HU-MATHS-IN miniproject proposal Dimension Reduction of High Frequency and High Dimensional Data in Time and Space

Marianna Bolla, Máté Baranyi, Fatma Abdelkhalek
Department of Stochastics
Institute of Mathematics, BME
marib@math.bme.hu

November 22, 2019.

Background: Time Series and Random Fields

- Reduction of dimensionality in high frequency time series; for example, finding the underlying signals detected by many sensors: M. Bolla, Factor Analysis, Dynamic. Wiley StatsRef: Statistics Reference Online (2017) and book in preparation with T. Szabados.
- Graphical models and prediction along regression graphs (to develop artificial intelligence, e.g., for medical diagnosis): M. Bola, F. Abdelkhalek, M. Baranyi, Graphical models, regression graphs, and recursive linear regression in a unified way, Acta Univ. Szeged (2019) (related SEM, PLS papers and Python program in preparation).
- Image segmentation with spectral clustering tools (tumor-segmentation, computer aided surgery): M. Bolla, Spectral Clustering and Biclustering, Wiley (2013).

Background: Time Series and Random Fields

- Reduction of dimensionality in high frequency time series; for example, finding the underlying signals detected by many sensors: M. Bolla, Factor Analysis, Dynamic. Wiley StatsRef: Statistics Reference Online (2017) and book in preparation with T. Szabados.
- Graphical models and prediction along regression graphs (to develop artificial intelligence, e.g., for medical diagnosis): M. Bola, F. Abdelkhalek, M. Baranyi, Graphical models, regression graphs, and recursive linear regression in a unified way, Acta Univ. Szeged (2019) (related SEM, PLS papers and Python program in preparation).
- Image segmentation with spectral clustering tools (tumor-segmentation, computer aided surgery): M. Bolla, Spectral Clustering and Biclustering, Wiley (2013).

Background: Time Series and Random Fields

- Reduction of dimensionality in high frequency time series; for example, finding the underlying signals detected by many sensors: M. Bolla, Factor Analysis, Dynamic. Wiley StatsRef: Statistics Reference Online (2017) and book in preparation with T. Szabados.
- Graphical models and prediction along regression graphs (to develop artificial intelligence, e.g., for medical diagnosis): M. Bola, F. Abdelkhalek, M. Baranyi, Graphical models, regression graphs, and recursive linear regression in a unified way, Acta Univ. Szeged (2019) (related SEM, PLS papers and Python program in preparation).
- Image segmentation with spectral clustering tools (tumor-segmentation, computer aided surgery): M. Bolla, Spectral Clustering and Biclustering, Wiley (2013).

Singular (deterministic) time series: Cramér, Wold, Wiener, Kolmogorov

Singular

Type (0)
$\mathbf{F} \perp \lambda$

Type (1)
$\exists \mathbf{f} = \mathbf{F}'$
$\mathbf{f}(A) = 0, \lambda(A) > 0$

Type (2)
$\exists \mathbf{f} = \mathbf{F}', \;\; \mathbf{f} > 0$
$\int_{-\pi}^{\pi} \log |\mathbf{f}(\omega)| \, d\omega = -\infty$

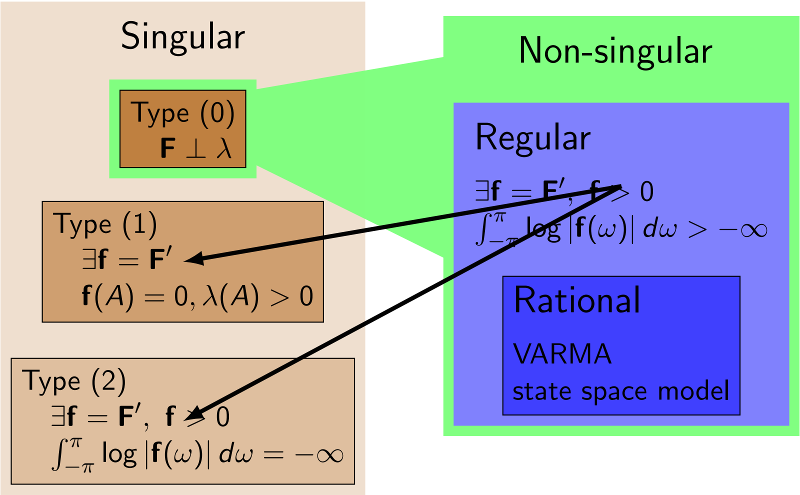
Non-singular

Regular
$\exists \mathbf{f} = \mathbf{F}', \;\; \mathbf{f} > 0$
$\int_{-\pi}^{\pi} \log |\mathbf{f}(\omega)| \, d\omega > -\infty$

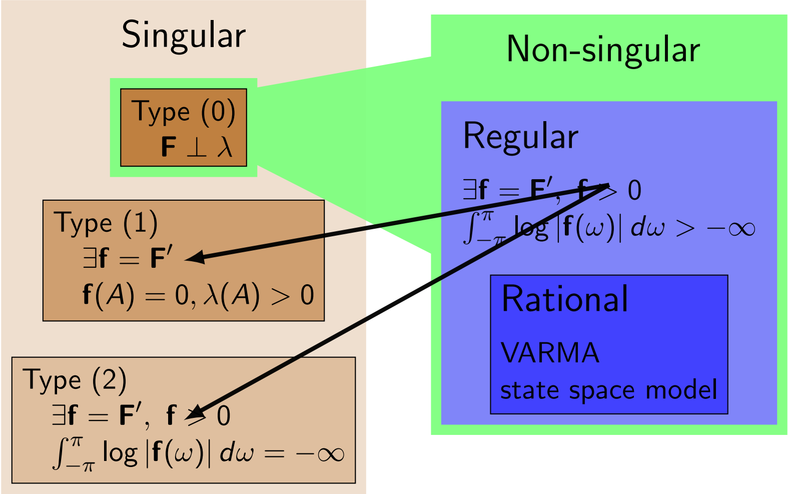
Rational
VARMA
state space model

e.g., $\mathbf{X}_t = \sum_{j=-\infty}^{\infty} \mathbf{B}_j \varepsilon_{t-j}$
$rank(\mathbf{f}) = r$ a.s.

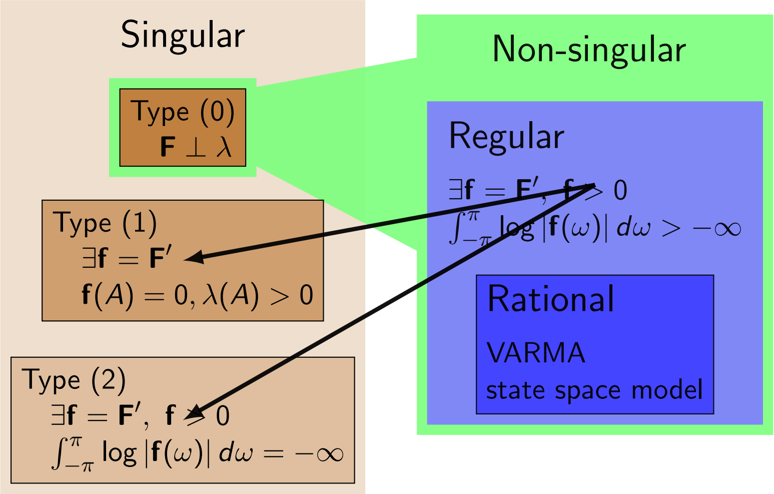
Singular-Regular



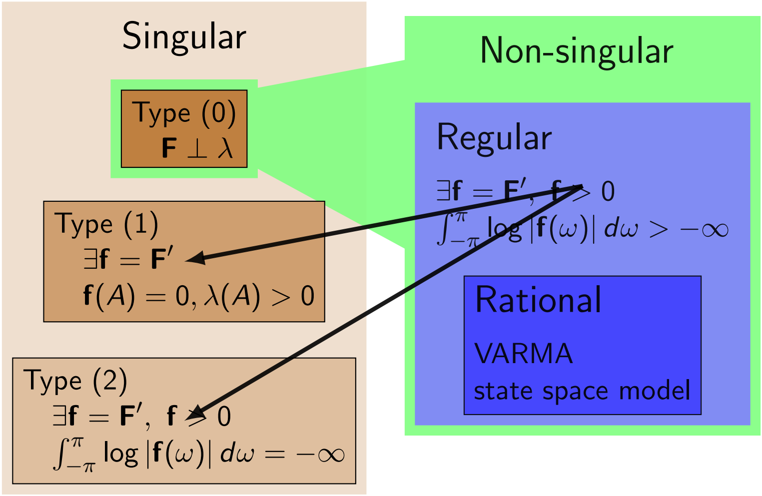
Singular-Regular



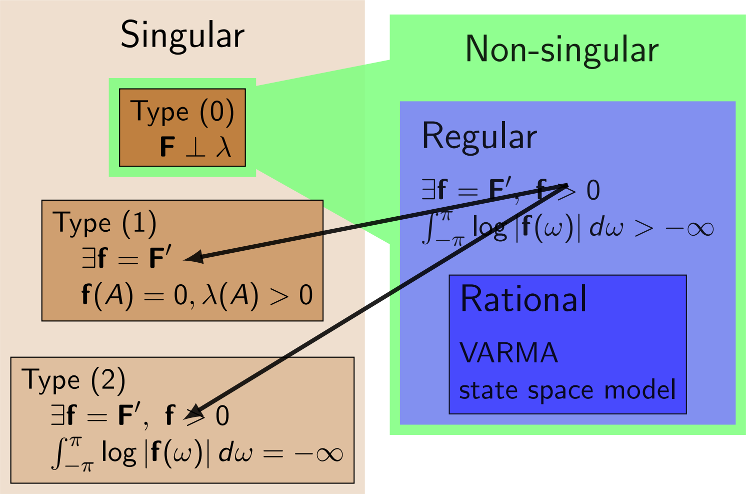
Singular-Regular



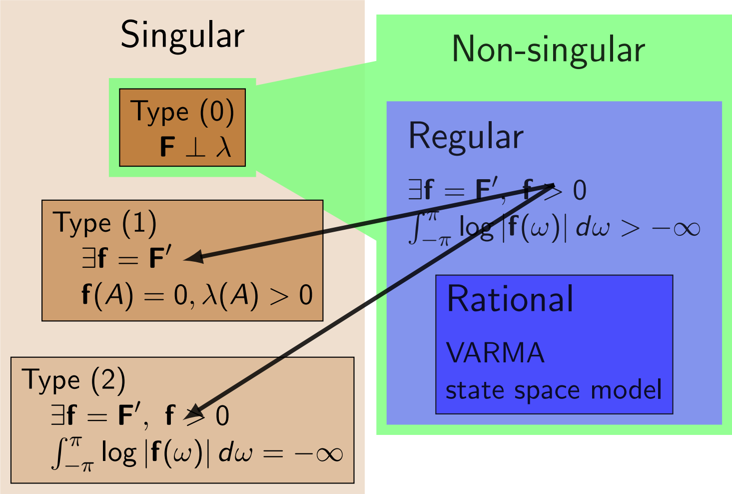
Singular-Regular



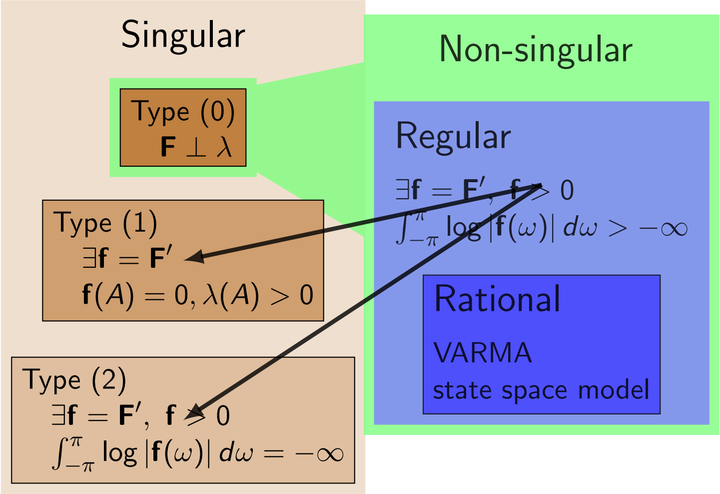
Singular-Regular



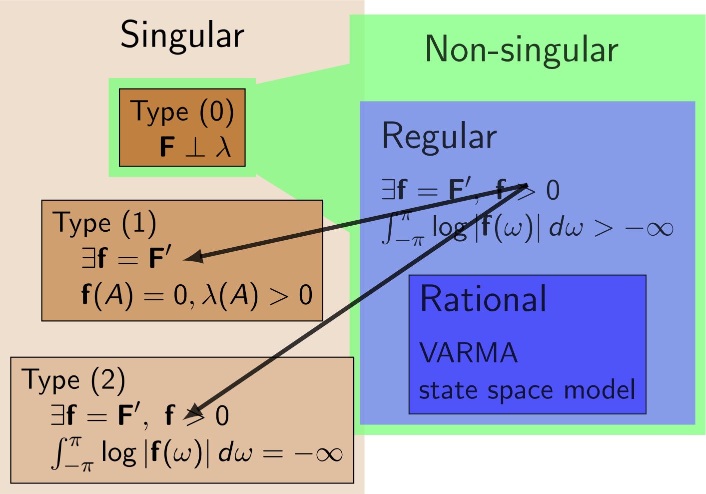
Singular-Regular



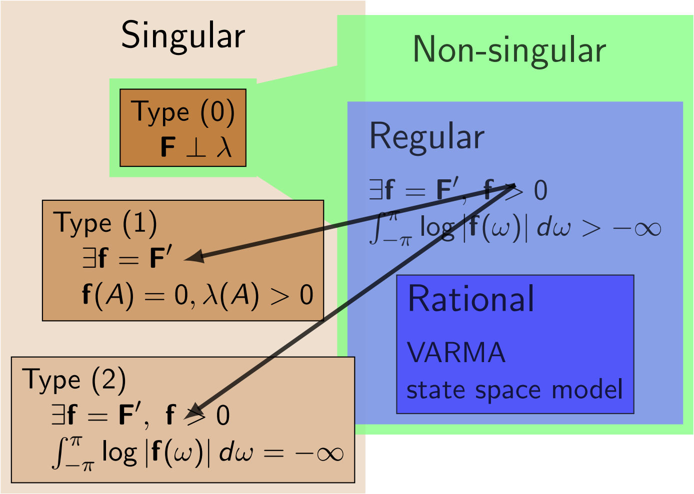
Singular-Regular



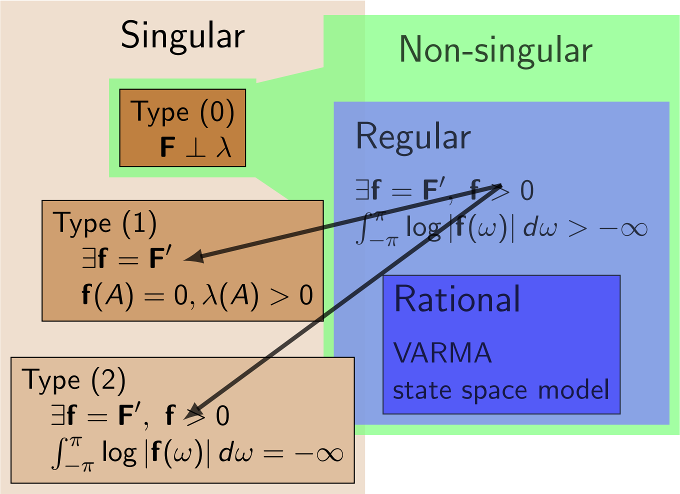
Singular-Regular



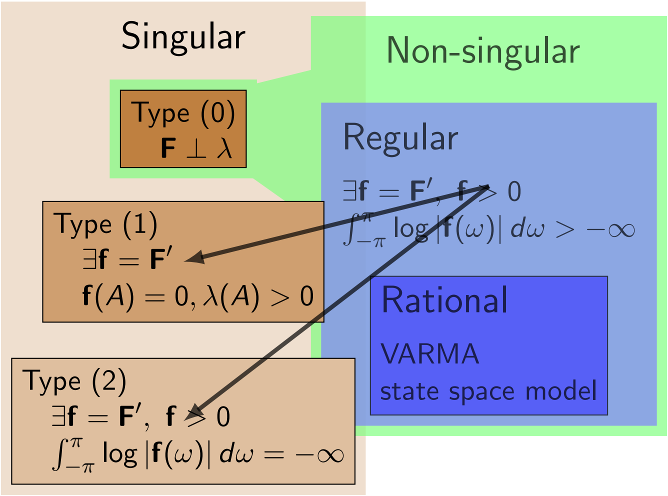
Singular-Regular



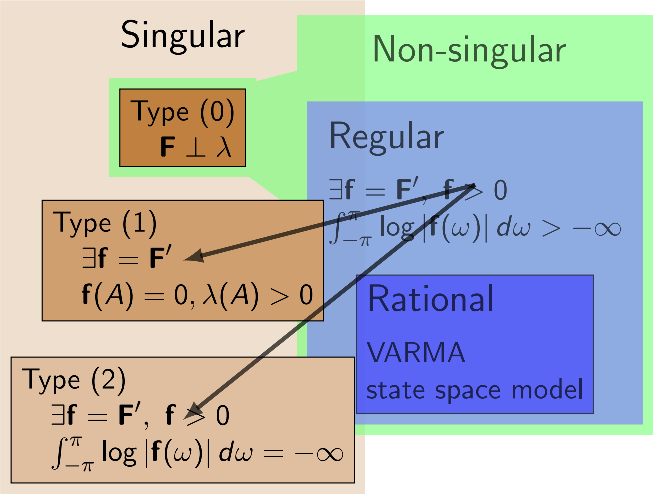
Singular-Regular



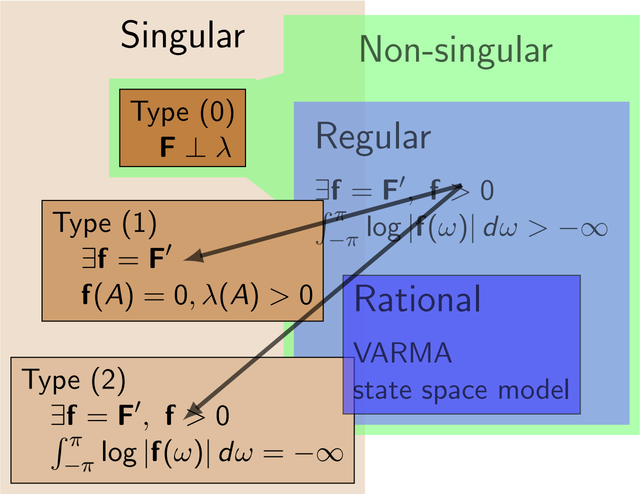
Singular-Regular



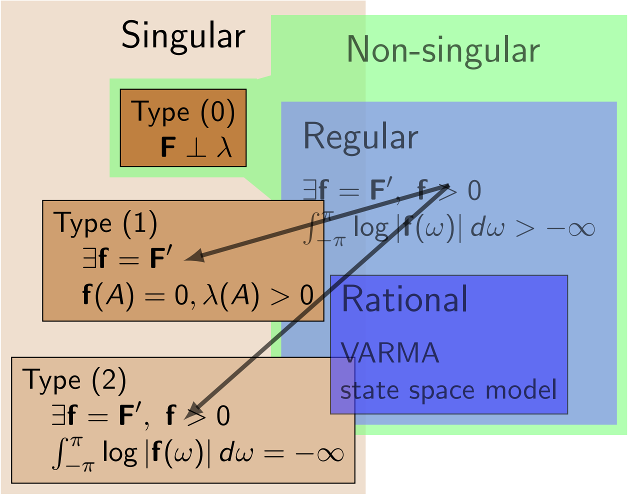
Singular-Regular



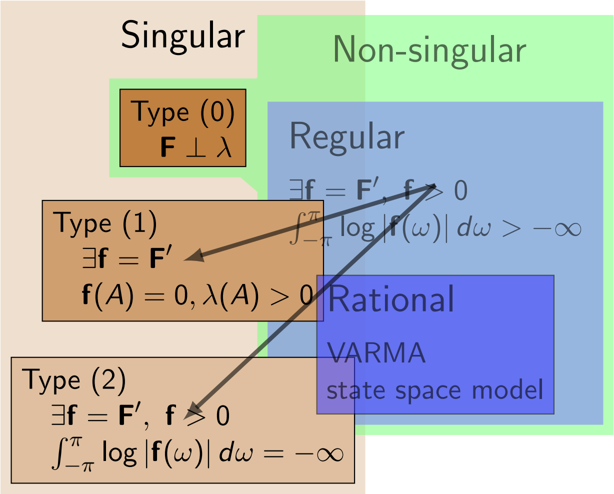
Singular-Regular



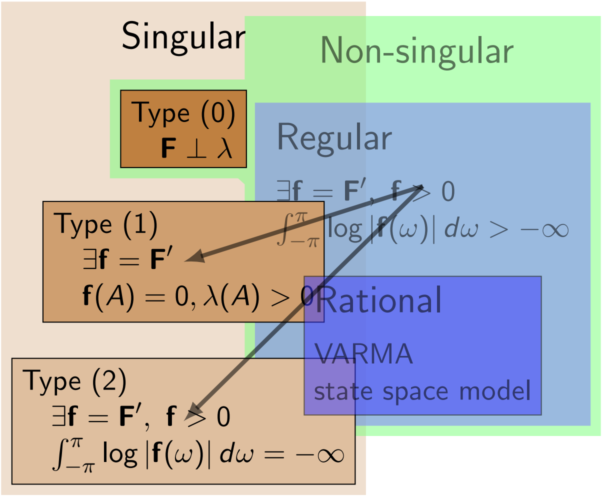
Singular-Regular



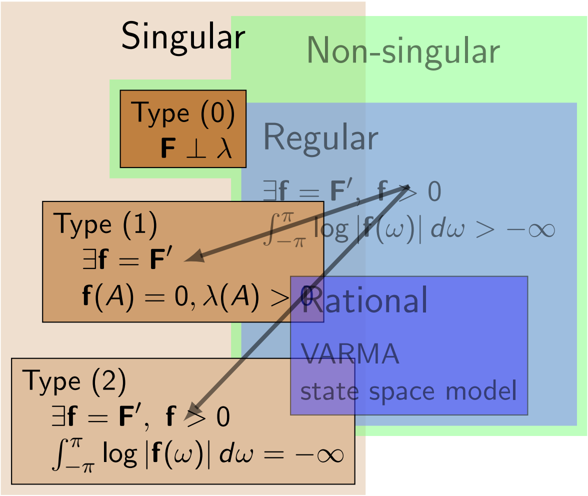
Singular-Regular



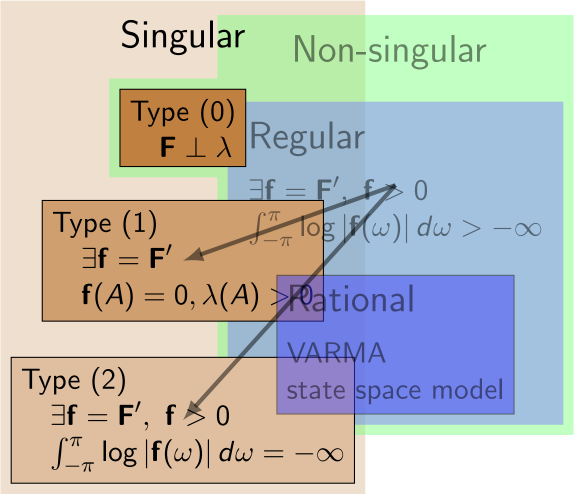
Singular-Regular



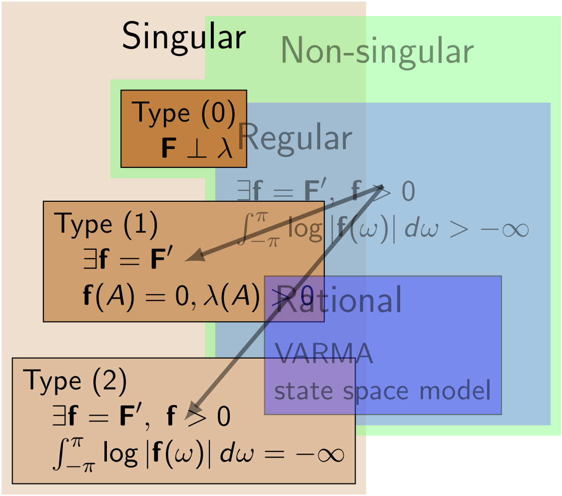
Singular-Regular



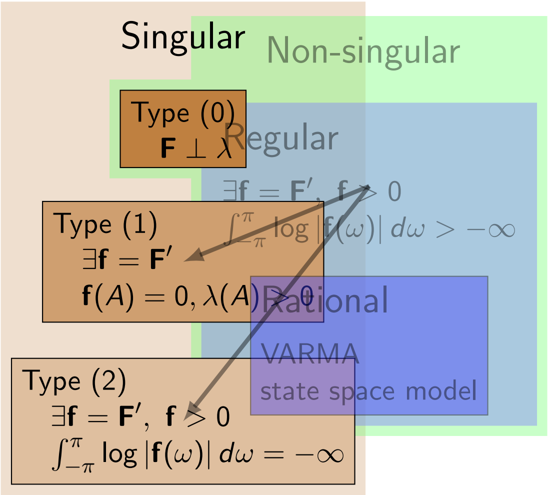
Singular-Regular



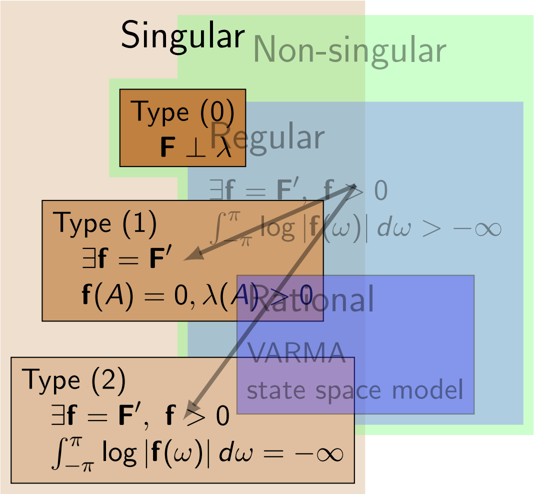
Singular-Regular



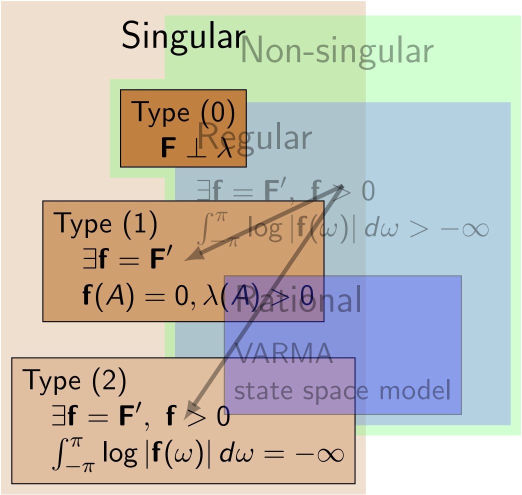
Singular-Regular



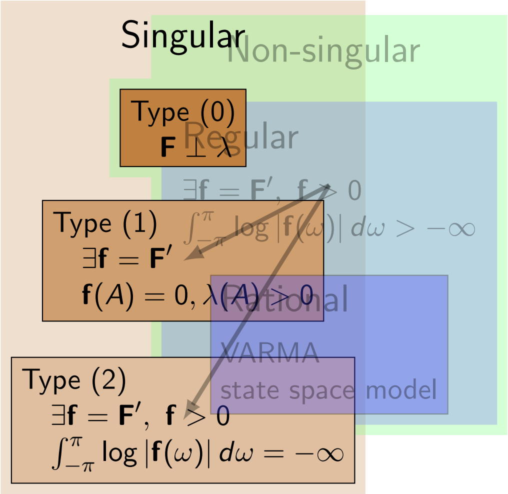
Singular-Regular



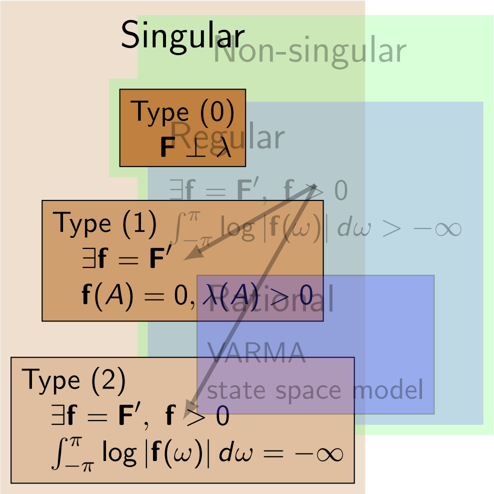
Singular-Regular



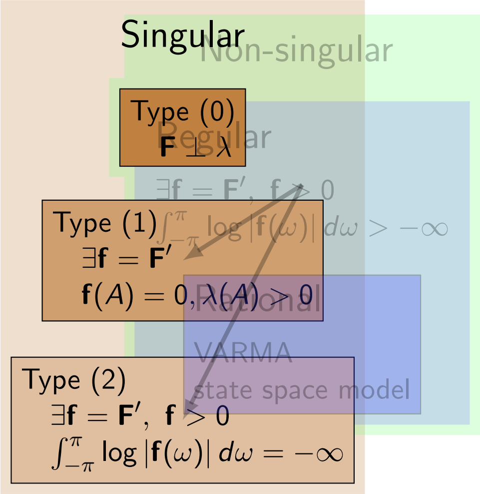
Singular-Regular



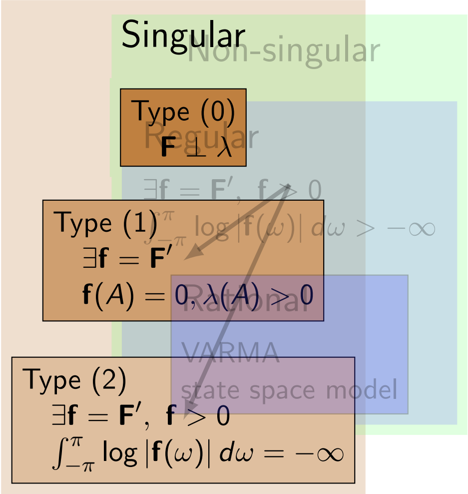
Singular-Regular



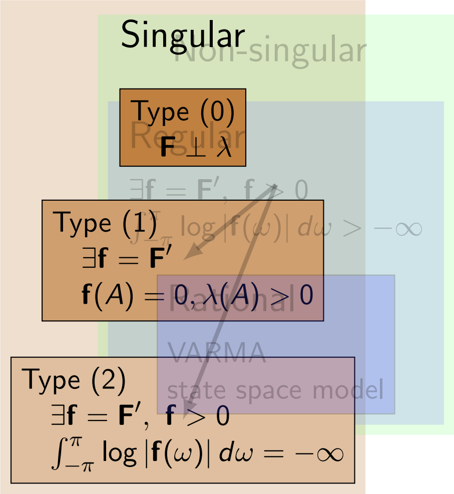
Singular-Regular



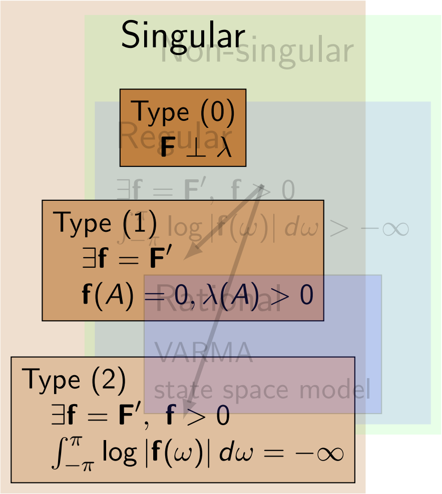
Singular-Regular



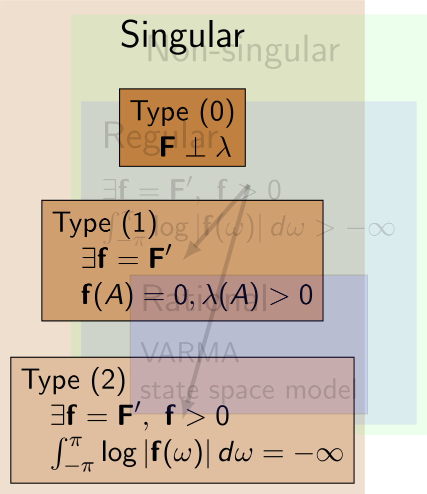
Singular-Regular



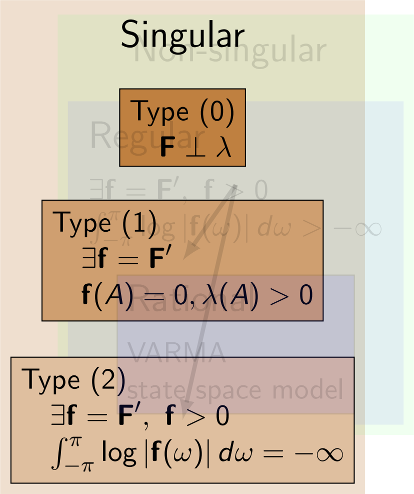
Singular-Regular



Singular-Regular



Singular-Regular



Singular-Regular

Singular

Type (0)
$\mathbf{F} \perp \lambda$

Type (1)
$\exists \mathbf{f} = \mathbf{F}'$
$\mathbf{f}(A) = 0, \lambda(A) > 0$

Type (2)
$\exists \mathbf{f} = \mathbf{F}', \ \mathbf{f} > 0$
$\int_{-\pi}^{\pi} \log |\mathbf{f}(\omega)| \, d\omega = -\infty$

...back to the original picture

Non-singular time series: Wold decomposition

H. Wold's Theorem:

$$\mathbf{X}_t = \sum_{j=0}^{\infty} \mathbf{B}_j \boldsymbol{\varepsilon}_{t-j} + \text{Type (0) singular},$$

where $\mathbf{B}_j$: $n \times r$ matrix (with square-summable entries), and $\{\varepsilon_t\}$ r-dimensional *white noise* with covariance matrix $\boldsymbol{\Sigma}$.
If $\{\mathbf{X}_t\}$ is **regular** with $n \times n$ symmetric, positive semidefinite *spectral density matrix* $\mathbf{f}$, $\mathtt{rang}\, \mathbf{f}(\omega) = r \leq n$ (a.s.) and

$$\mathbf{f}(z) \Longleftrightarrow (\mathbf{B}(z), \boldsymbol{\Sigma}) : \quad \mathbf{f}(z) = \frac{1}{2\pi} \mathbf{B}(z) \boldsymbol{\Sigma} \mathbf{B}^*(z),$$

where

$$\mathbf{B}(z) = \sum_{j=0}^{\infty} \mathbf{B}_j z^j, \quad |z| \leq 1$$

transfer-function.

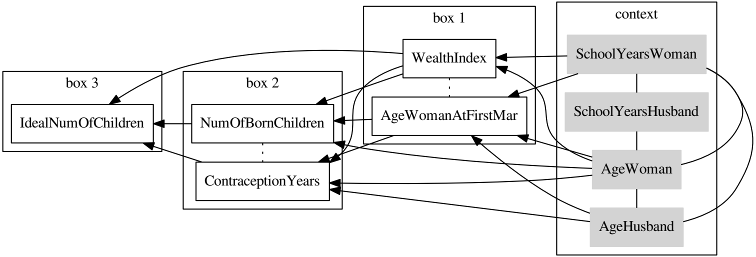
General Dynamic Factor Model (GDFM)

If $r = n$: $\mathbf{X}_t = \widehat{\mathbf{X}}_{t|t-1} + \varepsilon_t$, then
$\widehat{\mathbf{X}}_{t|t-1}$: *one-step ahead prediction* (based on infinite past) and ε_t: **innovation**.

If $r < n$: only the innovation subspace is unique, and
dim (innovation subspace) =rank $(\mathbf{f}) = r$.
GDFM (Forni, Lippi, Deistler): $n, t \to \infty$ (from a starting time).
Number of dynamic factors: essential rank of $\mathbf{f}$ (number of structural eigenvalues in the frequency domain).
Dynamic factors: standardized innovations.
After eliminating the idiosyncratic part, we predict in the time domain (based on finite past, block Cholesky decomposition, then static factor analysis in the innovation subspaces).
Application: finding underlying signals detected by many sensors.
Telecommunication systems (NOKIA), EEG, biomedical sensors.

Graphical models

Chain graph model, EDHS 2014 data



Application: artificial intelligence, medical diagnosis, health, and demography planning in developing countries.

Non-parametric regression: iteration along a DAG

$X_1, \dots, X_d$: random variables (vertices), $\mathbf{X}_{>k} := (X_{k+1}, \dots, X_d)$: context variables.

$$P_j(\mathbf{x}_{>j}) := \mathbb{E}(X_j | \mathbf{X}_{>j} = \mathbf{x}_{>j}), \quad j = k, k-1, \dots, 1;$$

$$S_j^{(n)}(\mathbf{x}_{>j}) := \sum_{i=1}^{n} x_j^{(i)} W^{(n,i)}(\mathbf{x}_{>j}), \quad j = k, k-1, \dots, 1$$

smoother (local averaging), e.g., with product kernels.
Consistency Theorem: If

$$\mathbb{E}[S_j^{(n)}(\mathbf{X}_{>j}) - P_j(\mathbf{X}_{>j})]^2 \to 0, \quad n \to \infty, \quad j = k, k-1, \dots, 1$$

then

$$\mathbb{E}[S_j^{(n)}(S_{j+1}^{(n)}(S_{j+2}^{(n)}(\dots(S_k^{(n)}(\mathbf{X}_{>k}), \dots), \mathbf{X}_{>k}), \mathbf{X}_{>k}), \mathbf{X}_{>k})-$$
$$- P_j(P_{j+1}(P_{j+2}(\dots(P_k(\mathbf{X}_{>k}), \dots), \mathbf{X}_{>k}), \mathbf{X}_{>k}), \mathbf{X}_{>k})]^2 \to 0, \quad n \to \infty$$

$\forall j = k, k-1, \dots, 1.$

Using Kálmán's filtering in SEM

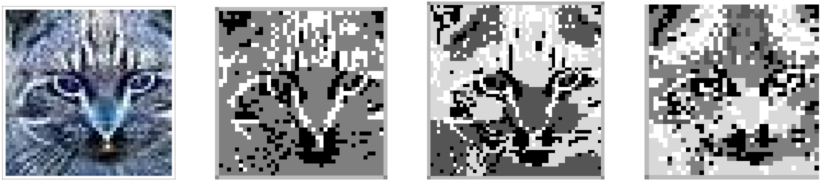
Structural Equation Model (SEM) as a linear dynamical system:

$$\begin{aligned}\mathbf{B}\boldsymbol{\eta}_t &= \mathbf{A}\boldsymbol{\xi}_t + \boldsymbol{\zeta}_t\\ \mathbf{X}_t &= \mathbf{C}\boldsymbol{\xi}_t + \boldsymbol{\varepsilon}_t\\ \mathbf{Y}_t &= \mathbf{D}\boldsymbol{\eta}_t + \boldsymbol{\delta}_t,\end{aligned}$$

where $\boldsymbol{\eta}_t$ is m- and $\boldsymbol{\xi}_t$ is n-dimensional latent vector; **B** and **A** are $m \times m$ and $m \times n$ coefficient matrices; $\boldsymbol{\zeta}_t$ is a random vector of residuals. It is uncorrelated with $\boldsymbol{\xi}_t$, and **B** is nonsingular. In the recursive models, **B** is upper triangular.
$\mathbf{X}_t$ and $\mathbf{Y}_t$ are q- and p-dimensional observable variables, $q \geq n$ and $p \geq m$. **C** is $q \times n$, **D** is $p \times m$; $\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D}$ are estimated from a learning sample (LISLER or block Cholesky decomposition).
$\boldsymbol{\zeta}_t$ is an orthogonal process, $\mathbb{E}\boldsymbol{\xi}_s^T\boldsymbol{\zeta}_t = \mathbf{0}$ for $s \leq t$; $\boldsymbol{\varepsilon}_t$ is independent of $\boldsymbol{\xi}_t$, $\boldsymbol{\delta}_t$ is independent of $\boldsymbol{\eta}_t$, they are also independent of each other and of $\boldsymbol{\zeta}_t$.
Prediction: $\mathbf{X}_t \to \widehat{\boldsymbol{\xi}}_t \to \widehat{\boldsymbol{\eta}}_t \to \widehat{\mathbf{Y}}_t$.

Spectral clustering

Original picture and pixels colored with 3, 4, 5 colors (clusters)



(48 × 48 pixels)

Structural eigenvalues of the normalized modularity matrix:

0.137259, 0.014255, 0.000925,

-0.0006707, $-0.0006706, \ldots$

Reproducing Kernel Hilbert Spaces, product Gaussian kernels

image segmentation

Normalized contingency table (e.g., microarray), directed graphs:

$\mathbf{C}_D = \mathbf{D}_{row}^{-1/2} \mathbf{C} \mathbf{D}_{col}^{-1/2}$ (correspondence analysis)

Application: image segmentation, microarrays (forensic data); e.g., tumor segmentation, computer aided surgery.

Thank you for your attention.