

# Package ‘aspline’

June 9, 2022

**Type** Package

**Title** Spline Regression with Adaptive Knot Selection

**Version** 0.2.0

**Maintainer** Vivien Goepf <vivien.goepf@gmail.com>

**Description** Perform one-dimensional spline regression with automatic knot selection.  
This package uses a penalized approach to select the most relevant knots.  
B-splines of any degree can be fitted. More details in 'Goepf et al. (2018)',  
"Spline Regression with Automatic Knot Selection", <arXiv:1808.01770>.

**Depends** R (>= 2.10)

**License** GPL-3

**Encoding** UTF-8

**LazyData** true

**URL** <https://github.com/goepf/aspline>

**BugReports** <https://github.com/goepf/aspline/issues>

**Imports** magrittr, ggplot2, dplyr, tidyr, splines2, Rcpp, mgcv, rlang

**RoxygenNote** 7.1.1

**LinkingTo** Rcpp

**Suggests** knitr, markdown, rmarkdown, covr

**VignetteBuilder** knitr

**NeedsCompilation** yes

**Author** Vivien Goepf [aut, cre] (<<https://orcid.org/0000-0001-6961-4260>>),  
Grégoire Nuel [aut]

**Repository** CRAN

**Date/Publication** 2022-06-09 08:00:02 UTC

**R topics documented:**

aspline . . . . .	2
bandsolve . . . . .	4
band_weight . . . . .	5
bladder . . . . .	5
block_design . . . . .	6
coal . . . . .	6
fossil . . . . .	7
helmet . . . . .	8
hessian_solver . . . . .	8
LDL . . . . .	9
lidar . . . . .	9
mat2rot . . . . .	10
montreal . . . . .	11
nmr . . . . .	11
rot2mat . . . . .	12
titanium . . . . .	12
weight_design_band . . . . .	13
wridge_solver . . . . .	14
<b>Index</b>	<b>15</b>

---

aspline	<i>Fit B-splines with automatic knot selection.</i>
---------	---

---

**Description**

Fit B-splines with automatic knot selection.

**Usage**

```
aspline(
  x,
  y,
  knots = seq(min(x), max(x), length = 42)[-c(1, 42)],
  pen = 10^seq(-3, 3, length = 100),
  degree = 3L,
  family = c("gaussian", "binomial", "poisson"),
  maxiter = 1000,
  epsilon = 1e-05,
  verbose = FALSE,
  tol = 1e-06
)

aridge_solver(
  x,
  y,
```

```

knots = seq(min(x), max(x), length = 42)[-c(1, 42)],
pen = 10^seq(-3, 3, length = 100),
degree = 3L,
family = c("gaussian", "binomial", "poisson"),
maxiter = 1000,
epsilon = 1e-05,
verbose = FALSE,
tol = 1e-06
)

```

### Arguments

x, y	Input data, numeric vectors of same length
knots	Knots
pen	A vector of positive penalty values. The adaptive spline regression is performed for every value of pen
degree	The degree of the splines. Recommended value is 3, which corresponds to natural splines.
family	A description of the error distribution and link function to be used in the model. The "gaussian", "binomial", and "poisson" families are currently implemented, corresponding to the linear regression, logistic regression, and Poisson regression, respectively.
maxiter	Maximum number of iterations in the main loop.
epsilon	Value of the constant in the adaptive ridge procedure (see <i>Frommlet, F., Nuel, G. (2016) An Adaptive Ridge Procedure for L0 Regularization.</i> )
verbose	Whether to print details at each step of the iterative procedure.
tol	The tolerance chosen to diagnostic convergence of the adaptive ridge procedure.

### Value

A list with the following elements:

- sel: list giving for each value of lambda the vector of the knot selection weights (a knot is selected if its weight is equal to 1.)
- knots\_sel: list giving for each value of lambda the vector of selected knots.
- model: list giving for each value of lambda the fitted regression model.
- par: parameters of the models for each value of lambda.
- sel\_mat: matrix of booleans whose columns indicate whether each knot is selected.
- aic, bic, and ebic: Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), and Extended BIC (EBIC) scores, for each value of lambda.
- dim: number of selected knots for each value of lambda.
- loglik: log-likelihood of the selected model, for each value of lambda.

### Functions

- aridge\_solver: Alias for aspline, for backwards compatibility.

---

bandsolve

*bandsolve*


---

### Description

Main function to solve efficiently and quickly a symmetric bandlinear system. These systems are solved much faster than standards system, dropping from complexity  $O(n^3)$  to  $O(0.5*nk^2)$ , where  $k$  is the number of sub diagonal.

### Usage

```
bandsolve(A, b = NULL, inplace = FALSE)
```

### Arguments

A	Band square matrix in rotated form. The rotated form can be obtained with the function <code>as.rotated</code> : it's the visual rotation by 90 degrees of the matrix, where subdiagonal are discarded.
b	right hand side of the equation. Can be either a vector or a matrix. If not supplied, the function return the inverse of A.
inplace	Should results overwrite pre-existing data? Default set to false.

### Value

Solution of the linear problem.

### Examples

```
A = diag(4)
A[2,3] = 2
A[3,2] = 2
R = mat2rot(A)
solve(A)
bandsolve(R)

set.seed(100)

n = 1000
D0 = rep(1.25, n)
D1 = rep(-0.5, n-1)
b = rnorm(n)
```

---

band_weight	<i>Create the penalty matrix</i>
-------------	----------------------------------

---

**Description**

Create the penalty matrix

**Usage**

```
band_weight(w, diff)
```

**Arguments**

w	Vector of weights
diff	Order of the differences to be applied to the parameters. Must be a strictly positive integer

**Value**

Weighted penalty matrix  $D^T \text{diag}(w)D$  where  $D \leftarrow \text{diff}(\text{diag}(\text{length}(w) + \text{diff}), \text{differences} = \text{diff})$ . Only the non-null superdiagonals of the weight matrix are returned, each column corresponding to a diagonal.

---

bladder	<i>Bladder Cancer aCGH profile data</i>
---------	---

---

**Description**

A dataset of 500 observations corresponding to 500 probes of the aCGH profile of a bladder cancer patient. The original data are provided by *Stransky et al. (2006)*. This dataset consists of probes 1 through 500 of individual 1.

**Usage**

```
bladder
```

**Format**

A data frame with 500 observations and 2 variables:

x probe number

y aCGH profile value

**Source**

Stransky, N., Vallot, C., Reyal, F., Bernard-Pierrot, I., de Medina, S. G. D., Segraves, R., de Rycke, Y., Elvin, P., Cassidy, A., Spraggon, C., Graham, A., Southgate, J., Asselain, B., Allory, Y., Abbou, C. C., Albertson, D. G., Thiery, J. P., Chopin, D. K., Pinkel, D. and Radvanyi, F. (2006). Regional Copy Number Independent Deregulation of Transcription in Cancer', *Nature Genetics* 38(12), 1386-1396.

---

block_design	<i>Transform a Spline Design Matrix in block compressed form</i>
--------------	--

---

**Description**

Transform a Spline Design Matrix in block compressed form

**Usage**

```
block_design(X, degree)
```

**Arguments**

X	The design matrix, as given by <code>splines2::bSpline</code> .
degree	Degree of the spline regression, as used in function <code>splines2::bSpline</code> .

**Value**

A matrix B with all non-zero entries of X and a vector of indices alpha representing the positions of the non-zero blocks of X.

---

coal	<i>Yearly number of coal mine disasters in Britain</i>
------	--

---

**Description**

A data of 112 observations registering the yearly number of coal mine disasters in Britain from 1851 to 1962. The data comes from *Diggle et al. (1988)* and has been used for spline regression by *Eilers et al. (1996)*.

**Usage**

```
coal
```

**Format**

A data frame with 112 observations and 2 variables:

<b>year</b>	year
<b>n</b>	number of coal mine disasters

**Source**

Diggle, P. and Marron, J. S. (1988). 'Equivalence of Smoothing Parameter Selectors in Density and Intensity Estimation', *Journal of the American Statistical Association* 83(403), 793-800.

**References**

Eilers, P. H. C. and Marx, B. D. (1996). 'Flexible Smoothing with B-splines and Penalties', *Statistical Science* 11(2), 89-102.

---

fossil

*Fossil data*

---

**Description**

A dataset with 106 observations on fossil shells from the SemiPar package (<https://CRAN.R-project.org/package=SemiPar>).

**Usage**

```
fossil
```

**Format**

A data frame with 106 observations and 2 variables:

**age** The age of fossils, in millions of years

**strontium.ratio** Ratio of strontium isotopes ...

**Source**

Bralower, T.T, Fullagar, P.D., Paull, C.K, Dwyer, G.S. and Leckie, R.M. (1997). Mid-cretaceous strontium-isotope stratigraphy of deep-sea sections. *Geological Society of America Bulletin*, 109, 1421-1442.

**References**

Ruppert, D., Wand, M.P. and Carroll, R.J. (2003). *Semiparametric Regression*, Cambridge University Press.

---

helmet	<i>Testing Crash Helmets</i>
--------	------------------------------

---

**Description**

A dataset containing the acceleration and time after impact of helmets from a simulated motorcycle accident.

**Usage**

helmet

**Format**

A data frame with 132 rows and 2 variables:

**x** Time after impact, in milliseconds

**y** Head acceleration, in units of  $g$  ...

**Source**

Dataset number 338 of *Hand, D. et al. (1993) A Handbook of Small Datasets.*

---

hessian_solver	<i>Inverse the hessian and multiply it by the score</i>
----------------	---

---

**Description**

Inverse the hessian and multiply it by the score

**Usage**

hessian\_solver(par, XX\_band, Xy, pen, w, diff)

**Arguments**

par	The parameter vector
XX_band	The matrix $X^T X$ where $X$ is the design matrix. This argument is given in the form of a band matrix, i.e., successive columns represent superdiagonals.
Xy	The vector of currently estimated points $X^T y$ , where $y$ is the y-coordinate of the data.
pen	Positive penalty constant.
w	Vector of weights. Has to be of length
diff	The order of the differences of the parameter. Equals degree + 1 in adaptive spline regression.



**Value**

The solution of the linear system:

$$(X^T X + pen D^T diag(w) D)^{-1} X^T y - par$$

LDL

*LDL***Description**

Fast inplace LDL decomposition of symmetric band matrix of length k.

**Arguments**

D Rotated row-wised matrix of dimensions n\*k, with first column corresponding to the diagonal, the second to the first super-diagonal and so on.

**Value**

List with D as solution of our LDL decomposition.

**Examples**

```
n=10;
D0=1:10;
D1=exp(-c(1:9));
D=cbind(D0,c(D1,0))
sol=LDL(D)
```

lidar

*Lidar data***Description**

Data from a light detection and ranging (LIDAR) experiment

**Usage**

```
lidar
```

**Format**

**range** distance travelled before the light is reflected back to its source

**logratio** logarithm of the ratio of received light from two laser sources

**Source**

- Sigrist, M. (Ed.) (1994). Air Monitoring by Spectroscopic Techniques (Chemical Analysis Series, vol. 197). New York: Wiley
- The R package <https://CRAN.R-project.org/package=SemiPar>

**References**

Ruppert, D., Wand, M.P. and Carroll, R.J. (2003). *Semiparametric Regression*, Cambridge University Press.

---

mat2rot

*Rotate a band matrix to get the rotated row-wised matrix associated.*

---

**Description**

Rotate a symmetric band matrix to get the rotated matrix associated. Each column of the rotated matrix correspond to a diagonal. The first column is the main diagonal, the second one is the upper-diagonal and so on. Artificial 0 are placed at the end of each column if necessary.

**Usage**

```
mat2rot(M)
```

**Arguments**

M                      Band square matrix or a list of diagonal.

**Value**

Rotated matrix.

**Examples**

```
A = diag(4)
A[2,3] = 2
A[3,2] = 2

## Original Matrix
A
## Rotated version
R = mat2rot(A)
R

rot2mat(mat2rot(A))
```

---

montreal	<i>Montreal Temperature Data</i>
----------	----------------------------------

---

**Description**

A dataset containing the temperature in Montreal for two years

**Usage**

```
montreal
```

**Format**

A data frame with 730 rows and 2 variables:

**day** The day of the year from January 1, 1961, to December 31, 1962

**temp** Temperature in Celsius ...

**Source**

```
fda: "MontrealTemp"
```

---

nmr	<i>Nuclear Magnetic Resonance data</i>
-----	--

---

**Description**

A signal of nuclear magnetic resonance.

**Usage**

```
nmr
```

**Format**

Data frame of 1024 rows and two columns: the index  $x$  and the signal  $y$ .

**Source**

- Data from <https://web.stanford.edu/~hastie/ElemStatLearn/datasets/nmr1.csv>.
- See also The Elements of Statistical Learning (2001, 2nd Ed.), *Hastie, T., Friedman, J., and Tibshirani, R.J.*, p. 176

---

rot2mat	<i>Get back a symmetric square matrix based on his rotated row-wised version.</i>
---------	---

---

**Description**

Get back a symmetric square matrix based on his rotated row-wised version. The rotated form of the input is such each column correspond to a diagonal, where the first column is the main diagonal and next ones are the upper/lower-diagonal. To match dimension, last element of these columns are discarded.

**Usage**

```
rot2mat(R)
```

**Arguments**

R                      Rotated matrix.

**Value**

Band square matrix.

**Examples**

```
D0 = 1:5;
D1 = c(0,1,0,0);
D2 = rep(2,3);

A = rot2mat(cbind(D0,c(D1,0),c(D2,0,0)))
A
mat2rot(rot2mat(cbind(D0,c(D1,0),c(D2,0,0))))
```

---

titanium	<i>Titanium heat data</i>
----------	---------------------------

---

**Description**

A data set of 49 samples expressing the thermal property of titanium

**Usage**

```
titanium
```

**Format**

49 observations and two variables:

**x** temperature

**y** physical property

**Source**

- de Boor, C., and Rice, J. R. (1986), Least-squares cubic spline approximation. II: variable knots. *Report CSD TR 21, Purdue U., Lafayette, IN.*
- Dierckx, P. (1993), *Curve and Surface Fitting with Splines*, Springer.
- Jupp, D. L. B. (1975), *Approximation to data by splines with free knots*, SIAM Journal on Numerical Analysis, 15: 328-343.

---

weight_design_band	<i>Fast computation of weighted design matrix for generalized linear model</i>
--------------------	--

---

**Description**

Fast computation of weighted design matrix for generalized linear model

**Usage**

```
weight_design_band(w, alpha, B)
```

**Arguments**

- |       |   |
|-------|---|
| w     | Vector of weights.  |
| alpha | Vector of indexes representing the start of blocks of the design matrix, as given by <a href="#">block_design</a> . |
| B     | Design matrix in compressed block format, as given by <a href="#">block_design</a> .                                |

**Value**

Weighted design matrix  $X^T \text{diag}(w)X$  where  $X$  is the design matrix and  $W = \text{diag}(w)$  is a diagonal matrix of weights.

---

wridge_solver	<i>Fit B-Splines with weighted penalization over differences of parameters</i>
---------------	--

---

### Description

Fit B-Splines with weighted penalization over differences of parameters

### Usage

```
wridge_solver(
  XX_band,
  Xy,
  degree,
  pen,
  w = rep(1, nrow(XX_band) - degree - 1),
  old_par = rep(1, nrow(XX_band)),
  maxiter = 1000,
  tol = 1e-08
)
```

### Arguments

XX_band	The matrix $X^T X$ where $X$ is the design matrix. This argument is given in the form of a band matrix, i.e., successive columns represent superdiagonals.
Xy	The vector of currently estimated points $X^T y$ , where $y$ is the y-coordinate of the data.
degree	The degree of the B-splines.
pen	Positive penalty constant.
w	Vector of weights. The case $\mathbf{w} = \mathbf{1}$ corresponds to fitting P-splines with difference #’ order degree + 1 (see <i>Eilers, P., Marx, B. (1996) Flexible smoothing with B-splines and penalties.</i> )
old_par	Initial parameter to serve as starting point of the iterating process.
maxiter	Maximum number of Newton-Raphson iterations to be computed.
tol	The tolerance chosen to diagnostic convergence of the adaptive ridge procedure.

### Value

The estimated parameter of the spline regression.

# Index

## \* datasets

- bladder, [5](#)
- coal, [6](#)
- fossil, [7](#)
- helmet, [8](#)
- lidar, [9](#)
- montreal, [11](#)
- nmr, [11](#)
- titanium, [12](#)

aridge\_solver (aspline), [2](#)

aspline, [2](#)

band\_weight, [5](#)

bandsolve, [4](#)

bladder, [5](#)

block\_design, [6](#), [13](#)

coal, [6](#)

fossil, [7](#)

helmet, [8](#)

hessian\_solver, [8](#)

LDL, [9](#)

lidar, [9](#)

mat2rot, [10](#)

montreal, [11](#)

nmr, [11](#)

rot2mat, [12](#)

titanium, [12](#)

weight\_design\_band, [13](#)

wridge\_solver, [14](#)