The Chi-squared Distribution of the Regularized Least Squares Functional for Regularization Parameter Estimation

Rosemary Renaut

Arizona State University
Department of Mathematics and Statistics

Prague 2008
Outline

Introduction and Motivation
  Some Standard (or NOT) Methods for Regularization Parameter Estimation

Statistical Results for Least Squares

Implications of Statistical Results for Regularized Least Squares

Newton algorithm

Algorithm with LSQR

Results

Conclusions and Future Work

Other Detailed Results
  Results for Validating LSQR Implementation with GSVD
Least Squares for $Ax = b$

- Consider discrete systems: $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$, $x \in \mathbb{R}^n$

$$Ax = b + e,$$

- **Classical Approach** Linear Least Squares

$$x_{LS} = \arg \min_x ||Ax - b||_2^2$$

- **Difficulty** $x_{LS}$ is sensitive to changes in the right hand side $b$ when $A$ is ill-conditioned.

**System is numerically ill-posed.**
Handling the ill-posedness: Given multiple measurements of data

Include some additional information about the solution and/or the data

- Usually error in $b$, $e$ is an $m-$vector of random measurement errors with mean 0 and **positive definite covariance** matrix $C_b = E(ee^T)$.

- Suppose $C_b$ is known. (Calculate if given multiple $b$)
  - For **uncorrelated** measurements $C_b$ is **diagonal** matrix of **standard deviations** of the errors. (Colored noise)

- Perhaps the fit to data can be calculated in a **weighted** norm.

- Let $W_b = C_b^{-1}$ and $L_b L_b^T = W_b$ be the Choleski factorization of $W_b$ and weight the equation: $L_b A x = L_b b + \tilde{e}$, ie

  \[
x_{WLS} = \arg \min_{x} \{ \| b - A x \|_W^2 \}
\]

- Then, theoretically, $\tilde{e}$ are uncorrelated. (**White noise**).

- But system ill-conditioning is usually **deteriorated** if noise is far from white (see Hansen).
Alternative: Introduce a Mechanism for Regularization

**Weighted Fidelity with Regularization**

- Regularize
  \[ x_{LS} = \arg \min_x \{ \| b - Ax \|^2_{W_b} + \lambda^2 R(x) \}, \]
  where \( R(x) \) is a regularization term
- \( \lambda \) is a regularization parameter which is unknown.
- Notice that the solution is \( x_{LS}(\lambda) \), dependent on \( \lambda \). It also depends on choice of \( R \).

**Requirements**

- Depends on \( R \) - what to chose?
- Depends on \( \lambda \) - what to chose?
A Specific Choice $R(x) = \|D(x - x_0)\|^2$: Tikhonov Regularized

Generalized Tikhonov regularization: Given matrix $D$ that is suitable.

\[ \hat{x} = \arg\min_x J(x) = \arg\min_x \{\|Ax - b\|^2_W + \lambda^2 \|D(x - x_0)\|^2\}. \]  

1. Assume $\mathcal{N}(A) \cap \mathcal{N}(D) = \emptyset$
2. Weighting matrix $W_b$ is inverse covariance matrix for data $b$.
3. $x_0$ is a reference solution, often $x_0 = 0$.
4. Solution

\[ \hat{x}(\lambda) = \arg\min_x J(x) = \arg\min_x \{\|Ax - b\|^2_W + \lambda^2 \|D(x - x_0)\|^2\}. \]  

Question

Given $D$, how do we find $\lambda$?
Some standard approaches I: L-curve - *Find the corner*

- Let $\mathbf{r}(\lambda) = (A(\lambda) - A)b$:
  
  Influence Matrix
  
  $A(\lambda) = A(A^T W_b A + \lambda^2 D^T D)^{-1} A^T$

- Plot
  
  $\log(\|D\mathbf{x}\|), \log(\|\mathbf{r}(\lambda)\|)$

  Trade off contributions.

- **Expensive** - requires range of $\lambda$.

- **GSVD** makes calculations **efficient**.

- **Not statistically based**

**Find corner**

**No corner**
Some standard approaches II: Generalized Cross-Validation (GCV)

- Minimizes GCV function
  \[ \frac{\|b - Ax(\lambda)\|_W^2}{\text{trace}(I_m - A(\lambda))^2} \]
  which estimates predictive risk.

- Expensive - requires range of \( \lambda \).
- GSVD makes calculations efficient.
- Statistically based
- Requires minimum

Multiple minima

Sometimes flat
Some standard approaches III: Unbiased Predictive Risk Estimation (UPRE)

- Minimize expected value of predictive risk: Minimize UPRE function

\[ \| b - Ax(\lambda) \|^2_{W_b} + 2 \text{trace}(A(\lambda)) - m \]

- Expensive - requires range of \( \lambda \).
- GSVD makes calculations efficient.
- Statistically based
- Minimum needed
Hybrid LSQR Methods to Project out the Noise

- Iterate LSQR to find solution and stop when noise starts to dominate (Hnetynkova and others)
- Solve the reduced system.
- Hybrid method - solve reduced system with additional regularization.
  - Cost of regularization of reduced system is minimal
  - Any regularization method may be used (Nagy)
- Talk to the local experts!
A More general formulation: Maximum A Posteriori Method

Formulation:

\[ \hat{x} = \arg\min J(x) = \arg\min \{ \|Ax - b\|^2_{W_b} + \| (x - x_0) \|^2_{W_D} \}. \] (3)

Notice the regularization term includes a new weighting matrix, but no mapping \( D \). (It is hidden in \( W_D \))

- **Standard:** \( W_D = \lambda^2 D^T D \), \( \lambda \) unknown penalty parameter.
- **Ideally, statistically,** \( W_D \) is inverse covariance matrix for the mapped model \( Dx \) i.e. \( \lambda = 1/\sigma_x \), \( \sigma^2_x \) the common variance in \( Dx \).
- Assumes the resulting estimates for \( Dx \) uncorrelated.
- \( \hat{x} \) is the standard **maximum a posteriori** (MAP) estimate of the solution, when all \( a \ priori \) information is provided.

Can this provide additional information?
Theorem (Rao73: First Fundamental Theorem)

Let $r$ be the rank of $A$ and for $b \sim N(Ax, \sigma^2_b I)$, (errors in measurements are normally distributed with mean 0 and covariance $\sigma^2_b I$), then

$$J = \min_x \|Ax - b\|^2 \sim \sigma^2_b \chi^2(m - r).$$

$J$ follows a $\chi^2$ distribution with $m - r$ degrees of freedom:

**Basically the Discrepancy Principle**

Corollary (Weighted Least Squares)

For $b \sim N(Ax, C_b)$, and $W_b = C_b^{-1}$ then

$$J = \min_x \|Ax - b\|^2_{W_b} \sim \chi^2(m - r).$$
Two New Results to Help Find the Regularization parameter:

**Theorem:** \( \chi^2 \) distribution of the regularized functional

\[
\hat{x} = \arg\min_{x} J_D(x) = \arg\min_{x} \{ \|Ax - b\|_{W_b}^2 + \|(x - x_0)\|_{W_D}^2 \}, \quad W_D = D^T W_x D.
\]

(4)

Assume

- \( W_b \) and \( W_x \) are symmetric positive definite.
- Problem is uniquely solvable \( \mathcal{N}(A) \cap \mathcal{N}(D) \neq 0 \).
- Moore-Penrose generalized inverse of \( W_D \) is \( C_D \)
- Statistics: \( (b - Ax) = e \sim N(0, C_b) \), \( (x - x_0) = f \sim N(0, C_D) \),
  - \( x_0 \) is the mean vector of the model parameters.

Then

\[
J_D \sim \chi^2(m + p - n)
\]
Corollary: a-priori information not mean value, e.g. $x_0 = 0$

Corollary: non-central $\chi^2$ distribution of the regularized functional

$$\hat{x} = \arg\min J_D(x) = \arg\min \left\{ \|Ax - b\|_{W_b}^2 + \|(x - x_0)\|_{W_D}^2 \right\}, \quad W_D = D^T W_x D. \quad (5)$$

Assume all assumptions as before, but $x_1 \neq x_0$ is the mean vector of the model parameters.

Let

$$c = \|c\|_2^2 = \|\tilde{Q}U^T W_b^{1/2} A(x_1 - x_0)\|_2^2$$

Then

$$J_D \sim \chi^2(m + p - n, c)$$
Implications of the Result

Statistical Distribution of the Functional

- Mean and Variance are prescribed

\[ E(J_D) = m + p - n + c \quad E(J_D J_D^T) = 2(m + p - n) + 4c \]

- Can we use this?
- **YES**

- Try to find \( W_D \) so that \( E(J) = m - n + p + 2c \) (Mead 2007)
- Mead algorithm finds \( W_D \) but is expensive
- Our proposal - find \( \lambda \) only.
What do we need to apply the Theory?

**Requirements**

- **Covariance** information $C_b$ on data parameters $b$ (or on model parameters $x$!)
- **A priori** information either $x_0$ is the mean, or mean value $x_1$.
- But $x_1$ and $x_0$ are not known.
- For repeated data measurements $C_b$ can be calculated. Also $b_1$ can be found, the mean of $b$.
- But $E(b) = AE(x)$ implies $b_1 = Ax_1$. Hence

\[
c = \|c\|^2 = \|\tilde{Q}U^T W_b^{1/2} (b_1 - Ax_0)\|^2_2
\]

Then we can use $E(J)$ to find $\lambda$
Assume $x_0$ is the mean (experimentalists do know something about the model parameters)

**DESIGNING THE ALGORITHM: I**

- Recall: if $C_b$ and $C_x$ are good estimates of covariance

\[
|J_D(\hat{x}) - (m + p - n)|
\]

should be small.

- Thus, let $\tilde{m} = m + p - n$ then we want

\[
\tilde{m} - \sqrt{2\tilde{m}z_{\alpha/2}} < J(x(W_D)) < \tilde{m} + \sqrt{2\tilde{m}z_{\alpha/2}}.
\]

(6)

- $z_{\alpha/2}$ is the relevant $z$-value for a $\chi^2$-distribution with $\tilde{m}$ degrees

**GOAL**

Find $W_D$ to make (6) tight: Single Variable case find $\lambda$

\[
J_D(\hat{x}(\lambda)) \approx \tilde{m}
\]
A Newton-line search Algorithm to find \( \lambda \).

**(Basic algebra)**

**Newton to Solve** 

\( F(\sigma) = J_D(\sigma) - \tilde{m} = 0 \)

- We use \( \sigma = 1/\lambda \), and \( y(\sigma^{(k)}) \) is the current solution for which

\[
x(\sigma^{(k)}) = y(\sigma^{(k)}) + x_0
\]

Then

\[
\frac{\partial}{\partial \sigma} J(\sigma) = -\frac{2}{\sigma^3} \|Dy(\sigma)\|^2 < 0
\]

- Hence we have a basic Newton Iteration

\[
\sigma^{(k+1)} = \sigma^{(k)} (1 + \frac{1}{2} \left( \frac{\sigma^{(k)}}{\|Dy\|} \right)^2 (J_D(\sigma^{(k)}) - \tilde{m})).
\]

- Add a line search

\[
\sigma^{(k+1)} = \sigma^{(k)} (1 + \frac{\alpha^{(k)}}{2} \left( \frac{\sigma^{(k)}}{\|Dy\|} \right)^2 (J_D(\sigma^{(k)}) - \tilde{m})).
\]
Algorithm Using the GSVD

**GSVD**

- Use GSVD of \([W_b^{1/2} A, D]\)
- For \(\gamma_i\) the generalized singular values, and \(s = U^T W_b^{1/2} r\)
- \(\tilde{m} = m - n + p\)
- \(\tilde{s}_i = s_i / (\gamma_i^2 \sigma_x^2 + 1), i = 1, \ldots, p, \quad t_i = \tilde{s}_i \gamma_i\).
- Find root of

\[
F(\sigma_x) = \sum_{i=1}^{p} \left( \frac{1}{\gamma_i^2 \sigma_x^2 + 1} \right) s_i^2 + \sum_{i=n+1}^{m} s_i^2 - \tilde{m} = 0
\]

- Equivalently: solve \(F = 0\), where

\[
F(\sigma_x) = s^T \tilde{s} - \tilde{m} \quad \text{and} \quad F'(\sigma_x) = -2\sigma_x \|t\|_2^2.
\]
Discussion on Convergence

- $F$ is monotonic decreasing $(F'(\sigma_x) = -2\sigma_x\|t\|_2^2)$
- Solution either exists and is unique for positive $\sigma$
- Or no solution exists $F(0) < 0$.
  - implies incorrect statistics of the model
- Theoretically, $\lim_{\sigma \to \infty} F > 0$ possible.
  - Equivalent to $\lambda = 0$. No regularization needed.
Find the parameter

- **Step 1:** Bracket the root by logarithmic search on $\sigma$ to handle the asymptotes: yields $\text{sigmamax}$ and $\text{sigmamin}$

- **Step 2:** Calculate step, with steepness controlled by $\text{tolD}$. Let $t = D y / \sigma^{(k)}$, where $y$ is the current update, given from the GSVD, then

  \[
  \text{step} = \frac{1}{2} \left( \frac{1}{\max \left\{ \|t\|, \text{tolD} \right\}} \right)^2 (J_D(\sigma^{(k)}) - \tilde{m})
  \]

- **Step 3:** Introduce line search $\alpha^{(k)}$ in Newton

  \[
  \text{sigmanew} = \sigma^{(k)} (1 + \alpha^{(k)} \text{step})
  \]

  $\alpha^{(k)}$ chosen such that $\text{sigmanew}$ within bracket.
Practical Details of Algorithm: Large Scale problems

Algorithm

Initialization

- Convert generalized Tikhonov problem to standard form.
- Use LSQR algorithm to find the bidiagonal matrix spanning appropriate space of solution.
- Obtain a solution of the bidiagonal problem for given $\sigma$.
- Resuse bidualization in update of $\sigma$ for Newton.
- Each $\sigma$ calculation of algorithm reuses saved information from the Lancos bidualization. The system is augmented if needed.
Comparison with Standard LSQR hybrid Algorithm

- Algorithm concurrently regularizes and solves the system.
- In contrast, standard hybrid LSQR solves projected system with regularization.
- Needs only cost of standard LSQR algorithm with some updates for solution solves for iterated $\sigma$.
- The regularization introduced by LSQR projection may be useful for preventing problems with GSVD expansion.
- Makes algorithm viable for large scale problems.
Recall: Implementation Assumptions

Covariance of Error: Statistics of Measurement Errors

- Information on the covariance structure of errors in $b$ needed.
- Use $C_b = \sigma_b^2 I$ for common covariance, white noise.
- Use $C_b = \text{diag}(\sigma_1^2, \sigma_2^2, \ldots, \sigma_m^2)$ for colored uncorrelated noise.
- With no noise information $C_b = I$.
- Use $b_1$ as the mean of measured $b$, when implemented with centrality parameter, $x_0 = 0$.

Tolerance on Convergence

- The convergence tolerance depends on the noise structure.
- Use $TOL = \sqrt{2\tilde{m}z_{\alpha/2}}$.
- No noise structure use $\alpha = .001$, generates large TOL
- Good noise information use $\alpha = .95$, generates small TOL
Recall: Implementation Assumptions

**Covariance of Error: Statistics of Measurement Errors**

- Information on the covariance structure of errors in $b$ needed.
- Use $C_b = \sigma_b^2 I$ for common covariance, white noise.
- Use $C_b = \text{diag}(\sigma_1^2, \sigma_2^2, \ldots, \sigma_m^2)$ for colored uncorrelated noise.
- With no noise information $C_b = I$.
- Use $b_1$ as the mean of measured $b$, when implemented with centrality parameter, $x_0 = 0$.

**Tolerance on Convergence**

- The convergence tolerance depends on the noise structure.
- Use $\text{TOL} = \sqrt{2\tilde{m}z_{\alpha/2}}$.
- No noise structure use $\alpha = .001$, generates large TOL
- Good noise information use $\alpha = .95$, generates small TOL
An example of the method: Seismic Signal Restoration

The Data Set and Goal

- Real data set of 48 signals of length 3000.
- The point spread function is derived from the signals.
- Calculate the signal variance pointwise over all 48 signals.
- Goal: restore the signal $x$ from $Ax = b$, where $A$ is psf matrix and $b$ is given blurred signal.

Method of Comparison- no exact solution known

- No exact solution.
- Downsample the signal and restore for different resolutions
  
  Resolution 2 : 1 5 : 1 10 : 1 20 : 1 100 : 1
  Points 1500 600 300 150 30
- Do results converge? Compare with UPRE and L-Curve.
An example of the method: Seismic Signal Restoration

The Data Set and Goal

- Real data set of 48 signals of length 3000.
- The point spread function is derived from the signals.
- Calculate the signal variance pointwise over all 48 signals.
- Goal: restore the signal $x$ from $Ax = b$, where $A$ is psf matrix and $b$ is given blurred signal.

Method of Comparison- no exact solution known

- No exact solution.
- Downsample the signal and restore for different resolutions
  - Resolution: 2 : 1, 5 : 1, 10 : 1, 20 : 1, 100 : 1
  - Points: 1500, 600, 300, 150, 30
- Do results converge? Compare with UPRE and L-Curve.
Greater contrast with $\chi^2$. UPRE is insufficiently regularized. L-curve severely undersmooths (not shown). Parameters not consistent across resolutions.
Regularization Parameters are consistent: $\sigma = 0.01005$ all resolutions
THE GSVD SOLUTION: White Noise (left) and Colored Noise (right)

\[ x_0 = 0 \]

Regularization Parameters are consistent:

\[ \sigma = 0.00058 \text{ (left), } \sigma = 0.00069 \text{ (right) all resolutions} \]
Regularization quite consistent resolution 2 to 100

\[ \sigma = 0.0000029, 0.0000029, 0.0000029, 0.0000057, 0.0000057 \text{ (left)} \]

\[ \sigma = 0.000007, 0.000007, 0.000007, 0.000007, 0.00012 \text{ (right)}. \]

Notice that colored noise eliminates second arrival of signal but excellent contrast to identify primary arrival.
More Signals! UPRE and L-curve exhibit under regularization.
Conclusions For the Case when $x_0$ known

Observations

- A new statistical method for estimating regularization parameter
  - Compares favorably with UPRE with respect to performance and compared to L-curve. (GCV is not competitive).
- Method can be used for large scale problems.
- Method is very efficient, Newton method is robust and fast.
- But $x_0$ is the mean of $x$ is needed.
Difficulties when central parameter is required

What are the issues?

- Function need not be monotonic
- More problematic for NonCentral version with $x_0$ not the mean. (ie $x_0 = 0$.
- $\sigma$ can be bounded by result of central case.
- Range of $\sigma$ given by range of $\gamma_i$.
- May oversmooth the solution if good range of $\sigma$ not found.
Future Work

Other Results and Future Work

- Degrees of freedom reduced when using the GSVD.
- How to apply Picard condition for GSVD to handle problems with robustness due to conditioning of $C_b$
- Image deblurring. (Implementation to use minimal storage)
- Diagonal Weighting Schemes
- Edge preserving regularization
- Constraint implementation (with Mead submitted).
Some Small Scale Experiments: Verify Robustness of LSQR/GSVD

Details

- Take example from Hansen’s toolbox, eg shaw, phillips, heat, ilaplace.
- Generate 500 copies for each noise level, here .005, .01, .05, .1.
- Solve for 500 cases using GSVD and LSQR Newton.
- Pairwise t test on obtained $\sigma$: verify equivalence GSVD and LSQR.
- Compare results with statistical technique: UPRE
  - Errors - relative least squares, and max error. Calculate over all errors less than .5.
  - Regularization parameter (not given).
Example of the Colored Noise distribution

The pointwise variance for each noise level
Noise and Exact Right Hand Side Vector

Noise distribution for problem shaw

Actual Noise Levels
Noise and Exact Right Hand Side Vector

Noise distribution for problem iLaplace

Actual Noise Levels
Noise and Exact Right Hand Side Vector

Noise distribution for problem heat

Actual Noise Levels

MATHEMATICS AND STATISTICS
Problem size 64, Regularization First Order Derivative, Colored Noise

Table: Convergence characteristics and P-values comparing GSVD and LSQR iteration steps and $\sigma$ values

<table>
<thead>
<tr>
<th></th>
<th>Bidiagonal</th>
<th>Newton Steps</th>
<th>P value</th>
<th>$\sigma$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\epsilon$</td>
<td>Average</td>
<td>GSVD</td>
<td>LSQR</td>
</tr>
<tr>
<td>$5e-02$</td>
<td>$7.2$</td>
<td>$7.1$</td>
<td>$7.1$</td>
<td>$1.000e+00$</td>
</tr>
<tr>
<td>$1e-01$</td>
<td>$6.2$</td>
<td>$7.0$</td>
<td>$7.0$</td>
<td>$9.166e-01$</td>
</tr>
<tr>
<td></td>
<td>$8.0$</td>
<td>$6.5$</td>
<td>$6.5$</td>
<td>$9.384e-01$</td>
</tr>
<tr>
<td>$1e-01$</td>
<td>$7.4$</td>
<td>$6.5$</td>
<td>$6.5$</td>
<td>$8.610e-01$</td>
</tr>
<tr>
<td></td>
<td>$29.4$</td>
<td>$5.8$</td>
<td>$5.8$</td>
<td>$1.000e+00$</td>
</tr>
<tr>
<td>$1e-01$</td>
<td>$22.4$</td>
<td>$6.0$</td>
<td>$6.0$</td>
<td>$1.000e+00$</td>
</tr>
<tr>
<td></td>
<td>$15.4$</td>
<td>$6.1$</td>
<td>$6.1$</td>
<td>$1.000e+00$</td>
</tr>
<tr>
<td>$1e-01$</td>
<td>$12.8$</td>
<td>$6.1$</td>
<td>$6.1$</td>
<td>$1.000e+00$</td>
</tr>
</tbody>
</table>
Problem size 64, Regularization First Order Derivative, Colored Noise

<table>
<thead>
<tr>
<th></th>
<th>Least Squares Error</th>
<th>Max Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\epsilon)</td>
<td>GSVD</td>
<td>LSQR</td>
</tr>
<tr>
<td>.05</td>
<td>.34(207)</td>
<td>.34(207)</td>
</tr>
<tr>
<td>.1</td>
<td>.39(105)</td>
<td>.39(105)</td>
</tr>
</tbody>
</table>

**ilaplace**

<table>
<thead>
<tr>
<th></th>
<th>Least Squares Error</th>
<th>Max Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\epsilon)</td>
<td>GSVD</td>
<td>LSQR</td>
</tr>
<tr>
<td>.05</td>
<td>.23(361)</td>
<td>.23(361)</td>
</tr>
<tr>
<td>.1</td>
<td>.25(317)</td>
<td>.25(317)</td>
</tr>
</tbody>
</table>

**heat**

<table>
<thead>
<tr>
<th></th>
<th>Least Squares Error</th>
<th>Max Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\epsilon)</td>
<td>GSVD</td>
<td>LSQR</td>
</tr>
<tr>
<td>.05</td>
<td>.27(500)</td>
<td>.27(500)</td>
</tr>
<tr>
<td>.1</td>
<td>.35(497)</td>
<td>.35(497)</td>
</tr>
</tbody>
</table>

**phillips**

<table>
<thead>
<tr>
<th></th>
<th>Least Squares Error</th>
<th>Max Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\epsilon)</td>
<td>GSVD</td>
<td>LSQR</td>
</tr>
<tr>
<td>.05</td>
<td>.10(500)</td>
<td>.10(500)</td>
</tr>
<tr>
<td>.1</td>
<td>.12(500)</td>
<td>.12(500)</td>
</tr>
</tbody>
</table>

**Table:** Comparison with UPRE, Relative Least Squares and max error, in parentheses the number of accepted values, large values (> .5) excluded from the average
### Summary of Results

#### Major Observations

| GSVD-LSQR | ▶ High correlation between $\sigma$ for both $\chi^2$ (GSVD and LSQR) obtained. $p$-values at or near 1.  
▶ Algorithms converge with very few regularization calculations, average $\approx 7$.  
▶ The bidiagonalized system is on average much smaller than the system size.  
▶ Errors distributions equivalent.  
▶ Total cost of LSQR is the cost of bidiagonalization plus use of bidiagonalization to obtain a solution, eg column 2 plus column 4 solves. |
|---|---|
| UPRE | ▶ Errors of UPRE are similar to $\chi^2$ approaches.  
▶ UPRE fails more often (more discarded solutions with high noise). |
THANK YOU!