Solving singularity issues in the estimation of econometric models

Michel Bierlaire   Michaël Thémans

Transport and Mobility Laboratory
École Polytechnique Fédérale de Lausanne

Outline

- Estimation of advanced discrete choice models
  - Singularity issues
  - Sources and drawbacks of singularity
- Structural singularity
- Singularity at the solution
  - Adaptation of optimization algorithms
  - Numerical results
- Conclusions and perspectives

Recent advances in DCM

- More Logit-like models in the GEV family
- Mixtures of Logit models
- Mixtures of GEV models
- Discrete mixtures of GEV models

$\Rightarrow$ Estimating those models (via maximum log-likelihood) becomes more and more problematic

TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Difficulties in the estimation

- Objective function becomes highly nonlinear and non concave
- Computational cost of evaluating the objective function and its derivatives can be significantly high
- Constraints imposed on parameters to overcome overspecification or to obtain meaningful values of parameters
- Identification issues
  $\Rightarrow$ singularity in the log-likelihood function
- Estimation requires specific optimization algorithms

Sources of singularity

- Structural
  - Identication issue for the parameters of the error terms
  - e.g. std. errors in heteroscedastic EC models
- Contextual
  - Identification issue for the parameters in the deterministic part
  - irrelevant attributes
  - data limitations

Types of singularity

We want to solve the maximum log-likelihood estimation problem

$$\max_{\beta \in \mathbb{R}^n} \bar{\mathcal{L}}(\beta)$$

- Structural singularity

  $\Rightarrow \quad \nabla^2 \bar{\mathcal{L}}(\beta)$ is singular $\forall \beta \in \mathbb{R}^n$

- Singularity at solution

  $\Rightarrow \quad \nabla^2 \bar{\mathcal{L}}(\beta)$ is singular at $\beta^* \in \mathbb{R}^n$ where β^* is the vector of estimated values of parameters

TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Impacts of singularity

- The convergence of the estimation process can be much slower
- The estimation time can be huge!
- The variance-covariance matrix cannot be obtained
  - Not possible to assess the quality of the calibrated model
- Numerical difficulties can arise in the estimation

$\Rightarrow$ Develop robust optimization algorithms able to solve efficiently singular problems

TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Dealing with singularity issues

- Some of the algorithms implemented in the optimization package BIOGEME are already robust to face structural singularity issues
- Focus our work on singularities issues at solution
  - Even if only $\nabla^2 \bar{\mathcal{L}}(\beta^*)$ is singular, the convergence of the overall sequences of iterates $\{\beta_k\}_{k \in \mathbb{N}}$ is significantly deteriorated
  - The associated optimization problem is ill-conditioned

$\Rightarrow$ Adapt existing optimization algorithms

Adaptation of optimization algorithms

- Two main issues have to be addressed
  - Detection of a singularity during the course of the optimization algorithm
  - Strategies to fix the singularity by adding constraints
- No details about optimization algorithms themselves
  - Description of the identification process
  - Fixing the singularity by adding constraints to the problem

Problem formulation

Defining $x = \beta$ and $f(x) = -\bar{\mathcal{L}}(\beta)$

$$\left\{ \begin{array}{l} \max \bar{\mathcal{L}}(\beta) \\ \beta \in \mathbb{R}^n \end{array} \right. \iff \left\{ \begin{array}{l} \min f(x) \\ x \in \mathbb{R}^n \end{array} \right.$$

- $\bar{\mathcal{L}}$ is the log-likelihood function
- β is the vector of parameters to be estimated
- $f : \mathbb{R}^n \to \mathbb{R}$ twice continuously differentiable

Problem formulation

Defining $x = \beta$ and $f(x) = -\bar{\mathcal{L}}(\beta)$

$$\begin{cases} \max \bar{\mathcal{L}}(\beta) \\ \beta \in \mathbb{R}^n \end{cases} \iff \begin{cases} \min f(x) \\ x \in \mathbb{R}^n \end{cases}$$

- $\bar{\mathcal{L}}$ is the log-likelihood function
- β is the vector of parameters to be estimated
- $f : \mathbb{R}^n \to \mathbb{R}$ twice continuously differentiable
- We assume that the problem is singular at x^*, that is
  - $\nabla^2 f(x^*)$ is singular
  - x^* is a local minimum of the problem

TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Ideas of the algorithm

- Identification of the singularity
  - Issue 1: must analyze $\nabla^2 f(x^*)$, without knowing x^*
  - Issue 2: eigenstructure analysis is time consuming
- Fixing the singularity
  - Issue 1: how to correct the singularity?
  - Issue 2: how to adapt existing algorithms?

Ideas of the algorithm

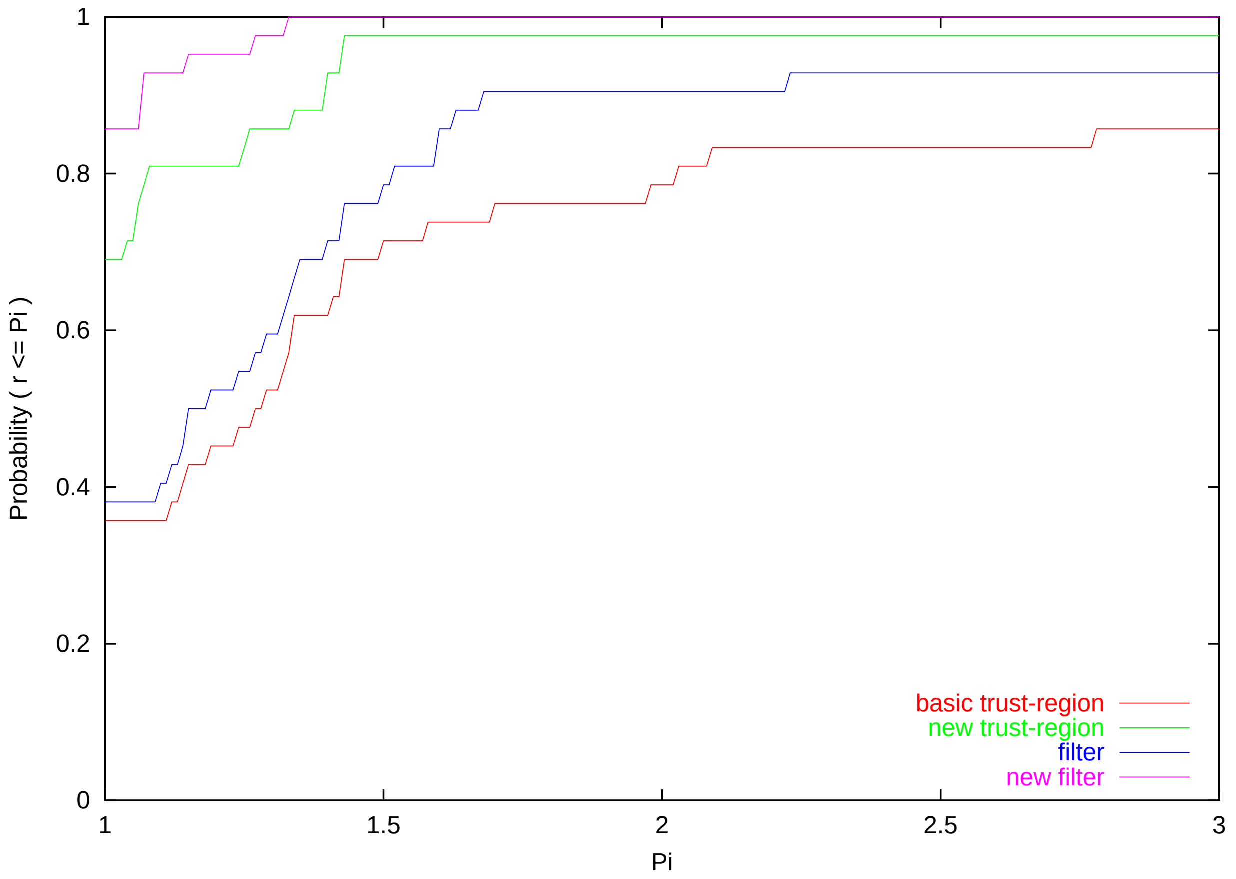
- Identification of the singularity
  - Solution 1: analyze $\nabla^2 f(x_k)$ instead of $\nabla^2 f(x^*)$
  - Solution 2: generalized inverse iteration
- Fixing the singularity
  - Solution 1: add curvature
  - Solution 2: trust-region framework is well designed for easy adaptations

Numerical tests

- Around 75 test problems
  - All containing a singularity at solution
  - Dimension between 2 and 40
- 4 optimization algorithms
  - Basic trust-region algorithm (in BIOGEME)
  - Trust-region algorithm designed to handle singularity
  - Filter-trust-region algorithm
  - Filter-trust-region algorithm designed to handle singularity
- 2 measures of performance
  - Number of iterations
  - CPU time

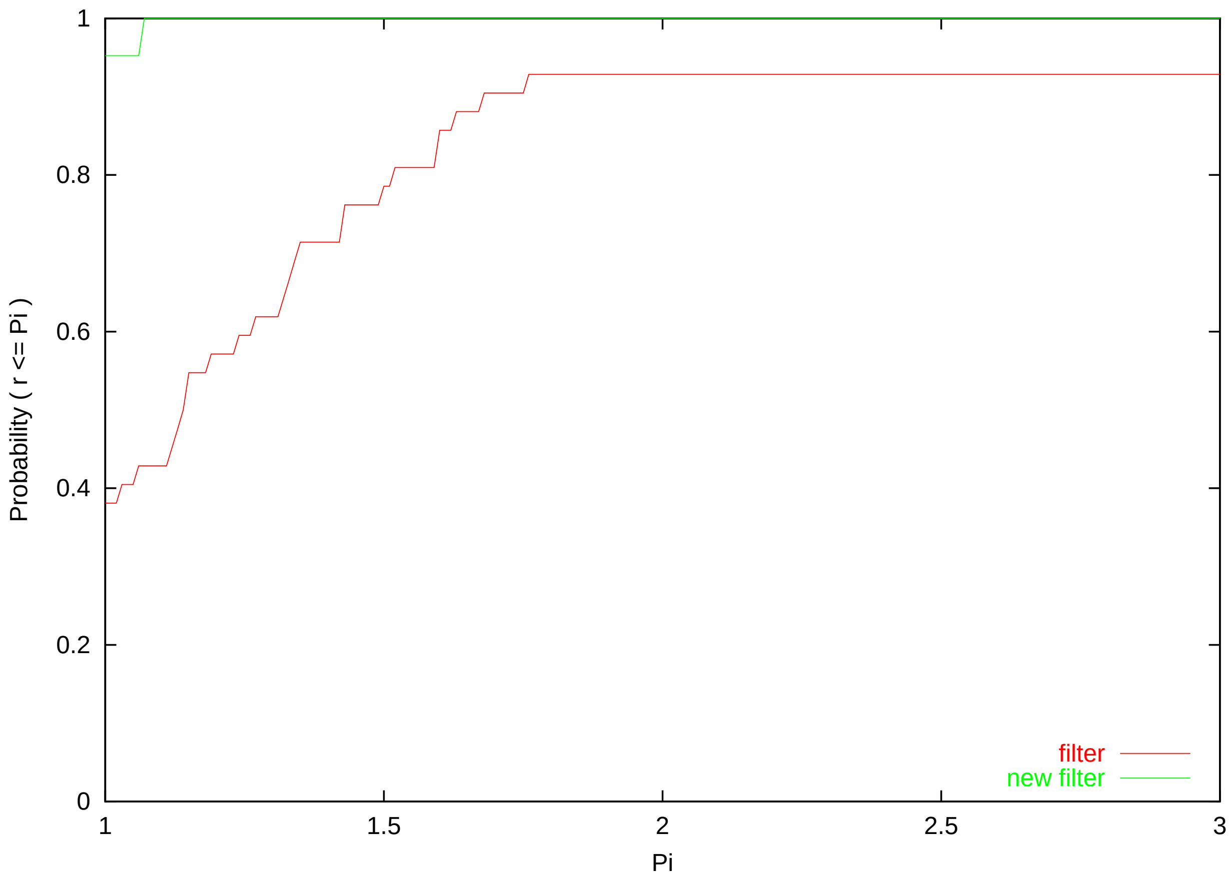
Performance profile 1

- All algorithms
- Number of iterations



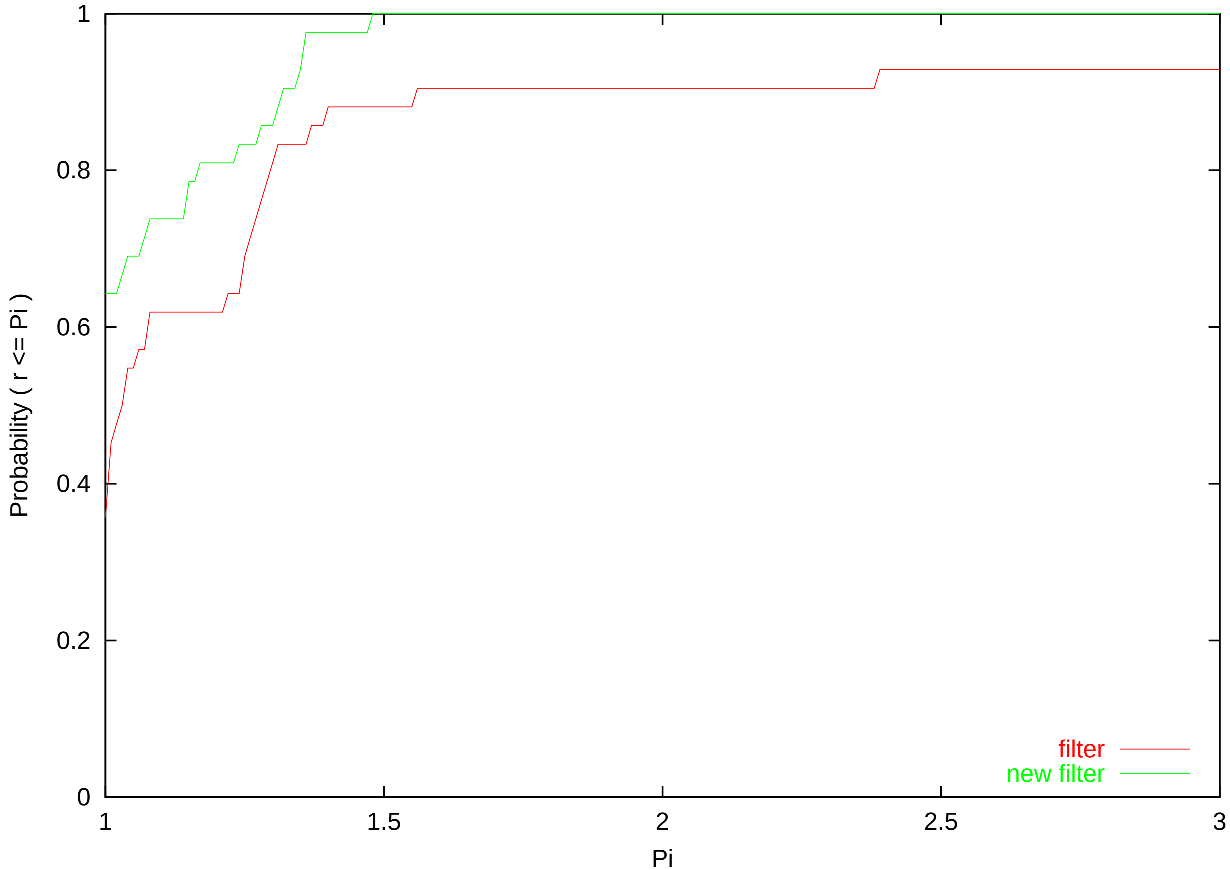
Performance profile 2

- Two filter-trust-region variants
- Number of iterations



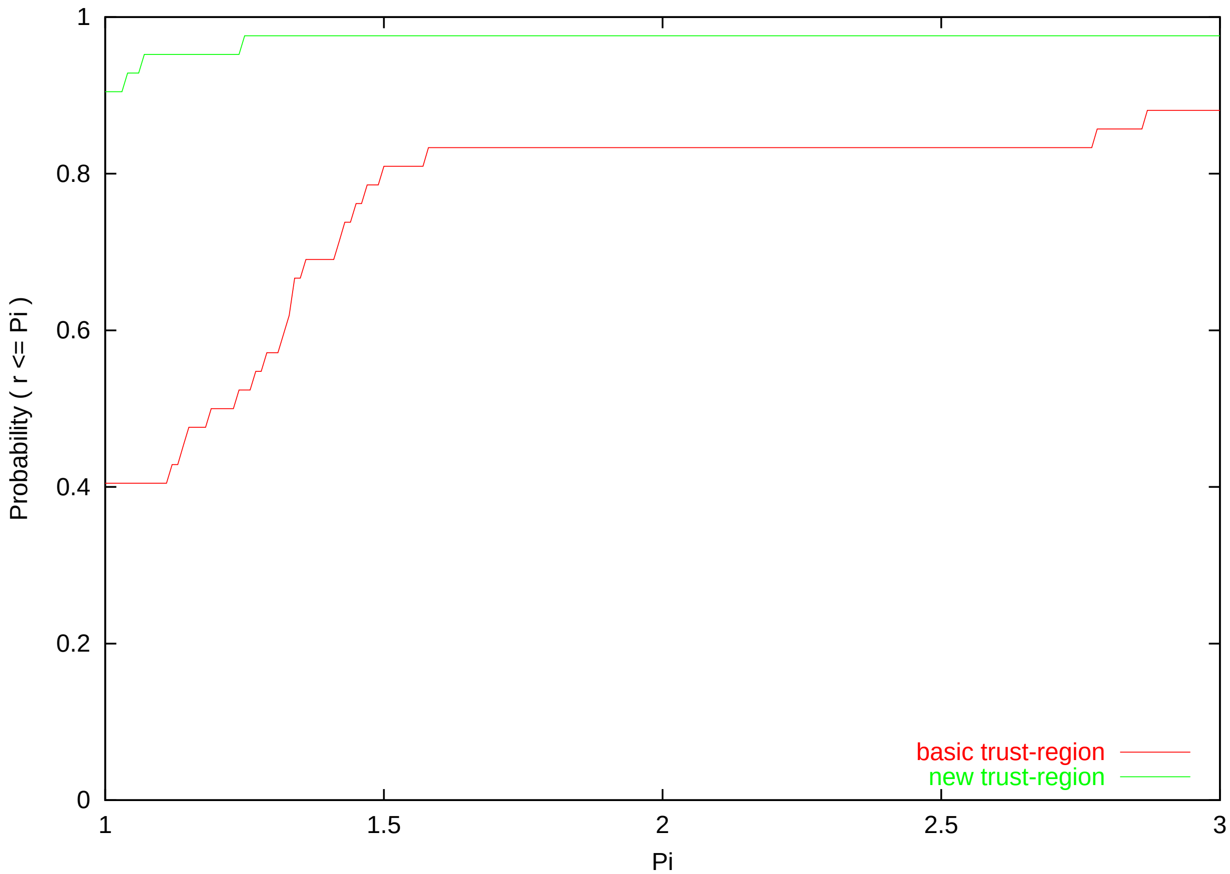
Performance profile 3

- Two filter-trust-region variants
- CPU time



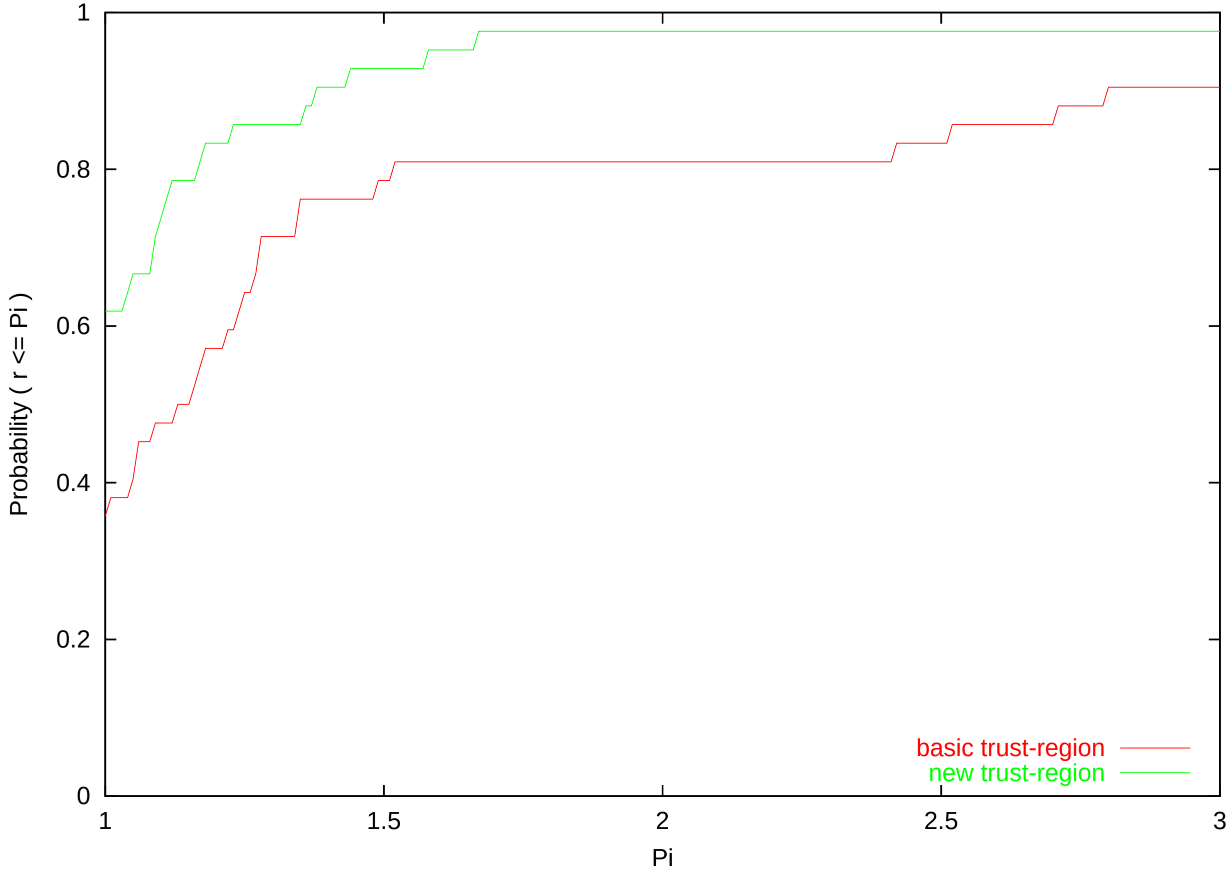
Performance profile 4

- Two trust-region variants
- Number of iterations



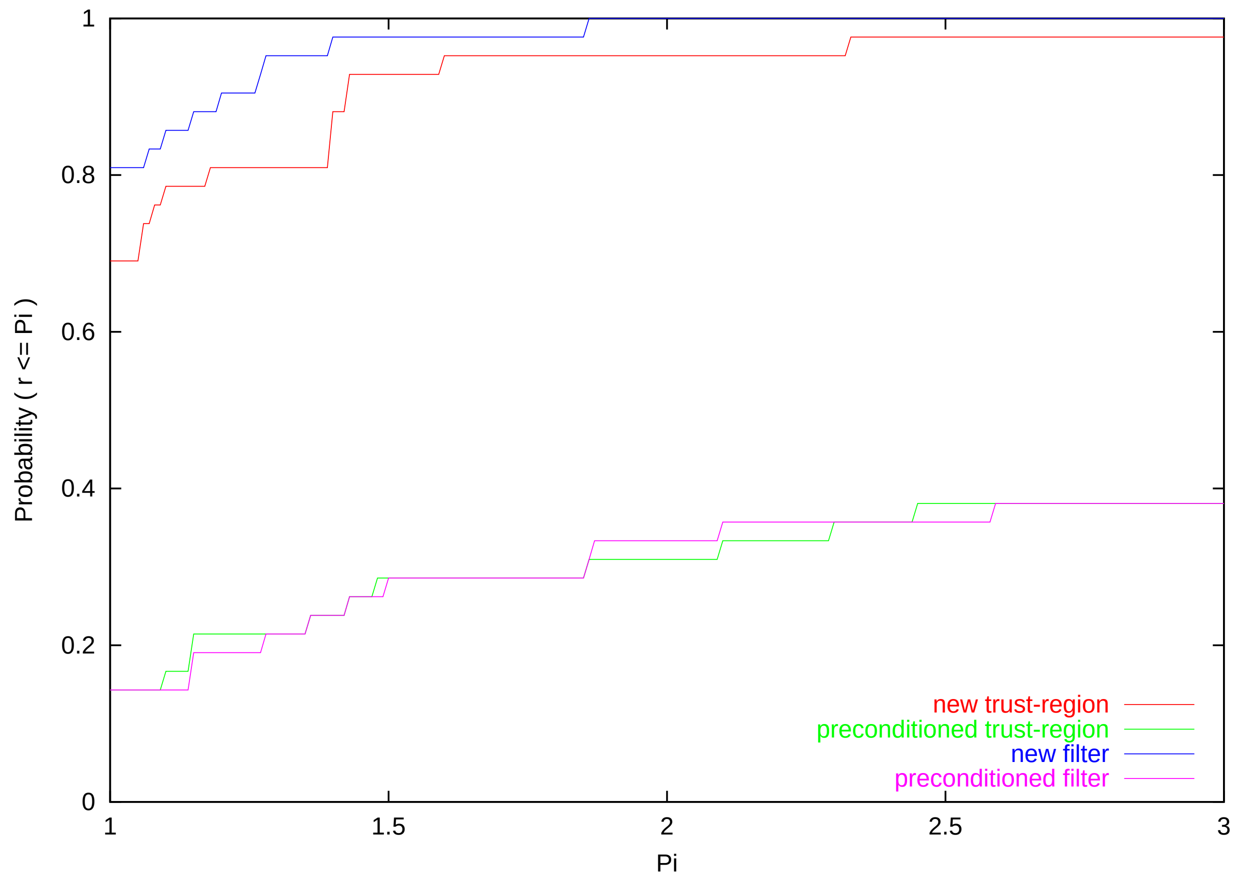
Performance profile 5

- Two trust-region variants
- CPU time



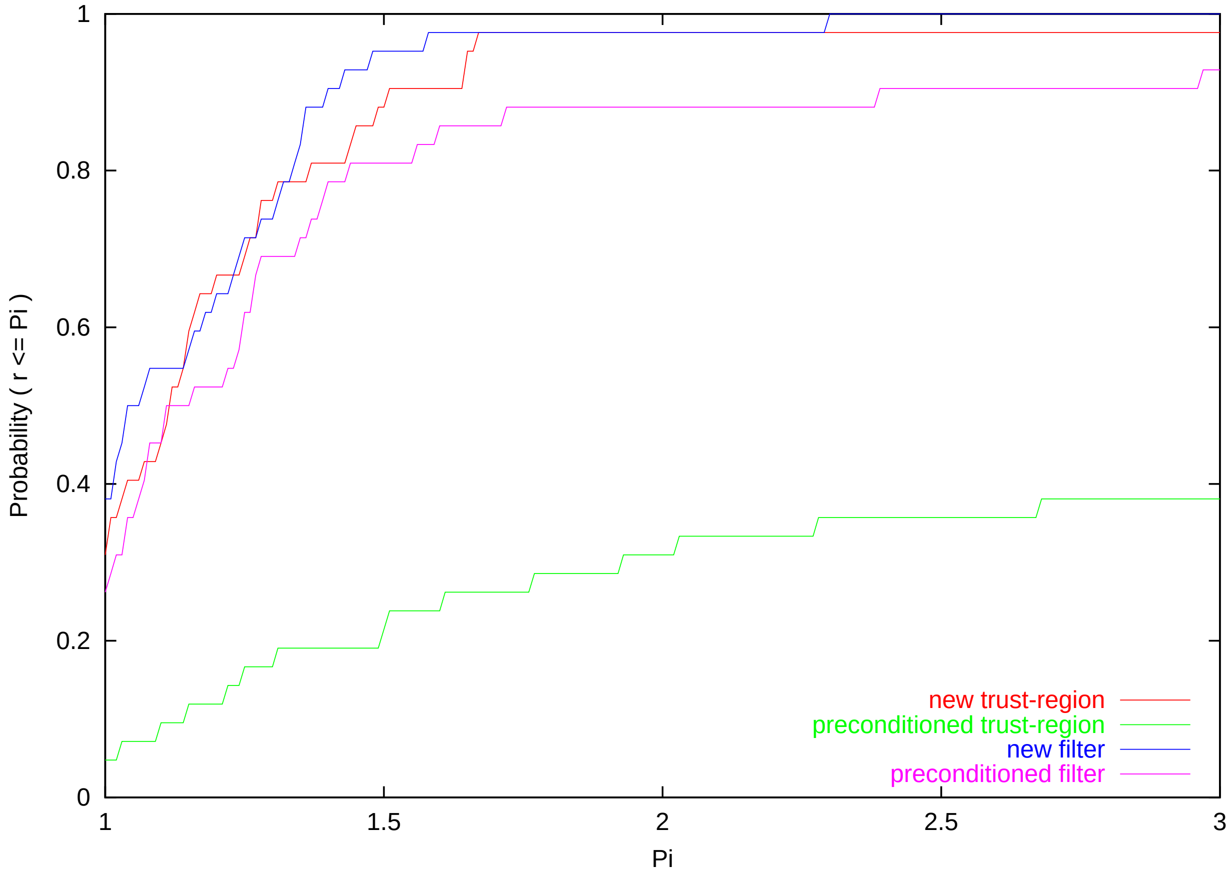
Test against preconditioning 1

- New variants vs preconditioned version of algorithms
- Number of iterations



Test against preconditioning 2

- New variants vs preconditioned version of algorithms
- CPU time



Conclusions

- Issues of singularity often arise in the estimation of DCM
- Adaptation of optimization algorithms to deal with a type of singularity
- Numerical results are very good
  - Significant improvement in term of number of iterations
  - Computational overhead highly compensed by the better efficiency
  - Significant gain expected in the estimation time of advanced DCM (Simulated Maximum Likelihood)

Perspectives

- Singularity issues
  - Perform tests on real DCM involving singularities at the solution
  - Generalization of both theoretical and algorithmic ideas to singular constrained nonlinear optimization
- Non-concavity issues
  - Nonlinear global optimization
  - Adapt optimization algorithms in order to be able to identify the global maximum of the optimization problem
  - A global maximum makes much more sense

TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Thank you for your attention !