Some computational issues in nonparametric circular-circular regression

Charles Taylor

charles@maths.leeds.ac.uk

University of Leeds

joint work with

Marco Di Marzio (Pescara), Agnese Panzera (Firenze)

Brussels, May 2014

UNIVERSITY OF LEEDS

Outline

1. Nonparametric circular-circular regression: weird behaviour
2. Nonparametric Spherical regression using rotation

Outline

1. Nonparametric circular-circular regression: weird behaviour

2. Nonparametric Spherical regression using rotation

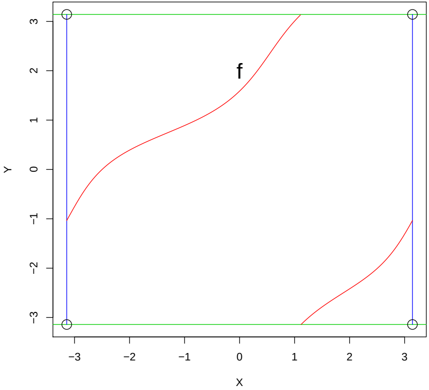
Circular Regression

(Start in $p = 1$ dimensions.)

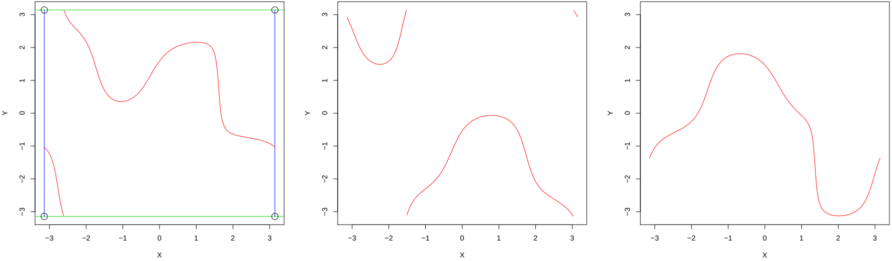
Let $Y = f(X)$ be a circular response variable and X a circular explanatory variable.

$(X, Y) \in [-\pi, \pi) \times [-\pi, \pi)$

Consider $f(X)$.



More examples



Nonparametric Estimator

Given a model

$$Y_i = [m(x_i) + \varepsilon_i] \mod 2\pi, \quad i = 1, \ldots, n$$

we initially compute estimates of $\mathrm{E}(\sin \mathrm{Y} \mid \mathrm{x})$ and $\mathrm{E}(\cos \mathrm{Y} \mid \mathrm{x})$:

$$\hat{g}_1(x) = \frac{1}{n}\sum_{i=1}^{n} \sin(Y_i) W(x_i - x) \quad \text{and} \quad \hat{g}_2(x) = \frac{1}{n}\sum_{i=1}^{n} \cos(Y_i) W(x_i - x),$$

where W is a local weight, and then the estimator of m is

$$\hat{m}(x) = \mathrm{atan2}[\hat{g}_1(x), \hat{g}_2(x)].$$

Nadaraya-Watson (local constant)

For a von Mises kernel, the weights take the form

$$W(x_i - x) = W(x_i - x; \kappa) \propto \exp(\kappa \cos(x_i - x))$$

Here κ is (the inverse of) a smoothing parameter.

Note that, in order to obtain $\hat{m}(x)$, we can forget the normalizing constant since

$$\mathrm{atan2}[y, x] = \mathrm{atan2}[cy, cx],$$

when $c > 0$.

Nadaraya-Watson (local constant)

For a von Mises kernel, the weights take the form

$$W(x_i - x) = W(x_i - x; \kappa) \propto \exp(\kappa \cos(x_i - x))$$

Here κ is (the inverse of) a smoothing parameter.

Note that, in order to obtain $\hat{m}(x)$, we can forget the normalizing constant since

$$\mathrm{atan2}[y, x] = \mathrm{atan2}[cy, cx],$$

when $c > 0$.

Nadaraya-Watson (local constant)

For a von Mises kernel, the weights take the form

$$W(x_i - x) = W(x_i - x; \kappa) \propto \exp(\kappa \cos(x_i - x))$$

Here κ is (the inverse of) a smoothing parameter.

Note that, in order to obtain $\hat{m}(x)$, we can forget the normalizing constant since

$$\mathrm{atan2}[y, x] = \mathrm{atan2}[cy, cx],$$

when $c > 0$.

Local linear

For local linear polynomial fitting we take

$$W(\theta_i - \theta) \propto K_\kappa(\theta_i - \theta)\left\{\sum_{j=1}^{n} K_\kappa(\theta_j - \theta)\sin^2(\theta_j - \theta)\sin(\theta_i - \theta)\sum_{j=1}^{n} K_\kappa(\theta_j - \theta)\sin(\theta_j - \theta)\right\},$$

with K_κ a circular kernel which, in general

i) admits a uniformly convergent Fourier series $\{1 + 2\sum_{j=1}^{\infty}\gamma_j(\kappa)\cos(j\theta)\}/(2\pi)$, $\theta \in \mathbb{T}$, where $\gamma_j(\kappa)$ is a strictly monotonic function of κ;

ii) $\int_{\mathbb{T}} K_\kappa = 1$, and, if K_κ takes negative values, there exists $0 < M < \infty$ such that, for all $\kappa > 0$

$$\int_{\mathbb{T}} |K_\kappa(\theta)|\, d\theta \leq M\,;$$

iii) for all $0 < \delta < \pi$,

$$\lim_{\kappa\to\infty} \int_{\delta \leq |\theta| \leq \pi} |K_\kappa(\theta)|\, d\theta = 0.$$

Local linear

For local linear polynomial fitting we take

$$W(\theta_i-\theta) \propto K_\kappa(\theta_i-\theta)\left\{\sum_{j=1}^{n} K_\kappa(\theta_j-\theta)\sin^2(\theta_j-\theta)\sin(\theta_i-\theta)\sum_{j=1}^{n} K_\kappa(\theta_j-\theta)\sin(\theta_j-\theta)\right\},$$

with K_κ a circular kernel which, in general

i) admits a uniformly convergent Fourier series $\{1+2\sum_{j=1}^{\infty}\gamma_j(\kappa)\cos(j\theta)\}/(2\pi)$, $\theta\in\mathbb{T}$, where $\gamma_j(\kappa)$ is a strictly monotonic function of κ;

ii) $\int_{\mathbb{T}} K_\kappa = 1$, and, if K_κ takes negative values, there exists $0 < M < \infty$ such that, for all $\kappa > 0$

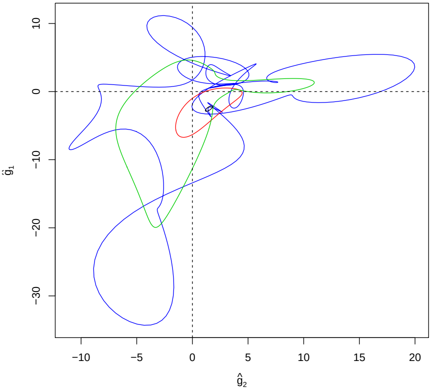
$$\int_{\mathbb{T}} |K_\kappa(\theta)|\, d\theta \le M\,;$$

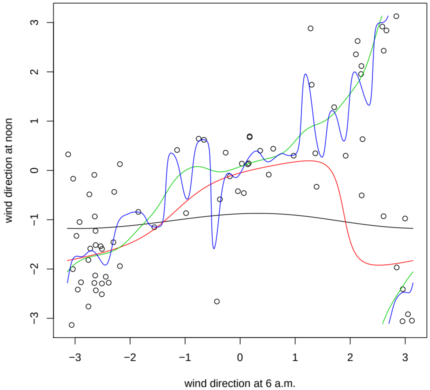
iii) for all $0<\delta<\pi$,

$$\lim_{\kappa\to\infty}\int_{\delta\le|\theta|\le\pi} |K_\kappa(\theta)|\, d\theta = 0.$$

Local Constant Example — and Issues

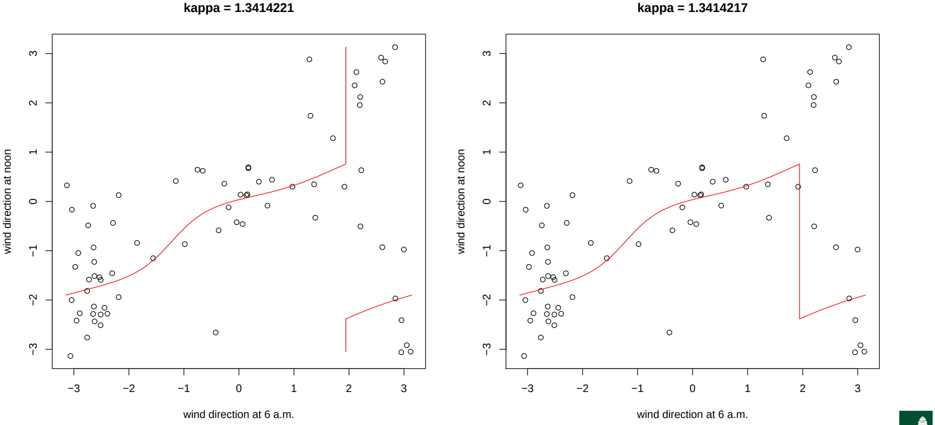
Wind direction at noon (Y_i), and at 6am (x_i).
Given κ (100,10,1,0.1), we can plot $\hat{g}_1(x)$ vs $\hat{g}_2(x)$ and $\hat{m}(x)$:





Weird Behaviour

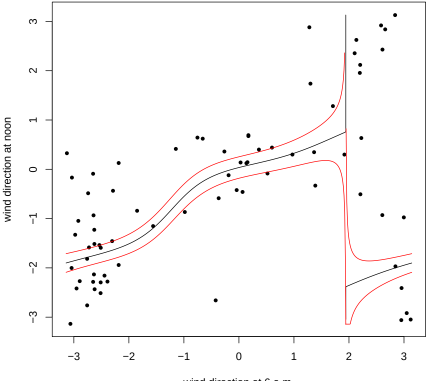
When $\hat{g}_1(x) = \hat{g}_2(x) = 0$ for some x (i.e. the curve passes through the origin) the resulting estimate is unstable:

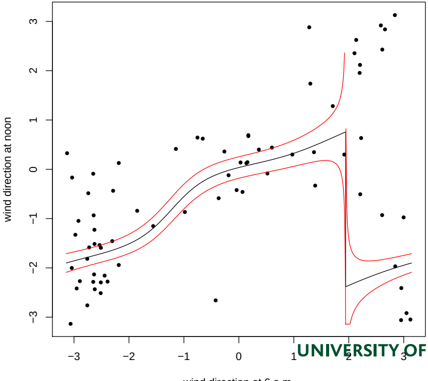


Weird Behaviour CIs

Assuming von Mises errors (observed – predicted), and estimating the concentration parameter, $\hat{\nu}$, then the variance can be estimated by

$$\hat{\tau}(\theta) = \frac{A_1(\hat{\nu})^2 R(K_\kappa)}{n\,\hat{\nu}\,\{\hat{g}_1^2(\theta) + \hat{g}_2^2(\theta)\}^{1/2}} \quad \text{where} \quad A_i(\nu) = \frac{I_i(\nu)}{I_0(\nu)}$$





Outline

1. Nonparametric circular-circular regression: weird behaviour
2. Nonparametric Spherical regression using rotation

Parametric Approach

Given “paired” data $(x_i, y_i), i = 1, \ldots, n$ where $x, y \in \mathbb{R}^p$ are on the unit sphere, then a possible model is

$$y_i \sim M_p(Rx_i, \kappa) \quad i = 1, \ldots, n$$

We can estimate a rotaion matrix R to minimize

$$\sum ||y_i - Rx_i||^2$$

by using a SVD. That is, if $Y'X = UDV'$ with D a diagonal matrix containing the singuar values of $Y'X$ and U and V orthogonal matrices with determinant 1, then $\hat{R} = U'V$.

Nonparametric Version

Now we suppose that

$$y_i \sim M_p(R_i x_i, \kappa) \quad i = 1, \ldots, n$$

noting that R_i is, in general, not identifiable. However, if we suppose that $R_i \approx R_j$ when $x_i \approx x_j$, then, to predict y for a new point x we find R_x to be a weighted least squares solution to minimize

$$\sum ||y_i - R_x x_i||^2 w_i(x)$$

with weights $w_i(x) = w(x_i - x)$ chosen to reflect the distance from x_i to x.

Easily solved (given w) by $\hat{R}_x = U'V$ where UDV' is the SVD of $Y'W'WX$, and $W = \text{diag}(w_1, \ldots, w_n)$.

Nonparametric Version

Now we suppose that

$$y_i \sim M_p(R_i x_i, \kappa) \quad i = 1, \ldots, n$$

noting that R_i is, in general, not identifiable. However, if we suppose that $R_i \approx R_j$ when $x_i \approx x_j$, then, to predict y for a new point x we find R_x to be a weighted least squares solution to minimize

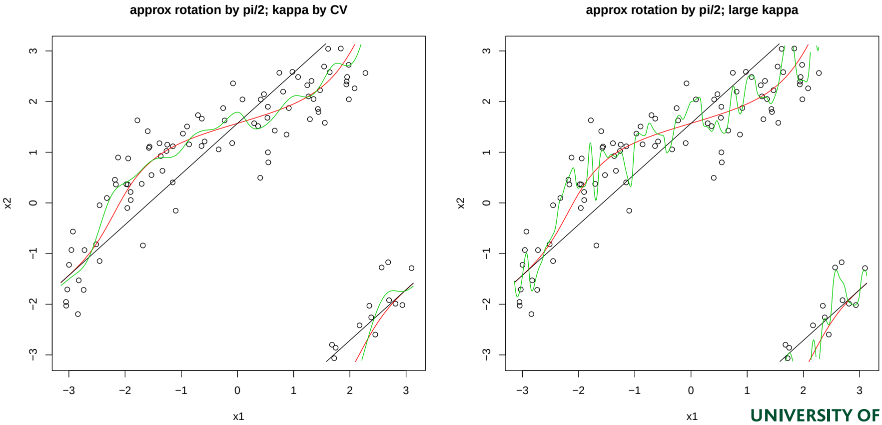
$$\sum ||y_i - R_x x_i||^2 w_i(x)$$

with weights $w_i(x) = w(x_i - x)$ chosen to reflect the distance from x_i to x.

Easily solved (given w) by $\hat{R}_x = U'V$ where UDV' is the SVD of $Y'W'WX$, and $W = \text{diag}(w_1, \ldots, w_n)$.

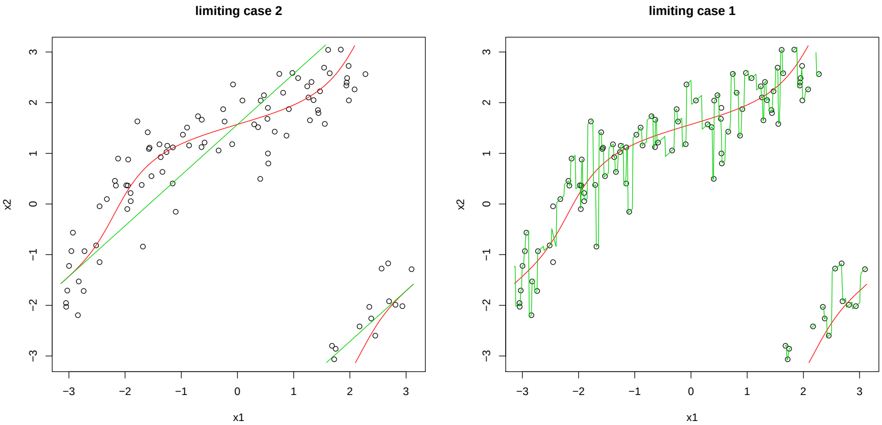
Weights, and $2 - d$ simulated example

The weights w_i can be determined using a kernel function, with a smoothing (concentration) parameter, κ. Obvious choice is to use a circular, or spherical density, for example $w_i(x) \propto \exp(\kappa x' x_i)$ As usual, κ may be chosen by cross-validation.



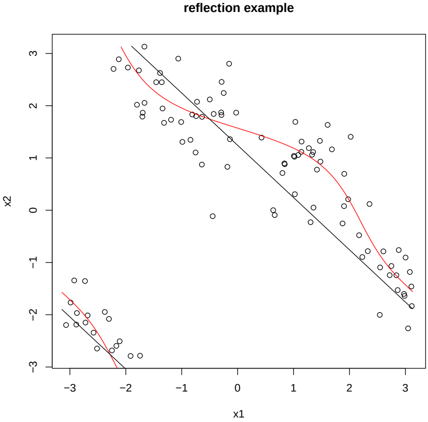
Limiting cases

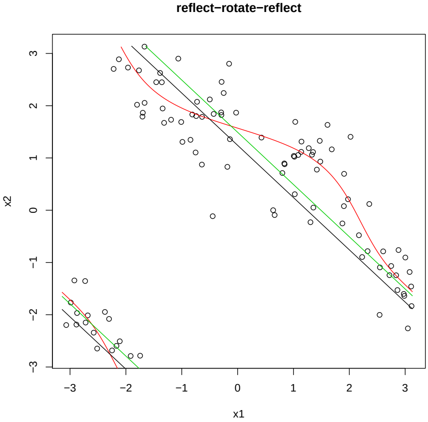
As expected, when $\kappa = 0$ we obtain the rigid solution. The other extreme is limited by the **dvm** function.



What about reflections (rigid case)?

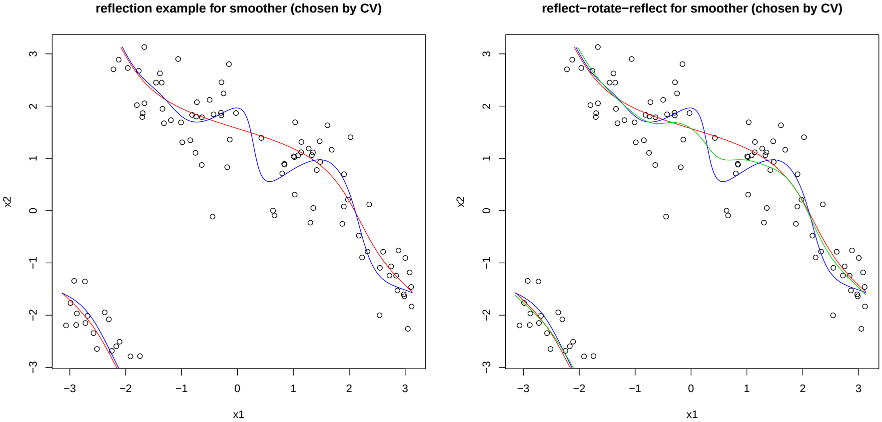
In this case we would obtain $|\hat{R}| = -1$. This solution has mean error $(\sum(1 - \hat{y}_i' y_i)/n)$ equal to 0.161. Alternatively, if we reflect y_i then find $\hat{R}$ then reflect back, the mean error is 0.117.





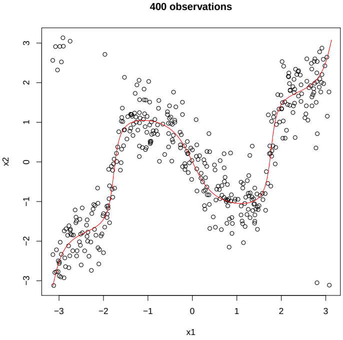
What about reflections (weighted case)?

The same thing happens for the weighted solution.
Errors are 0.040 and 0.009 respectively.



A reflection strategy for weighted solution

Choosing the smoothing parameter by cross-validation also gives an estimate (with a constant) of the error for this choice.
So, we can consider two solutions: one based on the original data, and one on the reflected data.

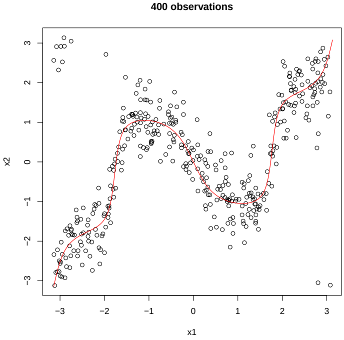


We take several samples of size n. Each sample we obtain the CV error, and the error based on the CV solution. We also compute the circular correlation between the angles.

In this example, we have:
$\rho = 0.118$,
error (CV)$= 0.136, 0.135$
errors$= 0.0064, 0.0062$

A reflection strategy for weighted solution

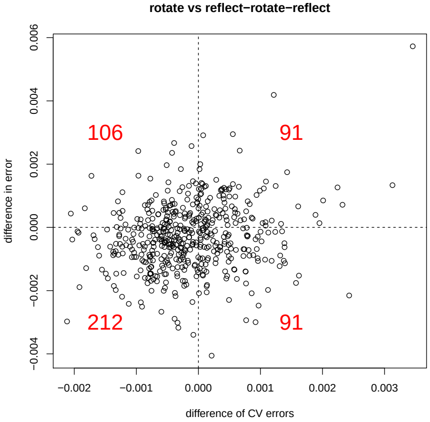
Choosing the smoothing parameter by cross-validation also gives an estimate (with a constant) of the error for this choice.
So, we can consider two solutions: one based on the original data, and one on the reflected data.



We take several samples of size n. Each sample we obtain the CV error, and the error based on the CV solution. We also compute the circular correlation between the angles.
In this example, we have:
$\rho = 0.118$,
error (CV)$= 0.136, 0.135$
errors$= 0.0064, 0.0062$

Proposed strategy results

For 500 samples of size 400 we have these average errors ($\times 10^{-3}$)



Fixed strategy		Adapt. strategy	
Normal	Reflect	CV	$\rho > 0$
8.87	9.17	8.91	9.04

Some theory

The model is

$$y_i \sim M_p(R_i x_i, \kappa) \quad i = 1, \ldots, n$$

Now for x close to x_i we suppose $R_i = \Gamma(x, x_i) R_x$ with $\Gamma(x, x_i)$ another rotation matrix.

Now write $\Gamma(x, x_i) = \exp(\Lambda(x, x_i))$, where Λ is skew-symmetric, and let $\Lambda(x, x_i) = (1 - x_i' x) S_x$.

Then we have

$$R_i x_i \approx (I + (1 - x_i' x) S_x) R_x x_i + \cdots$$

Using only the first term gives a local estimate of R_x as the solution to the previous weighted orthogonal Procrustes problem, i.e.

$$R_x = \text{argmin}_{R \in SO(d)} \sum ||y_i - R x_i||^2 w_i(x)$$

Some theory

The model is

$$y_i \sim M_p(R_i x_i, \kappa) \quad i = 1, \ldots, n$$

Now for x close to x_i we suppose $R_i = \Gamma(x, x_i) R_x$ with $\Gamma(x, x_i)$ another rotation matrix.

Now write $\Gamma(x, x_i) = \exp(\Lambda(x, x_i))$, where Λ is skew-symmetric, and let $\Lambda(x, x_i) = (1 - x_i' x) S_x$.

Then we have

$$R_i x_i \approx (I + (1 - x_i' x) S_x) R_x x_i + \cdots$$

Using only the first term gives a local estimate of R_x as the solution to the previous weighted orthogonal Procrustes problem, i.e.

$$R_x = \text{argmin}_{R \in SO(d)} \sum ||y_i - R x_i||^2 w_i(x)$$

Some theory

The model is

$$y_i \sim M_p(R_i x_i, \kappa) \quad i = 1, \ldots, n$$

Now for x close to x_i we suppose $R_i = \Gamma(x, x_i)R_x$ with $\Gamma(x, x_i)$ another rotation matrix.
Now write $\Gamma(x, x_i) = \exp(\Lambda(x, x_i))$, where Λ is skew-symmetric, and let $\Lambda(x, x_i) = (1 - x_i' x)S_x$.
Then we have

$$R_i x_i \approx (I + (1 - x_i' x)S_x)R_x x_i + \cdots$$

Using only the first term gives a local estimate of R_x as the solution to the previous weighted orthogonal Procrustes problem, i.e.

$$R_x = \text{argmin}_{R \in SO(d)} \sum ||y_i - Rx_i||^2 w_i(x)$$

More theory for $2 - d$

Given $y_i = (\cos\phi_i, \sin\phi_i)', x_i = (\cos\theta_i, \sin\theta_i)'$ and $x = (\cos\gamma, \sin\gamma)'$ (for some γ). Then suppose R depends on x through some function α:

$$R_x = \begin{pmatrix} \cos\alpha(\gamma) & -\sin\alpha(\gamma) \\ \sin\alpha(\gamma) & \cos\alpha(\gamma) \end{pmatrix}.$$

The solution is given by

$$\hat{\alpha}(\gamma) = \text{atan2}(\hat{h}_1(\gamma), \hat{h}_2(\gamma)),$$

where

$$\hat{h}_1(x) = \frac{1}{n}\sum_{i=1}^{n} \sin(\phi_i - \theta_i) w_i(\gamma) \quad \text{and} \quad \hat{h}_2(x) = \frac{1}{n}\sum_{i=1}^{n} \cos(\phi_i - \theta_i) w_i(\gamma)$$

More theory for $2-d$

This estimator then has bias (first-order term):

$$\frac{\eta_2(K_\kappa)}{2}\left(\alpha''(\gamma)+2\frac{\alpha'(\gamma)f'(\gamma)}{f(\gamma)}\right)$$

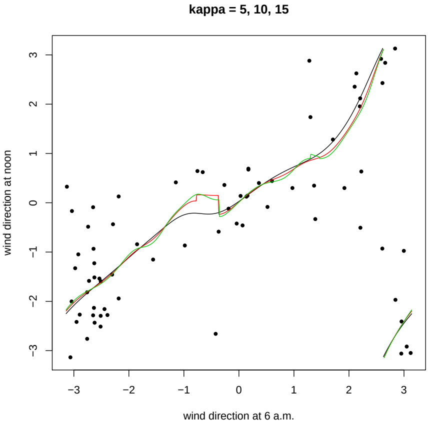
and variance (first-order term):

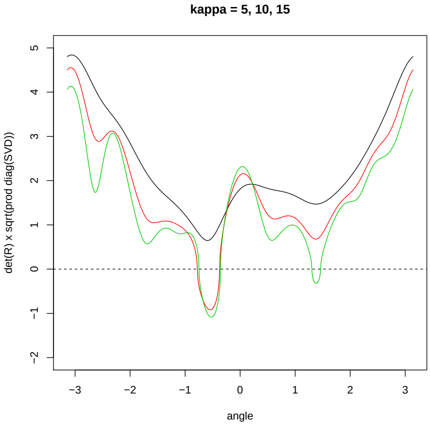
$$\frac{R(K_\kappa)}{nf(\gamma)}\left\{\cos^2 u\ s_1^2(\gamma)+\sin^2 u\ s_2^2(\gamma)-2\sin u\cos u\ c(\gamma)\right\}$$

where $u=\alpha(\gamma)+\gamma$, $\eta_2(K_\kappa)=\int\sin^2\theta K_\kappa(\theta)d\theta$, $R(K_\kappa)=\int K_\kappa^2(\theta)d\theta$
and
$s_1^2(\theta)=\mathrm{var}(\sin\phi\mid\theta)$, $s_2^2(\theta)=\mathrm{var}(\cos\phi\mid\theta)$, $c(\theta)=\mathrm{E}(\sin\phi\cos\phi\mid\theta)$

More weird behaviour

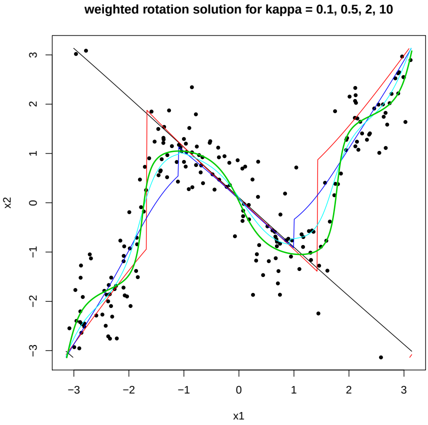
A small change in κ can give an unstable solution.

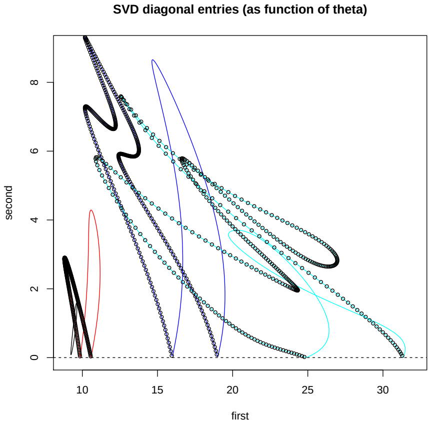




Weird behaviour (simulated data)

When the determinant of $\hat{R}$ is nearly zero, the estimate is unstable.





Extending the theory to higher dimensions

The solution to the weighted least squares solution is immediate.

The theory (bias and variance) is less obvious.

Extending the theory to higher dimensions

The solution to the weighted least squares solution is immediate.

The theory (bias and variance) is less obvious.

Extensions to Shape Analysis?

Could this weighted solution also be used for translation and scale?

Set x to each landmark in turn, obtaining a weighted solution for that landmark — then need to combine in some way, to obtain a "global" solution.

Extensions to Shape Analysis?

Could this weighted solution also be used for translation and scale?

Set x to each landmark in turn, obtaining a weighted solution for that landmark — then need to combine in some way, to obtain a "global" solution.

References

- M. Di Marzio, A. Panzera and C.C. Taylor (2011) Density estimation on the torus. *Journal of Statistical Planning and Inference*, **141**. 2156–2173.
- C.C.Taylor, K.V. Mardia, M. Di Marzio, and A. Panzera (2012) Validating protein structure using kernel density estimates. *Journal of Applied Statistics*, **39**, 2379–2388.
- J.K. MacKenzie (1957) The estimation of an orientation relationship. *Acta Cryst.* **10**, 61–62.
- M.A. Stephens (1979) Vector correlation. *Biometrika* **66**, 41–48.
- G. Wahba (1965) A least squares estimate of satellite attitude. *SIAM Review*, **7**, 409.