Parameter Estimation Algorithms for kinetic modelling from noisy measurements.

F. Zama
Dario Frascari   Davide Pinelli

Department of Mathematics,
Department of Civil, Chemical, Environmental and Materials Engineering (DICAM),
University of Bologna

June 29 - July 3, 2015

Overview

Comparative study of optimization algorithms for parameter estimation of kinetic terms in models of bio-chemical processes.

- Definition of a Test Problem to obtain synthetic data.
- Test of several optimization functions available in Matlab Optimization Toolbox and Tomlab library.
- Add Gaussian noise to synthetic data and apply Morozov's Discrepancy Principle.
- Corrupt synthetic data by impulsive noise and solve by L1 norm minimization.

Parameter Estimation Problem

- Estimation of the parameter q of a model (state equation) :

$$\min_q J(u,q) \quad s.t. \quad c(u,q) = 0 \qquad \text{ODE-PDE model}$$

  - $q \in Q$ parameter to be estimated.
  - $u \equiv u(q)$ (state variable) is the unique solution of state equation $c(u,q) = 0$.

- Reduced problem: $\boxed{\min_q \widehat{J}(q)}$ where $\widehat{J}(q) \equiv J(u(q), q)$
- Some measurements $y \in Y$ of the state variable $u(q)$ are available i.e. $y = \mathcal{C}(u(q))$ where $\mathcal{C}$ is the observation operator that maps the state $u \in U$ into the measurements space Y.
- The data-fidelity term $\widehat{J}(q)$ is: $\boxed{\widehat{J}(q) \equiv \|F(q) - y\|}$ where $F(q) \equiv \mathcal{C}(u(q))$

Discrete Optimization Problem

- $\mathbf{q} \in \mathbb{R}^P$ vector of the parameters to estimated.
- $\mathbf{y}^\delta \in \mathbb{R}^N$ Noisy data.
- Discrete minimization problem: $\min_{\mathbf{q}} \widehat{J}(\mathbf{q})$
  - Gaussian Noise: Nonlinear Least Squares Problem

$$\widehat{J}(\mathbf{q}) \equiv \frac{1}{2}\|F(\mathbf{q}) - \mathbf{y}^\delta\|_2^2$$

  - Impulsive noise: L1 norm minimization

$$\widehat{J}(\mathbf{q}) \equiv \|F(\mathbf{q}) - \mathbf{y}^\delta\|_1$$

- $F(\mathbf{q})$ is obtained by:
  1. Solve the discrete state equation (ODE-PDE) model:

$$\text{find} \quad \mathbf{u} \quad s.t. \quad c(\mathbf{u}, \mathbf{q}) = 0$$

  2. Compute the values corresponding to the measurements $\mathbf{y}^\delta$ by applying the observation operator $\mathcal{C}(\mathbf{u}, \mathbf{q})$.

The Differential Model

$$\begin{aligned} \frac{\partial u}{\partial t} &= -\nu \frac{\partial u}{\partial z} + D \frac{\partial^2 u}{\partial z^2} - R(u, v, \boldsymbol{\theta}) \\ \delta_v \frac{\partial v}{\partial t} &= R(u, v, \boldsymbol{\theta}) \end{aligned} \tag{1}$$

where

$$R(u, v, \boldsymbol{\theta}) = \theta_1 \left(u - \frac{\theta_2 v}{\theta_3 - v} \right)$$

- Spatial Domain: $z \in [z_a, z_b]$
- δ_v retardation factor
- Boundary Conditions: $u(t, z_a) = u_a$, $\frac{\partial u}{\partial z}(t, z_b) = 0$,
- Initial Condition: $u(0, z) = u_0(z)$, $v(0, z) = v_0(z)$

Ref. D. Frascari et al. *Olive mill wastewater valorisation through Phenolic compounds adsorption in a continuous flow column* Chemical Engineering Journal, submitted.

Test Problem

- Model state equation:

$$\begin{cases} \frac{\partial u}{\partial t} = -\nu \frac{\partial u}{\partial z} + D\frac{\partial^2 u}{\partial z^2} - R(u, v, \boldsymbol{\theta}) + f_u \\ \delta_v \frac{\partial v}{\partial t} = R(u, v, \boldsymbol{\theta}) + f_v \end{cases} \quad (2)$$

- Parameter vector $\mathbf{q}_{true} = [\theta_1, \theta_2, \theta_3] = [1, 2, 3]$
- f_u and f_v defined by the given solution:

$$u(t, z) = e^{(-\pi^2 t)}(\sin(\pi z^2)) \cos(0.5\pi z^2), \quad v(t, z) = e(-\pi^2 t)(\sin(\pi z))$$

- Initial and boundary conditions defined by the solution (u, v).
- Retardation factor $\delta_v = 2.2$
- Time interval $[0, 0.1]$.
- Solved by the method of lines using matlab `ode15s`.
- Data $\mathbf{y} = F(\mathbf{q}_{true})$ computed in a uniformly spaced grid of $N_t \times N_z$ points.

Nonlinear Least Squares Problem

Gauss Newton - Levemberg Marquardt Iteration step:

$$\mathbf{q}^{(k+1)} = \mathbf{q}^{(k)} + \alpha_k \mathbf{s}_k$$

where

$$\left((J_F^{(k)})^t J_F^{(k)} + \lambda_k I\right) \mathbf{s}_k = -(J_F^{(k)})^t (F(\mathbf{q}^{(k)}) - \mathbf{y}^\delta)$$

- $0 < \alpha_k \leq 1$ damping parameters.
- $\lambda_k = 0$ for the Gauss Newon method.
- $J_F^{(k)}$ is the Jacobian matrix $(J_F^{(k)})_{i,j} = \partial F_i(\mathbf{q}) / \partial \mathbf{q}_j$
- Main computation step: approximation of the Jacobian
  - Finite differences method (forward-central): solve P or $2P$ forward problems (state equation).
  - General: available in optimization libraries.
  - Time-consuming.

Nonlinear Least Squares Problem

Quasi Newton Iteration step:

$$\mathbf{q}^{(k+1)} = \mathbf{q}^{(k)} + \alpha_k \mathbf{s}_k$$

where

$$H^{(k)}\mathbf{s}_k = -\nabla_{\mathbf{q}} \widehat{J}(\mathbf{q}^{(k)})$$
$$\text{compute} \quad H^{(k+1)} = H^{(k)} + S(\mathbf{q}^{(k)})$$

- BFGS update: $$S(\mathbf{q}^{(k)}) = \frac{v_k v_k^t}{v_k^t \mathbf{s}_k} - \frac{H^{(k)}\mathbf{s}_k \mathbf{s}_k^t H^{(k)}}{\mathbf{s}_k^t H^{(k)} \mathbf{s}_k}$$

  where $v_k = \nabla_{\mathbf{q}}(\widehat{J}(\mathbf{q}^{(k+1)})) - \nabla_{\mathbf{q}}(\widehat{J}(\mathbf{q}^{(k)}))$
- Initial definition $H^{(0)} = \gamma I$, γ scaling of the variables.
- $0 < \alpha_k \leq 1$ damping parameters.

Ref.

Experiments without data noise

Methods used:

- Gauss Newton Method. Stopping criterion:

$$G_F \mathbf{s}^{(k)} < \tau_F \text{ and } \|G_F\|_\infty < 10(\tau_F + \tau_X), \quad \|s^{(k)}\|_\infty < \tau_X$$

  where $G_F = 2(J_F^{(k)})^t (F(\mathbf{q}^{(k)}) - \mathbf{y}^\delta)$

- Quasi Newton BFGS. Stopping criterion:

$$\|\nabla \widehat{J}(\mathbf{q}^{(k)})\|_\infty < \tau_F (1 + \|\nabla \widehat{J}(\mathbf{q}^{(0)})\|_\infty)$$

$$\max_i \left(\frac{|q_i^{(k+1)} - q_i^{(k)}|}{1 + |q_i|} \right) < \tau_X$$

$\tau_F = \tau_X = 10^{-6}$

Experiments without data noise

- Initial parameter guess $\mathbf{q}_0 = \mathbf{q}_{true}(1 + \delta_q)$

	N_t	N_z	$\delta_q(\%)$	PRE	ResN	fval (it)
Gauss Newton	5	5	25	3.7589e-13	1.9595e-29	24(5)
	20	10	25	2.3943e-13	1.7037e-27	24(5)
	20	20	25	3.2203e-13	3.9651e-26	24(5)
	5	5	40	8.6163e-14	8.4433e-29	32(8)
	20	10	40	3.5806e-12	5.9765e-26	32(7)
	20	20	40	3.3278e-12	1.6605e-24	32(8)

	N_t	N_z	$\delta_q(\%)$	*PRE*	*ResN*	fval (it)
Quasi Newton BFGS	5	5	25	1.2708e-02	2.8781e-09	80(18)
	20	10	25	1.3880e-02	3.0633e-08	80(18)
	20	20	25	1.2106e-04	5.4168e-12	116(28)
	5	5	40	5.1000e-02	3.9129e-08	104(24)
	20	10	40	4.0196e-04	3.0938e-11	132(28)
	20	20	40	6.9894e-04	1.9933e-10	120(26)

$PRE = \|\mathbf{q}_{true} - \mathbf{q}_{it}\| / \|\mathbf{q}_{true}\|$, $ResN = \|F(\mathbf{q}_{it}) - \mathbf{y}\|$

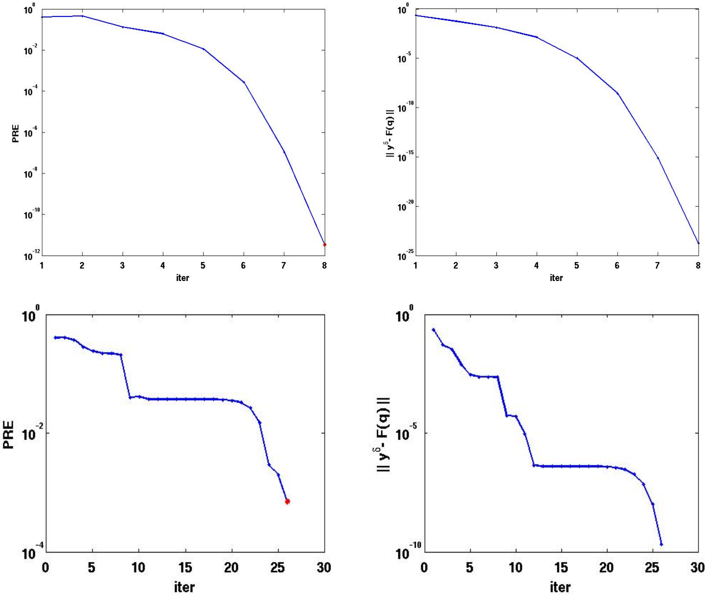
Convergence Plots: $\delta_q = 40\%$, $N_t = N_z = 20$

Gauss

Newton

Quasi

Newton



Case $N_t = 20$, $N_z = 10$, $\delta_q = 25\%$

Gaussian random noise of level δ:

$$\mathbf{y}^\delta = \mathbf{y} + \delta\|\mathbf{y}\|\boldsymbol{\eta}, \qquad \|\eta\| = 1$$

Gauss Newton

δ	PRE	$\|F(\mathbf{q}) - \mathbf{y}^\delta\|$	fval(it)
1e-6	8.6040e-5	9.5454e-11	24(6)
1e-4	8.6964e-3	9.5455e-07	24(6)
1e-3	**9.6303e-2**	9.5466e-05	24(6)
5e-3	9.1788e-1	2.3879e-03	28(7)

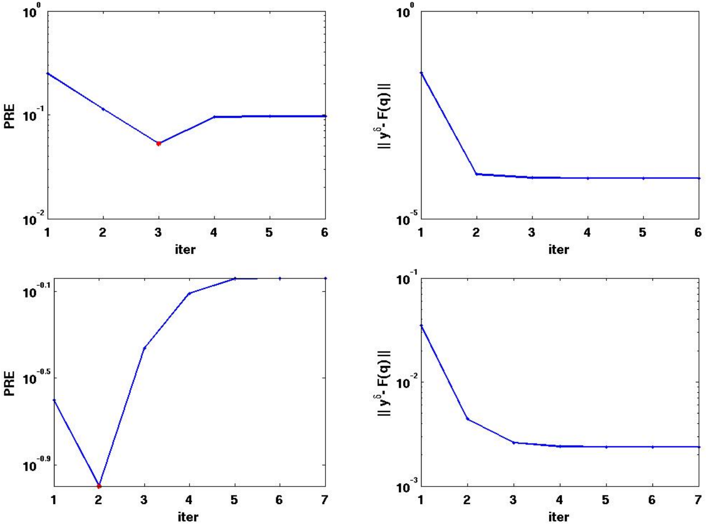
Quasi Newton

δ	PRE	$\|F(\mathbf{q}) - \mathbf{y}^\delta\|$	fval(it)
1e-6	1.3879e-02	3.1096e-8	80(18)
1e-04	1.3943e-02	9.8731e-7	80(18)
1e-3	**9.1184e-3**	9.7764e-5	124(28)
5e-03	2.0990e-01	2.4125e-3	132(31)
1e-02	2.9171e-01	9.7249e-3	136(32)

Convergence Plots: Gauss Newton

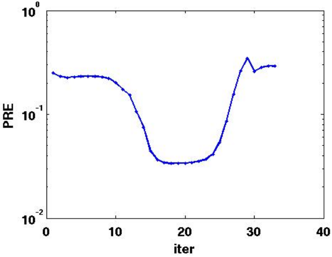
$\delta = 0.001$

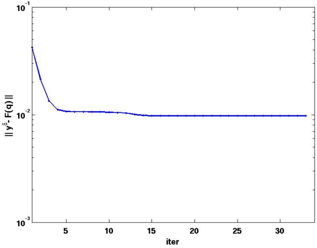
$\delta = 0.005$



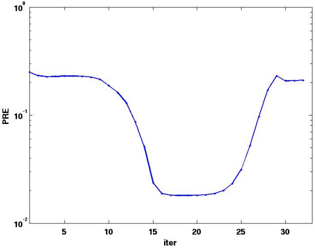
Convergence Plots: Quasi Newton BFGS

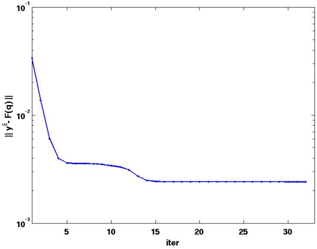
$\delta = 0.01$





$\delta = 0.005$





Regularization of semiconvergent iterations

- Semiconvergent error curve for BFGS.
- Improve the solution quality by suitable stopping rules.

Morozov's Discrepancy Principle (MDP).
When using MDP to compute a regularized solution of an iterative method, we stop the iterations at an index d s.t.

$$\|F(\mathbf{q}_d) - \mathbf{y}^\delta\| \leq \sigma(\delta), \quad \sigma(\delta) \simeq \delta \tag{3}$$

The success of this method is based on the correct estimation of the noise level δ.
Proposed stopping rule (SR_d):

$$\boxed{|J_k - J_{k-1}| < \tau|J_2 - J_1|, \quad \tau = 10^p, \quad p = \lceil \log_{10}(J_2^3) \rceil}$$

where $J_k = \widehat{J}(\mathbf{q}_k)$

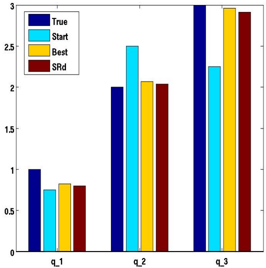
Quasi Newton BFGS with SR_d

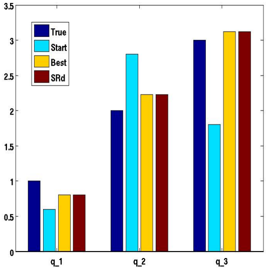
Results with $N_t = 20$, $N_z = 10$, $\delta_q = 25\%$

δ	PRE	fval(it)	PRE_{opt}	it	PRE_d	d
1e-6	1.3879e-2	80(18)	1.3853e-2	18	1.3963e-2	17
1e-4	1.3943e-2	80(18)	1.3912e-2	18	1.4037e-2	17
1e-3	1.5598e-1	132(30)	1.0628e-2	24	1.6370e-2	17
5e-3	2.0990e-01	132(31)	1.7999e-02	18	1.8132e-02	17
1e-2	2.9171e-01	136(32)	3.3698e-02	18	3.4146e-02	17

$\delta = 1e-2$

$\delta_q = 25\%$





$\delta = 1e-2$

$\delta_q = 40\%$

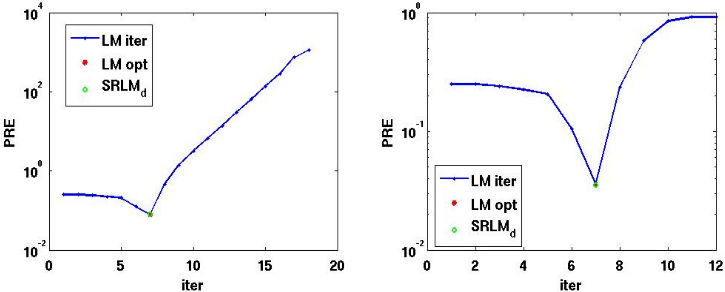
Regularization by Levemberg Marquardt

$$\left((J_F^{(k)})^t J_F^{(k)} + \lambda_k I\right) \mathbf{s}_k = -(J_F^{(k)})^t (F(\mathbf{q}^{(k)}) - \mathbf{y}^\delta)$$

- $\lambda_k = \lambda_0 10^{-k}$, $\lambda_0 > 0$ (Matlab)
- Proposed stopping rule ($SRLM_d$):

$$\boxed{|J_k - J_{k-1}| < \tau |J_2 - J_1|, \quad \tau = 10^p, \quad p = \lceil \log_{10}(J_2) \rceil}$$

$\delta = 0.01$ $\delta = 0.005$



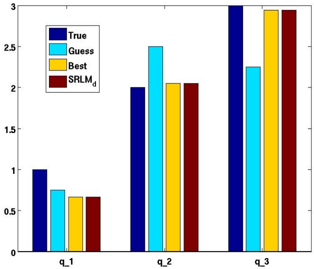
Regularization by Levemberg Marquardt

Levemberg Marquardt

δ	PRE	fval(it)	PRE_{opt}	it	PRE_d	d	λ_0
1e-4	9.5455e-07	32(7)	1.5835e-3	5	1.5835e-3	5	0.01
1e-3	9.6323e-2	32(7)	1.0942e-2	4	4.270019e-2	5	0.1
5e-3	9.1834e-1	48(11)	3.5532e-2	7	3.5532e-2	7	1
1e-2	1.1286e+3	72(17)	7.9780e-2	7	7.9780e-2	7	10

$\delta = 1e-2$

$\delta_q = 25\%$

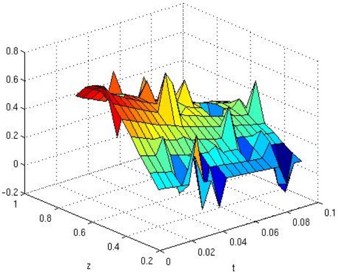


Data Corrupted by Impulsive Noise

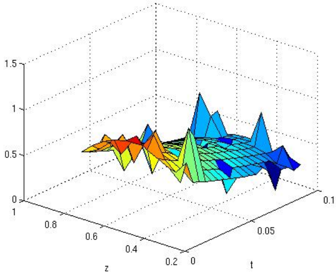
Given $\boldsymbol{\eta}$ vector with $\{0, \pm 1\}$ elements, define $\mathbf{y}^\delta$ corrupted by impulsive noise of level δ, as:

$$\mathbf{y}^\delta = \mathbf{y} + \delta \|\mathbf{y}\| \boldsymbol{\eta}, \quad \sum_i (|\eta_i|) = \lceil P \cdot N_t \cdot N_z \rceil, \quad P < 1$$

$\mathbf{y}_u^\delta$



$\mathbf{y}_v^\delta$



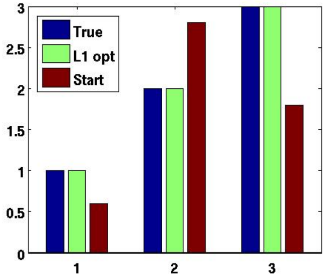
Case $N_t = 20,\ N_z = 10$

- Problem solved by `Tomlab` Solver `L1solve`:

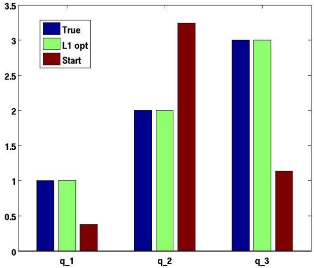
$$\min \|F(\mathbf{q}) - \mathbf{y}^\delta\|_1 \quad \Rightarrow \quad \min \sum_i w_i \ s.t. \ -\mathbf{w} \le F(\mathbf{q}) - \mathbf{y}^\delta \le \mathbf{w}$$

$P\%$	$\delta\%$	$\delta_p\%$	PRE	$\|F(\mathbf{q}) - \mathbf{y}\|_1$	fval(it)
80	50	40	1.4281e-12	8.3817e-12	15(13)
80	50	60	1.0567e-08	2.8787e-09	16(13)
80	50	62	1.9426e-06	7.5940e-07	39(27)

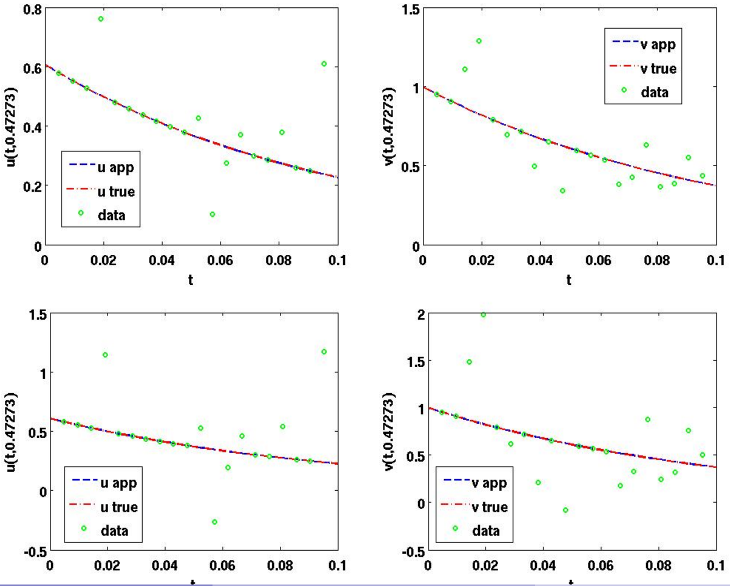
$\delta_p = 40\%$



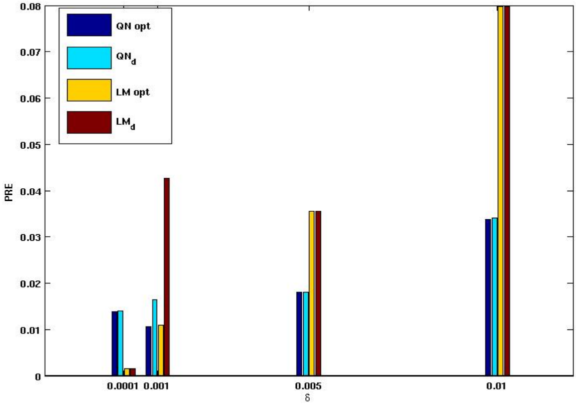
$\delta_p = 62\%$



Case $N_t = 20$, $N_z = 10$



Gaussian Random Noise



Low noise $\Rightarrow$ Levemberg Marquardt,

Medium-High noise $\Rightarrow$ Regularized BFGS

Final Remarks

- Computation Times ($\mathit{Intel}\ i5\ \mathit{CPU}, 1.60\mathit{GHz}\ x\ 4,\ 8\ \mathit{Gb}\ \mathit{RAM}$):
  - Gauss Noise $\delta = 0.01$:

$$QN\ BFGS + SR_d = 78\ sec. \quad LM + LMSR_d = 35\ sec.$$

  - Impulsive Noise $\delta_p = 40\%$:

$$\texttt{Tomlab}\ \ \texttt{L1Solve} = 63\ sec.$$

- Further steps:
  - Solve a regularized optimization problem:

$$\min_{\mathbf{q}} \widehat{J}(\mathbf{q}) + \alpha \mathcal{R}(\mathbf{q})$$

  - Application to Real Data.

Thanks for your attention