Convergence analysis of the GKB-GCV algorithm

by

Dai Togashi
Takashi Nodera

Dai Togashi
School of Fundamental Science and Technology
Keio University

Takashi Nodera
Department of Mathematics
Keio University
Convergence analysis of the GKB-GCV algorithm

Dai TOGASHI*   Takashi NODERA†

March 27, 2017

Abstract

This paper explores the application of the generic Tikhonov regularization used to stabilize large scale ill-posed problems in deblurring, hyper resolution and other applicable situations. Recently, a new solver for the generic Tikhonov regularization, called the GKB-GCV method was proposed by D. Togashi et al. [GSTF JMSR, Vol. 3, No. 2, pp. 53–58]. This paper, analyzes the convergence properties of the GKB-GCV method.

Key Words. ill-posed problem, Tikhonov regularization, GKB-GCV

AMS(MOS) subject classifications. 65F22

1 Introduction

The stable approximate solution for a large scale ill-posed problem of the form:

\[ x_{LS} = \arg \min_{x \in \mathbb{R}^n} \| b - Ax \|_2^2, \]

(1)

is computed, where matrix \( A \in \mathbb{R}^{m \times n}, m \geq n \), is ill-conditioned. The right-hand vector \( b \in \mathbb{R}^m \) contains the following error:

\[ b = Ax_{\text{exact}} + \epsilon, \]

(2)

where \( x_{\text{exact}} \in \mathbb{R}^n \) is the exact solution, and \( \epsilon \in \mathbb{R}^m \) is the unknown noise. A matrix of this form sometimes comes from image resolutions, e.g. image deblurring or hyper resolution. Because matrix \( A \) is ill-conditioned, \( x_{LS} \) is dependent on noise. The Tikhonov regularization [7] constructs stable approximations of \( x_{\text{exact}} \) by solving the least squares problem of the form:

\[ x_\lambda = \arg \min_{x \in \mathbb{R}^n} \{ \| b - Ax \|_2^2 + \lambda \| Lx \|_2^2 \}, \]

(3)

*School of Fundamental Science and Technology, Graduate School of Science and Technology, Keio University, JAPAN. mail: dai-togashi@keio.jp
†Department of Mathematics, Faculty of Science and Technology, Keio University, JAPAN. mail: nodera@math.keio.ac.jp
where \( L \in \mathbb{R}^{p \times n} \) is the regularization matrix, and \( \lambda > 0 \) is the regularization parameter. The standard form of the Tikhonov regularization is when \( L = I_n \), where \( I_n \) is the \( n \times n \) identity matrix. The general form of the Tikhonov regularization is when \( L \neq I_n \). When the common space between the null spaces of \( A \) and \( L \) is the zero space, the regularization problem (3) has a unique solution. To obtain a good approximate solution for (3), an appropriate regularization parameter is required. There are many methods for determining the regularization parameter without identifying the norm of the noise: \( \| \epsilon \|_2 \), [1, 5].

For \( L = I_n \), there are two hybrid methods, called GKB-FP [2] and W-GCV [4]. These methods do not require identifying the norm \( \| \epsilon \|_2 \), and contain a projection over the Krylov subspace generated by the Golub-Kahan Bidiagonalization (GKB) method. The difference between these two methods is in the approach, i.e. in terms of determining the regularization parameter. The GKB-FP uses the FP scheme, whereas the W-GCV uses the weighed GCV. Bazán et al. [3] proposed an approach without identifying norm \( \| \epsilon \| \), which is created by the extension of the GKB-FP method. In a recent study, the W-GCV method was extended to a general form of the Tikhonov regularization. This was called the GKB-GCV method [8].

This paper explores the uses of the GKB-GCV method which is a solver for a large scale general form of the Tikhonov regularization. A convergence property of the GKB-GCV is described succinctly. The conclusions are summed-up in Section 4.

2 The GKB-GCV method

The GKB-GCV is one of the algorithms for a general form of the Tikhonov regularization, which is based on the GKB and GCV. When \( k < n \) GKB steps are applied to matrix \( A \) with the initial vector \( b/\| b \|_2 \), it results in two matrices \( Y_{k+1} = [y_1, \ldots, y_{k+1}] \in \mathbb{R}^{m \times (k+1)} \) and \( W_k = [w_1, \ldots, w_k] \in \mathbb{R}^{n \times k} \) with orthonormal columns, and a lower bidiagonal matrix as follows:

\[
B_k = \begin{pmatrix}
\alpha_1 & \beta_2 & \alpha_3 & \cdots & \alpha_k \\
& \beta_3 & \ddots & \alpha_{k+1} \\
& & \ddots & \beta_{k+1}
\end{pmatrix} \in \mathbb{R}^{(k+1) \times k},
\]

such that,

\[
\beta_1 Y_{k+1} e_1 = b = \beta_1 y_1,
\]
\[
AW_k = Y_{k+1} B_k,
\]
\[
A^T Y_{k+1} = W_k B_k^T + \alpha_{k+1} w_{k+1} e_{k+1}^T,
\]

where \( e_i \) denotes the \( i \)-th unit vector in \( \mathbb{R}^{k+1} \). Columns of \( W_k \) are the orthonormal basis for the generalized Krylov subspace \( \mathcal{K}_k(A^T A, A^T b) \). The general form of regularization
over the generated Krylov subspace is as follows:

\[ x^{(k)}_\lambda = \arg\min_{x \in X_k(\mathbf{A}^T \mathbf{A}, \mathbf{A}^T \mathbf{b})} \{ \| \mathbf{A} x - \mathbf{b} \|_2^2 + \lambda \| L x \|_2^2 \}. \] (4)

Since the columns of \( W_k \) are the orthonormal basis for the generated Krylov subspace, equation (4) is rewritten as follows:

\[ x^{(k)}_\lambda = W_k y^{(k)}_\lambda, \quad y^{(k)}_\lambda = \arg\min_{y \in \mathbb{R}^k} \{ \| \mathbf{B}_k y - \beta_1 e_1 \|_2^2 + \lambda \| LW_k y \|_2^2 \}. \] (5)

GKB-GCV uses the same reduction to PROJ-L when solving the general form of the Tikhonov regularization of Bazán [3]. By using the reduction QR factorization for matrix products \( LW_k \), equation (5) is rewritten as follows:

\[ y^{(k)}_\lambda = \arg\min_{y \in \mathbb{R}^k} \{ \| \mathbf{B}_k y - \beta_1 e_1 \|_2^2 + \lambda \| L R_k y \|_2^2 \}. \] (6)

where \( Q_k R_k = LW_k \) and \( Q_k \) has orthogonal columns. To increase \( k \), the QR factorization can be updated computing \( k + 1 \) elements by using the summation and a product of the vectors. This reduction technique is a good choice for large scale problems, because this approach reduces the size of the least squares problem: \((m + p) \times n \to (2k + 1) \times k\).

The GCV determines the regularization parameter for equation (3) by searching for the minimum point of function as follows:

\[ G(\lambda) = \frac{\| (\mathbf{I}_m - \mathbf{A} A_{\lambda,L}^+ \mathbf{b}) \|_2^2}{(\text{trace}(\mathbf{I}_m - \mathbf{A} A_{\lambda,L}^+ \mathbf{b}))^2} \] (7)

where \( A_{\lambda,L}^+ = (\mathbf{A}^T \mathbf{A} + \lambda \mathbf{L}^T \mathbf{L})^{-1} \mathbf{A}^T \). Using the GSVD for the matrix pair \((\mathbf{A}, \mathbf{L})\), equation (7) is written as follows:

\[ G(\lambda) = \frac{\sum_{i=1}^{n} \left( \frac{\sigma_i^2 \mathbf{u}_i^T \mathbf{b}}{\sigma_i + \sigma_i^2} \right)^2 + \sum_{i=n+1}^{m} (\mathbf{u}_i^T \mathbf{b})^2}{(m - \sum_{i=1}^{n} \frac{\sigma_i^2}{\sigma_i + \sigma_i^2})^2}. \] (8)

where \( \mathbf{A} = \mathbf{U} \mathbf{S} \mathbf{Z}^{-1}, \mathbf{L} = \mathbf{V} \mathbf{C} \mathbf{Z}^{-1} \). At the \( k \) step, the GKB-GCV uses the same approach as AT-GCV [6] for determining \( \lambda \). The regularization parameter \( \lambda \) is chosen to minimize the following function:

\[ G_k(\lambda) = \frac{\| (\mathbf{I}_m - \mathbf{A} W_k (B_k)_{\lambda,R_k} Y_{k+1}^T) \mathbf{b} \|_2^2}{(\text{trace}(\mathbf{I}_m - \mathbf{A} W_k (B_k)_{\lambda,R_k} Y_{k+1}^T))^2} \]

\[ = \frac{\beta_1^2 \left( \sum_{i=1}^{k} \left( \frac{\lambda_k \mathbf{u}_{i(k)}^T e_1}{\sigma_{i(k)} + \lambda_k} \right)^2 + (\mathbf{u}_{k+1(k)}^T e_1)^2 \right)}{(m - \sum_{i=1}^{k} \frac{\sigma_{i(k)}^2}{\sigma_{i(k)} + \lambda_k})^2}. \]

where \( B_k = \mathbf{U}_k S_k \mathbf{Z}_k^{-1}, R_k = \mathbf{V}_k \mathbf{C}_k \mathbf{Z}_k^{-1} \) by using the reduction GSVD(\( B_k, R_k \)).

The GKB-GCV method is compactly summarized in Algorithm 1.
Algorithm 1 GKB-GCV Method

Require: $A, b, L, \text{tol}$

Ensure: Regularized solution $x_{\lambda^*}^{(k)}$

1. Apply the GKB step to $A$ with starting vector $b$ at $k = 0$ and set $k = 1$.
2. Perform one more GKB step and update the QR factorization of $LW_k$.
   
   $$LW_k = Q_k R_k.$$

3. Compute GSVD$(B_k, R_k)$. 
   
   $$B_k = U_k S_k Z^{-1}, R_k = V_k C_k Z^{-1}.$$  

4. Compute the minimized point $\lambda_k$ of $G_k(\lambda)$.
5. If the stopping criteria is satisfied do

   $$\lambda^* = \lambda_k.$$

else do

   $k \leftarrow k + 1$
   Go to step 2.

end if

6. Solve subproblem $y_{\lambda^*}^{(k)}$.
7. Compute the regularized solution $x_{\lambda^*}^{(k)}$.

3 Convergence property

It will be assumed that when $\lambda_k = \arg\min G_k(\lambda)$, $\lambda_k$ will not have a monotone convergence. Therefore, the following theorem must be proven to analyze the convergence property of the GKB-GCV method.

Theorem 3.1 Assume $\text{rank}(A) = q < m$ and $Ax_{\text{exact}} \neq 0$. Whenever the noise’s norm converges to 0, the regularization parameter determined by GKB-GCV method converges to 0 at the $q$ step:

$$||\epsilon||_2 \to 0 \Rightarrow (\lambda_q \to 0).$$

Proof: 1 At the $q$ step, the GKB method generates these matrices:

$$B_q = \begin{bmatrix}
\alpha_1 \\
\beta_2 \\
\alpha_2 \\
\beta_3 \\
\ddots \\
\alpha_{q-1} \\
\beta_q \\
\alpha_q \\
0
\end{bmatrix} = \begin{bmatrix}
\hat{B}_q \\
0
\end{bmatrix} \text{ with } \alpha_q \neq 0,$$

$$b = \beta_1 Y_{q+1} e_1 = A x_{\text{exact}} + \epsilon,$$

$$A = Y_{q+1} B_q W_q^T = \hat{Y}_q \hat{B}_q W_q^T.$$

The reduction GSVD for matrix pair $(\hat{B}_q, R_q)$ is considered, and then the diagonal matrix $D_q$ is defined as follows:

$$[D_q]_{i,i} = \frac{[\hat{C}_q]_{i,i}^2}{[\hat{S}_q]_{i,i}^2 + \lambda [\hat{C}_q]_{i,i}^2} \text{ with } \hat{B}_q = \hat{U}_q \hat{S}_q \hat{X}_q, R_q = \hat{V}_q \hat{C}_q \hat{X}_q.$$
From the triangle inequality, it follows that:

\[ \| r_{q,\lambda} \|_2 = \| (I_m - Y_q \hat{B}_q^T \hat{B}_q + \lambda R_q^T R_q)^{-1} \hat{B}_q^T Y_q^T \|_2, \]
\[ = \| (I_m - Y_q \hat{U}_q \hat{S}_q^2 (\hat{S}_q^2 + \lambda \hat{C}_q^2)^{-1} \hat{U}_q^T Y_q) \|_2, \]
\[ \leq \| (I_m - Y_q \hat{U}_q \hat{S}_q^2 (\hat{S}_q^2 + \lambda \hat{C}_q^2)^{-1} \hat{U}_q^T Y_q) A x_{\text{exa}} \|_2, \]
\[ + \| (I_m - Y_q \hat{U}_q \hat{S}_q^2 (\hat{S}_q^2 + \lambda \hat{C}_q^2)^{-1} \hat{U}_q^T Y_q) \|_2. \]

Since \( Y_q \hat{U}_q \hat{U}_q^T Y_q^T \) is an orthogonal projection from \( Y_q^T Y_q = \hat{U}_q^T \hat{U}_q = I_q \), the following inequality is approved:

\[ \| (I_m - Y_q \hat{U}_q \hat{S}_q^2 (\hat{S}_q^2 + \lambda \hat{C}_q^2)^{-1} \hat{U}_q^T Y_q) e \|_2 \]
\[ \leq \| (I_q - \hat{S}_q^2 (\hat{S}_q^2 + \lambda \hat{C}_q^2)^{-1} \hat{U}_q^T Y_q) e \|_2 + \| e \|_2, \]
\[ = \| \lambda \hat{C}_q^2 (\hat{S}_q^2 + \lambda \hat{C}_q^2)^{-1} \hat{U}_q^T Y_q e \|_2 + \| e \|_2, \]
\[ = \| \lambda D_q \hat{U}_q^T Y_q e \|_2 + \| e \|_2, \]
\[ \leq 2 \| e \|_2. \]

The last inequality comes from \( \| \lambda D_q f \|_2 \leq \| f \|_2 \). This follows for all \( \lambda \) and \( f \) from the definition of the \( D_q \).

From \( A = Y_q \hat{B}_q W_q^T \), and \( Y_q^T Y_q = \hat{U}_q^T \hat{U}_q = I_q \),

\[ \| (I_m - Y_q \hat{U}_q \hat{S}_q^2 (\hat{S}_q^2 + \lambda \hat{C}_q^2)^{-1} \hat{U}_q^T Y_q) A x_{\text{exa}} \|_2, \]
\[ = \| Y_q \hat{U}_q (I_q - \hat{S}_q^2 (\hat{S}_q^2 + \lambda \hat{C}_q^2)^{-1} \hat{S}_q \hat{X}_q W_q^T x_{\text{exa}}) \|_2, \]
\[ = \| \lambda D_q \hat{U}_q^T Y_q \hat{S}_q \hat{X}_q W_q^T x_{\text{exa}} \|_2, \]
\[ = \lambda \| D_q \hat{U}_q^T Y_q \|_2. \]

Note that since \( \| D_q \hat{U}_q^T Y_q x_{\text{exa}} \|_2 > 0 \) from the assumption and the definition of the matrices, it has a minimum value \( 0 \) at \( \lambda = 0 \). Therefore, from \( G_k(\lambda) \geq 0 \) for all \( k \) and trace\( (A^T A^+) A^T) = m - q > 0 \),

\[ (\| e \|_2 \to 0) \Rightarrow (\| r_{q,0} \|_2 \leq 2 \| e \|_2 \to 0) \Rightarrow (\lambda_q \to 0). \]

**Theorem 3.2** Assume that \( \text{rank}(A) = q < m \), \( x_{\text{exact}} \in (\text{Ker}(A))^\perp \). Then, if the \( \| e \| \to 0 \), GKB-GCV method converges to a true solution at most \( q \) iterations. That is:

\[ (\| e \|_2 \to 0) \Rightarrow (x_{q,\lambda_q} \to x_{\text{exa}}). \]

**Proof:** At the \( q \) step, from the triangle inequality:

\[ \| x_{q,\lambda_q} - x_{\text{exa}} \|_2 \leq \| (W_q (\hat{B}_q^T \hat{B}_q + \lambda_q R_q^T R_q)^{-1} \hat{B}_q^T Y_q^T A - I_q) x_{\text{exa}} \|_2, \]
\[ + \| W_q (\hat{B}_q^T \hat{B}_q + \lambda_q R_q^T R_q)^{-1} \hat{B}_q^T Y_q^T \|_2. \]
From $A = Y_q \hat{B}_q W_q^T$, and $W_q W_q^T x_{\text{exa}} = x_{\text{exa}}$, the first term in the inequality rewrites:

$$
\| (W_q (\hat{B}_q^T \hat{B}_q + \lambda_q R_q^T R_q) - I_n) x_{\text{exa}} \|_2,
$$

$$
= \| (W_q (\hat{B}_q^T \hat{B}_q + \lambda_q R_q^T R_q) - I_n) x_{\text{exa}} \|_2,
$$

$$
= \| (W_q \tilde{X}_q^{-1} (\tilde{S}_q^2 + \lambda_q \tilde{C}_q^2) - I_q) \tilde{X}_q W_q^T x_{\text{exa}} \|_2,
$$

$$
= \| - \lambda_q \tilde{X}_q^{-1} D_q \tilde{X}_q W_q^T x_{\text{exa}} \|_2,
$$

$$
= \lambda_q \| \tilde{X}_q^{-1} D_q \tilde{X}_q W_q^T x_{\text{exa}} \|_2.
$$

From Theorem 3.1: $(\| \epsilon \|_2 \to 0) \Rightarrow (\lambda_q \to 0)$. So,

$$
\lambda_q \| \tilde{X}_q^{-1} D_q \tilde{X}_q W_q^T x_{\text{exa}} \|_2 \to 0 \quad (\| \epsilon \|_2 \to 0)
$$

From the property of the norm, it follows that:

$$
\| W_q (\hat{B}_q^T \hat{B}_q + \lambda_q R_q^T R_q)^{-1} \hat{B}_q Y_q^T \epsilon \|_2,
$$

$$
= \| W_q \tilde{X}_q^{-1} (\tilde{S}_q^2 + \lambda_q \tilde{C}_q^2)^{-1} \tilde{S}_q U_q Y_q^T \epsilon \|_2,
$$

$$
= \| \hat{B}_q^{-1} \hat{U}_q (\tilde{S}_q^2 + \lambda_q \tilde{C}_q^2)^{-1} \tilde{S}_q^2 U_q Y_q^T \epsilon \|_2,
$$

$$
\leq \| \hat{B}_q^{-1} \|_2 \cdot \| (\tilde{S}_q^2 + \lambda_q \tilde{C}_q^2)^{-1} \tilde{S}_q^2 \|_2 \cdot \| \epsilon \|_2 \to 0.
$$

In the last limits, this is used since $\hat{B}_q$ is a full rank and in Theorem 3.1: $(\| \tilde{S}_q^2 + \lambda \tilde{C}_q^2 \|_2 \to 1$. From the above,

$$
\| x_{\text{exa}} - x_{\text{exa}} \|_2 \leq \lambda_q \| \tilde{X}_q^{-1} D_q \tilde{X}_q W_q^T x_{\text{exa}} \|_2 + \| \hat{B}_q^{-1} \|_2 \cdot \| \epsilon \|_2,
$$

$$
\to 0 \quad (\| \epsilon \|_2 \to 0).
$$

4 Conclusion

The GKB-GCV algorithm was explored and the convergence property of the GKB-GCV algorithm was analyzed. As a result, it was proven that when rank($A$) is less than $m$, the norm of the noise converged to 0 and the true solution is orthogonal to the kernel of $A$, then the GKB-GCV algorithm converges to produce the true solution at most matrix size iterations. The results show that if the GKB-GCV is applied to well-posed problems, the computed solution converges to a true solution.

References


Department of Mathematics
Faculty of Science and Technology
Keio University

Research Report

2016

[16/001] Shiro Ishikawa,
*Linguistic interpretation of quantum mechanics: Quantum Language [Ver. 2]*,
KSTS/RR-16/001, January 8, 2016

[16/002] Yuka Hashimoto, Takashi Nodera,
*Inexact shift-invert Arnoldi method for evolution equations*,
KSTS/RR-16/002, May 6, 2016

[16/003] Yuka Hashimoto, Takashi Nodera,
*A Note on Inexact Rational Krylov Method for Evolution Equations*,
KSTS/RR-16/003, November 9, 2016

[16/004] Sumiyuki Koizumi,
*On the theory of generalized Hilbert transforms (Chapter V: The spectre analysis and synthesis on the N.Wiener class S)*,
KSTS/RR-16/004, November 25, 2016

[16/005] Shiro Ishikawa,
*History of Western Philosophy from the quantum theoretical point of view*,
KSTS/RR-16/005, December 6, 2016

2017

[17/001] Yuka Hashimoto, Takashi Nodera,
*Inexact Shift-invert Rational Krylov Method for Evolution Equations*,
KSTS/RR-17/001, January 27, 2017

[17/002] Dai Togashi, Takashi Nodera,
*Convergence analysis of the GKB-GCV algorithm*,
KSTS/RR-17/002, March 27, 2017