

Package: SHIP (via r-universe)

October 15, 2026

Type Package

Title Shrinkage Covariance Estimation with Prior Knowledge

Version 2.1.0

Description Estimates shrinkage covariance matrices from data using user-selected targets, including diagonal, constant-correlation, and knowledge-based targets informed by functional gene groups. The shrinkage intensity is computed analytically.

License GPL (>= 2)

URL <https://github.com/vguillemot/SHIP>

BugReports <https://github.com/vguillemot/SHIP/issues>

Encoding UTF-8

NeedsCompilation no

Imports stats

Suggests knitr, quarto, testthat (>= 3.0.0)

VignetteBuilder quarto

Config/testthat/edition 3

Roxygen list(markdown = TRUE)

Config/roxygen2/version 8.1.0

Repository <https://vguillemot.r-universe.dev>

Date/Publication 2026-09-15 17:46:15 UTC

RemoteUrl <https://github.com/vguillemot/ship>

RemoteRef HEAD

RemoteSha 821aa31dc962123e6dced979efd2bc4565e1036f

Contents

SHIP-package	2
build.target	3
check.path	4
expl	5
shrink.estim	5
target.help	6
targetCor	7
targetD	8
targetF	9
targetG	10
targetGpos	11
targetGstar	12
Index	13

SHIP-package	<i>SHIP provides shrinkage covariance estimation with user-selected targets. The available targets include diagonal, constant-correlation, and knowledge-based structures informed by functional gene groups.</i>
--------------	---

Description

Start with `build.target` to construct a target matrix, then pass it to `shrink.estim` together with the data matrix.

Author(s)

Monika Jelizarow and Vincent Guillemot

References

- J. Schaefer and K. Strimmer, 2005. A shrinkage approach to large-scale covariance matrix estimation and implications for functional genomics. *Statist. Appl. Genet. Mol. Biol.* 4:32.
- M. Jelizarow, V. Guillemot, A. Tenenhaus, K. Strimmer, A.-L. Boulesteix, 2010. Over-optimism in bioinformatics: an illustration. *Bioinformatics*. Accepted.

See Also

Useful links:

- <https://github.com/vguillemot/SHIP>
- Report bugs at <https://github.com/vguillemot/SHIP/issues>

Examples

```
data("expl")
target <- build.target(expl$x, expl$genegroups, type = "G")
estimate <- shrink.estim(expl$x, target)
estimate$lambda
```

build.target	<i>Creating a covariance target, optionally by using information from KEGG pathways.</i>
--------------	--

Description

The function `build.target()` is a wrapper function to build the various types of covariance targets: diagonal ("D"), constant correlation ("F"), knowledge based ("G", "Gpos", and "Gstar"), correlation ("cor").

Usage

```
build.target(x, genegroups = NULL, type)
```

Arguments

<code>x</code>	An $n \times p$ matrix.
<code>genegroups</code>	List of the groups each gene belongs to: each entry of the list is dedicated to a gene (identified the same way as in x). Each item of the list is thus a vector of pathway IDs.
<code>type</code>	Character string specifying the wished target: "D" for a diagonal target, "cor" for a correlation target, "G", "Gpos" and "Gstar" for a G-type target (see Jelizarow et al, 2010) and "F" for a F-target.

Value

A $p \times p$ target covariance matrix of a certain type.

Author(s)

Vincent Guillemot

References

M. Jelizarow, V. Guillemot, A. Tenenhaus, K. Strimmer, A.-L. Boulesteix, 2010. Over-optimism in bioinformatics: an illustration. Bioinformatics. Accepted.

See Also

[targetCor](#), [targetD](#), [targetF](#), [targetG](#), [targetGpos](#), [targetGstar](#),
Other main functions: [shrink.estim\(\)](#)

Examples

```
# Simulate dataset
x <- matrix(rnorm(20*30), 20, 30)
# Try different targets
build.target(x, type = "D")
```

check.path

Check if two genes belong to any common KEGG pathway.

Description

Takes as arguments two vectors of IDs and test whether they have a common ID.

Usage

```
check.path(p1, p2)
```

Arguments

p1	Vector of pathways that gene 1 belongs to.
p2	Vector of pathways that gene 2 belongs to.

Value

Return 0 if the two genes don't belong to a common pathway, return 1 otherwise. This function is used by [target.help](#).

Author(s)

Monika Jelizarow and Vincent Guillemot

See Also

[target.help](#)

Examples

```
g1 <- c("path1", "path2", "path3", "path4")
g2 <- c("path5", "path6", "path3", "path11")
g3 <- c("path10", "path5", "path12", "path13")
check.path(g1, g2) # 1
check.path(g1, g3) # 0
```

expl

Small example extracted from a microarray data set.

Description

The microarray data set is the study on the prostate cancer by Singh et al. The collection of the microarray is hgu95av2, and the gene groups are thus given by the information in the hgu95av2.db Bioconductor library (see Carlson et al.).

Usage

```
data("expl")
```

Format

The dataset is a list containing:

- a 102×100 matrix x of 100 genes randomly chosen from the data set of Singh et al.,
- a list ‘genegroups’ containing 100 vectors of KEGG pathway IDs (which each gene belongs to).

Source

- M. Carlson, S. Falcon, H. Pages, N. Li. hgu95av2.db: Affymetrix Human Genome U95 Set annotation data (chip hgu95av2). R package version 2.2.12.
- D. Singh, P. G. Febbo, K. Ross, D. G. Jackson, J. Manola, C. Ladd, P. Tamayo, A. A. Renshaw, A. V. D’Amico, J. P. Richie, E. S. Lander, M. Loda, P. W. Kantoff, T. R. Golub, W. R. Sellers, 2002. Gene expression correlates of clinical prostate cancer behavior. Cancer Cell, Department of Adult Oncology, Brigham and Women’s Hospital, Harvard Medical School, Boston, MA 02115, USA., 1, 203-209.

shrink.estim

Shrinkage estimator of the covariance matrix, given a data set and a covariance target.

Description

The estimator combines the sample covariance matrix S with a target matrix T :

$$\hat{\Sigma} = \lambda T + (1 - \lambda)S$$

The estimated intensity λ is constrained to $[0, 1]$. Values near zero keep more of the sample covariance; values near one give more weight to the target.

Usage

```
shrink.estim(x, target)
```

Arguments

`x` A $n \times p$ matrix (the data set) .
`target` A $p \times p$ matrix (the covariance target).

Value

A list with components `shrink.cov`, the estimated covariance matrix, and `lambda`, the estimated shrinkage intensity.

Author(s)

Monika Jelizarow and Vincent Guillelot

References

J. Schaefer and K. Strimmer, 2005. A shrinkage approach to large-scale covariance matrix estimation and implications for functional genomics. *Statist. Appl. Genet. Mol. Biol.* 4:32. Jelizarow, M., Guillelot, V., Tenenhaus, A., Strimmer, K. and Boulesteix, A.-L. (2010). Over-optimism in bioinformatics: an illustration. *Bioinformatics*.

See Also

Other main functions: [build.target\(\)](#)

Examples

```
# Simulate dataset
x <- matrix(rnorm(20*30),20,30)
# Try different targets
shrink.estim(x, target = build.target(x, type = "D"))
```

target.help

Transform a list of Pathway IDs into a binary matrix.

Description

This function transforms a list of p (one vector of pathway IDs per gene) groups into a binary matrix.

Usage

```
target.help(genes)
```

Arguments

`genes` List of p items. Each item is the vector of Pathway IDs a gene belongs to.

Value

A $p \times p$ binary matrix: the coefficient (i,j) is 1 if genes i and j belong to a common pathway and 0 otherwise.

Author(s)

Monika Jelizarow and Vincent Guillemot

See Also

[targetF](#), [targetG](#), [targetGpos](#), [targetGstar](#).

Examples

```
g1 <- c("path1", "path2", "path3", "path4")
g2 <- c("path5", "path6", "path3", "path11")
g3 <- c("path10", "path5", "path12", "path13")
target.help(list(g1, g2, g3))
```

targetCor

Significance-filtered knowledge-based target.

Description

The $p \times p$ target Cor is computed from the $n \times p$ data matrix. It is a modified version of target G. In particular, it tests the correlations (with a significance level of 0.05) and sets the non-significant correlations to zero before the mean correlation $\bar{r}$ is computed.

Usage

```
targetCor(x, genegroups)
```

Arguments

x	A $n \times p$ data matrix.
genegroups	A list of genes obtained using the database KEGG, where each entry itself is a list of pathway names this gene belongs to. If a gene does not belong to any gene functional group, the entry is NA.

Details

For retained linked pairs, the off-diagonal target is

$$T_{ij} = \bar{r}_{sig} \sqrt{s_{ii}s_{jj}}$$

where $\bar{r}_{sig}$ is the mean correlation among links whose correlation test p-value is below 0.05.

Value

A $p \times p$ matrix.

Author(s)

Monika Jelizarow and Vincent Guillemot

See Also

Other covariance targets: [targetD\(\)](#), [targetF\(\)](#), [targetG\(\)](#), [targetGpos\(\)](#), [targetGstar\(\)](#)

Examples

```
# A short example on a toy dataset
# require(SHIP)
data(expl)
target_matrix <- targetCor(expl$x, expl$genegroups)
which(target_matrix[upper.tri(target_matrix)] != 0) # not many non zero coefficients !
```

targetD	<i>Diagonal covariance target.</i>
---------	------------------------------------

Description

The target keeps the sample variances and sets all covariances to zero. If $S = (s_{ij})$ is the sample covariance matrix, the target is

$$T_{ij} = \begin{cases} s_{ii} & \text{if } i = j \\ 0 & \text{otherwise} \end{cases}$$

Usage

```
targetD(x, genegroups = NULL)
```

Arguments

x A $n \times p$ data matrix.
genegroups The genegroups are not used for this target.

Value

A $p \times p$ diagonal matrix.

Author(s)

Monika Jelizarow and Vincent Guillemot

See Also

Other covariance targets: [targetCor\(\)](#), [targetF\(\)](#), [targetG\(\)](#), [targetGpos\(\)](#), [targetGstar\(\)](#)

Examples

```
x <- matrix(rnorm(10*30),10,30)
target_matrix <- targetD(x, NULL)
```

targetF

*Constant-correlation covariance target.***Description**

The target keeps the sample variances and replaces every off-diagonal correlation by their average. If $S = (s_{ij})$ is the sample covariance matrix and $\bar{r}$ is the mean sample correlation, then

$$T_{ij} = \begin{cases} s_{ii} & \text{if } i = j \\ \bar{r} \sqrt{s_{ii}s_{jj}} & \text{otherwise} \end{cases}$$

Usage

```
targetF(x, genegroups = NULL)
```

Arguments

x A $n \times p$ data matrix.
genegroups The genegroups are not used for this target.

Value

A $p \times p$ matrix.

Author(s)

Monika Jelizarow and Vincent Guillemot

See Also

Other covariance targets: [targetCor\(\)](#), [targetD\(\)](#), [targetG\(\)](#), [targetGpos\(\)](#), [targetGstar\(\)](#)

Examples

```
# A short example on a toy dataset
# require(SHIP)
data(expl)
target_matrix <- targetF(expl$x, NULL)
which(target_matrix[upper.tri(target_matrix)] != 0) # many non zero coefficients !
```

targetG

*Knowledge-based constant-correlation target.***Description**

This target applies the constant-correlation model only to pairs of genes that share a functional group. Let $C = (c_{ij})$ be the binary group matrix and let $\bar{r}$ be the mean correlation among linked pairs. Then

$$t_{ij} = \begin{cases} s_{ii} & \text{if } i = j \\ \bar{r} \sqrt{s_{ii}s_{jj}} & \text{if } i \neq j, i \sim j \end{cases}$$

where $i \sim j$ means that genes i and j share a functional group. Unlinked off-diagonal entries are zero.

Usage

```
targetG(x, genegroups)
```

Arguments

x	A $n \times p$ data matrix.
genegroups	A list of genes obtained using the database KEGG, where each entry itself is a list of pathway names this genes belongs to. If a gene does not belong to any gene functional group, the entry is NA.

Value

A $p \times p$ matrix.

Author(s)

Monika Jelizarow and Vincent Guillemot

See Also

Other covariance targets: [targetCor\(\)](#), [targetD\(\)](#), [targetF\(\)](#), [targetGpos\(\)](#), [targetGstar\(\)](#)

Examples

```
# require(SHIP)
data(expl)
target_matrix <- targetG(expl$x, expl$genegroups)
which(target_matrix[upper.tri(target_matrix)] != 0) # not many non zero coefficients !
```

targetGpos

*Positive-correlation knowledge-based target.***Description**

This is a version of `targetG()` that estimates the common correlation from positive linked correlations only. Negative linked correlations contribute zero to the off-diagonal target.

$$T_{ij} = \bar{r}_+ \sqrt{s_{ii}s_{jj}}$$

for linked pairs with positive correlation, where $\bar{r}_+$ is their mean positive correlation.

Usage

```
targetGpos(x, genegroups)
```

Arguments

<code>x</code>	A $n \times p$ data matrix.
<code>genegroups</code>	A list of genes obtained using the database KEGG, where each entry itself is a list of pathway names this genes belongs to. If a gene does not belong to any gene functional group, the entry is NA.

Value

A $p \times p$ matrix.

Author(s)

Monika Jelizarow and Vincent Guillemot

See Also

Other covariance targets: `targetCor()`, `targetD()`, `targetF()`, `targetG()`, `targetGstar()`

Examples

```
# require(SHIP)
data(expl)
target_matrix <- targetGpos(expl$x, expl$genegroups)
which(target_matrix[upper.tri(target_matrix)] != 0) # not many non zero coefficients !
```

targetGstar	<i>Signed-correlation knowledge-based target.</i>
-------------	---

Description

The $p \times p$ target Gstar is computed from the $n \times p$ data matrix. It is a modified version of target G. In particular, it involves two parameters for the correlation (a positive and a negative one) instead of the single parameter $\bar{r}$ in order to account for negatively correlated genes within the same pathway.

Usage

```
targetGstar(x, genegroups)
```

Arguments

x	A $n \times p$ data matrix.
genegroups	A list of genes obtained using the database KEGG, where each entry itself is a list of pathway names this gene belongs to. If a gene does not belong to any gene functional group, the entry is NA.

Details

The target uses separate mean correlations for positive and negative linked pairs:

$$T_{ij} = \begin{cases} \bar{r}_+ \sqrt{s_{ii}s_{jj}} & \text{if } r_{ij} > 0 \\ \bar{r}_- \sqrt{s_{ii}s_{jj}} & \text{if } r_{ij} < 0 \end{cases}$$

Here $\bar{r}_+$ and $\bar{r}_-$ are the mean positive and negative linked correlations.

Value

A $p \times p$ matrix.

Author(s)

Monika Jelizarow and Vincent Guillemot

See Also

Other covariance targets: [targetCor\(\)](#), [targetD\(\)](#), [targetF\(\)](#), [targetG\(\)](#), [targetGpos\(\)](#)

Examples

```
# require(SHIP)
data(expl)
target_matrix <- targetGstar(expl$x, expl$genegroups)
which(target_matrix[upper.tri(target_matrix)] != 0) # not many non zero coefficients !
```

Index

* covariance targets

- targetCor, [7](#)
- targetD, [8](#)
- targetF, [9](#)
- targetG, [10](#)
- targetGpos, [11](#)
- targetGstar, [12](#)

* main functions

- build.target, [3](#)
- shrink.estim, [5](#)

* methods

- build.target, [3](#)
- shrink.estim, [5](#)
- target.help, [6](#)
- targetCor, [7](#)
- targetD, [8](#)
- targetF, [9](#)
- targetG, [10](#)
- targetGpos, [11](#)
- targetGstar, [12](#)

* multivariate

- shrink.estim, [5](#)
- targetCor, [7](#)
- targetD, [8](#)
- targetF, [9](#)
- targetG, [10](#)
- targetGpos, [11](#)
- targetGstar, [12](#)

* package

- SHIP-package, [2](#)

build.target, [2](#), [3](#)

build.target(), [6](#)

check.path, [4](#)

expl, [5](#)

SHIP (SHIP-package), [2](#)

SHIP-package, [2](#)

shrink.estim, [2](#), [5](#)

shrink.estim(), [3](#)

target.help, [4](#), [6](#)

targetCor, [3](#), [7](#)

targetCor(), [9–12](#)

targetD, [3](#), [8](#)

targetD(), [8–12](#)

targetF, [3](#), [7](#), [9](#)

targetF(), [8–12](#)

targetG, [3](#), [7](#), [10](#)

targetG(), [8](#), [9](#), [11](#), [12](#)

targetGpos, [3](#), [7](#), [11](#)

targetGpos(), [8–10](#), [12](#)

targetGstar, [3](#), [7](#), [12](#)

targetGstar(), [8–11](#)