Iterated Tikhonov with Generalized Singular Value Decomposition

Mirjeta PASHA

Kent State University

mpasha1@kent.edu

May 23, 2018

What comes next in this presentation

1. Motivation on working with inverse problems
2. Introduction and insight on inverse problems
3. Iterated Tikhonov with GSVD
4. Comparison of some zero finders
5. Numerical tests and results

Motivation. Why inverse problems? Some examples where they arise

Inverse problems arise in many fields of science and life:

1. Medical imaging
2. Geophysics
3. Industrial process monitoring
4. Remote sensing
5. Pricing financial instruments, etc.

GOAL: Design and analyze reliable and computationally effective mathematical solution methods for inverse problems.

Inverse problems $\equiv$ Troubles

Naive solution of an equation with the blurring operator.
What we usually do?
$x = A^{\dagger} b$, where $A^{\dagger}$ is the Moore-Penrose pseudoinverse.

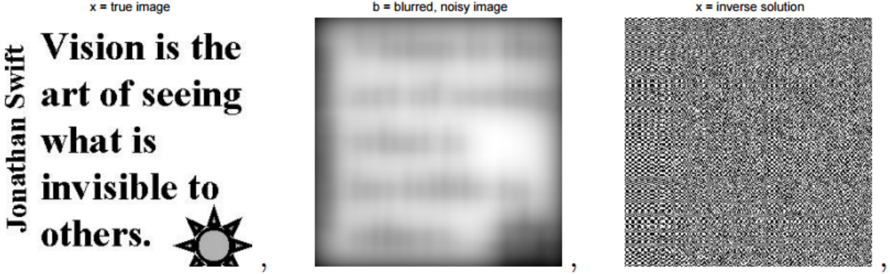
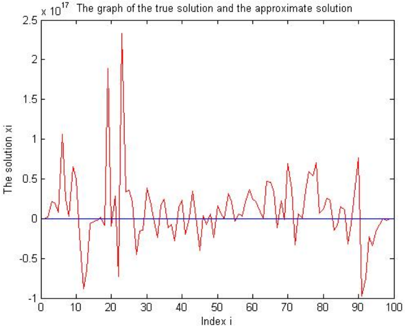


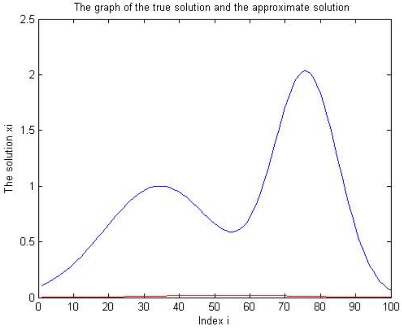
Figure: Direct solving ill-posed problems (example by James Nagy, Emory University, Atlanta).

GOAL: Solve $Ax = b$. $A \in \mathbb{R}^{100x100}$, x, b,$\in \mathbb{R}^{100}$ generated by Matlab code [A, x,b]= shaw(100).
Naive solution: $x = A^{-1}b$



Question: What will happen if I calculate the solution a little bit more carefully, i.e, x=A$\backslash b$?

GOAL: Solve $Ax = b$. $A \in \mathbb{R}^{100x100}$, x, b,$\in \mathbb{R}^{100}$ generated by Matlab code [A, x,b]= shaw(100).



Question: What will happen if there is some noise in the right- hand side b???

Motivation. Why inverse problems

Definition: (Well- posedness, Hadamard 1865-1963) A problem is called **well-posed** if:

- a solution exists
- the solution is unique
- the solution depends continuously on the given data (the solution is stable).

If at least one of the above conditions is violated, then the problem becomes ill-posed.

Problem with ill-conditioned matrices

- GOAL: Solve $Ax + e = b_{true}$. $A \in \mathbb{R}^{nxn}$, x, $b_{true}, \in \mathbb{R}^n$
- A is large ill-conditioned and maybe rank deficient.
- b represents the measured data
- e is the noise from measurements or other sources.
- Problems with the properties above are called linear ill-posed problems
- Methods to solve?? Some..
  Generally we regularize and then solve.

Notation

- A $\in \mathbb{R}^{mxn}$
- b $\in \mathbb{R}^{m}$ where $b = b_{true} + e$
- x $\in \mathbb{R}^{n}$ is the solution we are looking for.
- $\|\cdot\|$ is the Euclidian norm, i.e. $\|\cdot\| = \|\cdot\|_2$
- $\mu = \dfrac{1}{\beta}$

Introduction:

Consider the following problem:

$$\min_{x \in \mathbb{R}^n} \|Ax - b\|$$

DREAM: Would like to compute a solution of $Ax = b_{true}$
REALITY: Would like to compute an approximate solution of $Ax = b$ since b_{true} is not known.

Tikhonov regularization

The possibly most popular regularization method is Tikhonov regularization.

$$\textit{Standard form}: \quad \min_{x \in \mathbb{R}^n} \{\|Ax - b\|^2 + \mu\|x - x_0\|^2\} \tag{3}$$

$$\textit{General form}: \quad \min_{x \in \mathbb{R}^n} \{\|Ax - b\|^2 + \mu\|L(x - x_0)\|^2\} \tag{4}$$

- If L$=$ I (Identity), then the Tikhonov minimization problem is said to be in the *standard form*.
- If $L \neq I$ (Identity), then the Tikhonov minimization problem is said to be in the *general form*.
- The matrix L is chosen such that

$$\mathcal{N}(A) \cap \mathcal{N}(L) = 0 \tag{5}$$

- For $\mu > 0$ the Tikhonov minimization problem has unique solution.

The discrepancy principle

- Assume that a fairly estimate for $\delta = \|b - b_{true}\|_2$ is known. The discrepancy principle prescribes that $\mu > 0$ be chosen so that $\|Ax_\mu - b\|_2 = \eta\delta$ for some constant $\eta > 1$ independent of δ.
- Let define the function $\phi(\mu) = \|Ax_\mu - b\|^2$.

From Tikhonov to Iterated Tikhonov

- Consider x_k is our solution at some iteration k.
- Our direct solution is given by $x^\dagger = A^\dagger b$ and let $e_k = x^\dagger - x_k$
- $x^\dagger = x_k + e_k \approx x_k + h_k = x_{k+1}$
- $A(e_k) = A(x^\dagger - x_k) = b_{true} - Ax_k \approx b - Ax_k = r_k$
- $h_k = \min_{x \in \mathbb{R}^n} \|Ah - r_k\|^2 + \mu \|Lh\|^2$
- $h_k = (A^T A + \mu L^T L)^{-1} A^T (b - Ax_k)$
- $x_{k+1} = x_k + (A^T A + \mu L^T L)^{-1} A^T (b - Ax_k)$

GSVD (Generalized Singular Value Decomposition)

- The GSVD is a generalization of the SVD of A and the generalized singular values of the pair (A,L) are essentially the square roots of the generalized eigenvalues of the matrix pair $(A^T A, AA^T)$
- Let $A \in \mathbb{R}^{mxn}$ and let $L \in \mathbb{R}^{pxn}$ satisfy $m \geq n \geq p$
  Assume that $N(A) \cap N(L) = 0$ and that L has full row rank.
- The columns of U and V are orthonormal, X is nonsingular with columns that are $A^T A$ orthonormal and Σ and M are pxp diagonal matrices.
- The diagonal elements of Σ and M are nonnegative and ordered such that $0 \leq \sigma_1 \leq \sigma_2 \leq ... \leq \sigma_p \leq 1$, $0 < \mu_p \leq \mu_{p-1} \leq ... \leq \mu_1 < 1$.

GSVD (Generalized Singular Value Decomposition)

They are normalized such that $\sigma_i^2 + \mu_i^2 = 1, i = 1, 2, .., p.$

- Then, the generalized singular values γ_i of (A,L) are defined as the ratios $\gamma_i = \frac{\sigma_i}{\mu_i}, i = 1, 2, ..., p.$
- The pairs (σ_i, μ_i) are well conditioned with respect to perturbations in A and B.

Standard Tikhonov Regularization in general form

Assume that A and L are square matrices in $\mathbb{R}^{nxn}$ and define the factorizations as:

$A = U\Sigma Y^T$ *and* $L = V\Lambda Y^T$, where

- $U, V \in \mathbb{R}^{nxn}$ are orthogonal matrices
- $\Sigma = diag[\sigma_1, \sigma_2, ..., \sigma_n] \in \mathbb{R}^{nxn}$
- $\Lambda = diag[\lambda_1, \lambda_2, ..., \lambda_n] \in \mathbb{R}^{nxn}$
- The matrix Y is nonsingular

Due to (5) the unique solution will be given by:

$$x_\mu = (A^T A + \mu_k I)^{-1} A^T b \tag{12}$$

Formulas for Iterated Tikhonov with GSVD

Consider the Iterated Tikhonov formula in the general case:

- $x_{k+1} = x_k + (A^T A + \mu L^T L)^{-1} A^T (b - A x_k)$ or equivalently:
- $(A^T A + \mu L^T L) x_{k+1} = A^T b + \mu L^T L x_k$,
  $x_{k+1} = (A^T A + \mu L^T L)^{-1} (A^T b + \mu L^T L x_k)$

Formulas for Iterated Tikhonov

- Let $A = U\Sigma Y^T$ and $L = U\Lambda Y^T$. Substituting onto the above formula will get the simplified version of the iterations: $Y(\Sigma^T\Sigma + \mu\Lambda^T\Lambda)Y^T x_{k+1} = Y(\Sigma^T U^T b + \mu\Lambda^T\Lambda Y^T x_k)$
- let $Z_k = Y^T x_k$ and $\hat{b} = U^T b$ the formulas will become: $Z_{k+1} = (\Sigma^T\Sigma + \mu\Lambda^T\Lambda)^{-1}(\Sigma^T\hat{b} + \mu\Lambda^T\Lambda Z_k)$

$$\boxed{Z_{k+1} = (\Sigma^T\Sigma + \mu\Lambda^T\Lambda)^{-1}(\Sigma^T\hat{b} + \mu\Lambda^T\Lambda Z_k)} \; (\star)$$

Formulas for Discrepancy Principle, Iterated Tikhonov GSVD

- Use the idea of the iterated formula: $x_{k+1} = x_k + (A^T A + \mu L^T L)^{-1} A^T (b - Ax_k)$ and the Discrepancy principle: $\|Ax_{k+1} - b\|^2 = (\eta\delta)^2$
- Use the GSVD decomposition of A and L and plug in the iterated formula will get:
- $\|\Sigma(\Sigma^T\Sigma + \mu\Lambda^T\Lambda)^{-1}\Sigma^T(\hat{b} - \Sigma Z_k) + \Sigma Z_k - \hat{b}\|^2 = (\eta\delta)^2$
- $\sum_{j=1}^{m}(\frac{\sigma_j^2}{\sigma_j^2+\mu\lambda_j^2}(\hat{b}_j - \sigma_j(\hat{x}_k)_j) + \sigma_j(\hat{x}_k)_j - \hat{b}_j)^2 = (\eta\delta)^2$
- Using the fact $\mu = \frac{1}{\beta}$ will get the final formula:

$$\boxed{\sum_{j=1}^{m}(\frac{\hat{b}_j\lambda_j^2 - \sigma_j(\hat{x}_k)_j\lambda_j^2}{\beta\sigma_j^2 + \lambda_j^2})^2 = (\eta\delta)^2} \quad (\star\star)$$

All final formulas together

1. $Z_{k+1} = (\Sigma^T\Sigma + \mu\Lambda^T\Lambda)^{-1}(\Sigma^T\hat{b} + \mu\Lambda^T\Lambda Z_k)$ $(\star)$

2. $\phi(\beta) = \sum_{j=1}^{m} \left(\frac{\hat{b}_j\lambda_j^2 - \sigma_j(\hat{x_k})_j\lambda_j^2}{\beta\sigma_j^2 + \lambda_j^2} \right)^2 - (\eta\delta)^2$ $(\star\star)$

3. $\phi'(\beta) = \sum_{j=1}^{m} \frac{-2\sigma_j^2(\hat{b}_j\lambda_j^2 - \sigma_j(\hat{x_k})_j\lambda_j^2)^2}{(\beta\sigma_j^2 + \lambda_j^2)^3}$ $(\star\star\star)$

4. $\phi''(\beta) = \sum_{j=1}^{m} \frac{6\sigma_j^4(\hat{b}_j\lambda_j^2 - \sigma_j(\hat{x_k})_j\lambda_j^2)^2}{(\beta\sigma_j^2 + \lambda_j^2)^3}$ $(\star\star\star\star)$

Problem with ill-conditioned matrices

Algorithm 1 (Iterated Tikhonov with GSVD)

Input: Measurement matrix A, regularization parameter L and data b.
Output: Approximate solution $x_k \approx x$

0. Calculate the GSVD of the pair (A,L), $A = U\Sigma Y^T$, $L = V\Lambda Y^T$

1. Initialize $\mu = \dfrac{1}{\beta}$ (in general $\beta = 0$), $x_0 = 0$.

Let $Z_k = Y^T X_k$ and $\hat{b} = U^T b$

for k=1, 2, .. until stopping criteria do:

2. Calculate μ_k to satisfy the Discrepancy Principle using the function $\phi(\beta)$

3. Update $Z_{k+1} = (\Sigma^T\Sigma + \dfrac{1}{\beta}\Lambda^T\Lambda)^{-1}(\Sigma^T\hat{b} + \dfrac{1}{\beta}\Lambda^T\Lambda Z_k)$

end

Comparing Tikhonov with Iterated Tikhonov.

Table: Relative error (1% noise added)

.	Tikhonov GSVD	ITGSVD
shaw(100)	0.2503	0.1002
baart(100)	0.0905	0.0356
deriv2(100,2)	0.0202	0.0152

Comparing!

Comparing our new algorithm with a previous algorithm [A. Buccini, M.Donatelli, L.Reichel] where they use a sequence of regularization parameters which satisfy the condition:

$$\sum_{k=0}^{\infty} \alpha_k^{-1} = \frac{1}{\alpha_0} \sum_{k=0}^{\infty} q_k^{-1} = \infty$$

Note 1: Their method needs a good approximation of the parameter q, and a good choice of α_0 which will depend on the matrix L .
Note 2: They use q=0.8 and $\alpha_0 = 10^6$

Comparing GIT and ITGSVD with baart(100) and shaw(100)

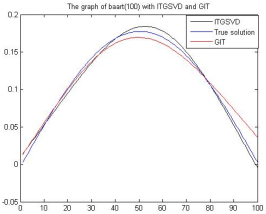


Figure: a. Example of baart(100)

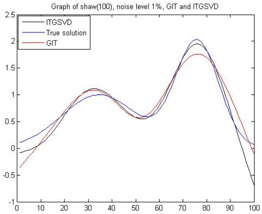


Figure: b. Example of shaw(100)

Table: Relative error (1% noise added)

.	GIT	ITGSVD
shaw(100)	0.0453	0.0815.
baart(100)	0.1663	0.1750

Zero finders

We consider 4 zero finders to compare the results. Why? Finding the regularization parameter in the best way is our GOAL.

$$\boxed{\textit{Good regularization parameter}} \Longrightarrow \boxed{\textit{Good approximate solution}}$$

- Since we use iterative methods to find β from $\phi(\beta) = 0$, the approximated solution depends on the method too.
- We will consider the iterative methods:
  1. **Bisection method** (It will find the solution, but not clear which should be the interval that we look for β and the convergence is very slow)
  2. **Newton method** (The convergence is he fast, quadratic?, but it may find solutions which are not accepted ($\beta < 0$) and the regularization is not valid.)
  3. **New convergence method** (L. Reichel, A. Shyshkov) which yields cubic convergence.
  4. **Newton method applied to** $1/\Phi(\beta)$

Comparing the zero finders

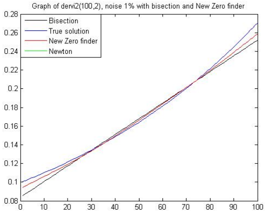


Figure: a. Example of baart(100)

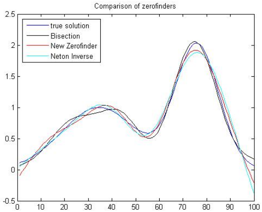


Figure: b. Example of shaw(100)

Table: Relative error shaw(100) (1% noise added)

.	Relative error	Time(s)
Bisection	0.0504	0.098
Newton	0.1663	0.073
Newton Inverse	0.1763	0.089
New zero	0.0486	0.0413

Advantages and disadvantages of IT-GSVD

Advantages:

- It uses the information of the matrices A and L to approximate the solution in a better way.
- It evaluates the regularization parameter without any prior knowledge.
- It is a relatively good method for small dimension problems.
- Compared to GIT there is no need to have prior knowledge of the 2 parameters that this method use.

Disadvantages:

- Might be a problem using this method if there are very high dimension matrices since the computation of the GSVD of the pair (A,L) is expensive.

References

- [1] Per Christian Hansen *Rank- Deficient and Discrete Ill-posed Problems*. SAIM, Philadelphia.
- [2] Alessandro Buccini, Marco Donatelli, Lothar Reichel. *Iterated Tikhonov regularization with general penalty term*.
- [3] Gaungxin Huang, Lothar Reichel, Feng Yin *On the choice of the solution subspace for nonstationary iterated Tikhonov regularization*
- [4] Gaungxin Huang, Lothar Reichel, Feng Yin *Projected Nonstationary Iterated Tikhonov Regularization*
- [5] Lothar Reichel Andriy Shyshkov *A new zero-finder for Tikhonov-Regularization*
- [6] https://www2.math.ethz.ch/education/bachelor/seminars/hs2010/ipip/s

Questions?

Questions?

Thank you!

Thank you!