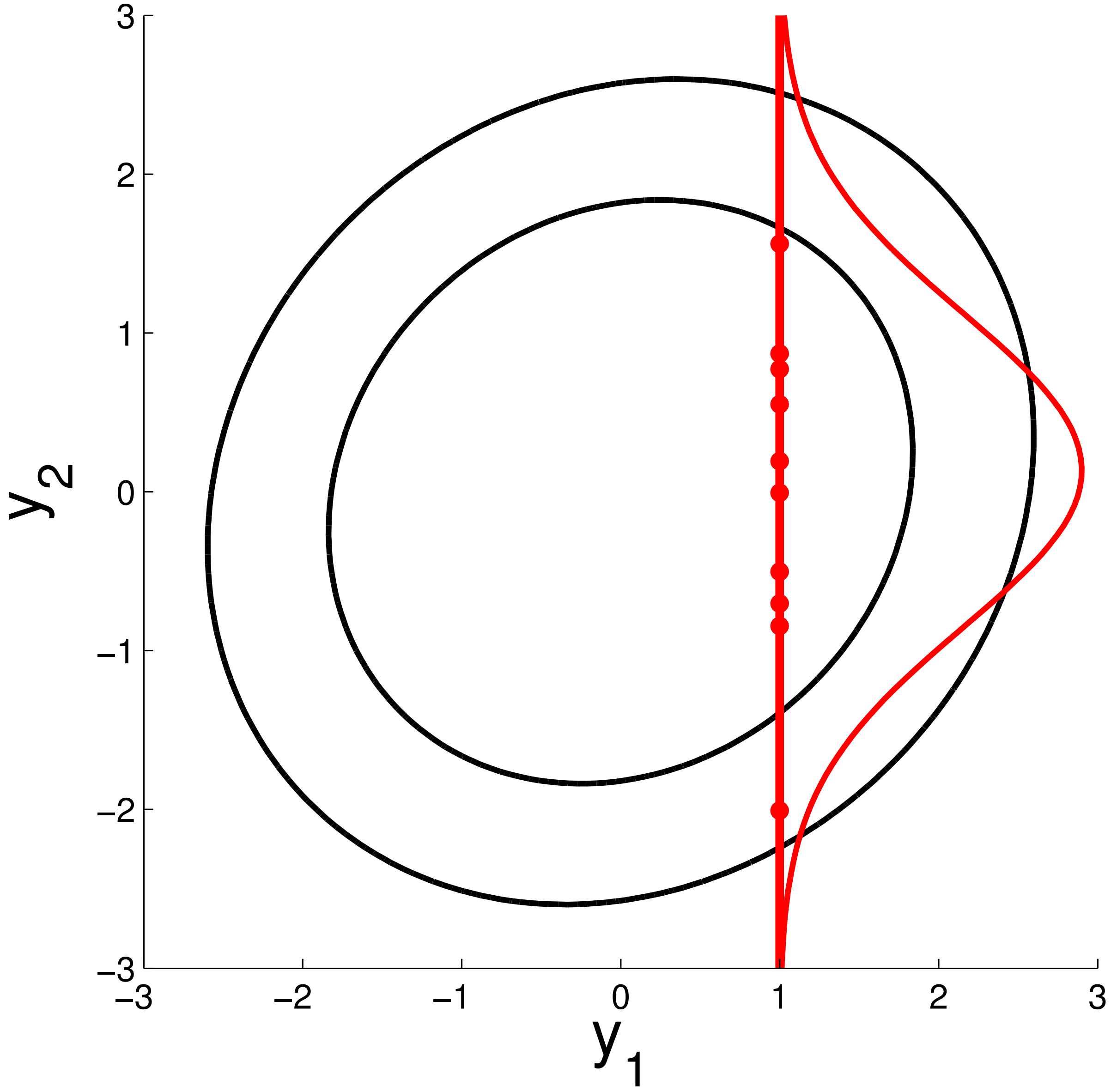
Gaussian distribution - Conditioning

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



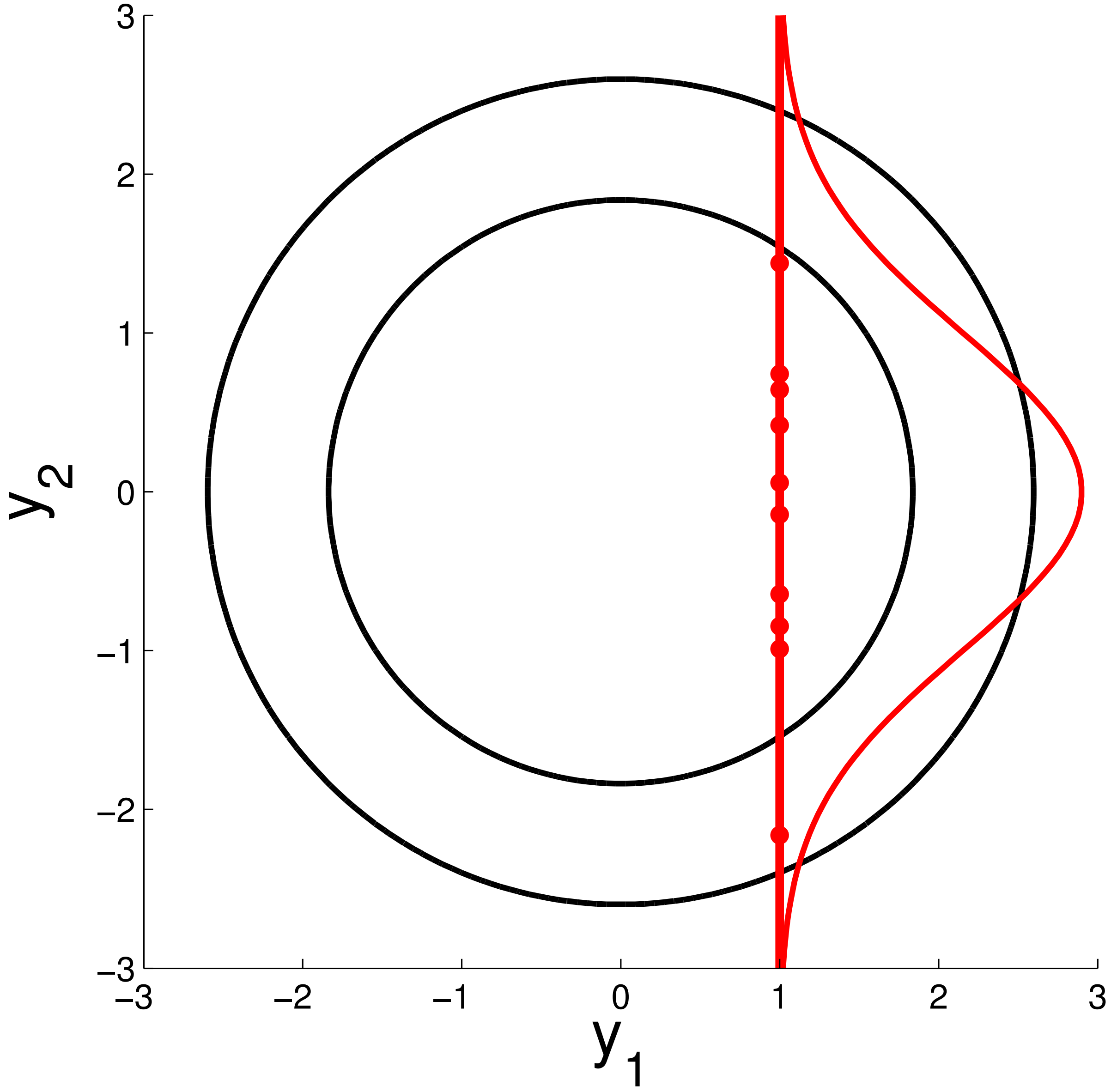
Gaussian distribution - Conditioning

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

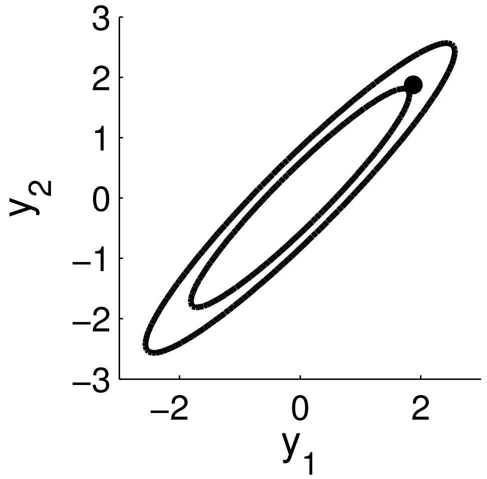


Gaussian distribution - Conditioning

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

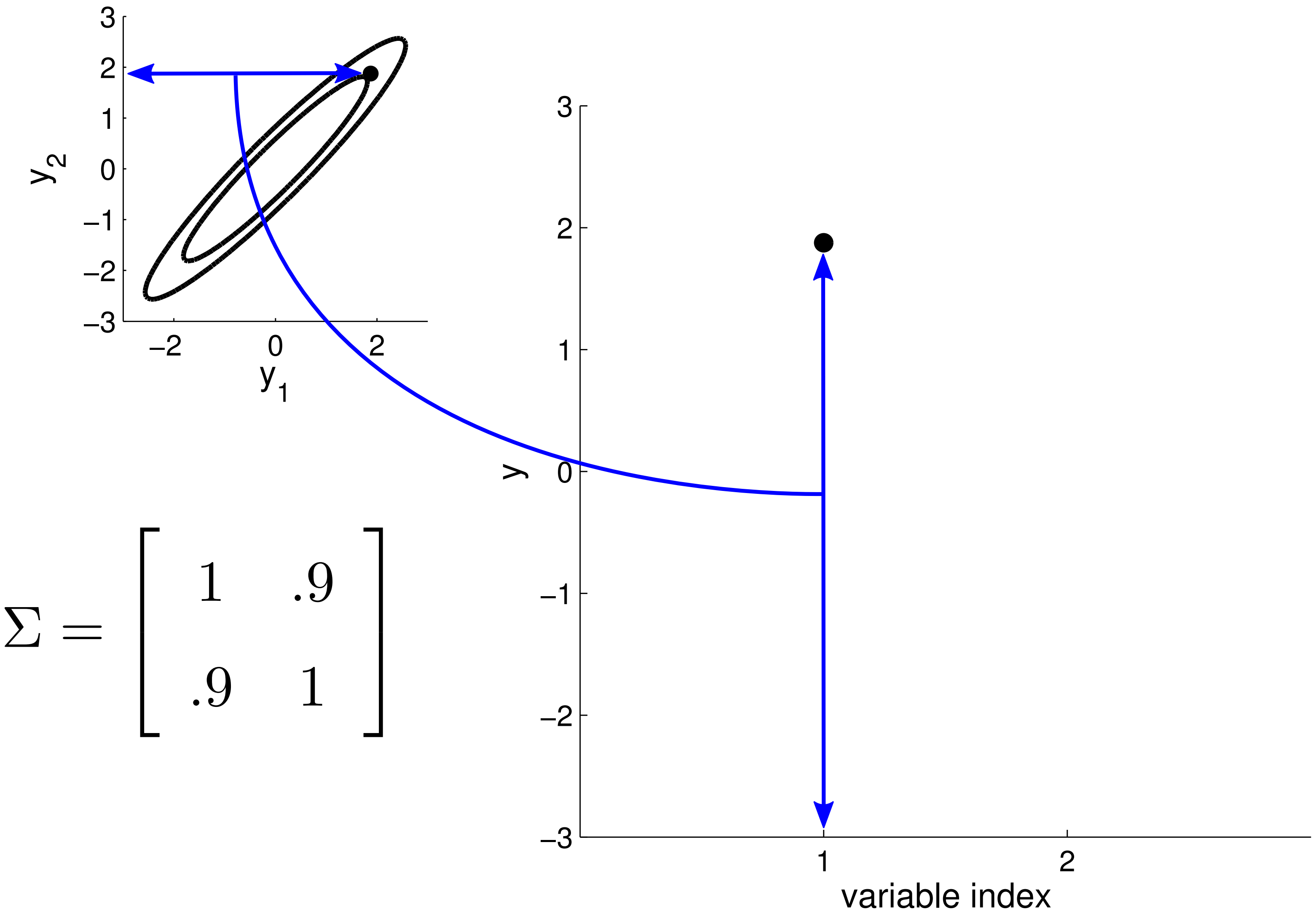


New visualisation



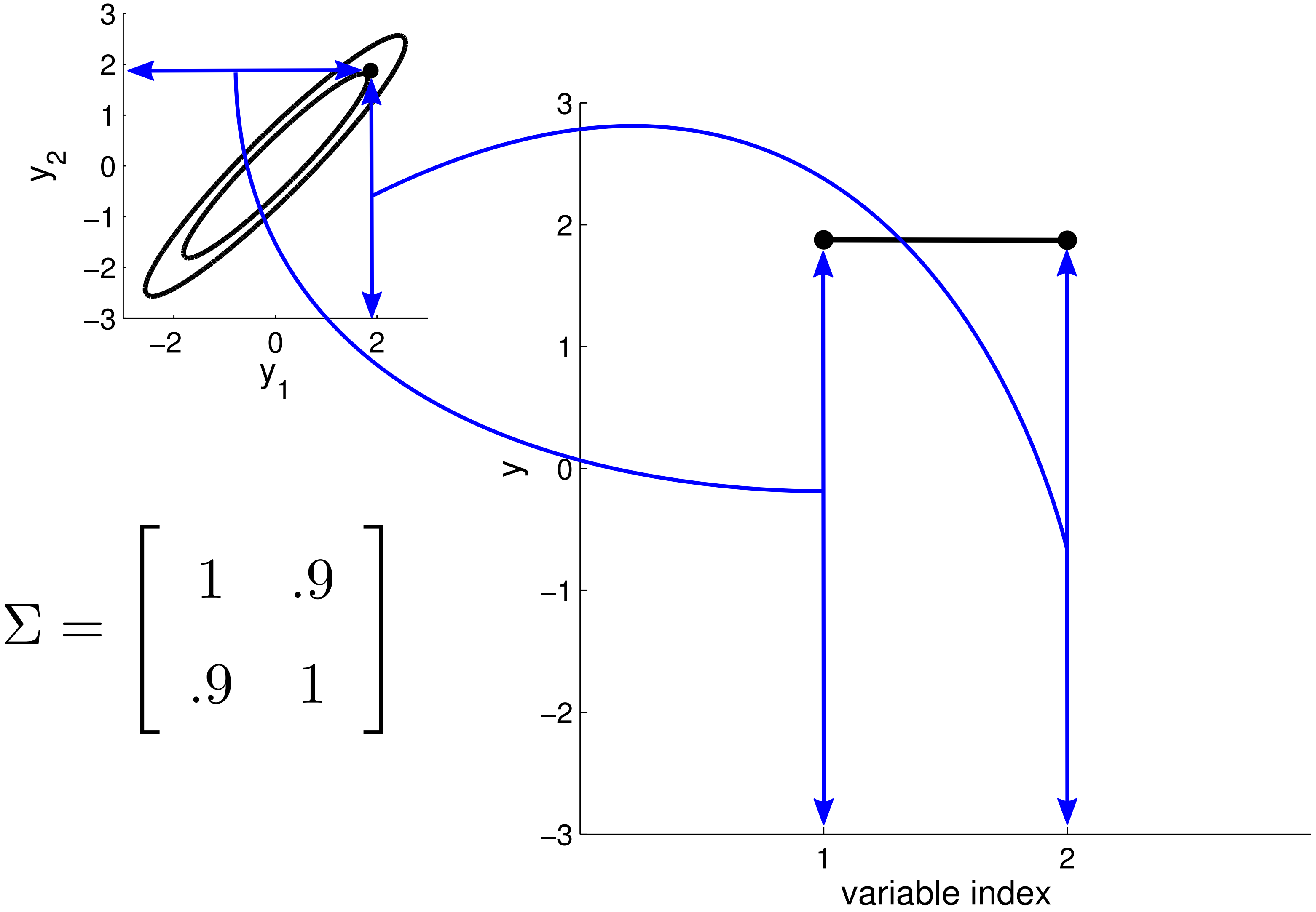
$$\Sigma = \left[\begin{array}{cc} 1 & .9 \\ .9 & 1 \end{array} \right]$$

New visualisation



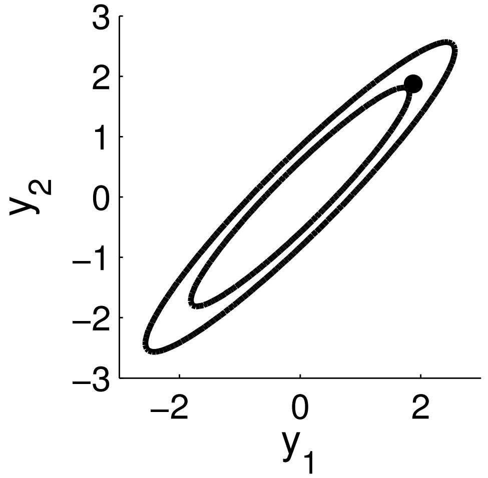
$$\Sigma = \begin{bmatrix} 1 & .9 \\ .9 & 1 \end{bmatrix}$$

New visualisation

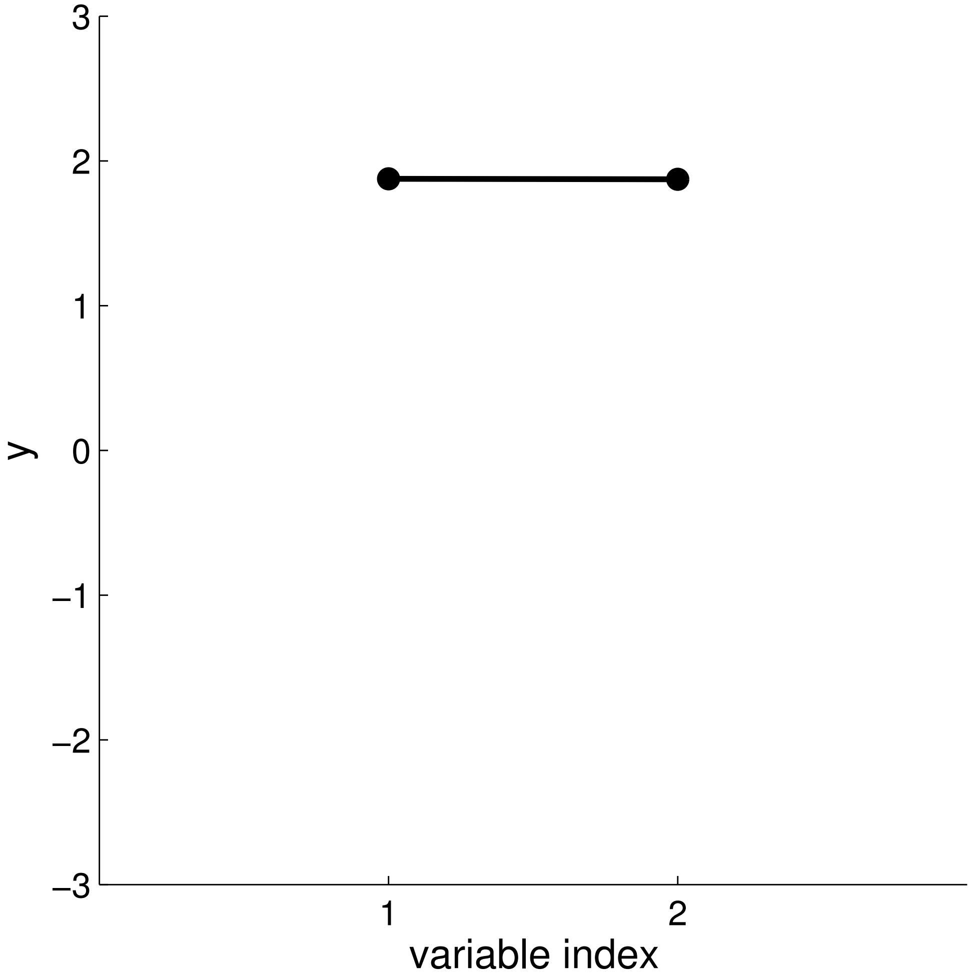


$$\Sigma = \begin{bmatrix} 1 & .9 \\ .9 & 1 \end{bmatrix}$$

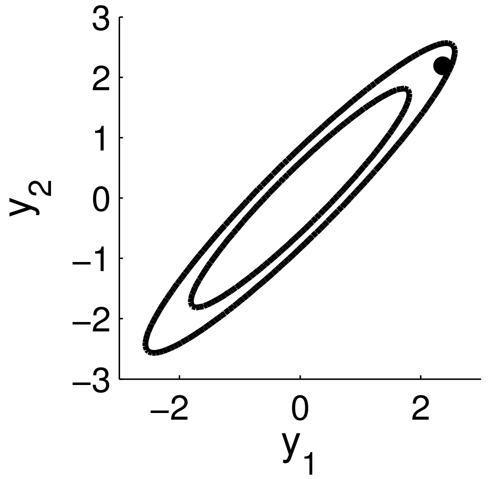
New visualisation



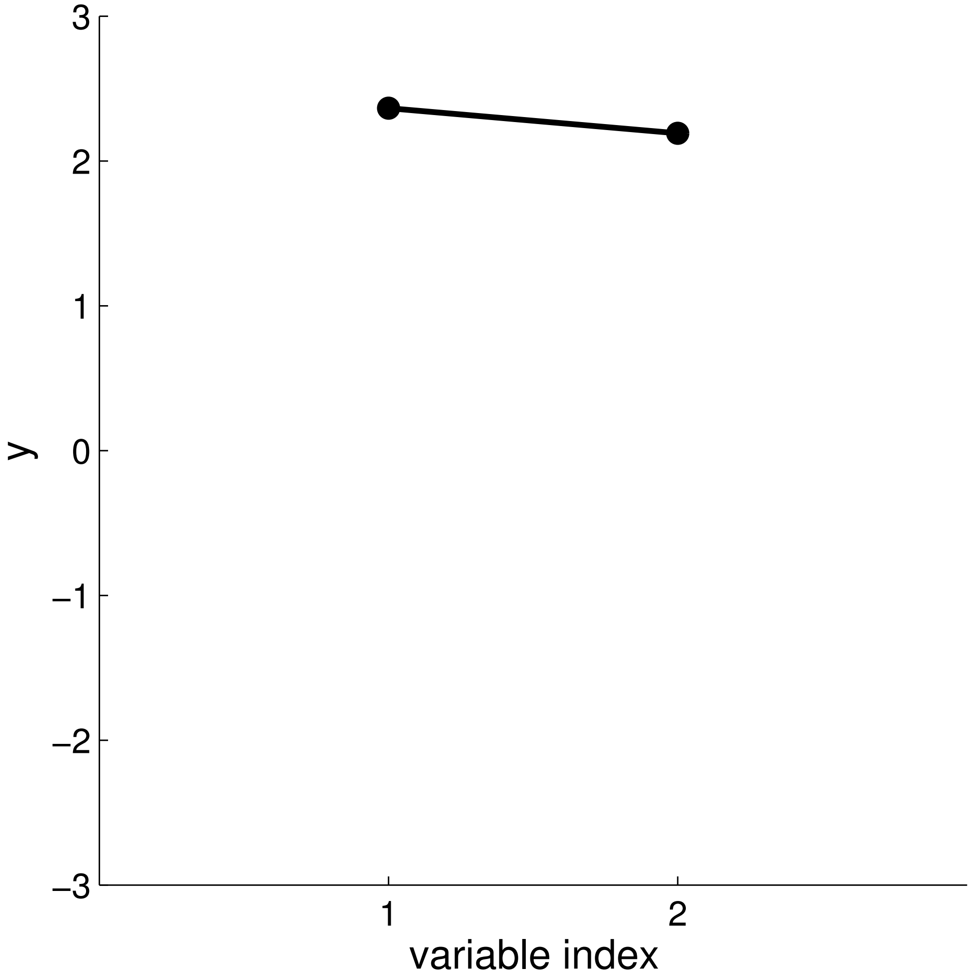
$$\Sigma = \begin{bmatrix} 1 & .9 \\ .9 & 1 \end{bmatrix}$$



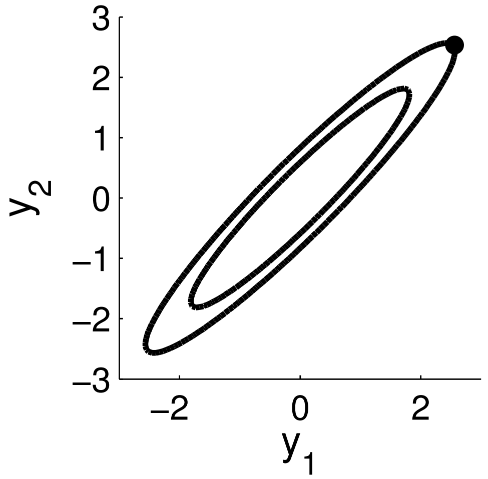
New visualisation



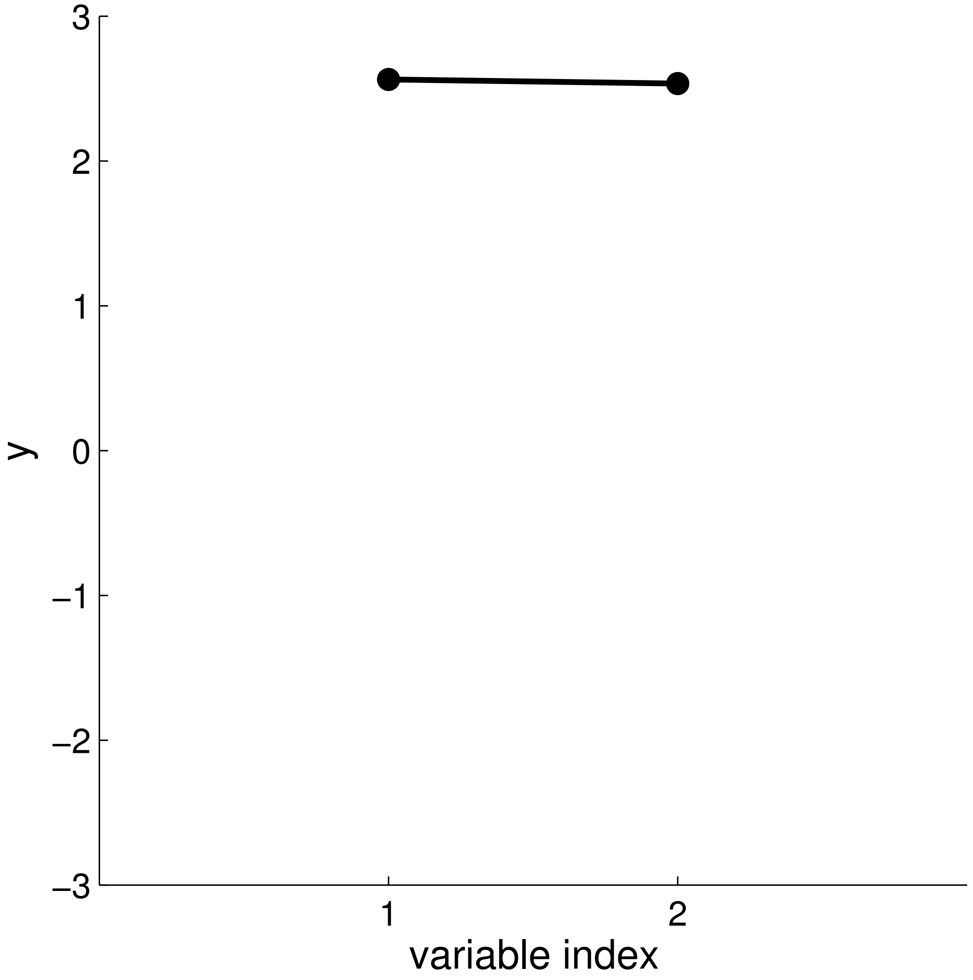
$$\Sigma = \begin{bmatrix} 1 & .9 \\ .9 & 1 \end{bmatrix}$$



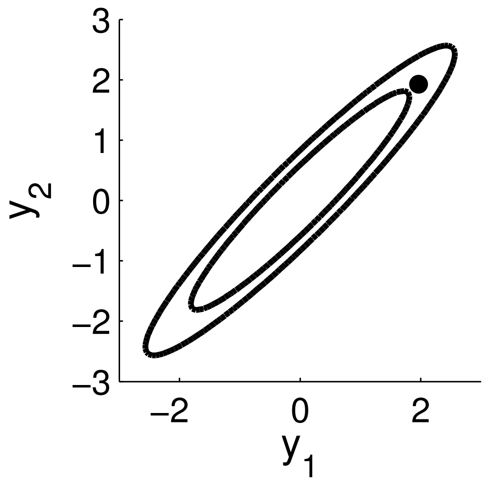
New visualisation



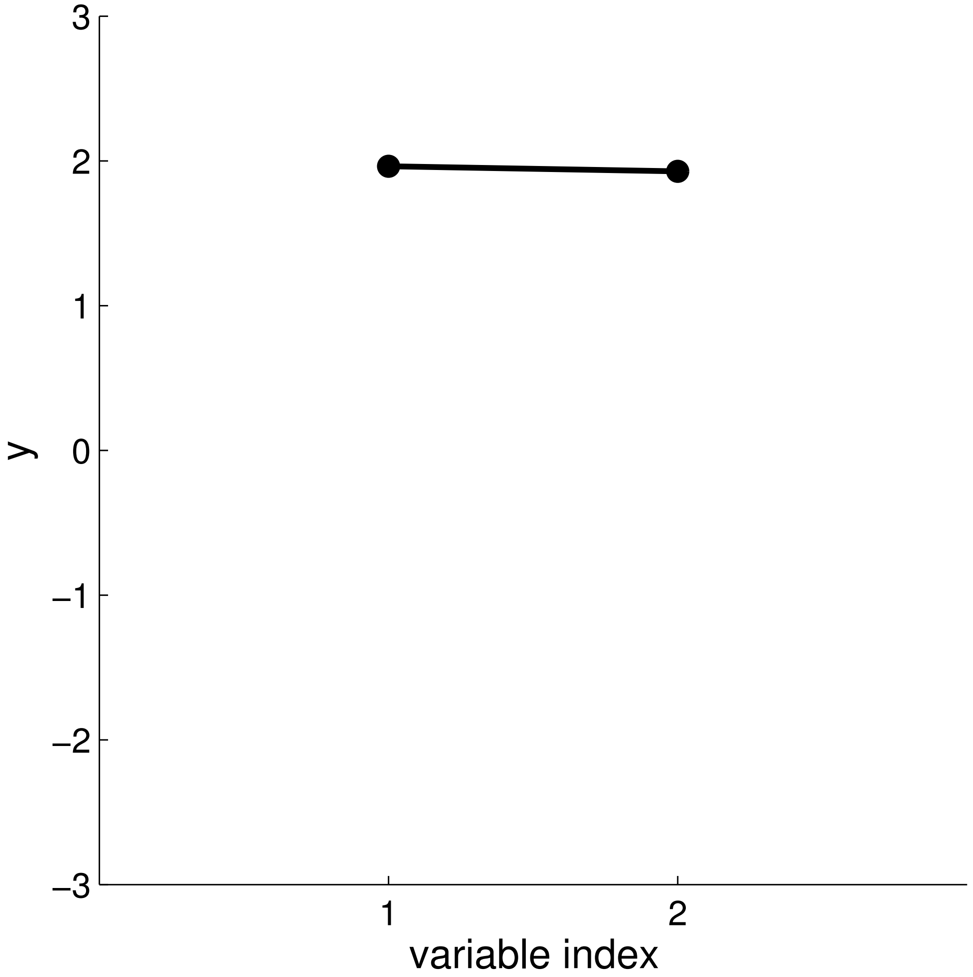
$$\Sigma = \begin{bmatrix} 1 & .9 \\ .9 & 1 \end{bmatrix}$$



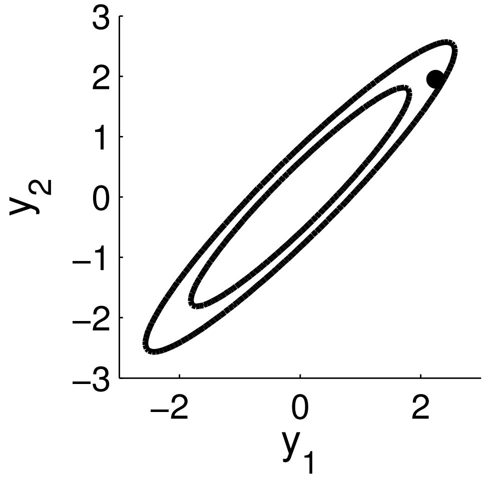
New visualisation



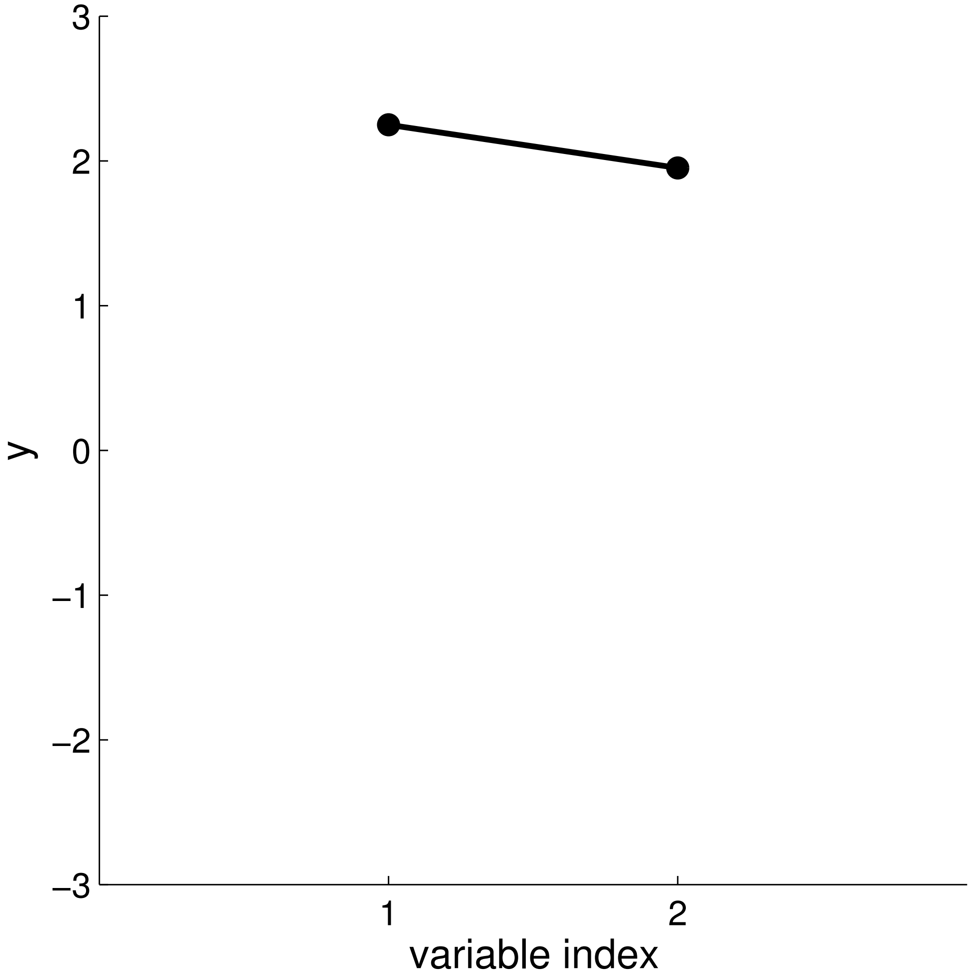
$$\Sigma = \begin{bmatrix} 1 & .9 \\ .9 & 1 \end{bmatrix}$$



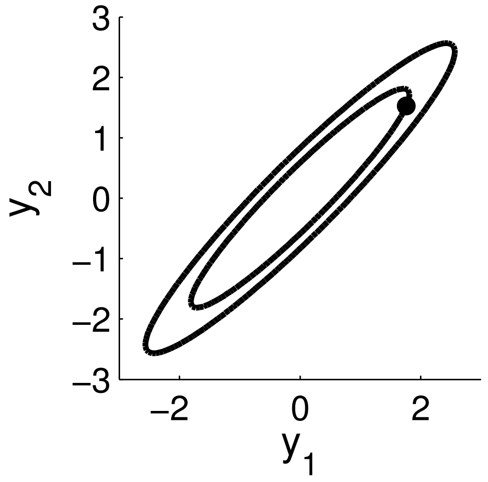
New visualisation



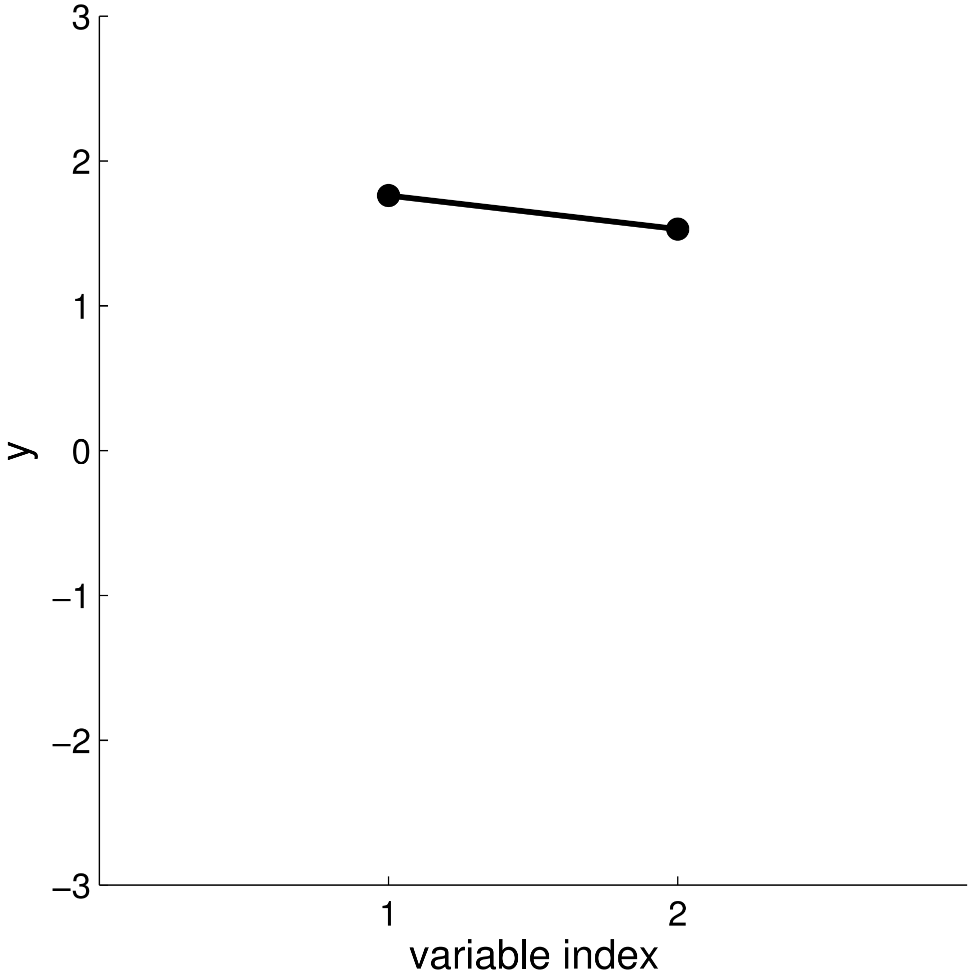
$$\Sigma = \begin{bmatrix} 1 & .9 \\ .9 & 1 \end{bmatrix}$$



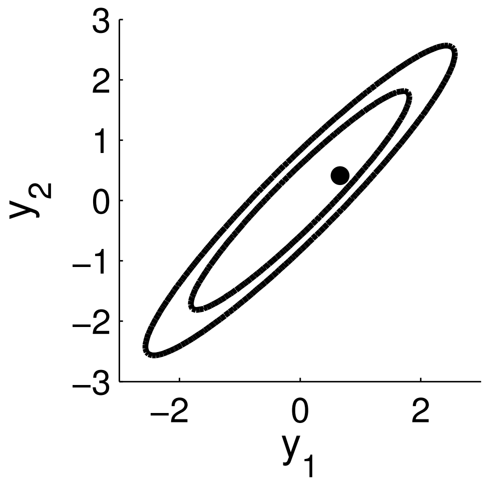
New visualisation



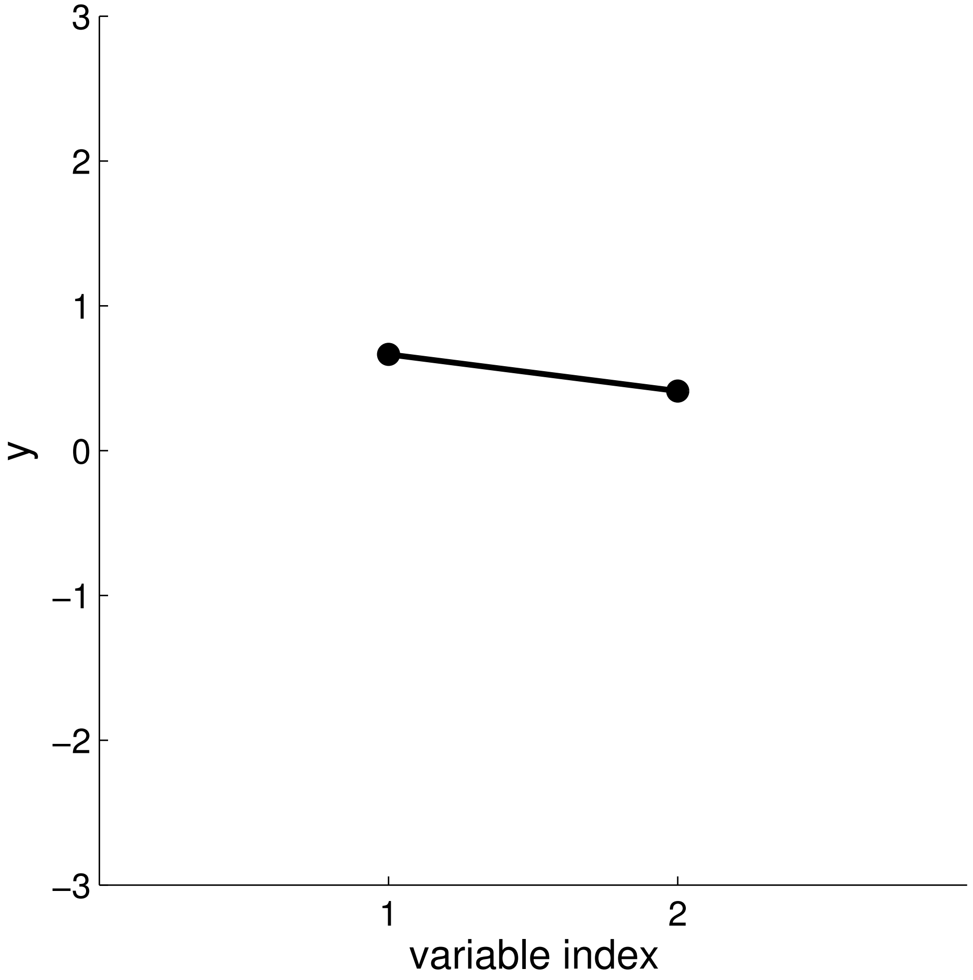
$$\Sigma = \begin{bmatrix} 1 & .9 \\ .9 & 1 \end{bmatrix}$$



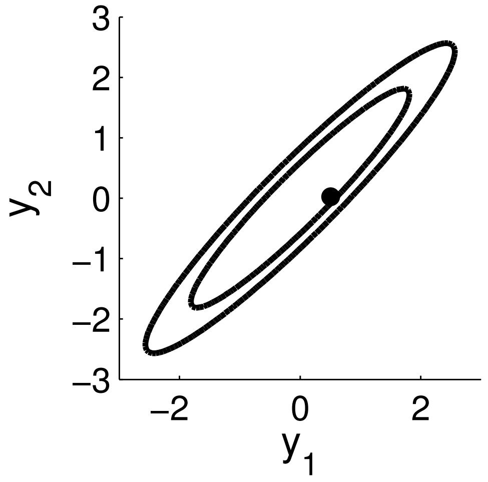
New visualisation



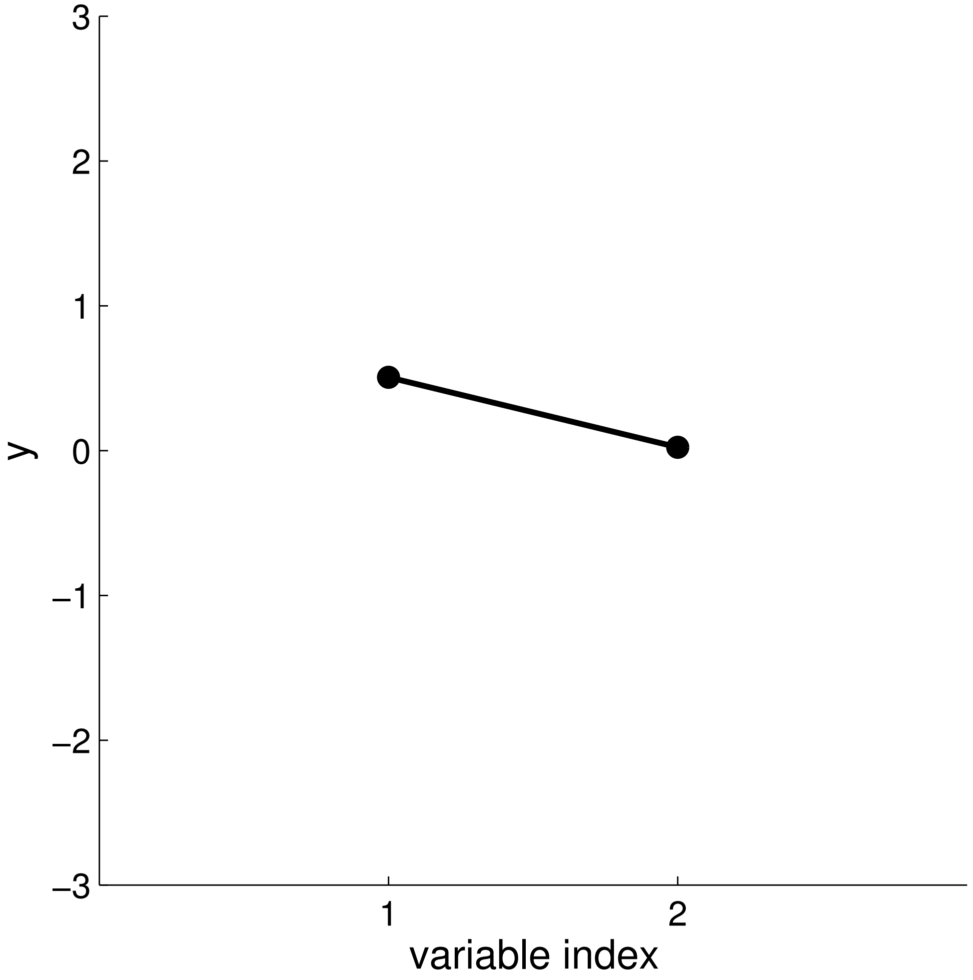
$$\Sigma = \begin{bmatrix} 1 & .9 \\ .9 & 1 \end{bmatrix}$$



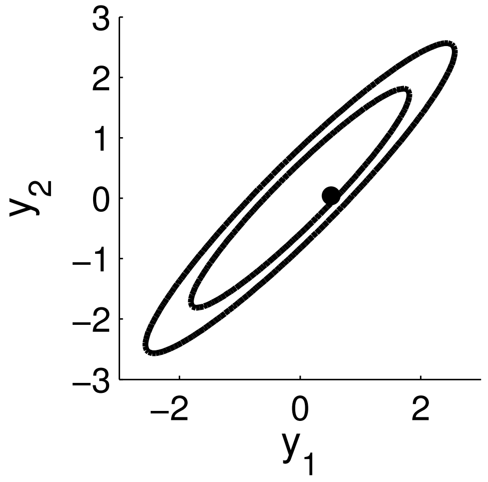
New visualisation



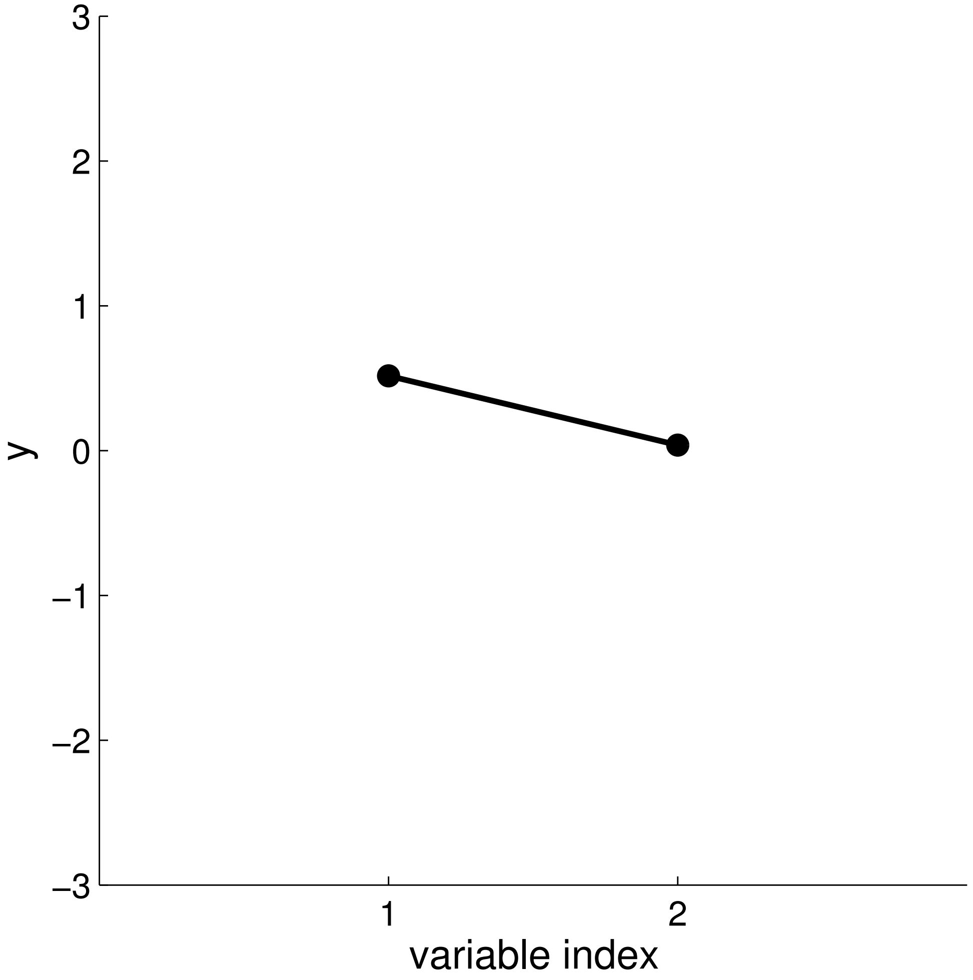
$$\Sigma = \begin{bmatrix} 1 & .9 \\ .9 & 1 \end{bmatrix}$$



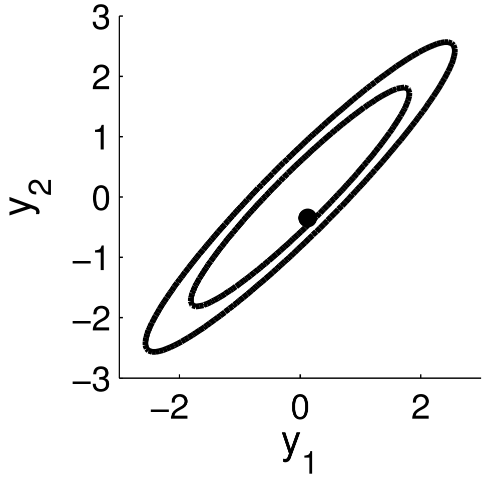
New visualisation

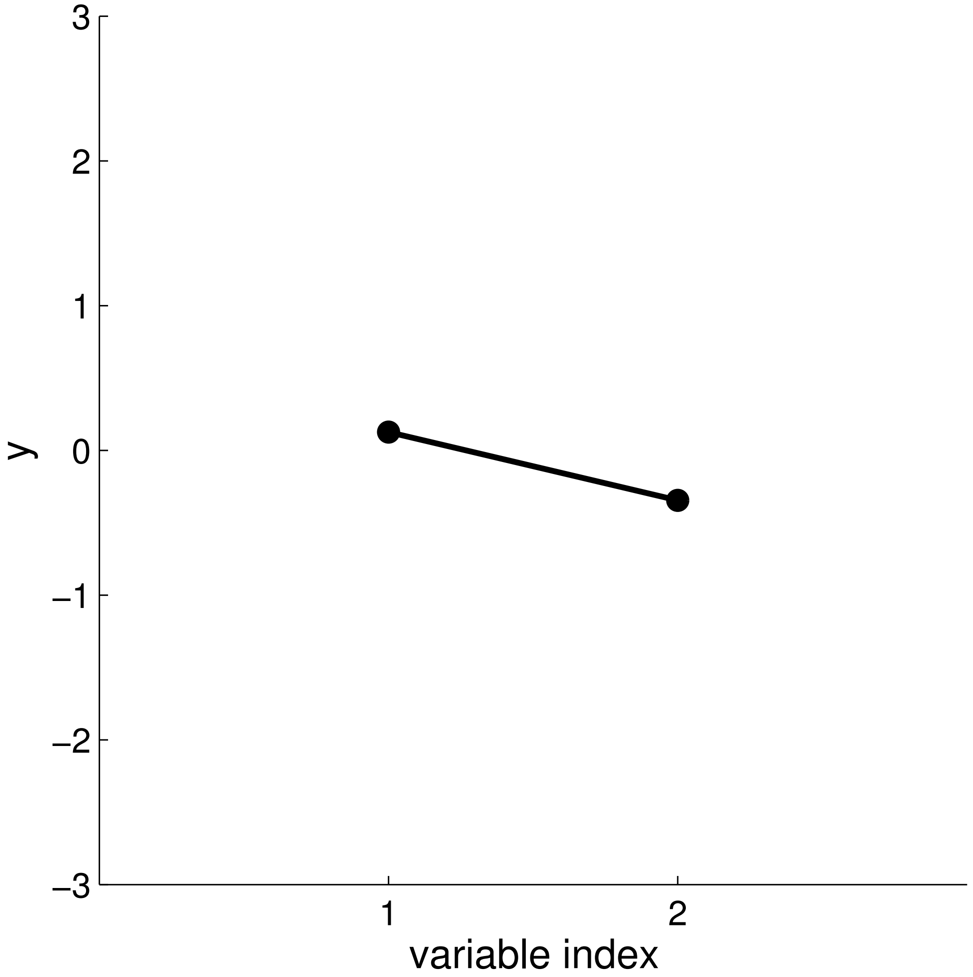


$$\Sigma = \begin{bmatrix} 1 & .9 \\ .9 & 1 \end{bmatrix}$$



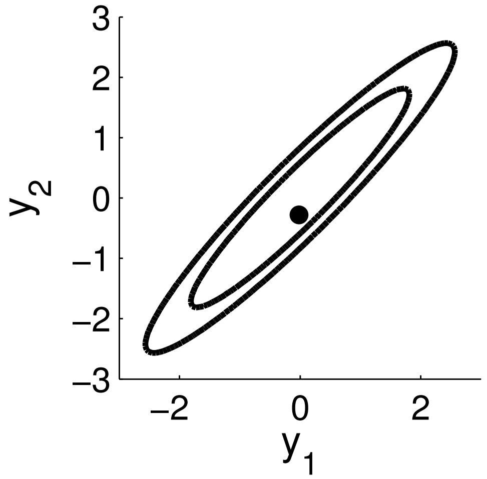
New visualisation



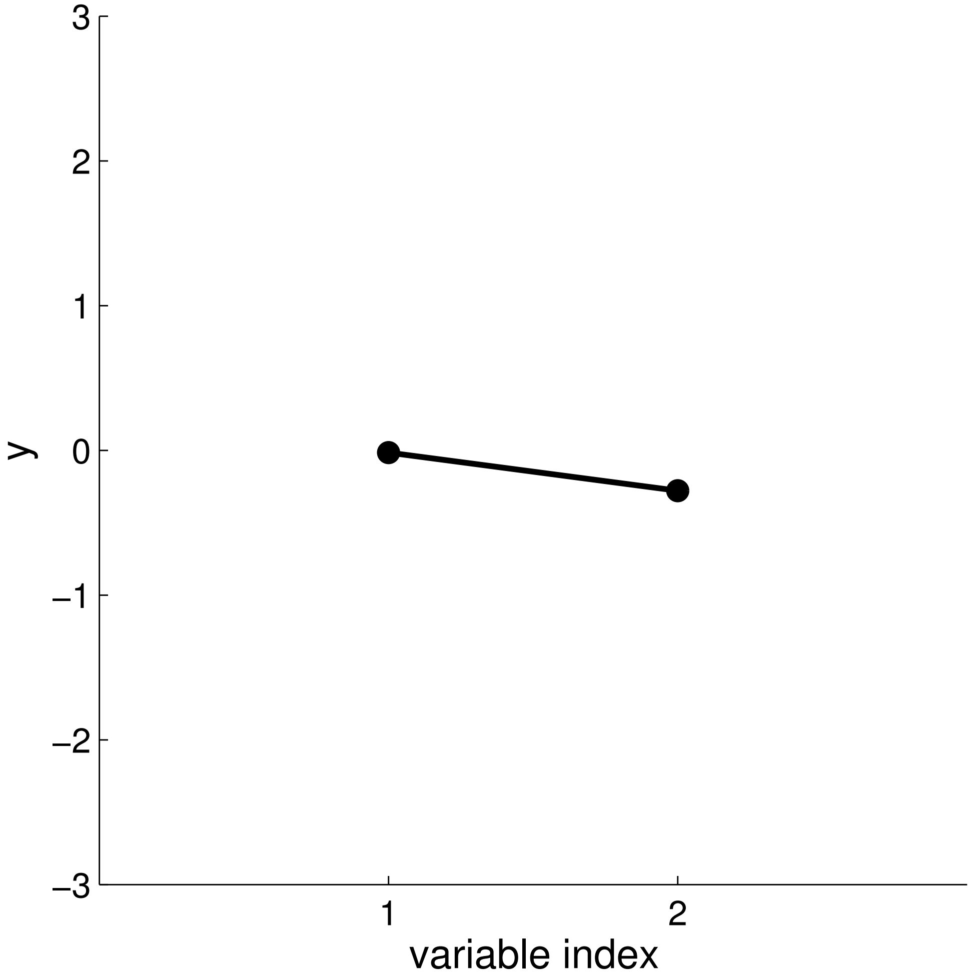


$$\Sigma = \begin{bmatrix} 1 & .9 \\ .9 & 1 \end{bmatrix}$$

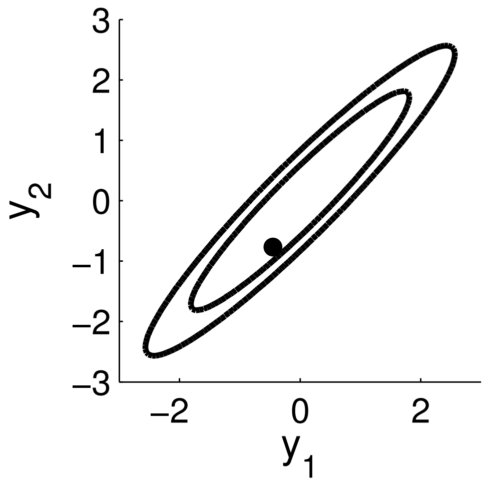
New visualisation



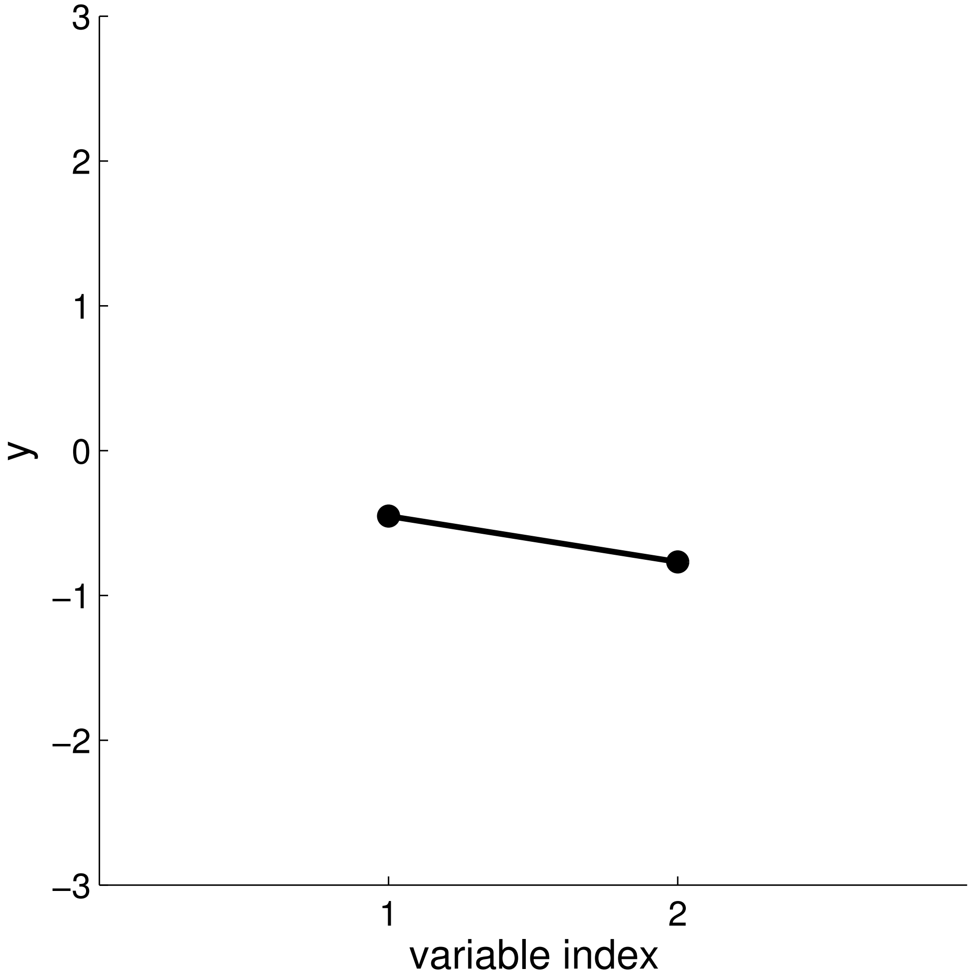
$$\Sigma = \begin{bmatrix} 1 & .9 \\ .9 & 1 \end{bmatrix}$$



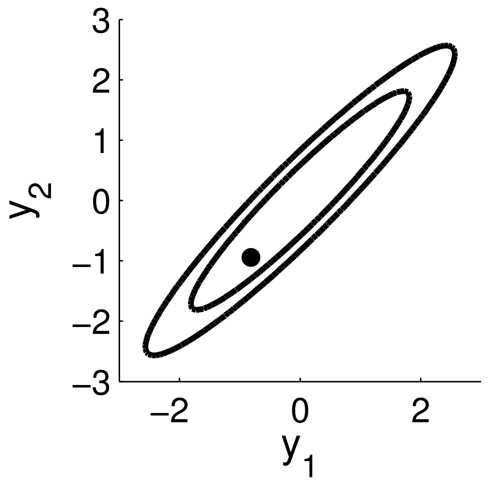
New visualisation



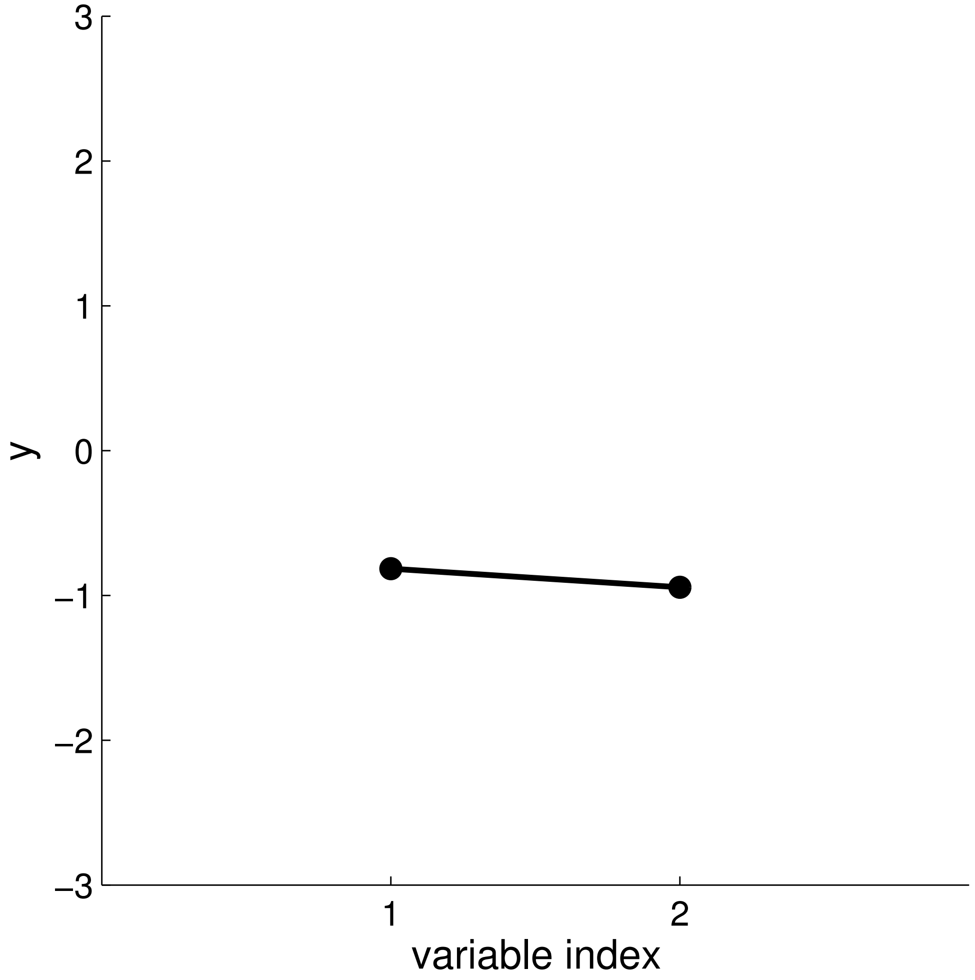
$$\Sigma = \begin{bmatrix} 1 & .9 \\ .9 & 1 \end{bmatrix}$$



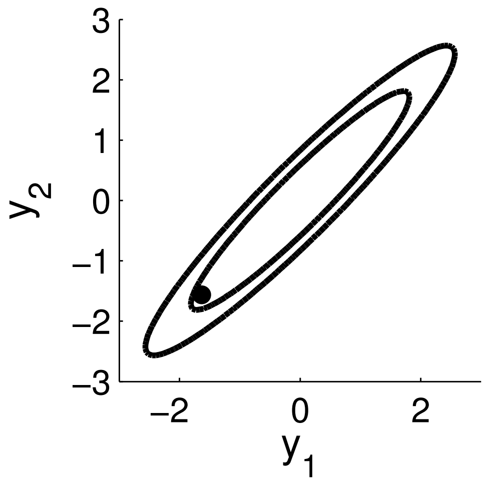
New visualisation



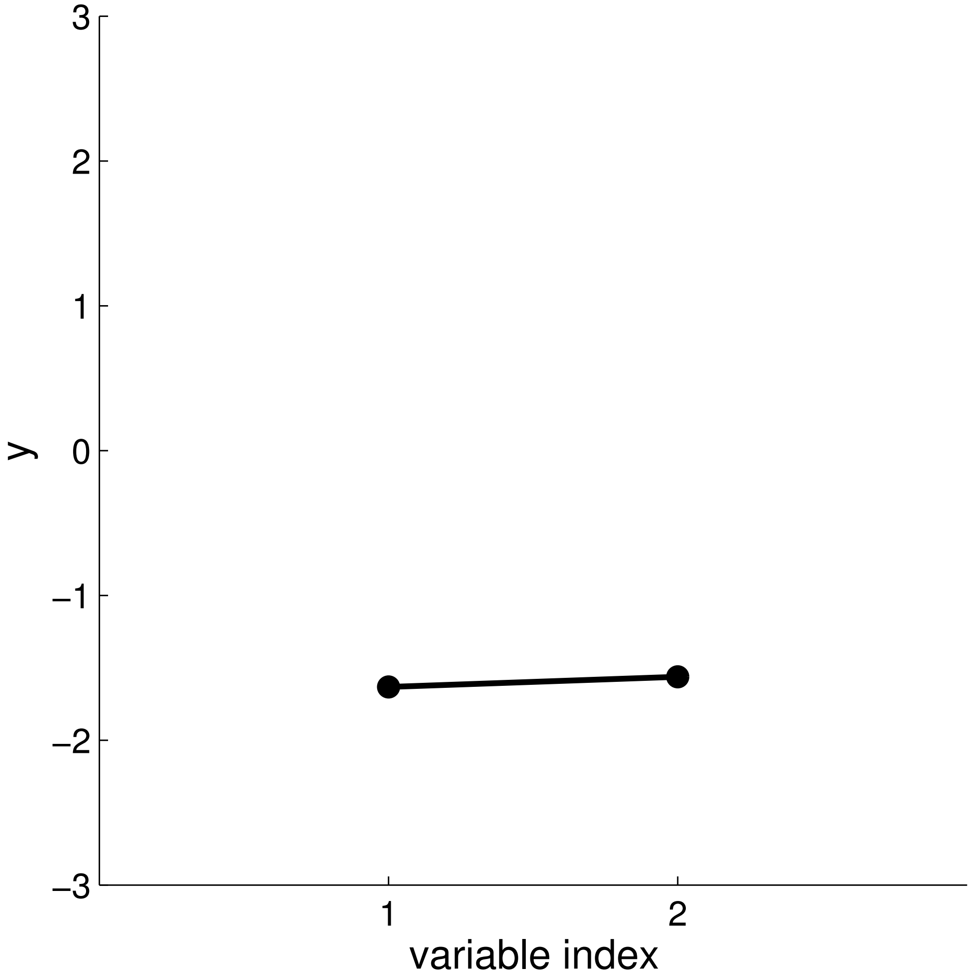
$$\Sigma = \begin{bmatrix} 1 & .9 \\ .9 & 1 \end{bmatrix}$$



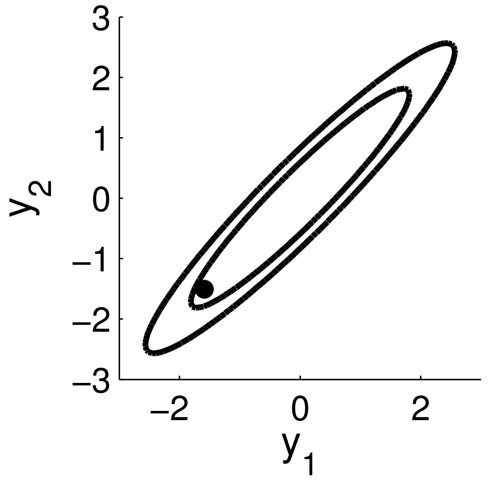
New visualisation



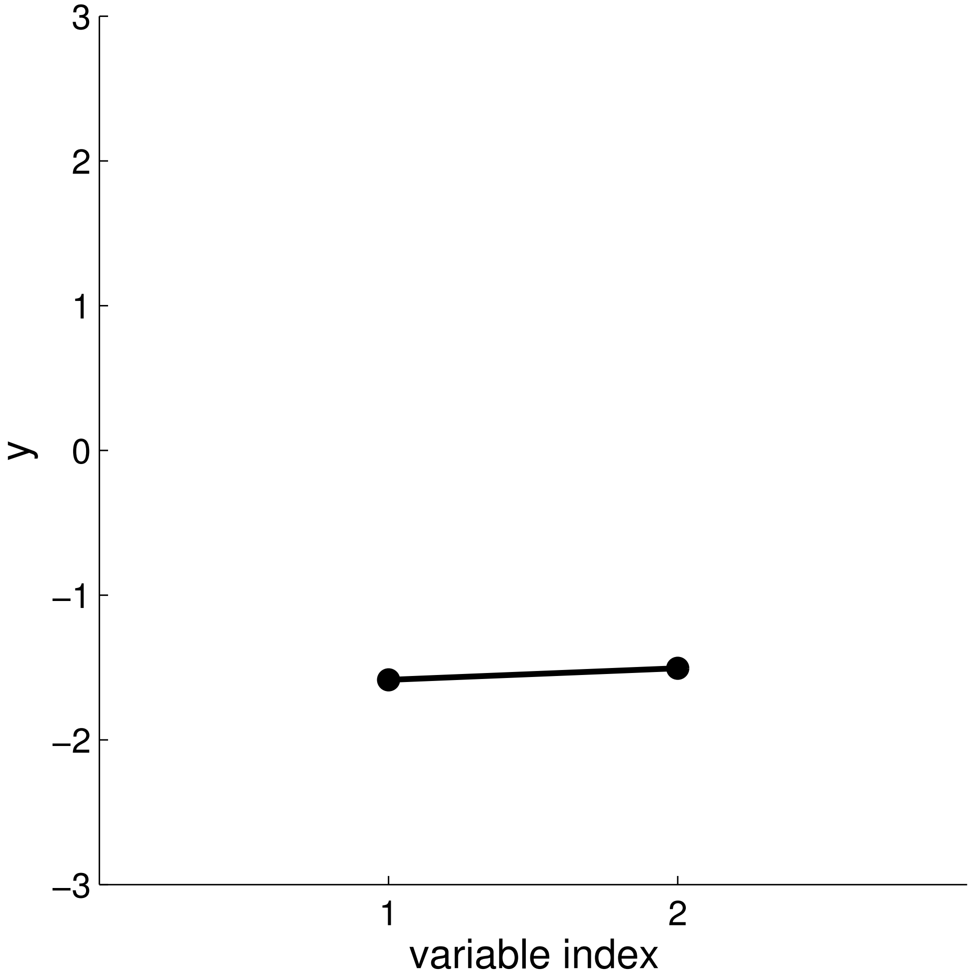
$$\Sigma = \begin{bmatrix} 1 & .9 \\ .9 & 1 \end{bmatrix}$$



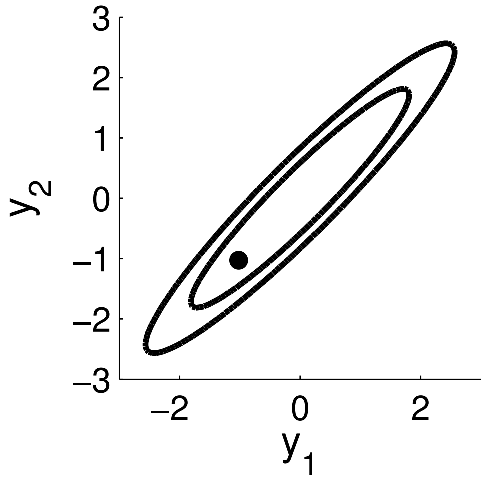
New visualisation



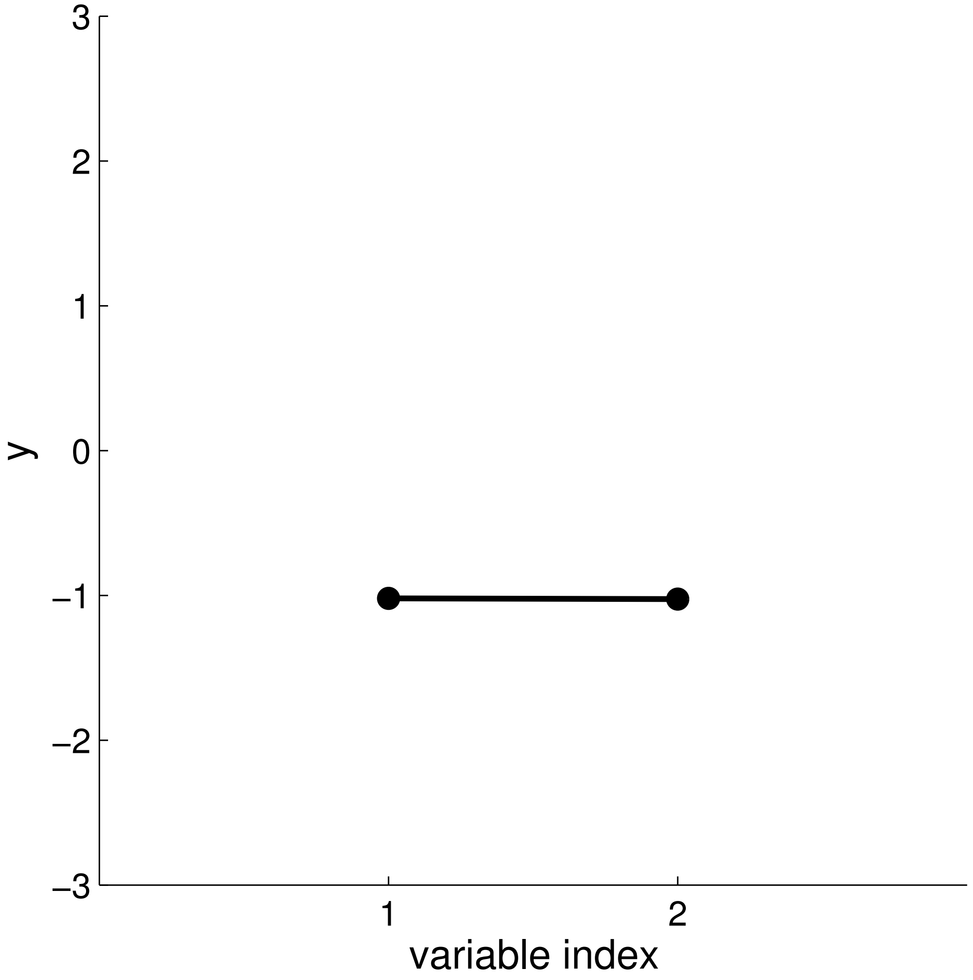
$$\Sigma = \begin{bmatrix} 1 & .9 \\ .9 & 1 \end{bmatrix}$$



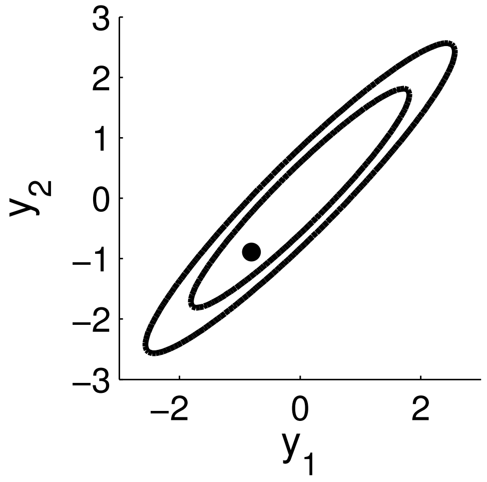
New visualisation



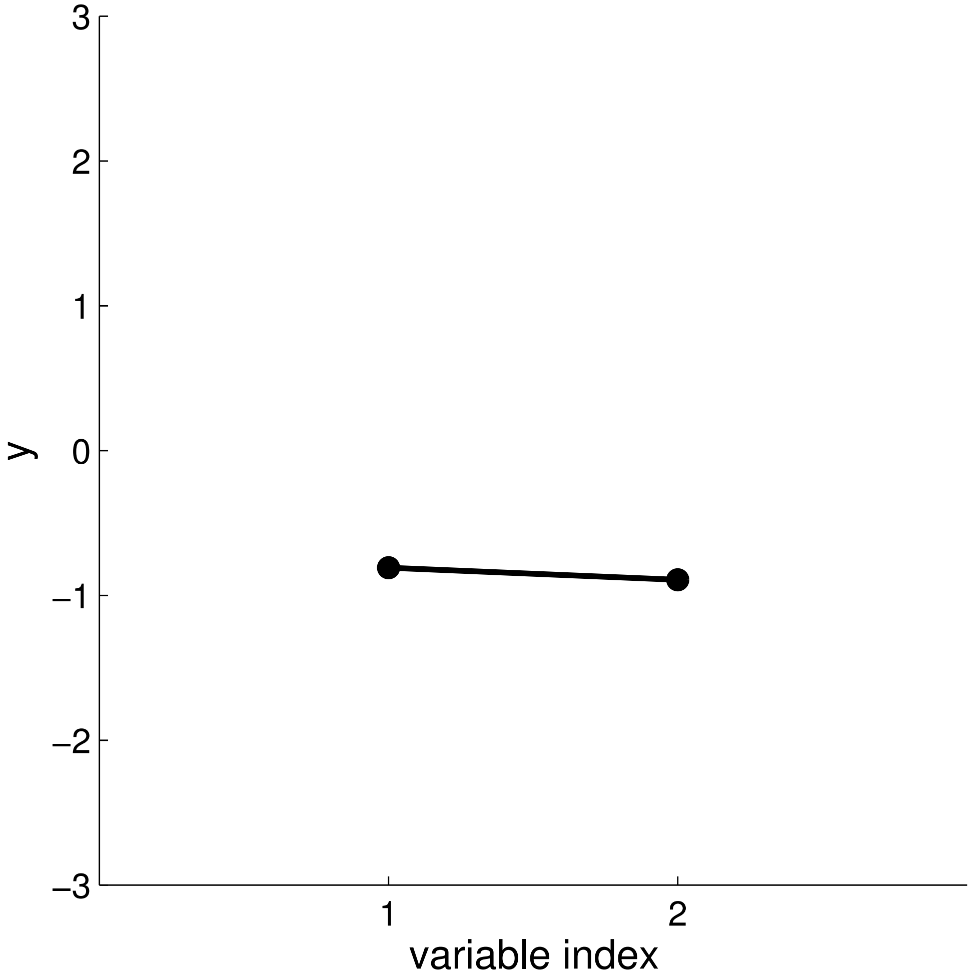
$$\Sigma = \begin{bmatrix} 1 & .9 \\ .9 & 1 \end{bmatrix}$$



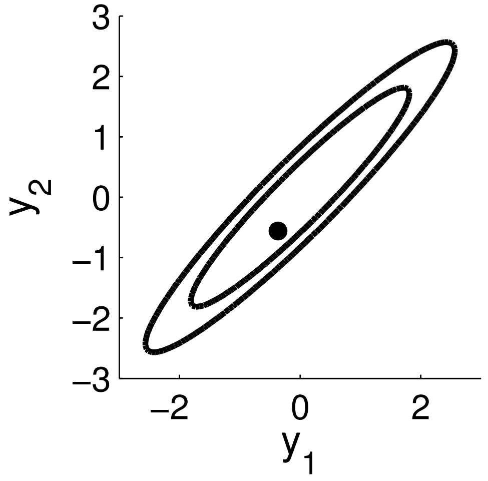
New visualisation



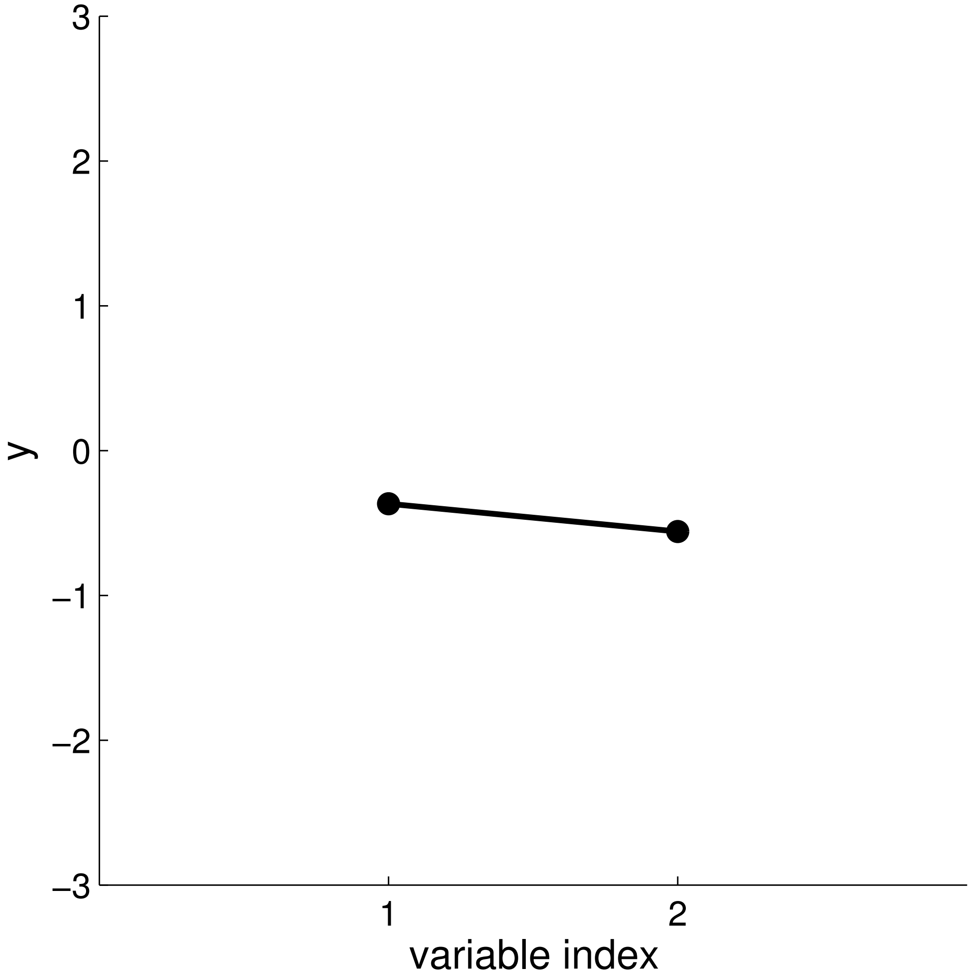
$$\Sigma = \left[\begin{array}{cc} 1 & .9 \\ .9 & 1 \end{array} \right]$$



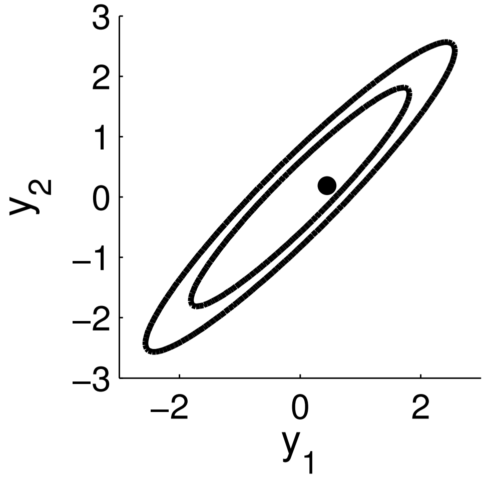
New visualisation



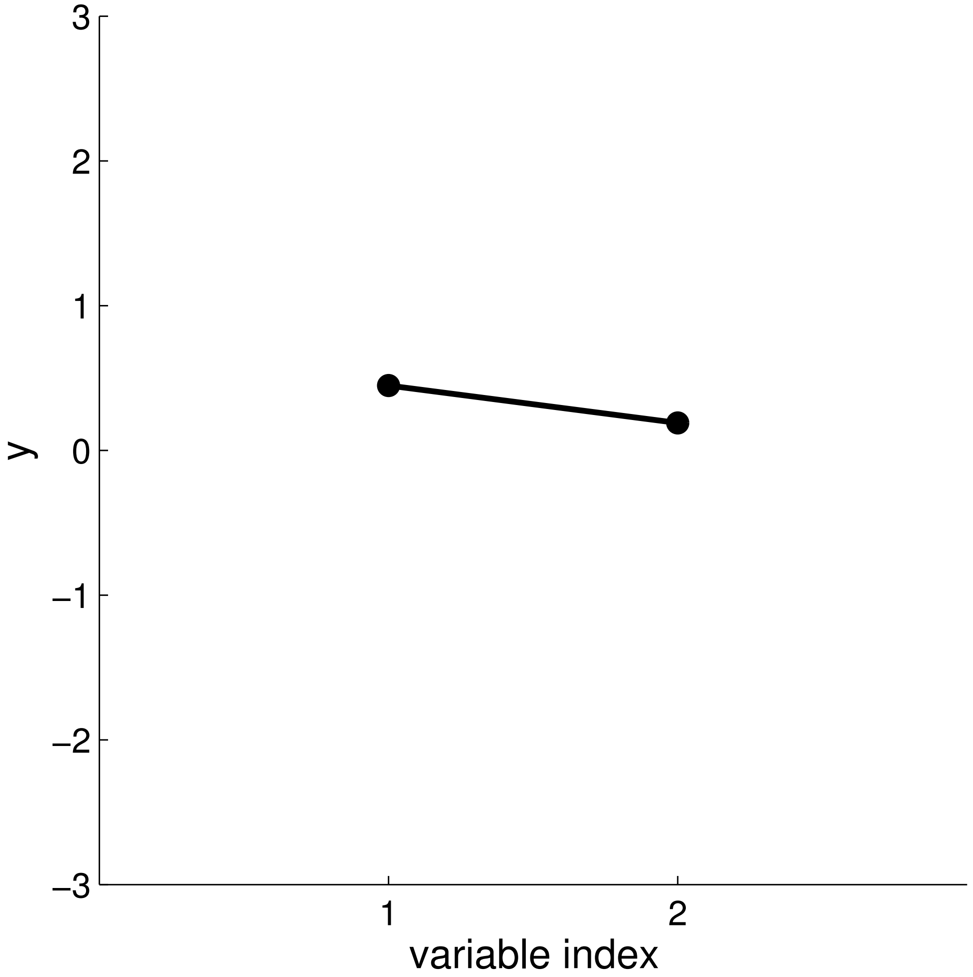
$$\Sigma = \begin{bmatrix} 1 & .9 \\ .9 & 1 \end{bmatrix}$$



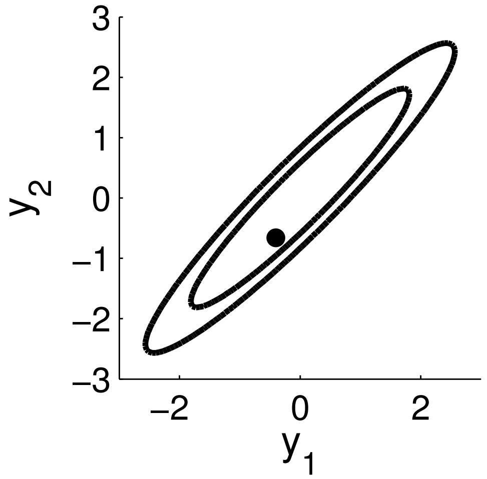
New visualisation



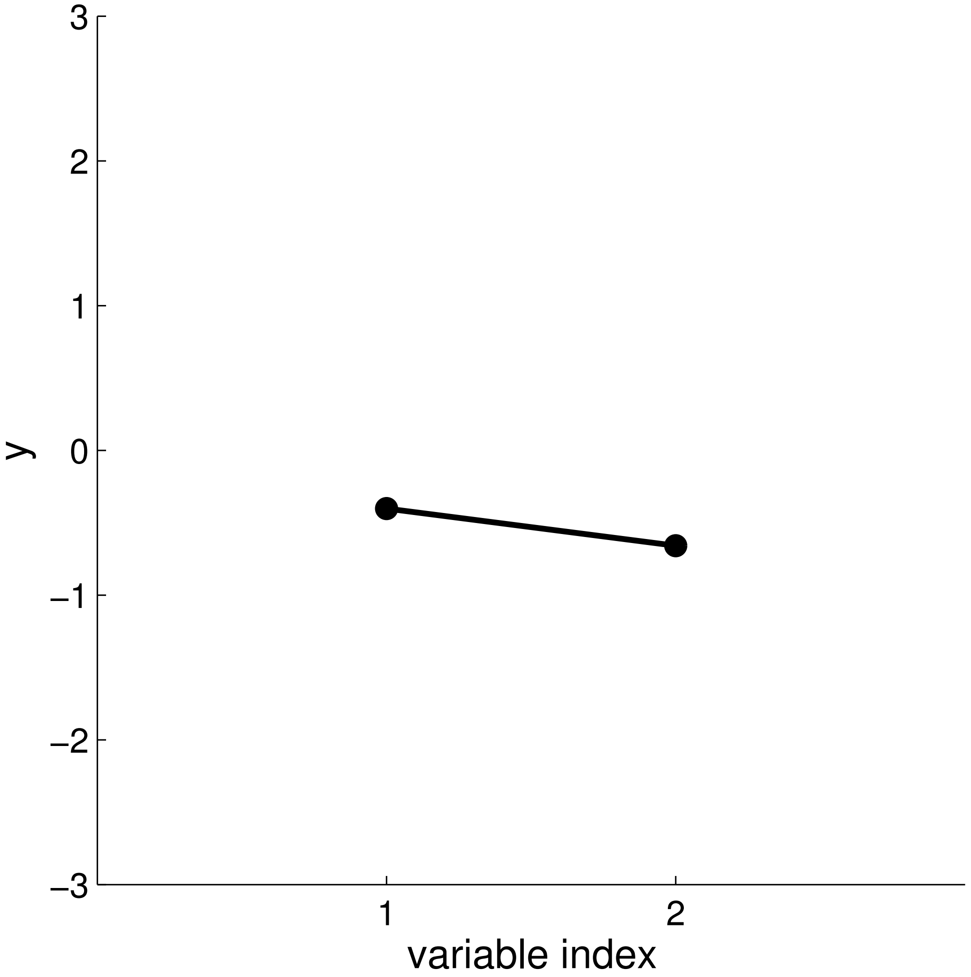
$$\Sigma = \begin{bmatrix} 1 & .9 \\ .9 & 1 \end{bmatrix}$$



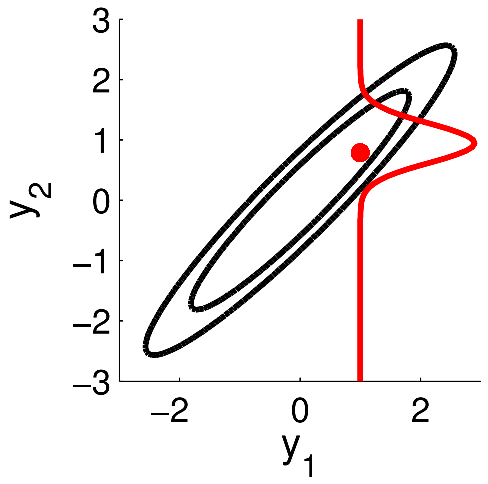
New visualisation



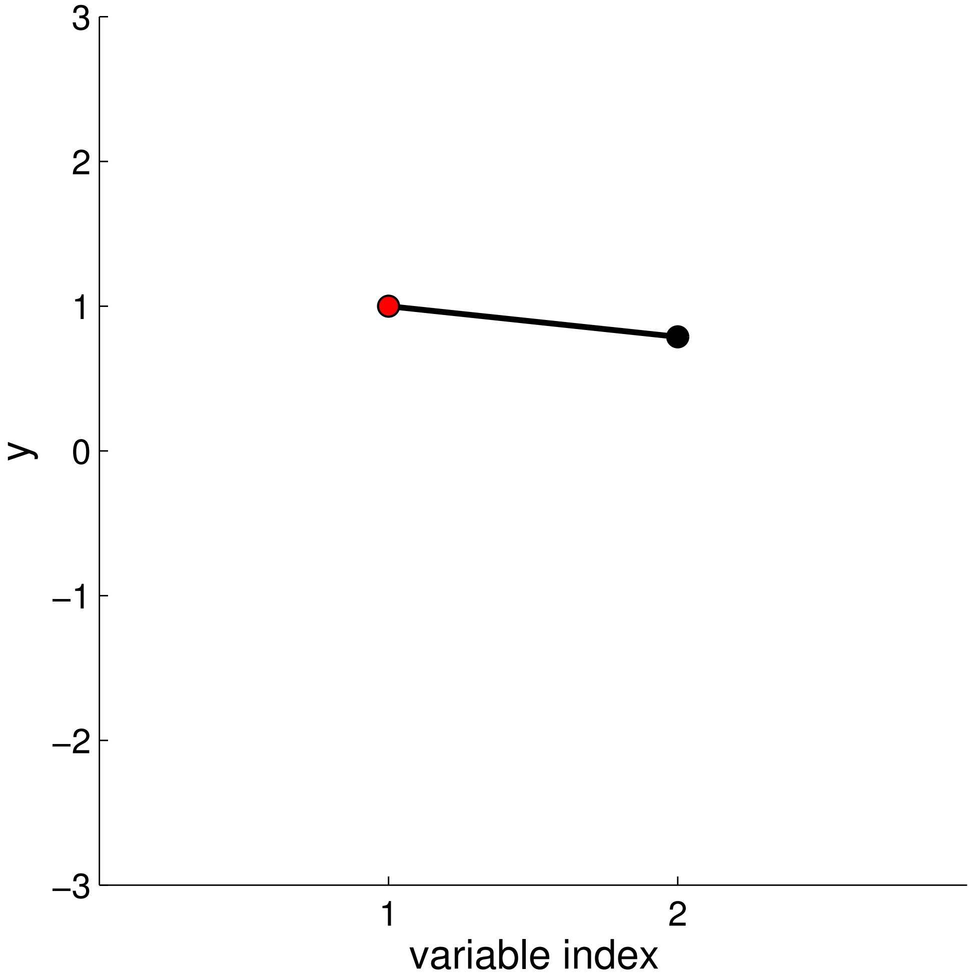
$$\Sigma = \begin{bmatrix} 1 & .9 \\ .9 & 1 \end{bmatrix}$$



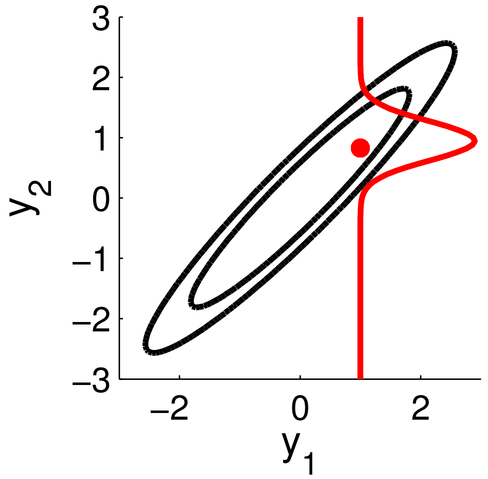
New visualisation



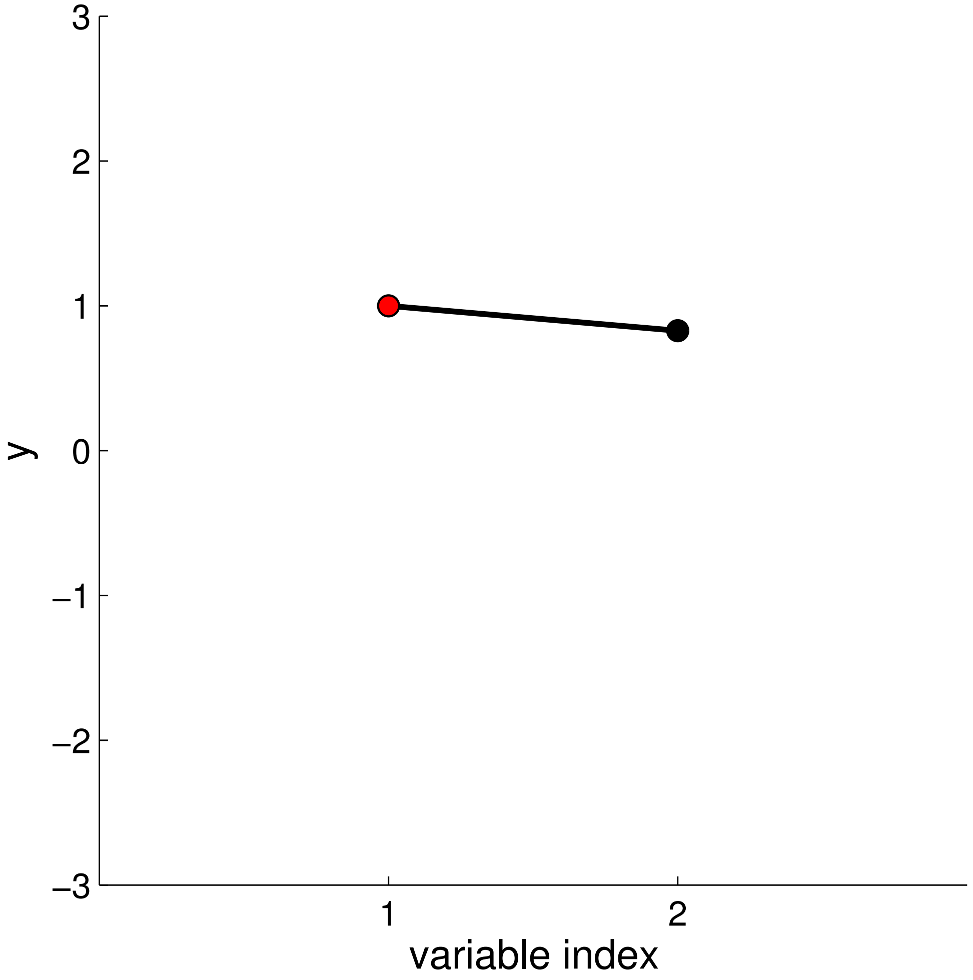
$$\Sigma = \begin{bmatrix} 1 & .9 \\ .9 & 1 \end{bmatrix}$$



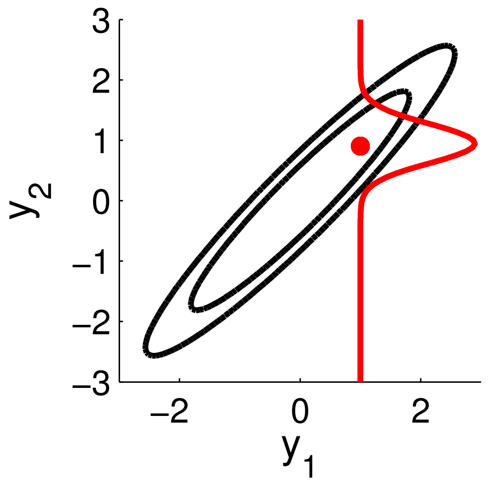
New visualisation



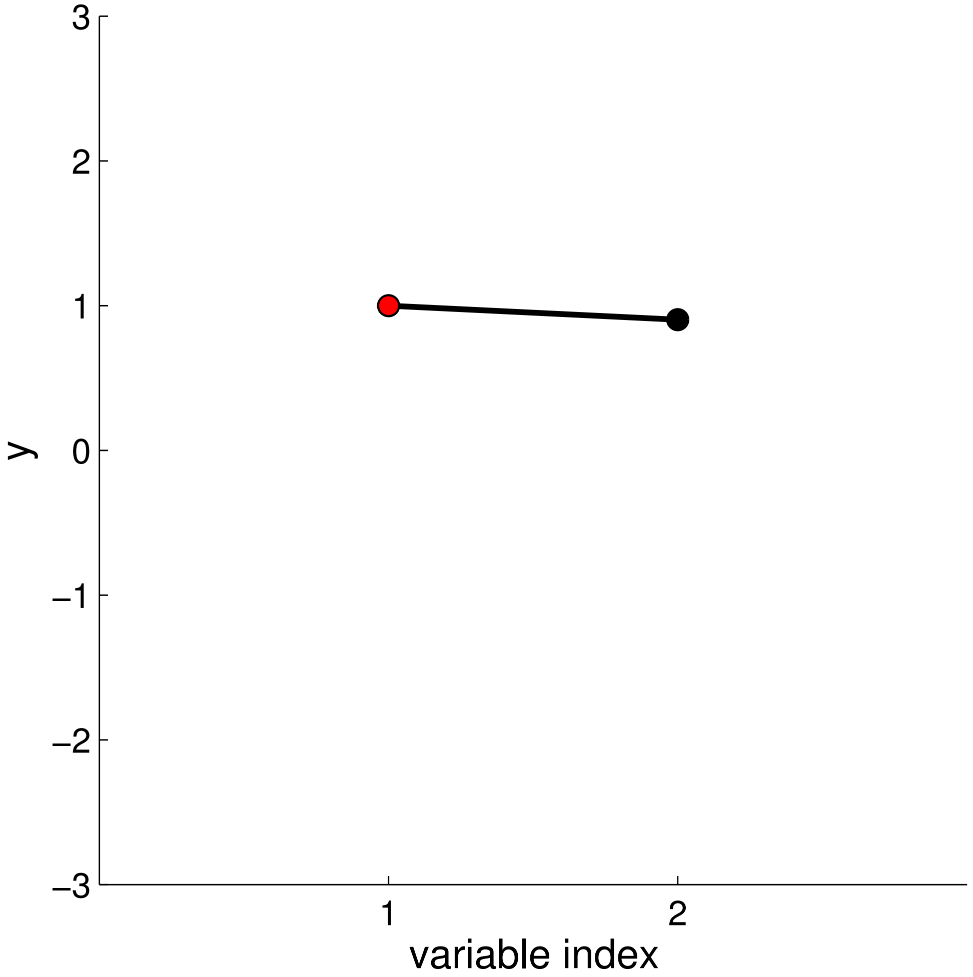
$$\Sigma = \begin{bmatrix} 1 & .9 \\ .9 & 1 \end{bmatrix}$$



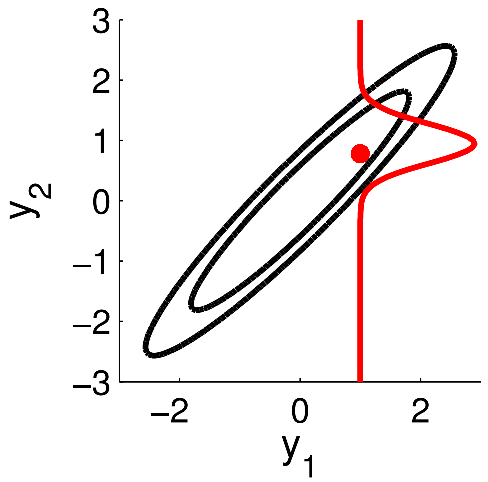
New visualisation



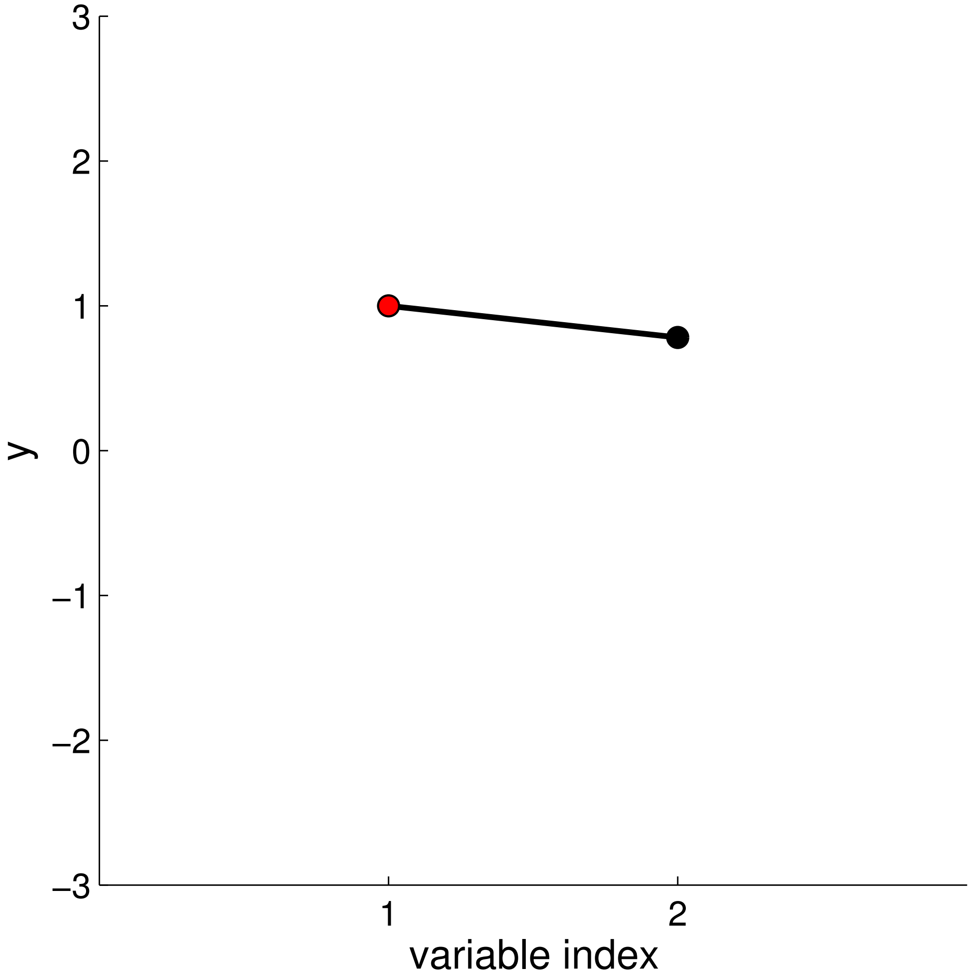
$$\Sigma = \begin{bmatrix} 1 & .9 \\ .9 & 1 \end{bmatrix}$$



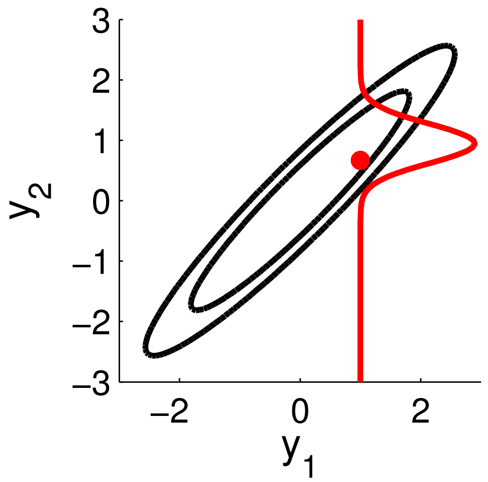
New visualisation



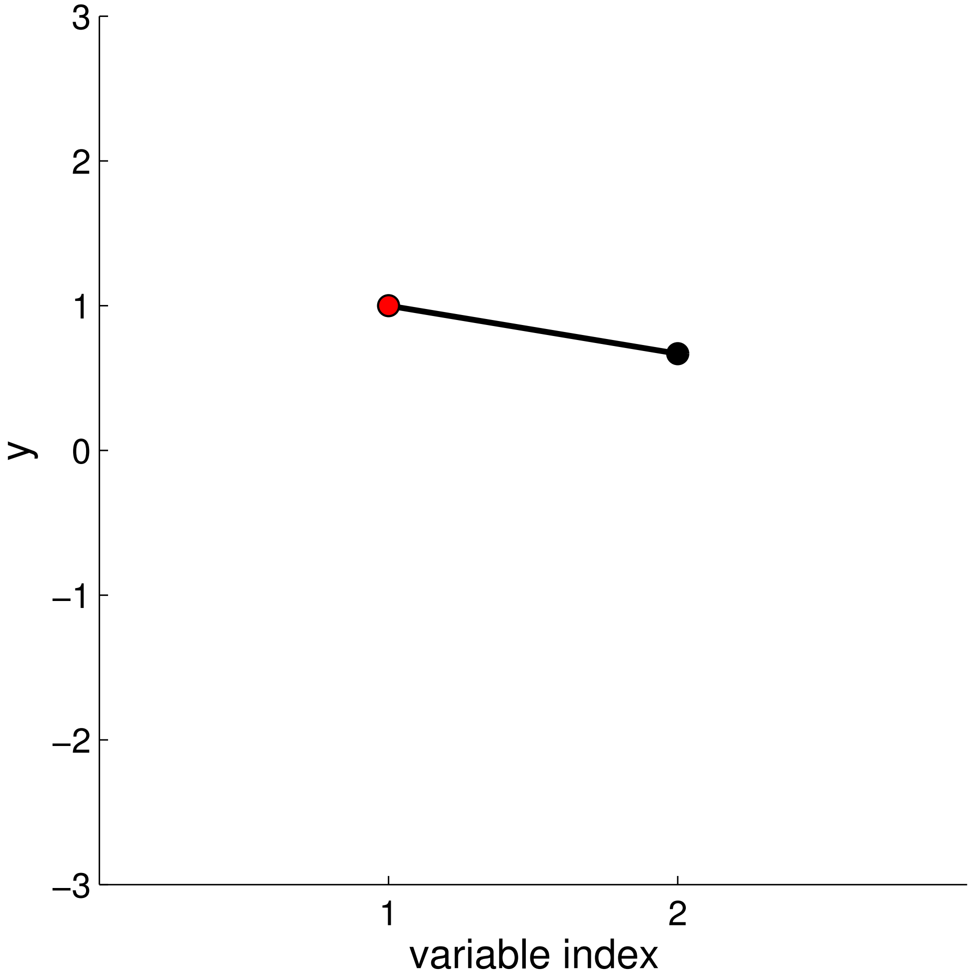
$$\Sigma = \begin{bmatrix} 1 & .9 \\ .9 & 1 \end{bmatrix}$$



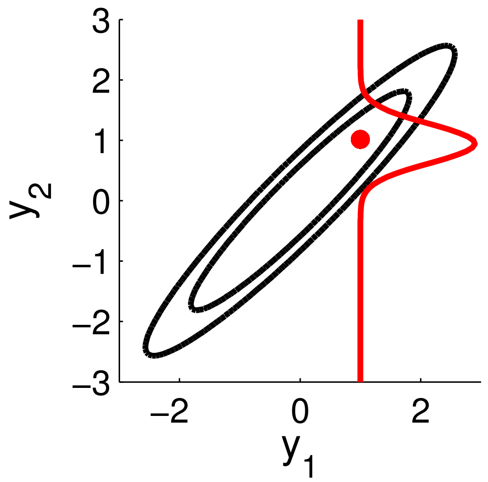
New visualisation



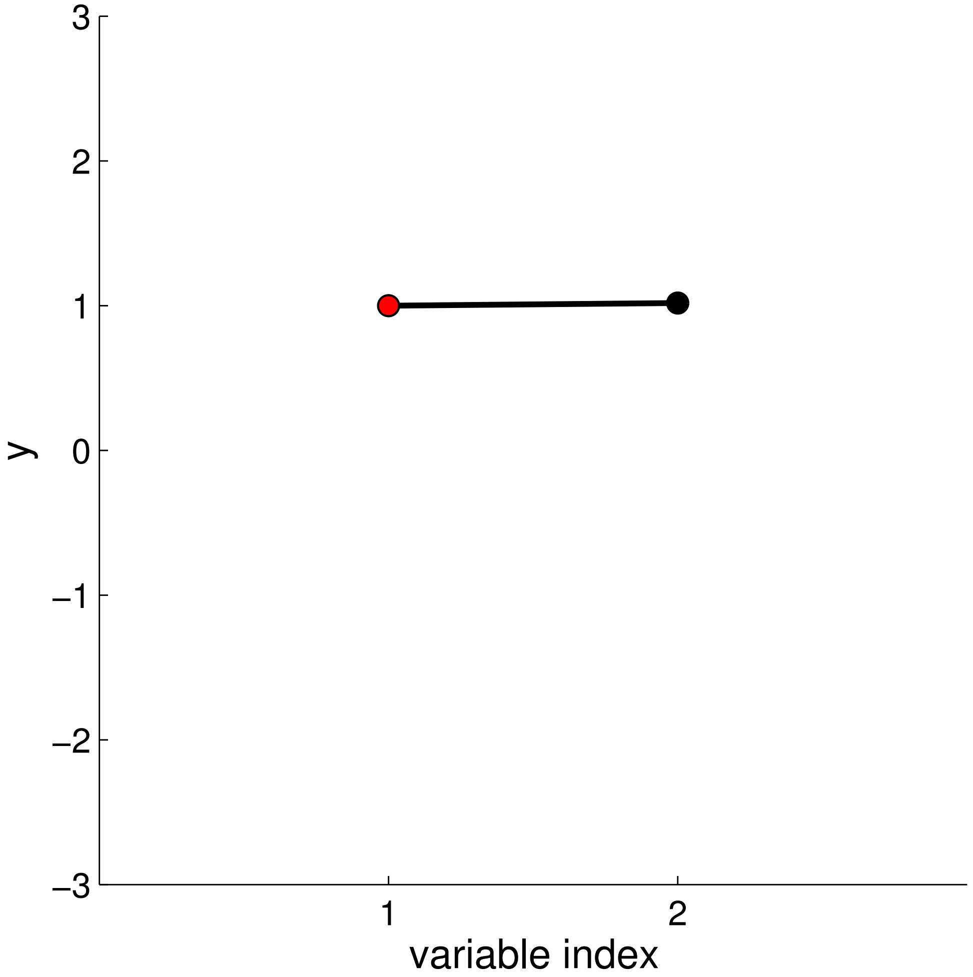
$$\Sigma = \begin{bmatrix} 1 & .9 \\ .9 & 1 \end{bmatrix}$$



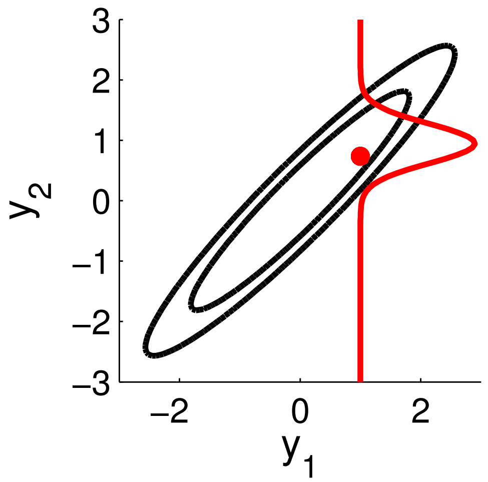
New visualisation



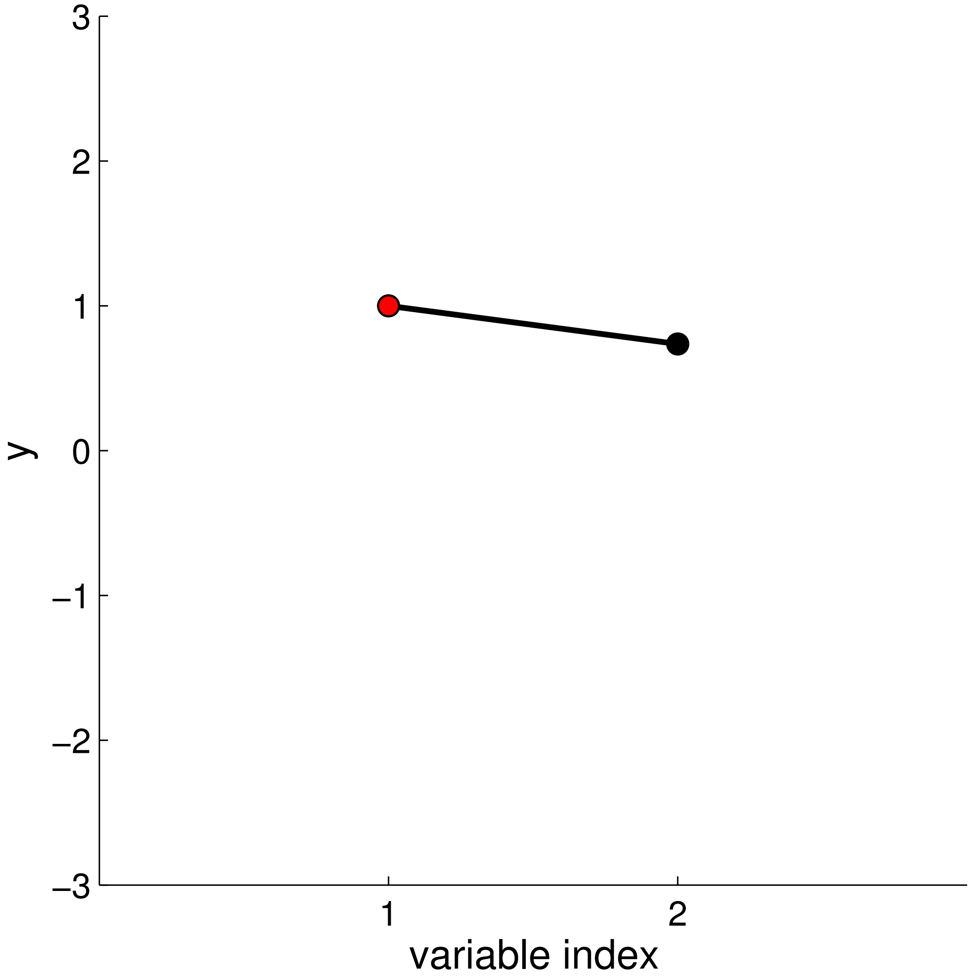
$$\Sigma = \begin{bmatrix} 1 & .9 \\ .9 & 1 \end{bmatrix}$$



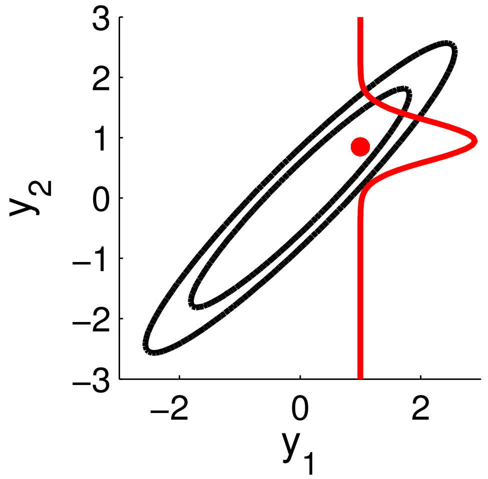
New visualisation



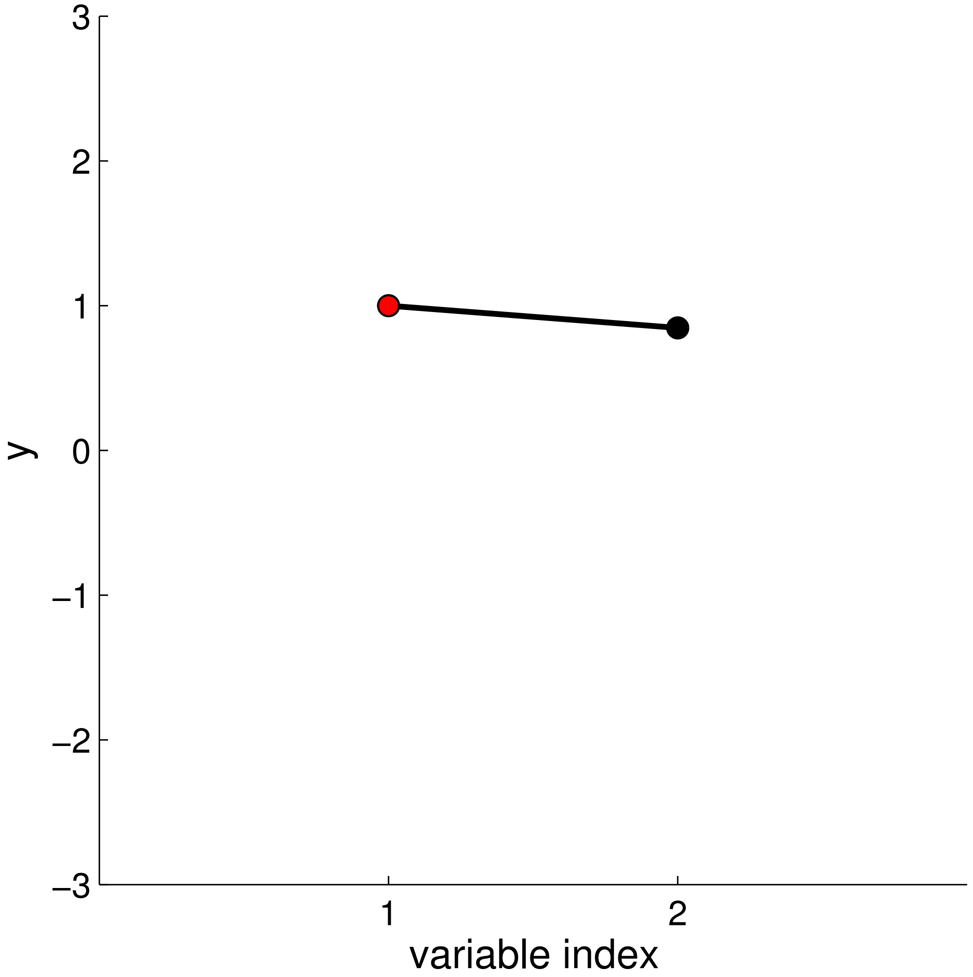
$$\Sigma = \begin{bmatrix} 1 & .9 \\ .9 & 1 \end{bmatrix}$$



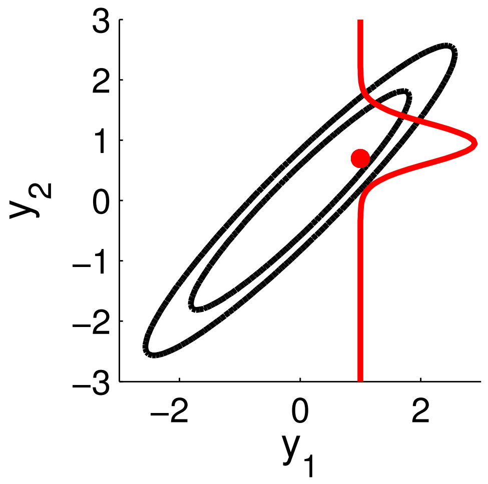
New visualisation



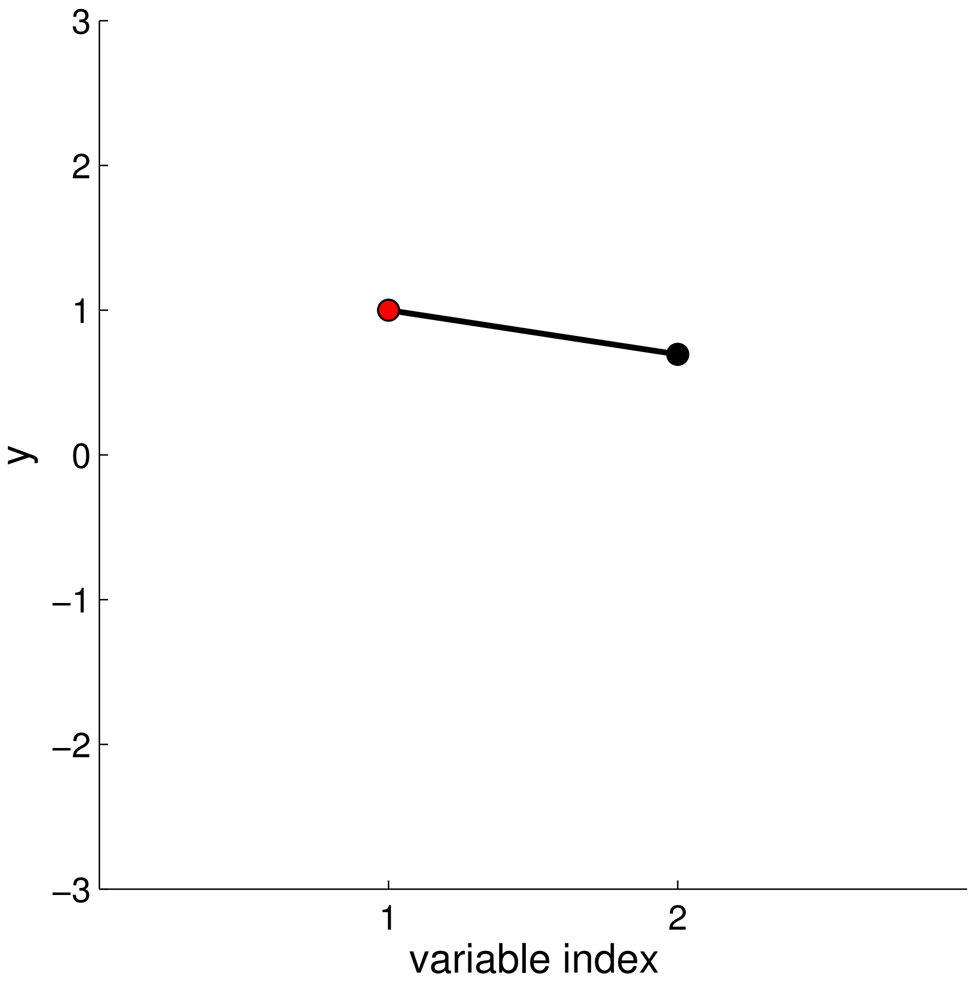
$$\Sigma = \begin{bmatrix} 1 & .9 \\ .9 & 1 \end{bmatrix}$$



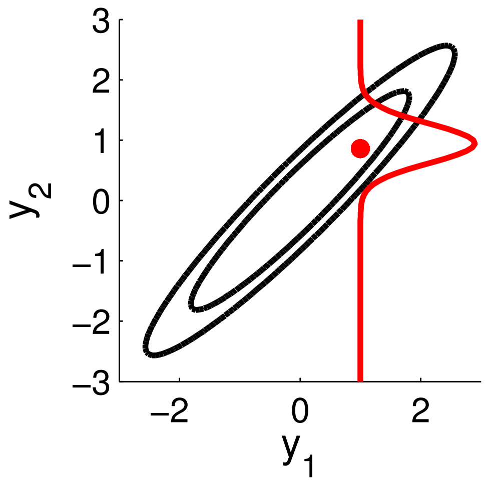
New visualisation



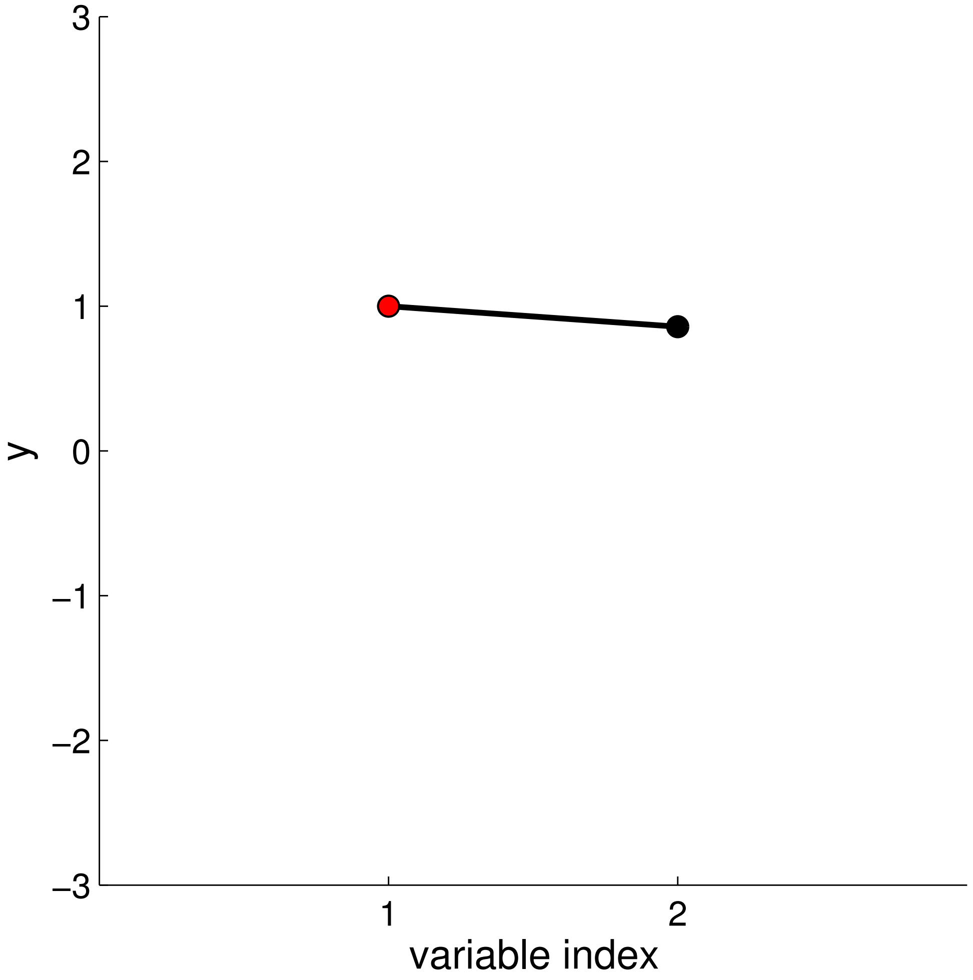
$$\Sigma = \begin{bmatrix} 1 & .9 \\ .9 & 1 \end{bmatrix}$$



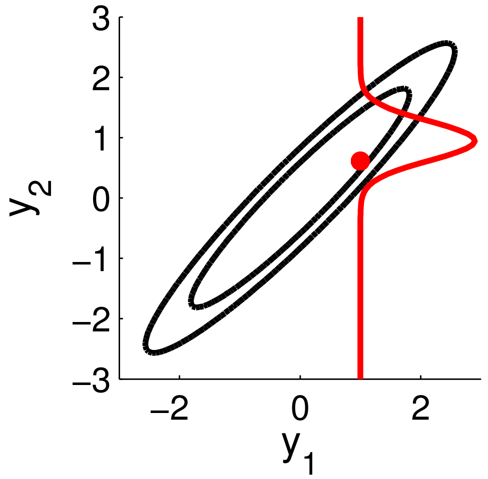
New visualisation

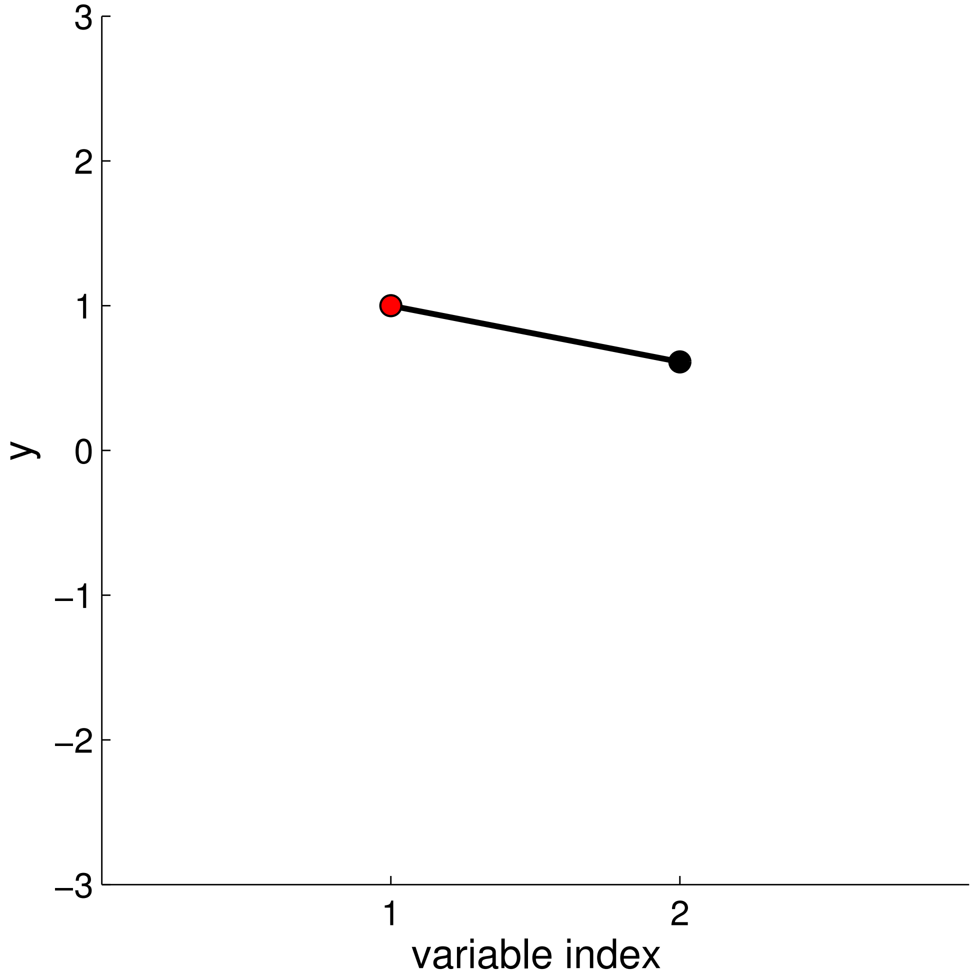


$$\Sigma = \begin{bmatrix} 1 & .9 \\ .9 & 1 \end{bmatrix}$$



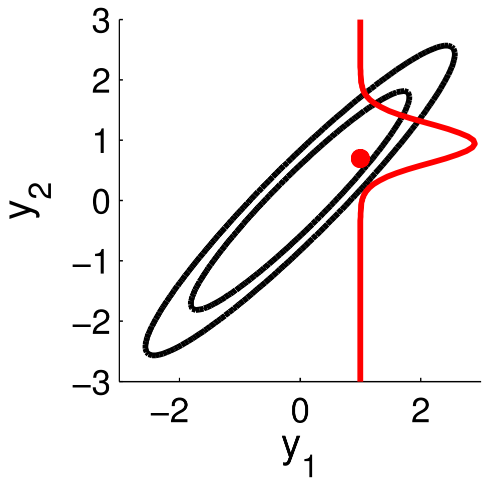
New visualisation



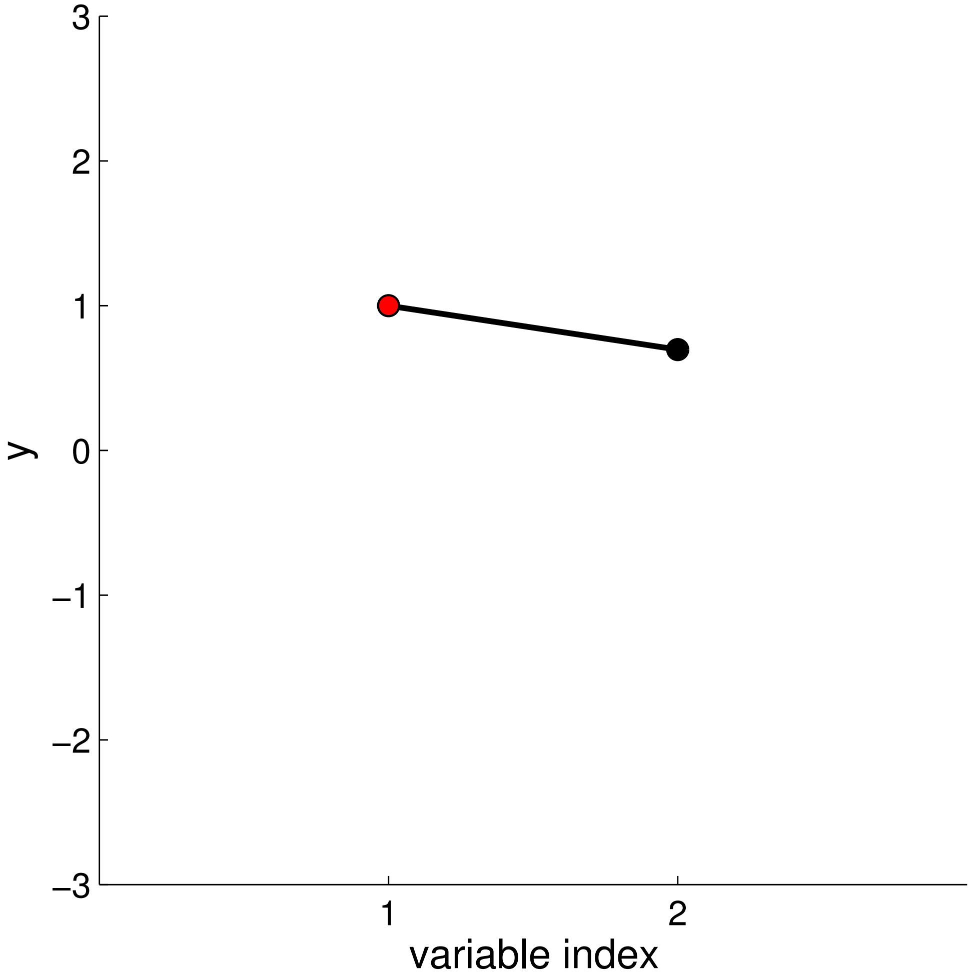


$$\Sigma = \begin{bmatrix} 1 & .9 \\ .9 & 1 \end{bmatrix}$$

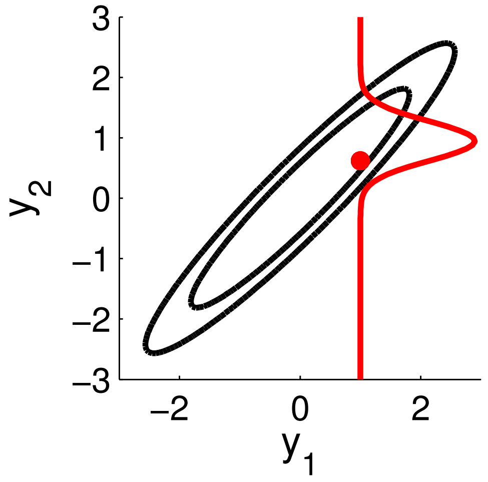
New visualisation



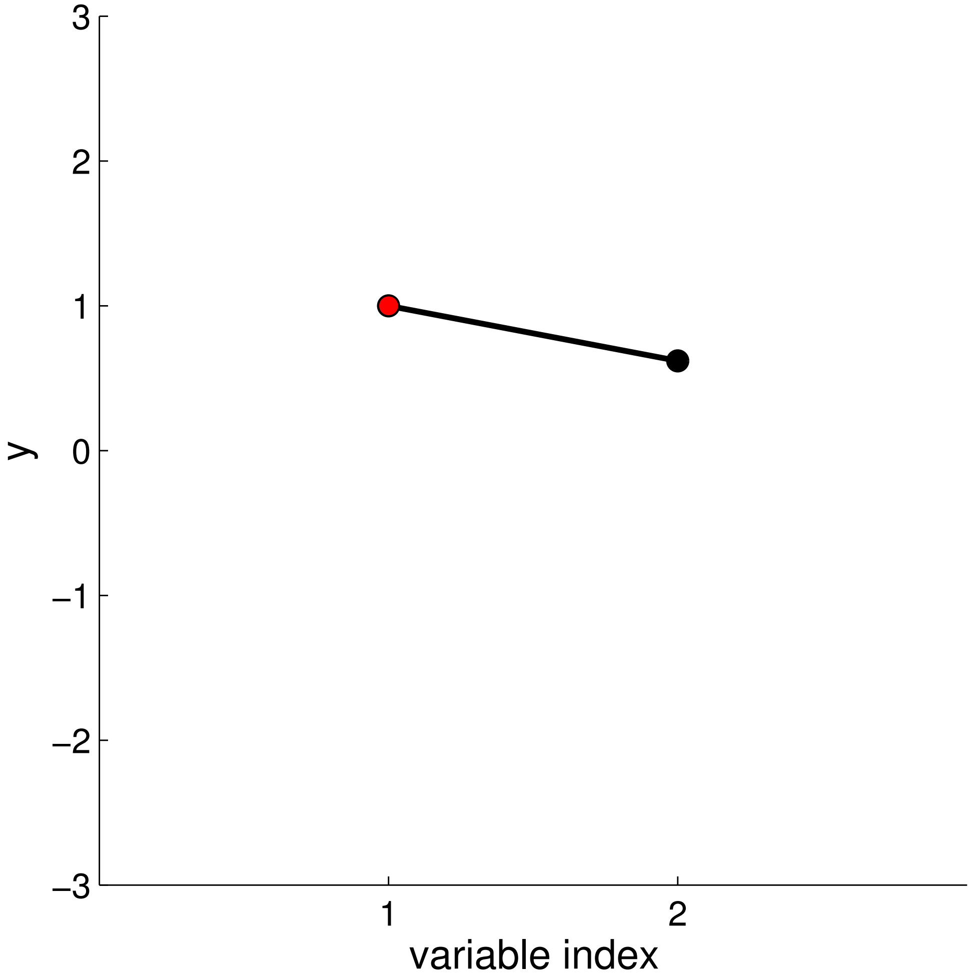
$$\Sigma = \begin{bmatrix} 1 & .9 \\ .9 & 1 \end{bmatrix}$$



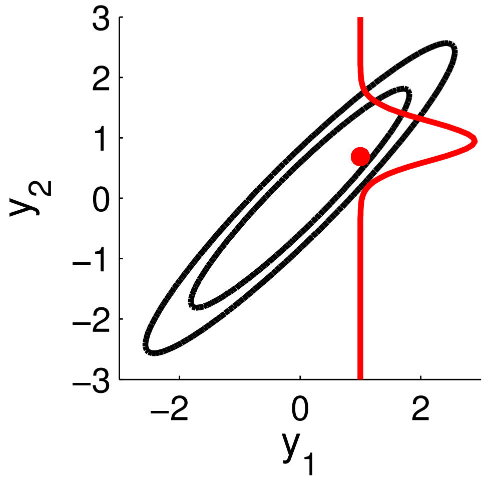
New visualisation



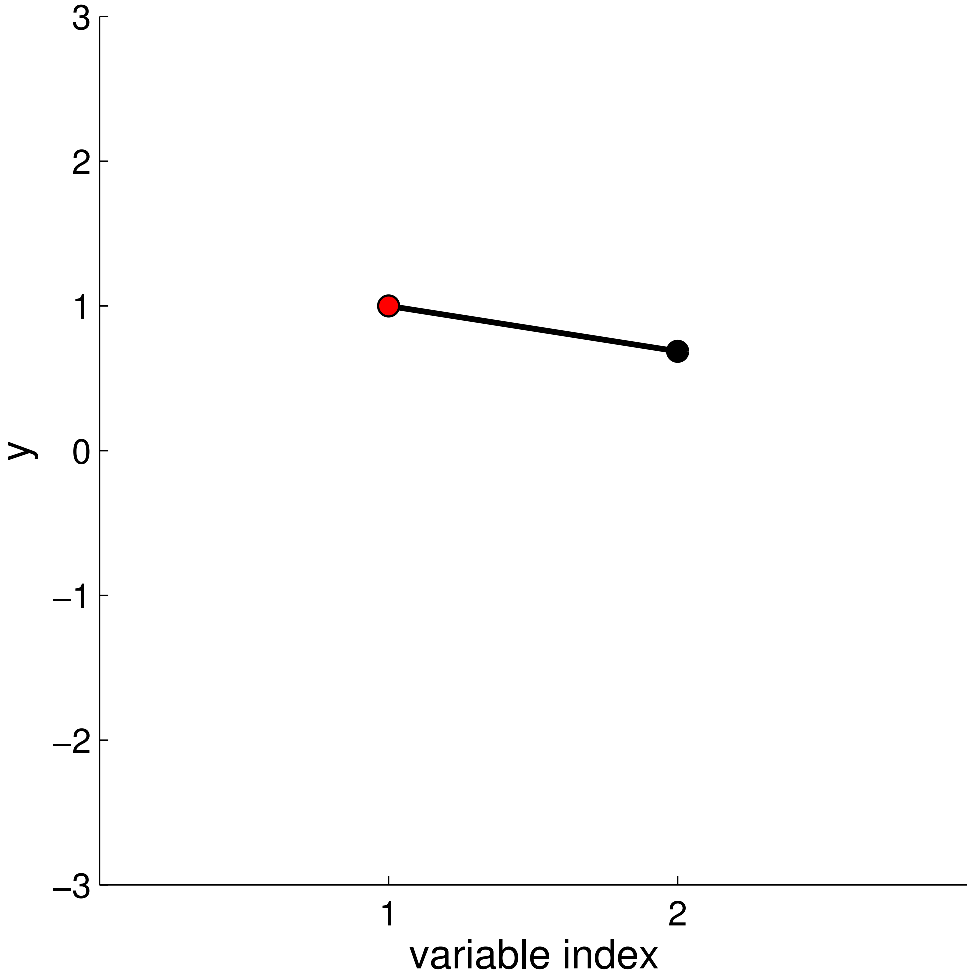
$$\Sigma = \begin{bmatrix} 1 & .9 \\ .9 & 1 \end{bmatrix}$$



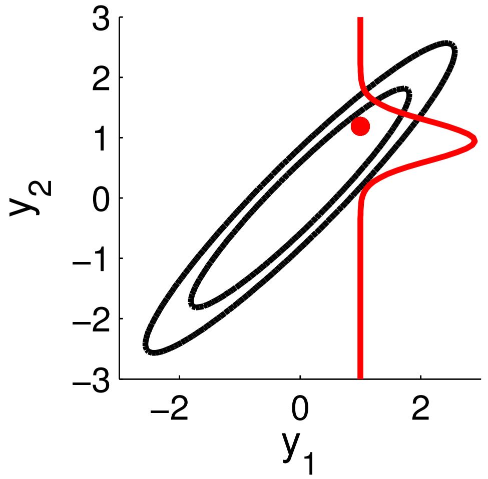
New visualisation



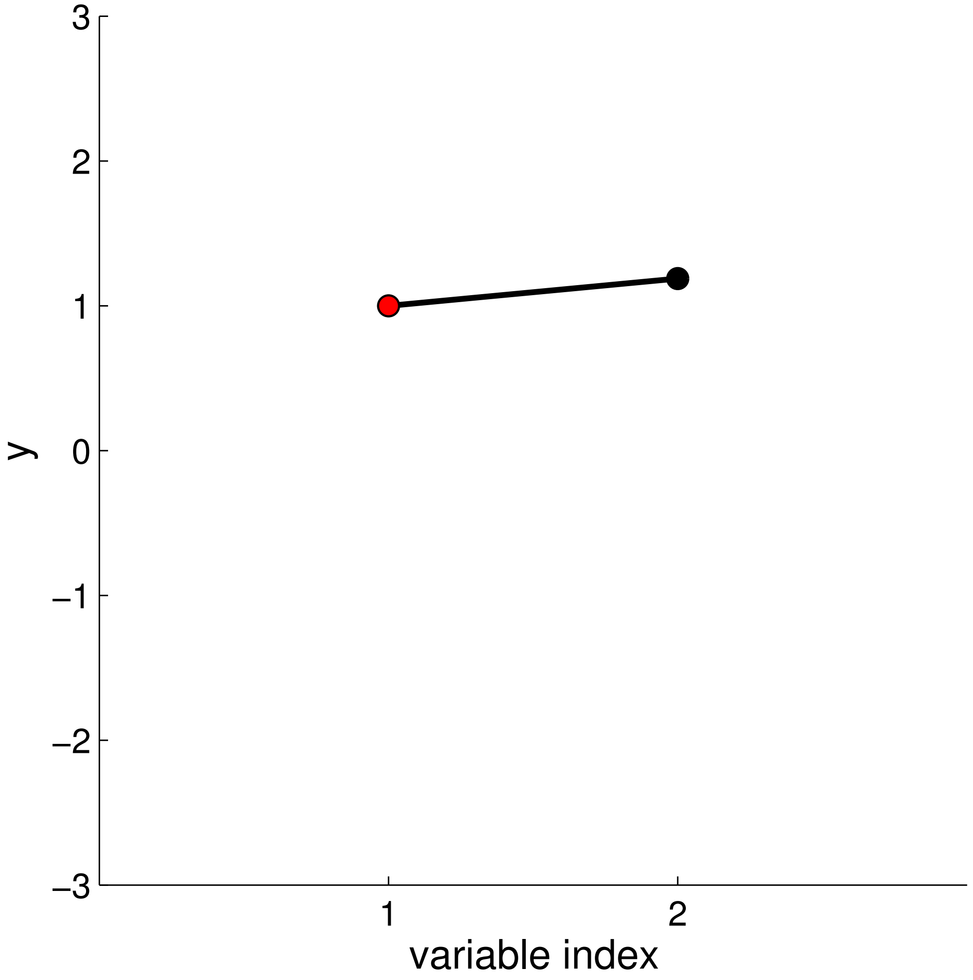
$$\Sigma = \begin{bmatrix} 1 & .9 \\ .9 & 1 \end{bmatrix}$$



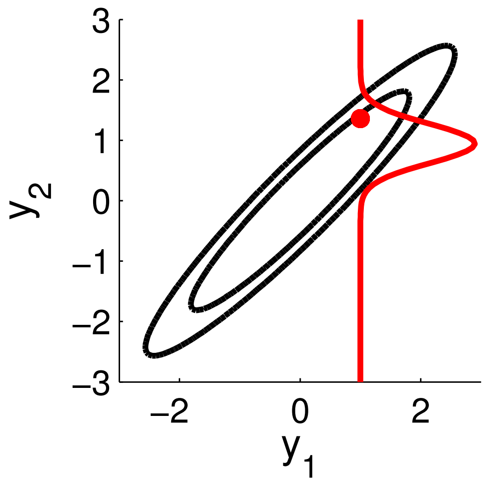
New visualisation



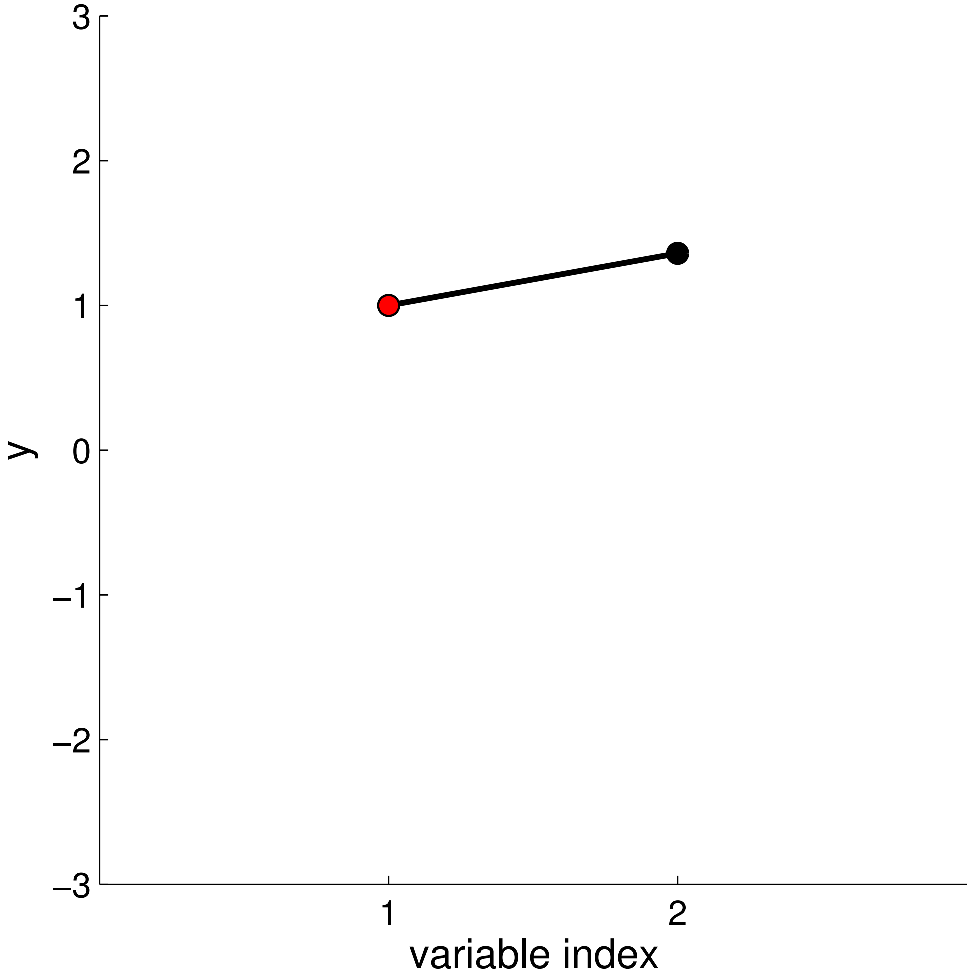
$$\Sigma = \begin{bmatrix} 1 & .9 \\ .9 & 1 \end{bmatrix}$$



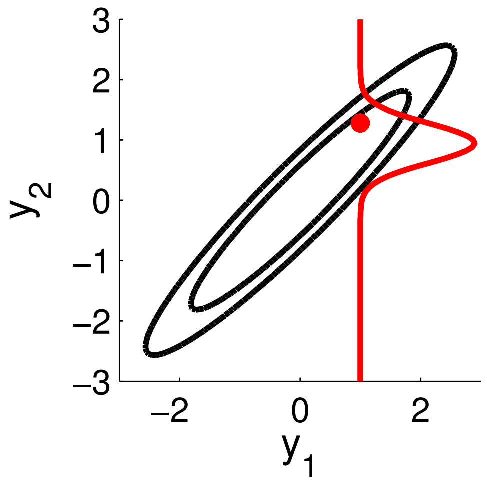
New visualisation



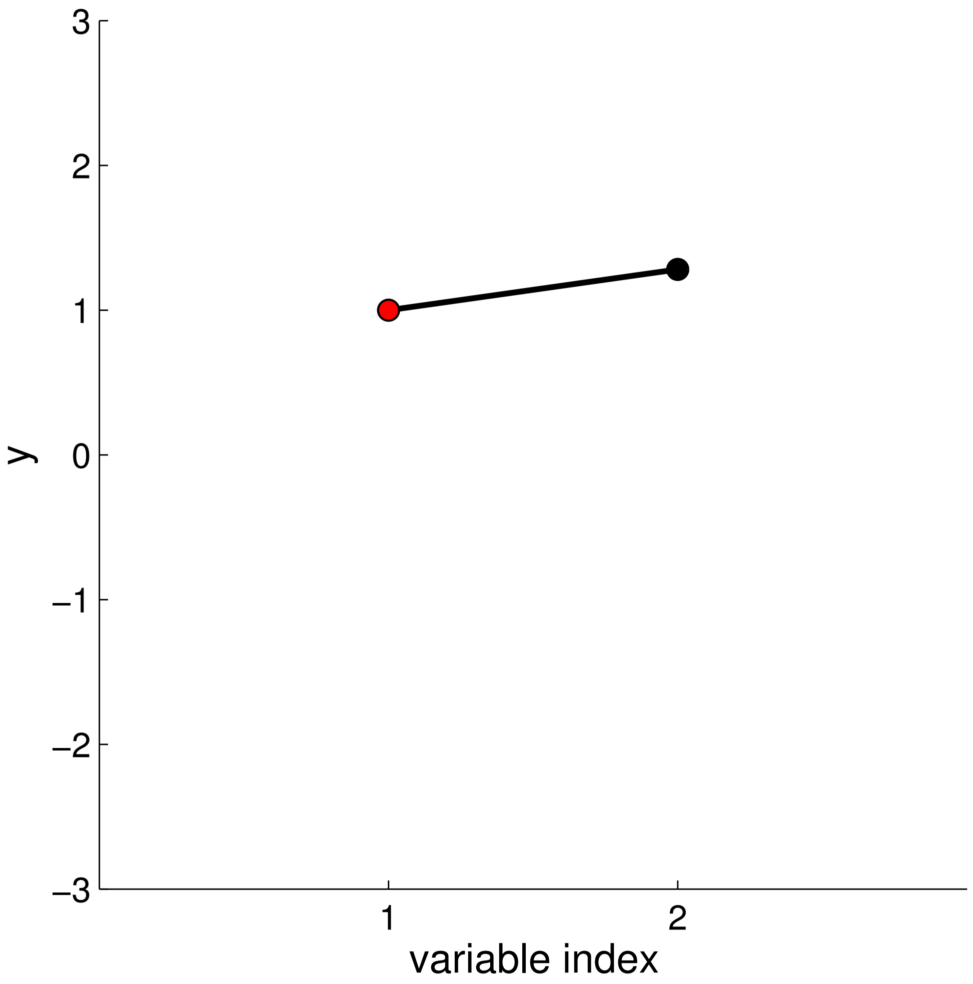
$$\Sigma = \begin{bmatrix} 1 & .9 \\ .9 & 1 \end{bmatrix}$$



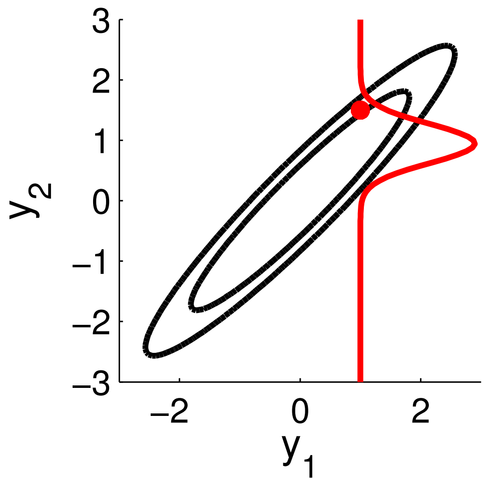
New visualisation

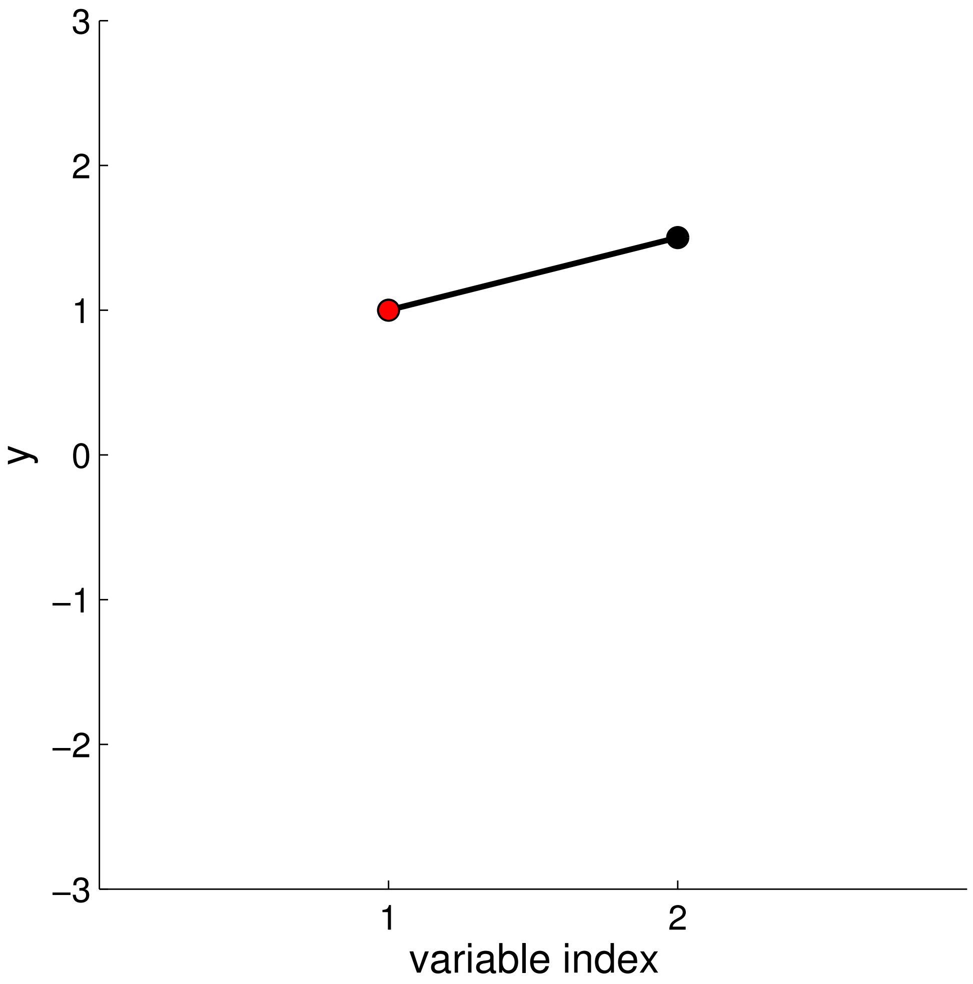


$$\Sigma = \begin{bmatrix} 1 & .9 \\ .9 & 1 \end{bmatrix}$$



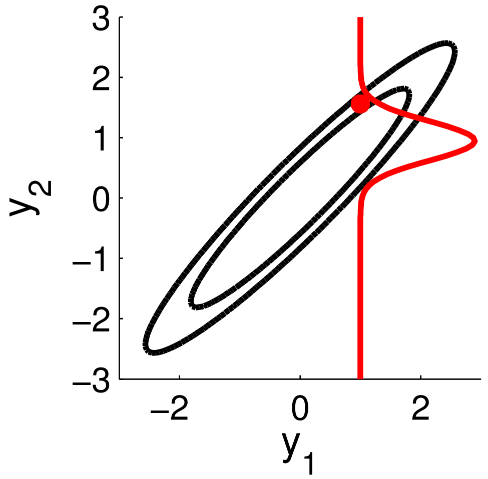
New visualisation



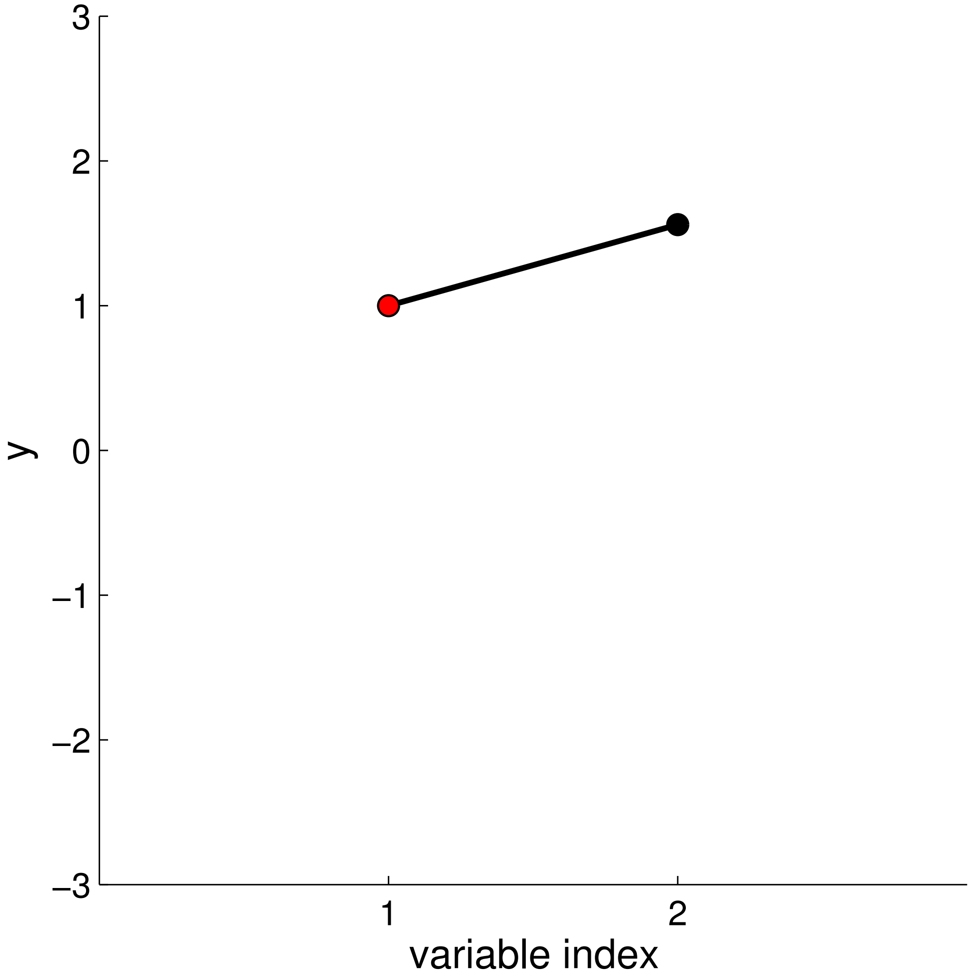


$$\Sigma = \begin{bmatrix} 1 & .9 \\ .9 & 1 \end{bmatrix}$$

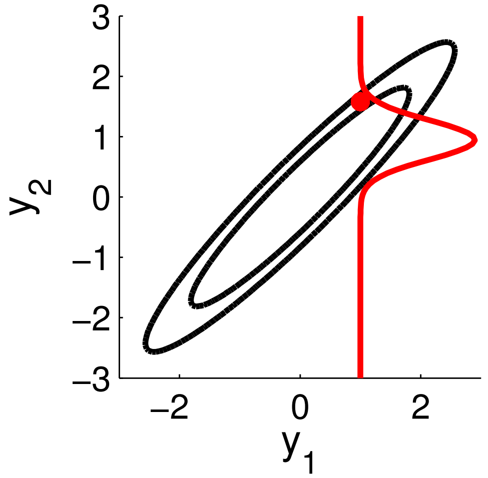
New visualisation

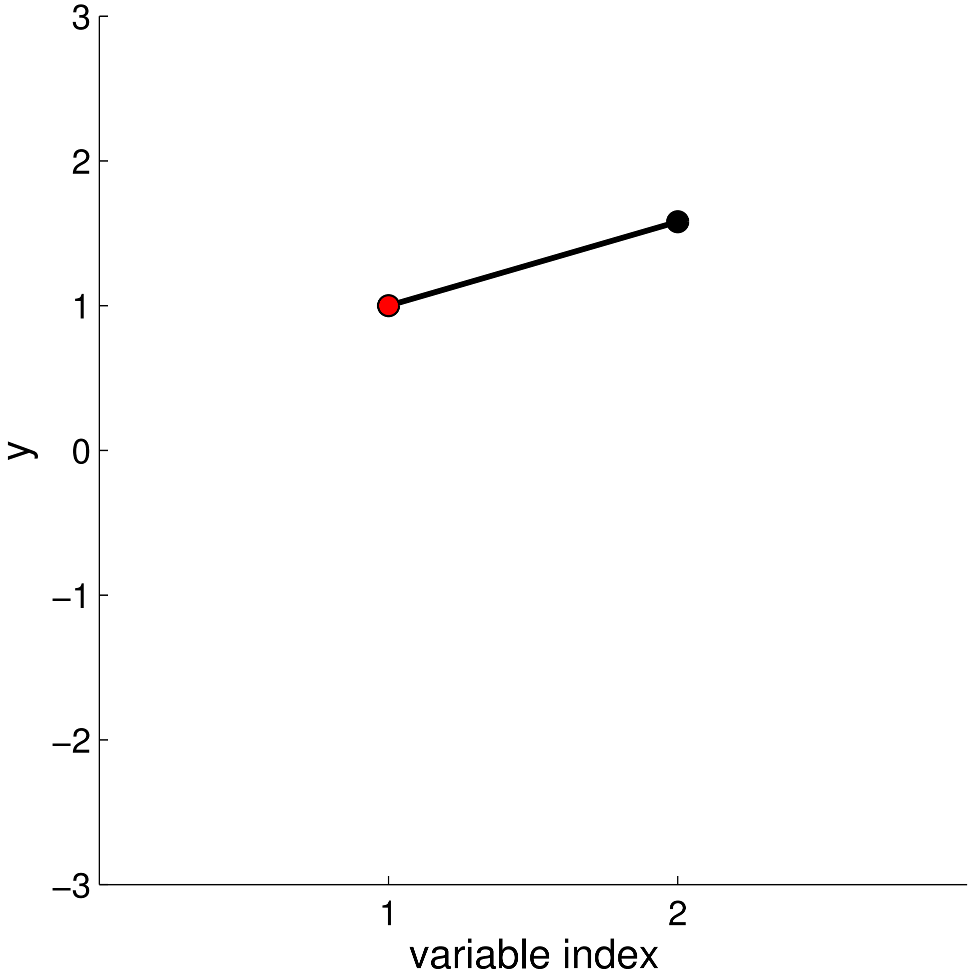


$$\Sigma = \begin{bmatrix} 1 & .9 \\ .9 & 1 \end{bmatrix}$$



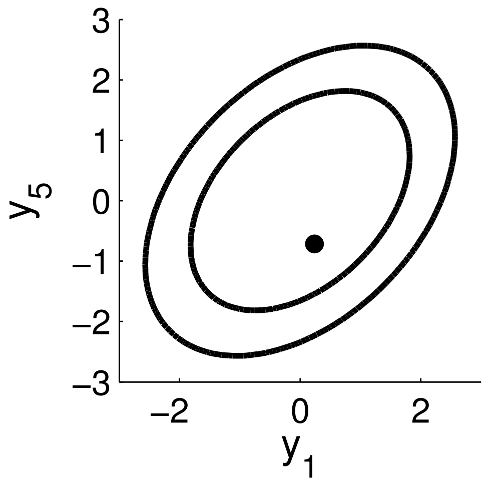
New visualisation

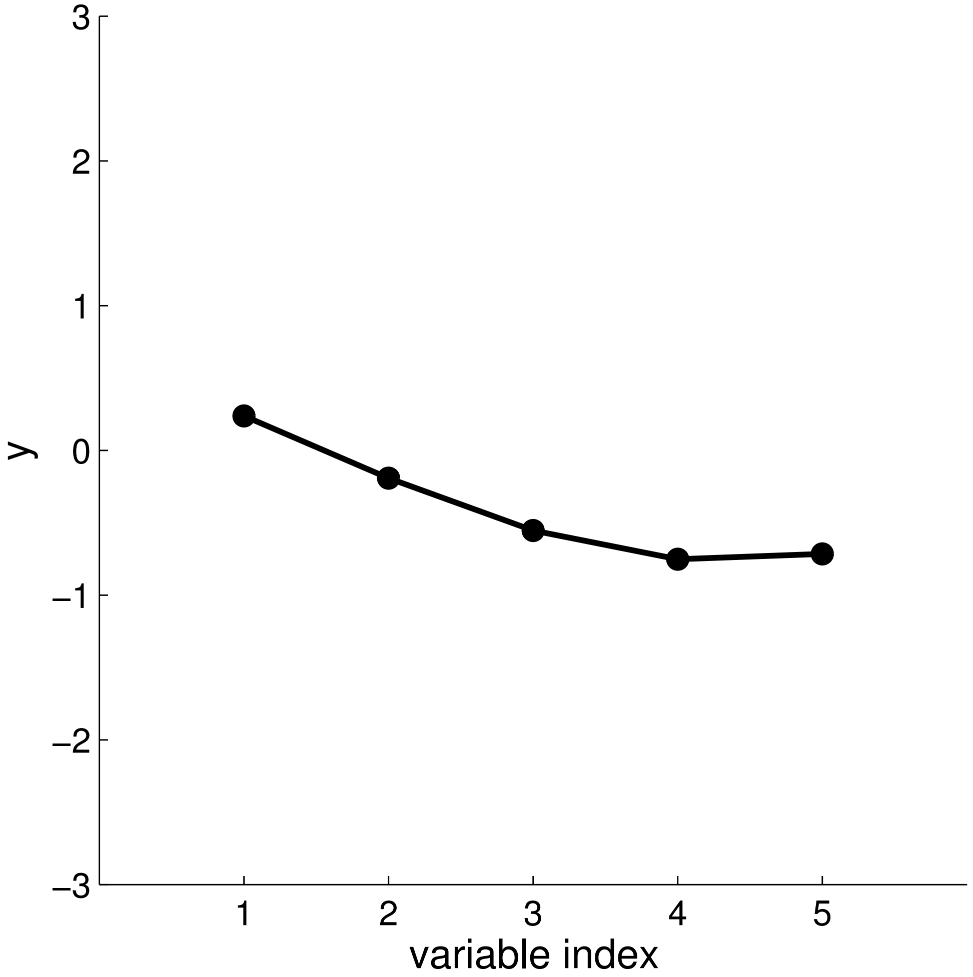




$$\Sigma = \begin{bmatrix} 1 & .9 \\ .9 & 1 \end{bmatrix}$$

New visualisation

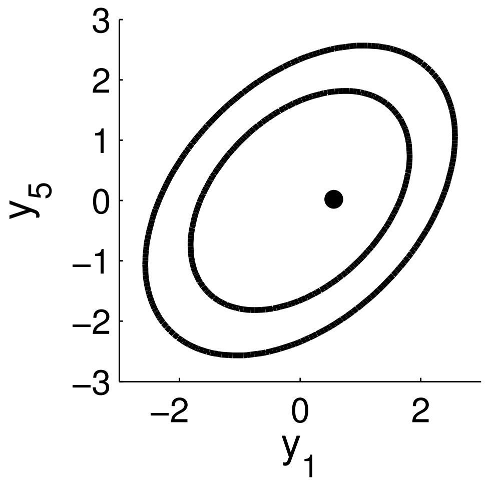




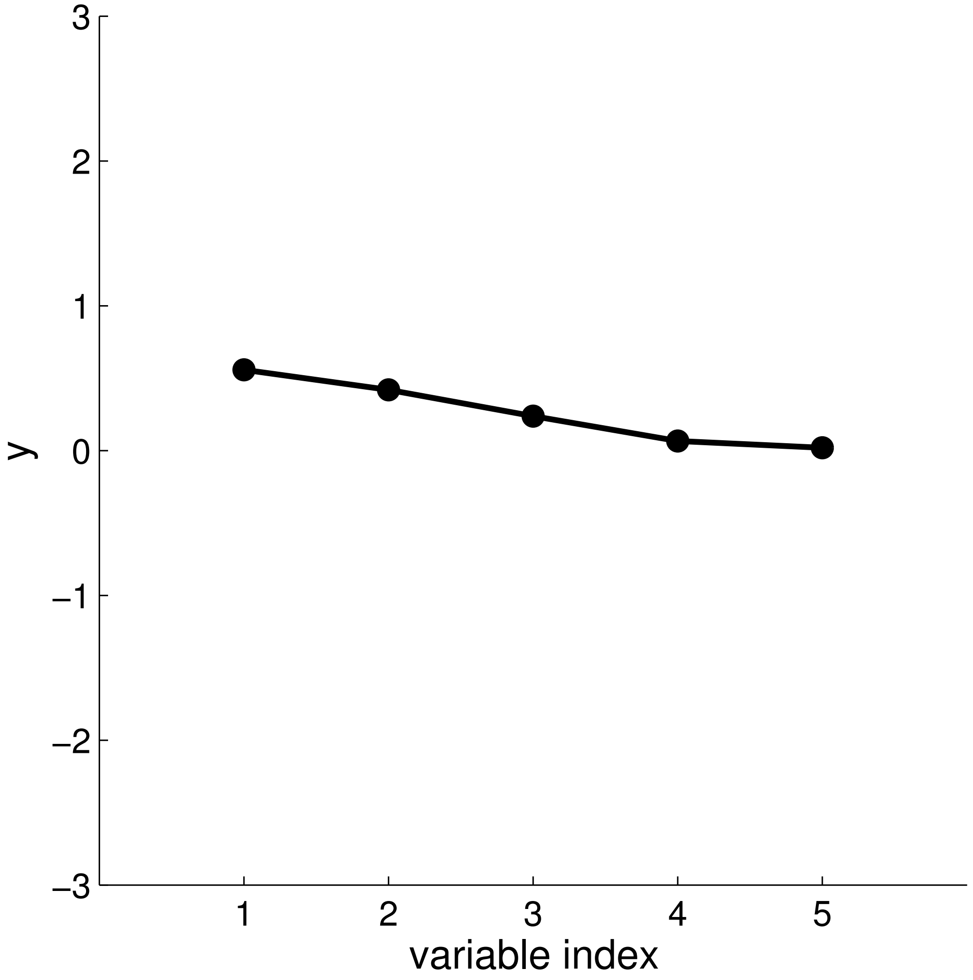
$$\Sigma = \begin{bmatrix} 1 & .9 & .8 & .6 & .4 \\ .9 & 1 & .9 & .8 & .6 \\ .8 & .9 & 1 & .9 & .8 \\ .6 & .8 & .9 & 1 & .9 \\ .4 & .6 & .8 & .9 & 1 \end{bmatrix}$$

- Special covariance matrix: correlations fall off the further the indices of the variables!

New visualisation

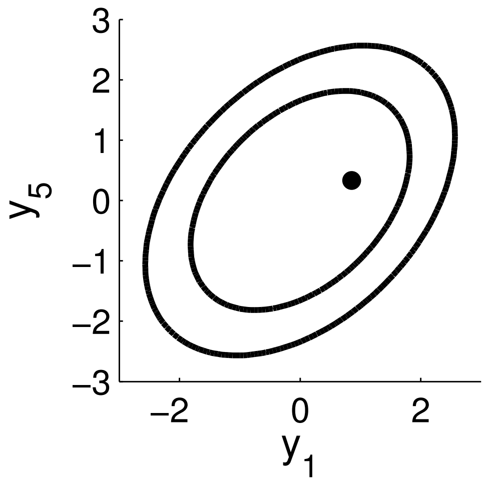


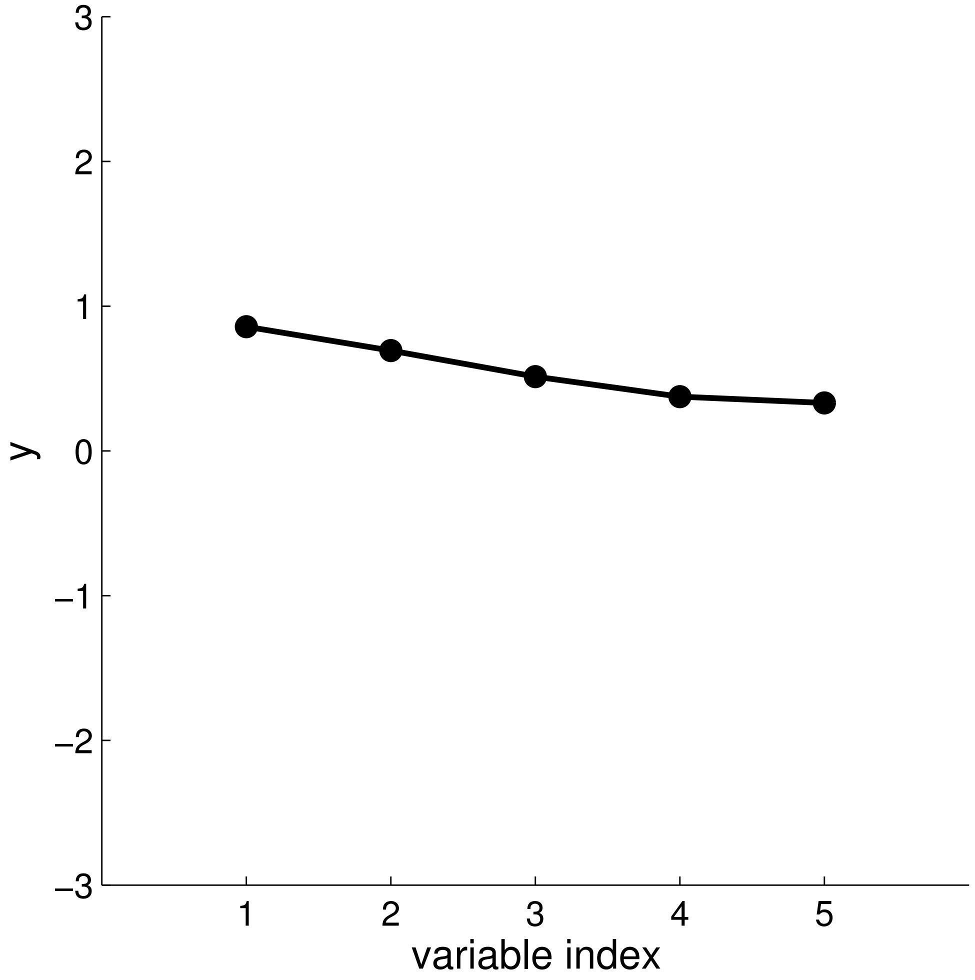
$$\Sigma = \begin{bmatrix} 1 & .9 & .8 & .6 & .4 \\ .9 & 1 & .9 & .8 & .6 \\ .8 & .9 & 1 & .9 & .8 \\ .6 & .8 & .9 & 1 & .9 \\ .4 & .6 & .8 & .9 & 1 \end{bmatrix}$$



- Special covariance matrix: correlations fall off the further the indices of the variables!

New visualisation

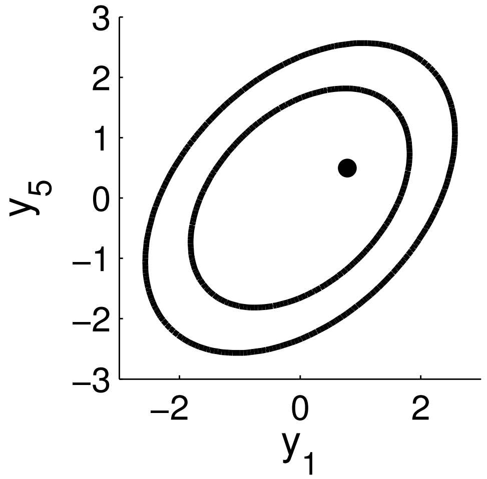


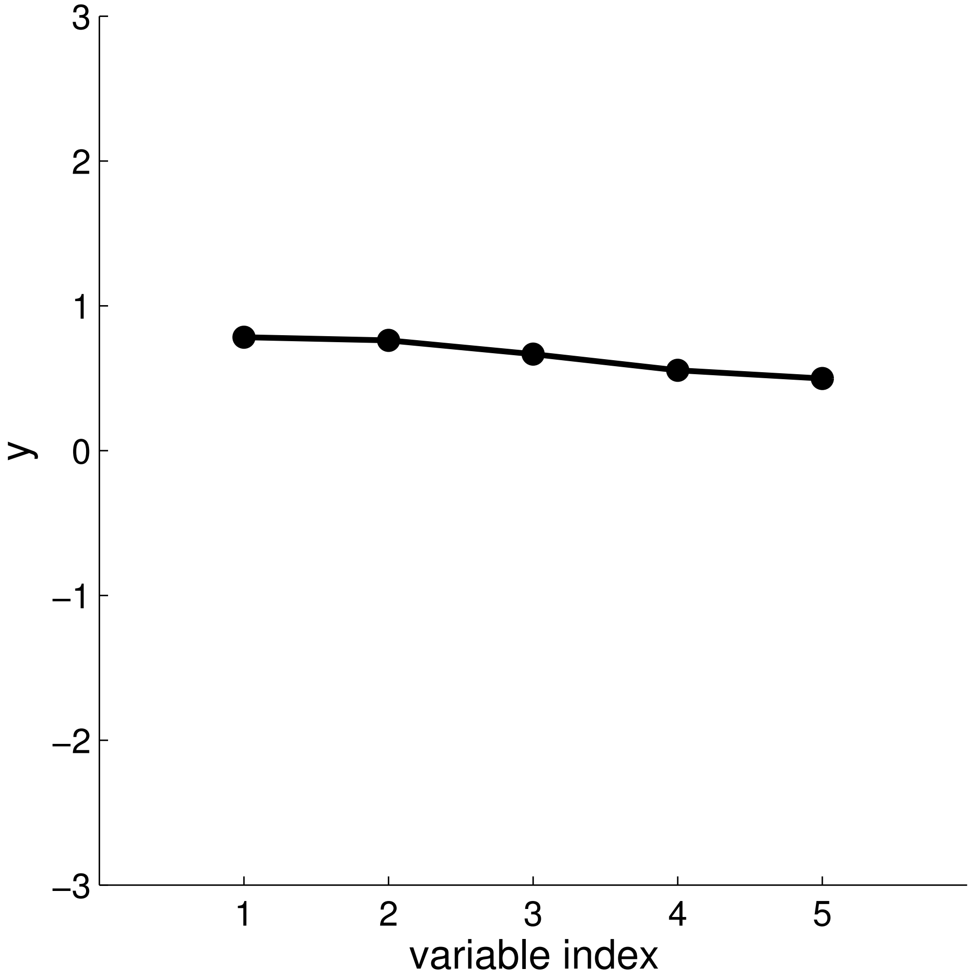


$$\Sigma = \begin{bmatrix} 1 & .9 & .8 & .6 & .4 \\ .9 & 1 & .9 & .8 & .6 \\ .8 & .9 & 1 & .9 & .8 \\ .6 & .8 & .9 & 1 & .9 \\ .4 & .6 & .8 & .9 & 1 \end{bmatrix}$$

- Special covariance matrix: correlations fall off the further the indices of the variables!

New visualisation

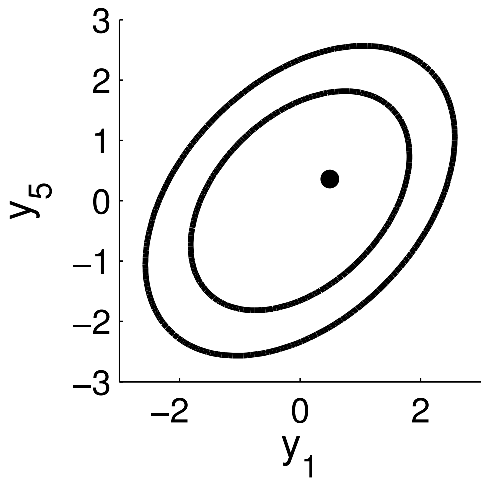


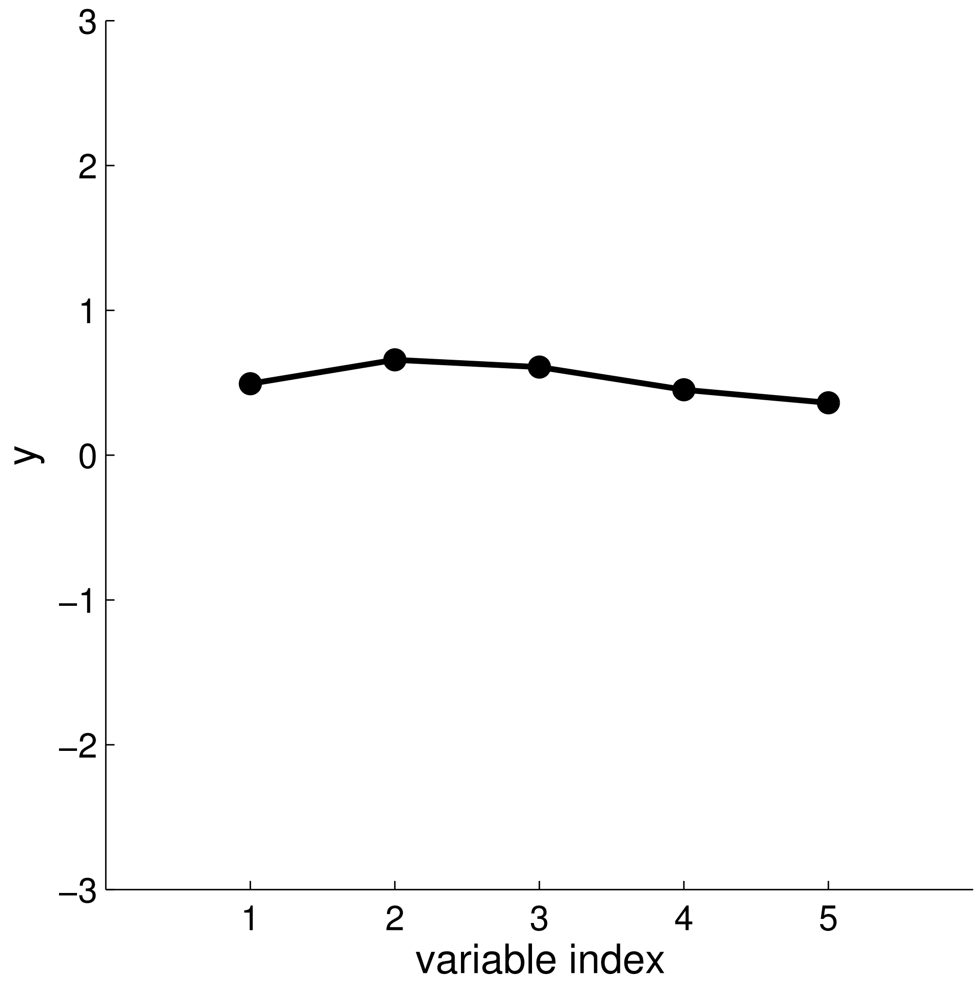


$$\Sigma = \begin{bmatrix} 1 & .9 & .8 & .6 & .4 \\ .9 & 1 & .9 & .8 & .6 \\ .8 & .9 & 1 & .9 & .8 \\ .6 & .8 & .9 & 1 & .9 \\ .4 & .6 & .8 & .9 & 1 \end{bmatrix}$$

- Special covariance matrix: correlations fall off the further the indices of the variables!

New visualisation

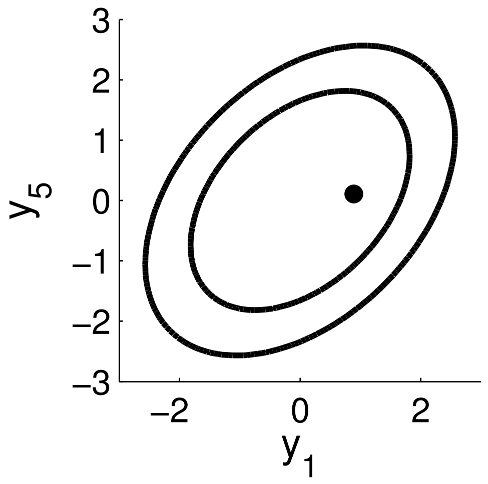


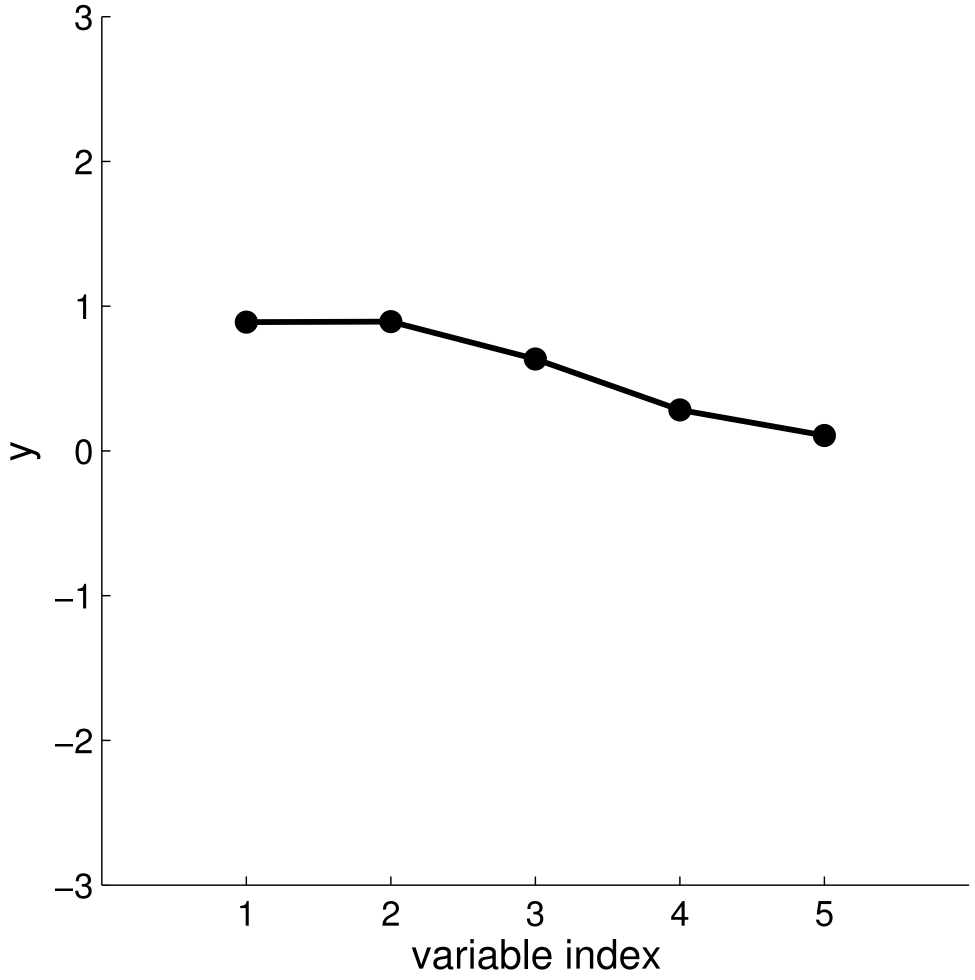


$$\Sigma = \begin{bmatrix} 1 & .9 & .8 & .6 & .4 \\ .9 & 1 & .9 & .8 & .6 \\ .8 & .9 & 1 & .9 & .8 \\ .6 & .8 & .9 & 1 & .9 \\ .4 & .6 & .8 & .9 & 1 \end{bmatrix}$$

- Special covariance matrix: correlations fall off the further the indices of the variables!

New visualisation

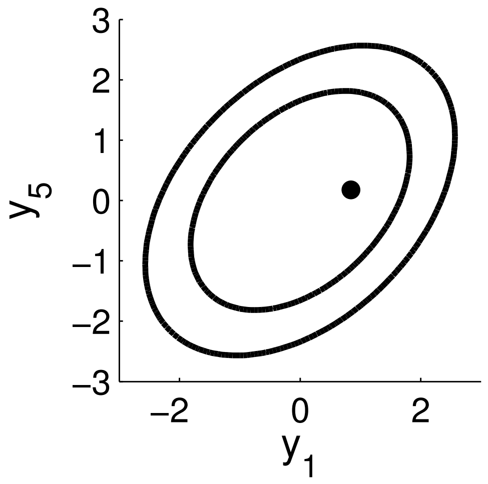




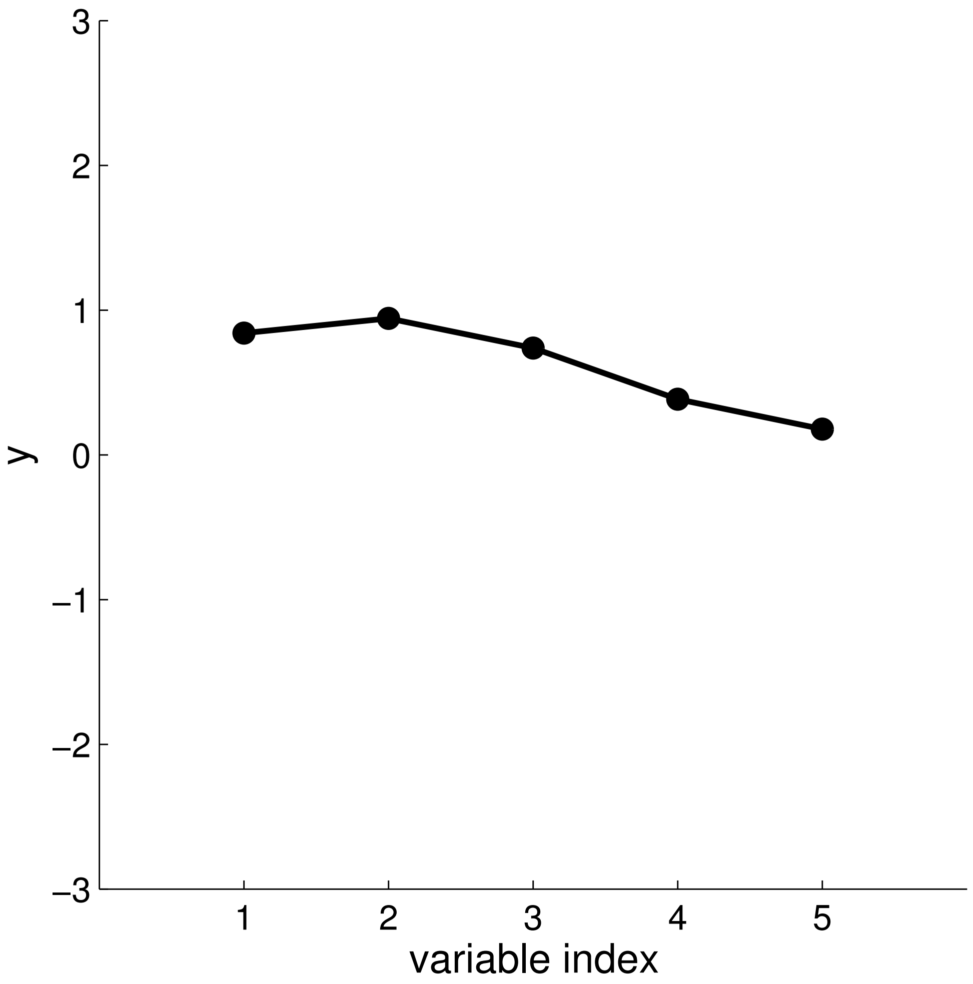
$$\Sigma = \begin{bmatrix} 1 & .9 & .8 & .6 & .4 \\ .9 & 1 & .9 & .8 & .6 \\ .8 & .9 & 1 & .9 & .8 \\ .6 & .8 & .9 & 1 & .9 \\ .4 & .6 & .8 & .9 & 1 \end{bmatrix}$$

- Special covariance matrix: correlations fall off the further the indices of the variables!

New visualisation

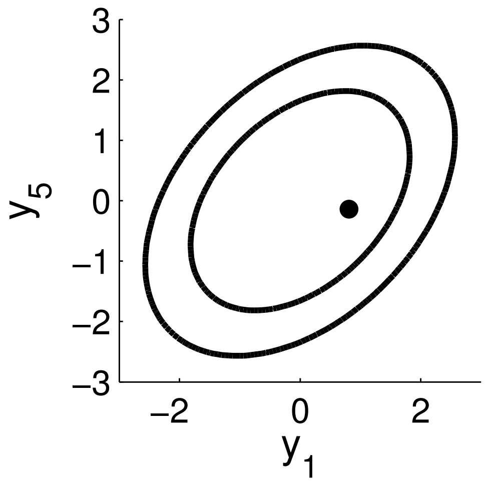


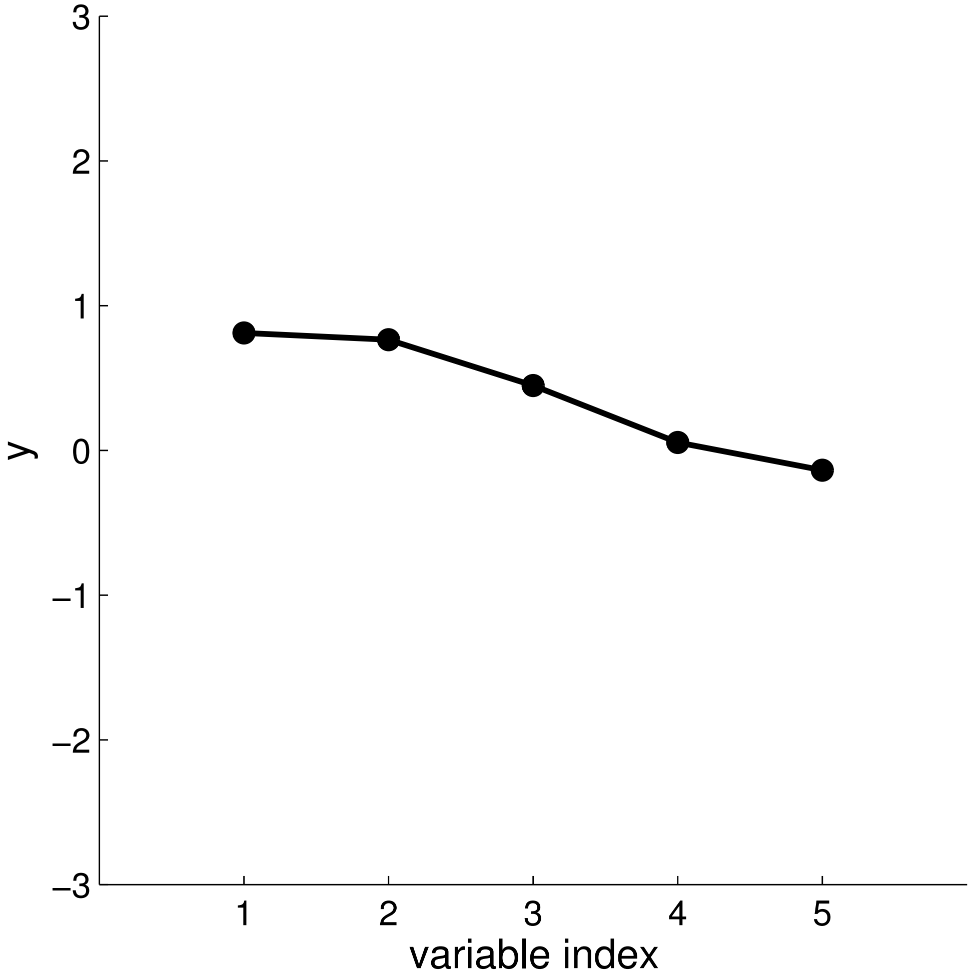
$$\Sigma = \begin{bmatrix} 1 & .9 & .8 & .6 & .4 \\ .9 & 1 & .9 & .8 & .6 \\ .8 & .9 & 1 & .9 & .8 \\ .6 & .8 & .9 & 1 & .9 \\ .4 & .6 & .8 & .9 & 1 \end{bmatrix}$$



- Special covariance matrix: correlations fall off the further the indices of the variables!

New visualisation

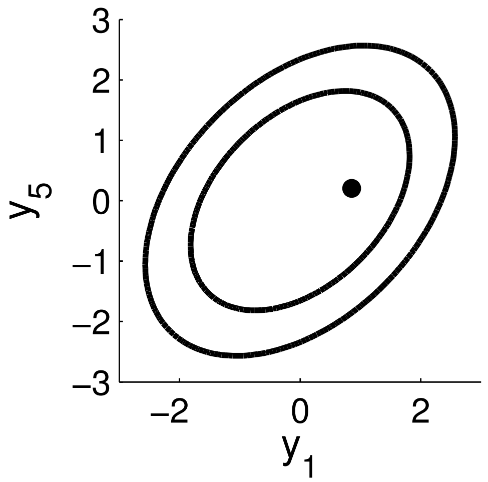




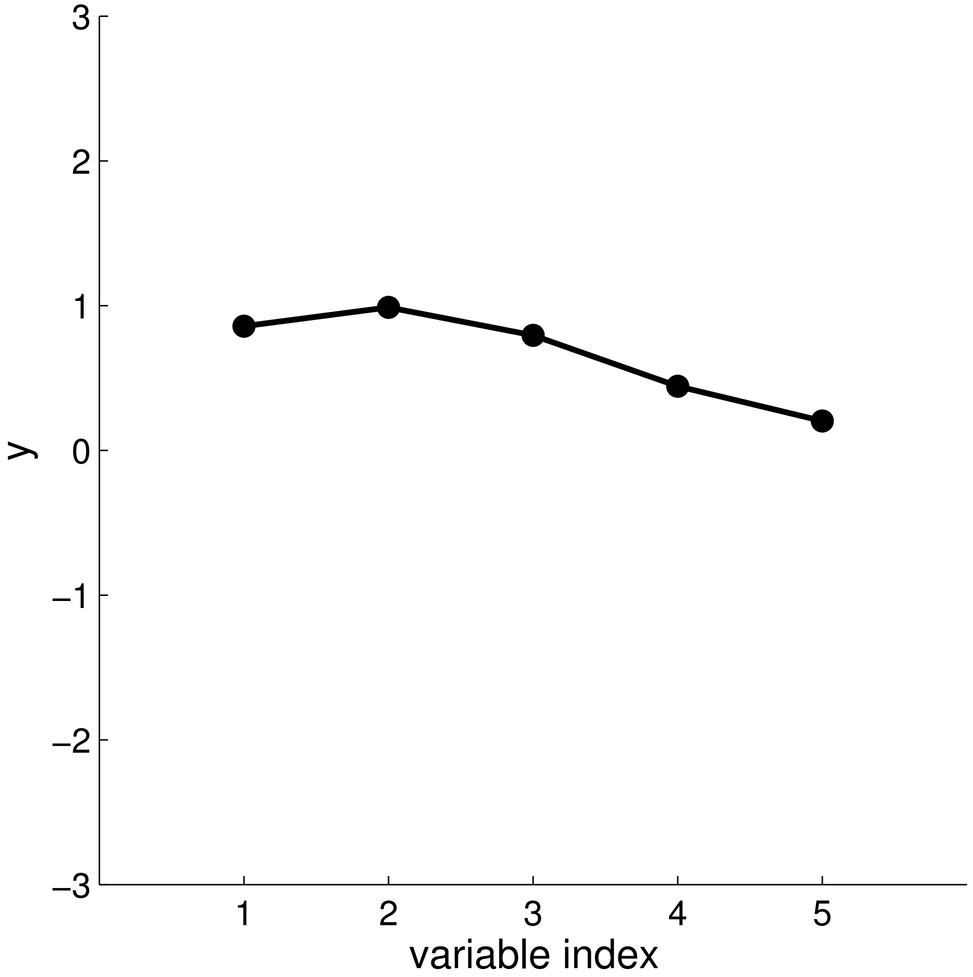
$$\Sigma = \begin{bmatrix} 1 & .9 & .8 & .6 & .4 \\ .9 & 1 & .9 & .8 & .6 \\ .8 & .9 & 1 & .9 & .8 \\ .6 & .8 & .9 & 1 & .9 \\ .4 & .6 & .8 & .9 & 1 \end{bmatrix}$$

- Special covariance matrix: correlations fall off the further the indices of the variables!

New visualisation

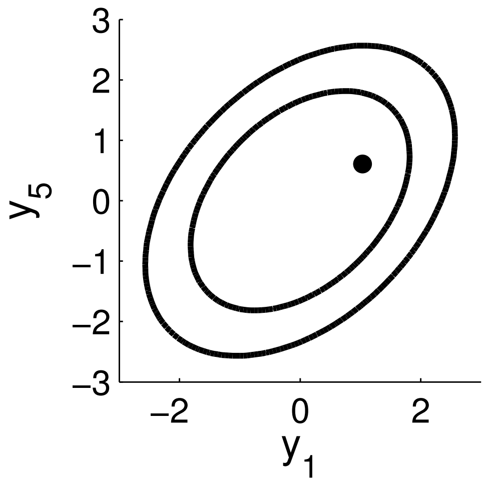


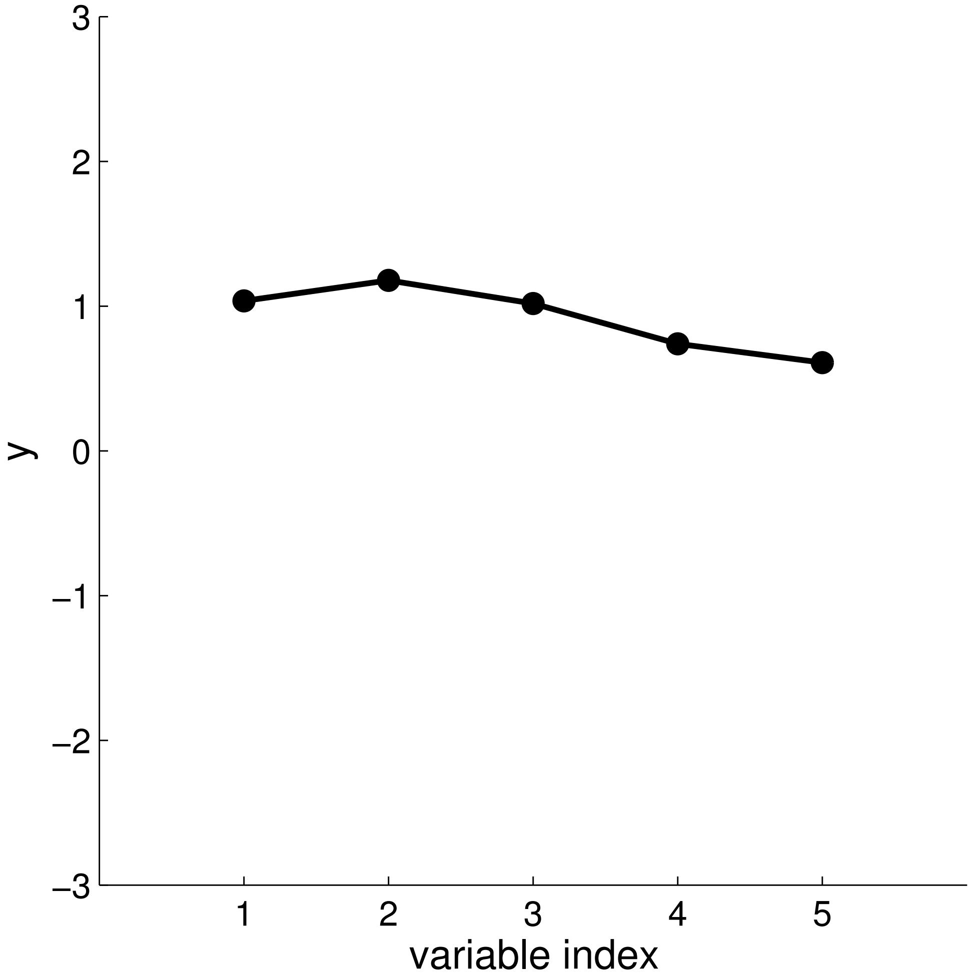
$$\Sigma = \begin{bmatrix} 1 & .9 & .8 & .6 & .4 \\ .9 & 1 & .9 & .8 & .6 \\ .8 & .9 & 1 & .9 & .8 \\ .6 & .8 & .9 & 1 & .9 \\ .4 & .6 & .8 & .9 & 1 \end{bmatrix}$$



- Special covariance matrix: correlations fall off the further the indices of the variables!

New visualisation

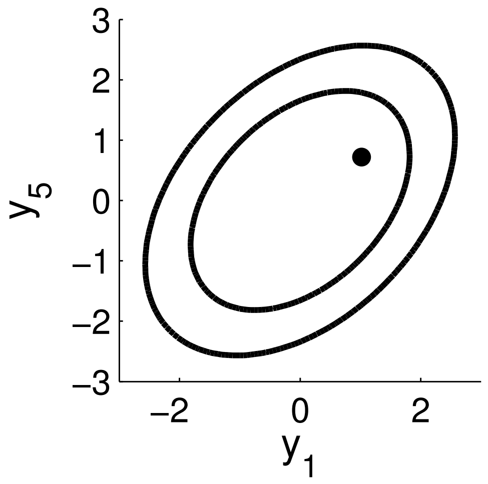




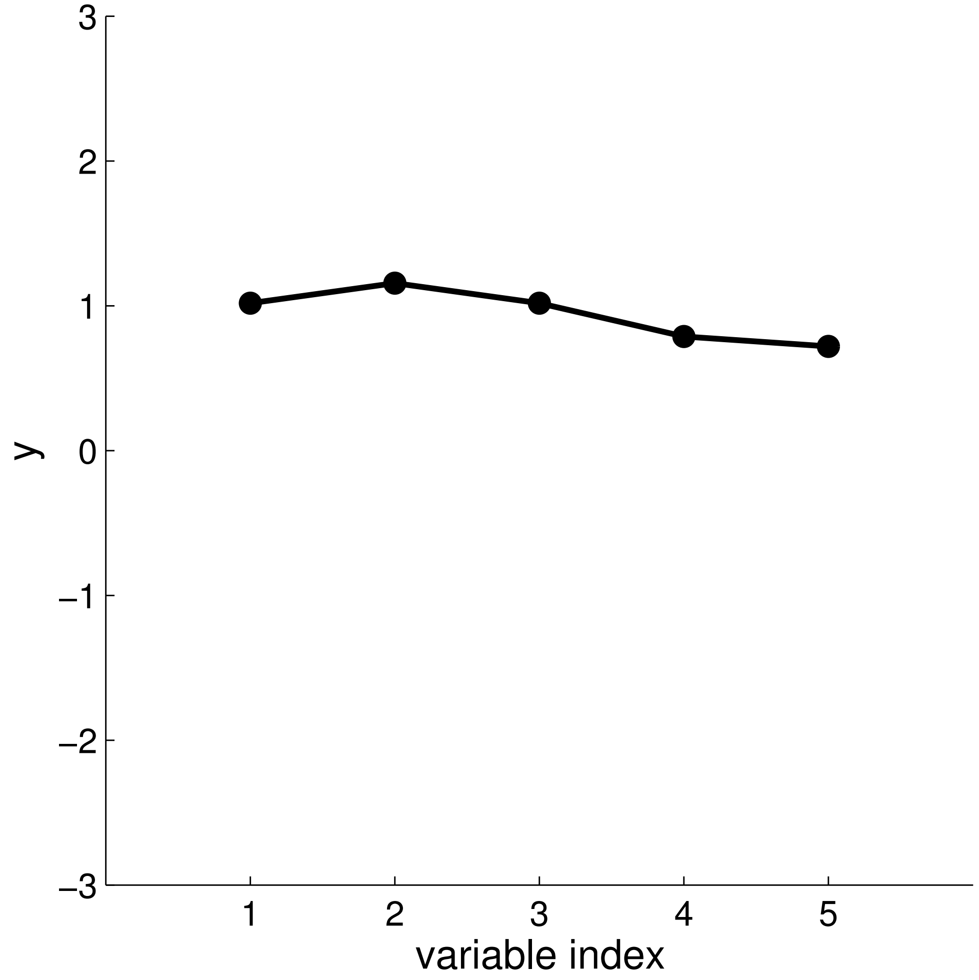
$$\Sigma = \begin{bmatrix} 1 & .9 & .8 & .6 & .4 \\ .9 & 1 & .9 & .8 & .6 \\ .8 & .9 & 1 & .9 & .8 \\ .6 & .8 & .9 & 1 & .9 \\ .4 & .6 & .8 & .9 & 1 \end{bmatrix}$$

- Special covariance matrix: correlations fall off the further the indices of the variables!

New visualisation

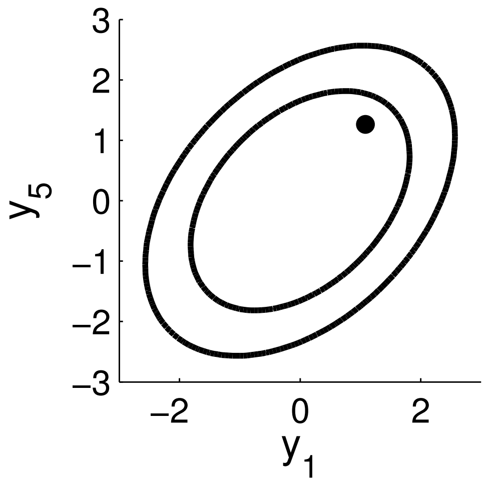


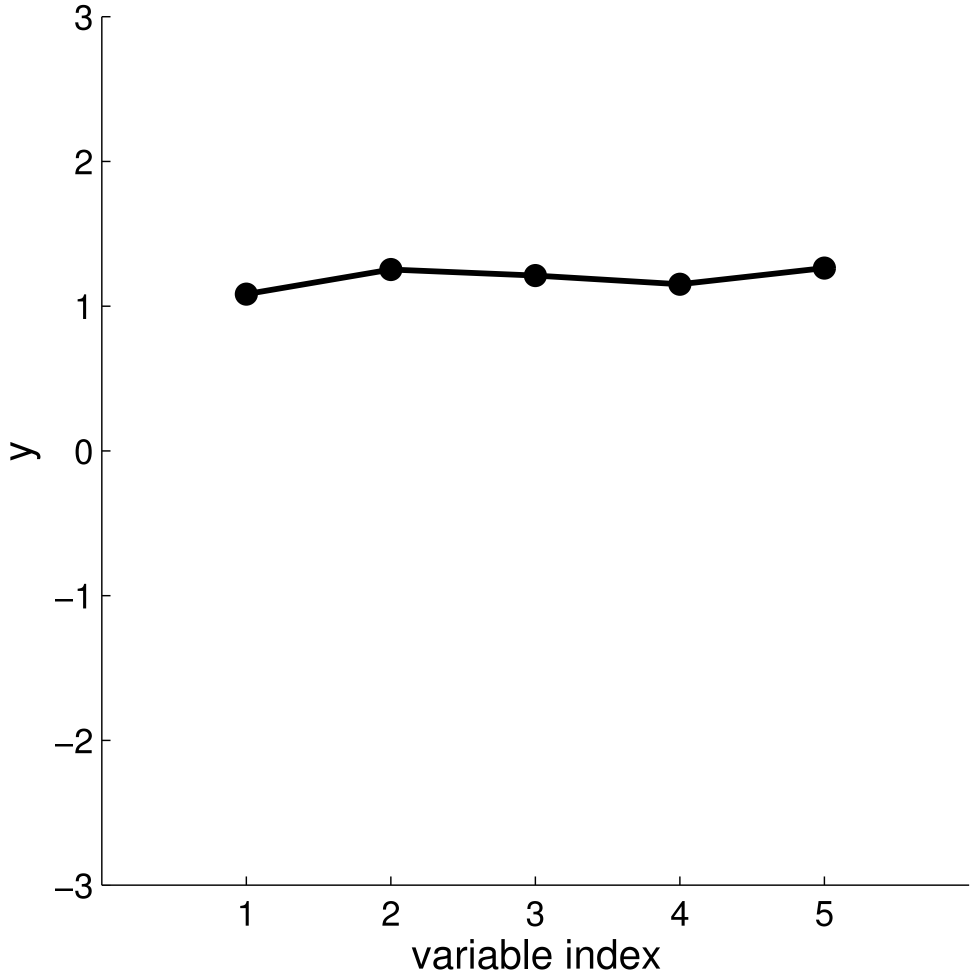
$$\Sigma = \begin{bmatrix} 1 & .9 & .8 & .6 & .4 \\ .9 & 1 & .9 & .8 & .6 \\ .8 & .9 & 1 & .9 & .8 \\ .6 & .8 & .9 & 1 & .9 \\ .4 & .6 & .8 & .9 & 1 \end{bmatrix}$$



- Special covariance matrix: correlations fall off the further the indices of the variables!

New visualisation

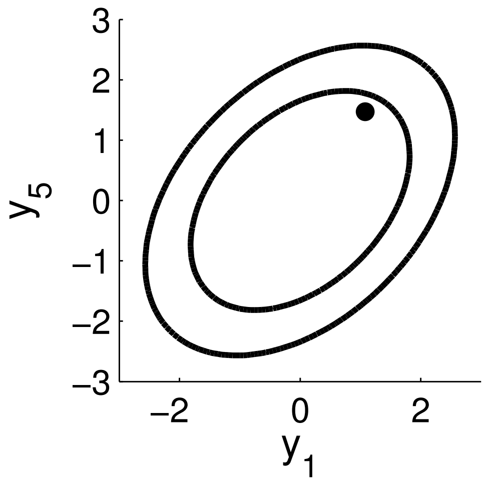


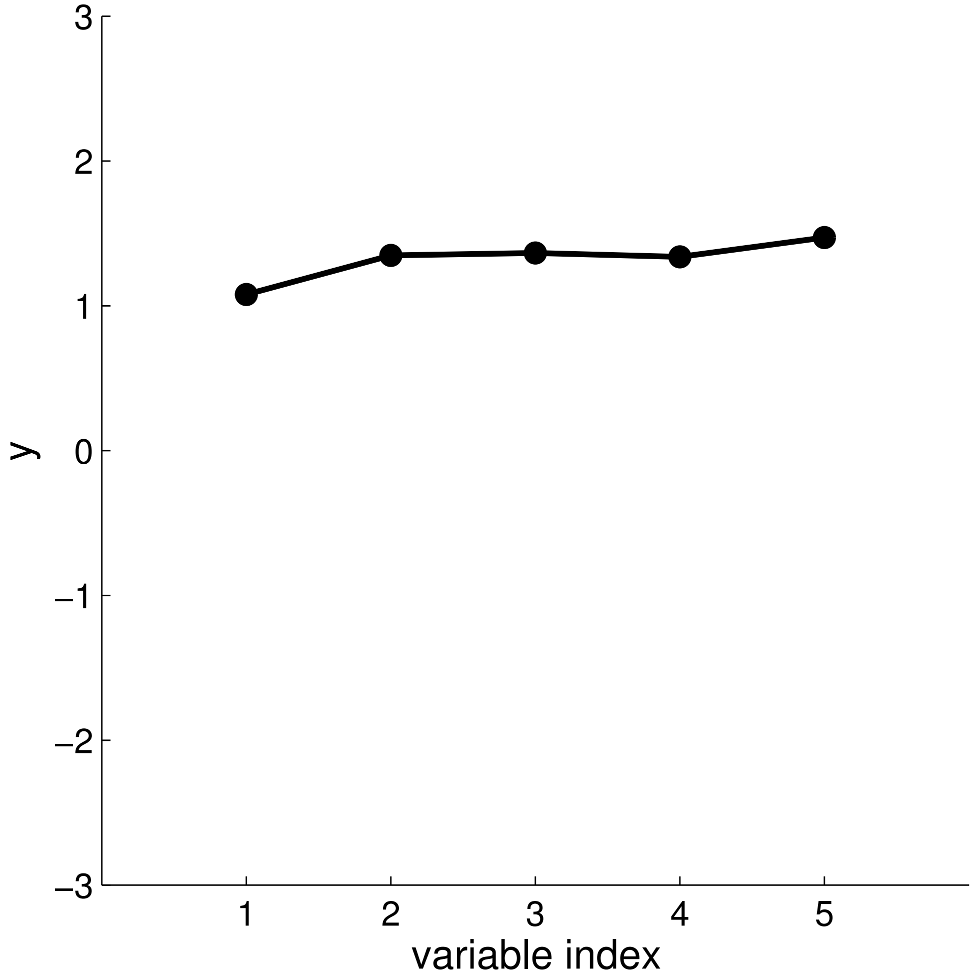


$$\Sigma = \begin{bmatrix} 1 & .9 & .8 & .6 & .4 \\ .9 & 1 & .9 & .8 & .6 \\ .8 & .9 & 1 & .9 & .8 \\ .6 & .8 & .9 & 1 & .9 \\ .4 & .6 & .8 & .9 & 1 \end{bmatrix}$$

- Special covariance matrix: correlations fall off the further the indices of the variables!

New visualisation

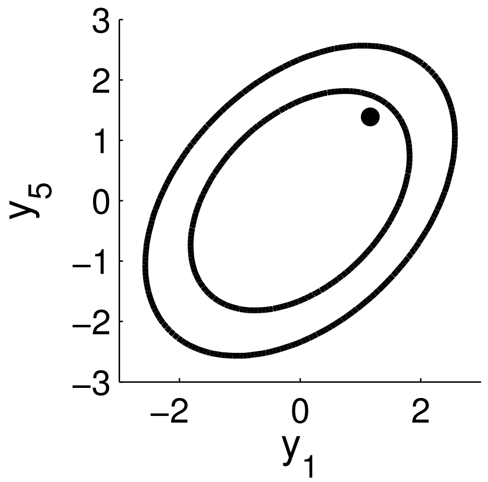


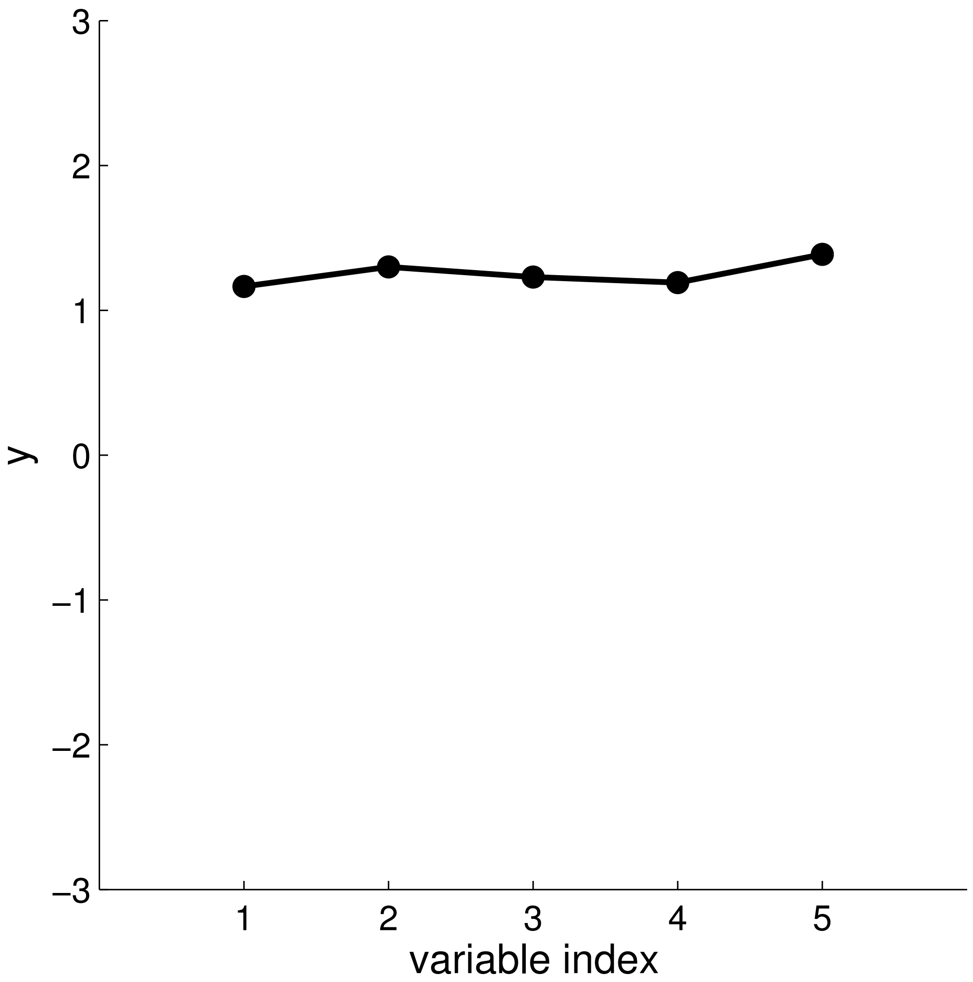


$$\Sigma = \begin{bmatrix} 1 & .9 & .8 & .6 & .4 \\ .9 & 1 & .9 & .8 & .6 \\ .8 & .9 & 1 & .9 & .8 \\ .6 & .8 & .9 & 1 & .9 \\ .4 & .6 & .8 & .9 & 1 \end{bmatrix}$$

- Special covariance matrix: correlations fall off the further the indices of the variables!

New visualisation

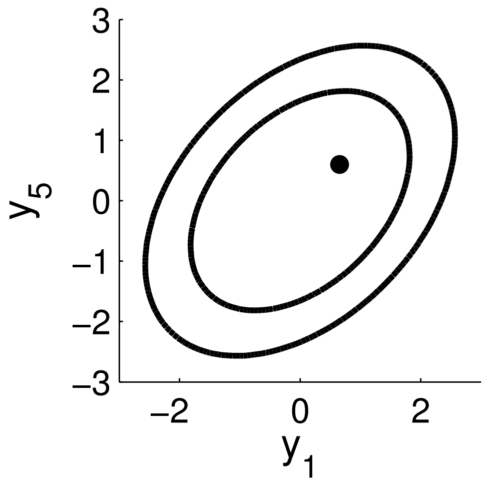


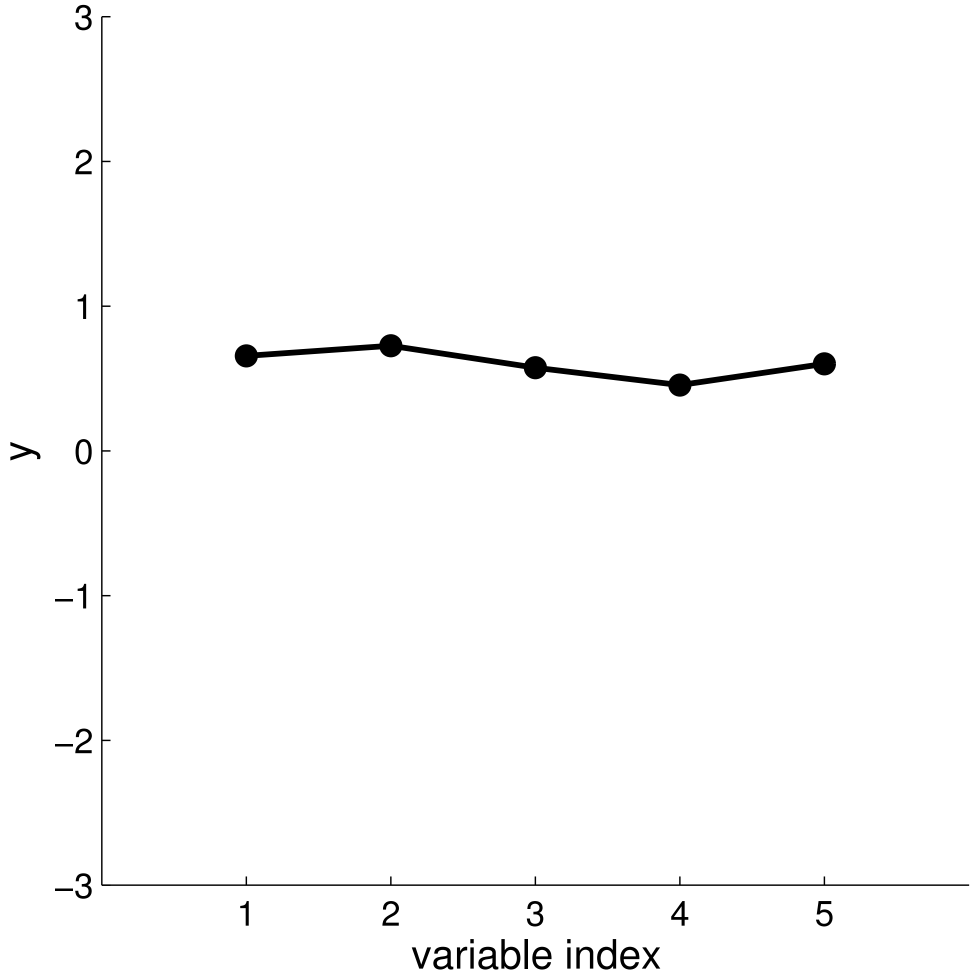


$$\Sigma = \begin{bmatrix} 1 & .9 & .8 & .6 & .4 \\ .9 & 1 & .9 & .8 & .6 \\ .8 & .9 & 1 & .9 & .8 \\ .6 & .8 & .9 & 1 & .9 \\ .4 & .6 & .8 & .9 & 1 \end{bmatrix}$$

- Special covariance matrix: correlations fall off the further the indices of the variables!

New visualisation

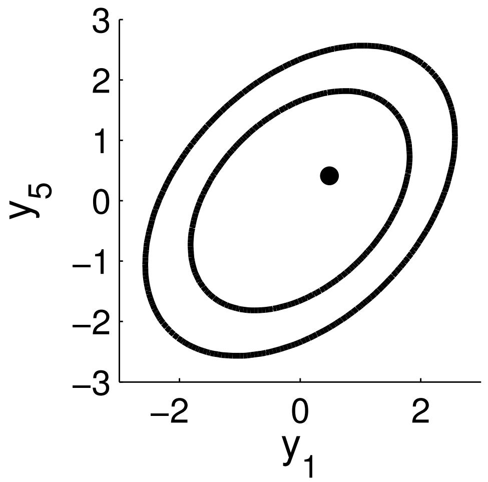


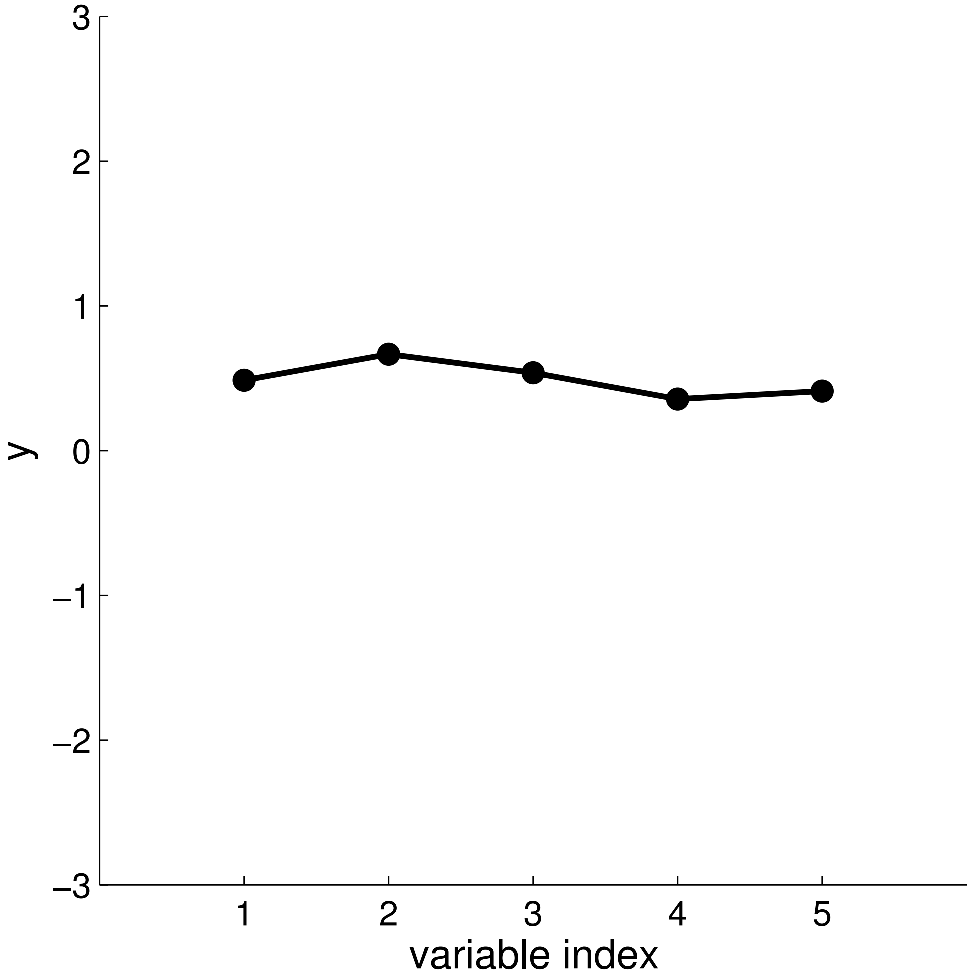


$$\Sigma = \begin{bmatrix} 1 & .9 & .8 & .6 & .4 \\ .9 & 1 & .9 & .8 & .6 \\ .8 & .9 & 1 & .9 & .8 \\ .6 & .8 & .9 & 1 & .9 \\ .4 & .6 & .8 & .9 & 1 \end{bmatrix}$$

- Special covariance matrix: correlations fall off the further the indices of the variables!

New visualisation

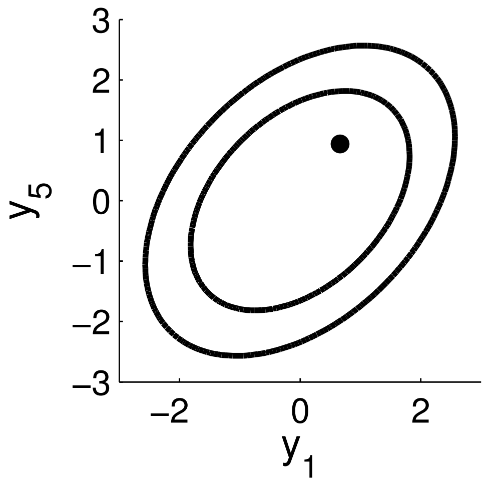


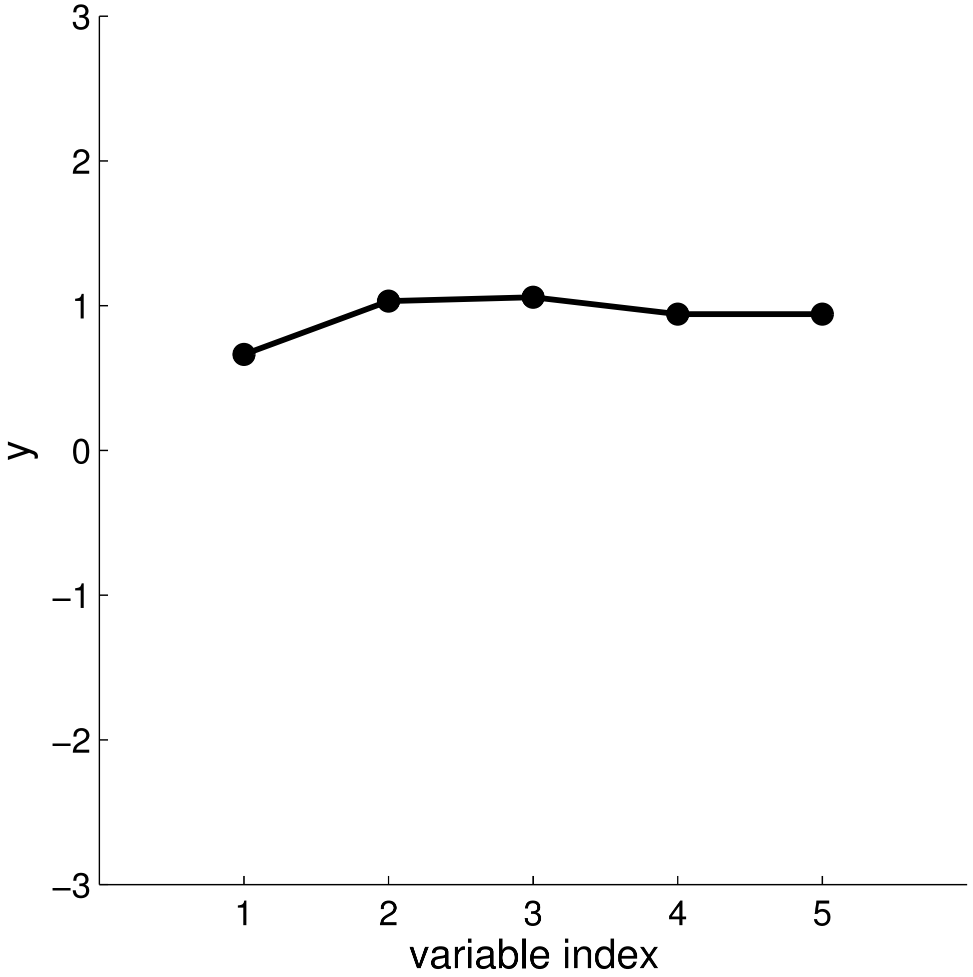


$$\Sigma = \begin{bmatrix} 1 & .9 & .8 & .6 & .4 \\ .9 & 1 & .9 & .8 & .6 \\ .8 & .9 & 1 & .9 & .8 \\ .6 & .8 & .9 & 1 & .9 \\ .4 & .6 & .8 & .9 & 1 \end{bmatrix}$$

- Special covariance matrix: correlations fall off the further the indices of the variables!

New visualisation

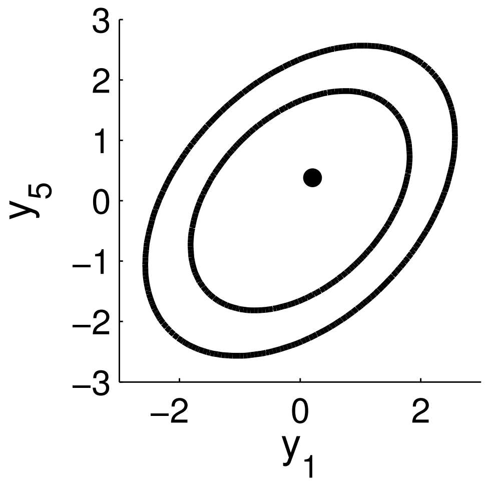


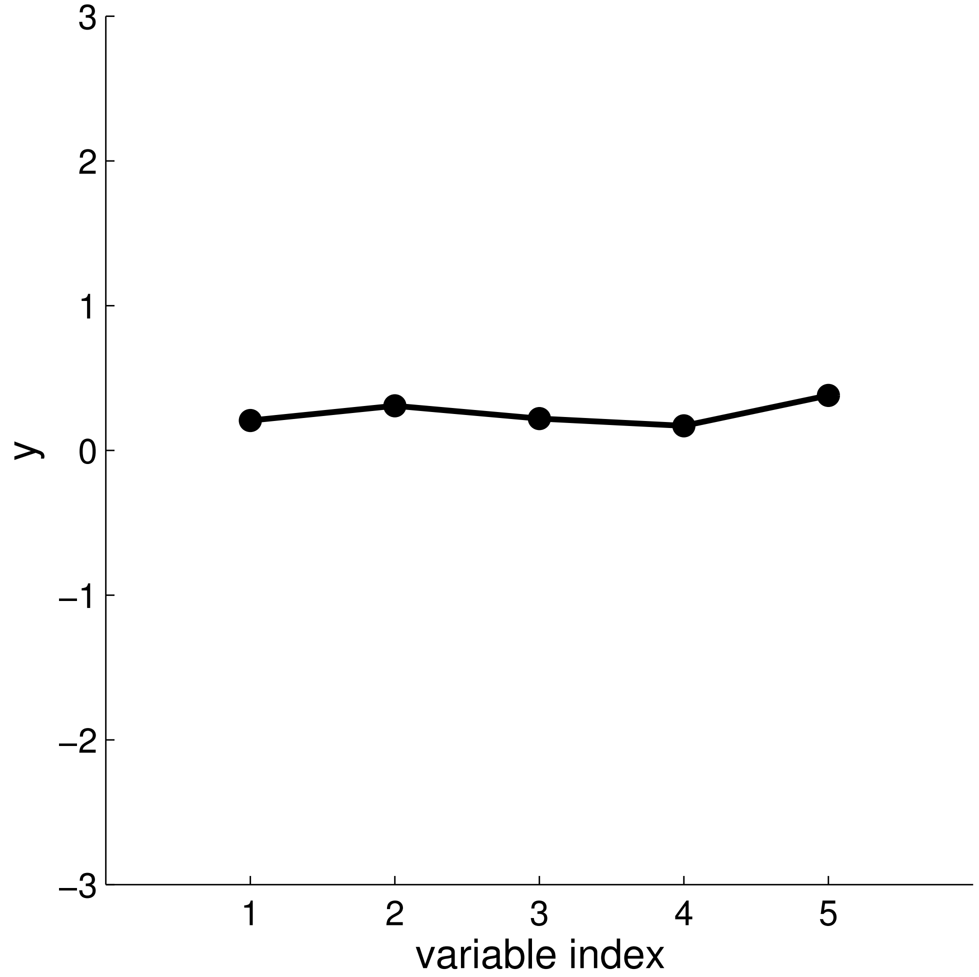


$$\Sigma = \begin{bmatrix} 1 & .9 & .8 & .6 & .4 \\ .9 & 1 & .9 & .8 & .6 \\ .8 & .9 & 1 & .9 & .8 \\ .6 & .8 & .9 & 1 & .9 \\ .4 & .6 & .8 & .9 & 1 \end{bmatrix}$$

- Special covariance matrix: correlations fall off the further the indices of the variables!

New visualisation

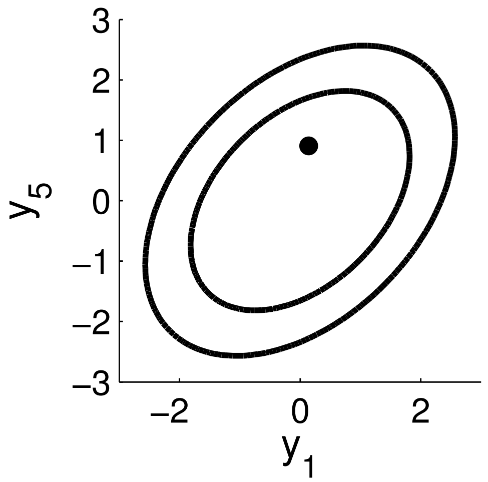


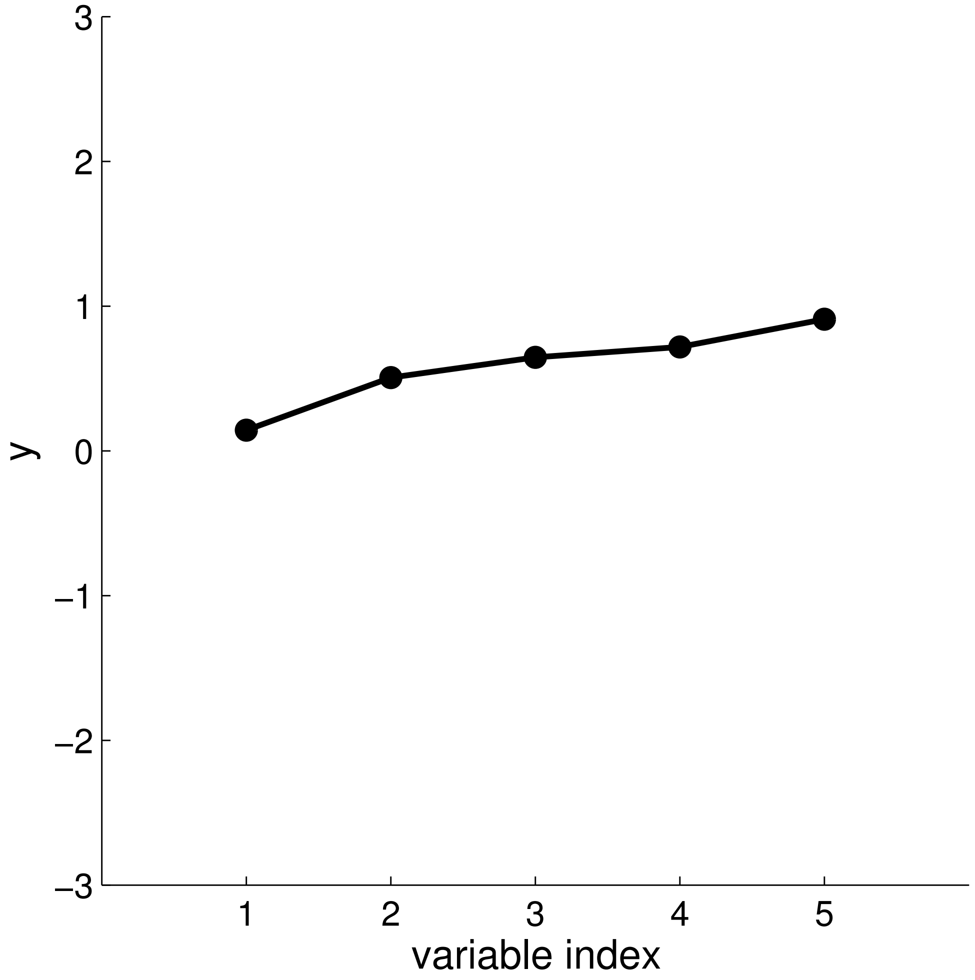


$$\Sigma = \begin{bmatrix} 1 & .9 & .8 & .6 & .4 \\ .9 & 1 & .9 & .8 & .6 \\ .8 & .9 & 1 & .9 & .8 \\ .6 & .8 & .9 & 1 & .9 \\ .4 & .6 & .8 & .9 & 1 \end{bmatrix}$$

- Special covariance matrix: correlations fall off the further the indices of the variables!

New visualisation

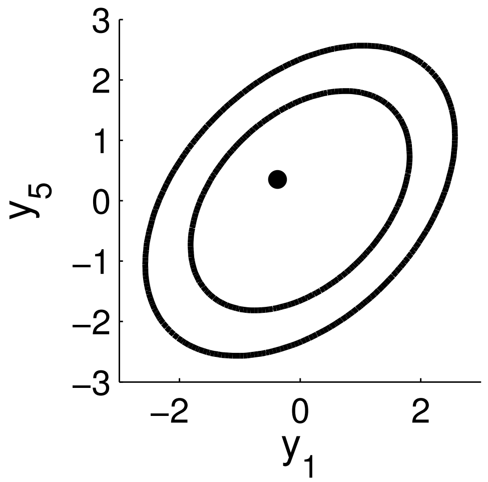




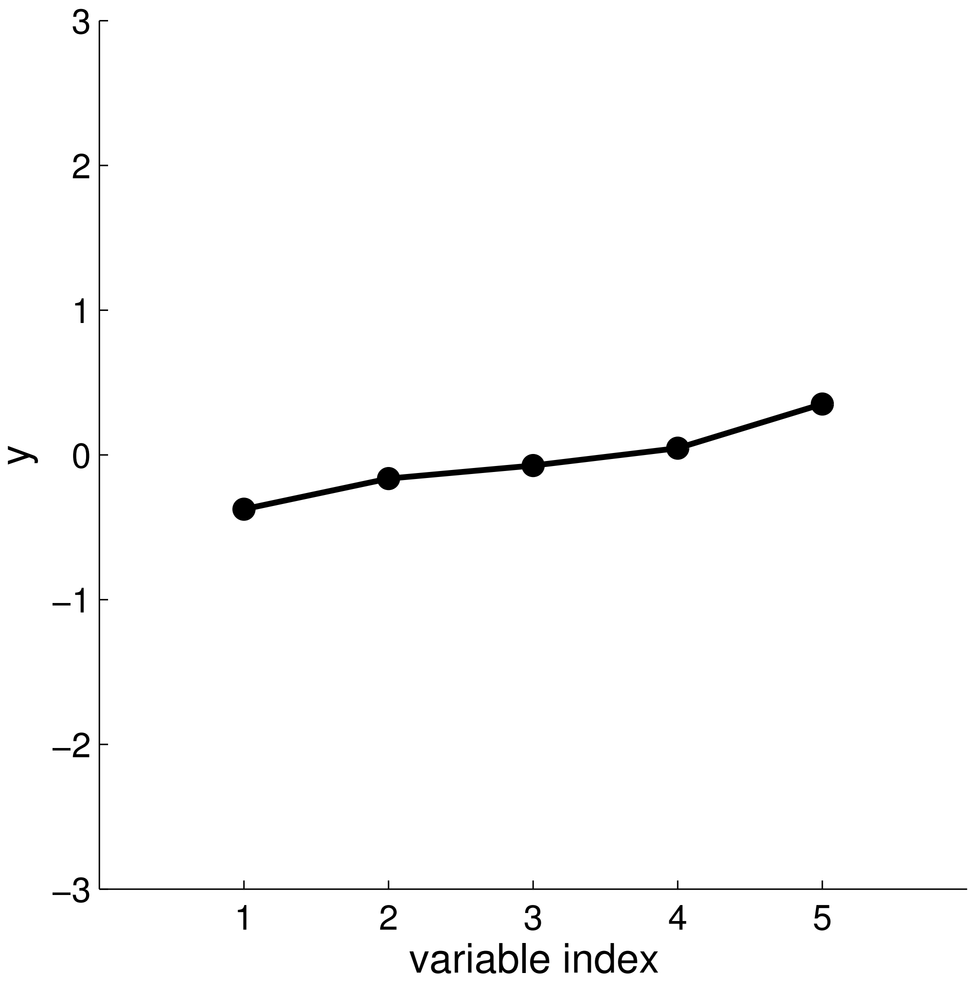
$$\Sigma = \begin{bmatrix} 1 & .9 & .8 & .6 & .4 \\ .9 & 1 & .9 & .8 & .6 \\ .8 & .9 & 1 & .9 & .8 \\ .6 & .8 & .9 & 1 & .9 \\ .4 & .6 & .8 & .9 & 1 \end{bmatrix}$$

- Special covariance matrix: correlations fall off the further the indices of the variables!

New visualisation

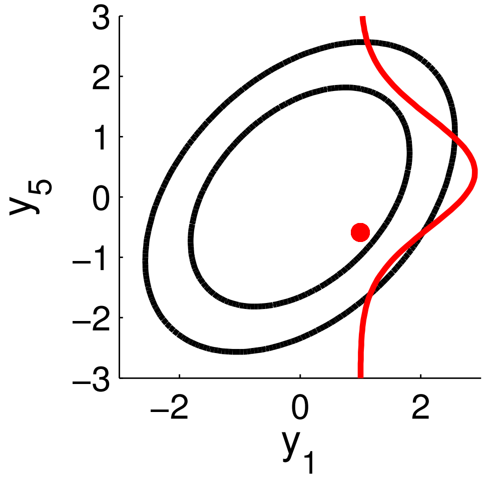


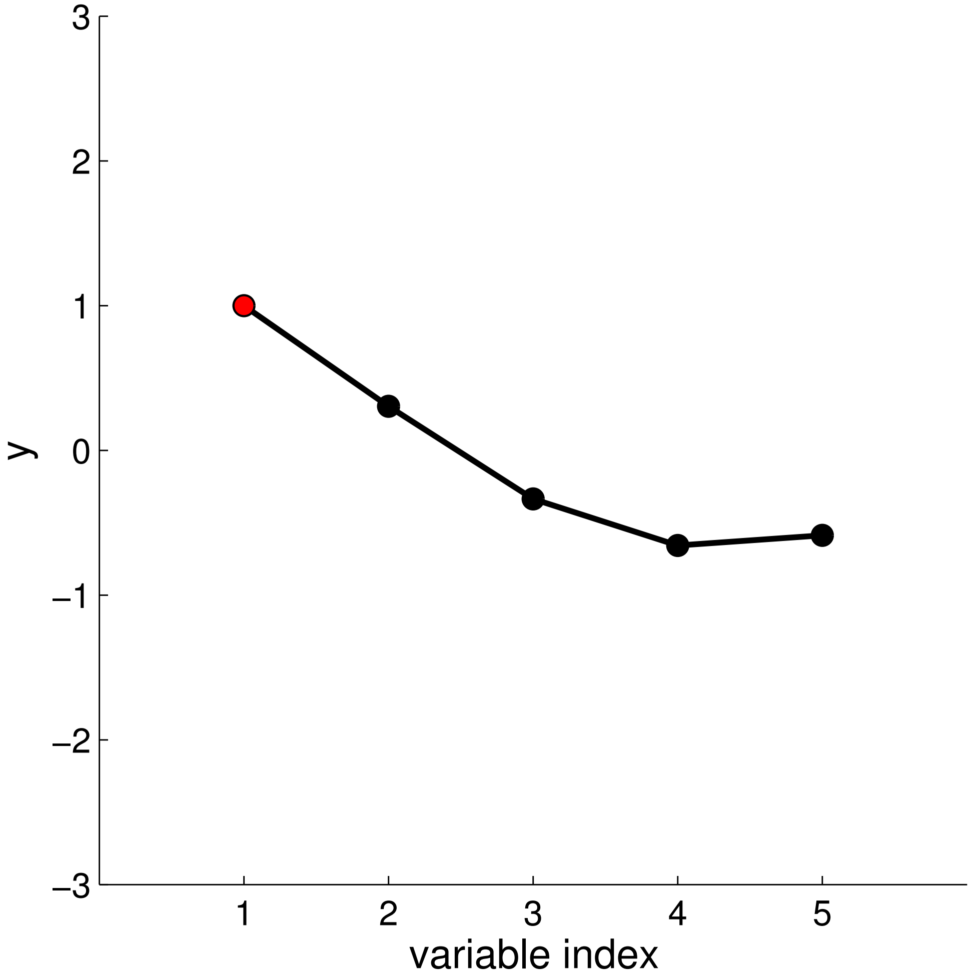
$$\Sigma = \begin{bmatrix} 1 & .9 & .8 & .6 & .4 \\ .9 & 1 & .9 & .8 & .6 \\ .8 & .9 & 1 & .9 & .8 \\ .6 & .8 & .9 & 1 & .9 \\ .4 & .6 & .8 & .9 & 1 \end{bmatrix}$$



- Special covariance matrix: correlations fall off the further the indices of the variables!

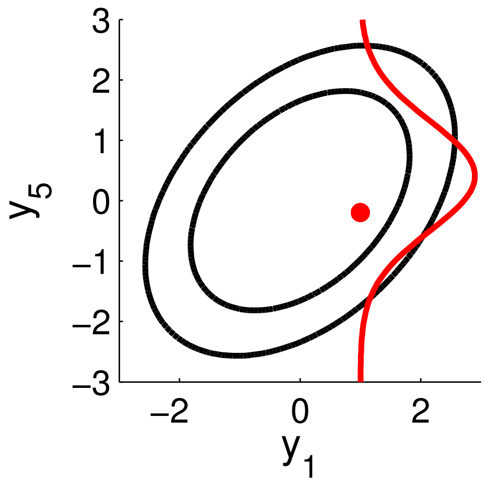
New visualisation

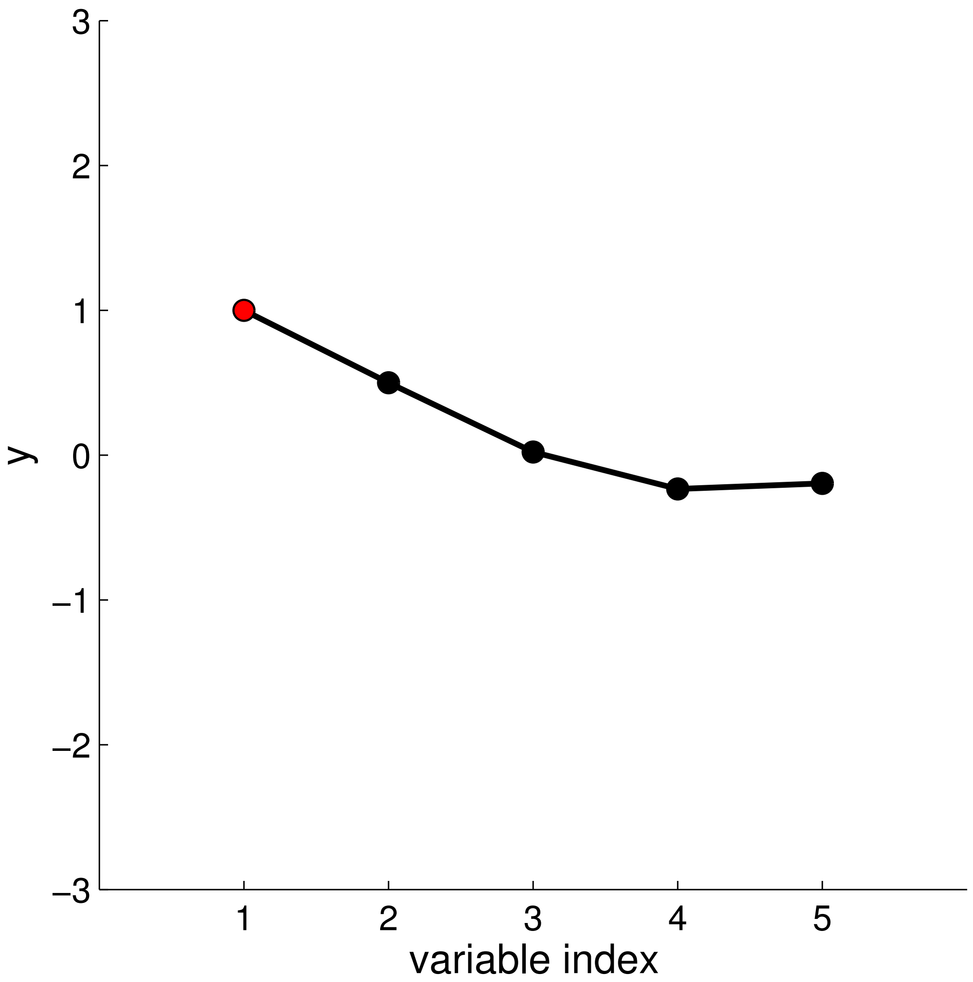




$$\Sigma = \begin{bmatrix} 1 & .9 & .8 & .6 & .4 \\ .9 & 1 & .9 & .8 & .6 \\ .8 & .9 & 1 & .9 & .8 \\ .6 & .8 & .9 & 1 & .9 \\ .4 & .6 & .8 & .9 & 1 \end{bmatrix}$$

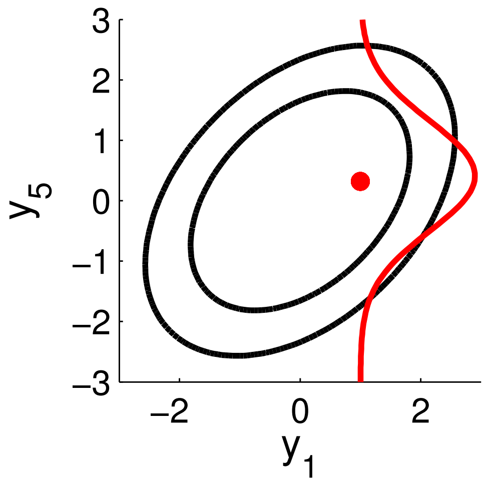
New visualisation

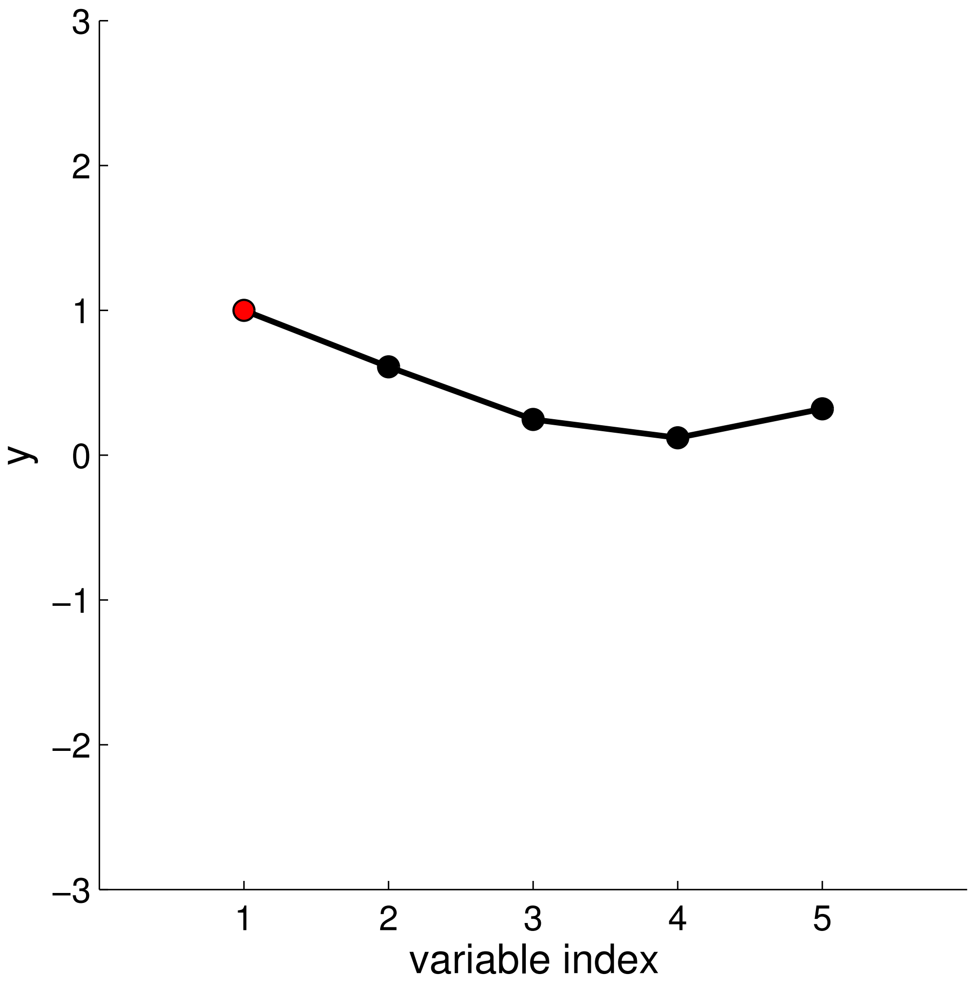




$$\Sigma = \begin{bmatrix} 1 & .9 & .8 & .6 & .4 \\ .9 & 1 & .9 & .8 & .6 \\ .8 & .9 & 1 & .9 & .8 \\ .6 & .8 & .9 & 1 & .9 \\ .4 & .6 & .8 & .9 & 1 \end{bmatrix}$$

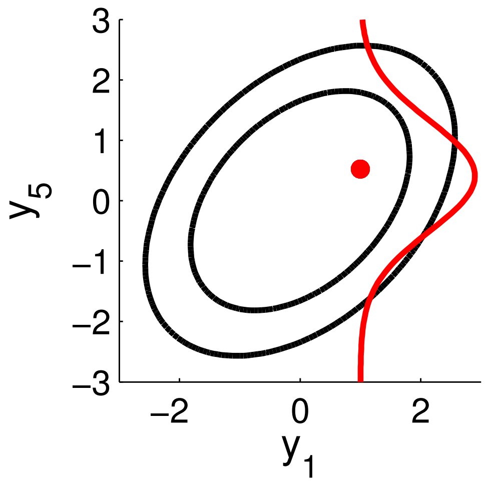
New visualisation

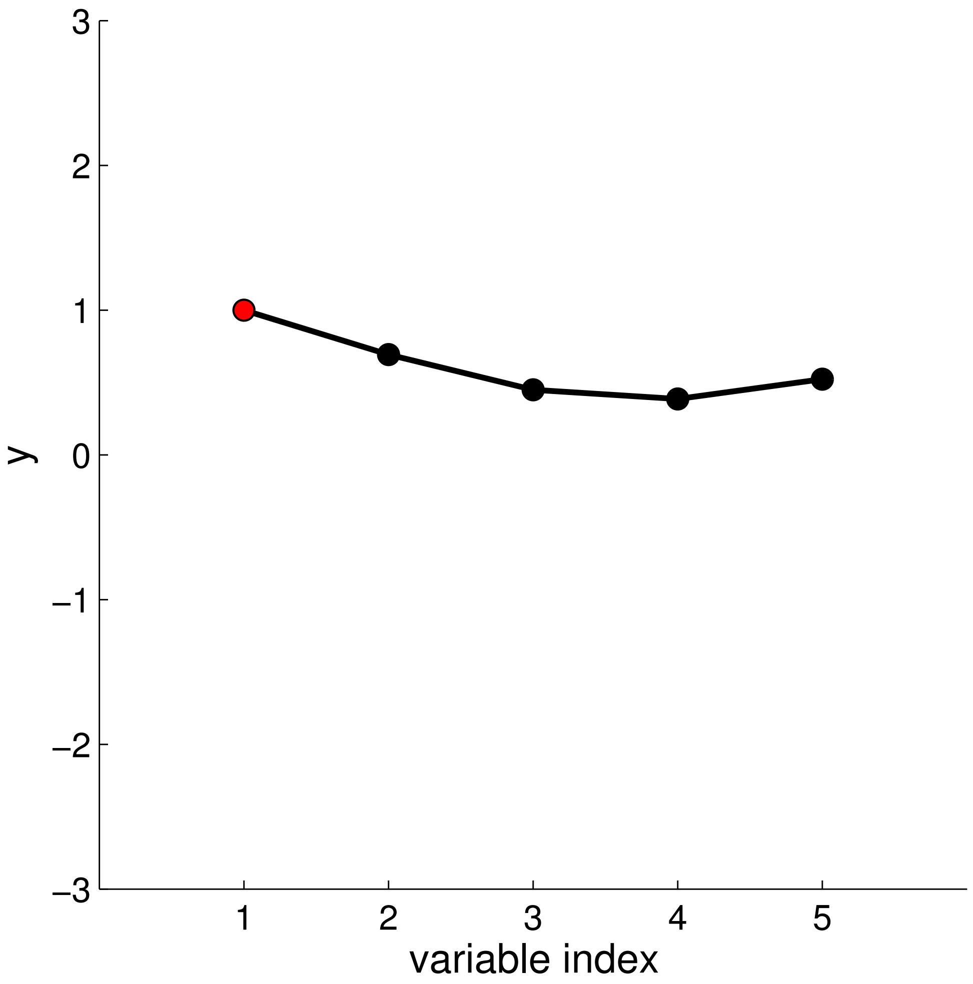




$$\Sigma = \begin{bmatrix} 1 & .9 & .8 & .6 & .4 \\ .9 & 1 & .9 & .8 & .6 \\ .8 & .9 & 1 & .9 & .8 \\ .6 & .8 & .9 & 1 & .9 \\ .4 & .6 & .8 & .9 & 1 \end{bmatrix}$$

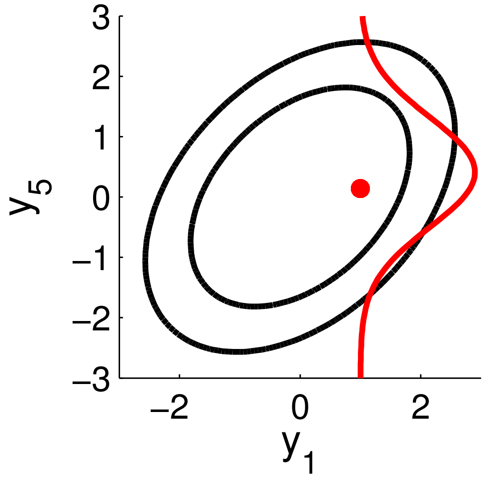
New visualisation

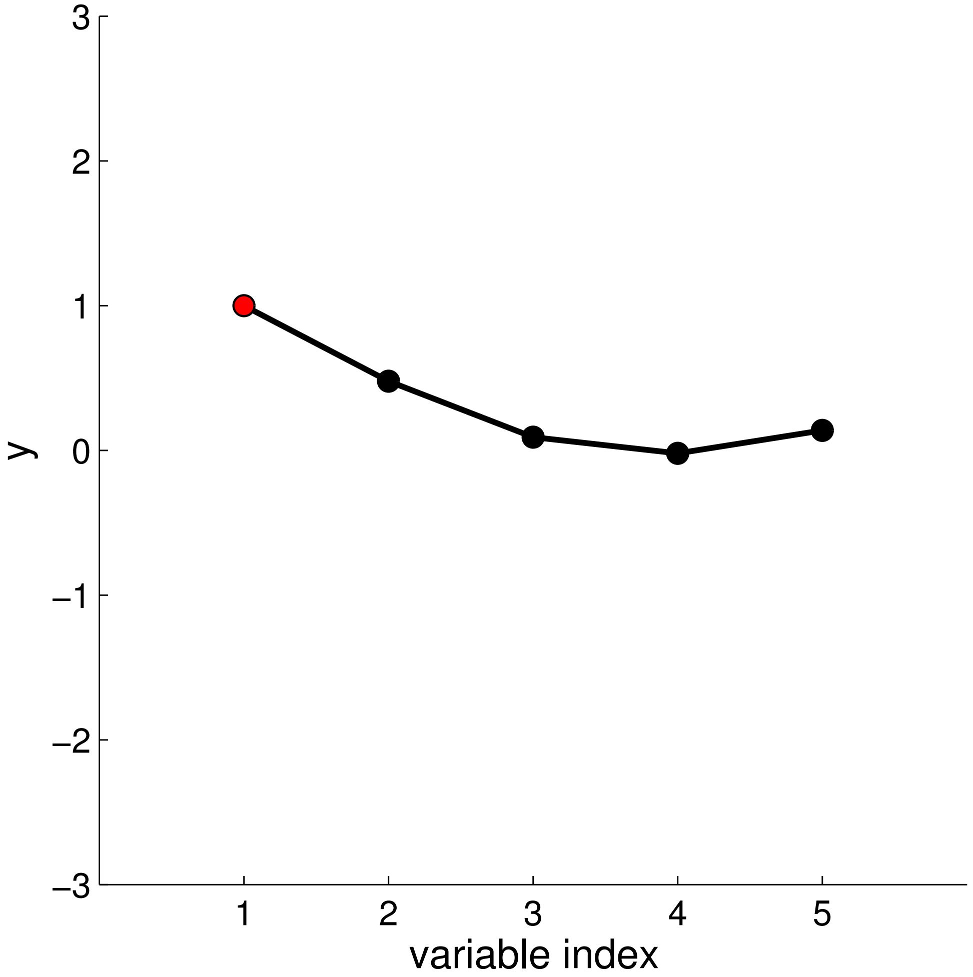




$$\Sigma = \begin{bmatrix} 1 & .9 & .8 & .6 & .4 \\ .9 & 1 & .9 & .8 & .6 \\ .8 & .9 & 1 & .9 & .8 \\ .6 & .8 & .9 & 1 & .9 \\ .4 & .6 & .8 & .9 & 1 \end{bmatrix}$$

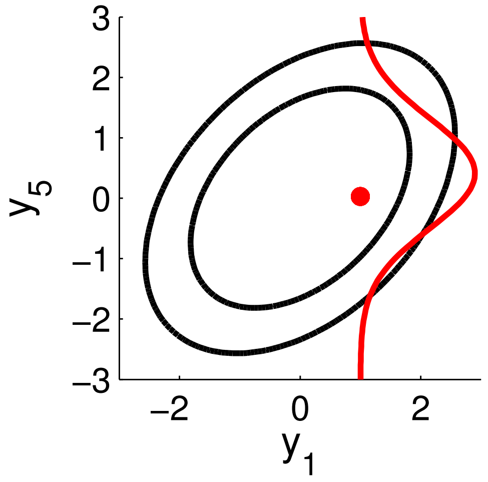
New visualisation

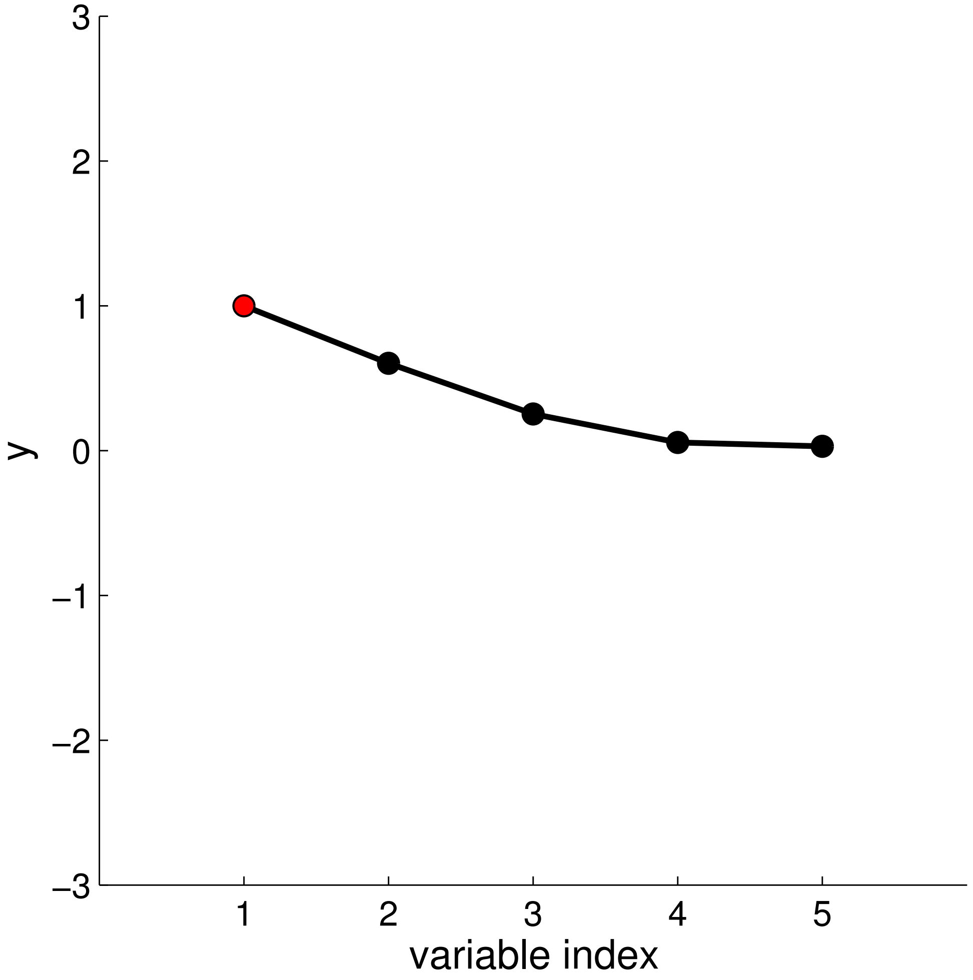




$$\Sigma = \begin{bmatrix} 1 & .9 & .8 & .6 & .4 \\ .9 & 1 & .9 & .8 & .6 \\ .8 & .9 & 1 & .9 & .8 \\ .6 & .8 & .9 & 1 & .9 \\ .4 & .6 & .8 & .9 & 1 \end{bmatrix}$$

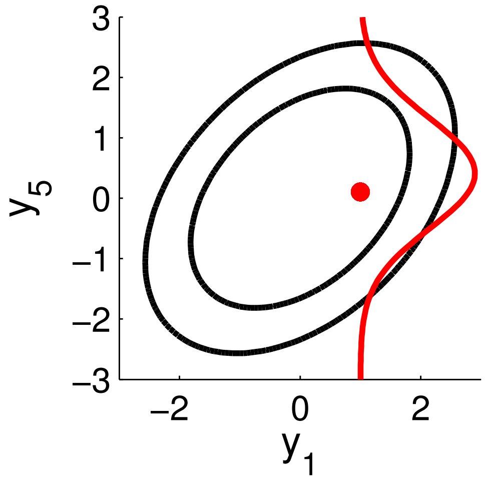
New visualisation

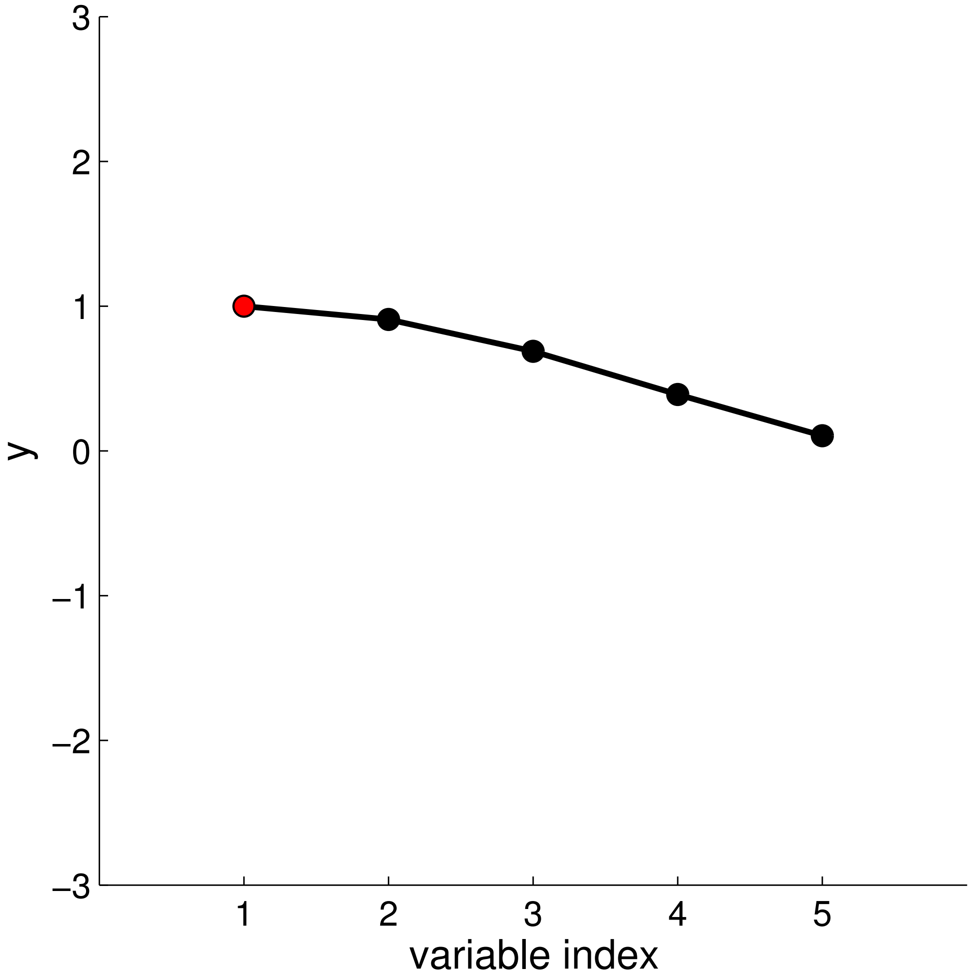




$$\Sigma = \begin{bmatrix} 1 & .9 & .8 & .6 & .4 \\ .9 & 1 & .9 & .8 & .6 \\ .8 & .9 & 1 & .9 & .8 \\ .6 & .8 & .9 & 1 & .9 \\ .4 & .6 & .8 & .9 & 1 \end{bmatrix}$$

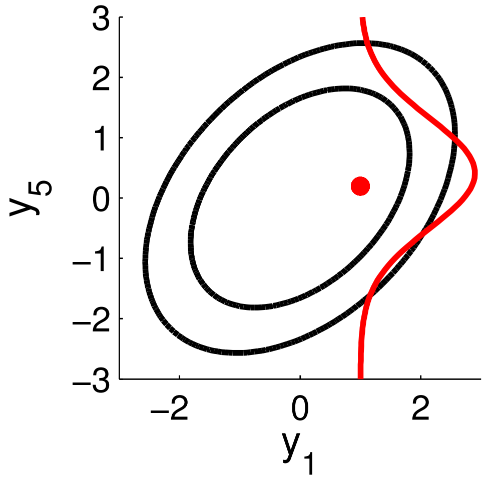
New visualisation

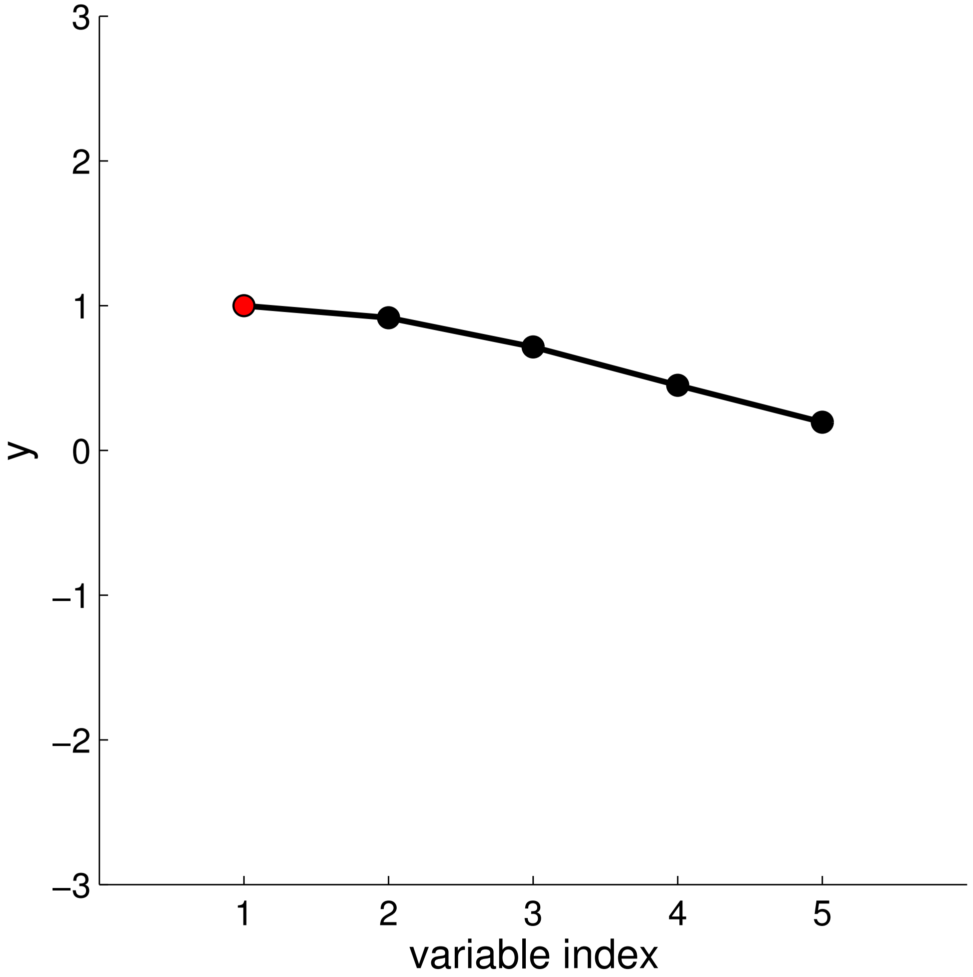




$$\Sigma = \begin{bmatrix} 1 & .9 & .8 & .6 & .4 \\ .9 & 1 & .9 & .8 & .6 \\ .8 & .9 & 1 & .9 & .8 \\ .6 & .8 & .9 & 1 & .9 \\ .4 & .6 & .8 & .9 & 1 \end{bmatrix}$$

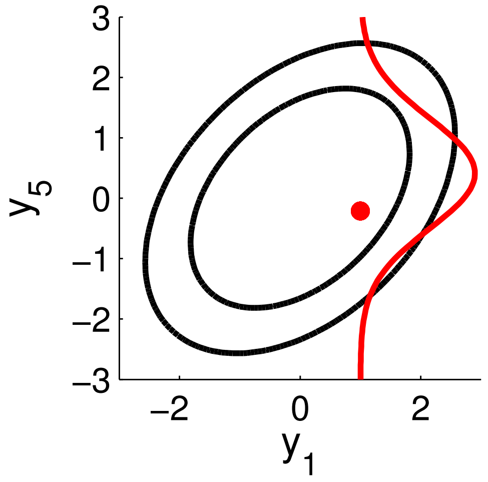
New visualisation

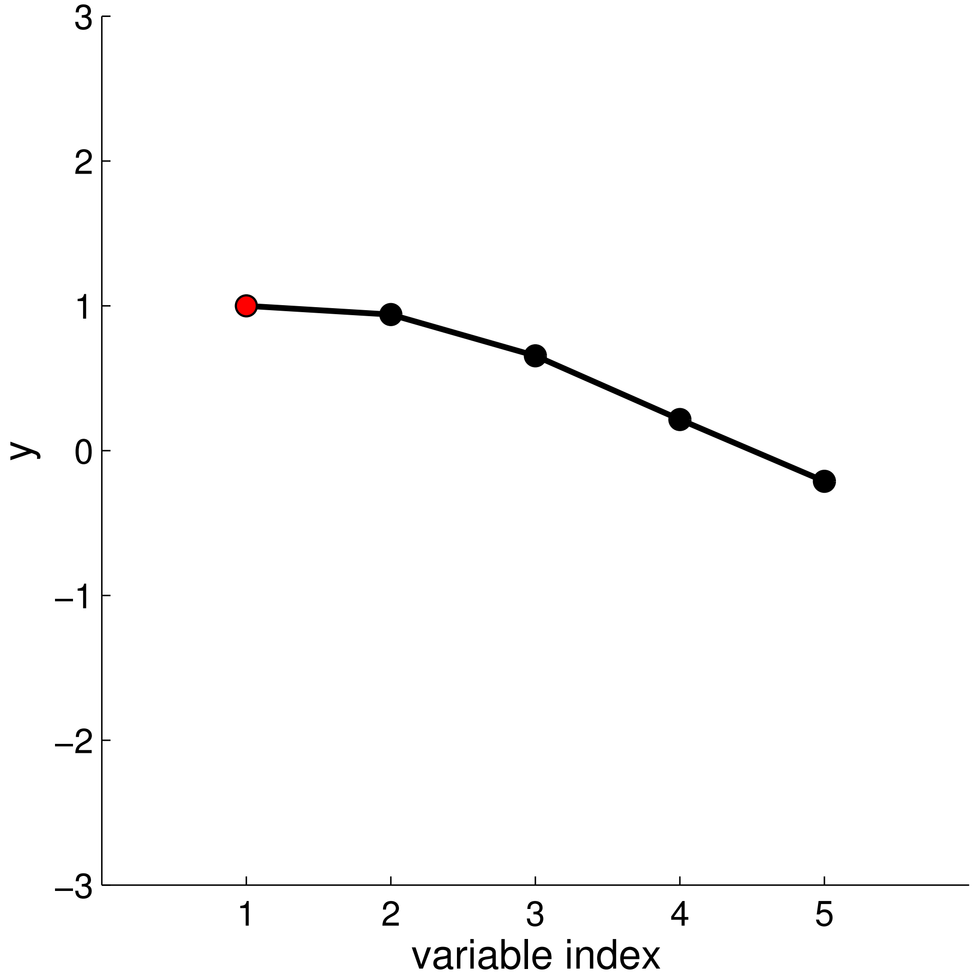




$$\Sigma = \begin{bmatrix} 1 & .9 & .8 & .6 & .4 \\ .9 & 1 & .9 & .8 & .6 \\ .8 & .9 & 1 & .9 & .8 \\ .6 & .8 & .9 & 1 & .9 \\ .4 & .6 & .8 & .9 & 1 \end{bmatrix}$$

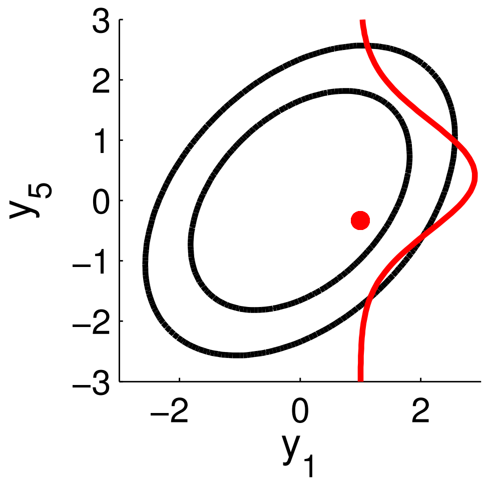
New visualisation

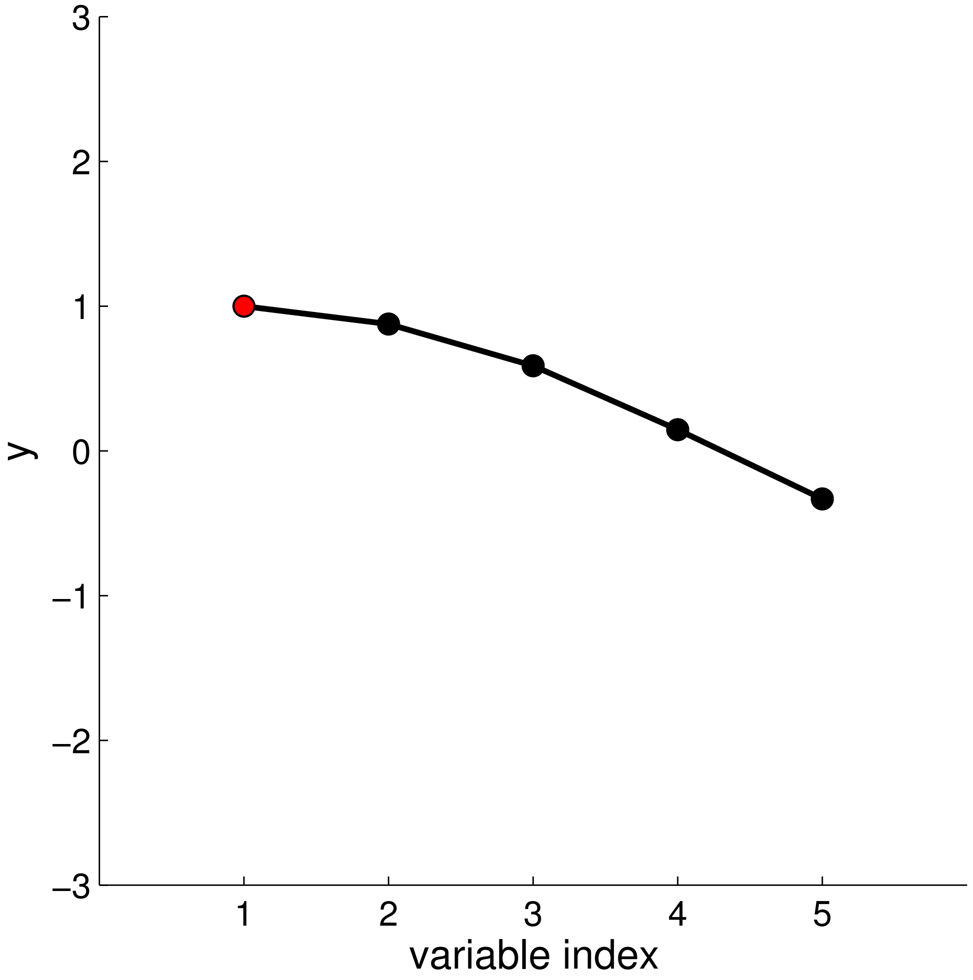




$$\Sigma = \begin{bmatrix} 1 & .9 & .8 & .6 & .4 \\ .9 & 1 & .9 & .8 & .6 \\ .8 & .9 & 1 & .9 & .8 \\ .6 & .8 & .9 & 1 & .9 \\ .4 & .6 & .8 & .9 & 1 \end{bmatrix}$$

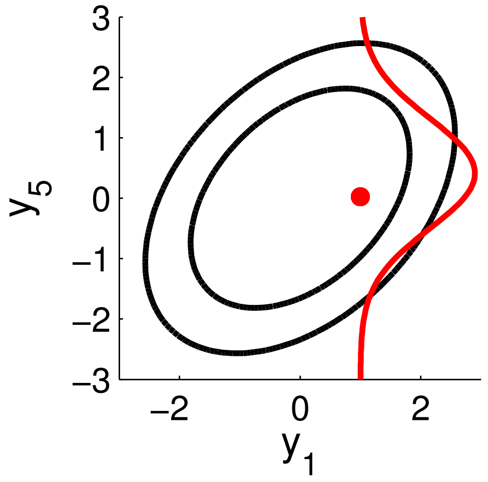
New visualisation

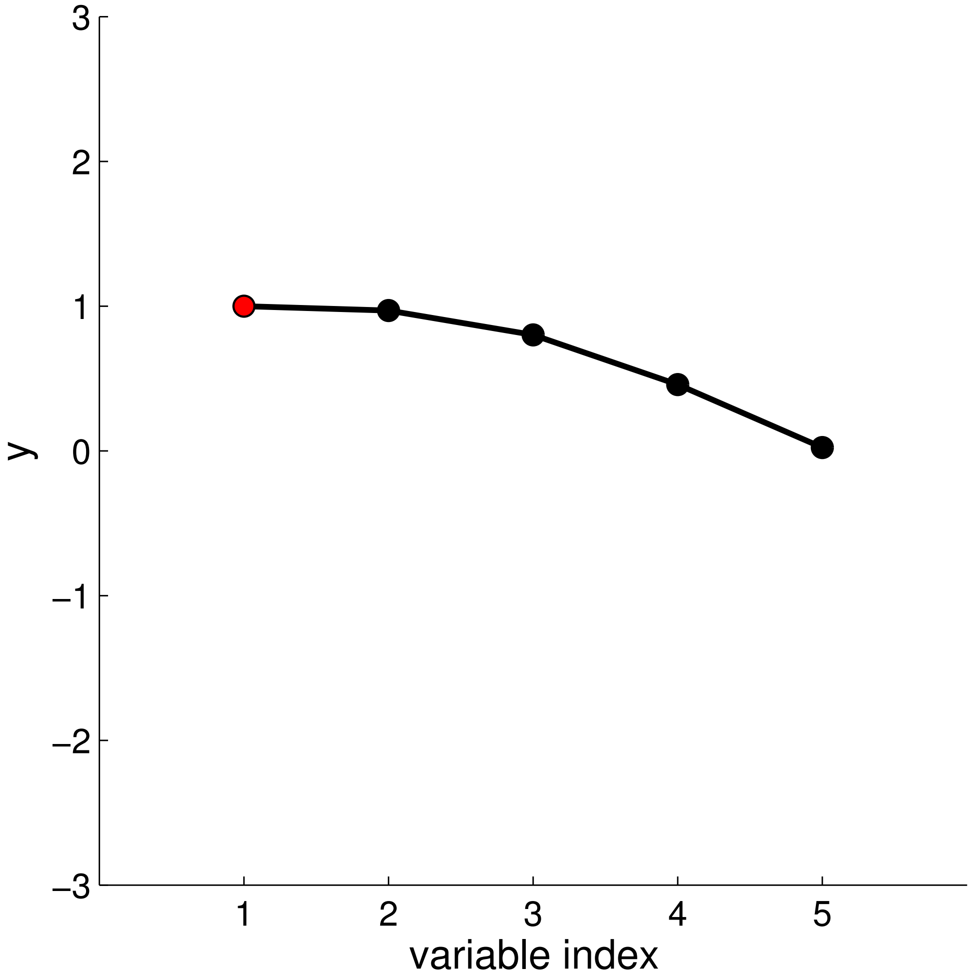




$$\Sigma = \begin{bmatrix} 1 & .9 & .8 & .6 & .4 \\ .9 & 1 & .9 & .8 & .6 \\ .8 & .9 & 1 & .9 & .8 \\ .6 & .8 & .9 & 1 & .9 \\ .4 & .6 & .8 & .9 & 1 \end{bmatrix}$$

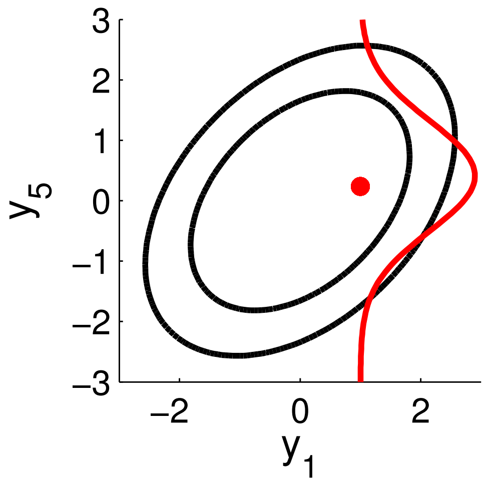
New visualisation

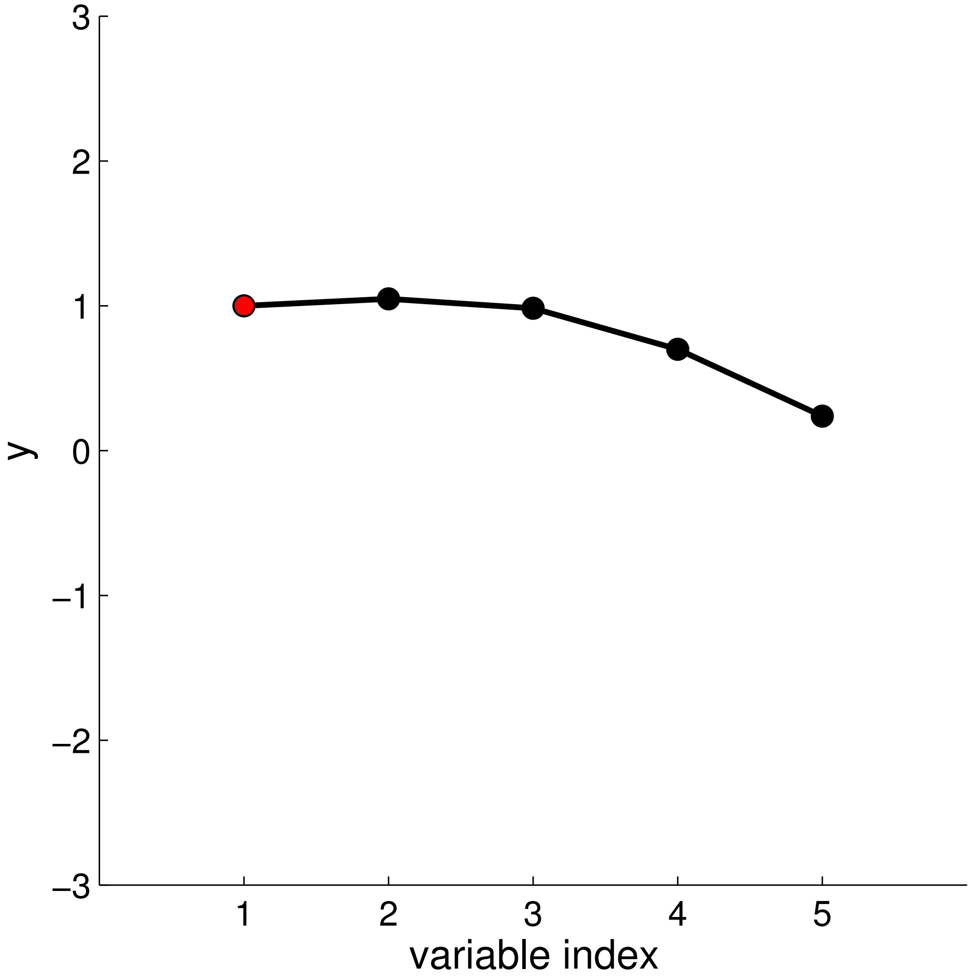




$$\Sigma = \begin{bmatrix} 1 & .9 & .8 & .6 & .4 \\ .9 & 1 & .9 & .8 & .6 \\ .8 & .9 & 1 & .9 & .8 \\ .6 & .8 & .9 & 1 & .9 \\ .4 & .6 & .8 & .9 & 1 \end{bmatrix}$$

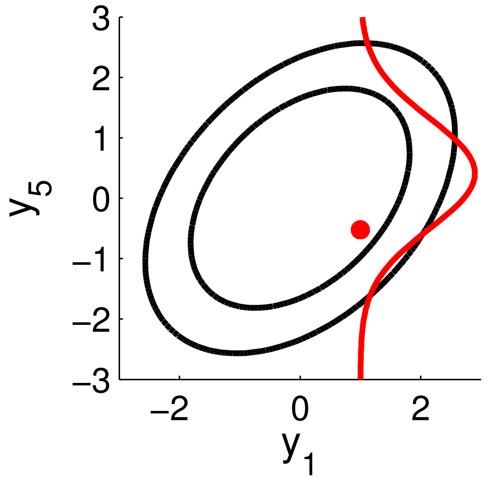
New visualisation

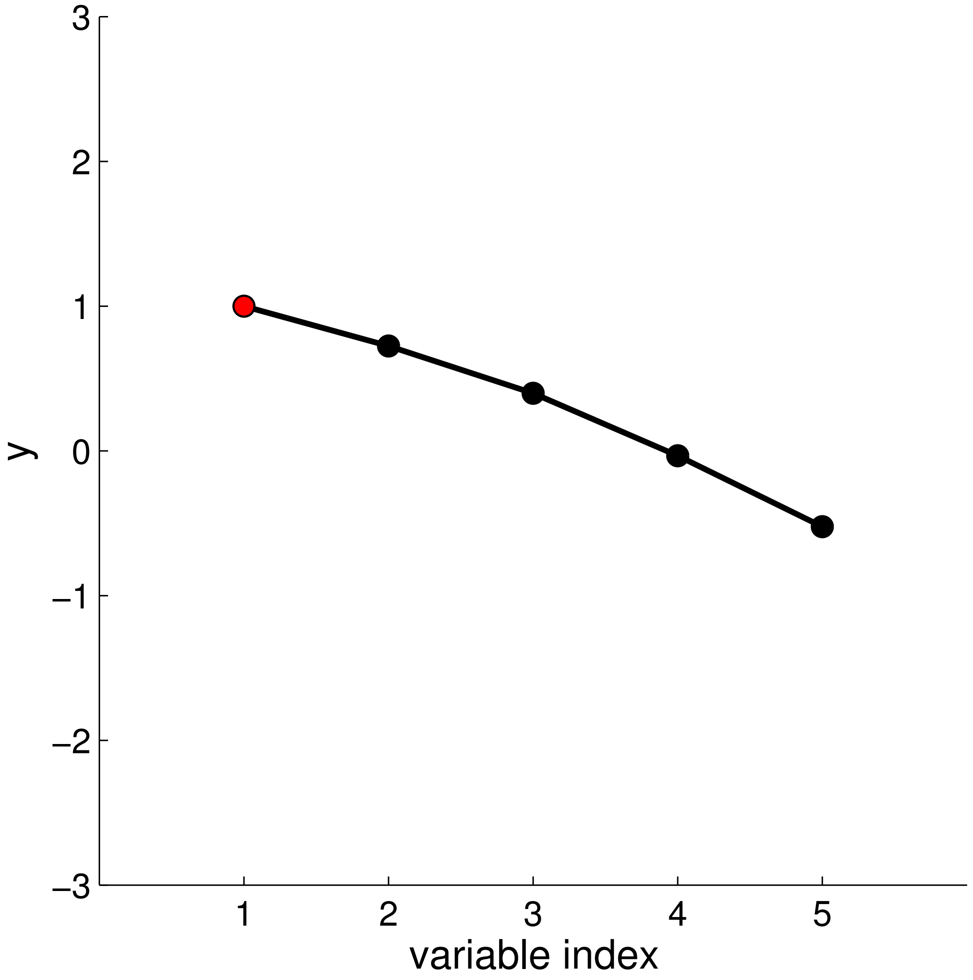




$$\Sigma = \begin{bmatrix} 1 & .9 & .8 & .6 & .4 \\ .9 & 1 & .9 & .8 & .6 \\ .8 & .9 & 1 & .9 & .8 \\ .6 & .8 & .9 & 1 & .9 \\ .4 & .6 & .8 & .9 & 1 \end{bmatrix}$$

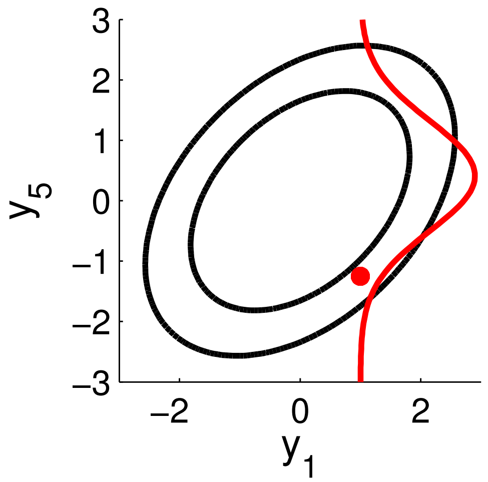
New visualisation

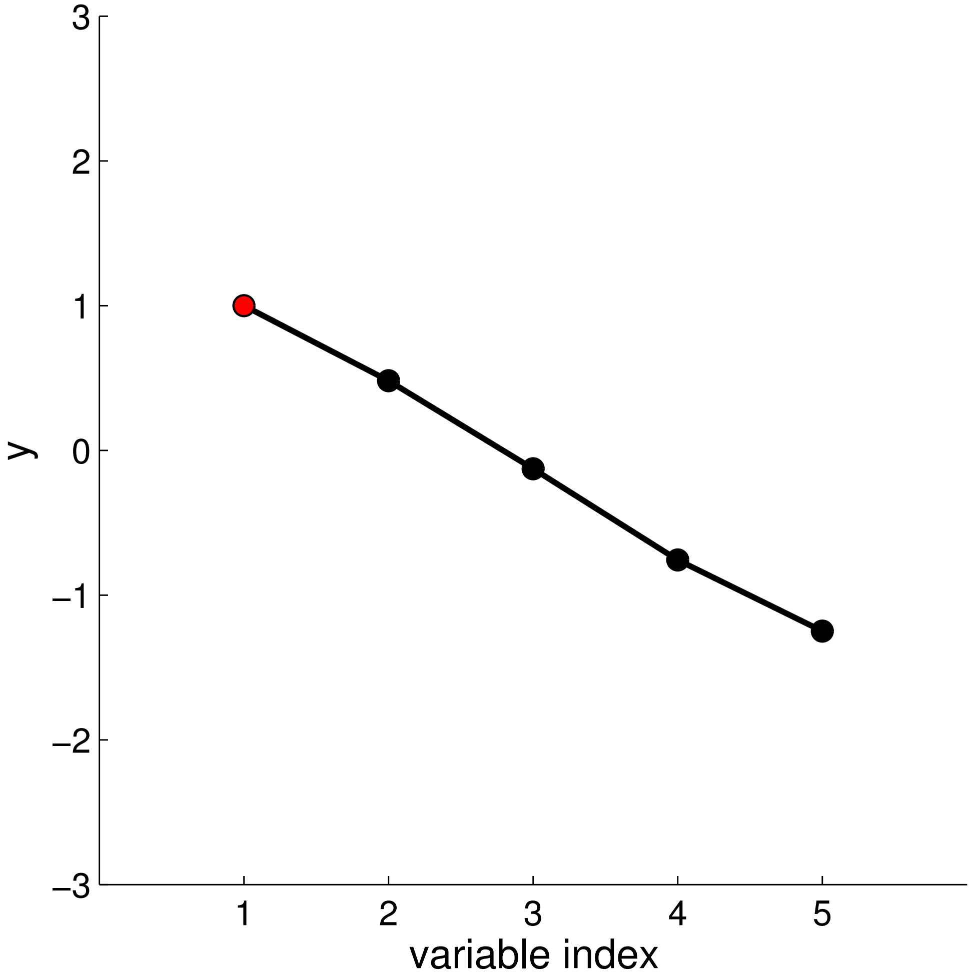




$$\Sigma = \begin{bmatrix} 1 & .9 & .8 & .6 & .4 \\ .9 & 1 & .9 & .8 & .6 \\ .8 & .9 & 1 & .9 & .8 \\ .6 & .8 & .9 & 1 & .9 \\ .4 & .6 & .8 & .9 & 1 \end{bmatrix}$$

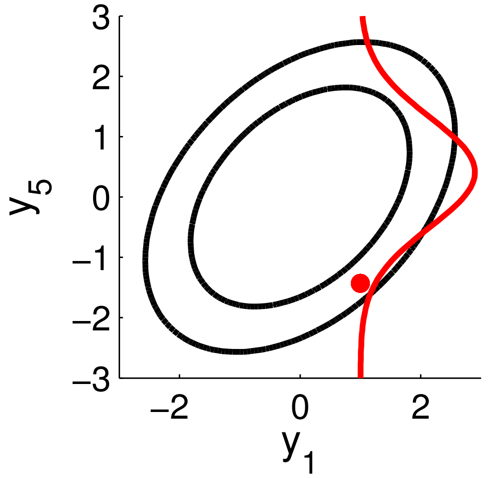
New visualisation

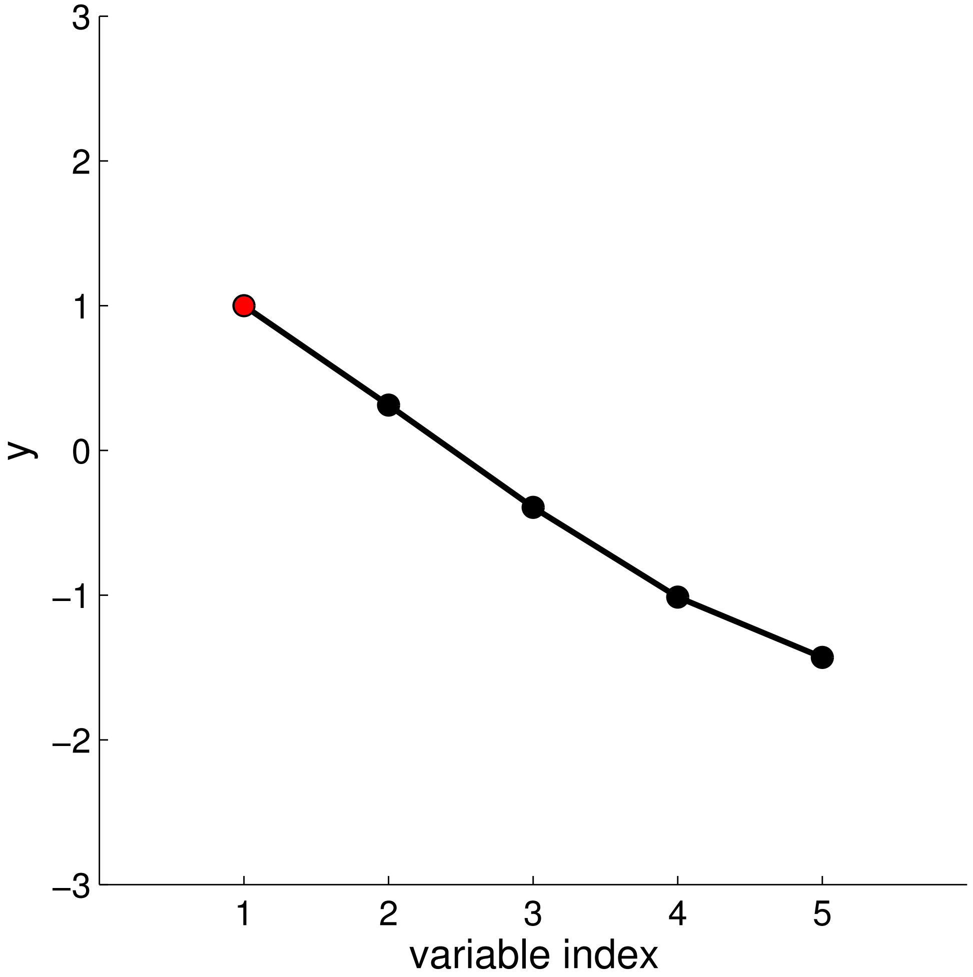




$$\Sigma = \begin{bmatrix} 1 & .9 & .8 & .6 & .4 \\ .9 & 1 & .9 & .8 & .6 \\ .8 & .9 & 1 & .9 & .8 \\ .6 & .8 & .9 & 1 & .9 \\ .4 & .6 & .8 & .9 & 1 \end{bmatrix}$$

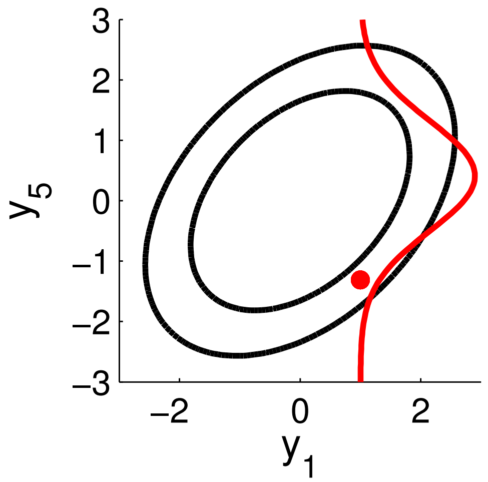
New visualisation



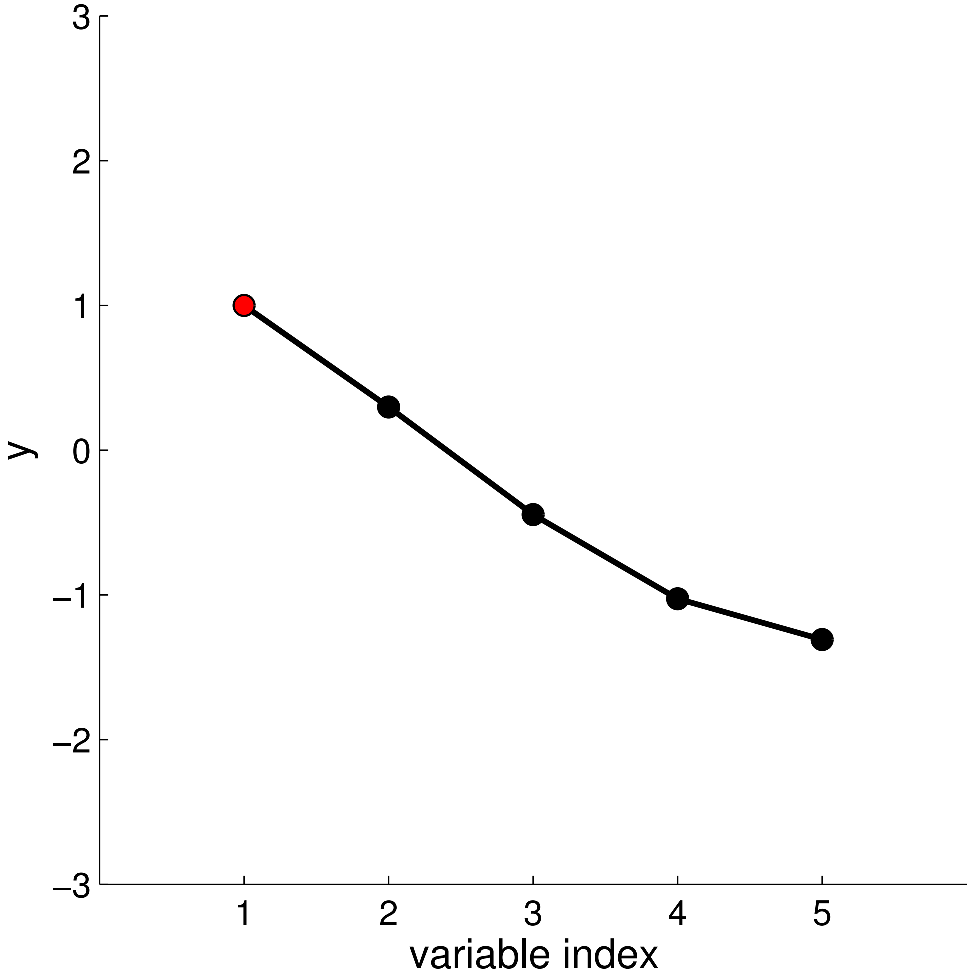


$$\Sigma = \begin{bmatrix} 1 & .9 & .8 & .6 & .4 \\ .9 & 1 & .9 & .8 & .6 \\ .8 & .9 & 1 & .9 & .8 \\ .6 & .8 & .9 & 1 & .9 \\ .4 & .6 & .8 & .9 & 1 \end{bmatrix}$$

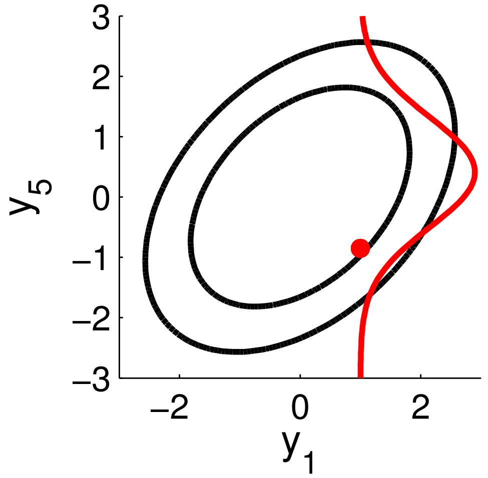
New visualisation

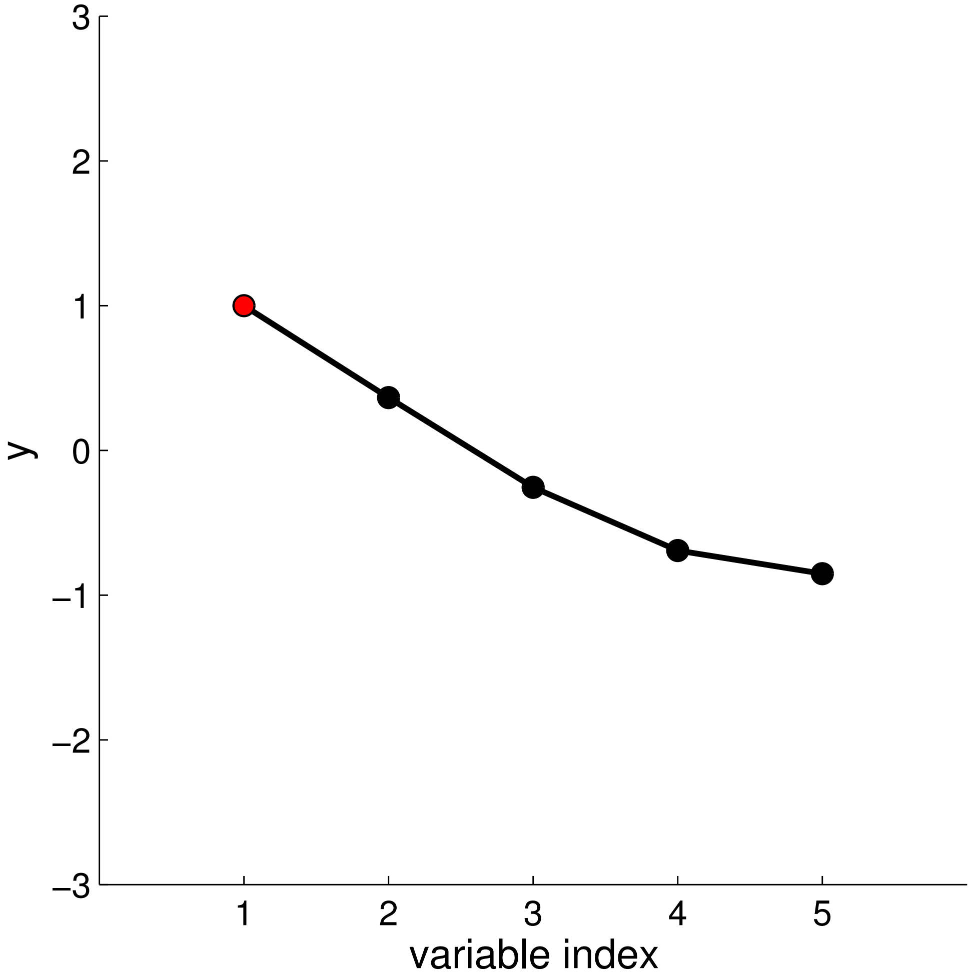


$$\Sigma = \begin{bmatrix} 1 & .9 & .8 & .6 & .4 \\ .9 & 1 & .9 & .8 & .6 \\ .8 & .9 & 1 & .9 & .8 \\ .6 & .8 & .9 & 1 & .9 \\ .4 & .6 & .8 & .9 & 1 \end{bmatrix}$$



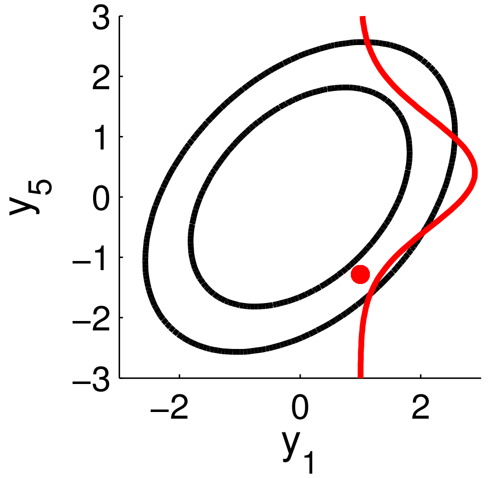
New visualisation



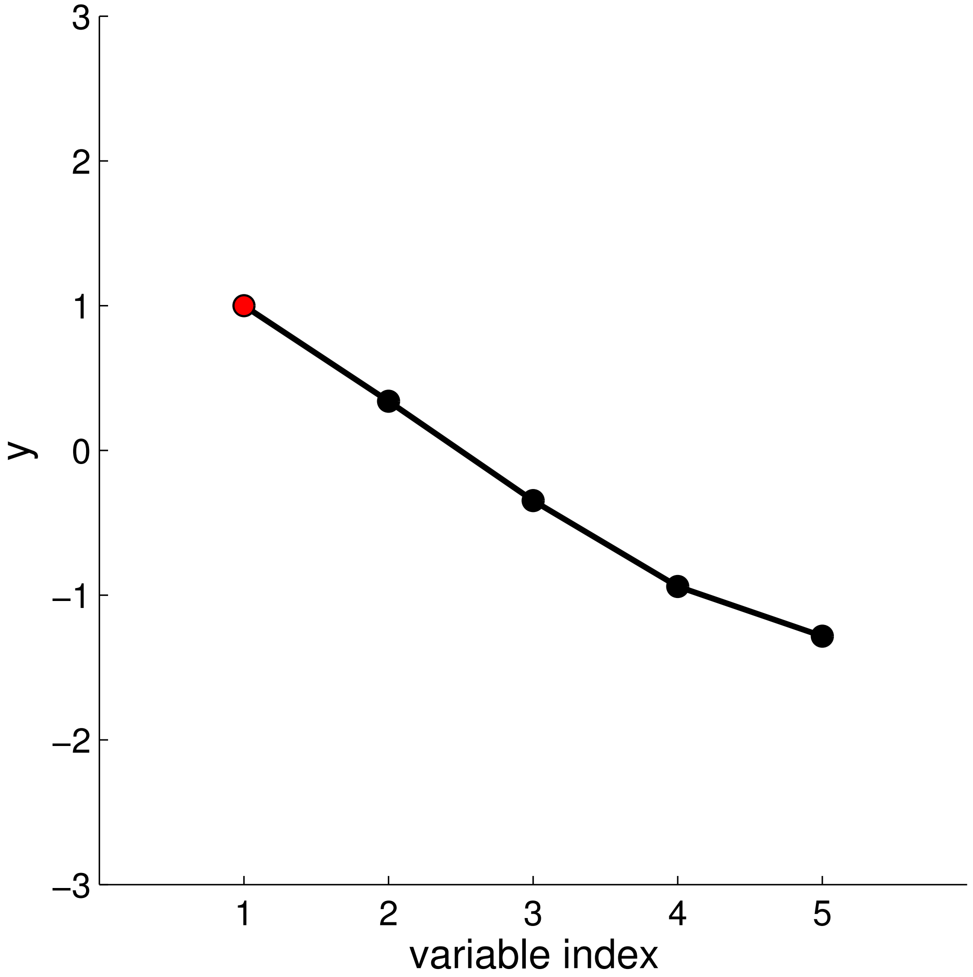


$$\Sigma = \begin{bmatrix} 1 & .9 & .8 & .6 & .4 \\ .9 & 1 & .9 & .8 & .6 \\ .8 & .9 & 1 & .9 & .8 \\ .6 & .8 & .9 & 1 & .9 \\ .4 & .6 & .8 & .9 & 1 \end{bmatrix}$$

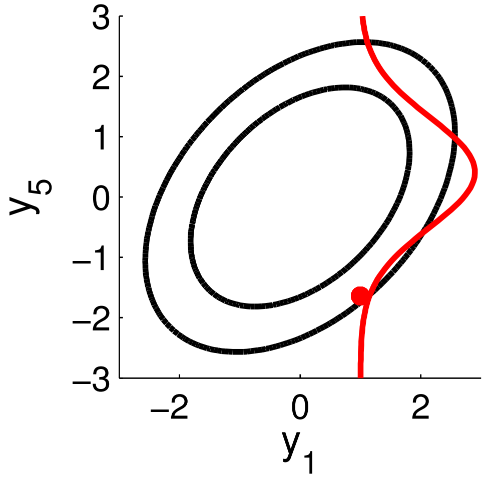
New visualisation



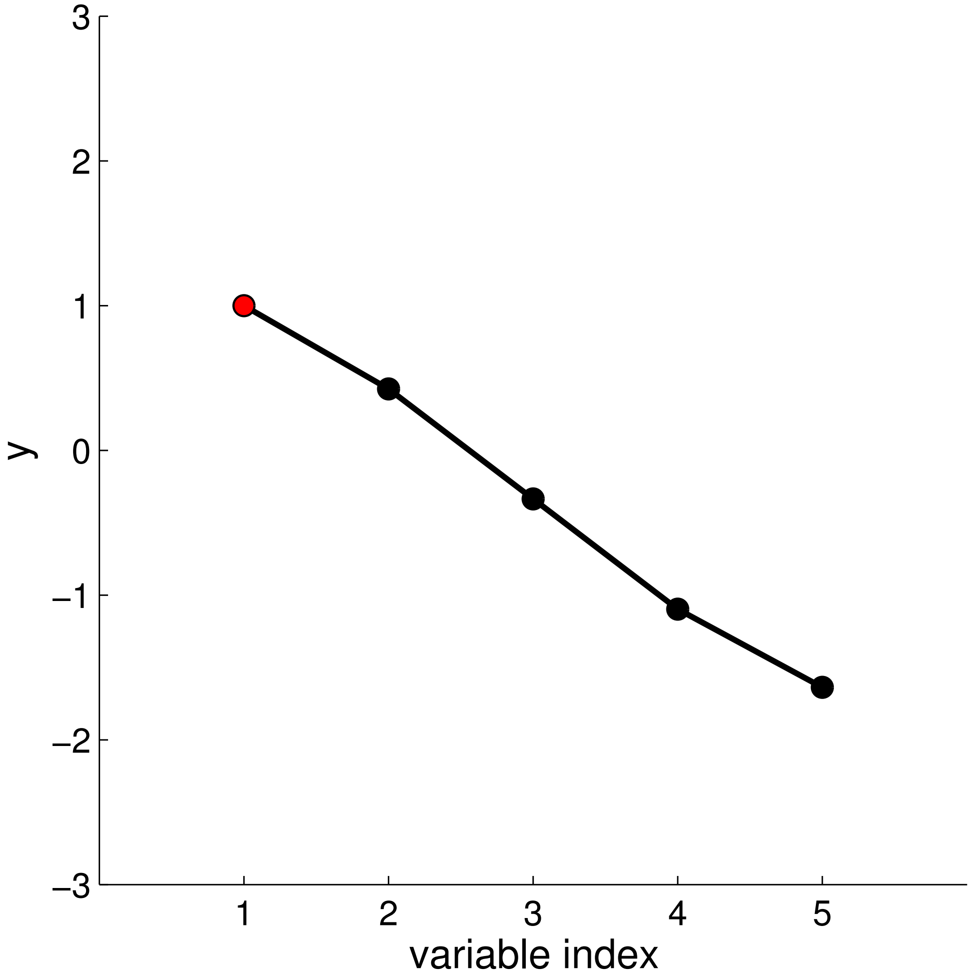
$$\Sigma = \begin{bmatrix} 1 & .9 & .8 & .6 & .4 \\ .9 & 1 & .9 & .8 & .6 \\ .8 & .9 & 1 & .9 & .8 \\ .6 & .8 & .9 & 1 & .9 \\ .4 & .6 & .8 & .9 & 1 \end{bmatrix}$$



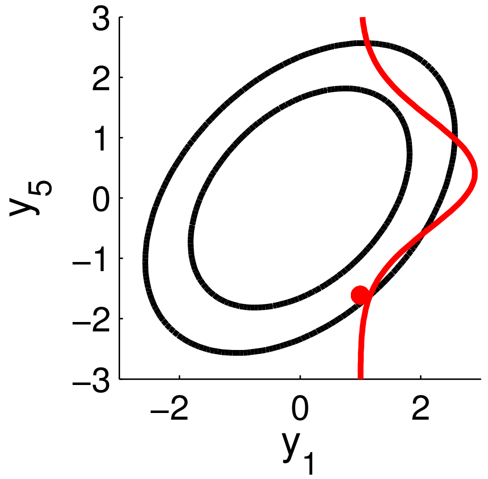
New visualisation

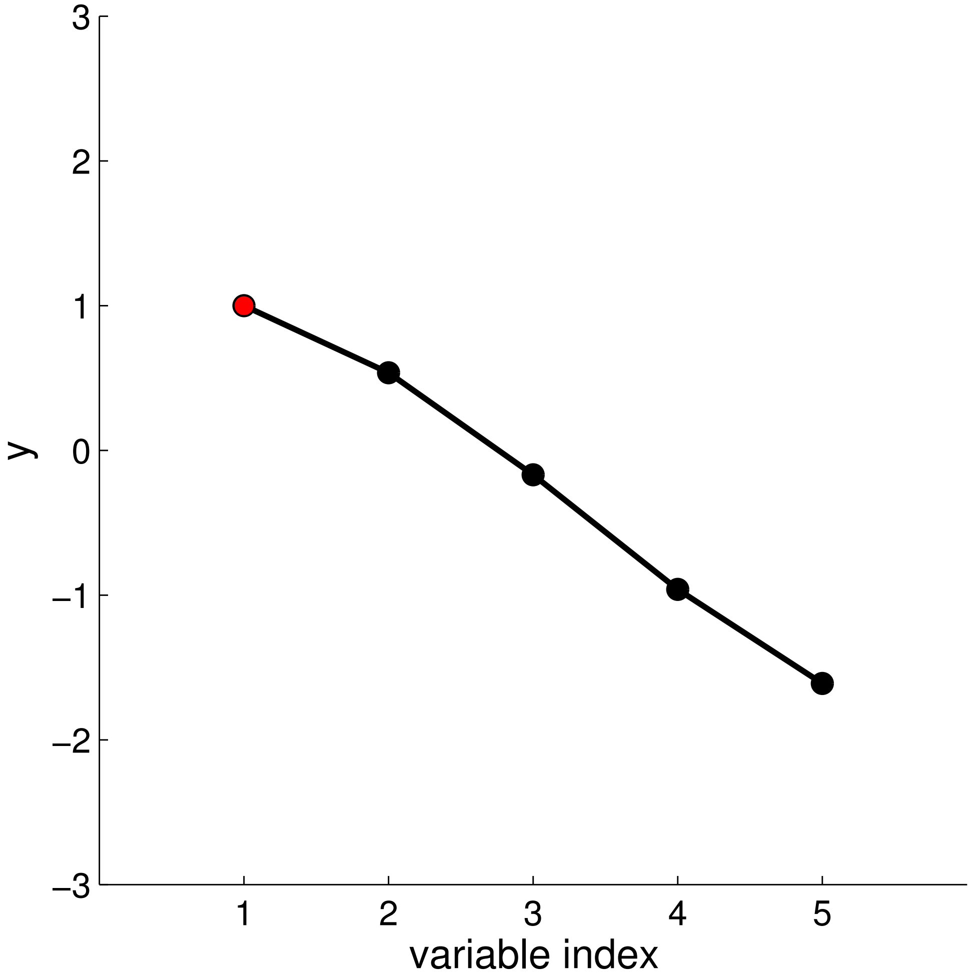


$$\Sigma = \begin{bmatrix} 1 & .9 & .8 & .6 & .4 \\ .9 & 1 & .9 & .8 & .6 \\ .8 & .9 & 1 & .9 & .8 \\ .6 & .8 & .9 & 1 & .9 \\ .4 & .6 & .8 & .9 & 1 \end{bmatrix}$$



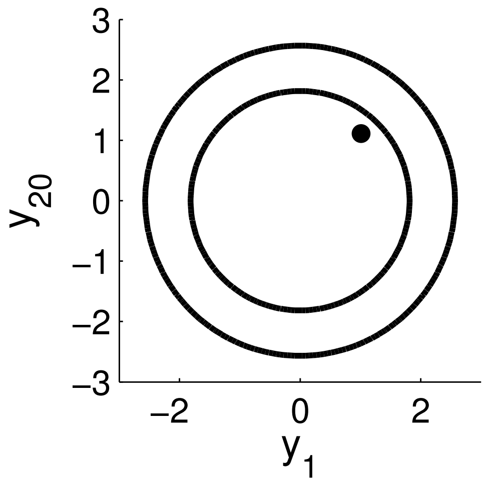
New visualisation

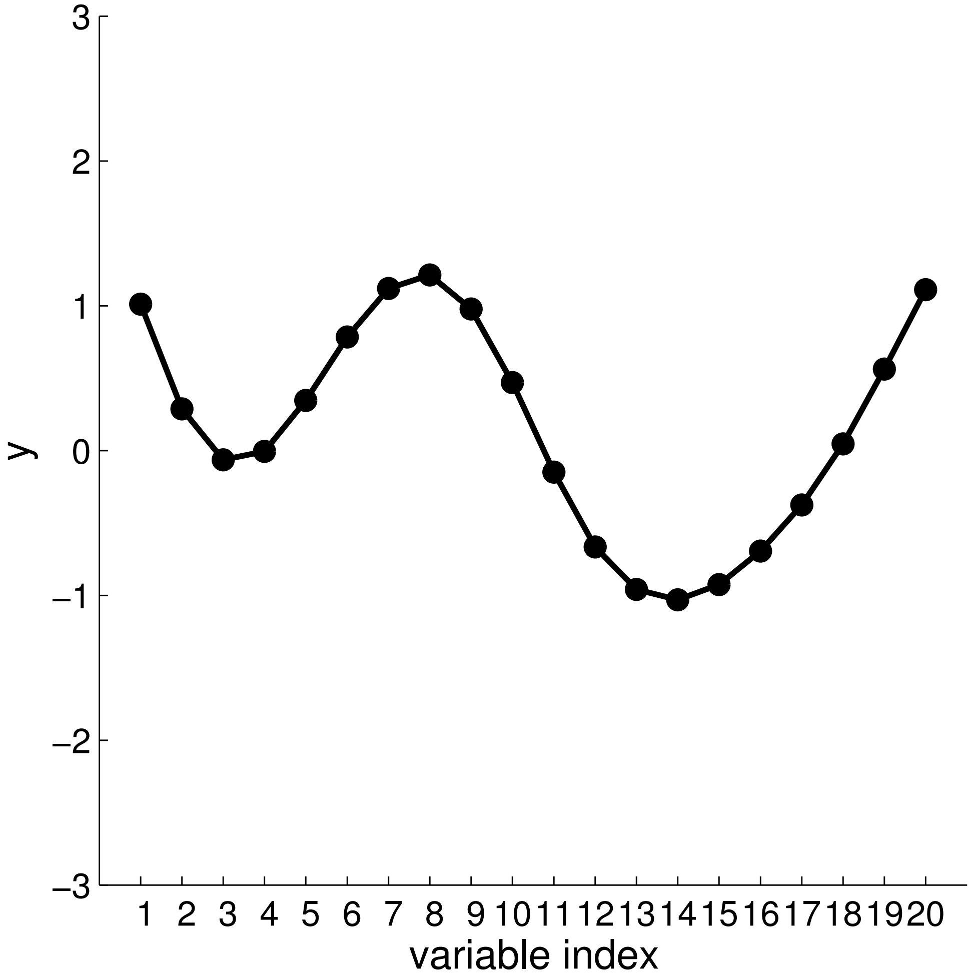




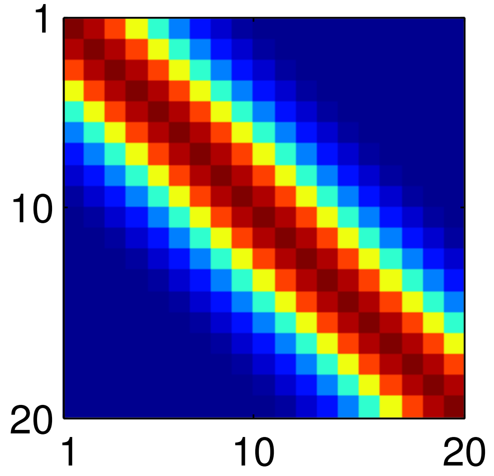
$$\Sigma = \begin{bmatrix} 1 & .9 & .8 & .6 & .4 \\ .9 & 1 & .9 & .8 & .6 \\ .8 & .9 & 1 & .9 & .8 \\ .6 & .8 & .9 & 1 & .9 \\ .4 & .6 & .8 & .9 & 1 \end{bmatrix}$$

New visualisation



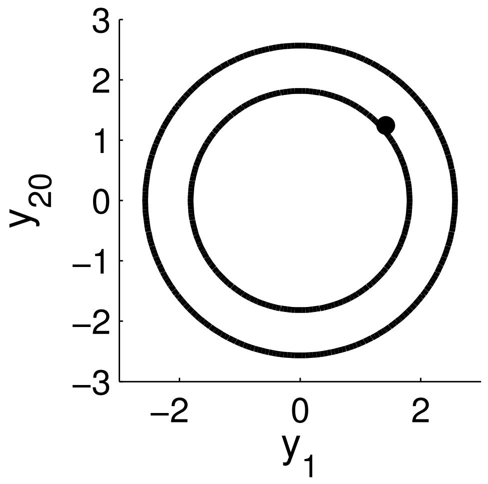


$\Sigma =$

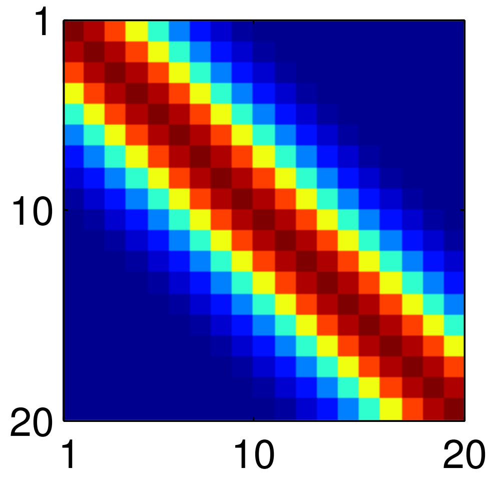


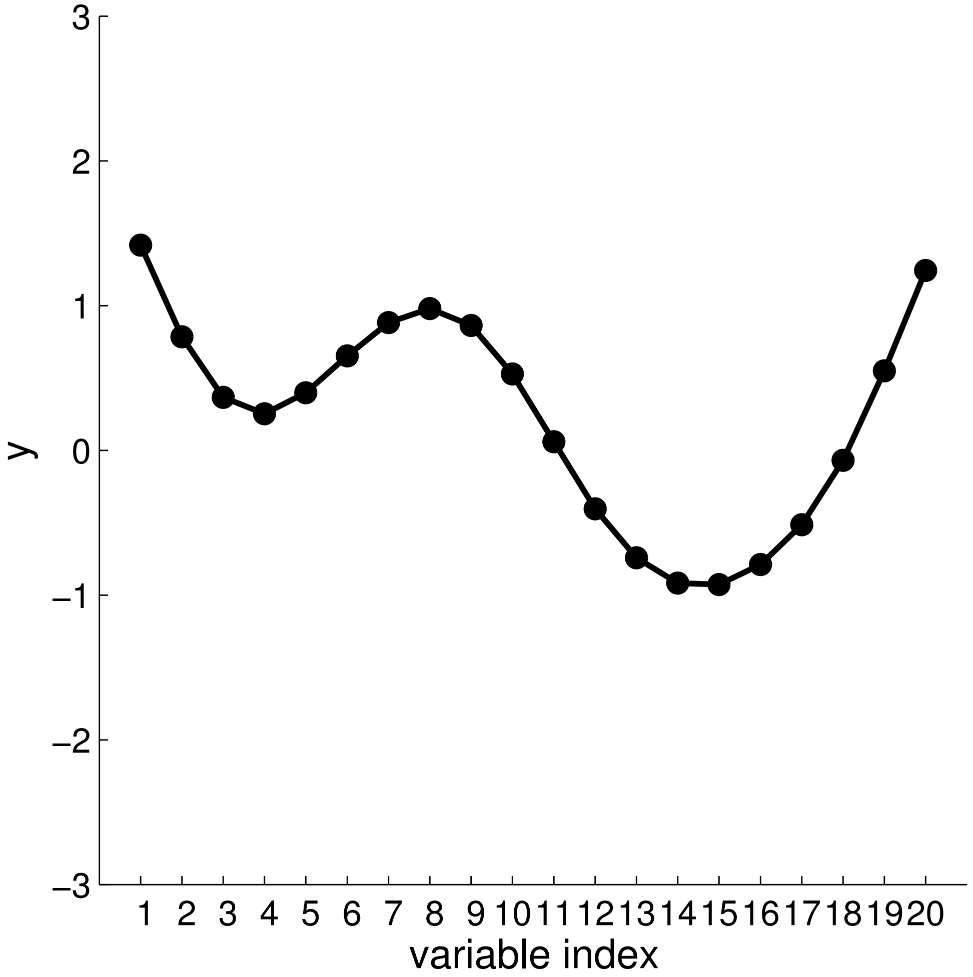
red is high, blue is low correlation

New visualisation

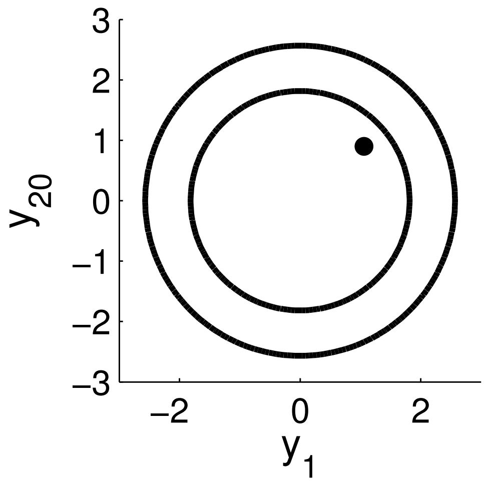


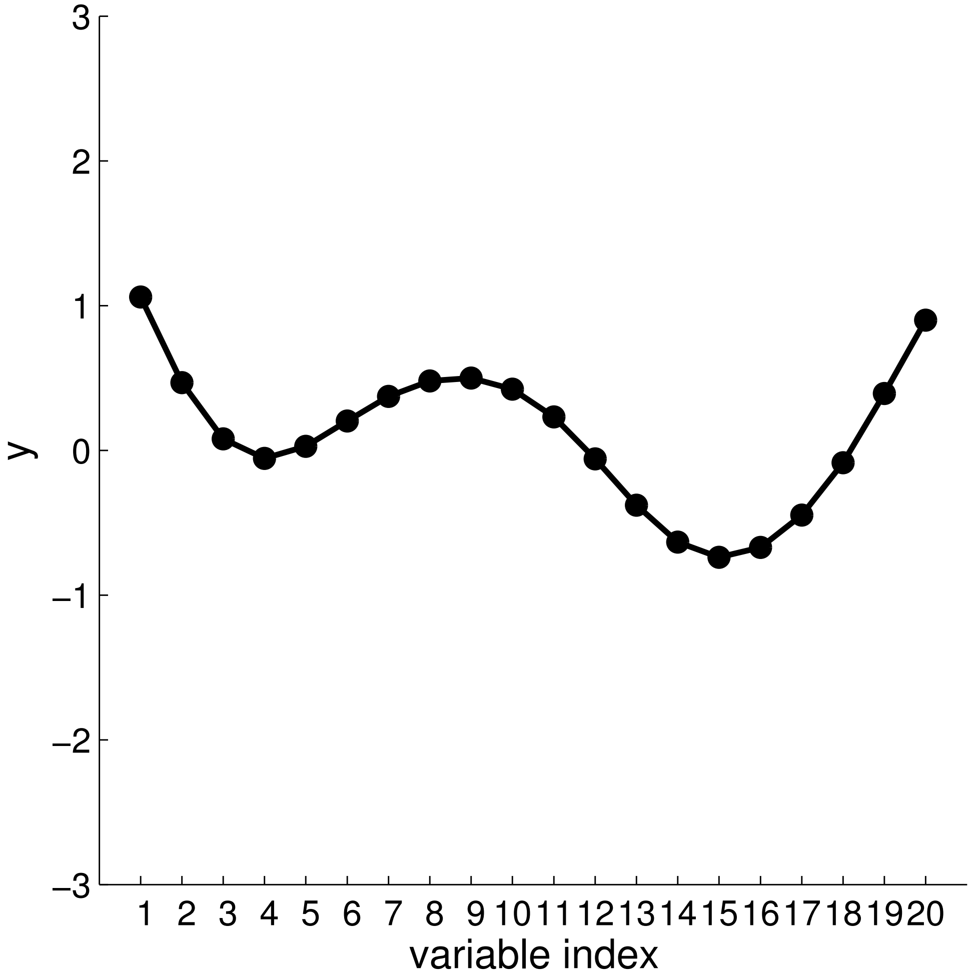
$\Sigma =$



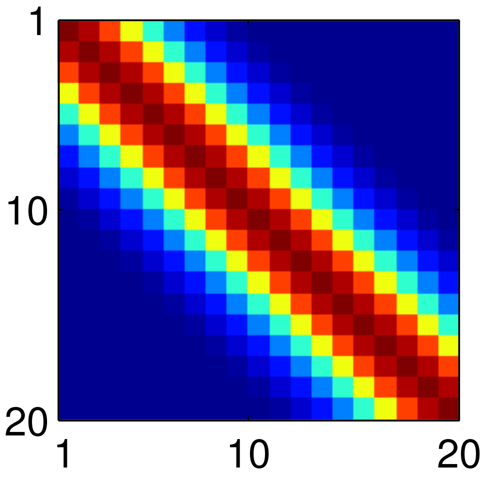


New visualisation

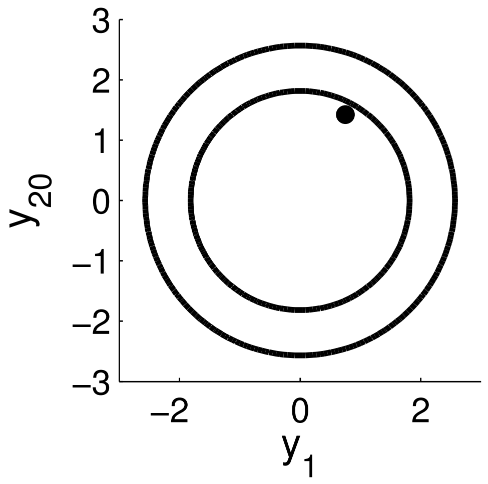


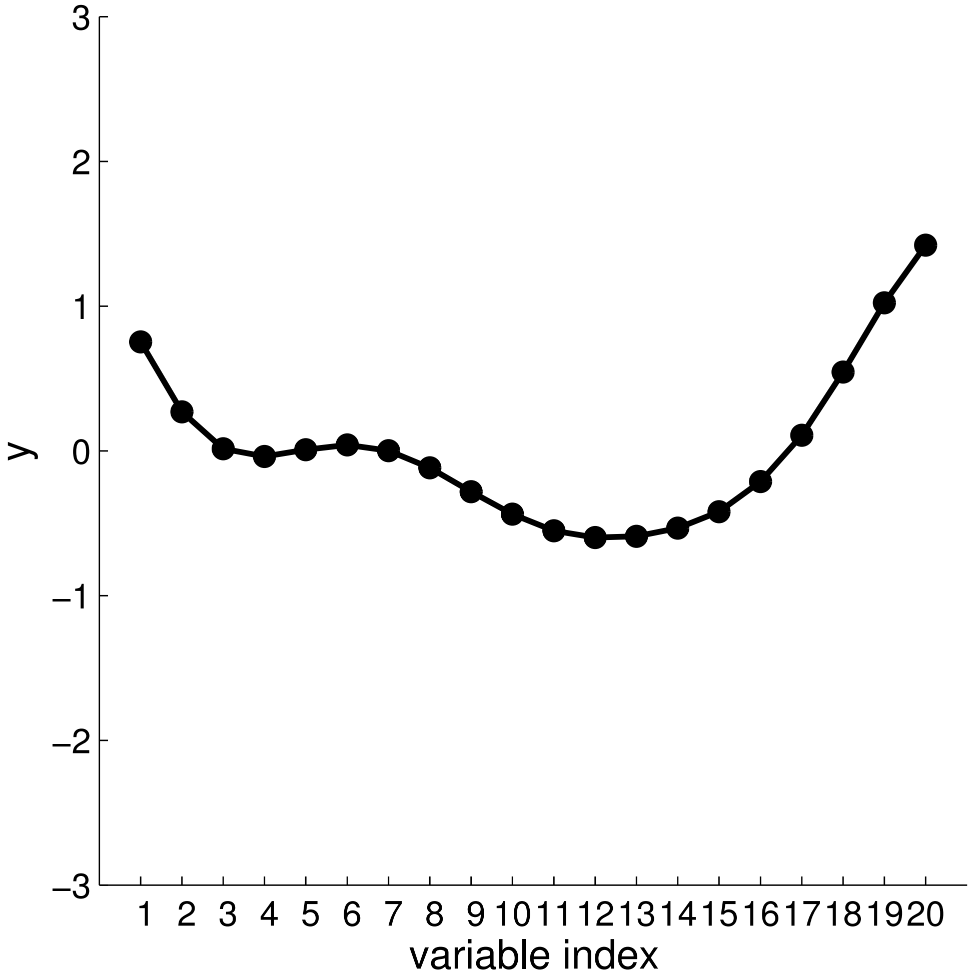


$\Sigma =$

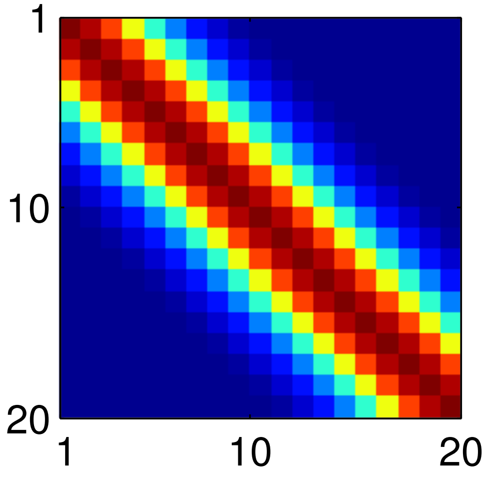


New visualisation

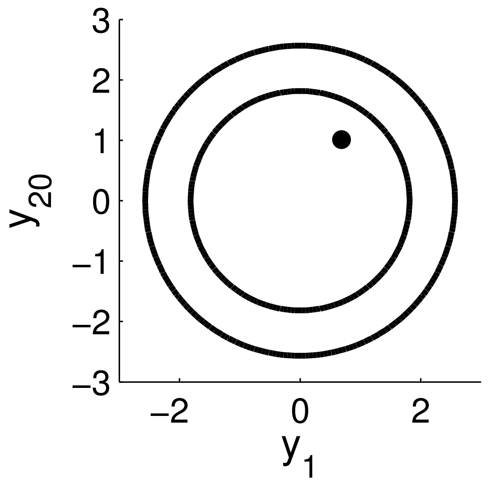


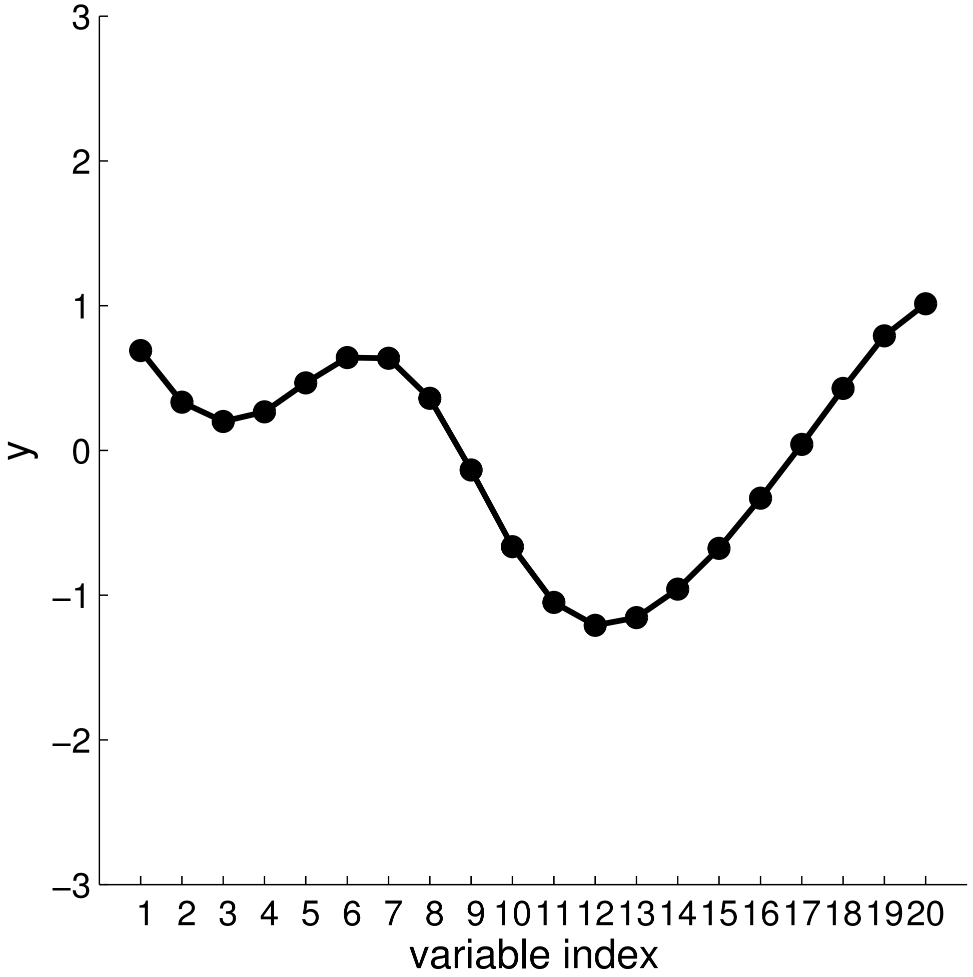


$\Sigma =$

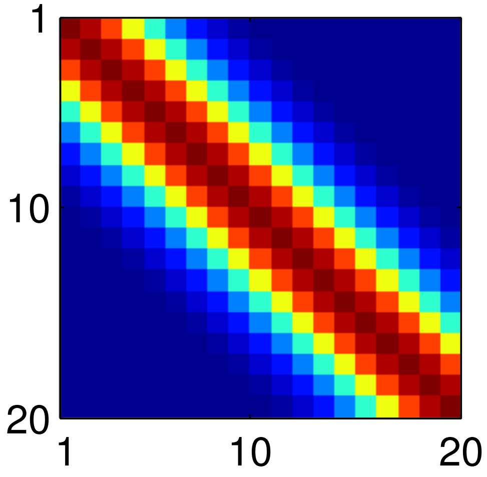


New visualisation

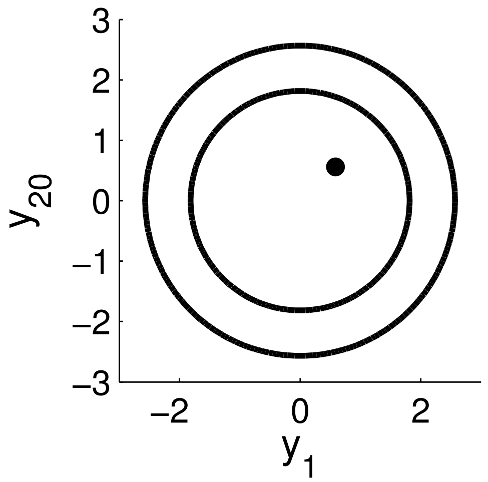




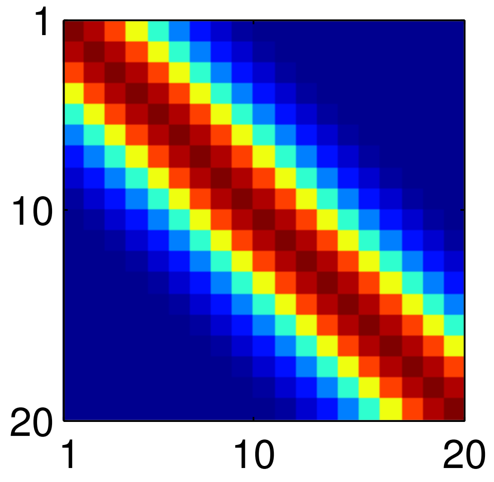
$\Sigma =$

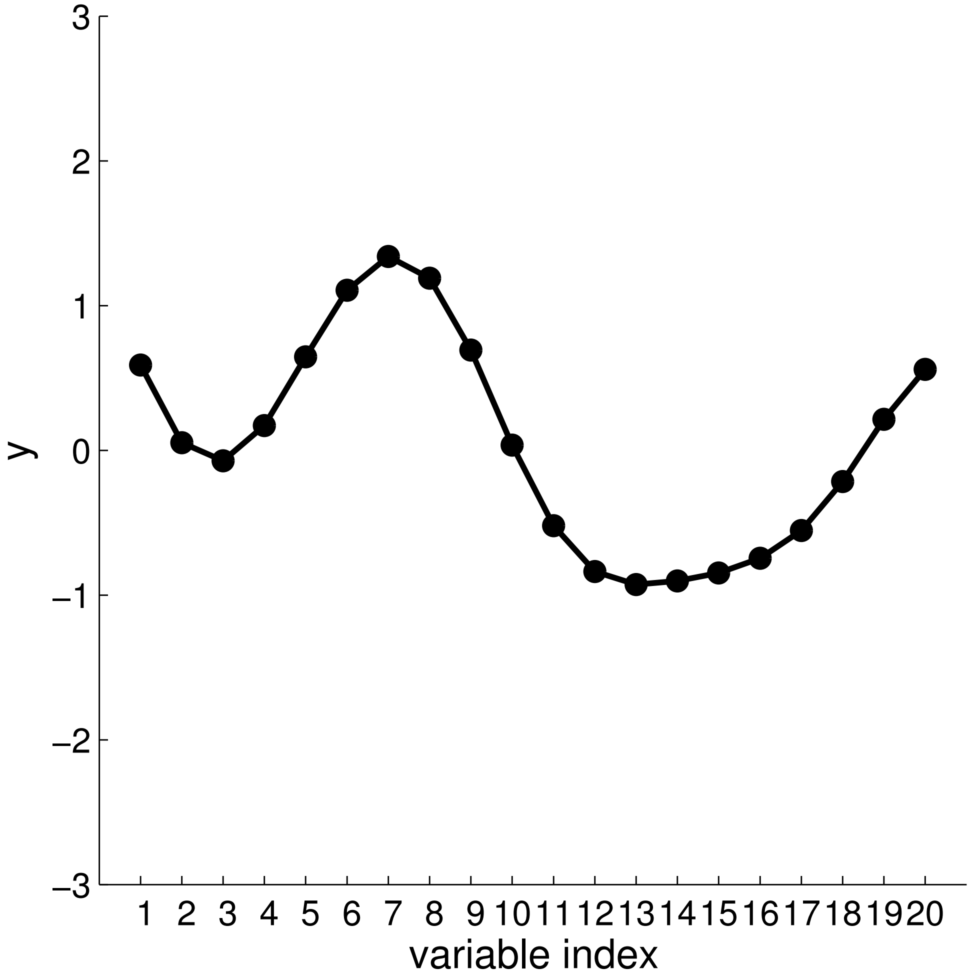


New visualisation

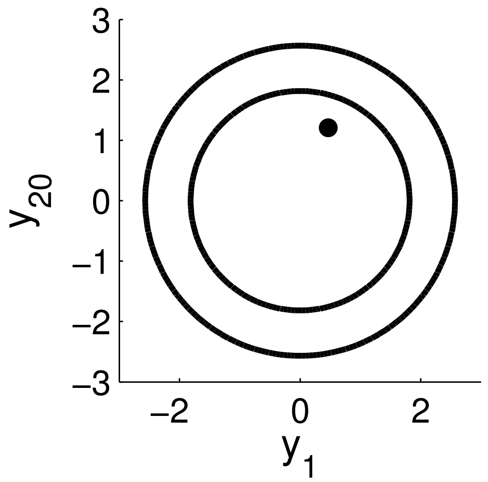


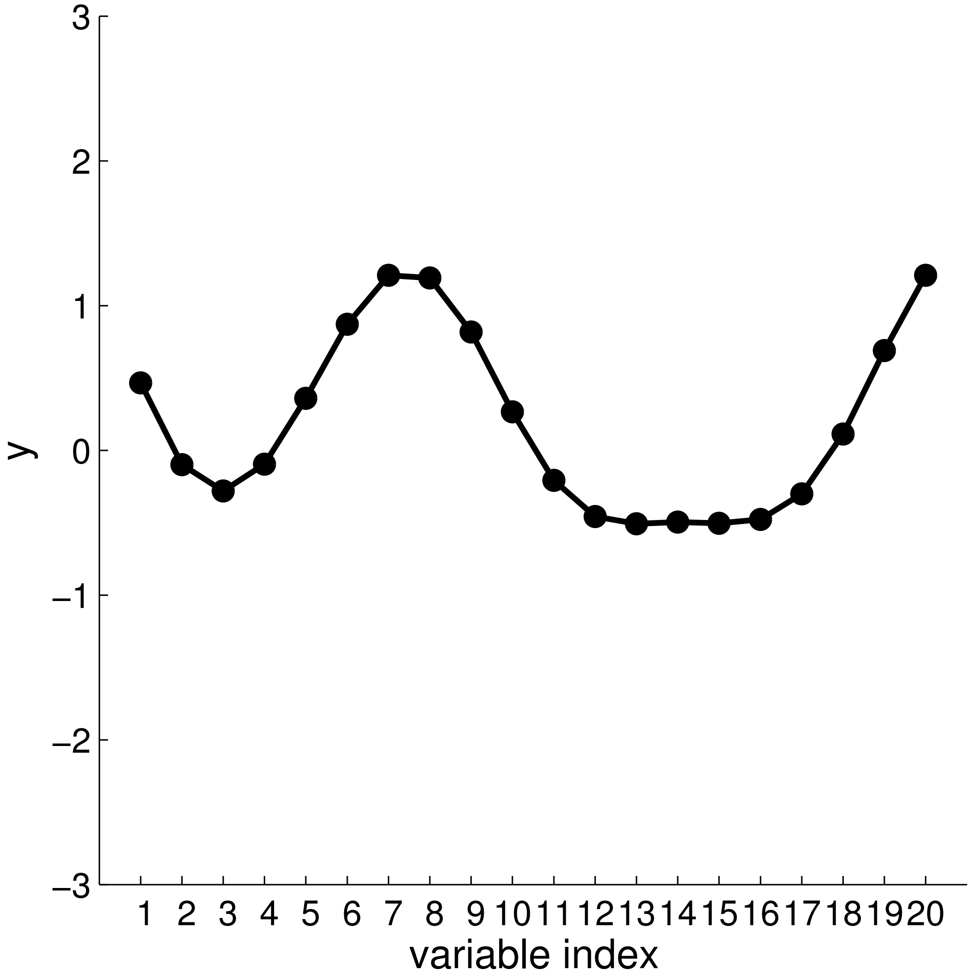
$\Sigma =$



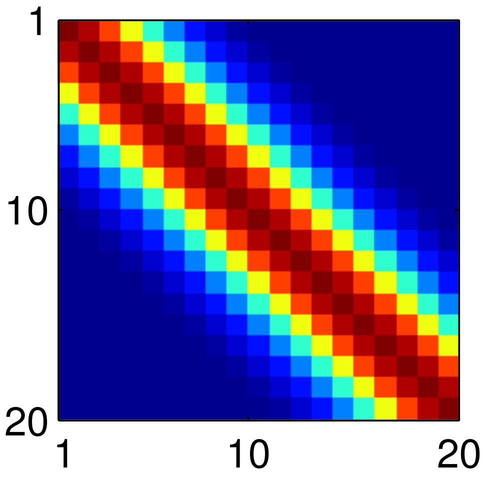


New visualisation

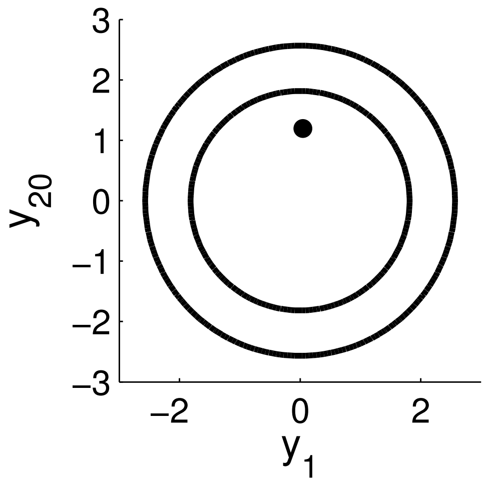


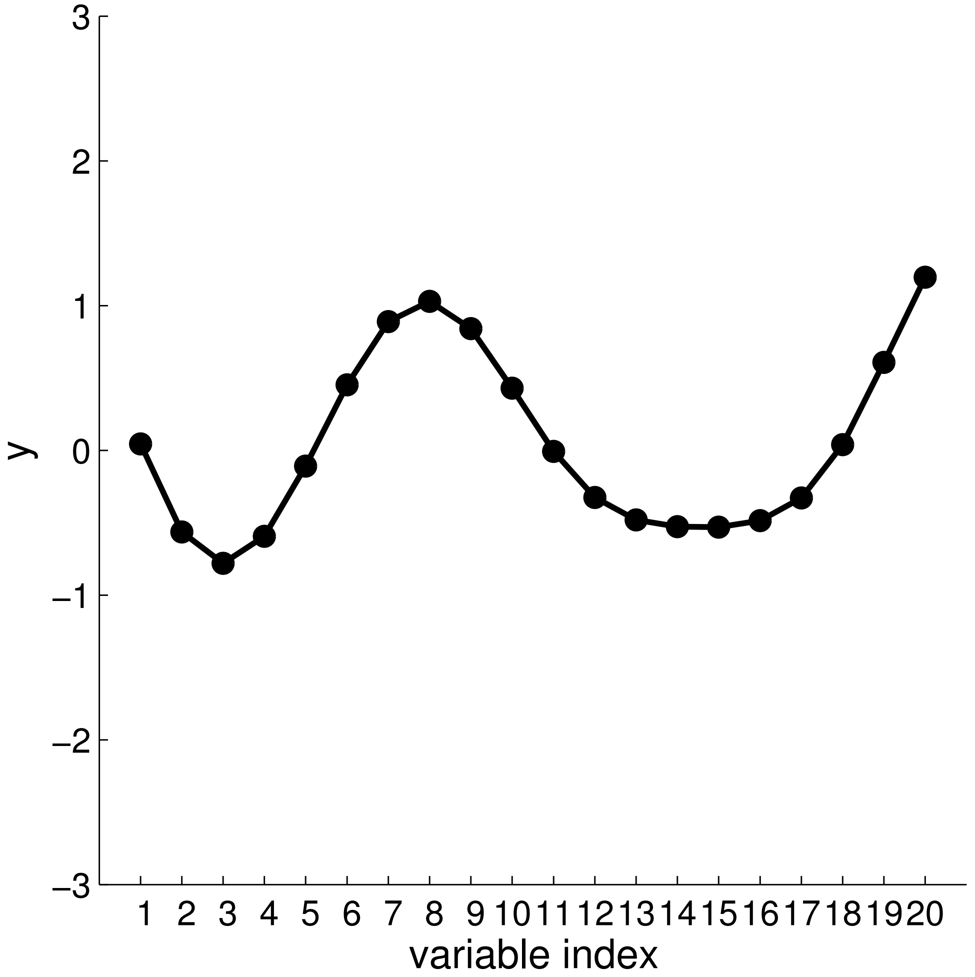


$\Sigma =$

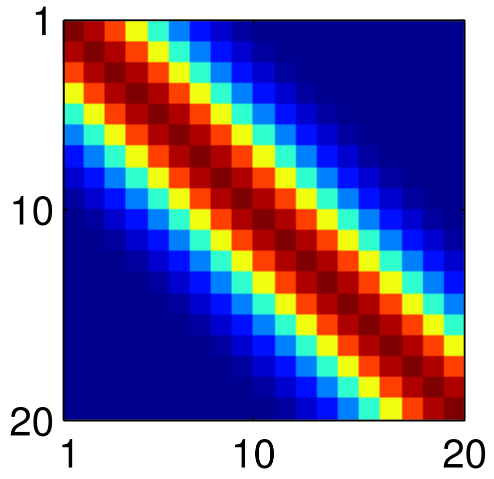


New visualisation

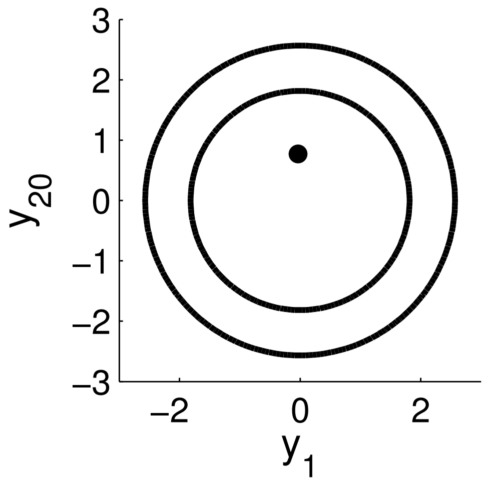




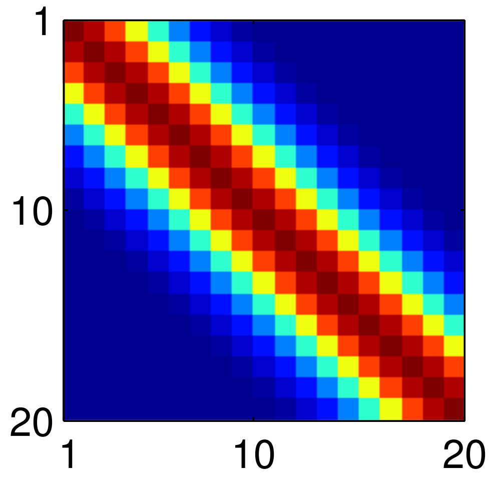
$\Sigma =$

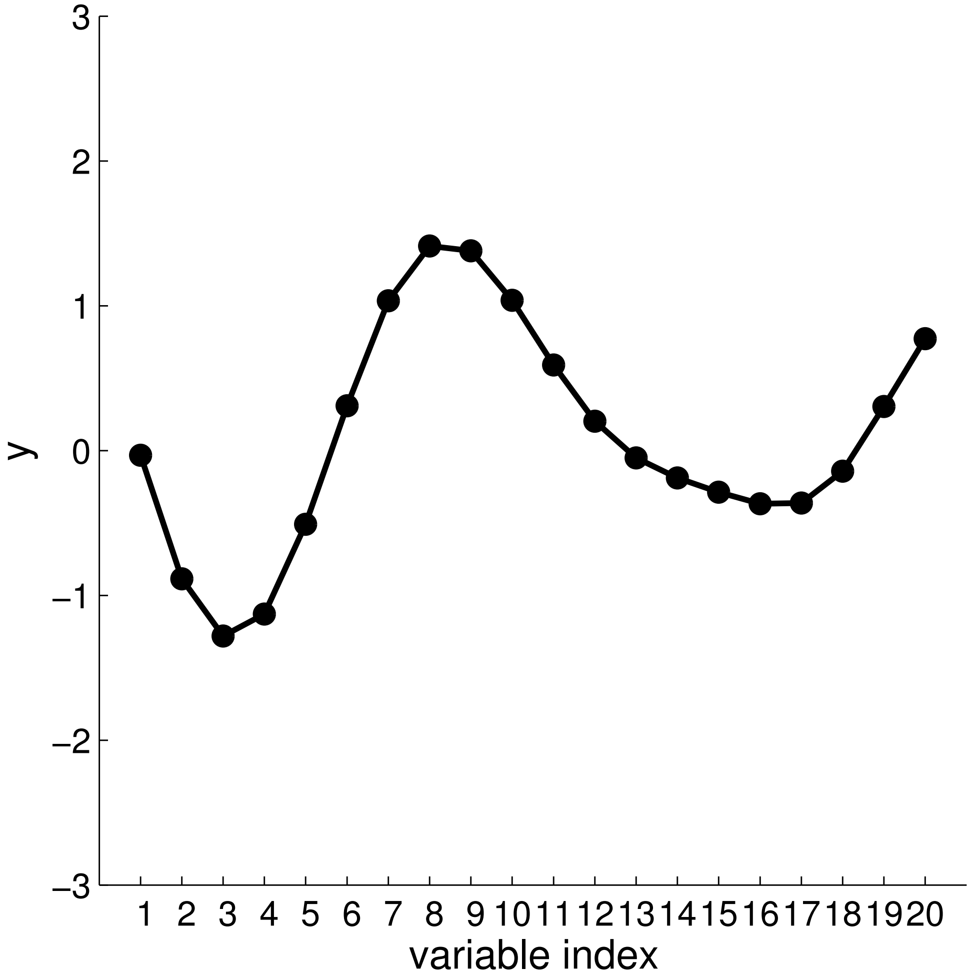


New visualisation

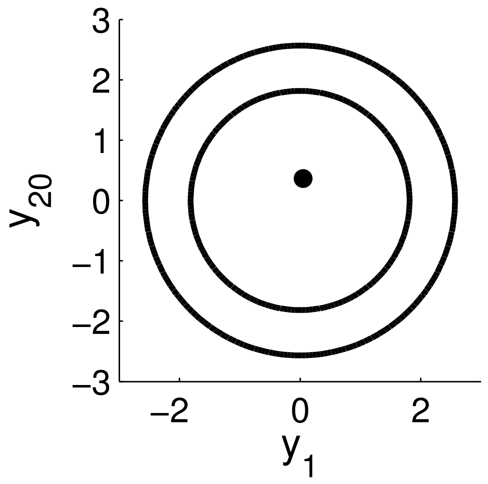


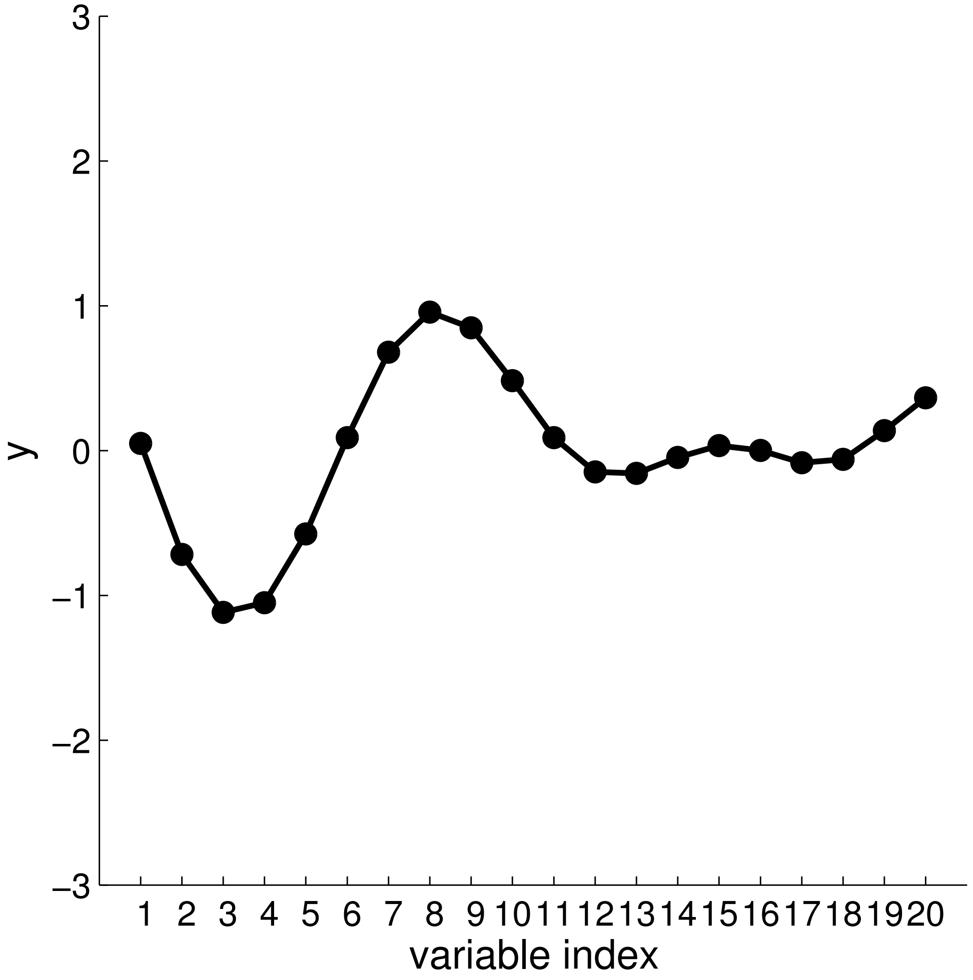
$\Sigma =$



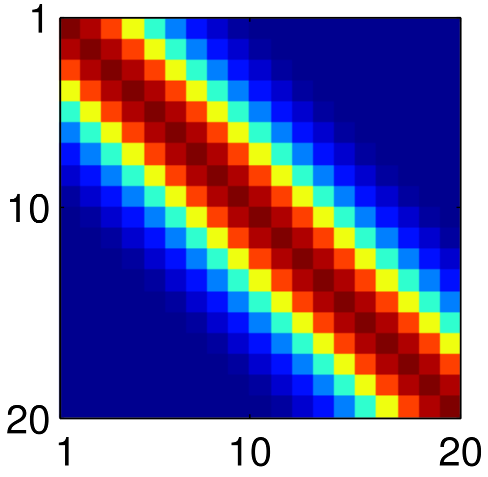


New visualisation

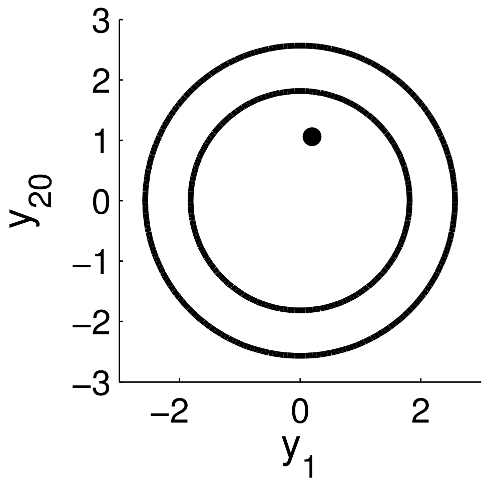


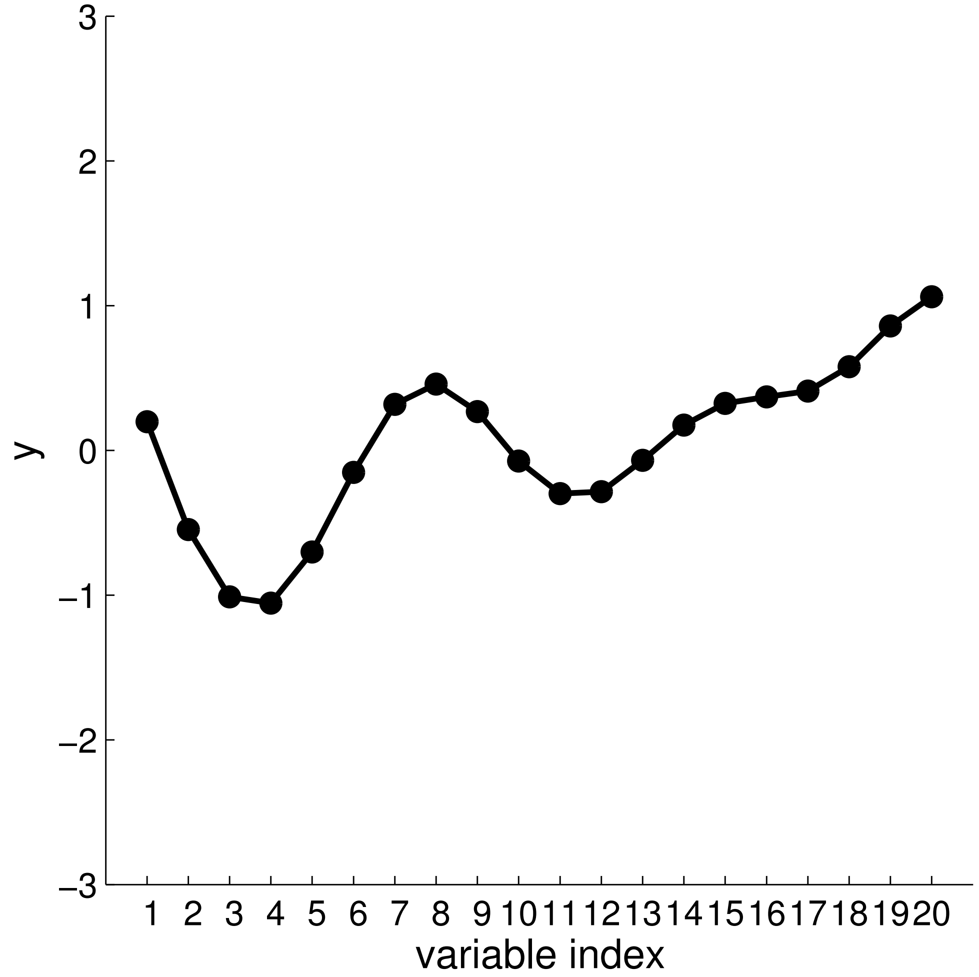


$\Sigma =$

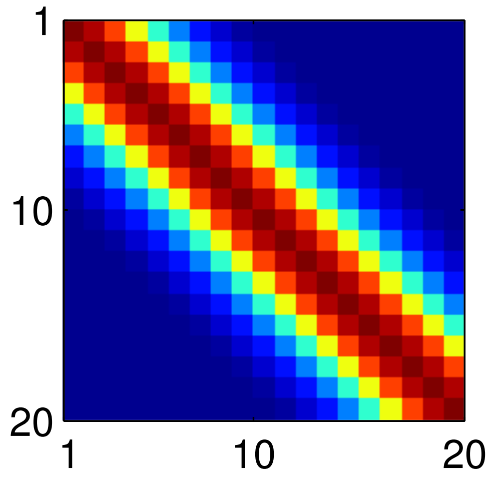


New visualisation

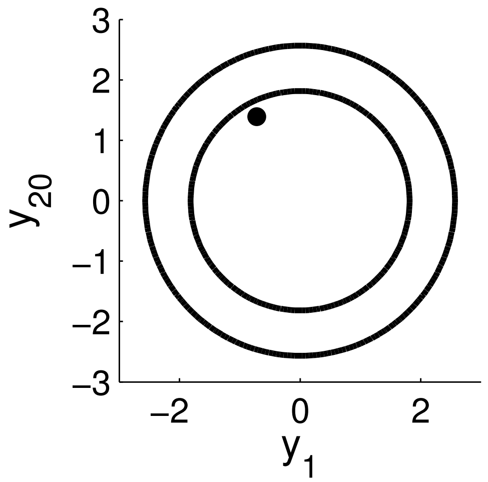




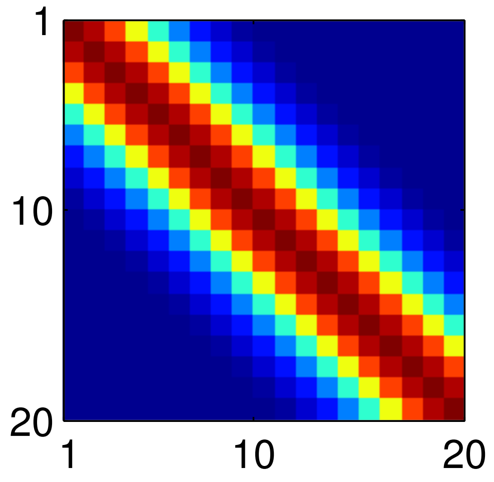
$\Sigma =$

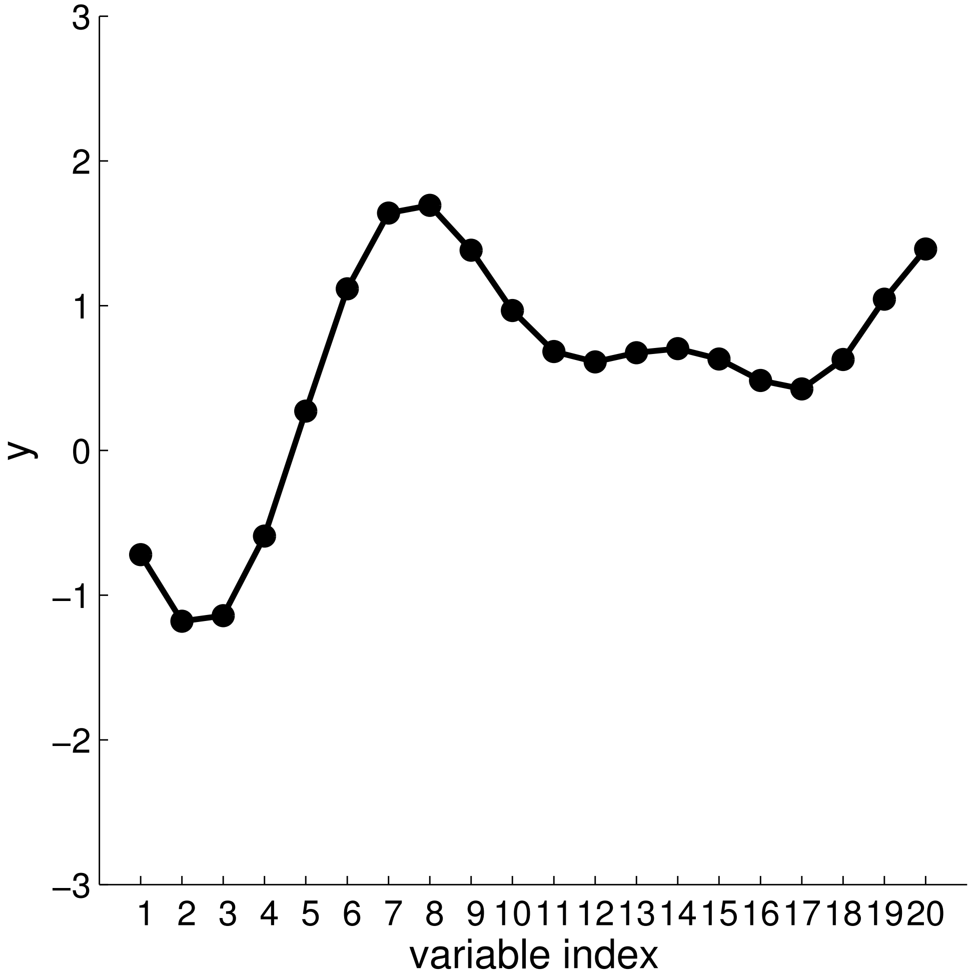


New visualisation

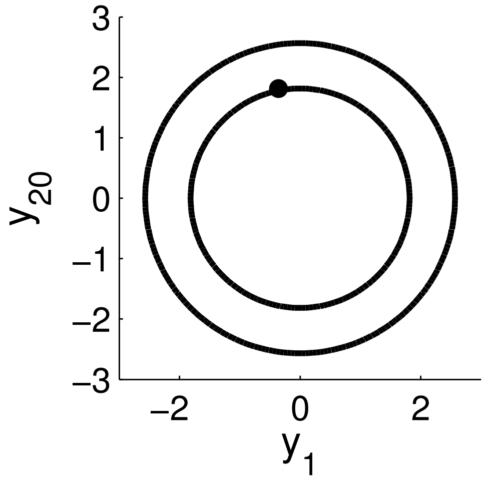


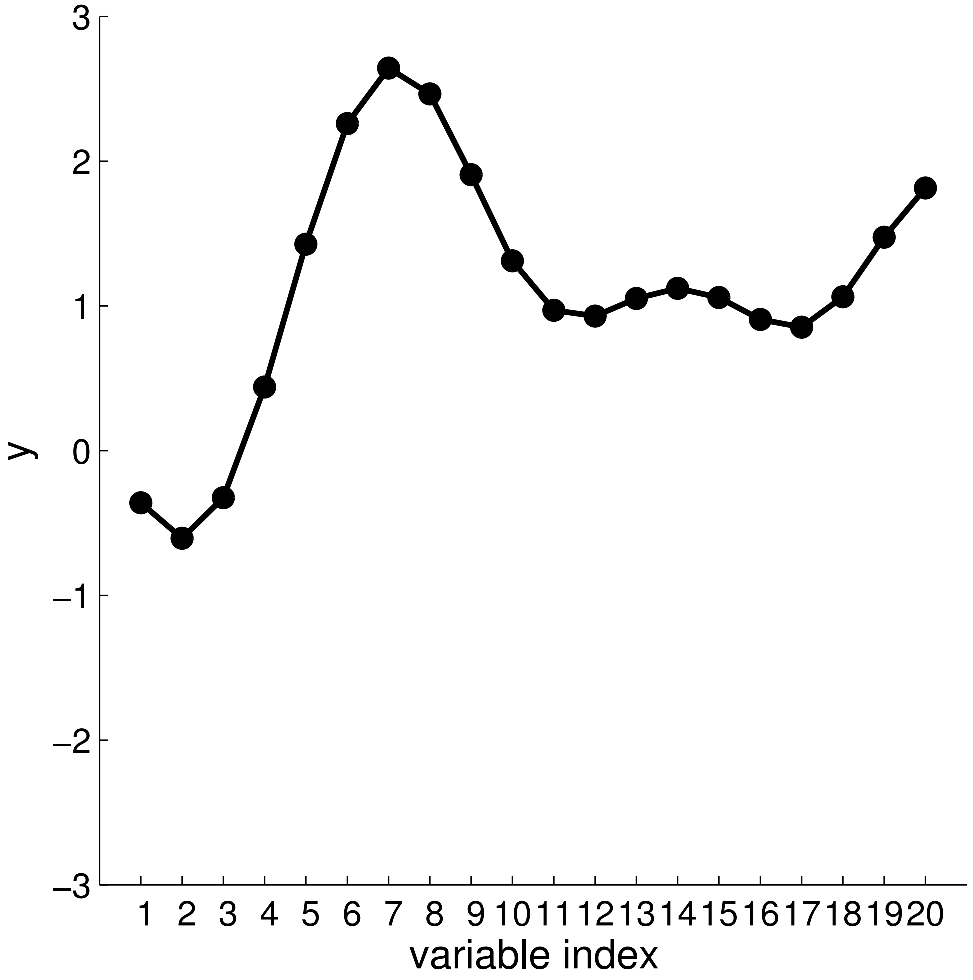
$\Sigma =$



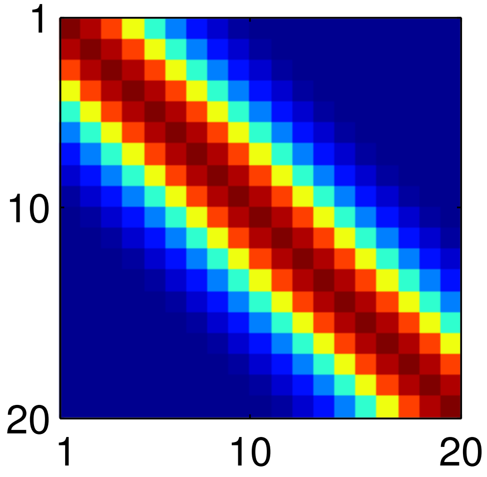


New visualisation

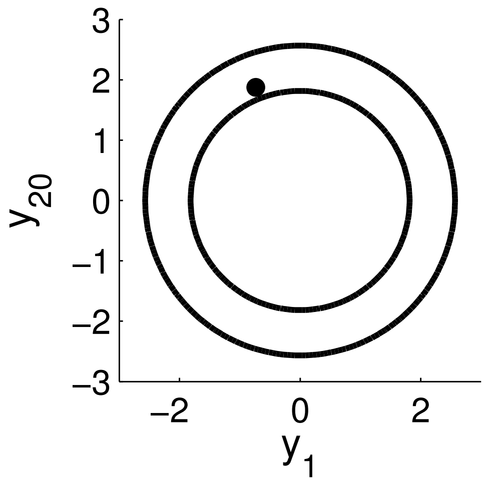




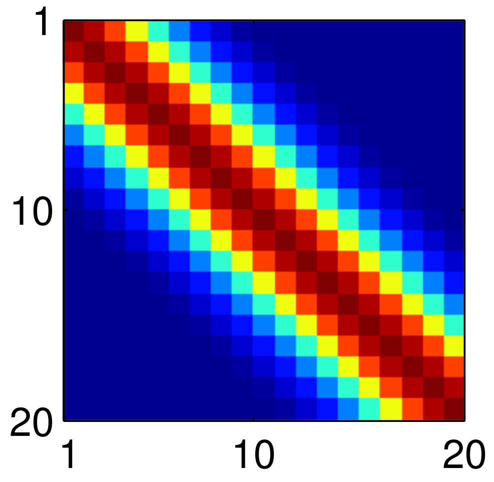
$\Sigma =$

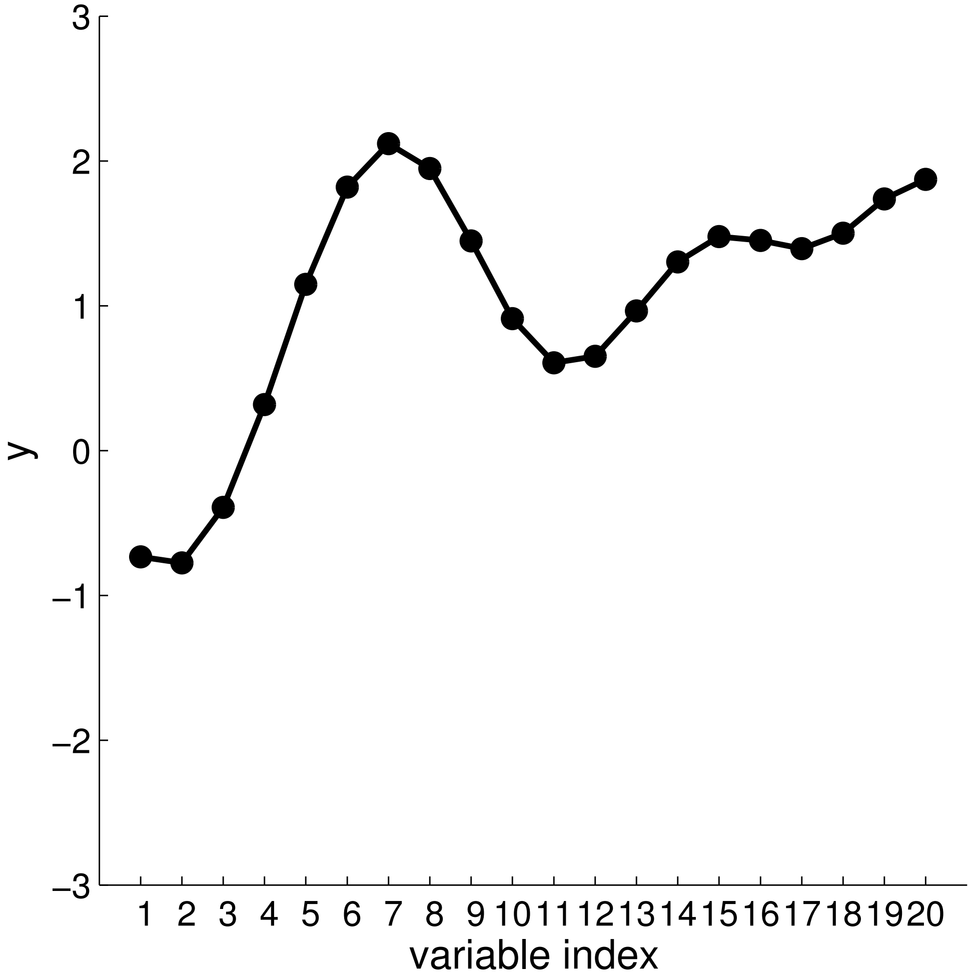


New visualisation

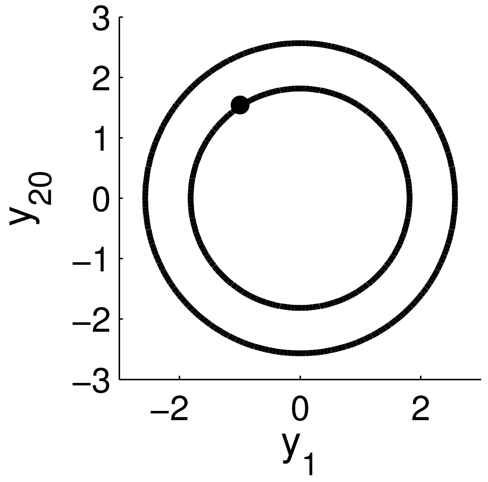


$$\Sigma =$$

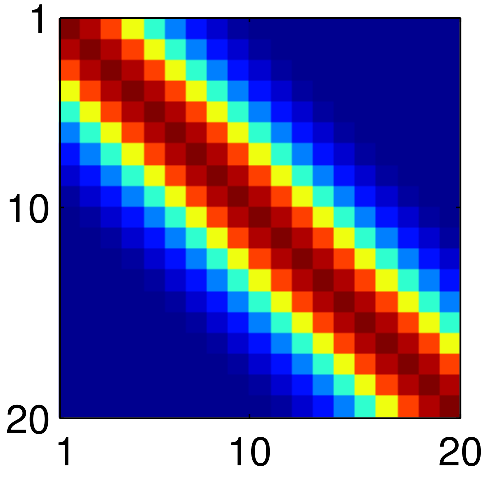


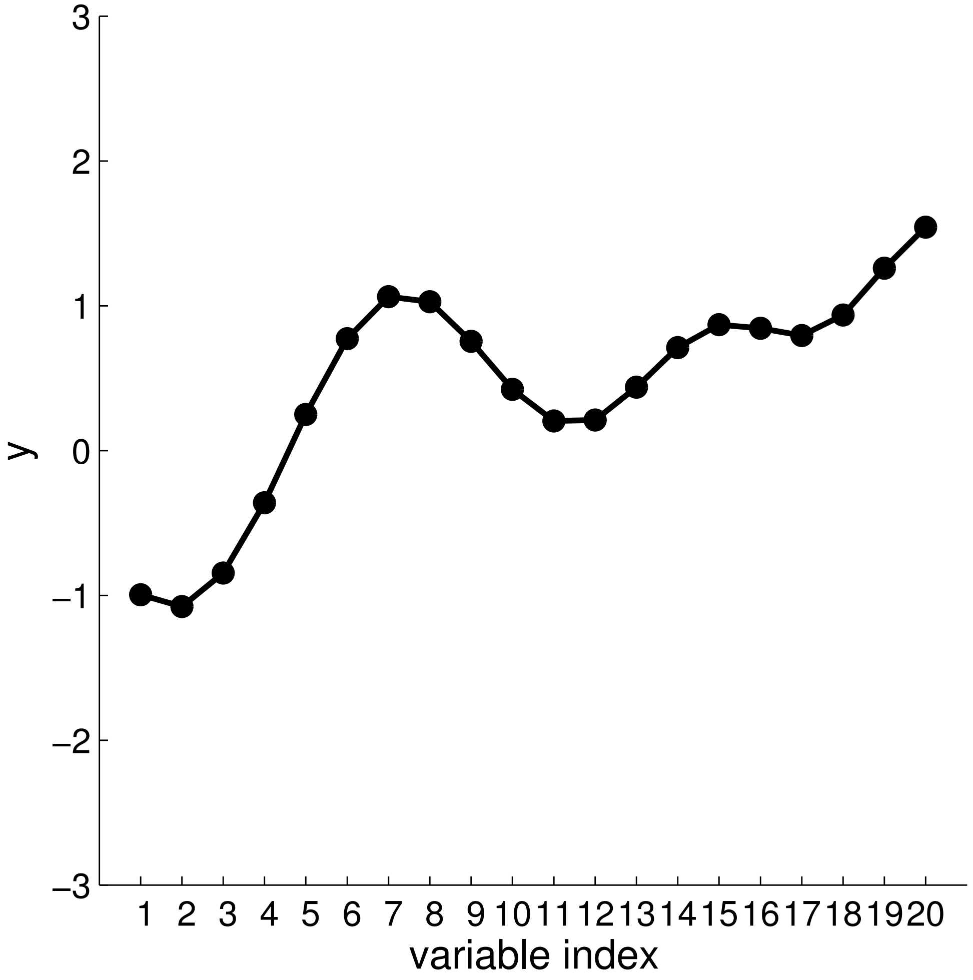


New visualisation

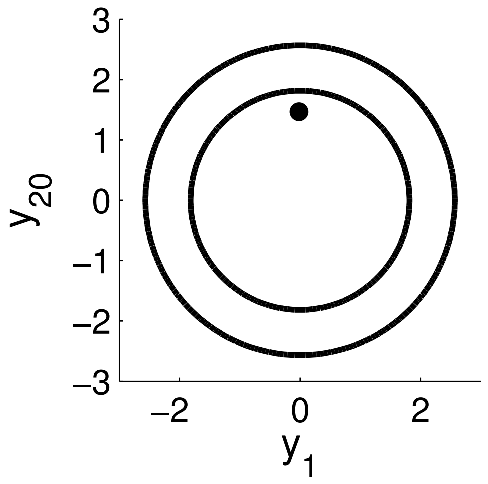


$\Sigma =$

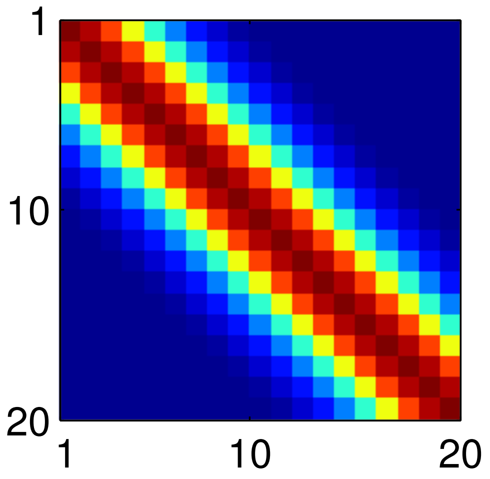


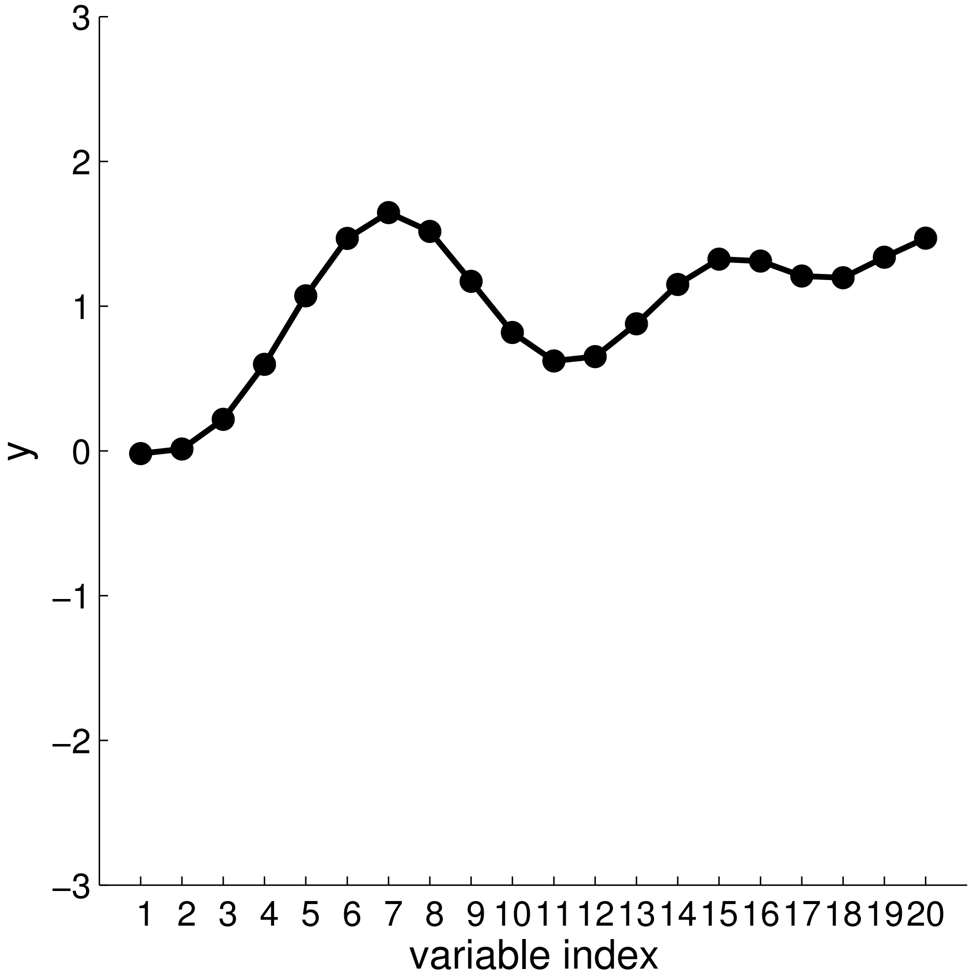


New visualisation

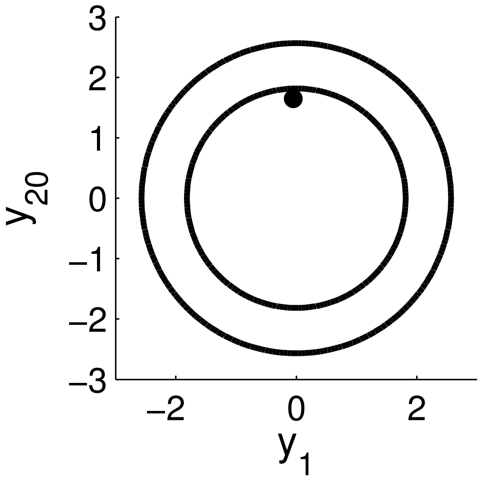


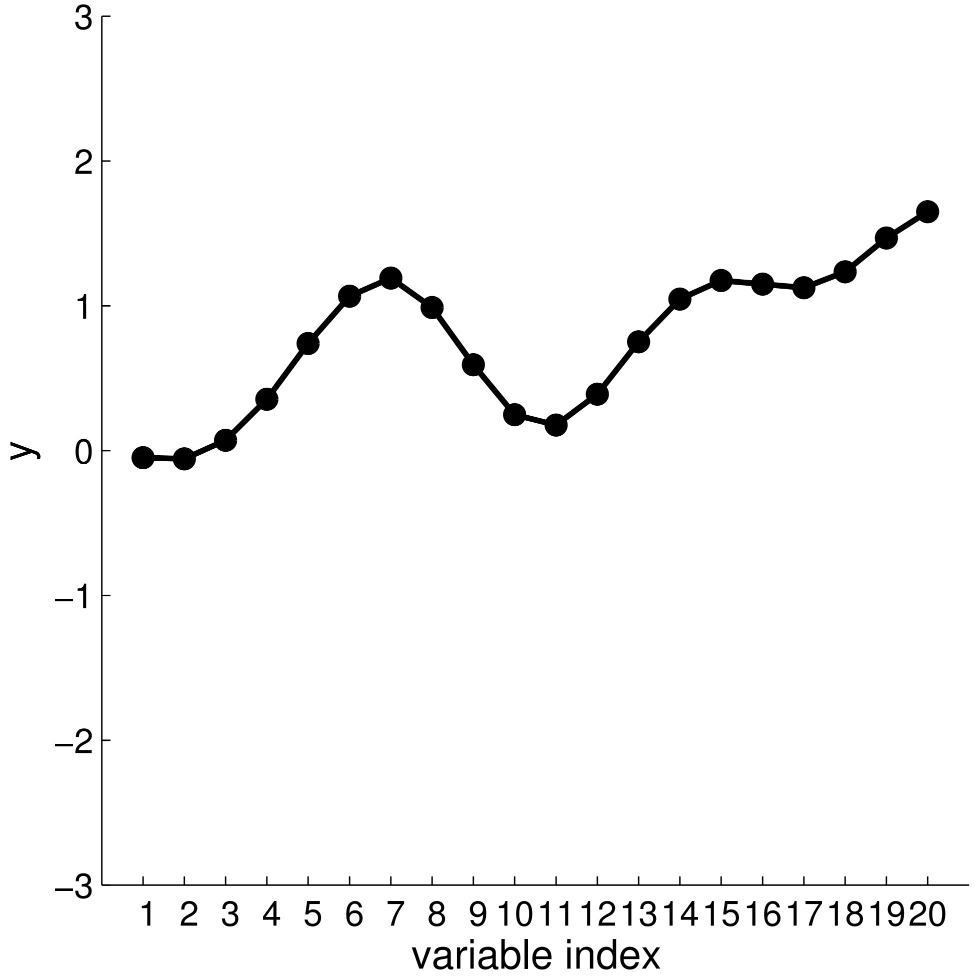
$\Sigma =$



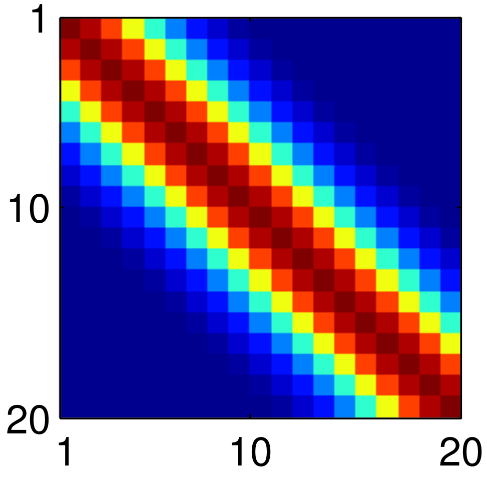


New visualisation

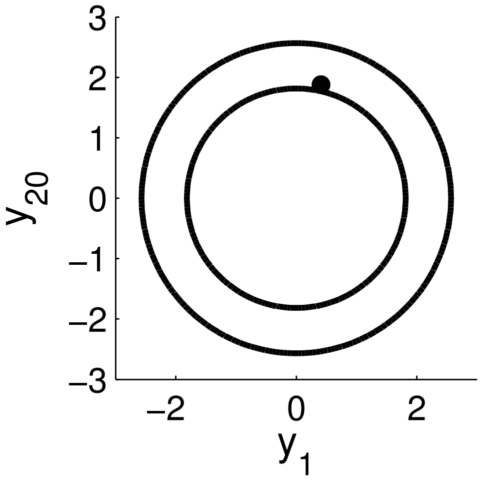




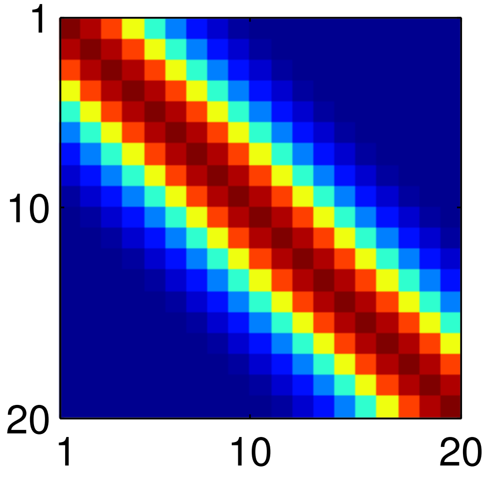
$\Sigma =$

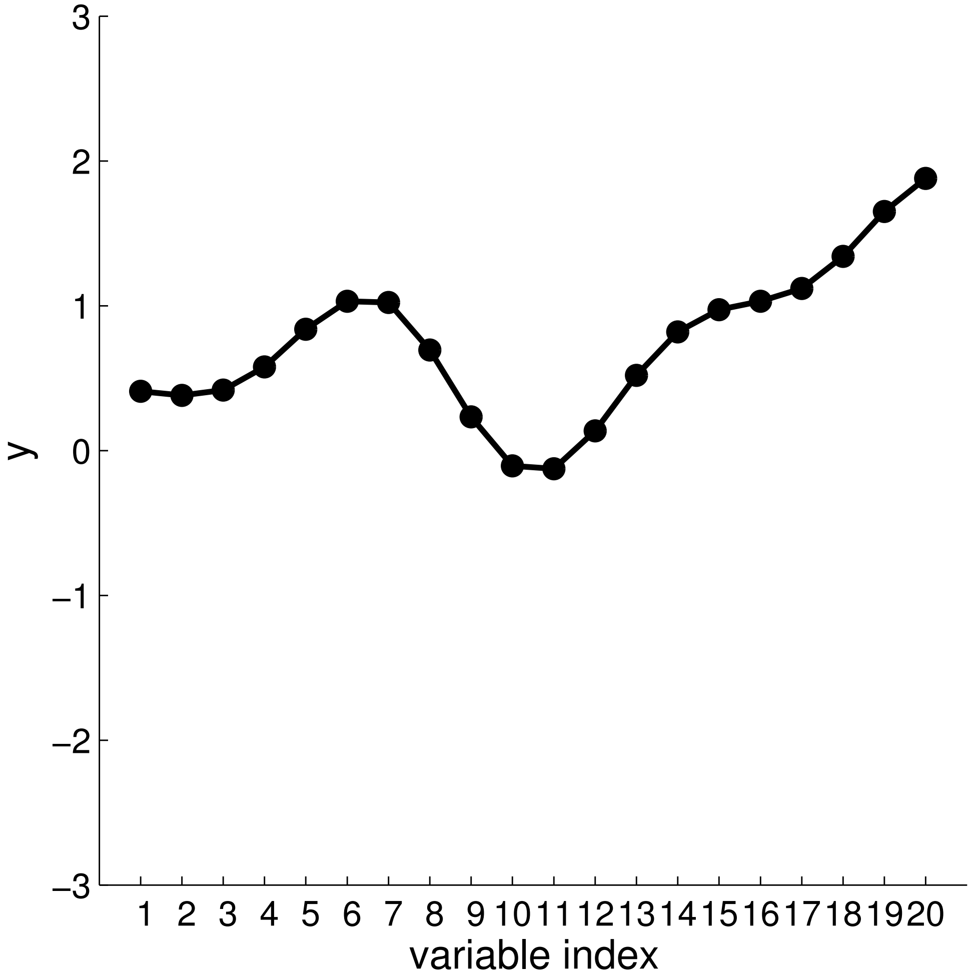


New visualisation

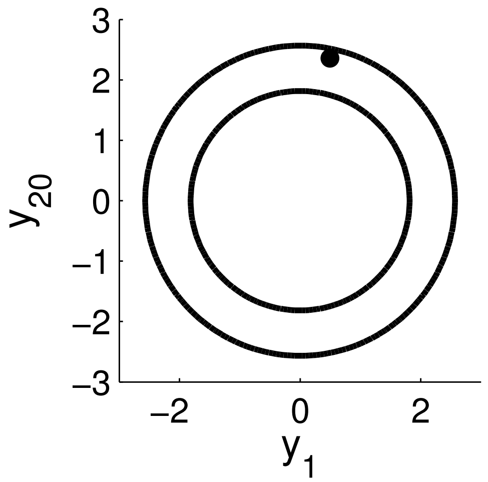


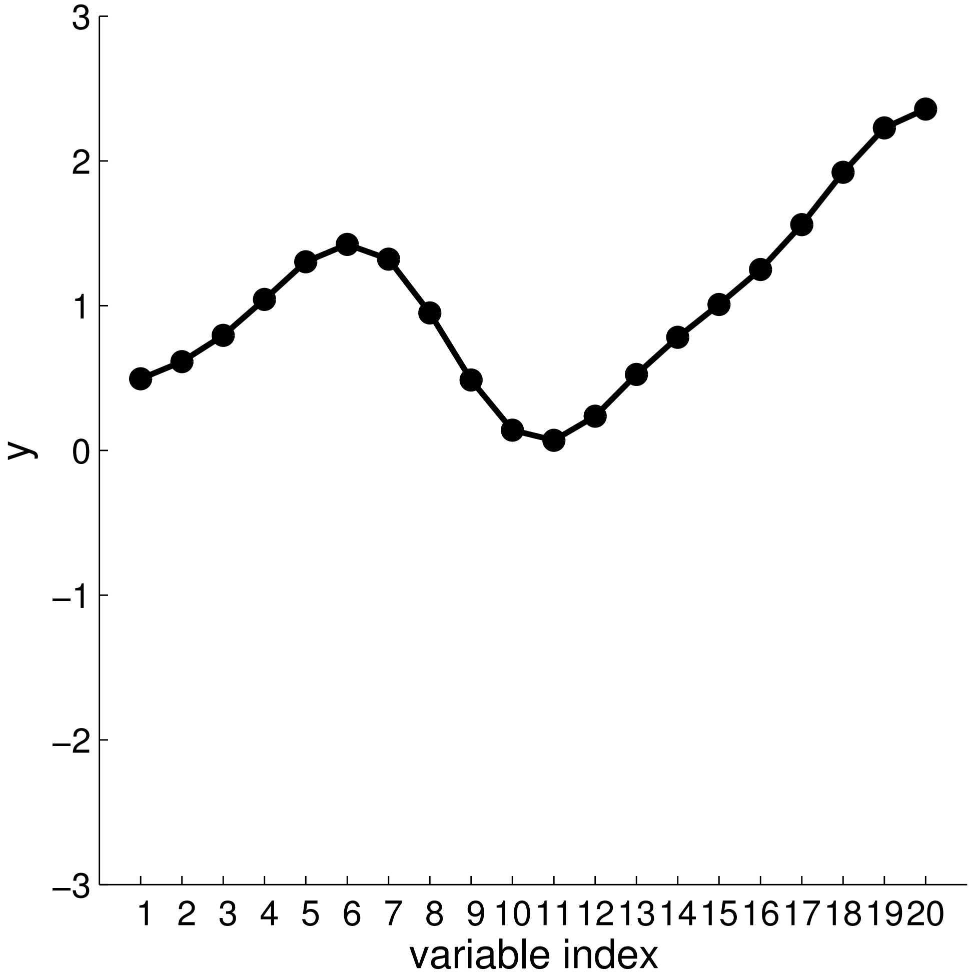
$\Sigma =$



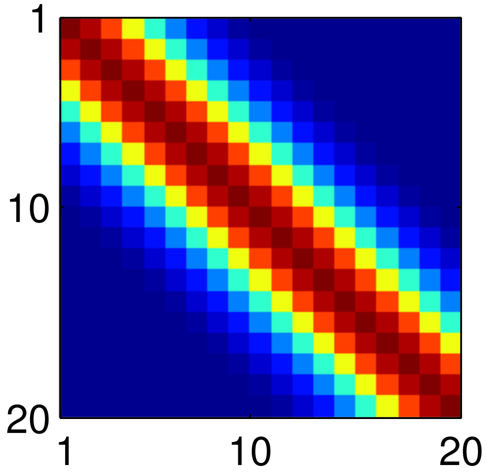


New visualisation

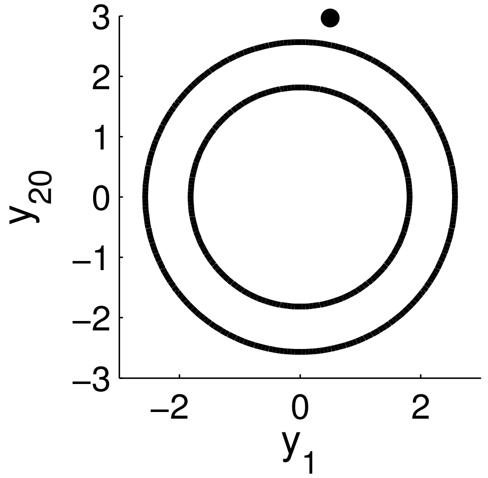


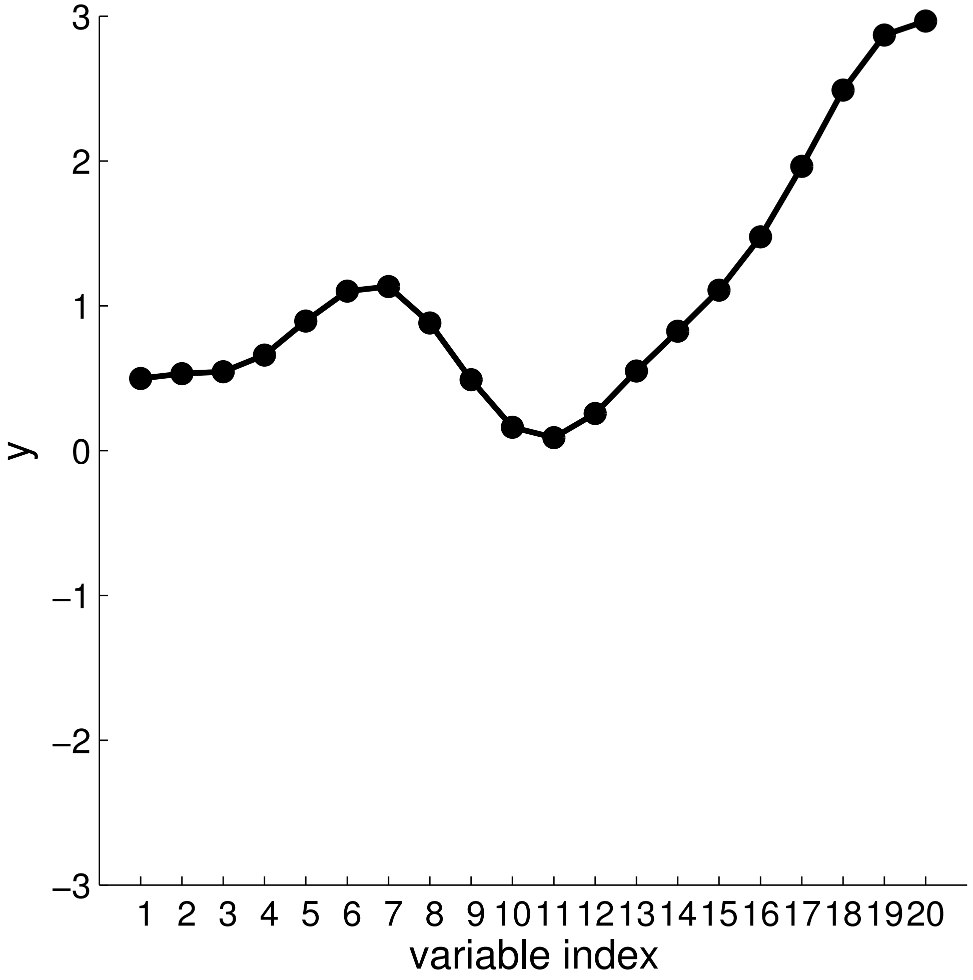


$\Sigma =$

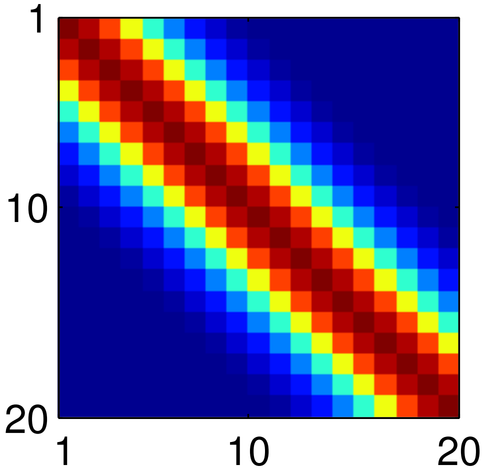


New visualisation

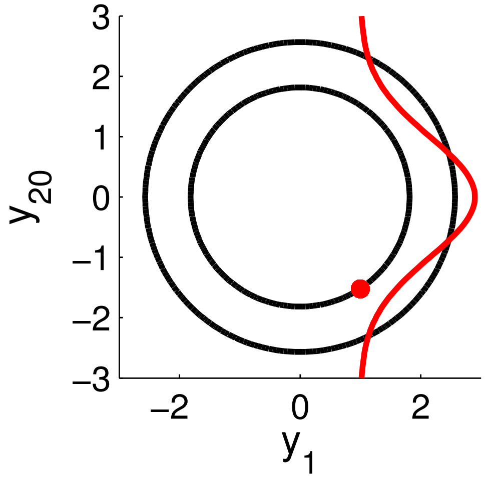


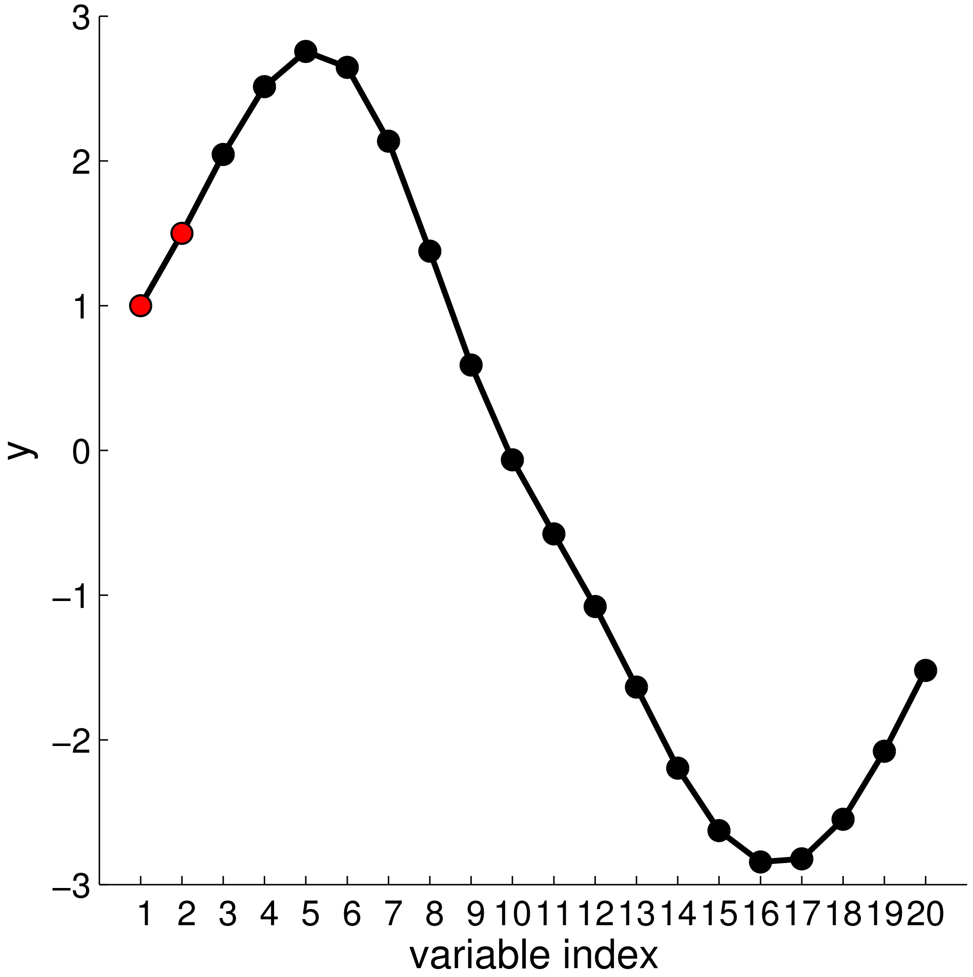


$\Sigma =$

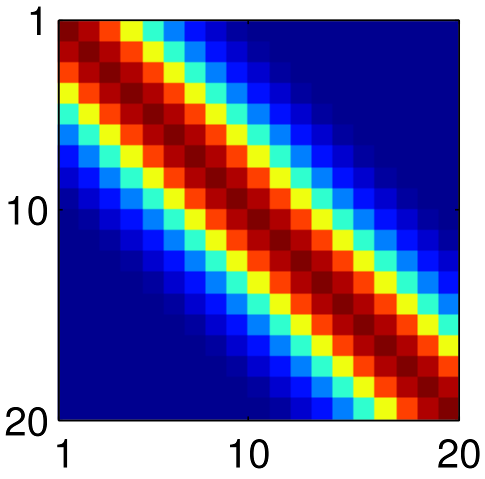


New visualisation



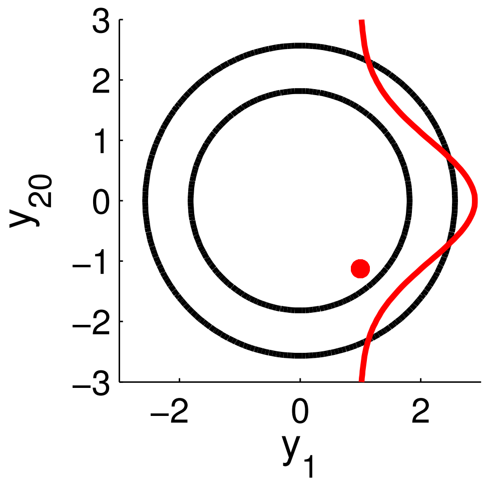


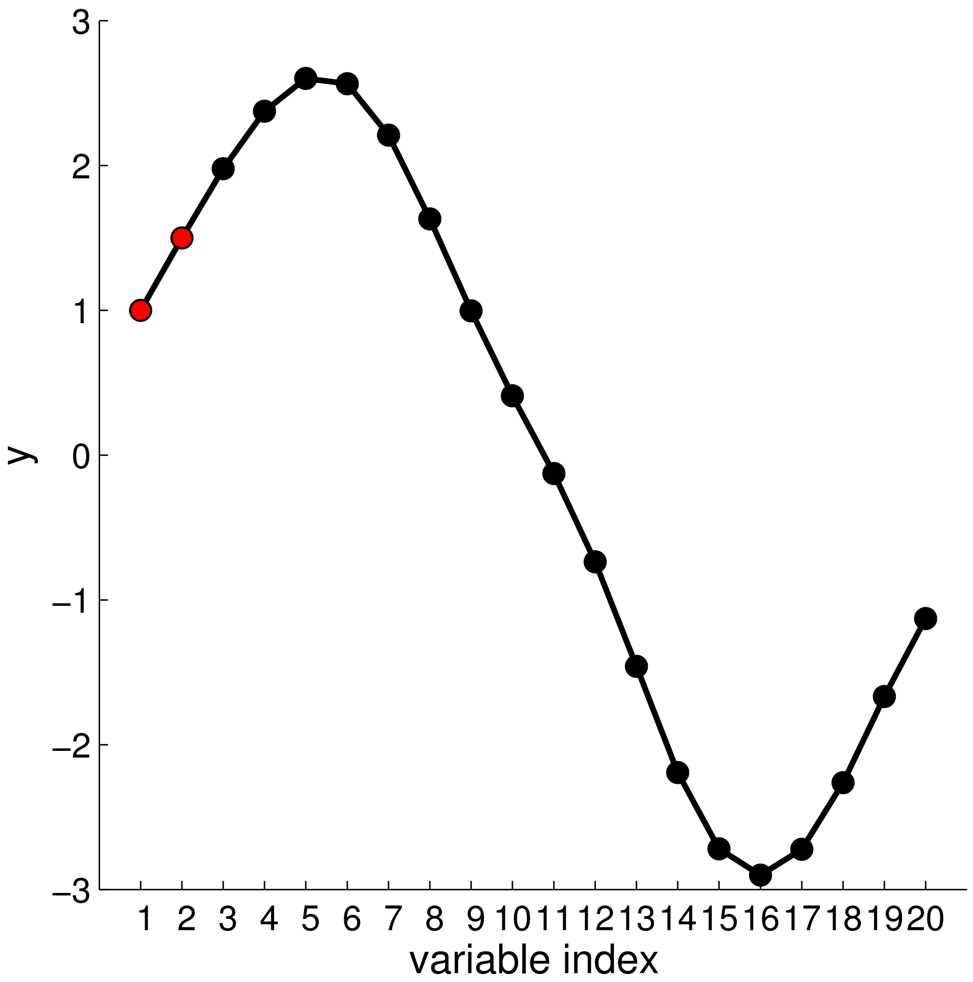
$\Sigma =$



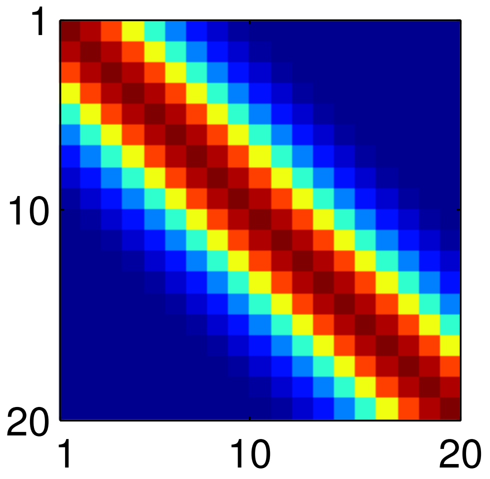
Conditioning on y1 and y2

New visualisation



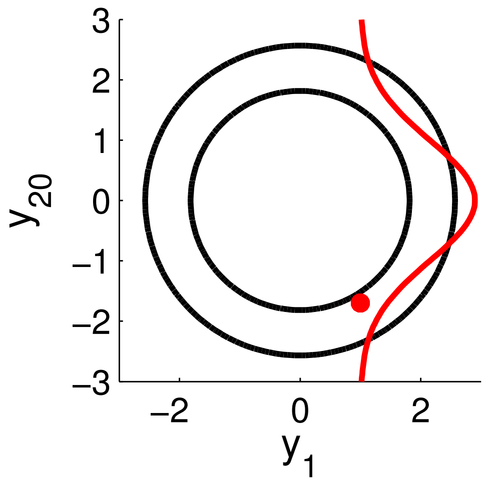


$\Sigma =$

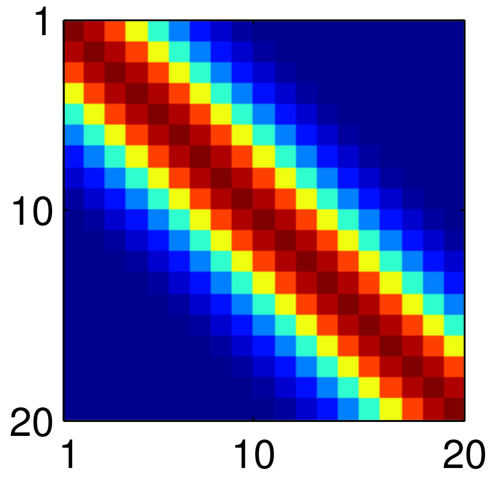


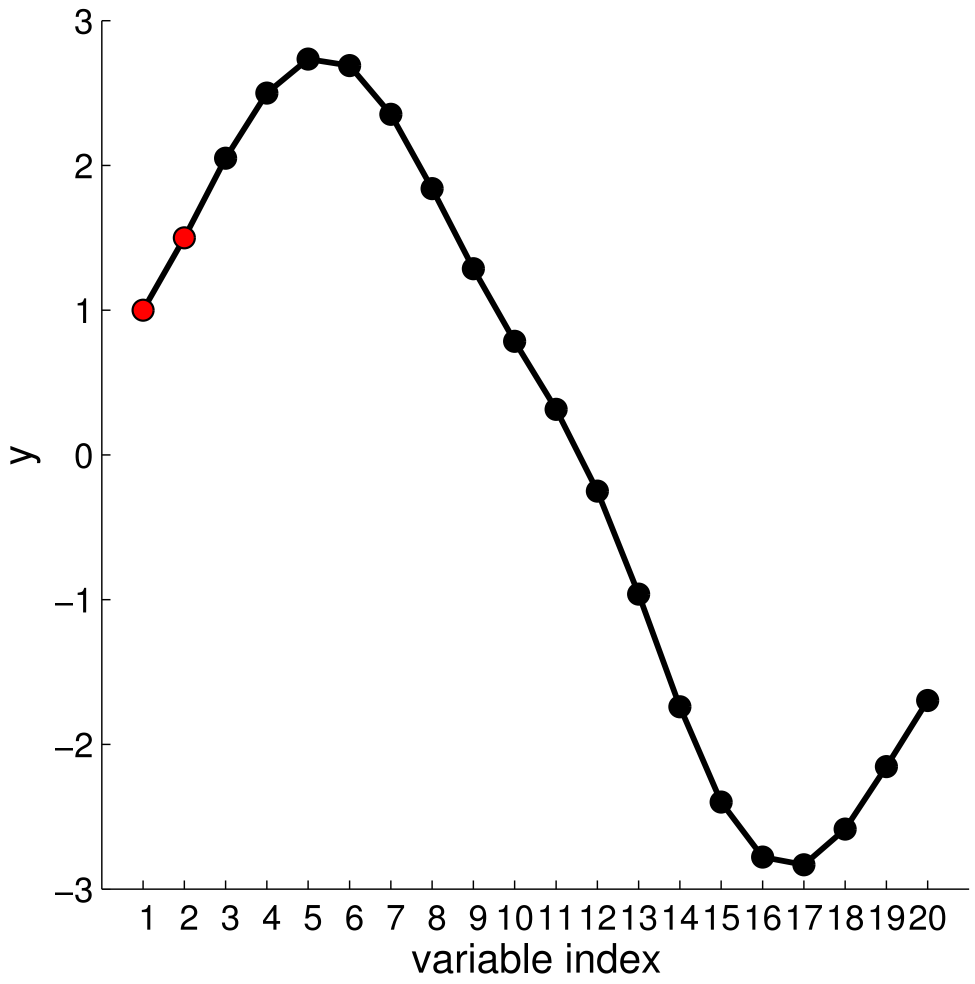
Conditioning on y1 and y2

New visualisation



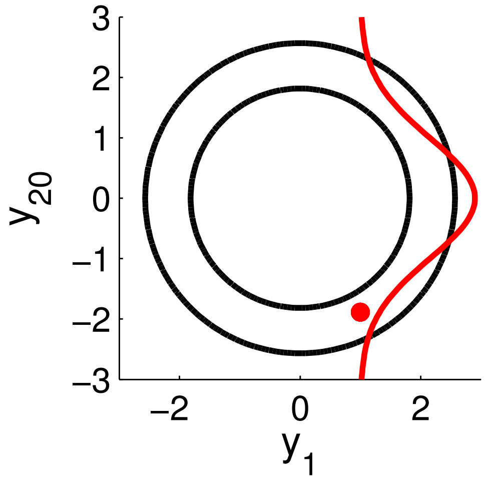
$\Sigma =$



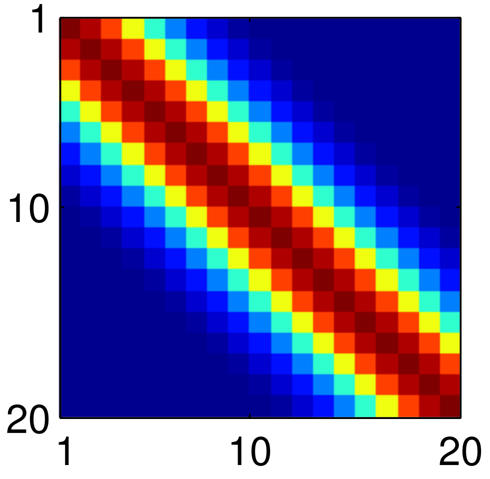


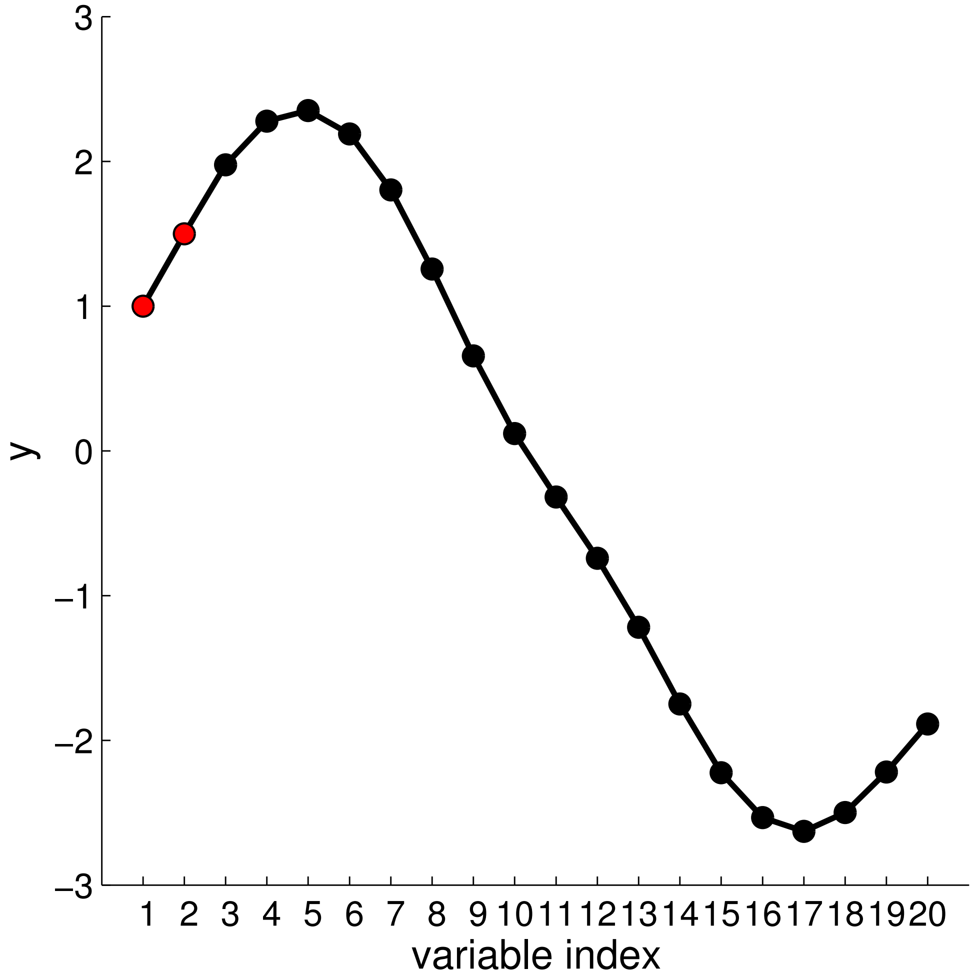
Conditioning on y1 and y2

New visualisation



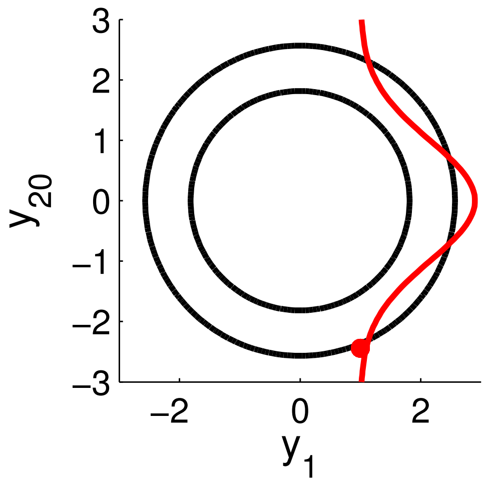
$\Sigma =$

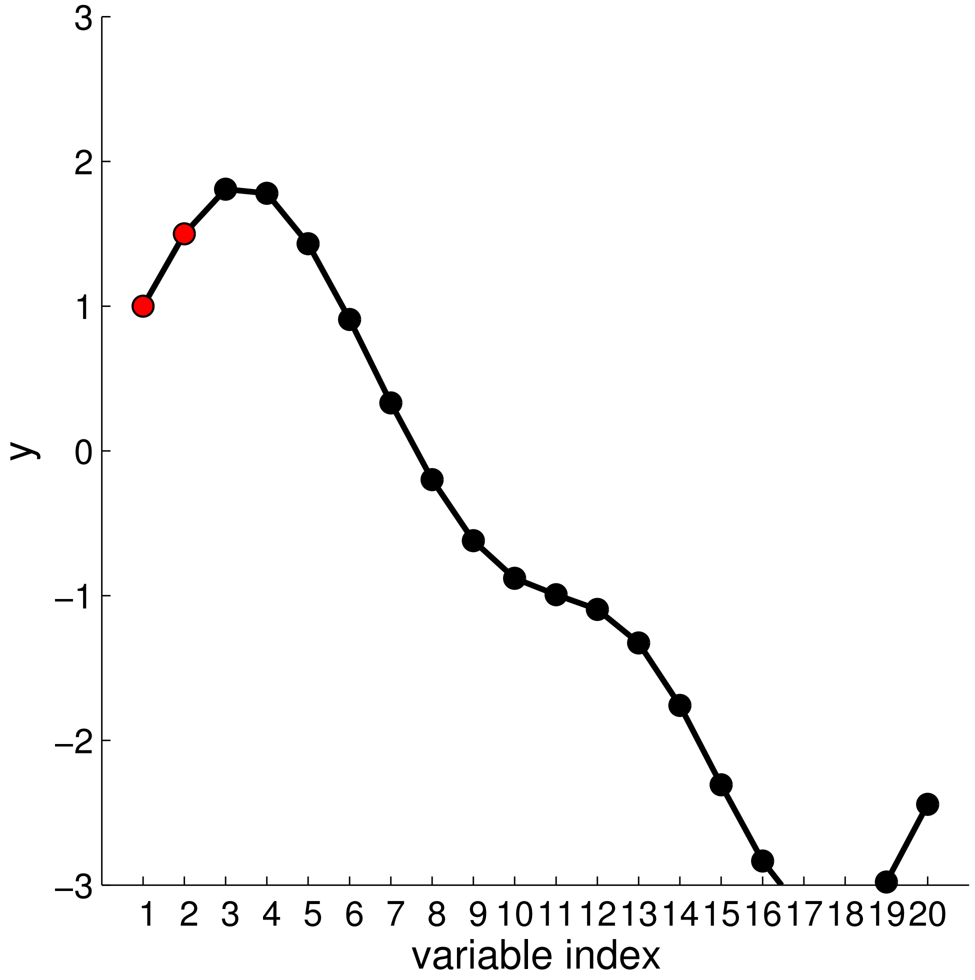




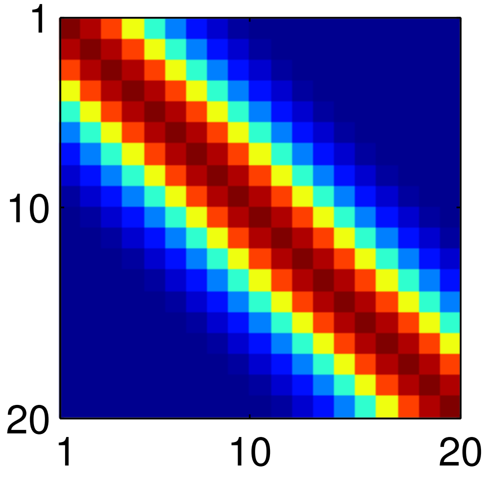
Conditioning on y1 and y2

New visualisation



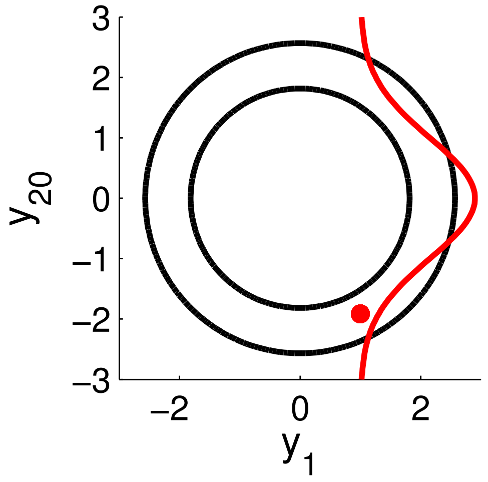


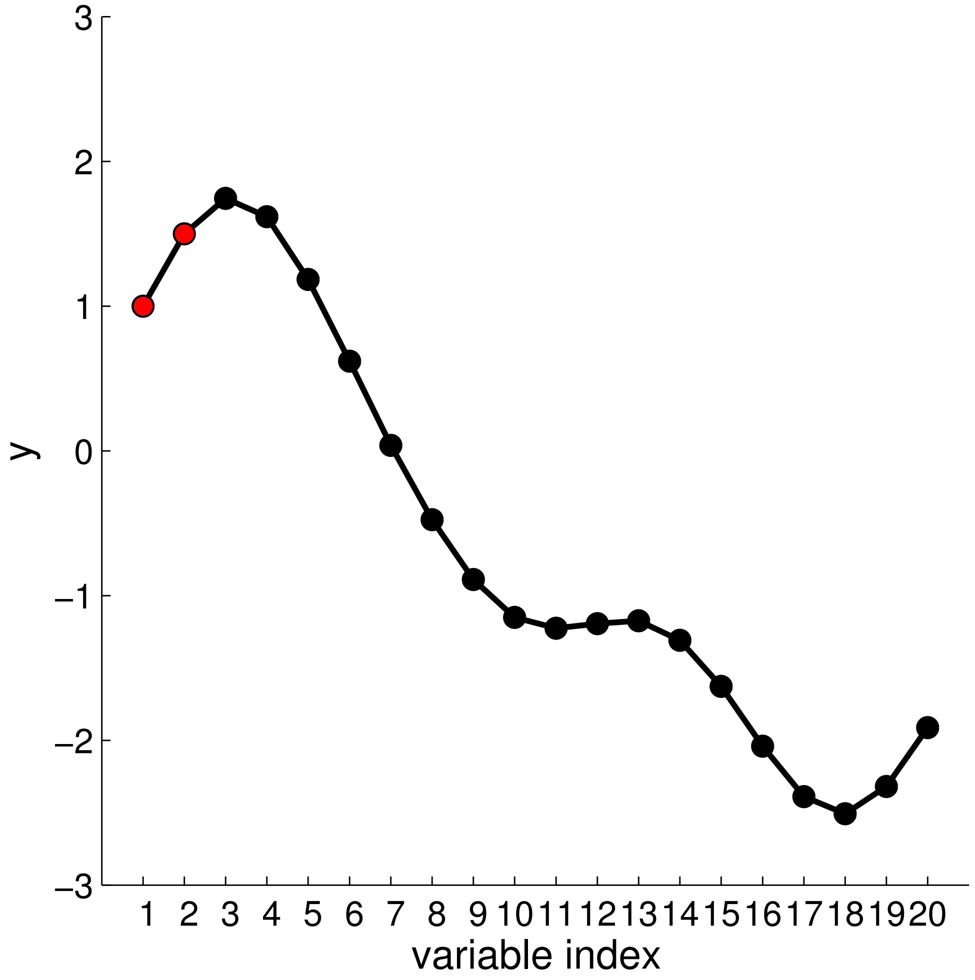
$\Sigma =$



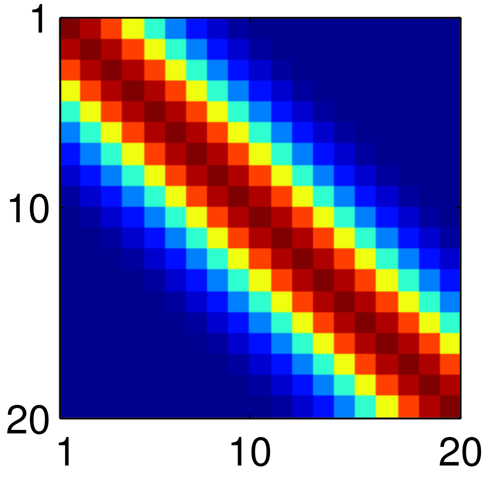
Conditioning on y1 and y2

New visualisation



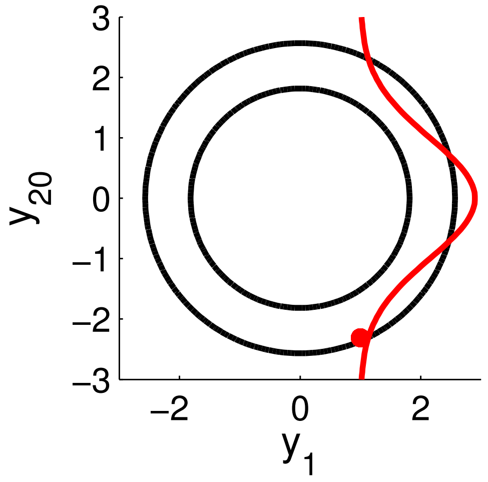


$\Sigma =$

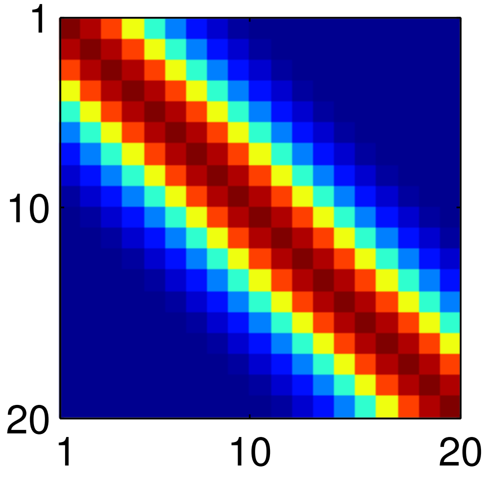


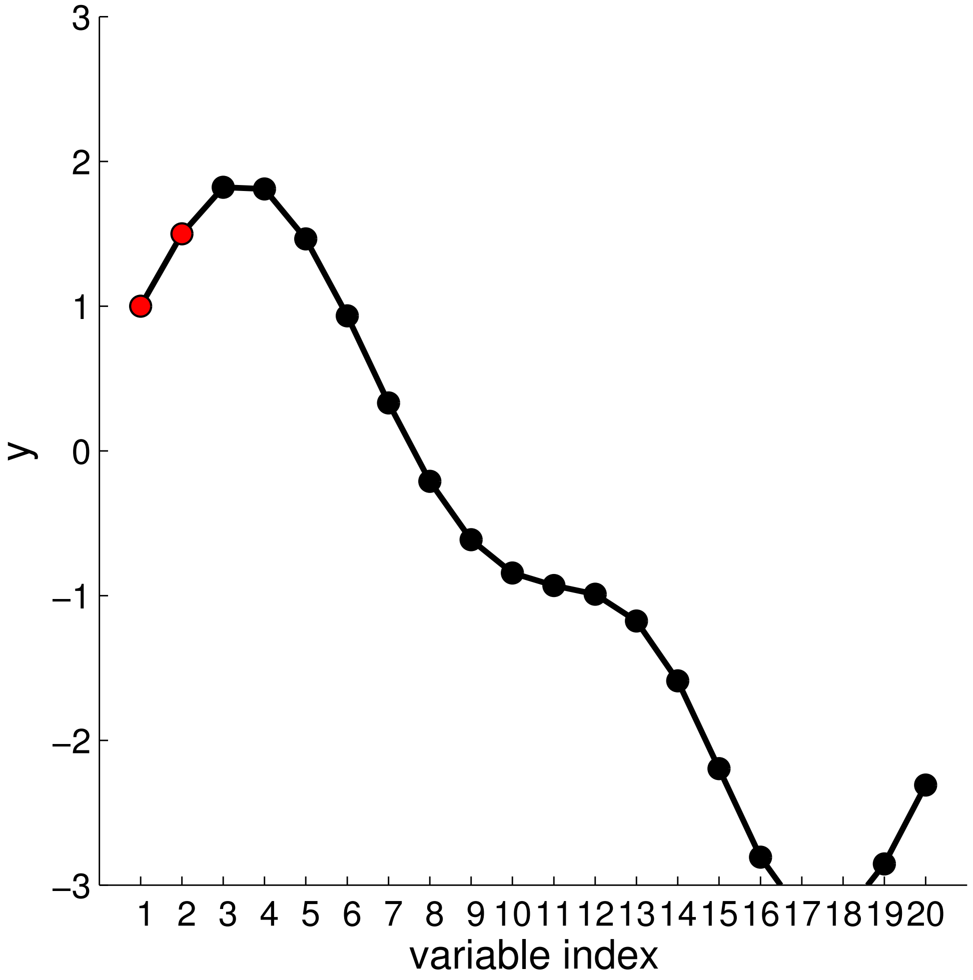
Conditioning on y1 and y2

New visualisation



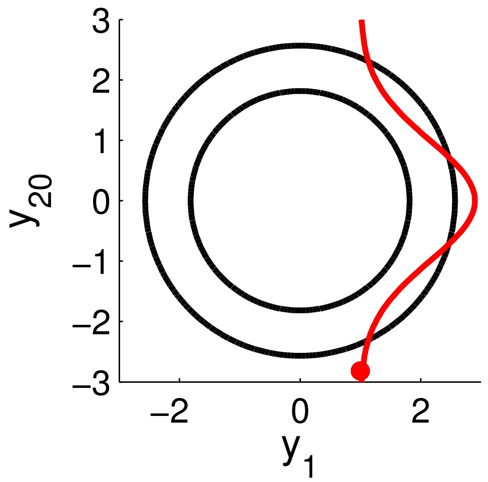
$\Sigma =$

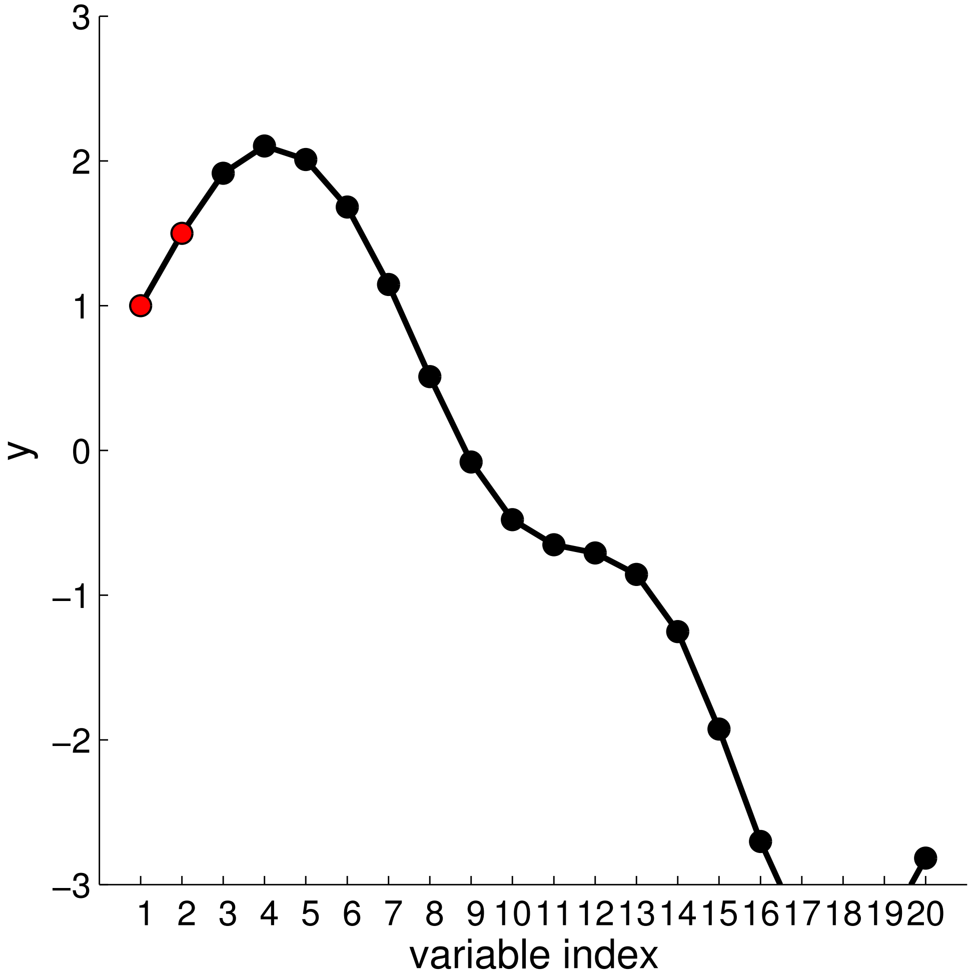




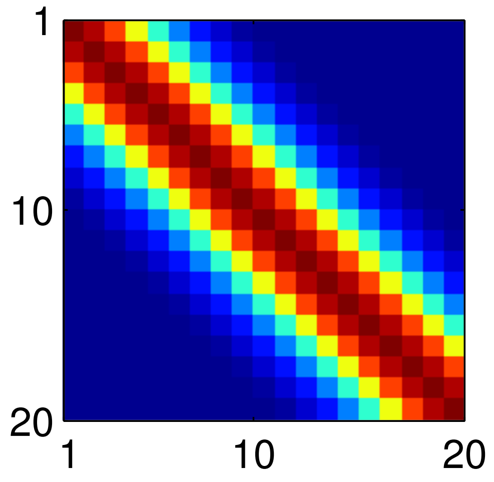
Conditioning on y1 and y2

New visualisation



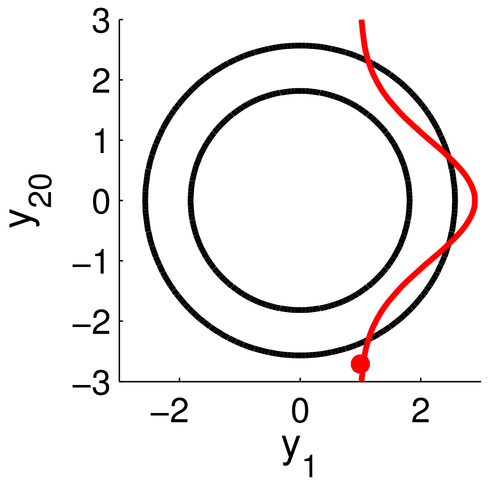


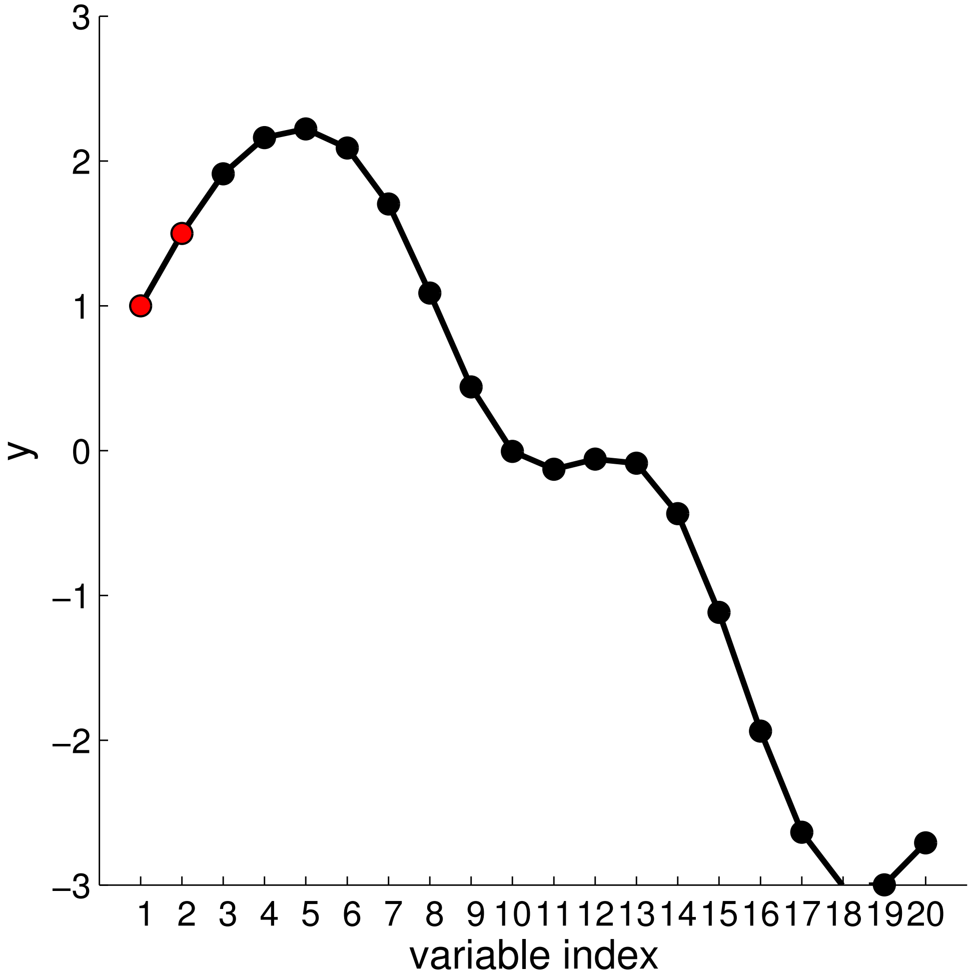
$\Sigma =$



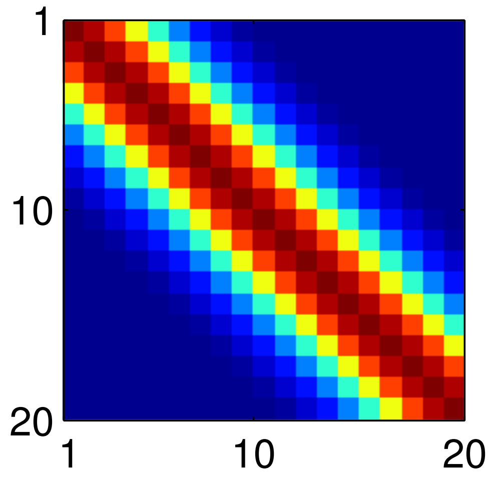
Conditioning on y1 and y2

New visualisation



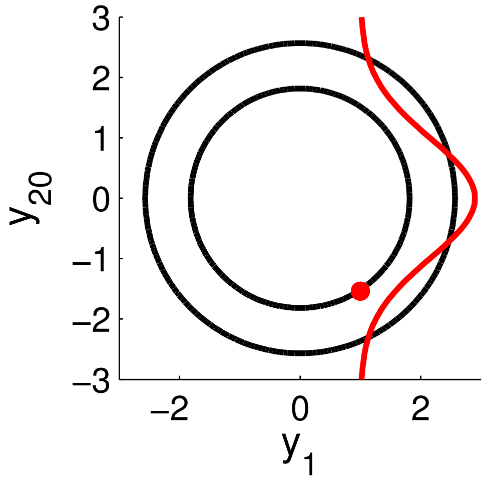


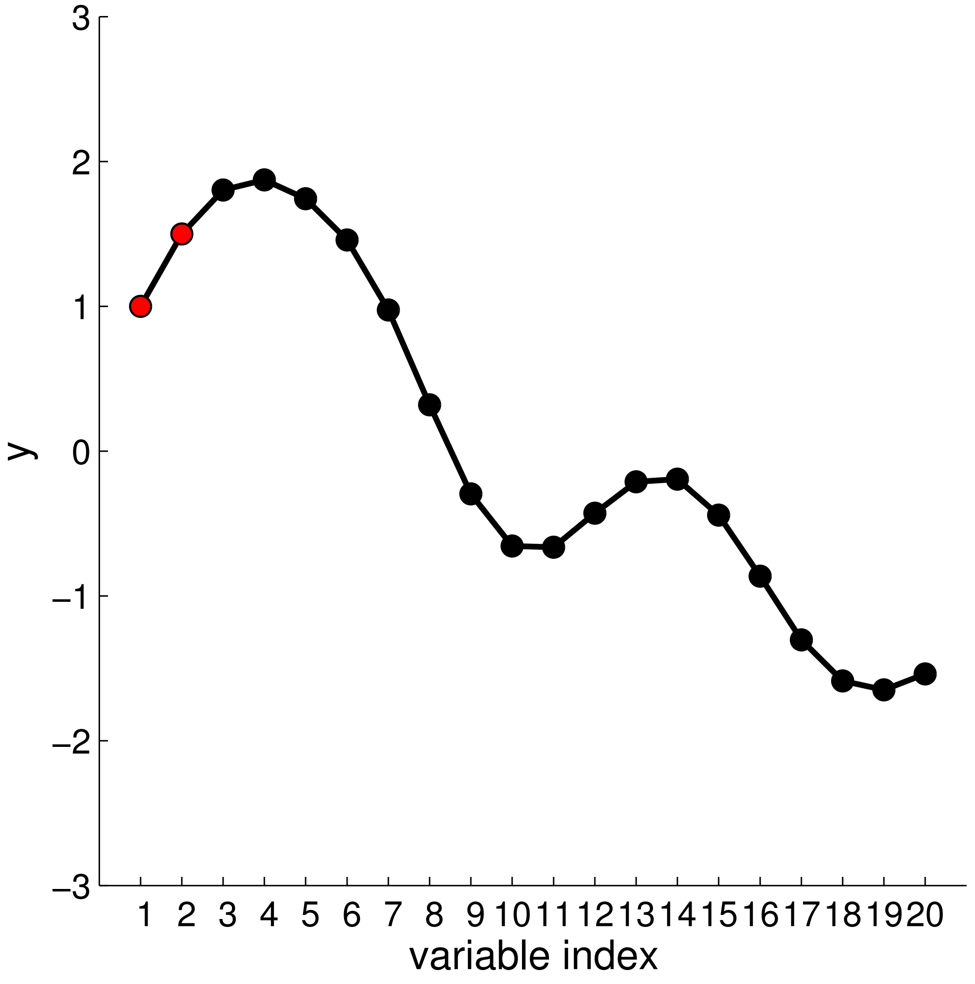
$\Sigma =$



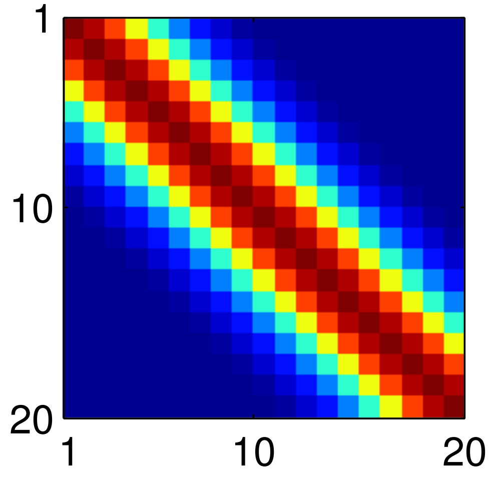
Conditioning on y1 and y2

New visualisation



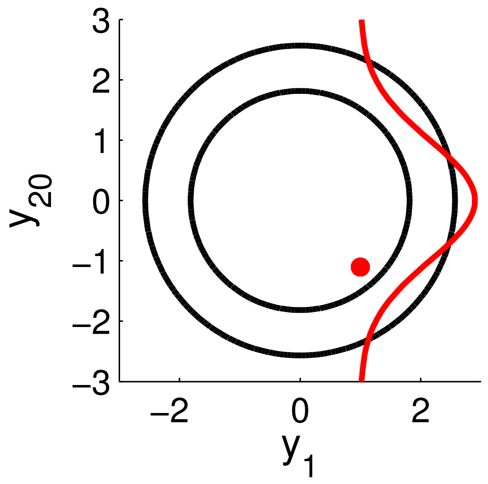


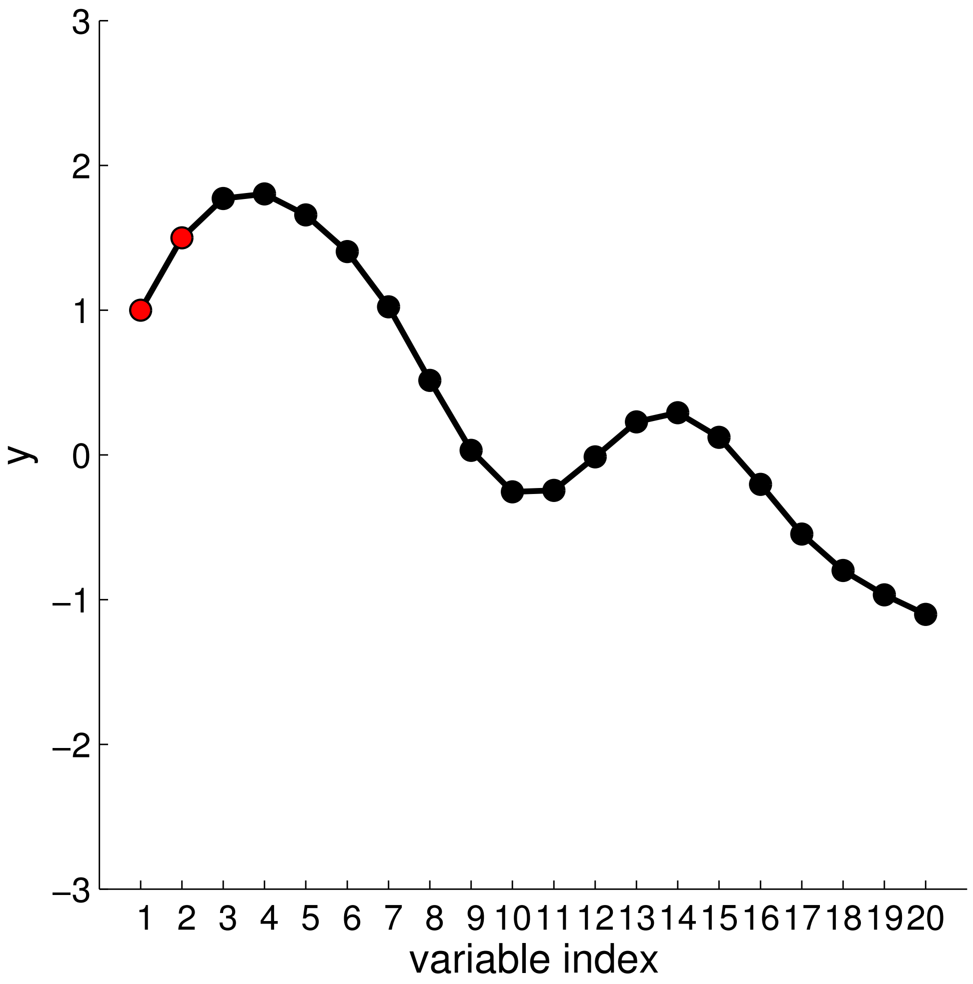
$$\Sigma =$$



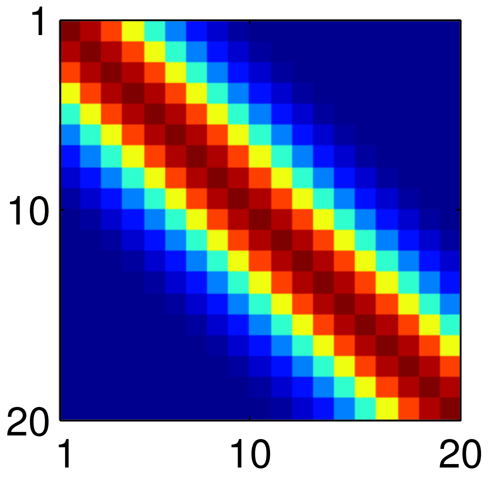
Conditioning on y1 and y2

New visualisation



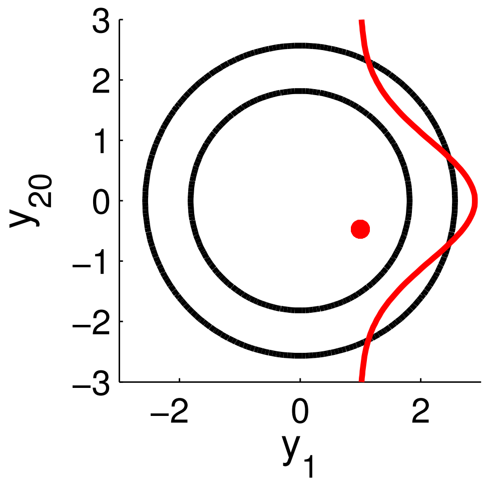


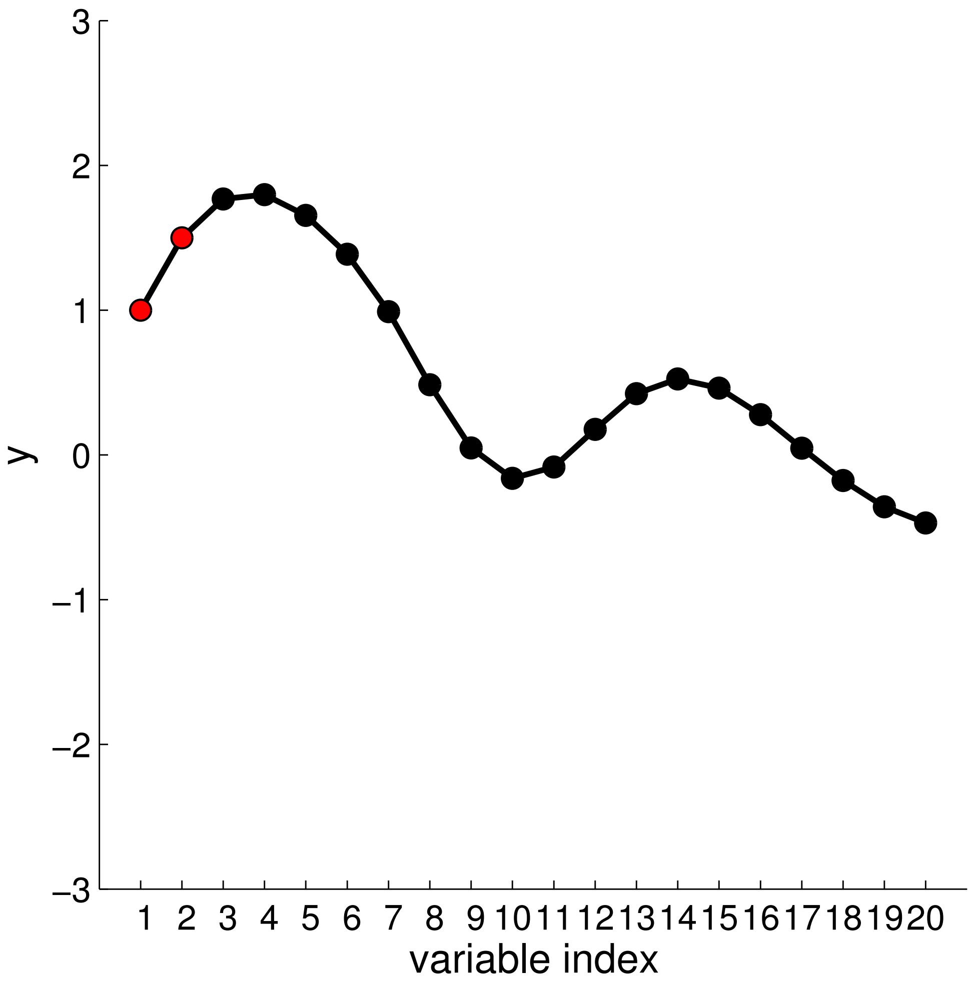
$\Sigma =$



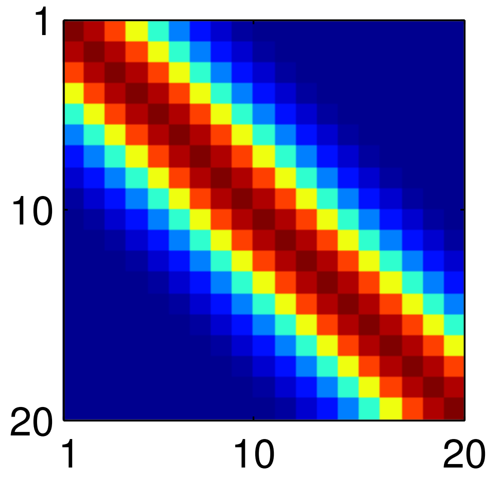
Conditioning on y1 and y2

New visualisation



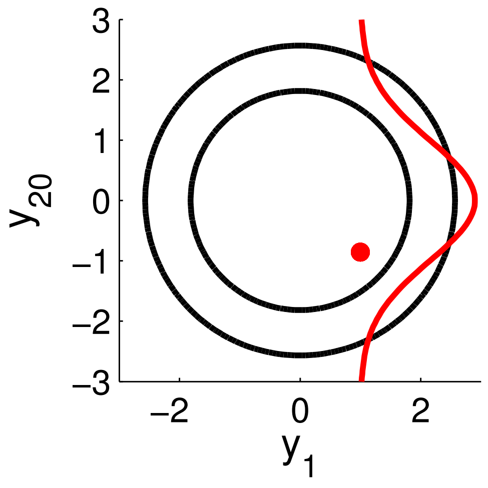


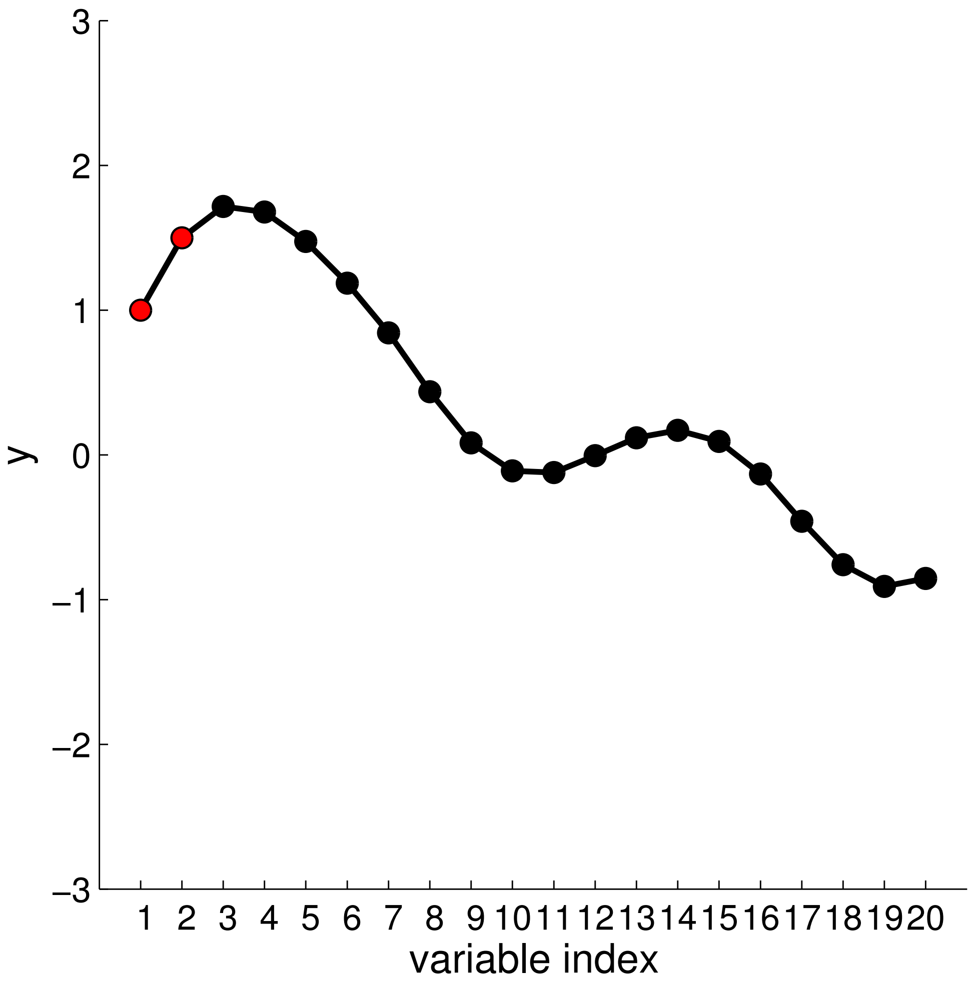
$\Sigma =$



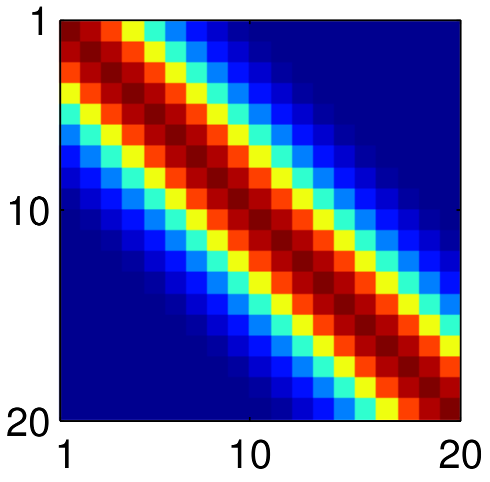
Conditioning on y1 and y2

New visualisation



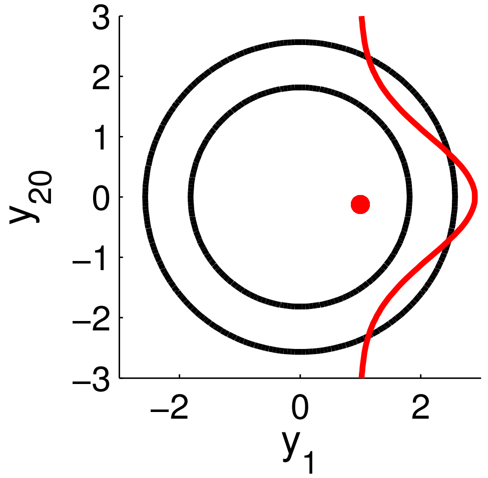


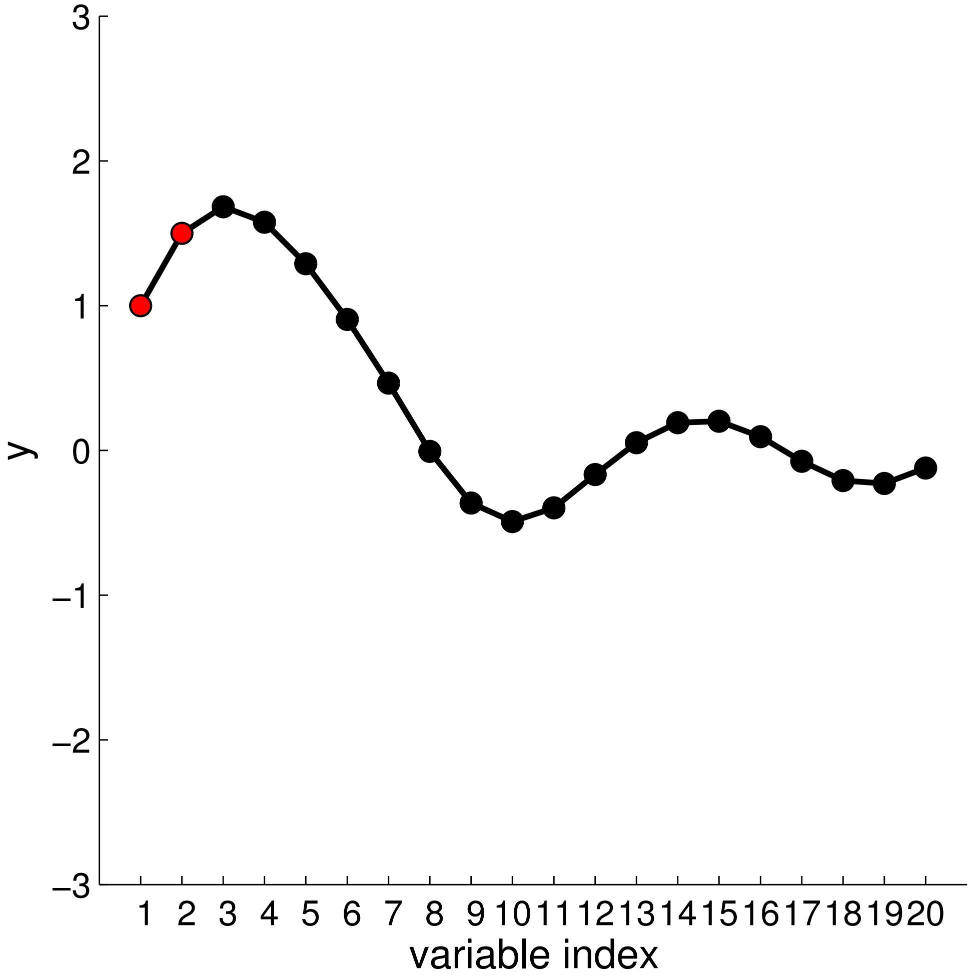
$\Sigma =$



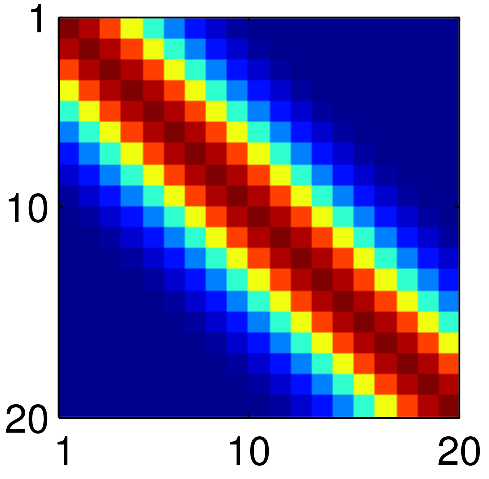
Conditioning on y1 and y2

New visualisation



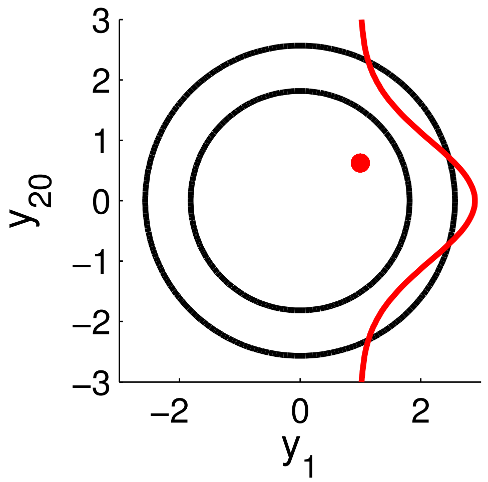


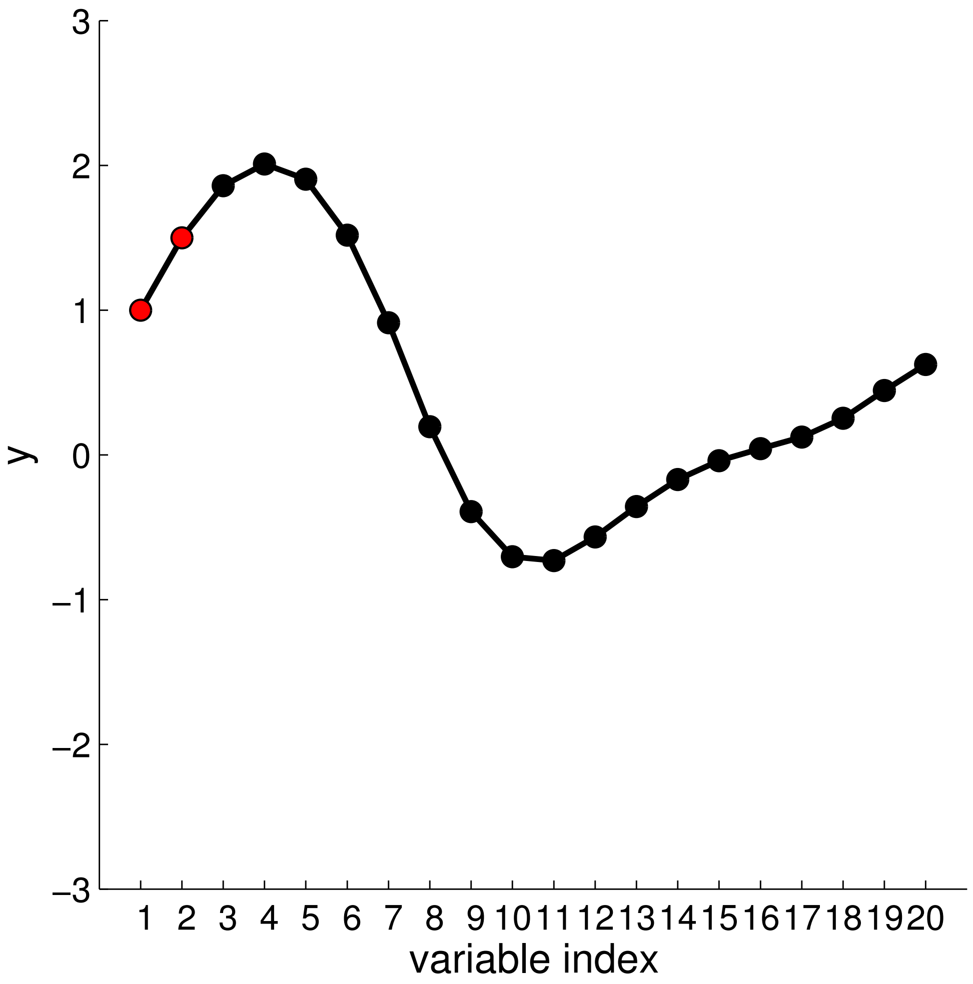
$\Sigma =$



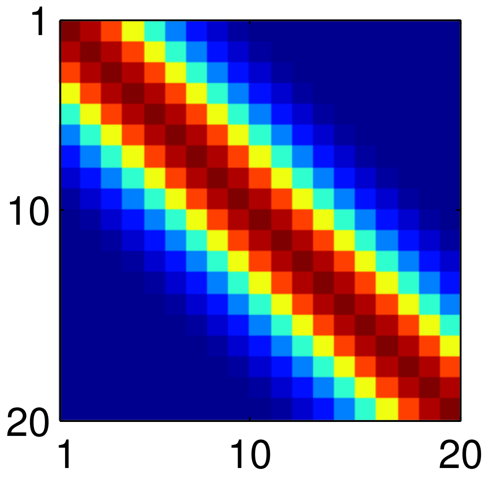
Conditioning on y1 and y2

New visualisation



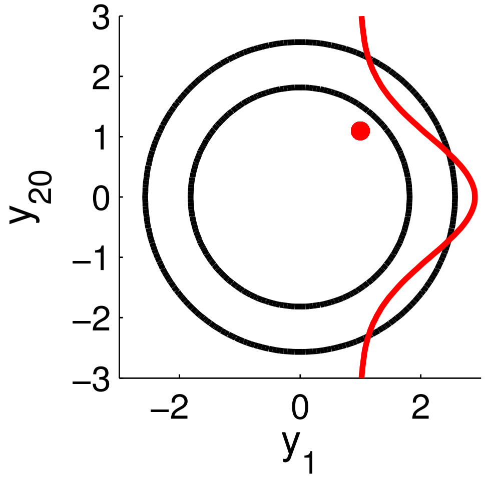


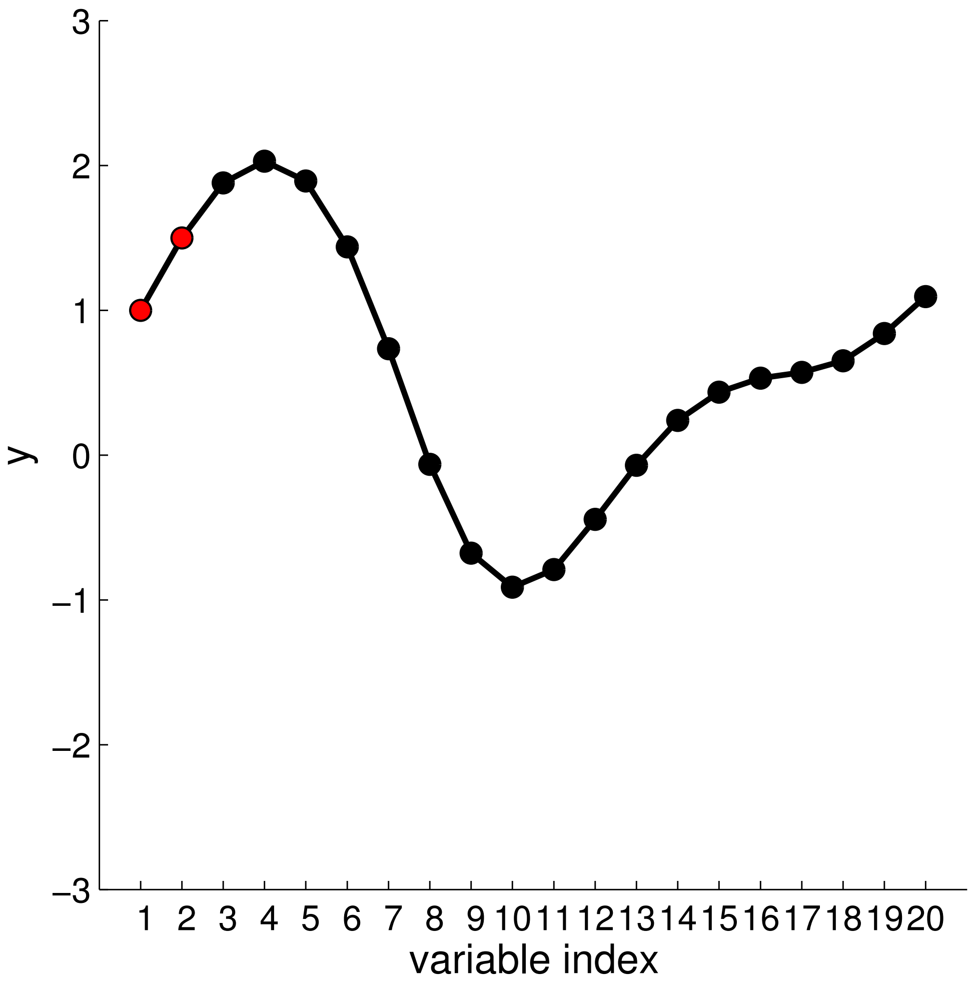
$\Sigma =$



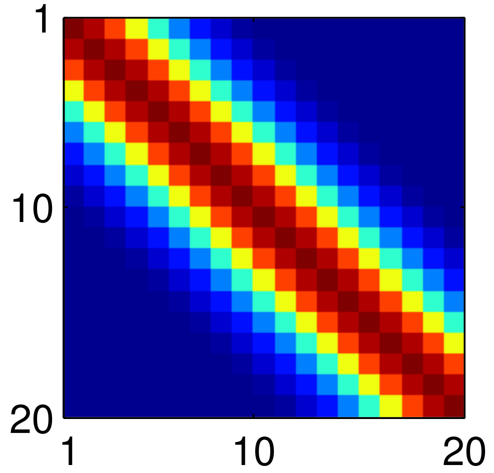
Conditioning on y1 and y2

New visualisation



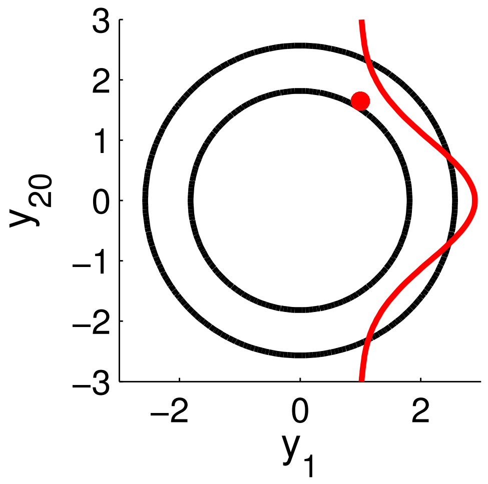


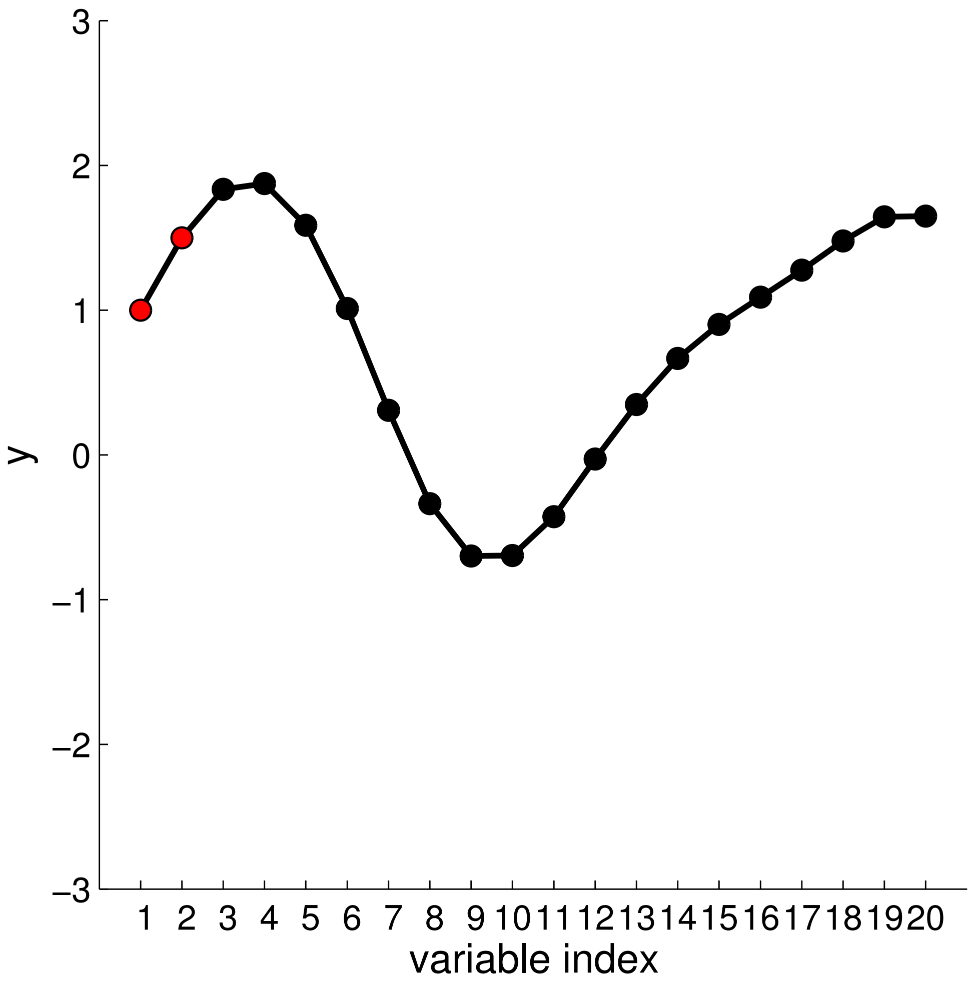
$\Sigma =$



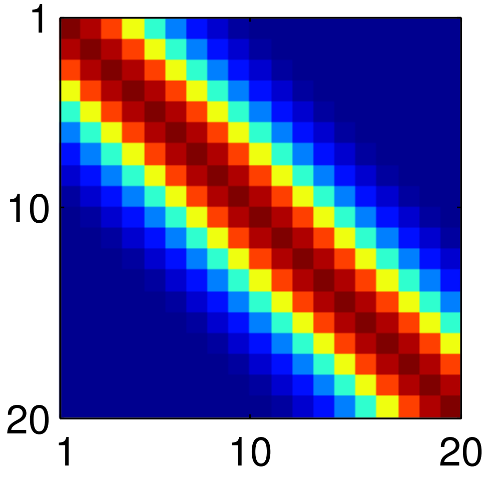
Conditioning on y1 and y2

New visualisation



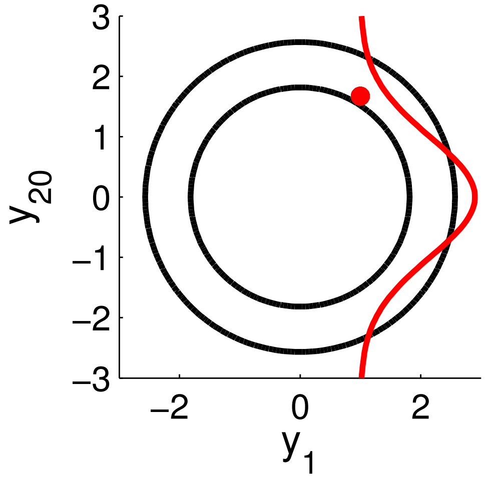


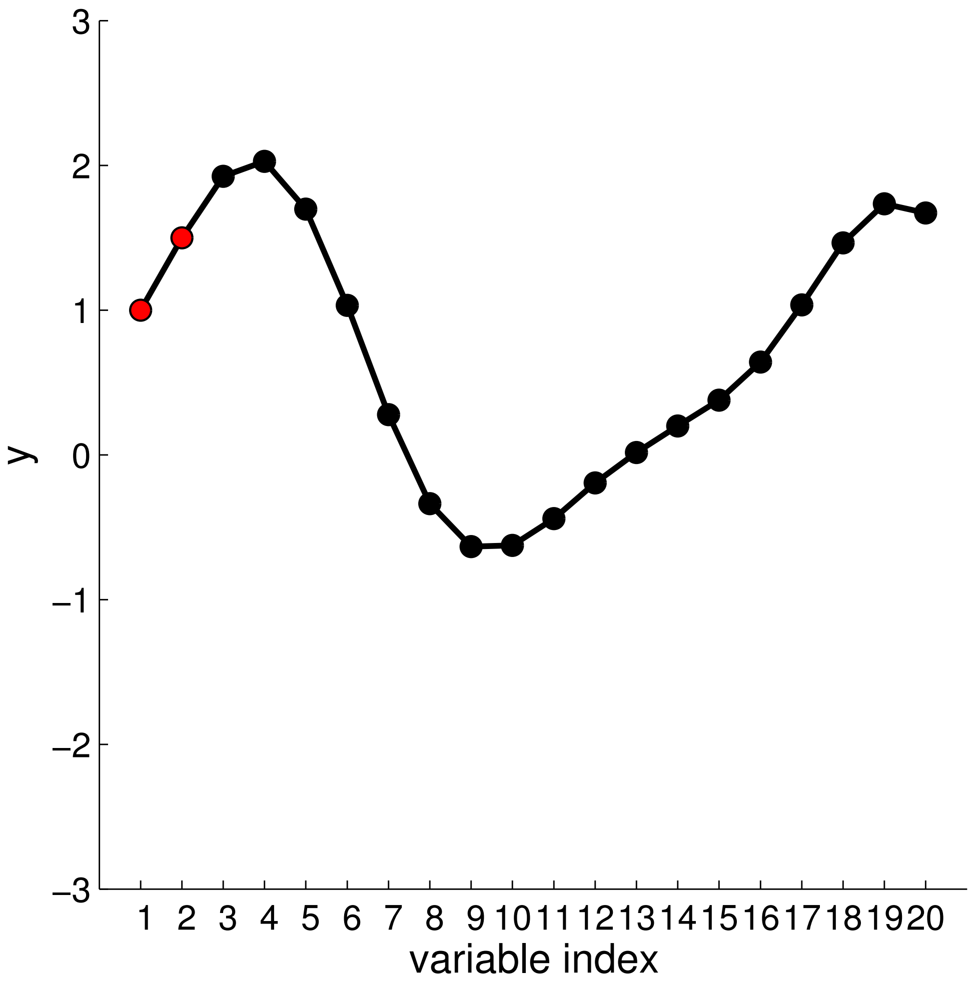
$\Sigma =$



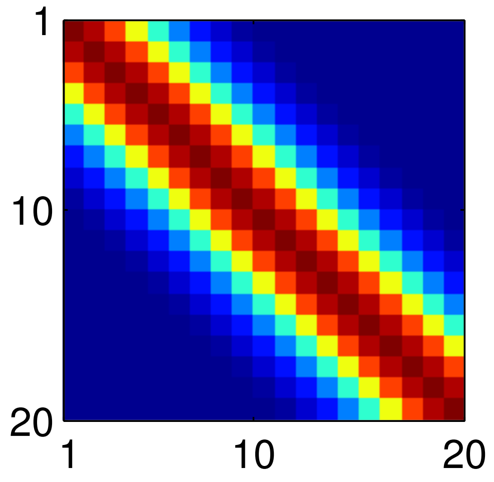
Conditioning on y1 and y2

New visualisation



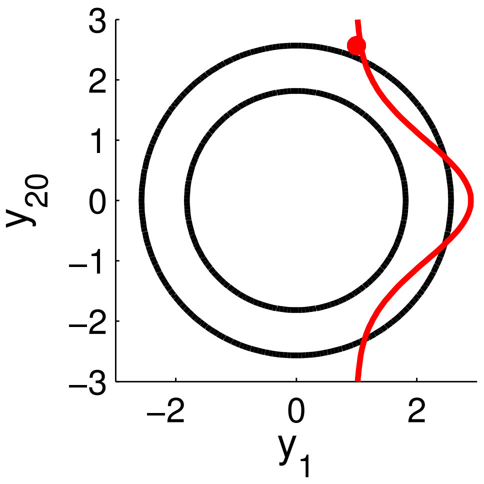


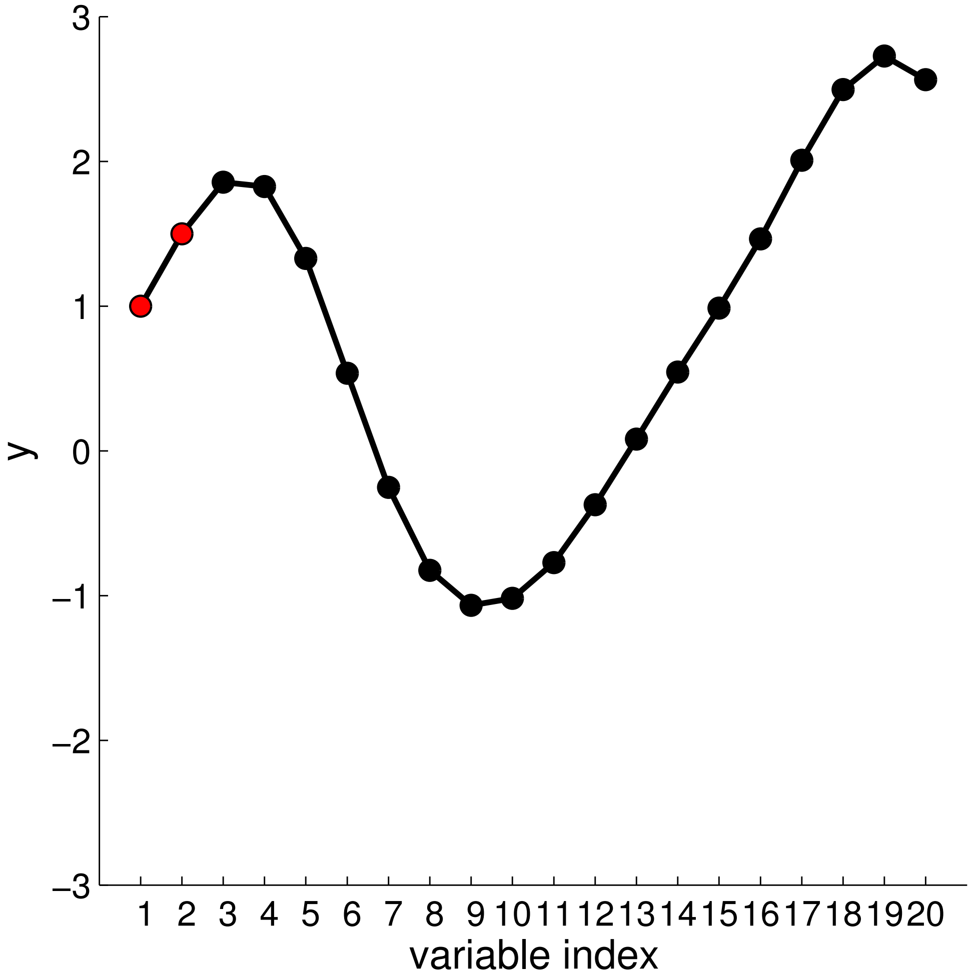
$\Sigma =$



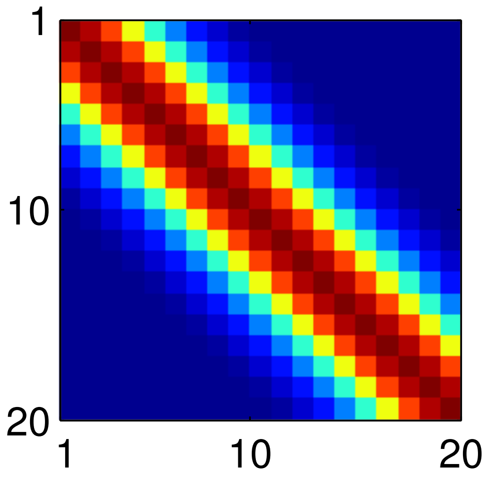
Conditioning on y1 and y2

New visualisation



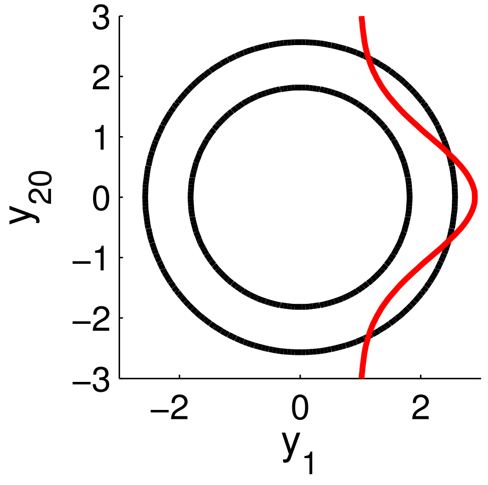


$\Sigma =$

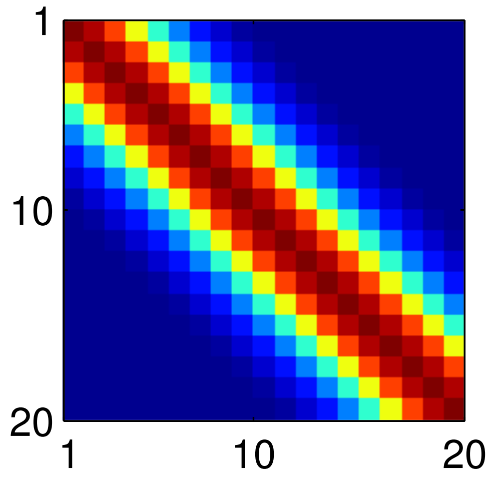


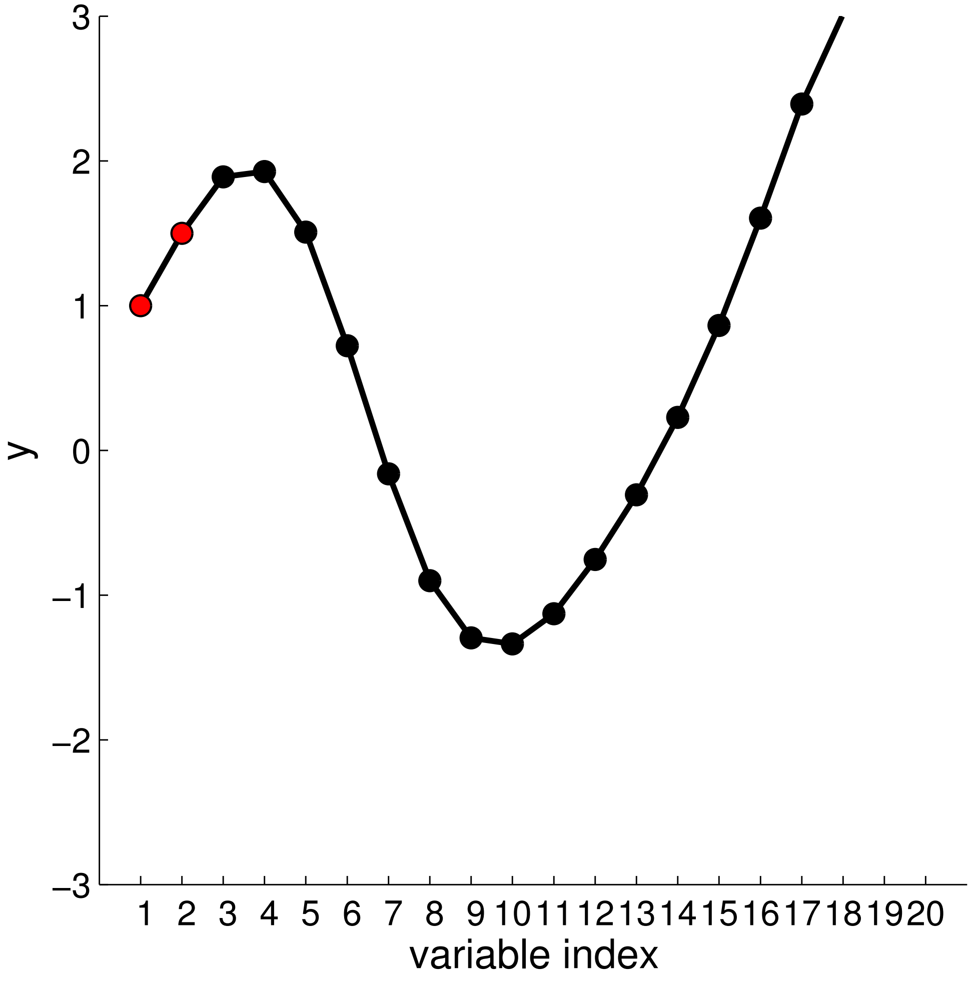
Conditioning on y1 and y2

New visualisation



$\Sigma =$

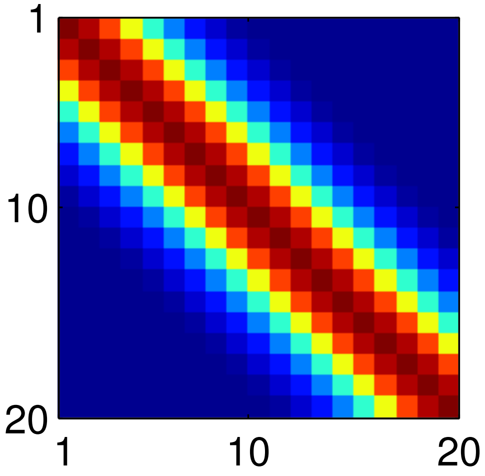


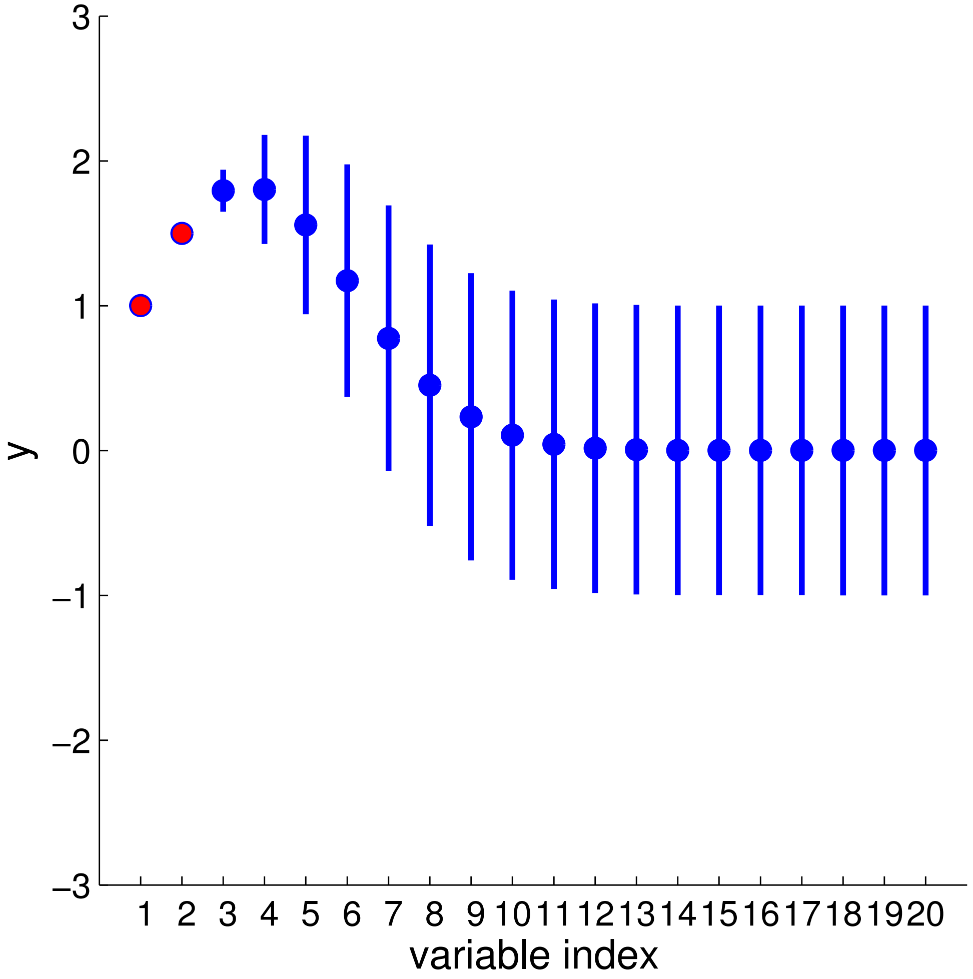


Conditioning on y1 and y2

Regression using Gaussians

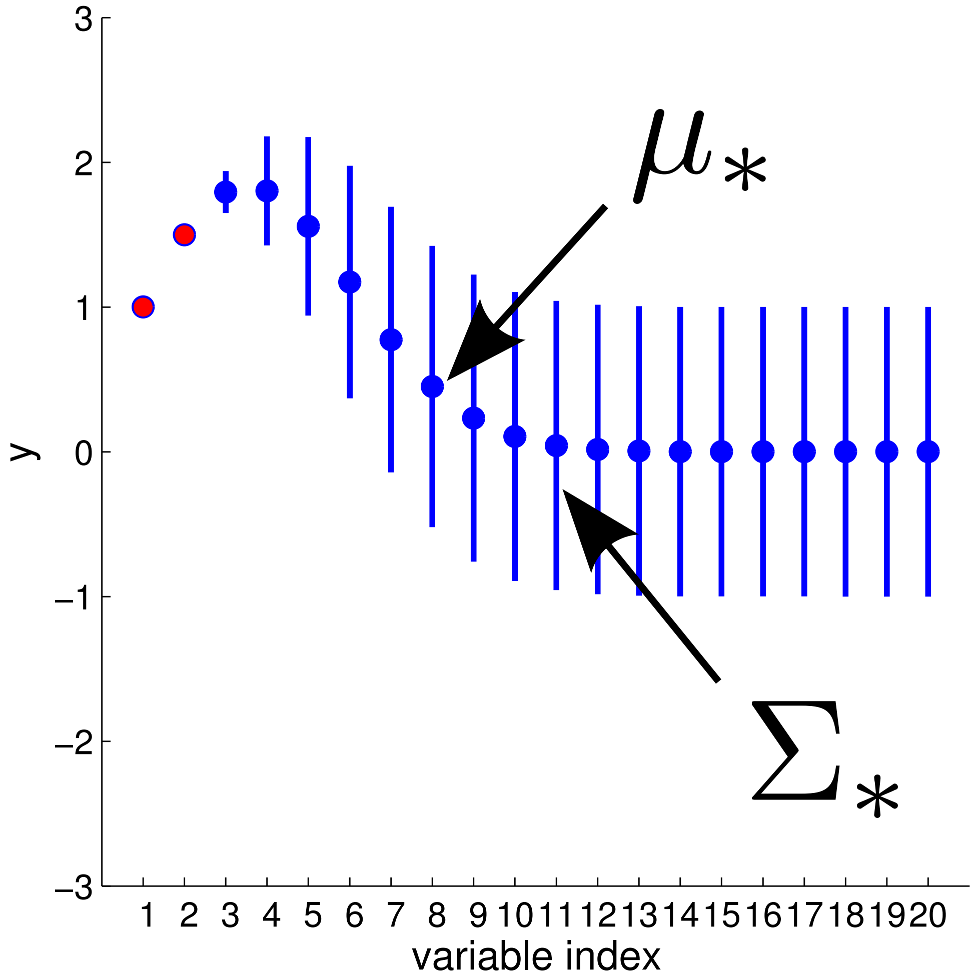
$\Sigma =$



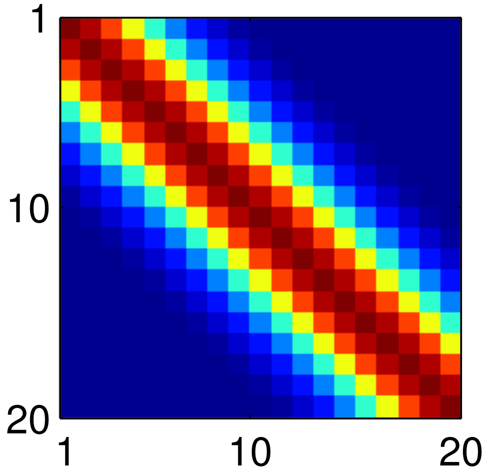


Conditioning on y1 and y2

Regression using Gaussians



$\Sigma =$

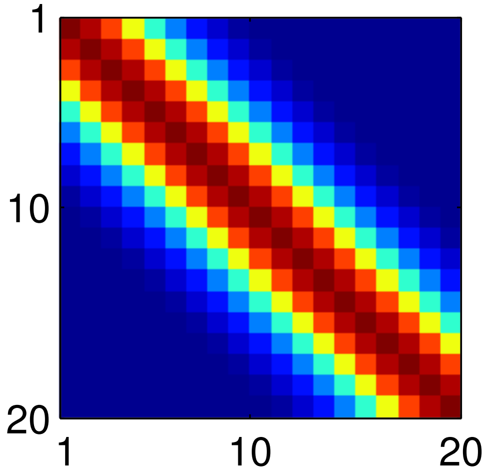


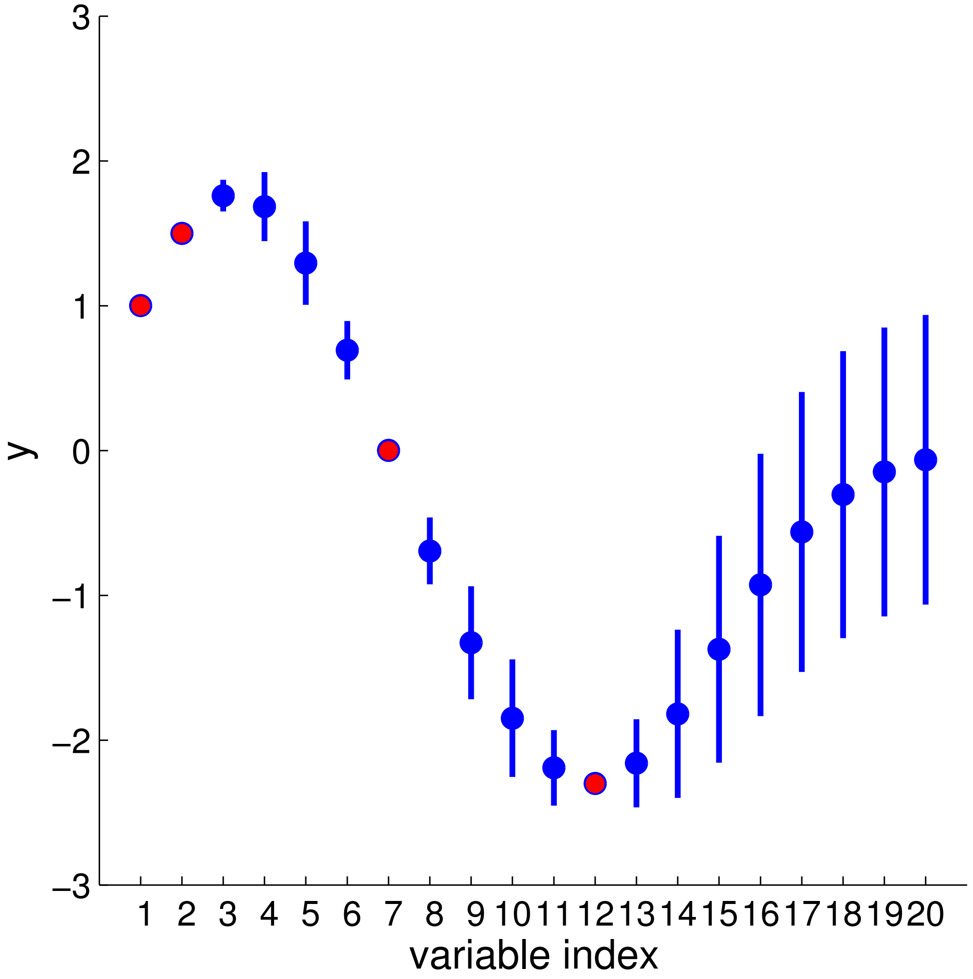
These quantities can be computed analytically: we do not need to average over many samples.

Q: why this is important?

Regression using Gaussians

$\Sigma =$



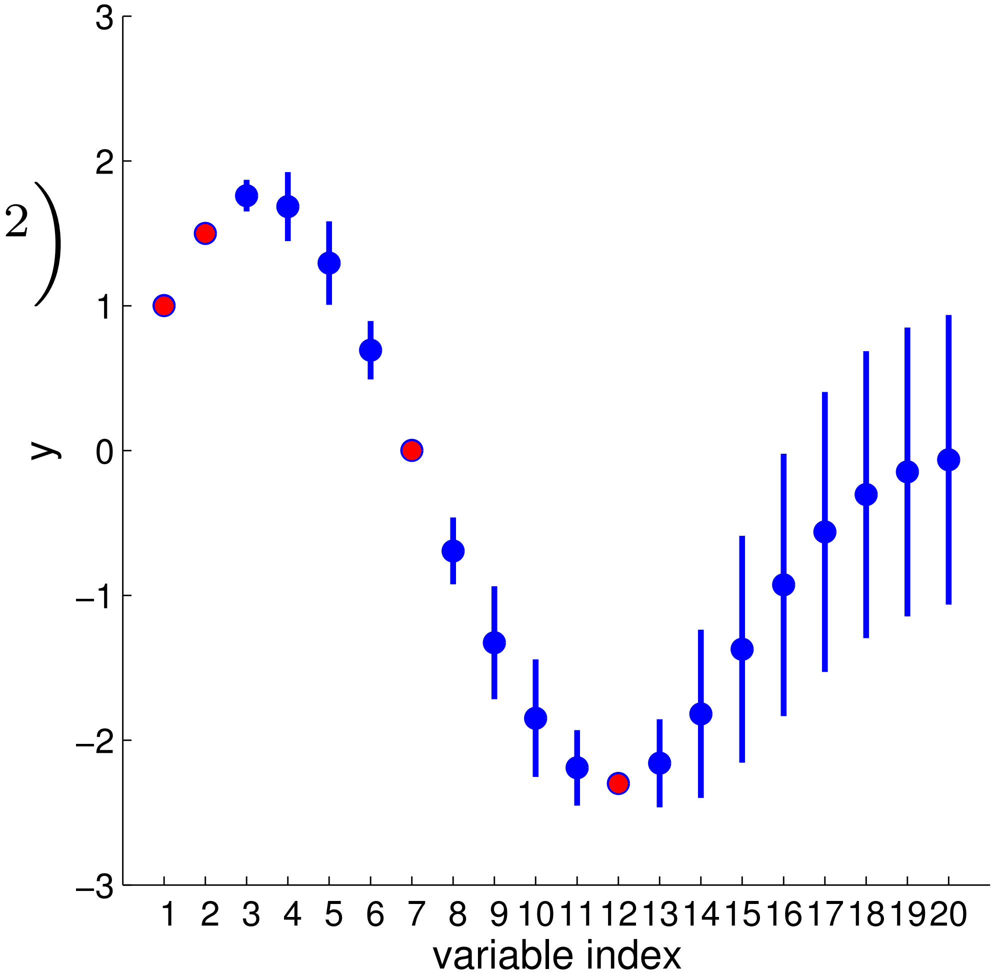


Regression using Gaussians

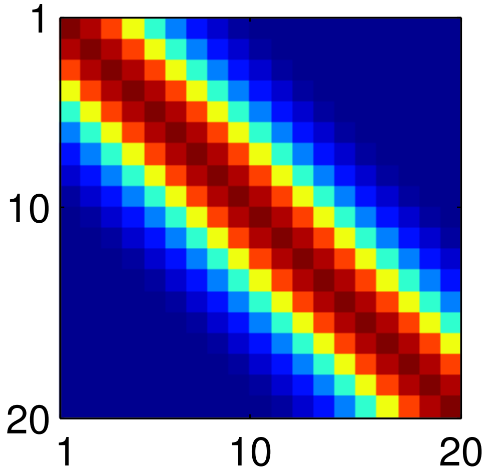
That’s how the covariance matrix was computed: the further x_1 from x_2, the smaller the correlation

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

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



$\Sigma =$



Q: Do x_1,x_2 need to be integers?

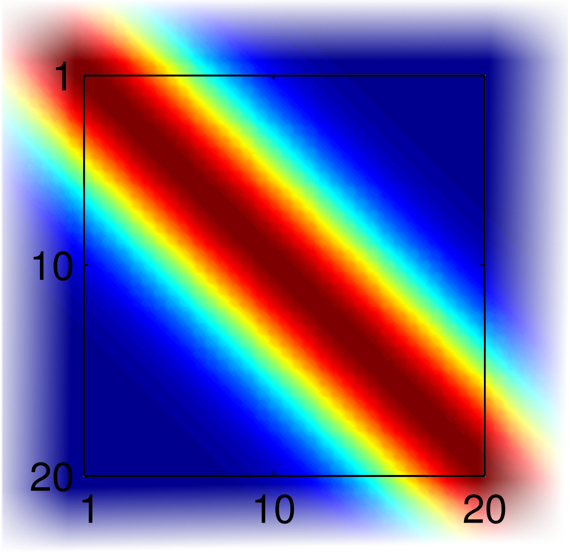
From multivariete Gaussian distributions to Gaussian Processes

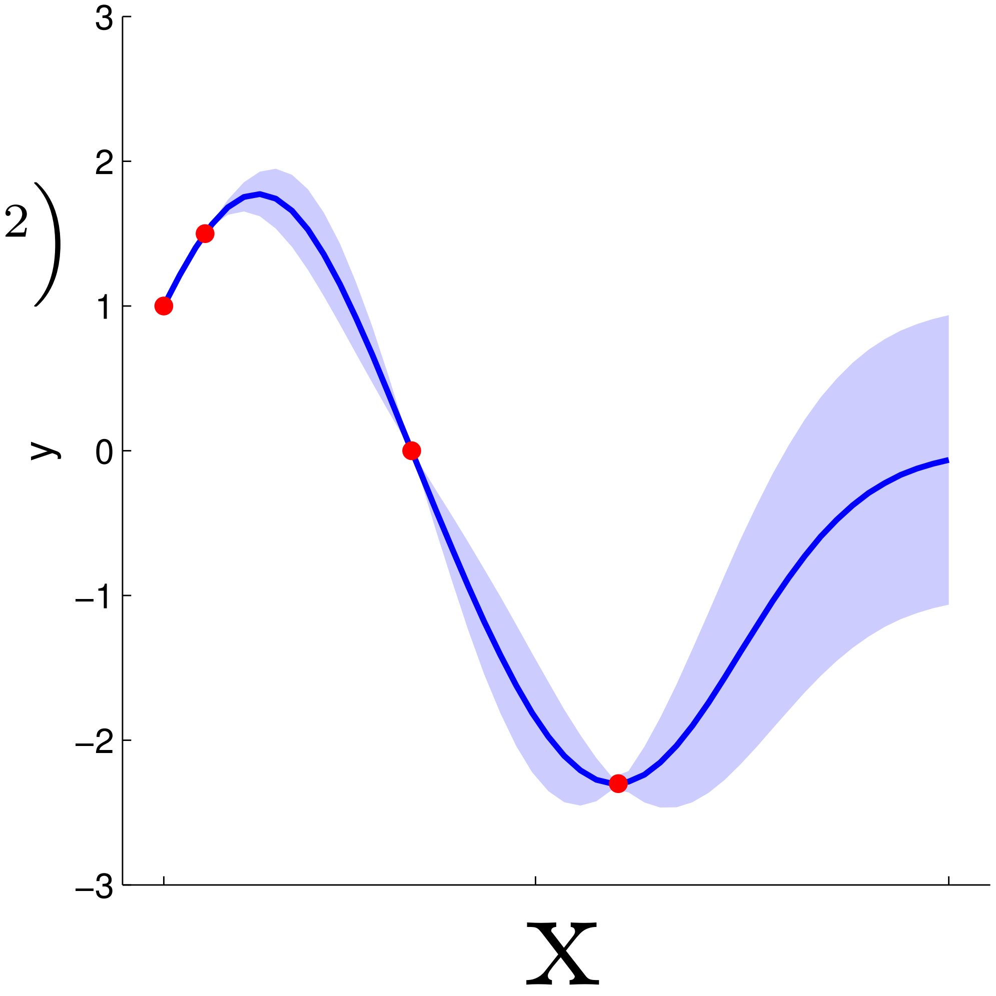
GP: *a multivariate Gaussian over an uncountably inf number of variables with inf mean vector and inf X inf covariance matrix*

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

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

$\Sigma =$





Regression: probabilistic inference in function space

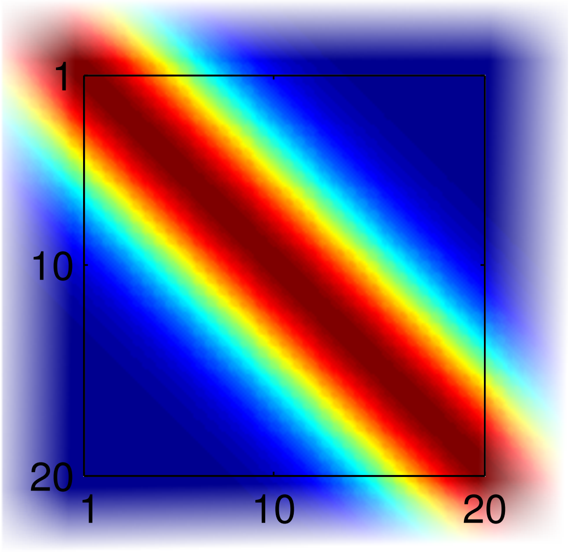
Non-parametric (∞-parametric)

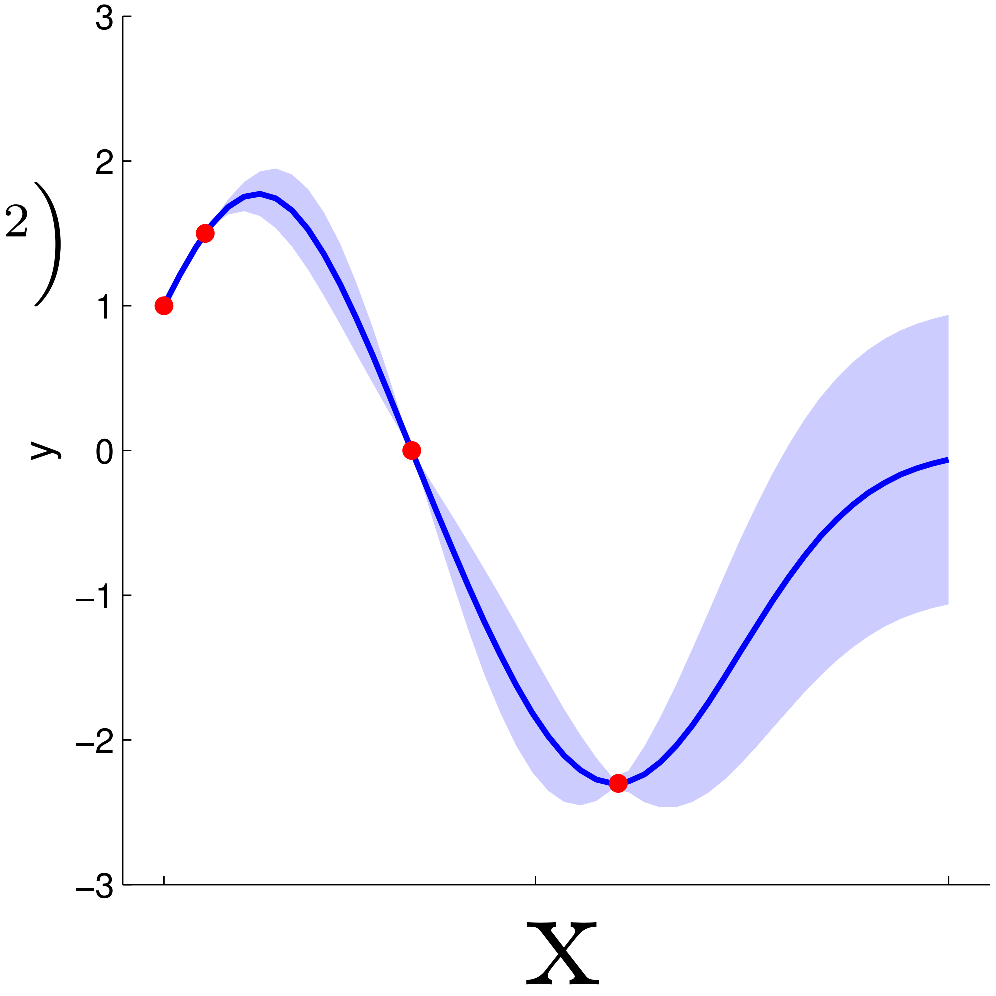
$$p(\mathrm{y}|\theta) = \mathcal{N}(0, \Sigma)$$

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

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

$$\Sigma =$$





Parametric model

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

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

Regression: probabilistic inference in function space

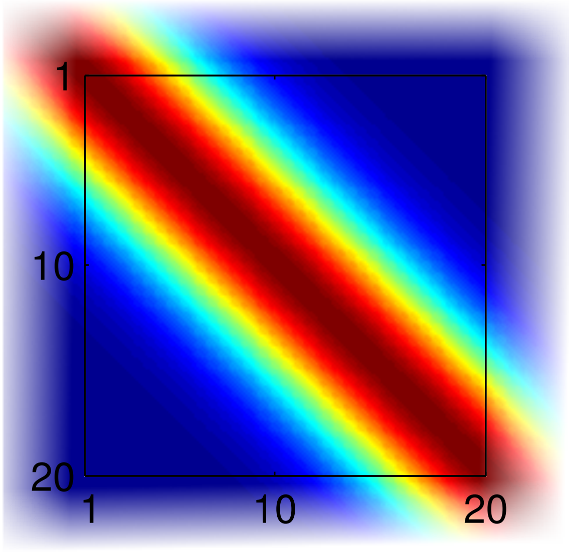
Non-parametric (∞-parametric)

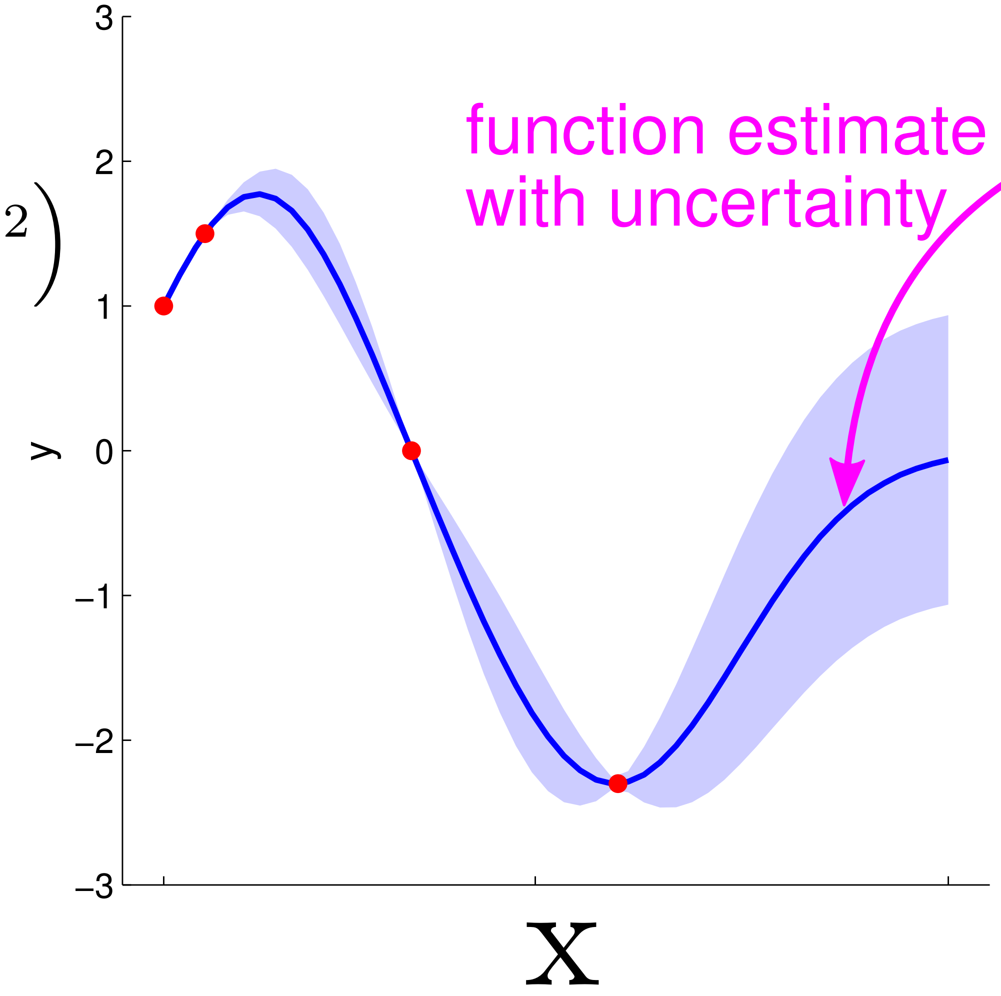
$$p(\mathsf{y}|\theta) = \mathcal{N}(0, \Sigma)$$

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

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

$\Sigma =$





function estimate with uncertainty

Parametric model

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

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

Regression: probabilistic inference in function space

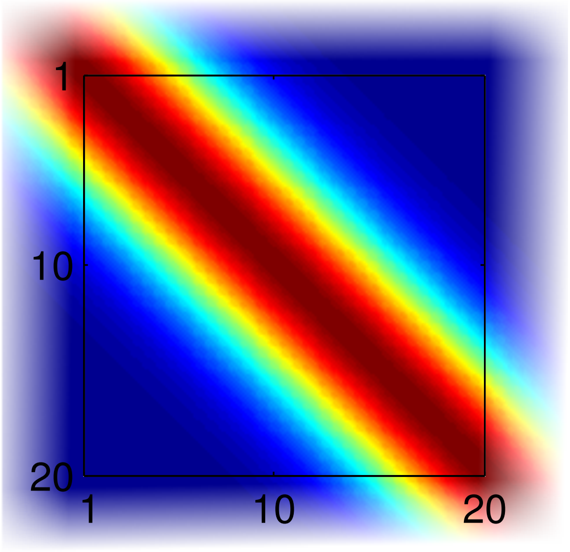
Non-parametric (∞-parametric)

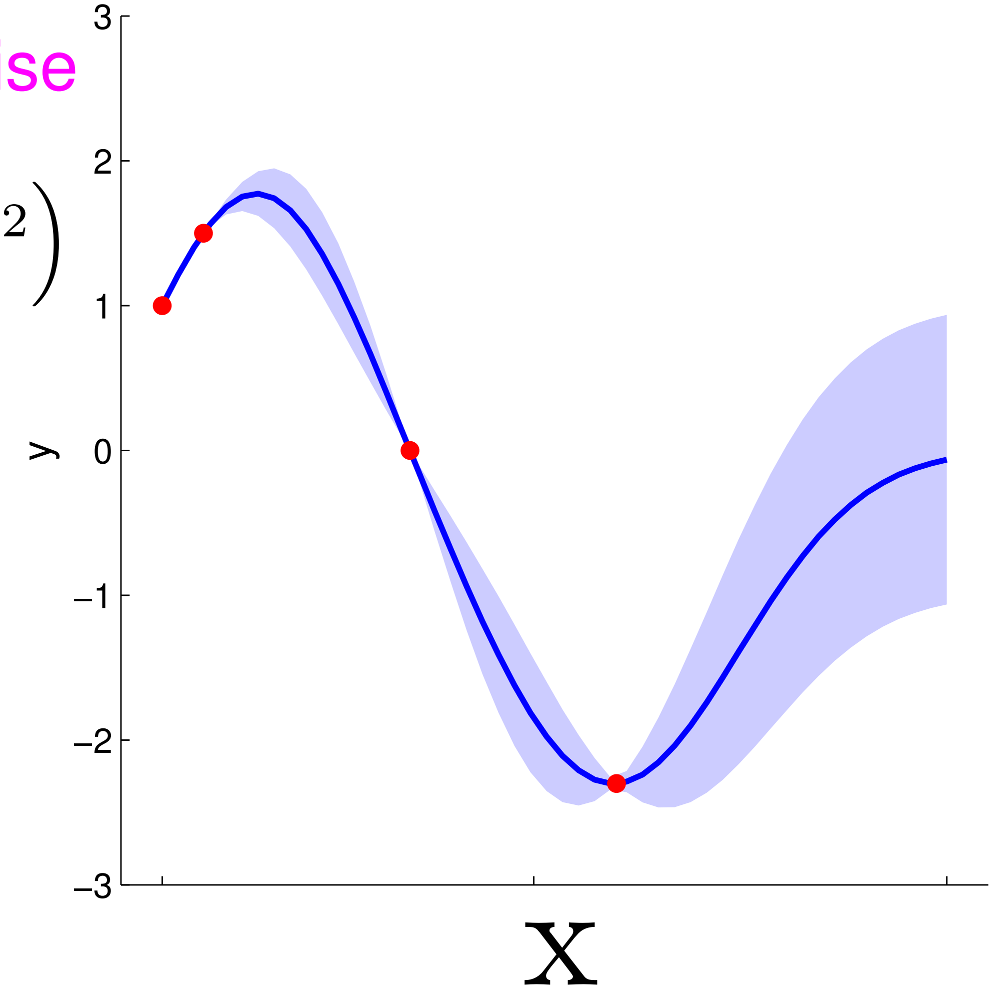
$$p(\mathsf{y}|\theta) = \mathcal{N}(0, \Sigma)$$

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

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

$\Sigma =$





Parametric model

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

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

Regression: probabilistic inference in function space

Non-parametric (∞-parametric)

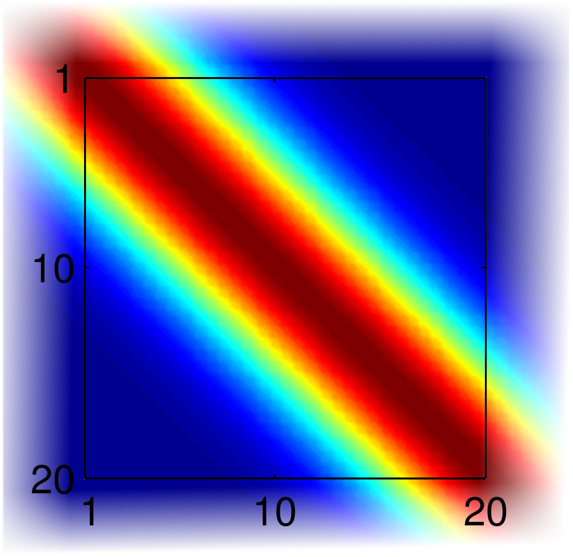
$$p(\mathsf{y}|\theta) = \mathcal{N}(0, \Sigma)$$

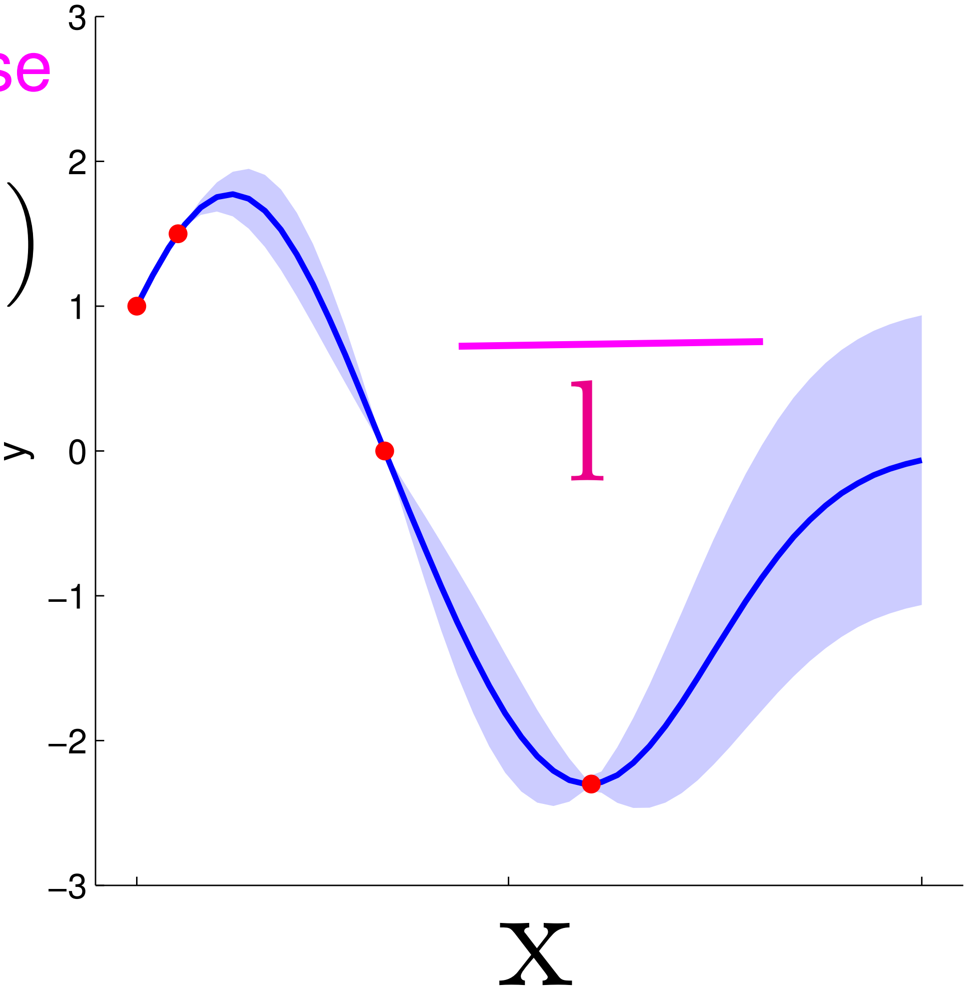
$$\Sigma(\mathsf{x}_1, \mathsf{x}_2) = \mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) + \mathsf{I}\sigma_{\mathsf{y}}^2 \leftarrow \text{noise}$$

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

↑
horizontal-scale

$\Sigma =$





Parametric model

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

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

How fast the correlations fall off..

Regression: probabilistic inference in function space

Non-parametric (∞-parametric)

Parametric model

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

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

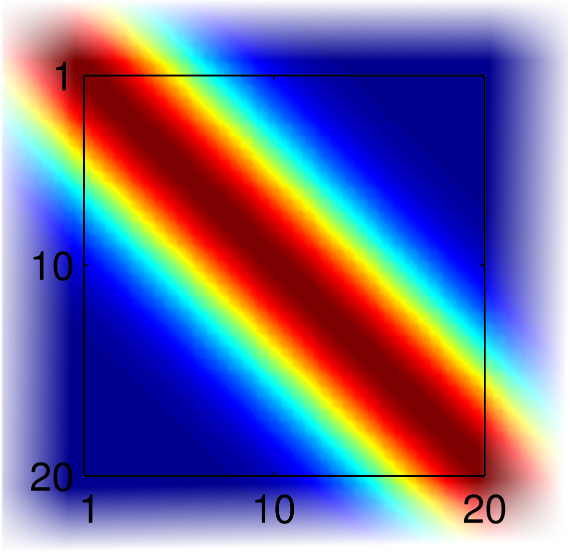
$$\mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) = \sigma^2 \exp\left(-\frac{1}{2\mathsf{l}^2}(\mathsf{x}_1 - \mathsf{x}_2)^2\right)$$

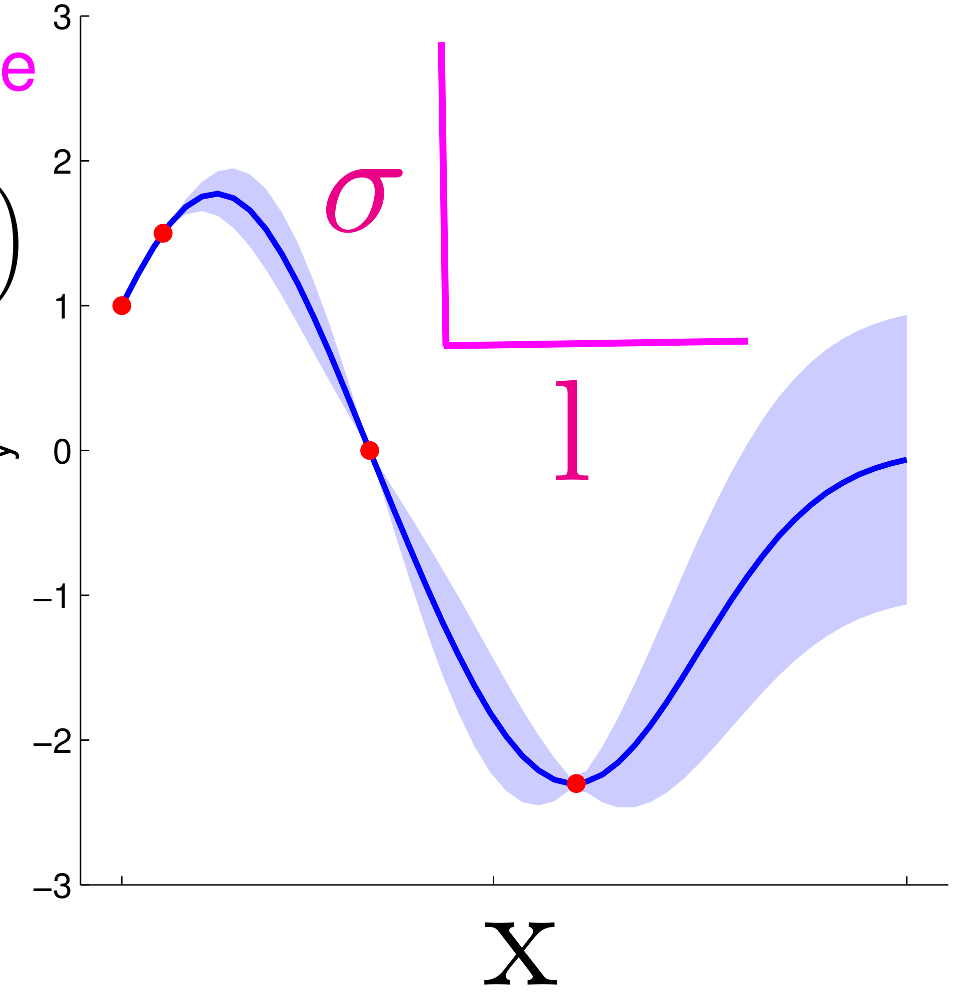
vertical-scale horizontal-scale

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

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

$$\Sigma =$$





How fast the correlations fall off..

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:

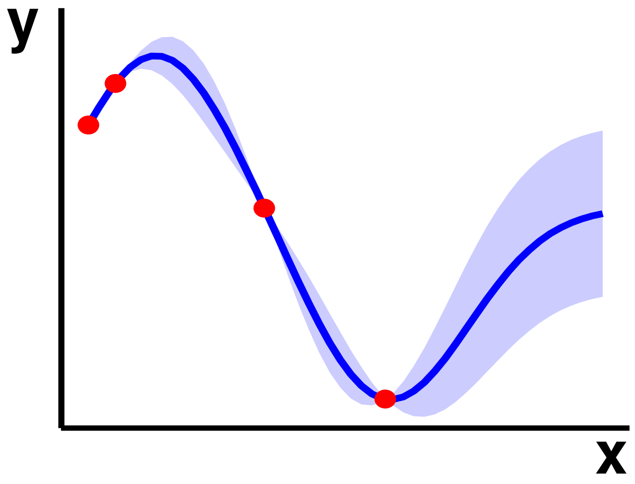
$$\mathbf{f} = (\mathrm{f}_1, \ldots, \mathrm{f}_n) \sim \mathcal{N}(\mu, \Sigma), \quad \text{indices} \quad i = 1, \ldots, n$$

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

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

Mathematical foundations: Prediction

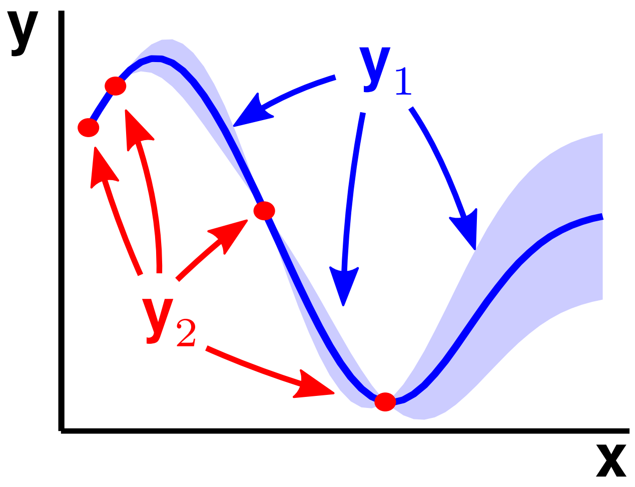
Q4. How do we make predictions?



Mathematical foundations: Prediction

Q4. How do we make predictions?

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

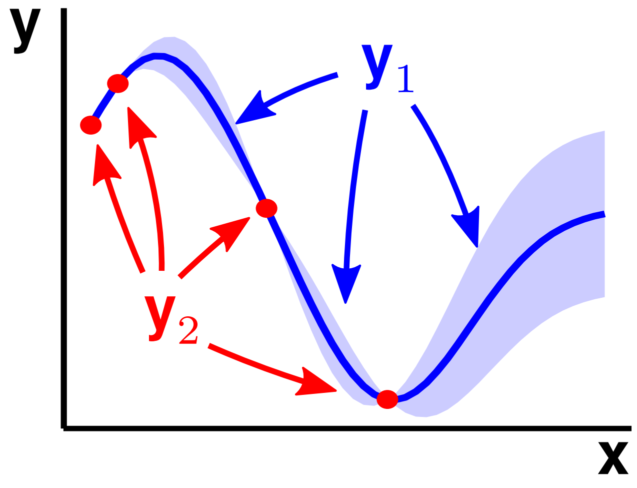


Mathematical foundations: Prediction

Q4. How do we make predictions?

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

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



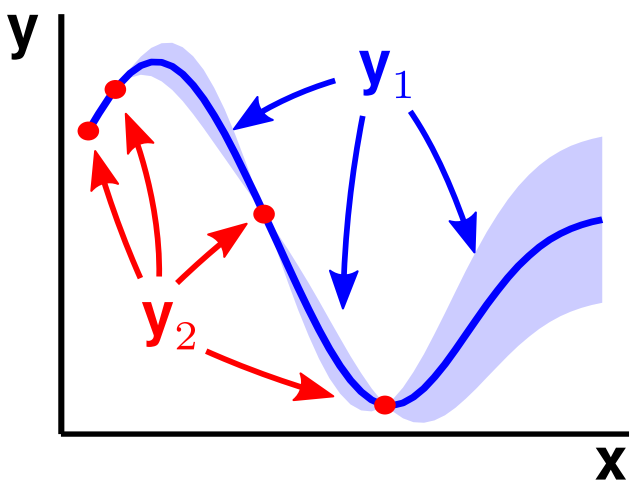
Mathematical foundations: Prediction

Q4. How do we make predictions?

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

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

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

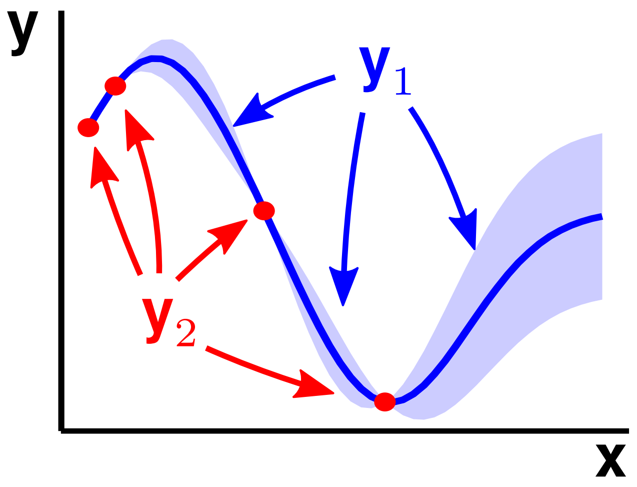


Mathematical foundations: Prediction

Q4. How do we make predictions?

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

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



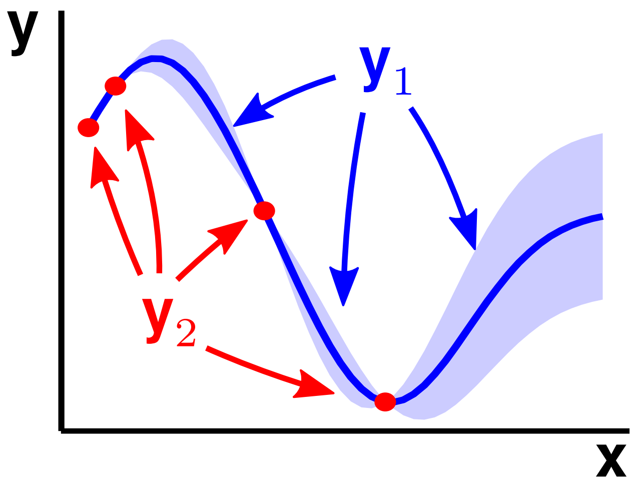
$$\Longrightarrow \quad p(\mathbf{y}_1|\mathbf{y}_2) = \mathcal{N}(\mathbf{a} + \mathsf{B}\mathsf{C}^{-1}(\mathbf{y}_2 - \mathbf{b}), \mathsf{A} - \mathsf{B}\mathsf{C}^{-1}\mathsf{B}^{\mathsf{T}})$$

Mathematical foundations: Prediction

Q4. How do we make predictions?

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

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



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

predictive mean

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

predictive covariance

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

Predictive uncertainty = prior uncertainty - reduction in uncertainty

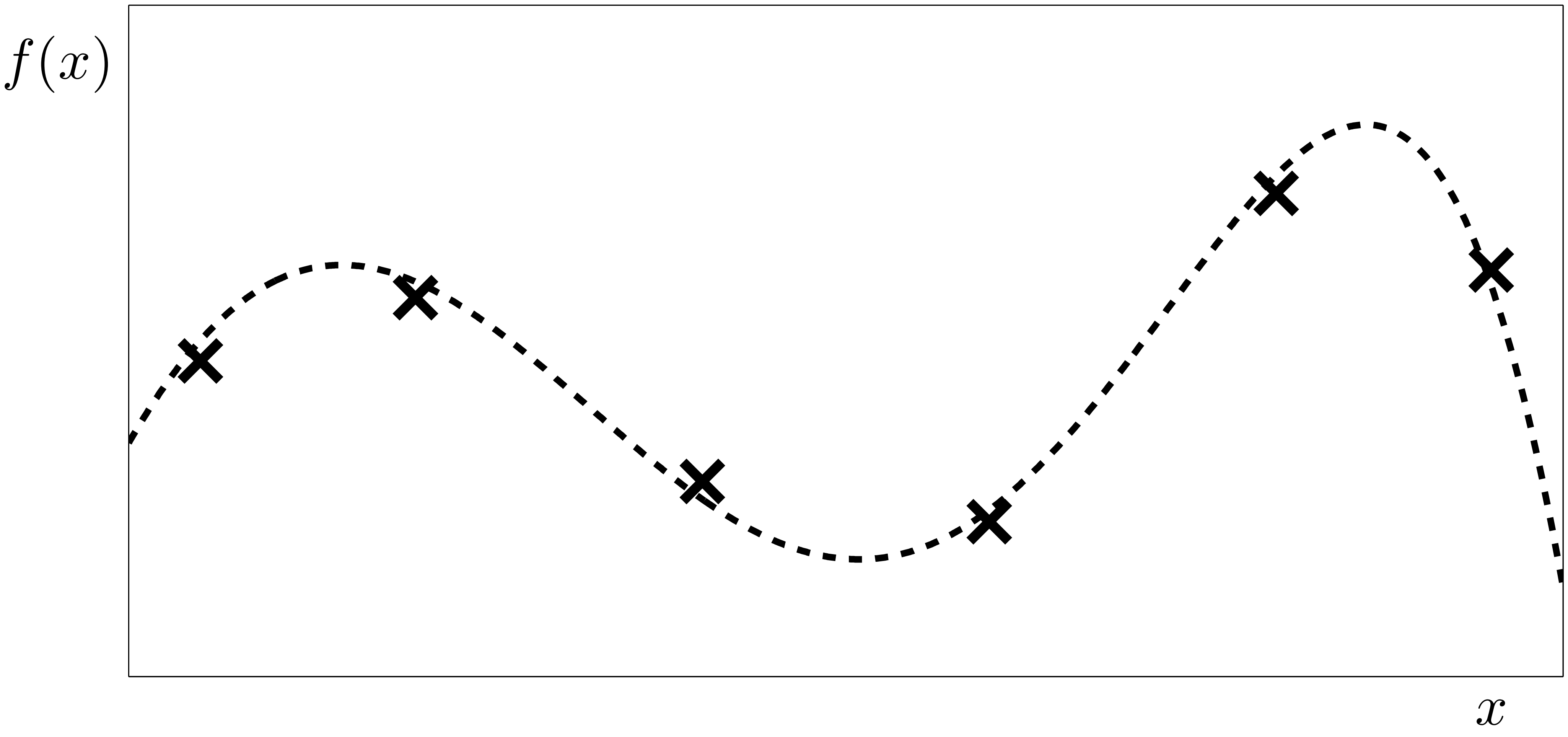
Bayesian Optimization

- Model f of the function I am trying to maximize (GPs for that)
- Aqcuisition function that takes as input the GP posteriror and suggests where to sample next
- We will see two acquisition functions:
  - UCB
  - Tompson sampling

Gaussian Processes ($\mathcal{GP}$)

$\mathcal{GP}(\mu, \kappa)$: A distribution over functions from $\mathcal{X}$ to $\mathbb{R}$.

Observations

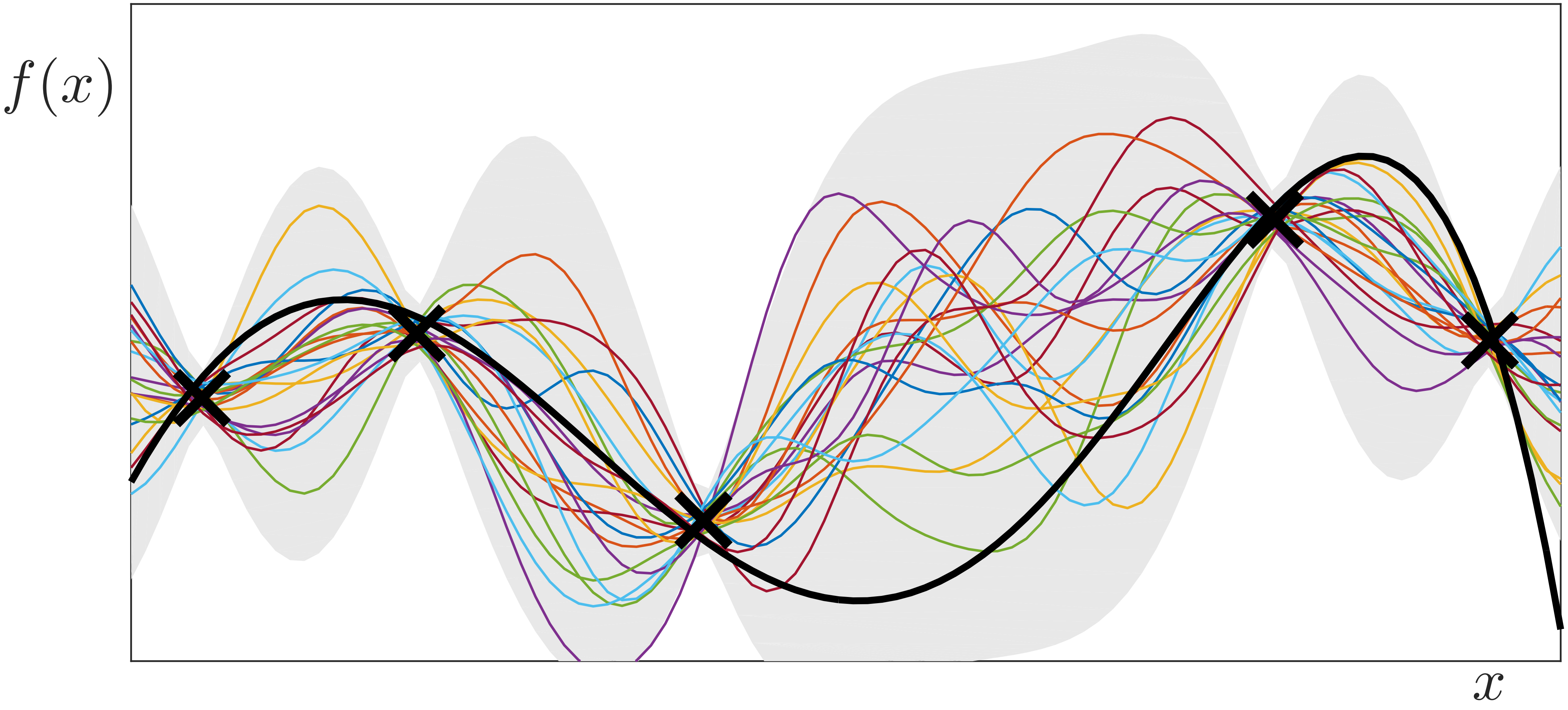


Algorithm 1: Upper Confidence Bounds in GP Bandits

Model $f \sim \mathcal{GP}(\mathbf{0}, \kappa)$.

Gaussian Process Upper Confidence Bound (GP-UCB)

(Srinivas et al. 2010)



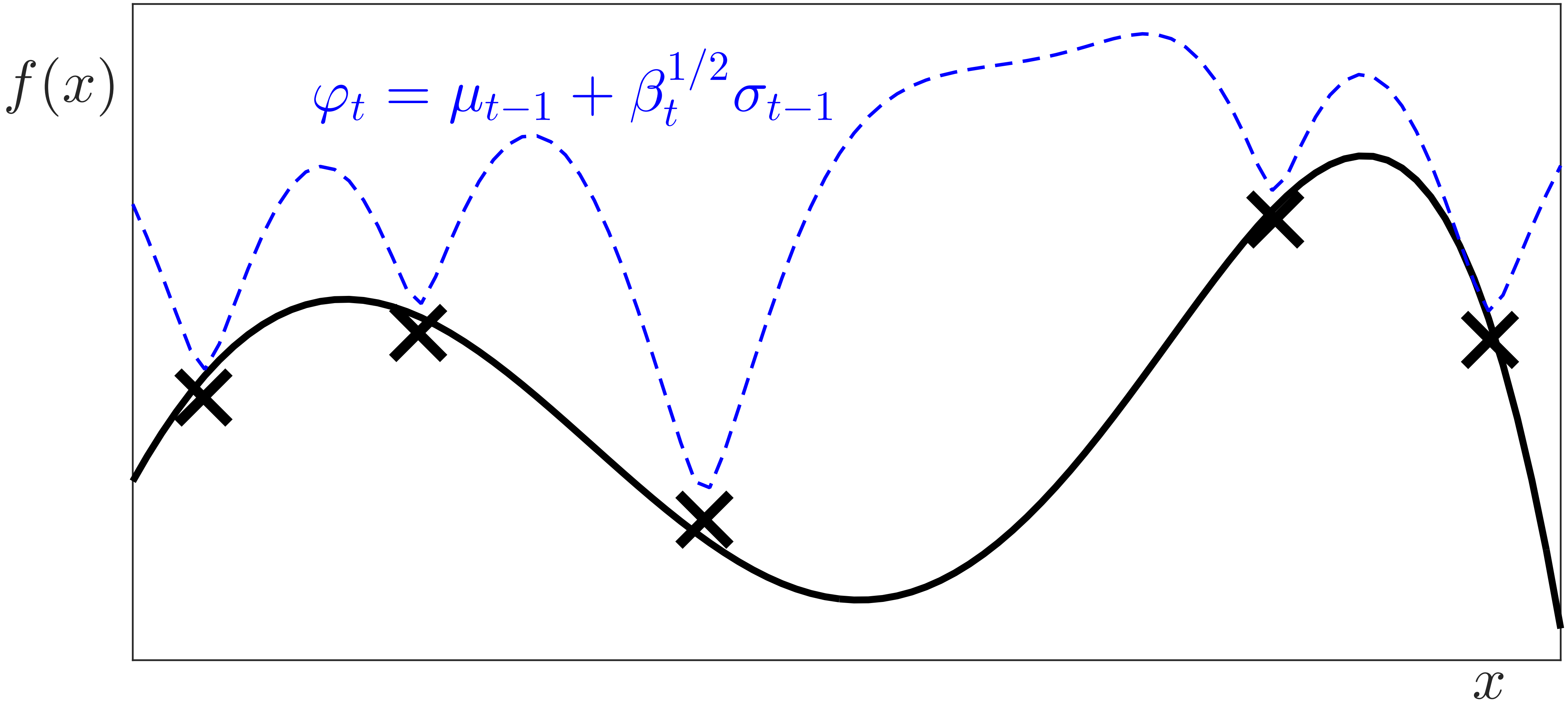
1) Compute posterior $\mathcal{GP}$.

Algorithm 1: Upper Confidence Bounds in GP Bandits

Model $f \sim \mathcal{GP}(\mathbf{0}, \kappa)$.

Gaussian Process Upper Confidence Bound (GP-UCB)

(Srinivas et al. 2010)



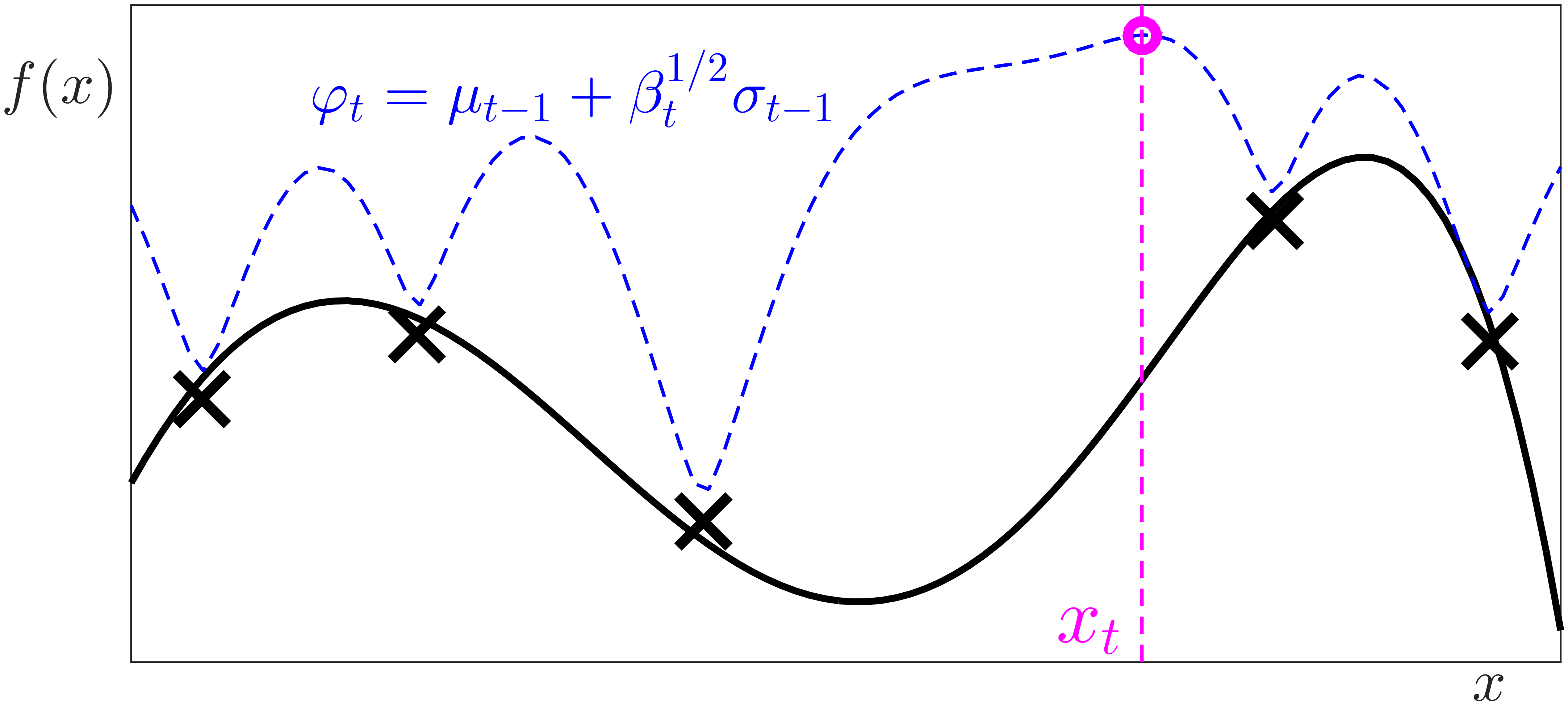
1) Compute posterior $\mathcal{GP}$. 2) Construct UCB φ_t.

Algorithm 1: Upper Confidence Bounds in GP Bandits

Model $f \sim \mathcal{GP}(\mathbf{0}, \kappa)$.

Gaussian Process Upper Confidence Bound (GP-UCB)

(Srinivas et al. 2010)



1) Compute posterior $\mathcal{GP}$.

2) Construct UCB φ_t.

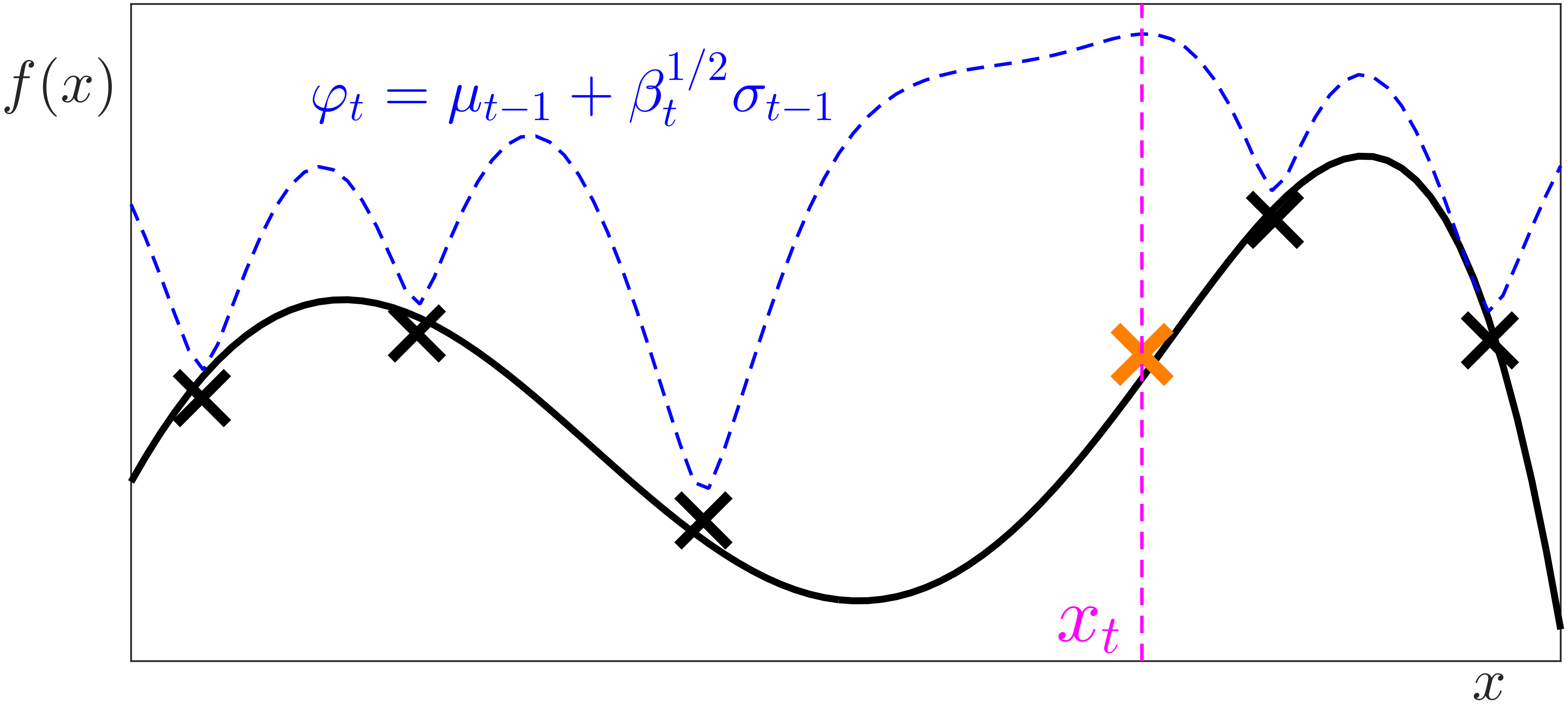
3) Choose $x_t = \mathrm{argmax}_x \, \varphi_t(x)$.

Algorithm 1: Upper Confidence Bounds in GP Bandits

Model $f \sim \mathcal{GP}(\mathbf{0}, \kappa)$.

Gaussian Process Upper Confidence Bound (GP-UCB)

(Srinivas et al. 2010)



1) Compute posterior $\mathcal{GP}$.

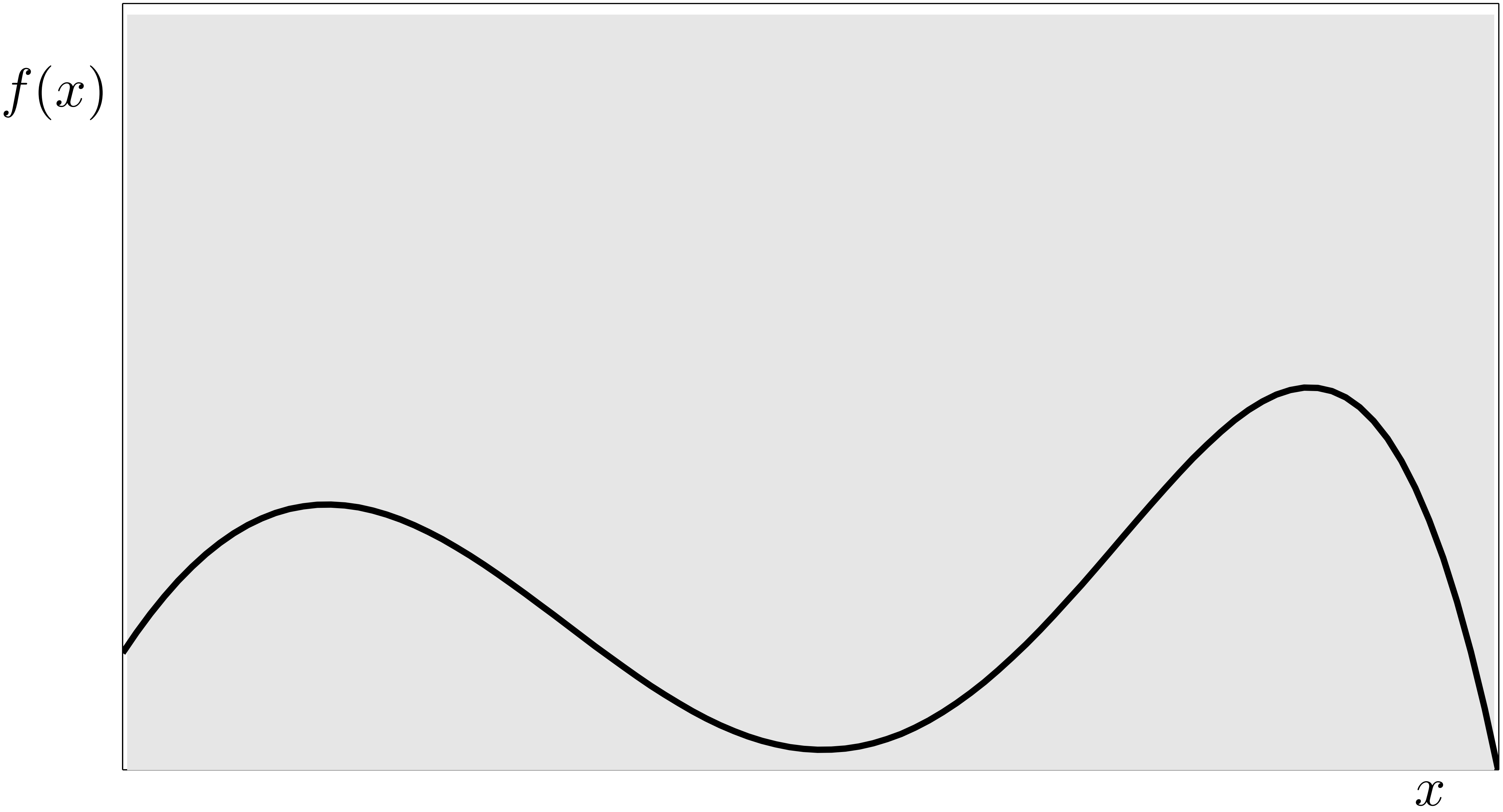
2) Construct UCB φ_t.

3) Choose $x_t = \operatorname{argmax}_x \varphi_t(x)$.

4) Evaluate f at x_t.

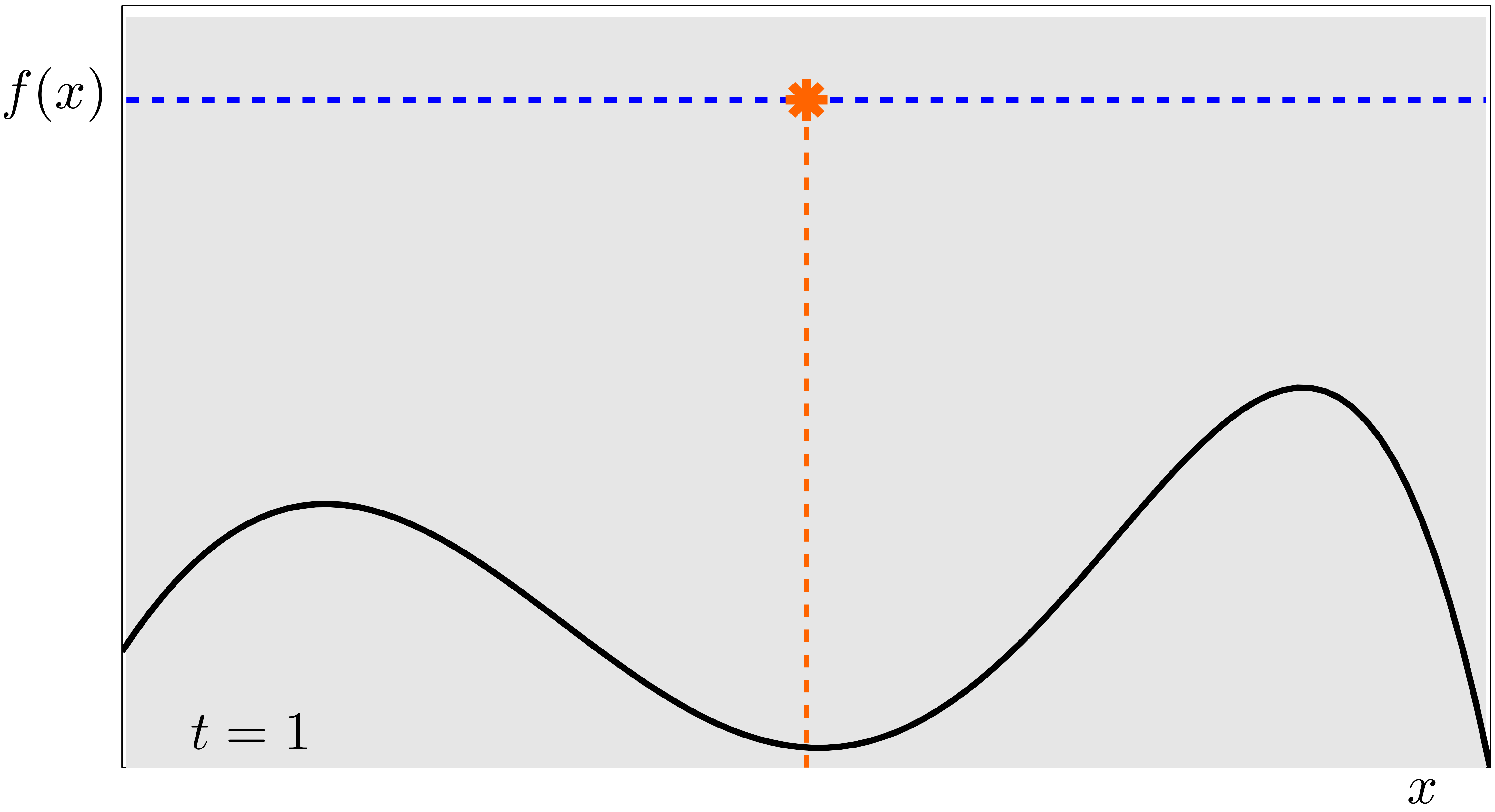
GP-UCB

(Srinivas et al. 2010)



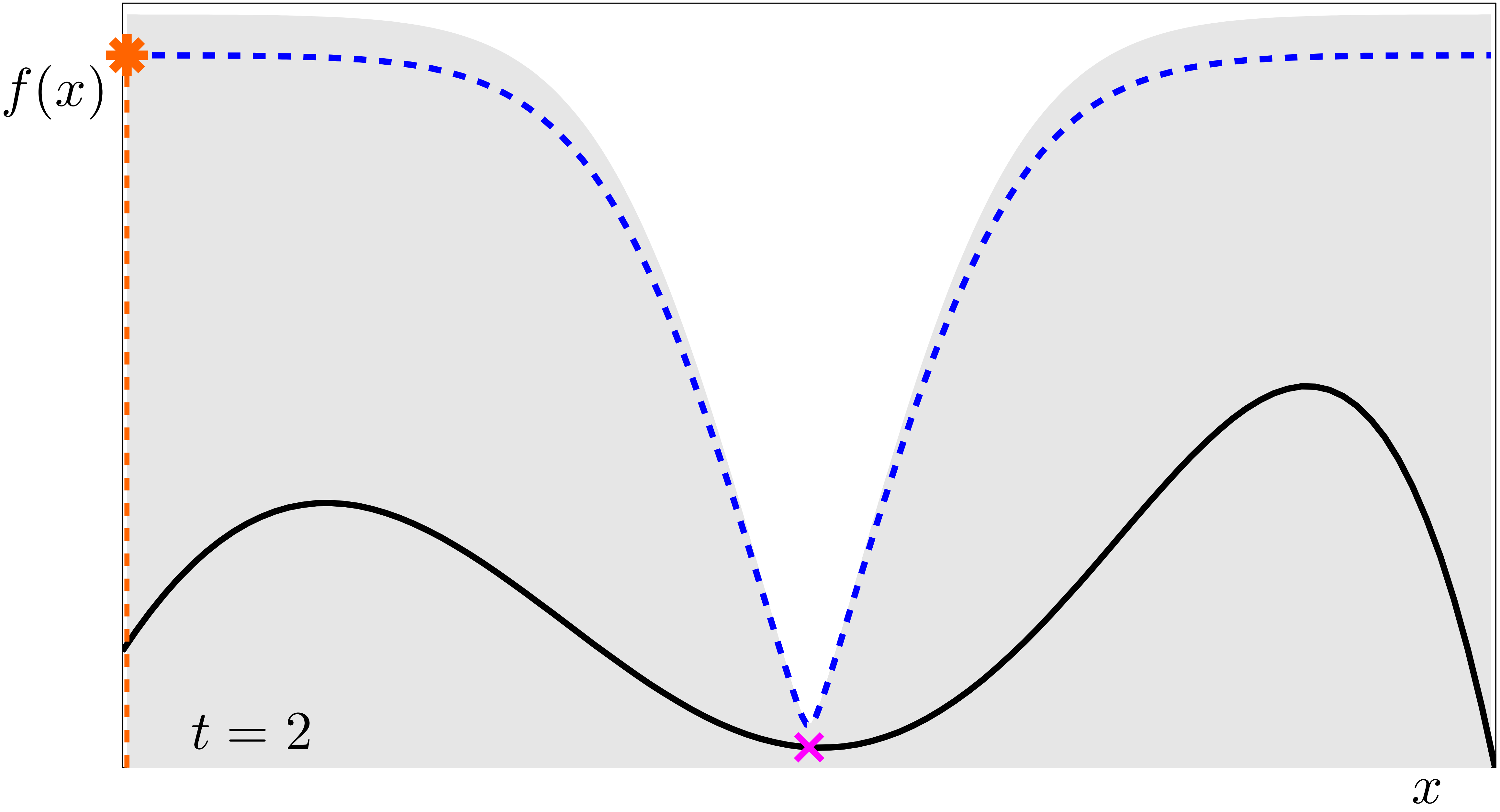
GP-UCB

(Srinivas et al. 2010)



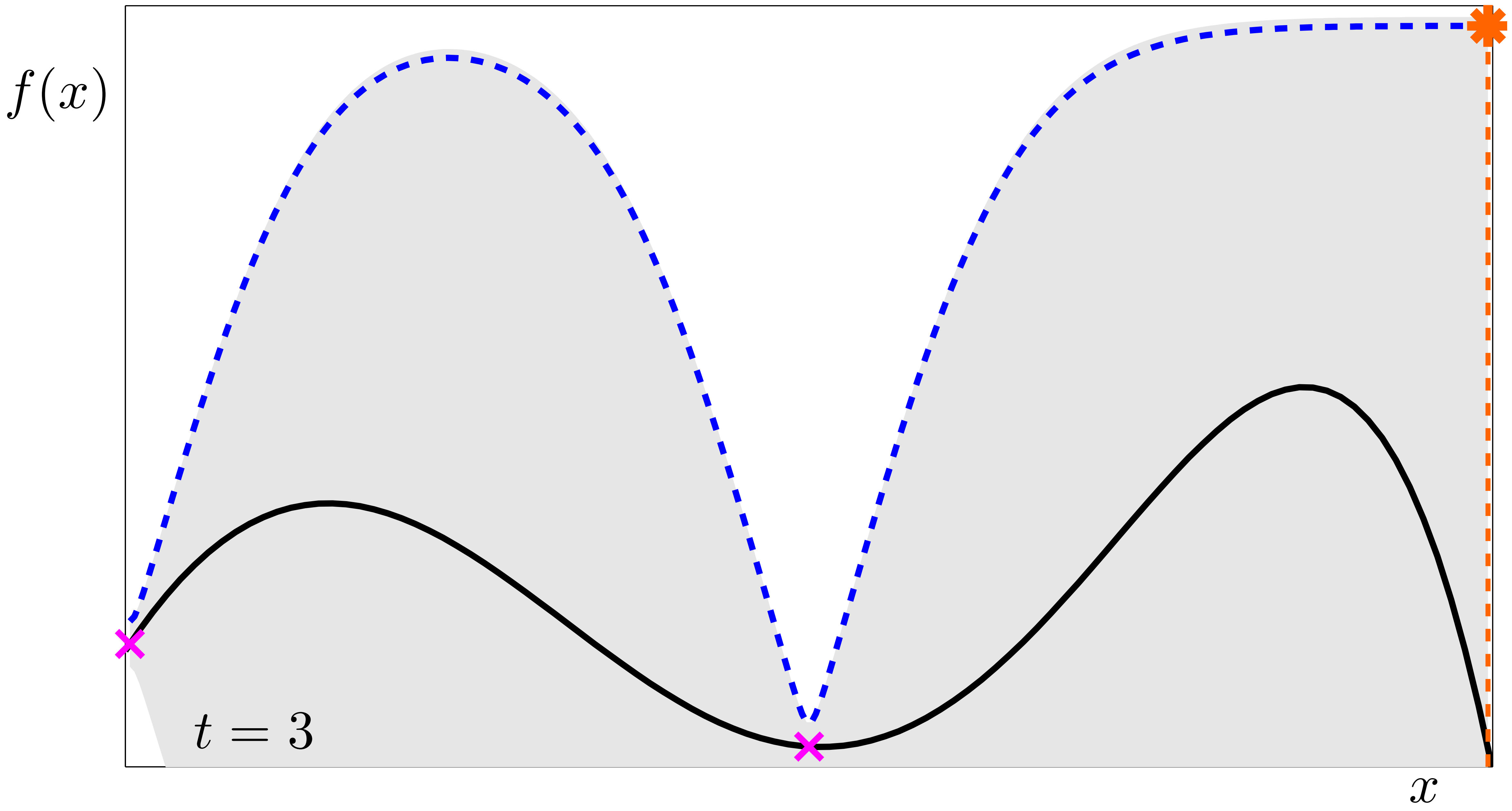
GP-UCB

(Srinivas et al. 2010)



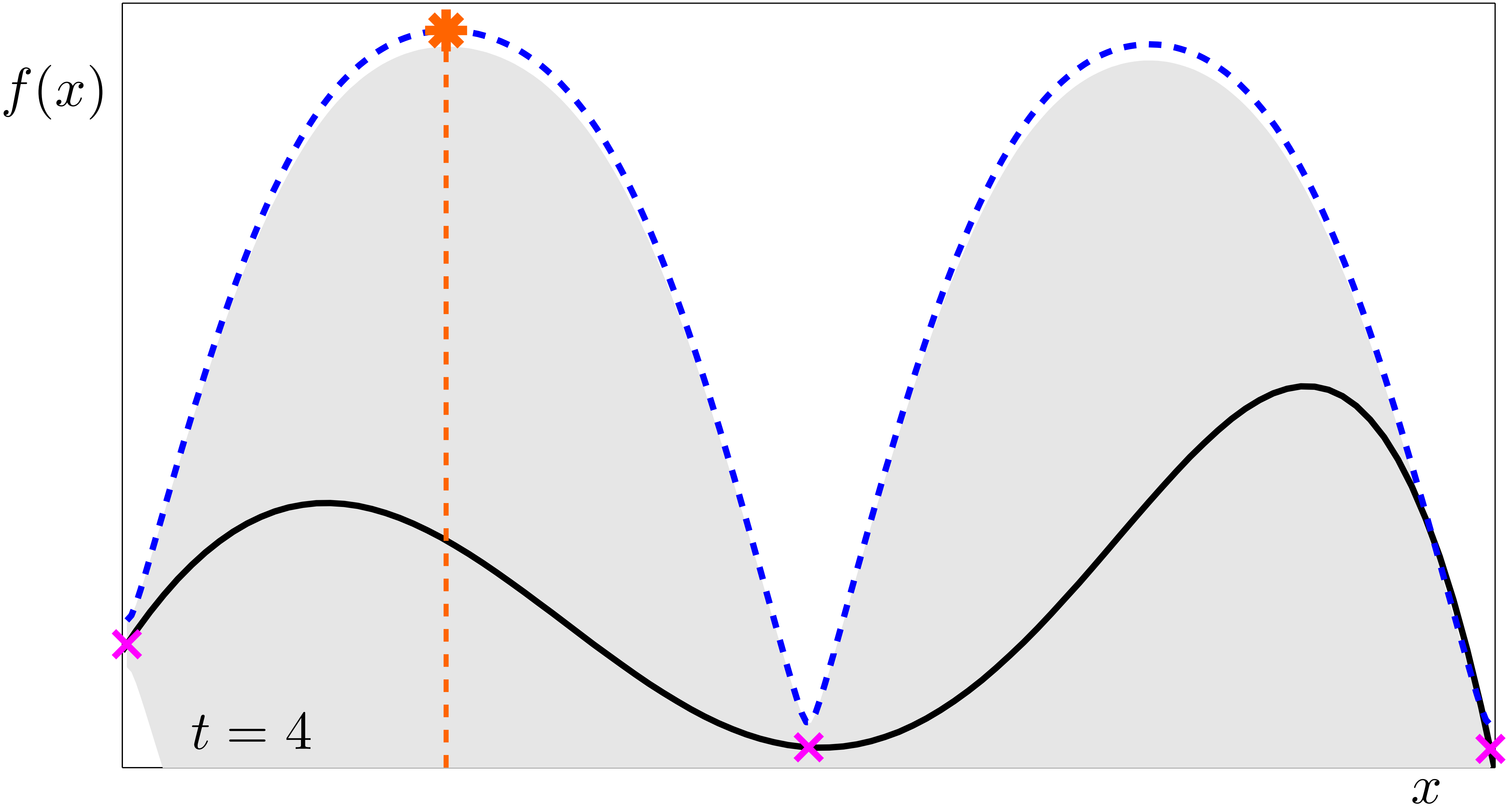
GP-UCB

(Srinivas et al. 2010)



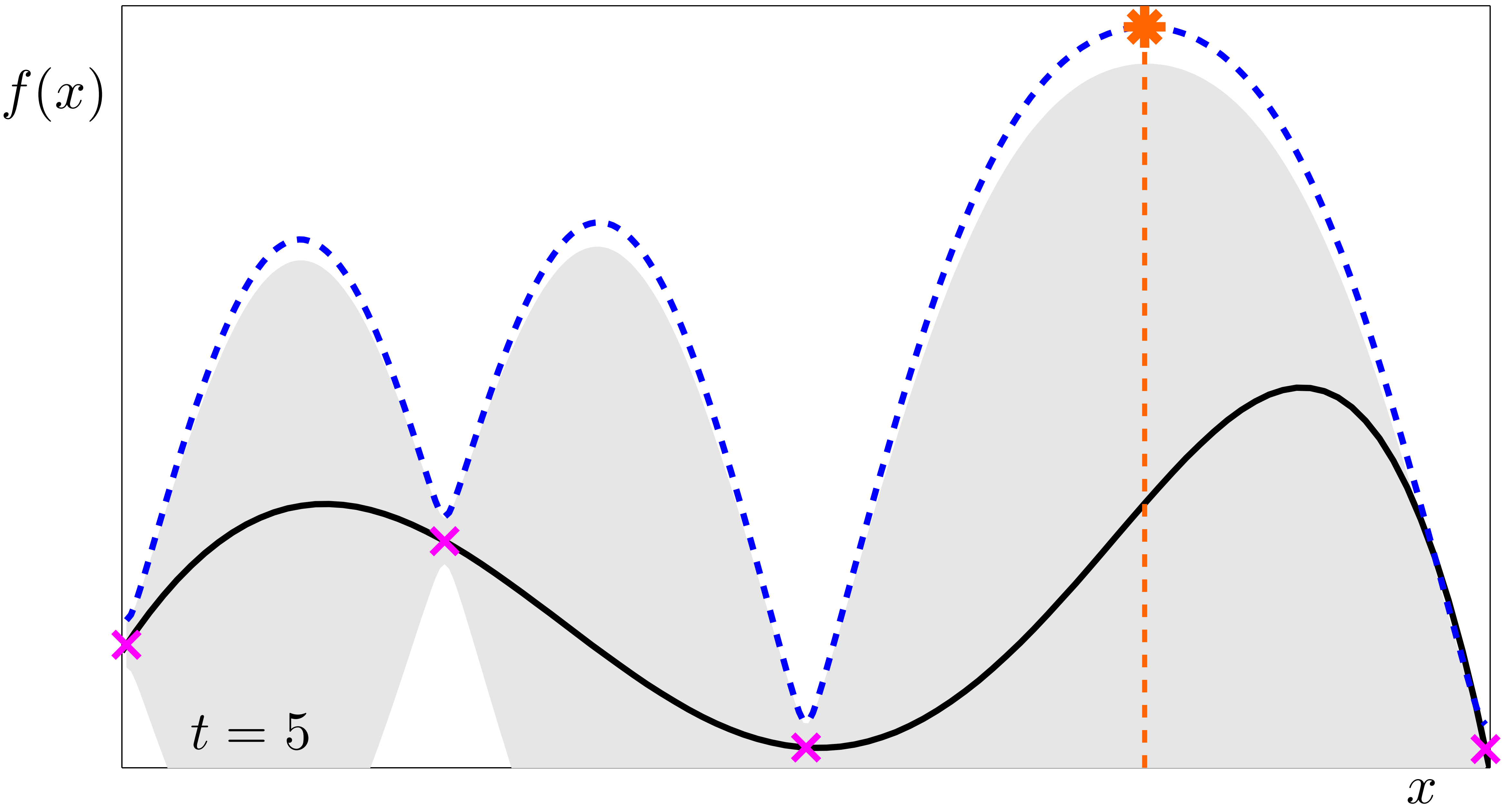
GP-UCB

(Srinivas et al. 2010)



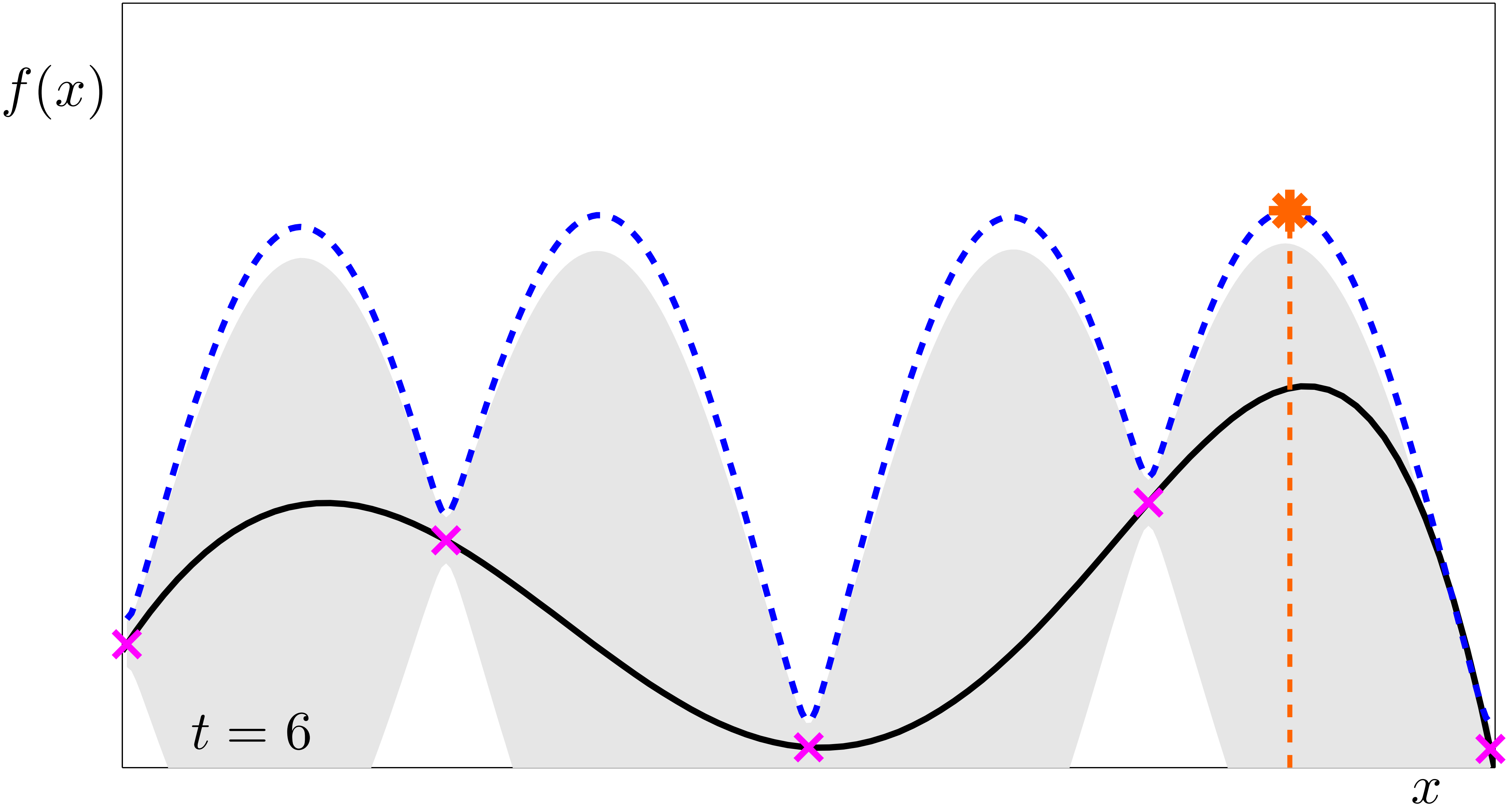
GP-UCB

(Srinivas et al. 2010)



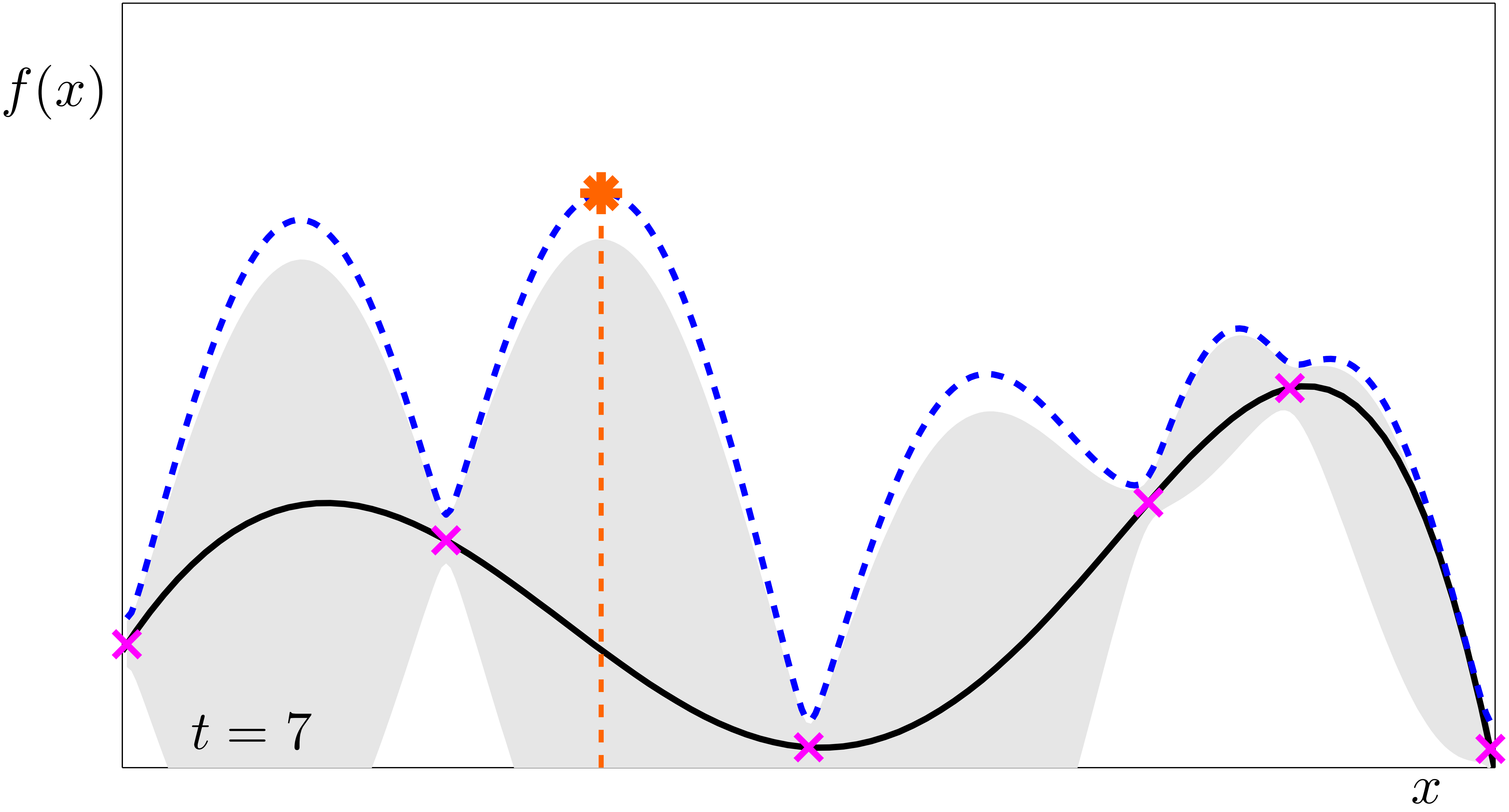
GP-UCB

(Srinivas et al. 2010)



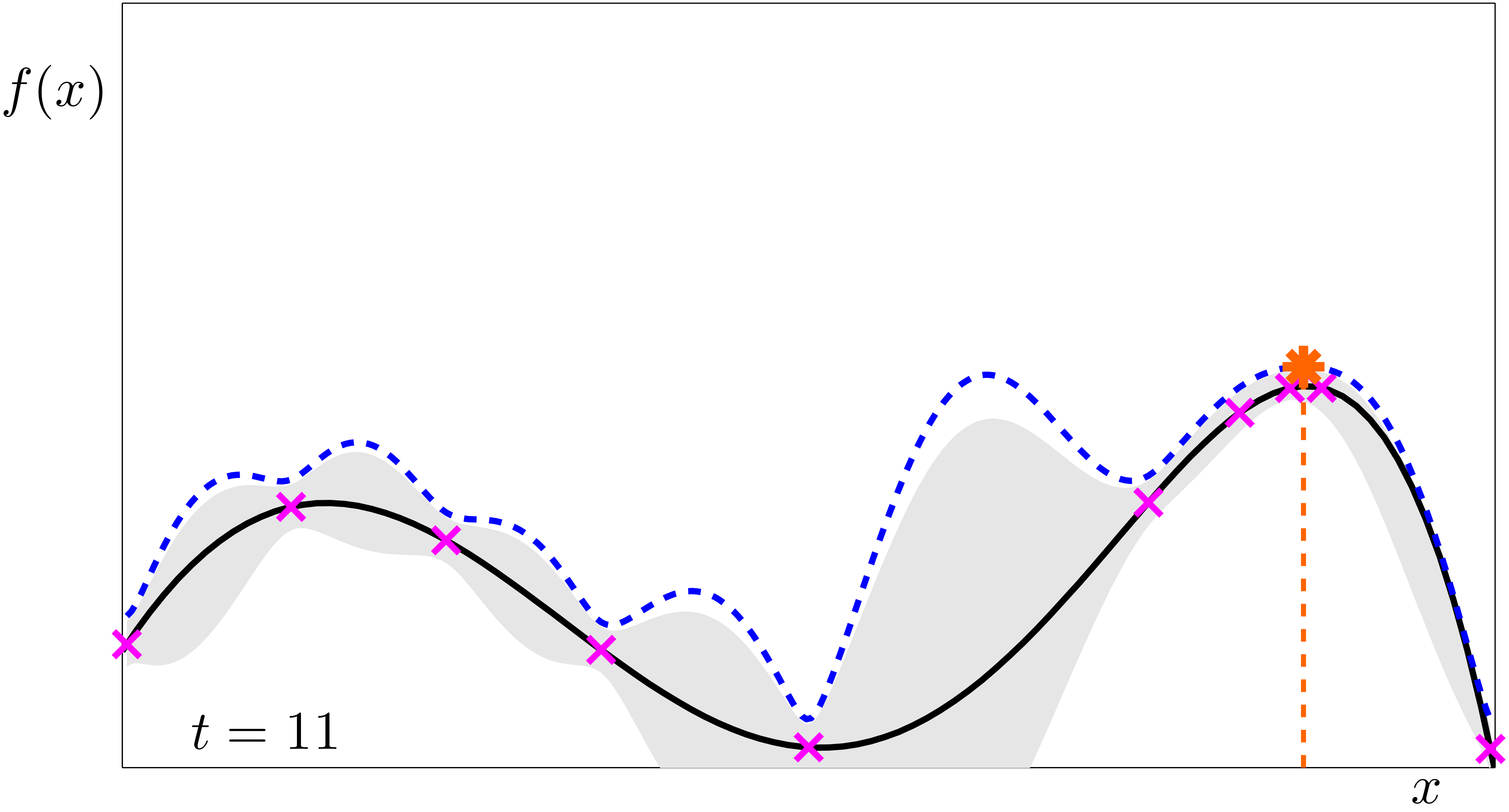
GP-UCB

(Srinivas et al. 2010)



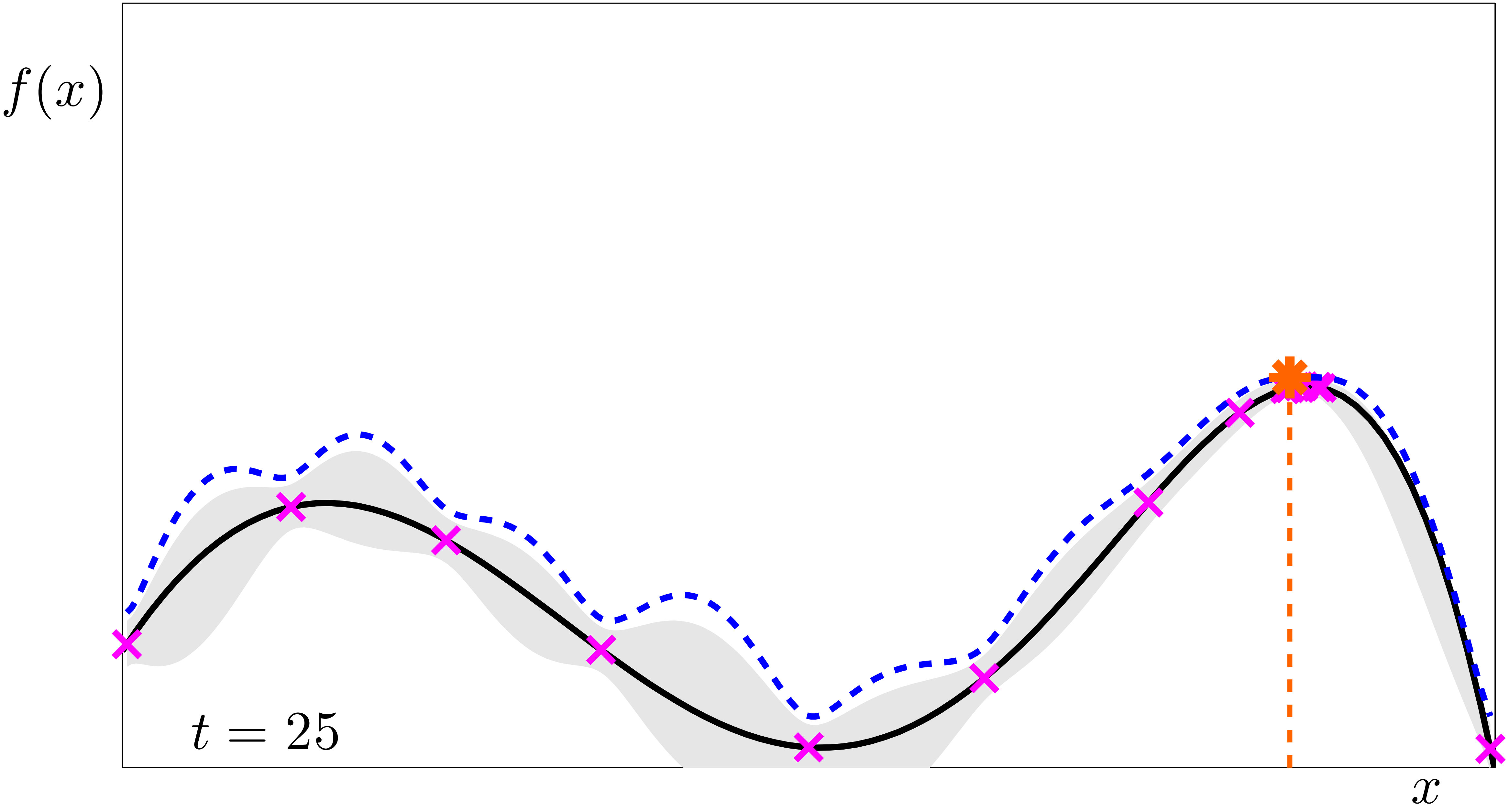
GP-UCB

(Srinivas et al. 2010)



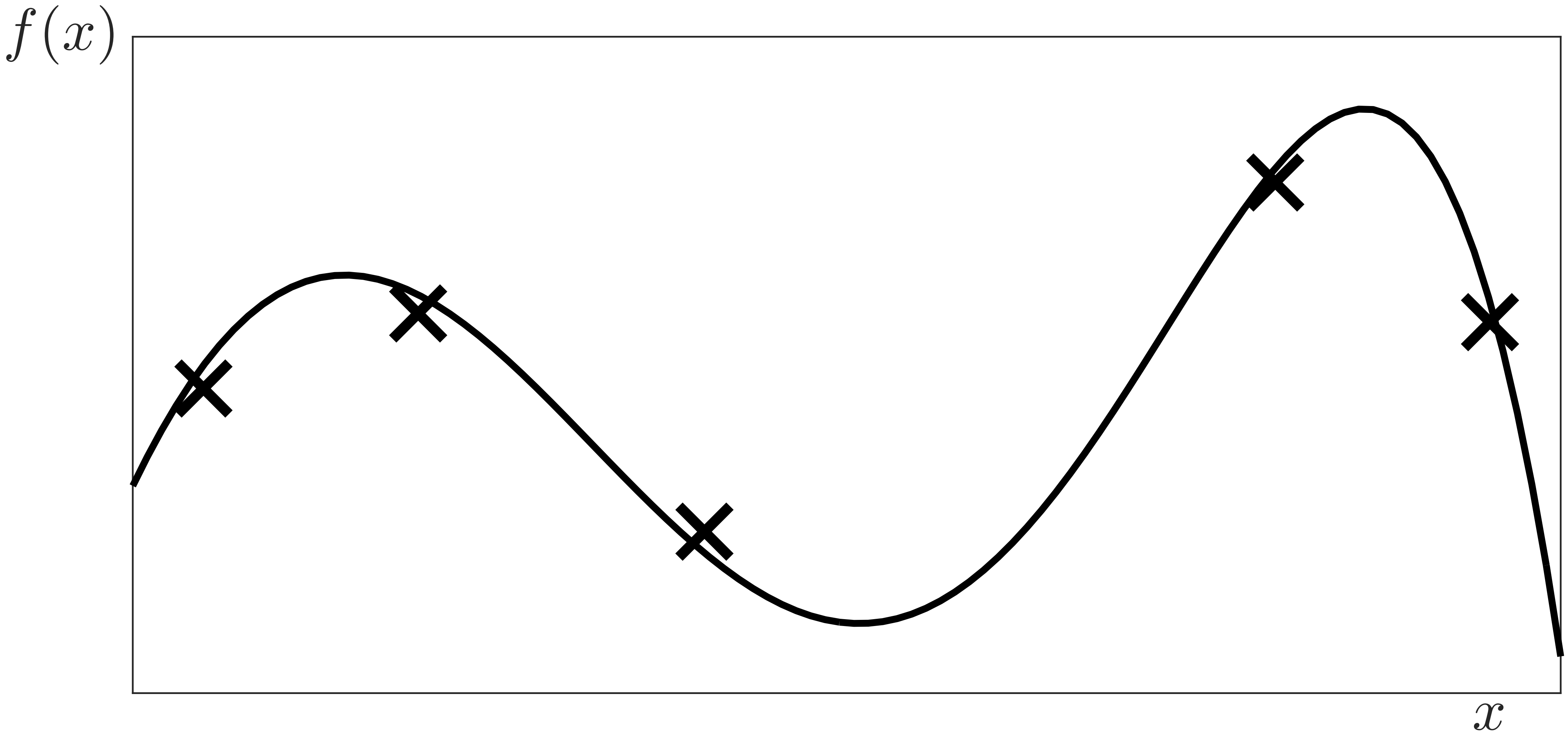
GP-UCB

(Srinivas et al. 2010)



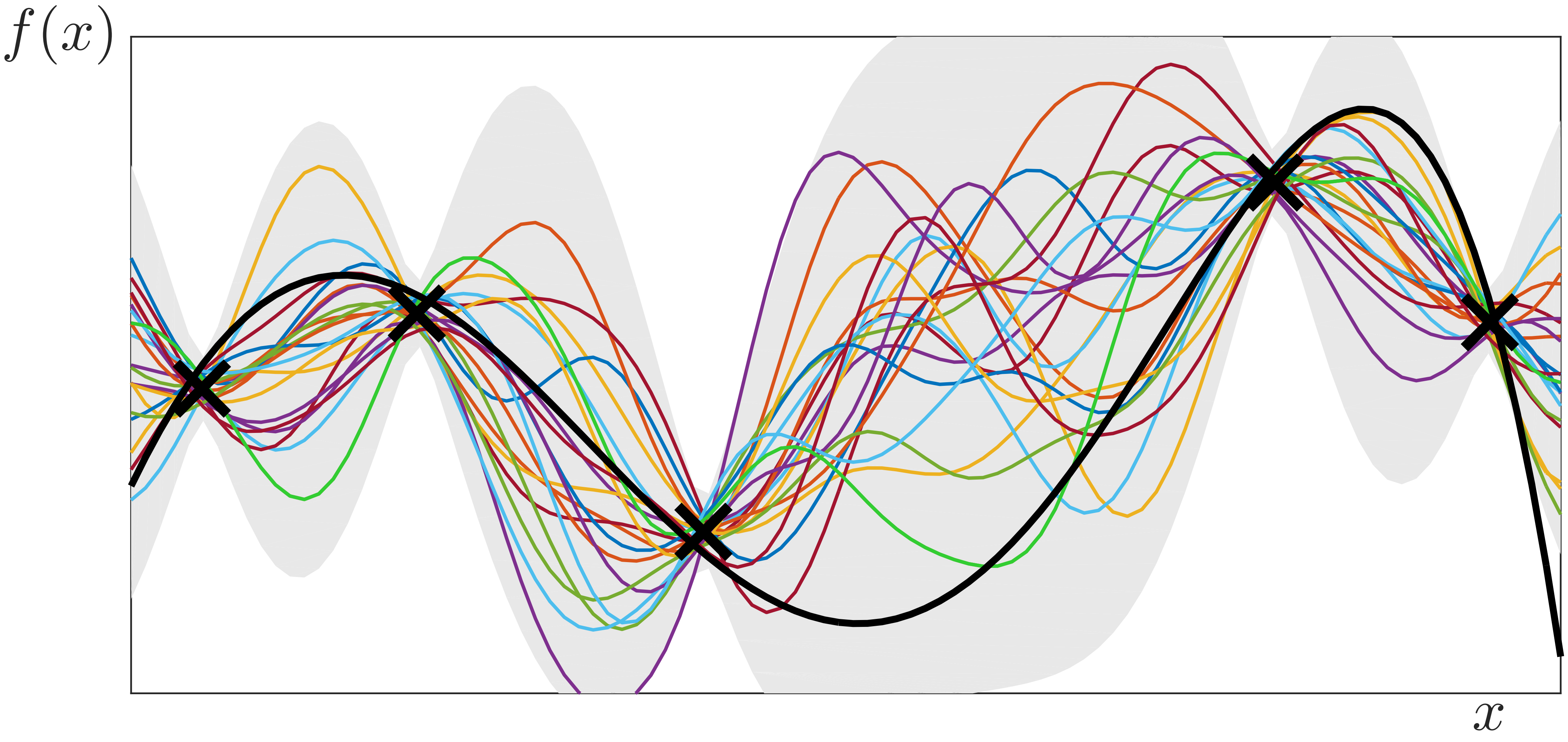
Algorithm 2: Thompson Sampling in GP Bandits

Model $f \sim \mathcal{GP}(\mathbf{0}, \kappa)$. (Thompson, 1933)



Algorithm 2: Thompson Sampling in GP Bandits

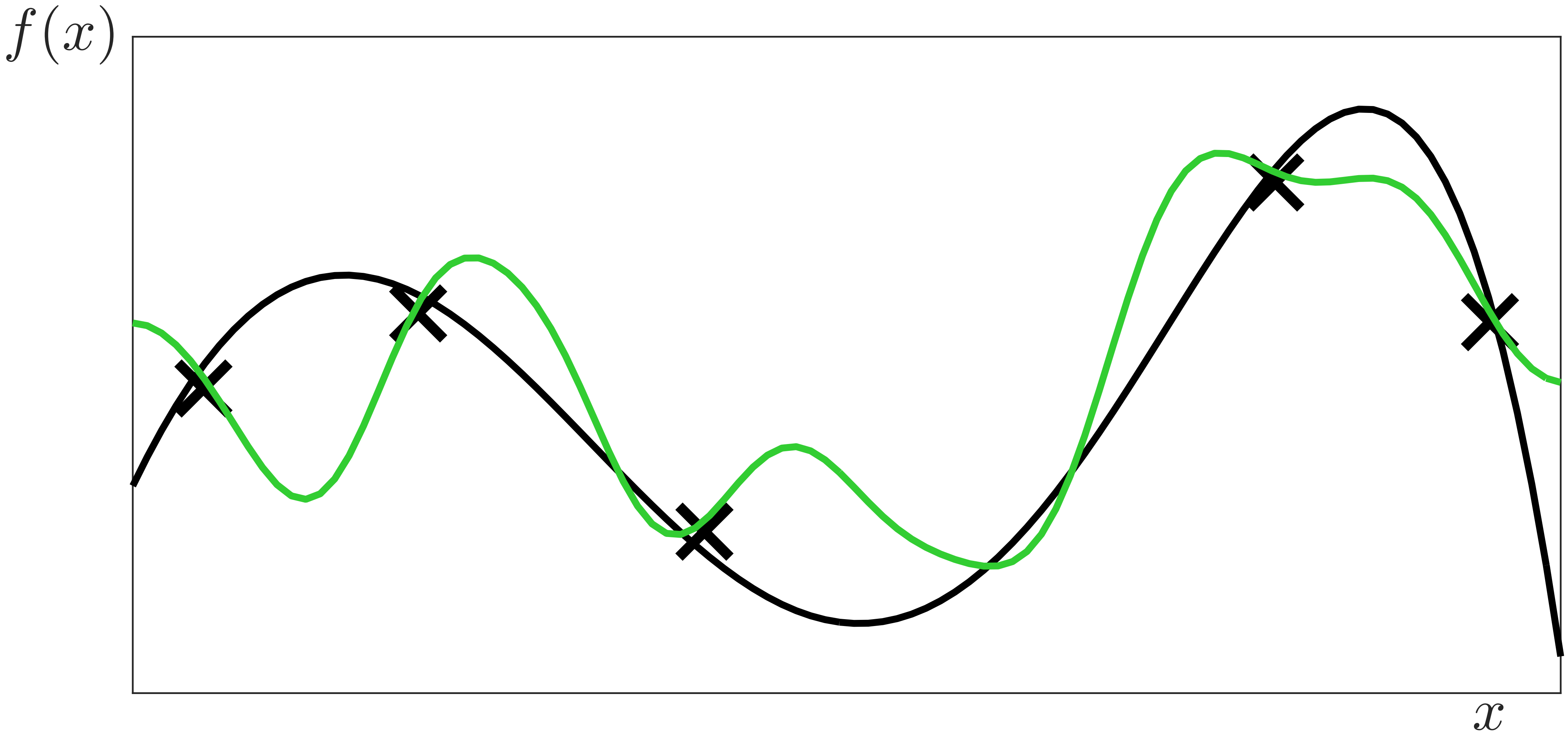
Model $f \sim \mathcal{GP}(\mathbf{0}, \kappa)$. (Thompson, 1933)



1) Construct posterior $\mathcal{GP}$.

Algorithm 2: Thompson Sampling in GP Bandits

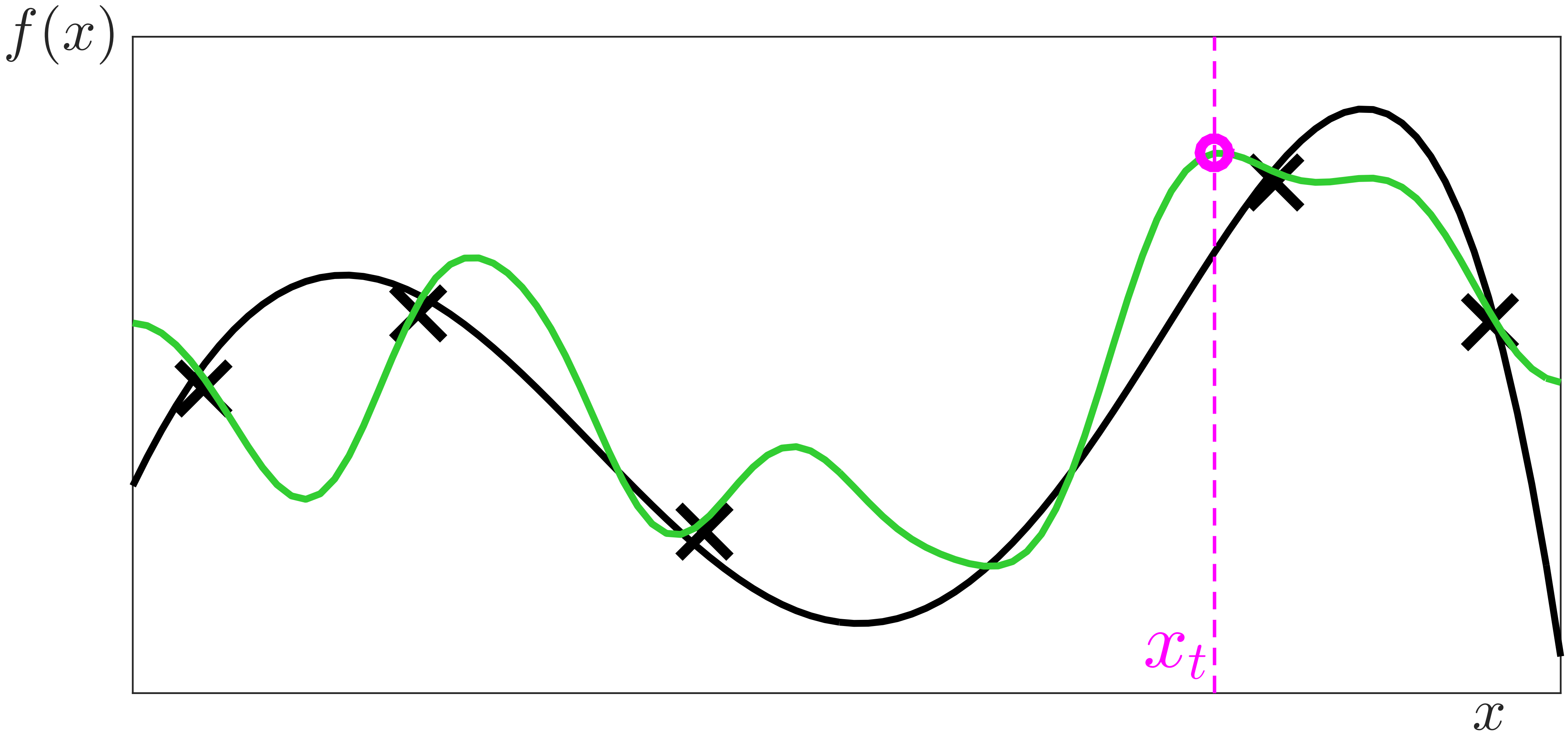
Model $f \sim \mathcal{GP}(\mathbf{0}, \kappa)$. (Thompson, 1933)



1) Construct posterior $\mathcal{GP}$. 2) Draw sample g from posterior.

Algorithm 2: Thompson Sampling in GP Bandits

Model $f \sim \mathcal{GP}(\mathbf{0}, \kappa)$. (Thompson, 1933)

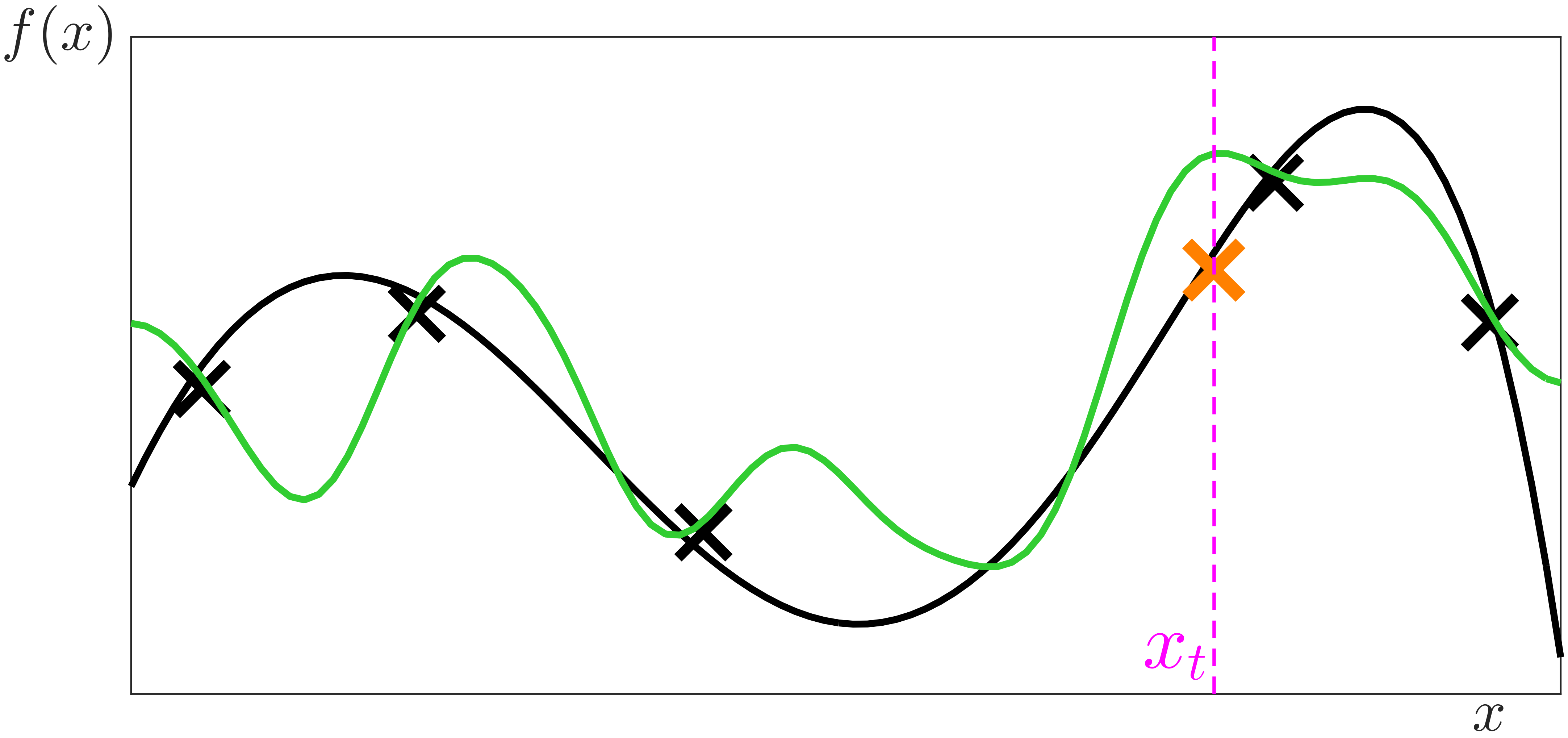


1) Construct posterior $\mathcal{GP}$. 2) Draw sample g from posterior.

3) Choose $x_t = \operatorname{argmax}_x g(x)$.

Algorithm 2: Thompson Sampling in GP Bandits

Model $f \sim \mathcal{GP}(\mathbf{0}, \kappa)$. (Thompson, 1933)



1) Construct posterior $\mathcal{GP}$.
2) Draw sample g from posterior.
3) Choose $x_t = \mathrm{argmax}_x\, g(x)$.
4) Evaluate f at x_t.