Iterative methods for Image Processing

Lothar Reichel

Como, May 2018.

Lecture 3: Block iterative methods, preconditioning, iterated Tikhonov.

Outline of Lecture 3:

- Block Krylov subspace methods, application to color image restoration
- Preconditining
- Iterated Tikhonov
- The method by Donatelli and Hanke

Color image restoration

Color images are represented by three channels: red, green, and blue. Hyperspectral images have more "colors" and require more channels. See Hansen, Nagy, and O'Leary.

Consider k-channel images. Let $b^{(i)} \in \mathbf{R}^{n^2}$ represent the blur- and noise-contaminated image in channel i, let $e^{(i)} \in \mathbf{R}^{n^2}$ describe the noise.

The contaminated images of all channels $b^{(i)}$ can be represented by

$$b = [(b^{(1)})^T, \ldots, (b^{(k)})^T]^T.$$

The degradation model is of the form

$$b = Hx_{\text{exact}} + e,$$

where

$$H = A_k \otimes A \in \mathbf{R}^{n^2k \times n^2k}$$

with

- $A \in \mathbf{R}^{n^2 \times n^2}$ modelling within-channel blurring,
- $A_k \in \mathbf{R}^{k \times k}$ modelling cross-channel blurring.

Determine approximation of x_{exact} by computing approximate solution of

$$Hx = b.$$

Alternatively, the contaminated images of all channels $b^{(i)}$ can be represented by

$$B = [b^{(1)}, \ldots, b^{(k)}].$$

Define the linear operator

$$\mathcal{A} : \mathbf{R}^{n^2 \times k} \to \mathbf{R}^{n^2 \times k} : \quad \mathcal{A}(X) := AXA_k^T.$$

The degradation model can be written as

$$B = \mathcal{A}(X_{\text{exact}}) + E,$$

where $X_{\text{exact}} = [x_{\text{exact}}^{(1)}, \ldots, x_{\text{exact}}^{(k)}]$.

Let $B_{\text{exact}} = \mathcal{A}(X_{\text{exact}})$. Denote $\mathcal{A}(X)$ by AX.

Tikhonov regularization

Solve the minimization problem

$$\min_X \{\|AX - B\|_F^2 + \mu\|X\|_F^2\},$$

where $\|\cdot\|_F$ denotes the Frobenius norm and $\mu > 0$ is a regularization parameter.

The normal equations, which are obtained by requiring the gradient with respect to X to vanish, are given by

$$(A^T A + \mu I)X = A^T B.$$

They have the unique solution

$$X_\mu = \left(A^T A + \mu I\right)^{-1} A^T B$$

for any $\mu > 0$. The discrepancy principle requires that $\mu > 0$ be determined so that

$$\|B - AX_\mu\|_F = \eta\delta, \qquad \delta = \|B - B_{\text{exact}}\|_F.$$

This is possible for most reasonable B.

Solution methods

- Trival method: Compute approximate solution of each system of equations

$$Ax = b^{(j)}, \quad j = 1, 2, \ldots, k, \tag{1}$$

  independently.

- Apply partial block Golub–Kahan bidiagonalization.
- Apply partial global Golub–Kahan bidiagonalization.
- Compute partial SVD of A and apply to each system (1) independently

Block Golub–Kahan bidiagonalization (BGKB)

Define the QR factorization

$$B = P_1 R_1,$$

where $P_1 \in \mathbf{R}^{n^2 \times k}$ has orthonormal columns, i.e.,

$$P_1^T P_1 = I$$

and $R_1 \in \mathbf{R}^{k \times k}$ is upper triangular.

ℓ steps of the BGKB applied to A with initial block vector P_1 gives the decompositions

$$AQ_\ell^{(k)} = P_{\ell+1}^{(k)} C_{\ell+1,\ell}^{(k)}, \qquad A^T P_\ell^{(k)} = Q_\ell^{(k)} C_{\ell,\ell}^{(k)^T},$$

where

$$\begin{aligned} P_{\ell+1}^{(k)} &= [P_\ell^{(k)}, P_\ell] = [P_1, \ldots, P_{\ell+1}] \in \mathbf{R}^{n^2 \times (\ell+1)k}, \\ Q_\ell^{(k)} &= [Q_1, \ldots, Q_\ell] \in \mathbf{R}^{n^2 \times \ell k} \end{aligned}$$

have orthonormal columns, i.e.,

$$(P_{\ell+1}^{(k)})^T P_{\ell+1}^{(k)} = I, \quad (Q_\ell^{(k)})^T Q_\ell^{(k)} = I.$$

The lower block bidiagonal matrix

$$C_{\ell+1,\ell}^{(k)} := \begin{bmatrix} L_1 & & & \\ R_2 & L_2 & & \\ & \ddots & \ddots & \\ & & R_\ell & L_\ell \\ & & & R_{\ell+1} \end{bmatrix} \in \mathbf{R}^{k(\ell+1)\times k\ell}$$

has lower triagular blocks $L_j \in \mathbf{R}^{k\times k}$ and upper triangular blocks $R_j \in \mathbf{R}^{k\times k}$; $C_{\ell,\ell}^{(k)} \in \mathbf{R}^{k\ell\times k\ell}$ is the leading submatrix of $C_{\ell+1,\ell}^{(k)}$.

Further,

$$\begin{aligned}\mathcal{R}(Q_\ell^{(k)}) &= \mathbf{K}_\ell(A^TA, A^TB) \\ &= \text{span}\{A^TB, (A^TA)A^TB, \ldots (A^TA)^{\ell-1}A^TB\}.\end{aligned}$$

Let $X = Q_\ell^{(k)}Y$ with $Y \in \mathbf{R}^{k\ell \times k\ell}$. Then

$$\begin{aligned}&\min_{X \in \mathbf{K}_\ell(A^TA, A^TB)} \{\|AX - B\|_F^2 + \mu\|X\|_F^2\} \\ &= \min_{Y \in \mathbf{R}^{k\ell \times k\ell}} \{\|AQ_\ell^{(k)}Y - B\|_F^2 + \mu\|Y\|_F^2\} \\ &= \min_{Y \in \mathbf{R}^{k\ell \times k\ell}} \left\{ \left\| C_{\ell+1,\ell}^{(k)} Y - \begin{bmatrix} R_1 \\ 0 \end{bmatrix} \right\|_F^2 + \mu\|Y\|_F^2 \right\}.\end{aligned}$$

Solve by QR factorization of $C_{\ell+1,\ell}^{(k)}$.

Gives Y_μ and $X_\mu = P_\ell^{(k)} Y_\mu$. Determine $\mu > 0$ by discrepancy principle, i.e., so that

$$\|AX_\mu - B\|_F = \left\| C_{\ell+1,\ell}^{(k)} Y_\mu - \begin{bmatrix} R_1 \\ 0 \end{bmatrix} \right\|_F = \eta\delta.$$

Requires that ℓ be sufficiently large and that error in B is reasonable ($< 100\%$). Then the desired $\mu > 0$ is the unique solution of a nonlinear equation determined by the reduced problem.

Global Golub–Kahan bidiagonalization (GGKB)

Define the matrix inner product

$$\langle M, N \rangle = \mathrm{tr}(M^T N), \qquad M, N \in \mathbf{R}^{n^2 \times k}.$$

Then

$$\|M\|_F = \langle M, M \rangle^{1/2}.$$

Application of ℓ steps of GGKB to A with initial block vector B determines the lower bidiagonal matrix

$$C_{\ell+1,\ell} = \begin{bmatrix} \rho_1 & & & & \\ \sigma_2 & \rho_2 & & & \\ & \ddots & \ddots & & \\ & & \sigma_{\ell-1} & \rho_{\ell-1} & \\ & & & \sigma_\ell & \rho_\ell \\ & & & & \sigma_{\ell+1} \end{bmatrix} \in \mathbf{R}^{(\ell+1)\times\ell}$$

and the matrices

$$\begin{aligned} U_{\ell+1}^{(k)} &= [U_1, U_2, \ldots, U_{\ell+1}] \in \mathbf{R}^{n^2\times(\ell+1)k}, \\ V_\ell^{(k)} &= [V_1, V_2, \ldots, V_\ell] \in \mathbf{R}^{n^2\times \ell k}, \end{aligned}$$

where $U_i, V_j \in \mathbf{R}^{n^2 \times k}$, $U_1 = B/\|B\|_F$, and

$$\langle U_i, U_j \rangle = \langle V_i, V_j \rangle = \begin{cases} 1 & i = j, \\ 0 & i \neq j. \end{cases}$$

Let $C_{\ell,\ell}$ be the leading $\ell \times \ell$ submatrix of $C_{\ell+1,\ell}$. If ℓ is small enough so that no breakdown occurs, then

$$\begin{aligned} A\big[V_1, V_2, \ldots, V_\ell\big] &= U_{\ell+1}^{(k)}(C_{\ell+1,\ell} \otimes I_k), \\ A^T\big[U_1, U_2, \ldots, U_\ell\big] &= V_\ell^{(k)}(C_{\ell,\ell}^T \otimes I_k). \end{aligned}$$

Recall that $A[V_1, V_2, \ldots, V_\ell]$ stands for $[\mathcal{A}(V_1), \mathcal{A}(V_2), \ldots, \mathcal{A}(V_\ell)]$; similarly for $A^T U_j$.

Determine an approximate solution of the form

$$X = V_\ell^{(k)}(y \otimes I_k), \quad y \in \mathbf{R}^\ell,$$

of the Tikhonov minimization problem

$$\min_{X=V_\ell^{(k)}(y\otimes I_k)} \{\|AX - B\|_F^2 + \mu\|X\|_F^2\}$$
$$= \min_{y\in\mathbf{R}^\ell}\{\|C_{\ell+1,\ell}y - e_1\|B\|_F^2\|_2^2 + \mu\|y\|_2^2\}$$

Denote the solution by y_{μ_ℓ}. Choose $\mu = \mu_\ell > 0$ so that y_{μ_ℓ} and therefore $X_{\mu_\ell} = V_\ell^{(k)}(y_{\mu_\ell} \otimes I_k)$ satisfy the discrepancy principle

$$\|AX_{\mu_\ell} - B\|_F = \|C_{\ell+1,\ell}y_{\mu_\ell} - e_1\|B\|_F^2\|_2 = \eta\delta.$$

Standard Golub–Kahan bidiagonalization for multiple right-hand sides

The largest singular triplets of A can be approximated well by carrying out a few GKB steps. This suggests the solution method:

- Apply ℓ bidiagonalization steps to A with initial vector $b^{(1)}$. Gives decompositions

$$AV_\ell = U_{\ell+1}C_{\ell+1,\ell}, \qquad A^T U_\ell = V_\ell C_{\ell,\ell}^T,$$

  with $V_\ell \in \mathbf{R}^{n^2\times\ell}$, $U_{\ell+1} \in \mathbf{R}^{n^2\times(\ell+1)}$ such that $V_\ell^T V_\ell = I$, $U_{\ell+1}^T U_{\ell+1} = I$, and $U_\ell e_1 = b/\|b\|_2$. Moreover, $C_{\ell+1,\ell} \in \mathbf{R}^{(\ell+1)\times\ell}$ lower bidiagonal.

- Then

$$\min_{x\in\mathcal{R}(V_\ell)}\{\|Ax-b^{(1)}\|_2^2+\mu\|x\|_2^2\}$$
$$=\min_{y\in\mathbf{R}^\ell}\{\|C_{\ell+1,\ell}y-U_{\ell+1}^Tb^{(1)}\|_2^2+\mu\|y\|_2^2\}.$$

Determine $\mu>0$ so that the solution y_μ satisfies the discrepancy principle

$$\|C_{\ell+1,\ell}y_\mu-U_{\ell+1}^Tb^{(1)}\|_2=\eta\delta^{(1)},$$

where $\delta^{(1)}$ is a bound for the error in $b^{(1)}$.

- Solve

$$\min_{x\in\mathcal{R}(V_\ell)} \{\|Ax - b^{(2)}\|_2^2 + \mu\|x\|_2^2\}$$
$$= \min_{y\in\mathbf{R}^\ell} \{\|C_{\ell+1,\ell}y - U_{\ell+1}^T b^{(2)}\|_2^2 + \mu\|y\|_2^2\}.$$

If discrepancy principle cannot be satisfied, then increase ℓ.

- Compute approximate solutions of

$$Ax = b^{(j)}, \qquad j = 3, 4, \ldots, k,$$

similarly.

Computations require the columns of $U_{\ell+1}$ to be numerically orthonormal to be able to accurately compute the Fourier coefficients

$$U_{\ell+1}^T b^{(j)}, \qquad j = 2, 3, \ldots, k.$$

Example: Let matrix $A \in \mathbf{R}^{70^2 \times 70^2}$ be determined by the function phillips in Regularization Tools by Hansen. The matrix is a discretization of a Fredholm integral equation of the first kind that describes a convolution on the interval $-6 \leq t \leq 6$. Generate 10 right-hand sides that model smooth functions. Add noise of same noise level to each right-hand side.

Noise level	Method	MVP	Relative error	CPU time (sec)
10^{-3}	BGKB	100	1.46×10^{-2}	0.30
	GGKB	200	1.31×10^{-2}	0.43
	1 GKB	16	2.28×10^{-2}	0.31
	10 GKBs	162	1.43×10^{-2}	2.08
10^{-2}	BGKB	80	2.54×10^{-2}	0.24
	GGKB	120	2.61×10^{-2}	0.30
	1 GKB	10	2.52×10^{-2}	0.19
	10 GKBs	140	2.60×10^{-2}	1.32

Example: Restoration of a 3-channel RGB color image that has been contaminated by blur and noise. The corrupted image is stored in a block vector B with three columns (one for each channel).

Original image (left), blurred and noisy image (right).

Restored image by BGKB (left), restored image by GGKB (right).

Noise level	Method	MVP	Relative error	CPU-time (sec)
10^{-3}	BGKB	492	6.93×10^{-2}	3.86
	GGKB	558	6.85×10^{-2}	3.95
	1 GKB	112	2.64×10^{-1}	1.66
	3 GKBs	632	1.29×10^{-1}	6.55
10^{-2}	BGKB	144	9.50×10^{-2}	1.13
	GGKB	156	9.44×10^{-2}	1.12
	1 GKB	20	2.91×10^{-1}	0.32
	3 GKBs	112	1.58×10^{-1}	1.10

Example: We restore an image that has been contaminated by noise, within-channel blur, and cross-channel blur. Same within-channel blur as above. The cross-channel blur is defined by the cross-channel blur matrix

$$A_3 = \begin{bmatrix} 0.7 & 0.2 & 0.1 \\ 0.25 & 0.5 & 0.25 \\ 0.15 & 0.1 & 0.75 \end{bmatrix}$$

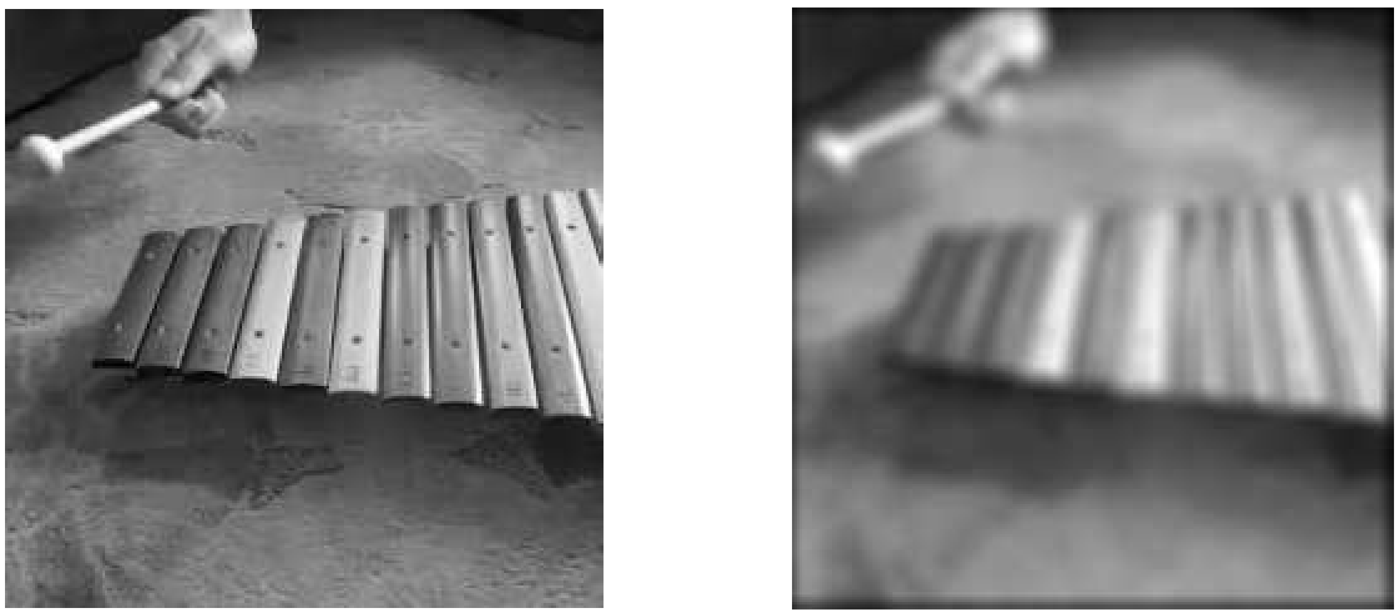
More details in book by Hansen, Nagy, and O'Leary.

Example: Cross-channel blurred and noisy image (left), restored image by GGKB (right).

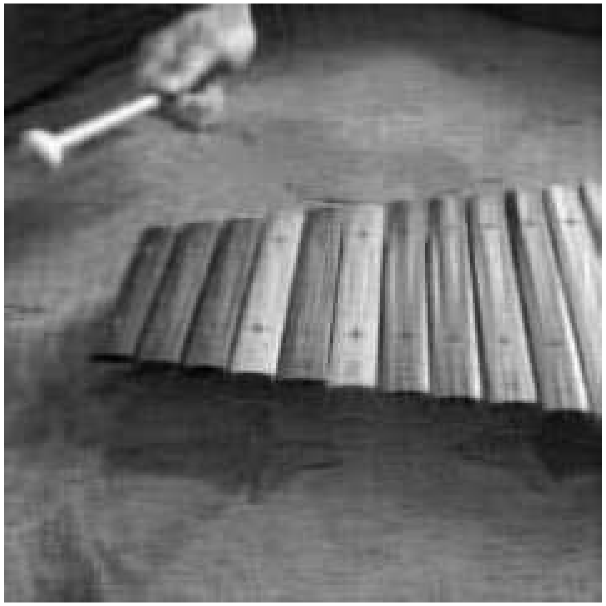
Noise level	Method	MVP	Relative error	CPU-time (sec)
10^{-3}	BGKB	354	7.56×10^{-2}	2.74
	GGKB	702	6.97×10^{-2}	4.99
	1 GKB	112	2.64×10^{-1}	1.63
	3 GKBs	556	1.35×10^{-1}	5.77

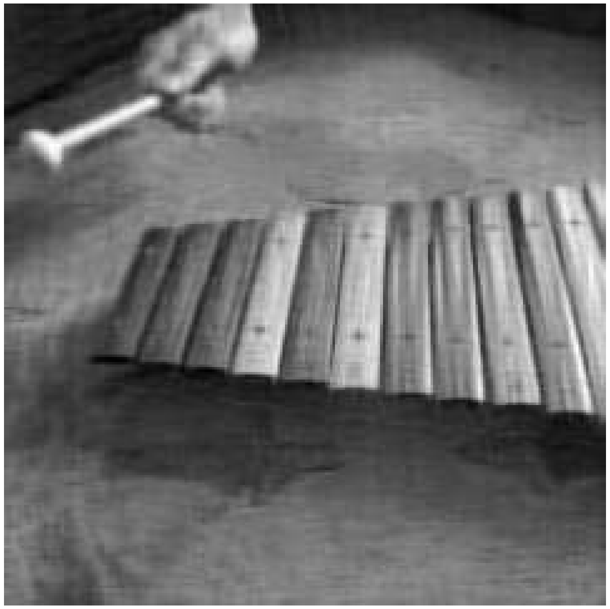
Example: Restoration of a video (from MATLAB). We have 6 frames with 240×240 pixels each.

Frame no. 3: Original frame (left), blurred and noisy frame (right).



Frame no. 3: Restored frame by BGKB (left), and restored frame by GGKB (right).





Noise level	Method	MVP	Relative error	CPU-time (sec)
10^{-3}	BGKB	1152	5.76×10^{-2}	8.72
	GGKB	1188	5.66×10^{-2}	6.23
	1 GKB	130	1.19×10^{-1}	1.69
	6 GKBs	1190	5.67×10^{-2}	10.79
10^{-2}	BGKB	264	9.48×10^{-2}	1.65
	GGKB	228	9.53×10^{-2}	1.21
	1 GKB	34	1.40×10^{-1}	0.44
	6 GKBs	250	9.48×10^{-2}	2.22

The global Arnoldi method

Compute approximate solution of

$$\min_{X \in \mathbf{R}^{m\times n}} \|G - \sum_{i=1}^{p} A_i\, X\, B_i\|_F,$$

At least one of the matrices $A_i \in \mathbf{R}^{m\times m}$ and $B_i \in \mathbf{R}^{n\times n}$ of each pair (A_i, B_i) is large and of ill-determined rank.

The matrix $G \in \mathbf{R}^{m\times n}$ represents available error-contaminated data, such as a blurred and noise-contaminated image.

Tikhonov regularization:

$$\min_{X \in \mathbf{R}^{m \times n}} \left\{ \left\| \sum_{i=1}^{p} A_i \, X \, B_i - G \right\|_F^2 + \mu \left\| \sum_{j=1}^{q} L_j^{(1)} \, X \, L_j^{(2)} \right\|_F^2 \right\},$$

where $L_j^{(1)} \in \mathbf{R}^{s \times m}$ and $L_j^{(2)} \in \mathbf{R}^{n \times t}$ are regularization matrices and $\mu > 0$ is a regularization parameter.

Let $g = \mathrm{vec}(G) \in \mathbf{R}^{mn}$ and $x = \mathrm{vec}(X) \in \mathbf{R}^{mn}$. Define

$$K = \sum_{i=1}^{p} B_i^T \otimes A_i, \quad L = \sum_{j=1}^{q} (L_j^{(2)})^T \otimes L_j^{(1)}.$$

with $\otimes$ denoting the Kronecker product. For matrices $C \in \mathbf{R}^{m \times m}$ and $D \in \mathbf{R}^{n \times n}$, we have

$$C \otimes D = [c_{ij} D] \in \mathbf{R}^{mn \times mn}.$$

Then the Tikhonov minimization problem can be written in the form

$$\min_{x \in \mathbf{R}^{mn}} \left\{ \|Kx - g\|_2^2 + \mu \, \|Lx\|_2^2 \right\}.$$

Define operator:

$$\mathcal{A} : \mathbf{R}^{m\times n} \longrightarrow \mathbf{R}^{m\times n} : \quad X \longrightarrow \mathcal{A}(X) = \sum_{i=1}^{p} A_i\, X\, B_i.$$

k steps of the global Arnoldi method applied to $\mathcal{A}$ with initial matrix G determines the decomposition

$$[\mathcal{A}(V_1), \ldots, \mathcal{A}(V_k)] = \mathcal{V}_{k+1}\, (H_{k+1,k} \otimes I_n),$$

where $H_{k+1,k} = [h_{i,j}] \in \mathbf{R}^{(k+1)\times k}$ is upper Hessenberg, $\mathcal{V}_{k+1} = [V_1, V_2, \ldots, V_{k+1}] \in \mathbf{R}^{m\times n(k+1)}$, $V_1 = G/\|G\|_F$, and $\{V_j\}_{j=1}^{k+1}$ is an F-orthonormal basis for the global Krylov subspace

$$\mathbf{K}_{k+1}(\mathcal{A}, G) = \text{span}\{G, \mathcal{A}(G), \ldots, \mathcal{A}^k(G)\}.$$

The global Arnoldi algorithm

1. Let $V_1 = G/\|G\|_F \in \mathbb{R}^{m\times n}$;
2. `for` $j = 1, \ldots, k$ `do`

   2.1. $V = \mathcal{A}(V_j)$;

   2.3. `for` $i = 1, \ldots, j$ `do`

   $$h_{i,j} = \langle V, V_i \rangle_F;$$

   $$V = V - h_{i,j} V_i;$$

   2.4. `end for`

   2.5. $h_{j+1,j} = \|V\|_F$;

   2.6. $V_{j+1} = V/h_{j+1,j}$;
3. `end for`

An element $X_k \in \mathbf{R}^{m\times n}$ in the global Krylov subspace $\mathbf{K}_{k+1}(\mathcal{A}, G)$ can be written as

$$X_k = \sum_{i=1}^{k} y_k^{(i)} V_i = \mathcal{V}_k \left(y_k \otimes I_n\right),$$

where $y_k = [y_k^{(1)}, y_k^{(2)}, \ldots, y_k^{(k)}]^T \in \mathbb{R}^k$.

Moreover,

$$\|\mathcal{A}(X_k) - G\|_F = \left\| H_{k+1,k} y_k - \|G\|_F e_1 \right\|_2.$$

Let

$$M_i = \sum_{j=1}^{q} L_j^{(1)} V_i L_j^{(2)}, \qquad 1 \le i \le k.$$

Then

$$\begin{aligned} \|\sum_{j=1}^{q} L_j^{(1)} X_k L_j^{(2)}\|_F^2 &= \sum_{i,j=1}^{k} y_k^{(i)} y_k^{(j)} \operatorname{trace}(M_i^T M_j) \\ &= y_k^T N y_k = \|R y_k\|_2^2, \quad N = R^T R. \end{aligned}$$

When N is singular, use spectral factorization instead of Choleski factorization.

The matrix Tikhonov regularization problem with solution restricted to $X \in \mathbf{K}_k(\mathcal{A}, G)$ can be written as

$$\min_{y \in \mathbf{R}^k} \left\{ \left\| H_{k+1,k} y - \|G\|_F e_1 \right\|_2^2 + \mu \left\| R y \right\|_2^2 \right\}.$$

The discrepancy principle prescribes that $\mu > 0$ be chosen so that

$$\| H_{k+1,k} y - \|G\|_F e_1 \|_2 = \eta \delta,$$

where

$$E = G - G_{\text{exact}}, \quad \delta = \|E\|_F, \quad \eta > 1.$$

Computed examples

Let

$$\nu = \frac{\|E\|_F}{\|G_{\text{exact}}\|_F},$$

and define the square regularization matrices

$$L_1 = \begin{bmatrix} 1 & -1 & & & & \mathbf{0} \\ & 1 & -1 & & & \\ & & 1 & -1 & & \\ & & & \ddots & \ddots & \\ & & & & 1 & -1 \\ \mathbf{0} & & & & & 0 \end{bmatrix} \in \mathbf{R}^{n\times n}$$

and

$$L_2 = \begin{bmatrix} 0 & 0 & 0 & & & \text{\Large 0} \\ -1 & 2 & -1 & & & \\ & -1 & 2 & -1 & & \\ & & \ddots & \ddots & \ddots & \\ & & & -1 & 2 & -1 \\ \text{\Large 0} & & & 0 & 0 & 0 \end{bmatrix} \in \mathbf{R}^{n \times n}.$$

Example. Restoration of the image `peppers`, which is represented by 256×256 pixels. We let $p = 1$ and $q = 1$. The available image G is corrupted by Gaussian blur and additive zero-mean white Gaussian noise. The blurring matrix $A_1 = [a_{i,j}] \in \mathbf{R}^{256\times 256}$ is Toeplitz with entries

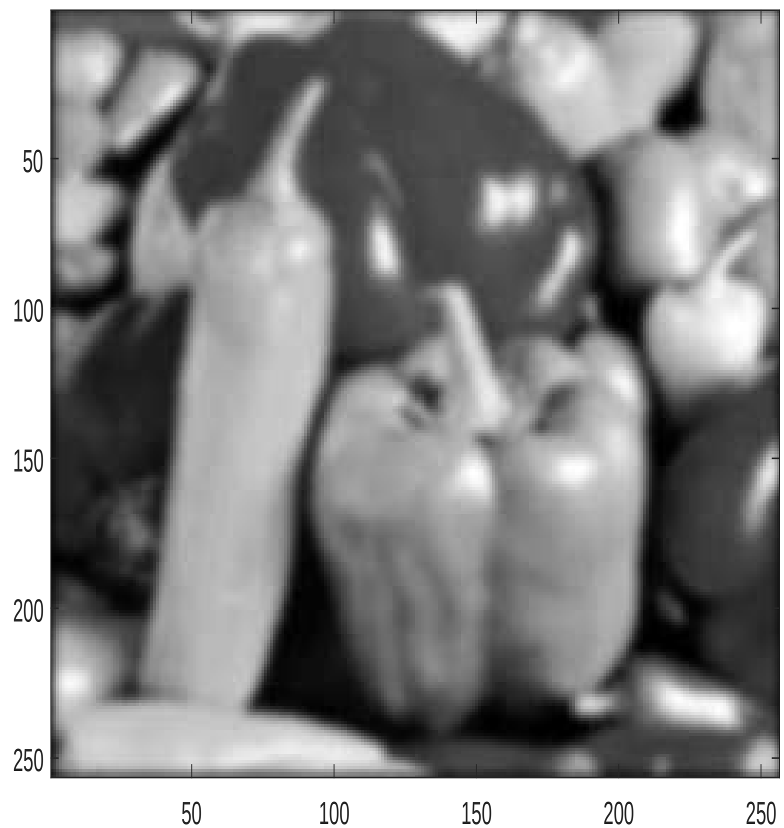
$$a_{i,j} = \begin{cases} \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(i-j)^2}{2\sigma^2}\right), & |i-j| \leq d, \\ 0, & \text{otherwise.} \end{cases},$$

with $d = 7$ and $\sigma = 2$. We let $B_1 = A_1$.

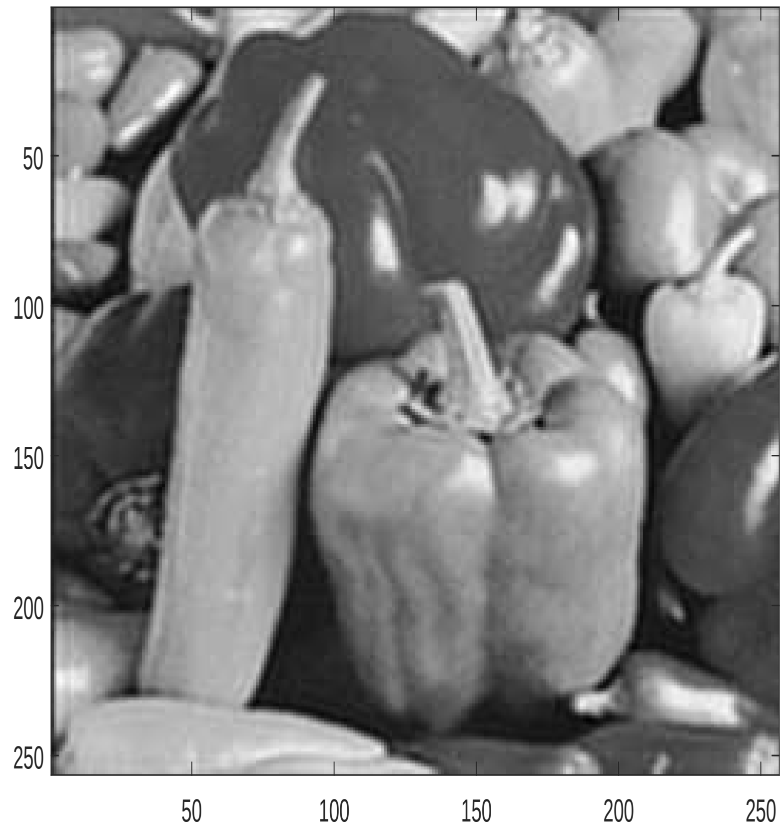
Restoration of peppers, noise level $\nu = 1 \cdot 10^{-2}$.

method	$(L_1^{(1)}, L_1^{(2)})$	k	CPU time (sec)	relative error e_k
SA	(L_1, L_1)	16	2.34	$9.59 \cdot 10^{-2}$
GA	(L_1, L_1)	16	1.42	$9.59 \cdot 10^{-2}$
SA	(L_1, L_2)	15	2.18	$9.64 \cdot 10^{-2}$
GA	(L_1, L_2)	15	1.13	$9.64 \cdot 10^{-2}$
SA	(L_2, L_2)	14	2.14	$9.70 \cdot 10^{-2}$
GA	(L_2, L_2)	14	1.08	$9.70 \cdot 10^{-2}$

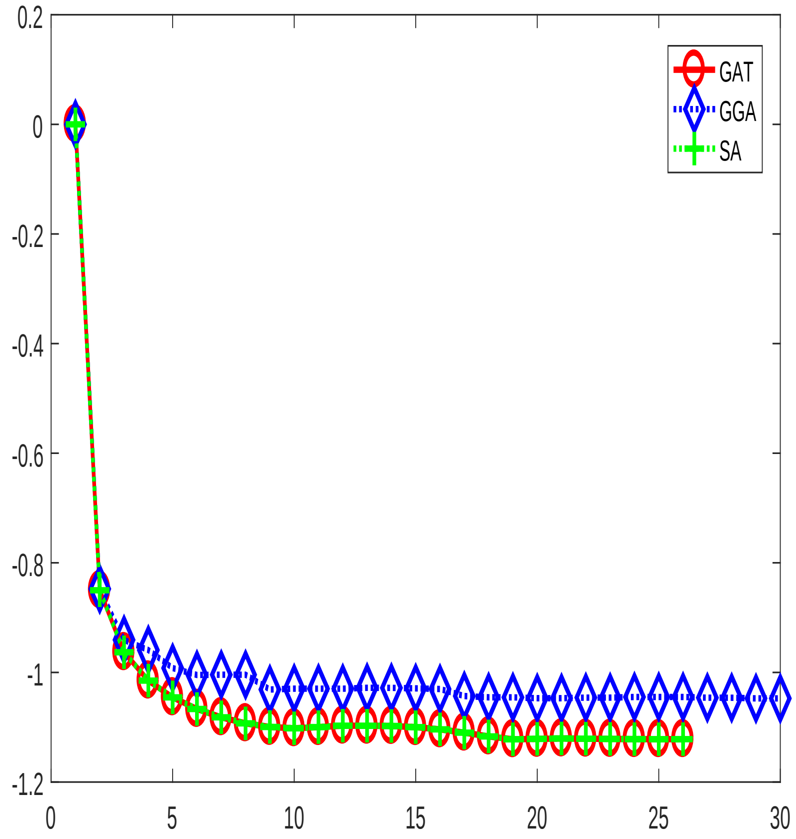
Blurred and noisy image



Restored image



Convergence history



Example. Restoration of the image `cameraman`, which is represented by 512×512 pixels. We let $p = 2$ and $q = 1$. The blurring operator is given by

$$\mathcal{A}(X) = A_1 X B_1 + A_2 X B_2,$$

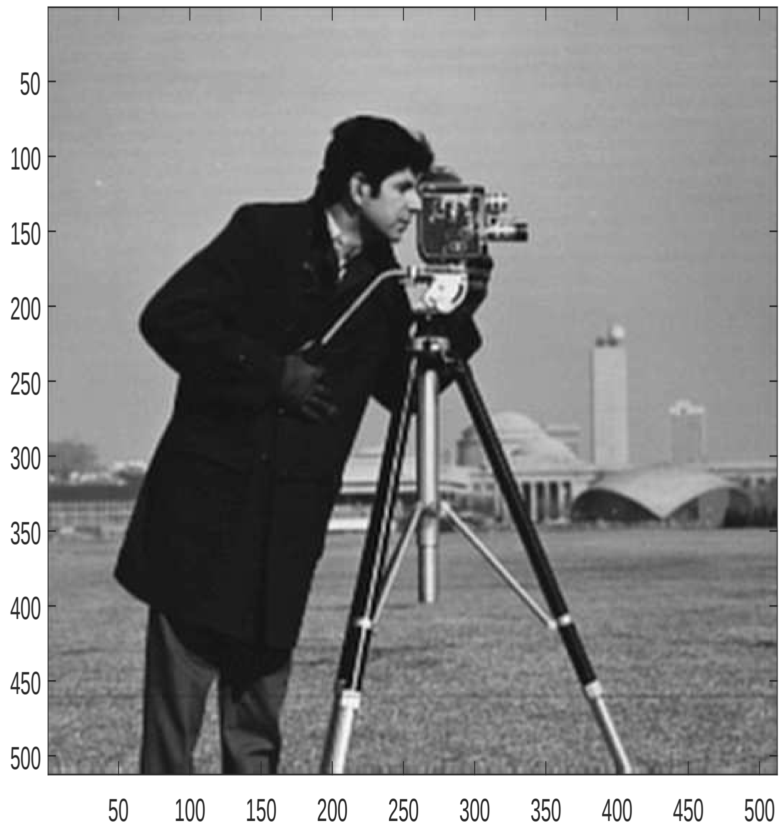
where A_i and B_i are Toeplitz matrices of the same form as previously. The matrix G represents the blurred and noisy image. The noise is white Gaussian.

Restoration of cameraman, noise level $\nu = 1 \cdot 10^{-3}$.

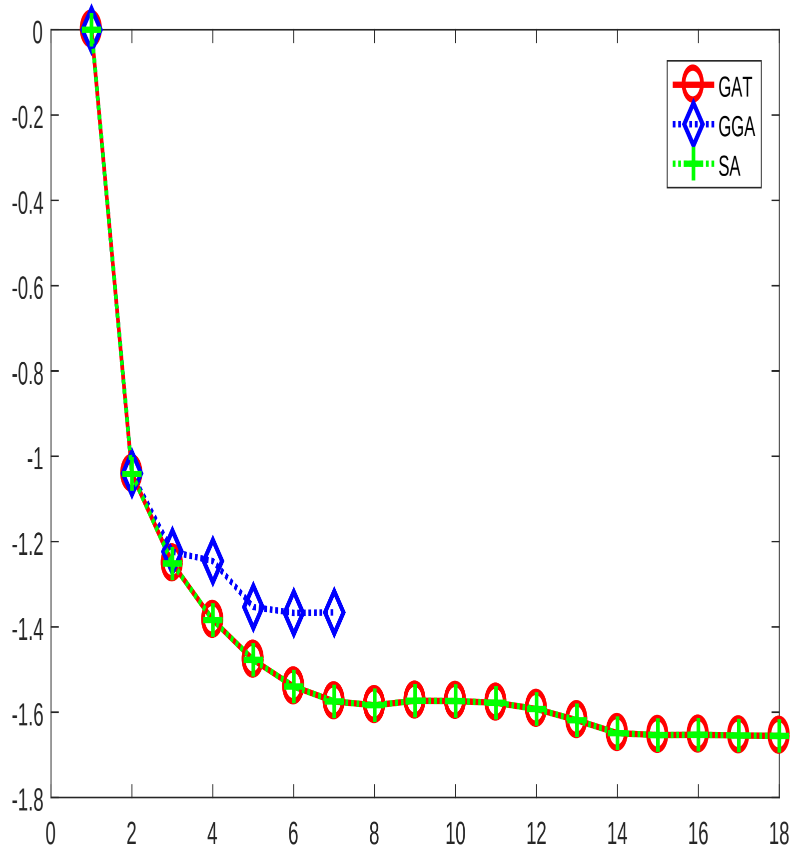
method	$(L_1^{(1)}, L_1^{(2)})$	k	CPU time (sec)	relative error e_k
SA	(L_1, L_1)	17	$1.02 \cdot 10^{-2}$	$2.21 \cdot 10^{-2}$
GA	(L_1, L_1)	17	$1.02 \cdot 10^{-2}$	$2.21 \cdot 10^{-2}$
SA	(L_1, L_2)	16	$9.23 \cdot 10^{-3}$	$2.22 \cdot 10^{-2}$
GA	(L_1, L_2)	16	$9.23 \cdot 10^{-3}$	$2.22 \cdot 10^{-2}$
SA	(L_2, L_2)	9	$1.33 \cdot 10^{-4}$	$2.66 \cdot 10^{-2}$
GA	(L_2, L_2)	9	$1.33 \cdot 10^{-4}$	$2.66 \cdot 10^{-2}$

Blurred and noisy image

Restored image



Convergence history



Iterated Tikhonov regularization

Tikhonov regularization in standard form

$$\min_{x \in \mathbf{R}^n} \{\|Ax - b\|_2^2 + \mu\|x - x_0\|_2^2\},$$

with $A \in \mathbf{R}^{m \times n}$, $b \in \mathbf{R}^m$, $x_0 \in \mathbf{R}^n$, and $\mu > 0$. has a unique solution

$$x_\mu = (A^T A + \mu I)^{-1}(A^T b + \mu x_0).$$

The discrepancy principle prescribes that $\mu_{\text{discr}} = \mu > 0$ be chosen so that

$$\|Ax_{\mu_{\text{discr}}} - b\|_2 = \eta\delta,$$

where $\eta > 1$ is independent of $\delta := \|b - b_{\text{exact}}\|_2$.

Tikhonov regularization in general form:

$$\min_{x \in \mathbf{R}^n} \{ \|Ax - b\|_2^2 + \mu \|L(x - x_0)\|_2^2 \}.$$

$\mu > 0$ regularization parameter, $L \in \mathbf{R}^{p \times n}$ regularization matrix, $x_0 \in \mathbf{R}^n$.

Assume that

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

Then the minimization problem has the unique solution

$$x_\mu = (A^T A + \mu L^T L)^{-1} (A^T b + \mu L^T L x_0)$$

for any $\mu > 0$.

The discrepancy principle prescribes that $\mu_{L,\text{discr}} = \mu > 0$ be chosen so that

$$\|Ax_{\mu_{L,\text{discr}}} - b\|_2 = \eta\delta.$$

The use of a suitable $L \neq I$ may enhance the quality of the computed approximation of x_{true} considerably.

Iterated Tikhonov regularization in standard form:

Let

$$h = x - x_0, \qquad r_0 = b - Ax_0.$$

Tikhonov regularization in standard form:

$$\min_{h \in \mathbf{R}^n} \{\|Ah - r_0\|_2^2 + \mu\|h\|_2^2\},$$

where

$$h \approx x_{\text{exact}} - x_0, \qquad x_{\text{exact}} \approx x_1 := x_0 + h.$$

Repeated application of this refinement strategy gives

Algorithm:

Given $x_0 \in \mathbf{R}^n$

for $k = 0, 1, \ldots$ do

1. compute $r_k = b - Ax_k$,
2. solve $\min_{h \in \mathbf{R}^n} \{\|Ah - r_k\|_2^2 + \mu_k \|h\|_2^2\}$ to obtain h_k,
3. update $x_{k+1} = x_k + h_k$,

where $\mu_0, \mu_1, \ldots$ denotes a sequence of positive regularization parameters.

The iterates can be expressed as

$$x_{k+1} = x_k + (A^T A + \mu_k I)^{-1} A^T (b - A x_k), \quad k = 0, 1, \ldots$$

The iteration method is said to be stationary when $\mu_k = \mu$ for all k, and nonstationary otherwise.

A common choice of regularization parameters for nonstationary iteration is

$$\mu_k = \mu_0 q^k, \qquad \mu_0 > 0, \quad 0 < q < 1.$$

The iterations can be terminated by the discrepancy principle, i.e., as soon as

$$\|A x_k - b\| \leq \eta \delta.$$

Nonstationay iterated Tikhonov regularization in standard form is known to generally determine more accurate approximations of x_{true} than (standard) Tikhonov regularization in standard form.

Nonstationary iterated Tikhonov regularization with a general regularization matrix:

$$x_{k+1} = x_k + (A^T A + \mu_k L^T L)^{-1} A^T (b - A x_k), \quad k = 0, 1, \ldots$$

This method combines the advantages of using a regularization matrix $L \neq I$ with those of nonstationary Tikhonov regularization.

A computed example:

Let $M \in \mathbf{R}^{300\times 300}$ determined discretization of the integral equation of the first kind "shaw" using software in Regularization Tool by Hansen. Let

$$A = \begin{bmatrix} M \\ M \end{bmatrix}, \qquad b \in \mathbf{R}^{600}, \quad \text{relative error } 0.1\%,$$

$$L = \begin{bmatrix} 1 & -1 & & \\ & \ddots & \ddots & \\ & & 1 & -1 \end{bmatrix} \in \mathbf{R}^{299\times 300} \quad \text{bidiagonal.}$$

Compute approximate solution using projection into generalized Krylov subspace.

Algorithm 1:

1. **Input:** $A \in \mathbf{R}^{m\times n}$, $b \in \mathbf{R}^m$, $L \in \mathbf{R}^{p\times n}$, $\eta > 1$, and δ;
2. **Initialize:** Columns of V_0 form orthonormal basis for Krylov subspace $\mathcal{K}_\ell(A^TA, A^Tb)$ for ℓ small; $y_0 = 0 \in \mathbf{R}^\ell$;
3. **for** $k = 1, 2, \ldots$ **until convergence**
4. Let $\bar{y}_k = \left[y_{k-1}^T, 0\right]^T$
5. Determine μ_k so that y_k satisfies $\|AV_ky_k - b\| = \eta\delta$
6. Compute $r_k = A^TAV_ky_k + \mu_k^{-1}L^TLV_k(y_k - \bar{y}_k) - A^Tb$
7. Normalize $v_{k+1} = r_k/\|r_k\|$
8. Enlarge search space $V_{k+1} = [V_k, v_{k+1}]$
9. **end for**
10. **Output:** approximate solution $x_k = V_ky_k$ and μ_k .

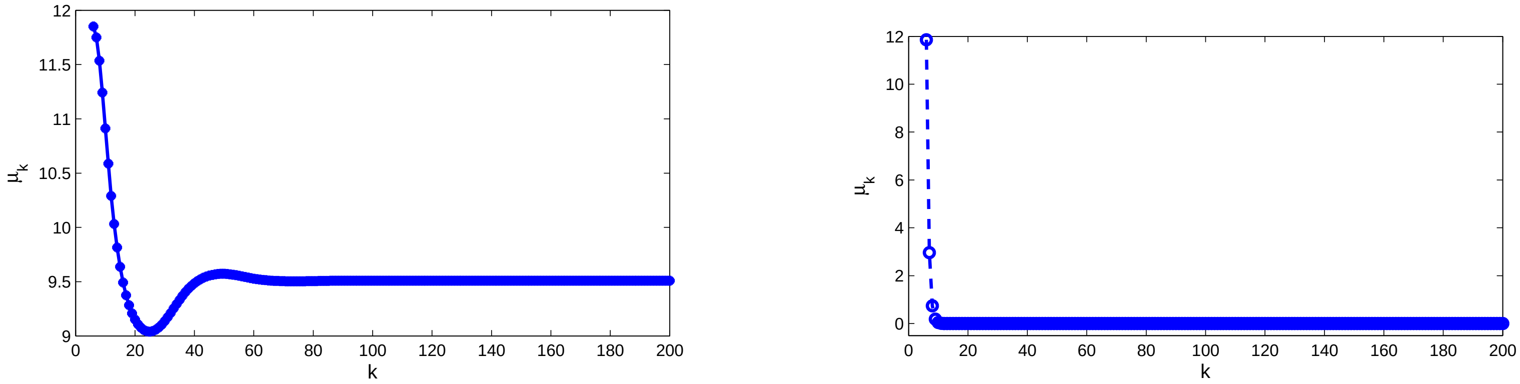


Figure 1: Convergence of regularization parameters for standard Tikhonov (left), and nonstationary iterated Tikhonov (right).

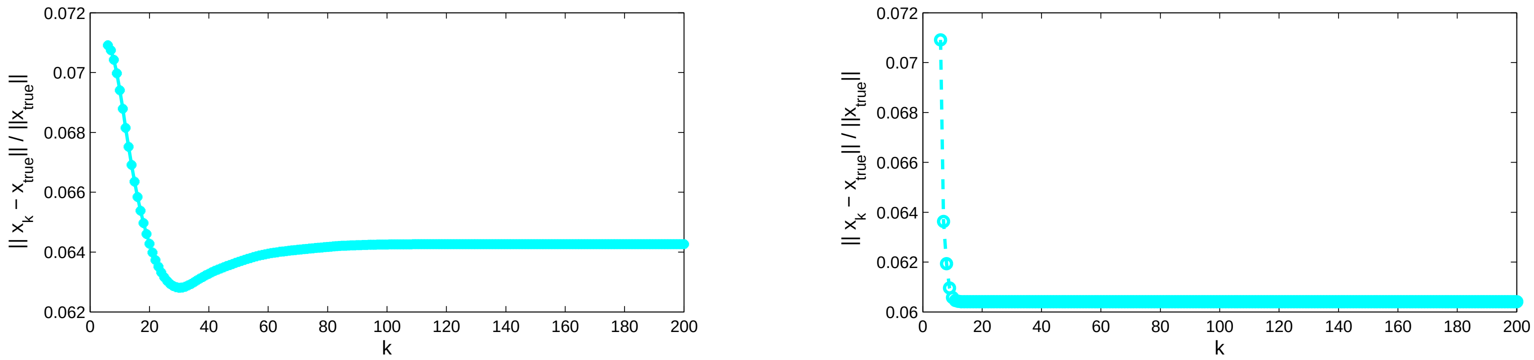


Figure 2: Convergence of computed solutions for standard Tikhonov (left), and nonstationary iterated Tikhonov (right).

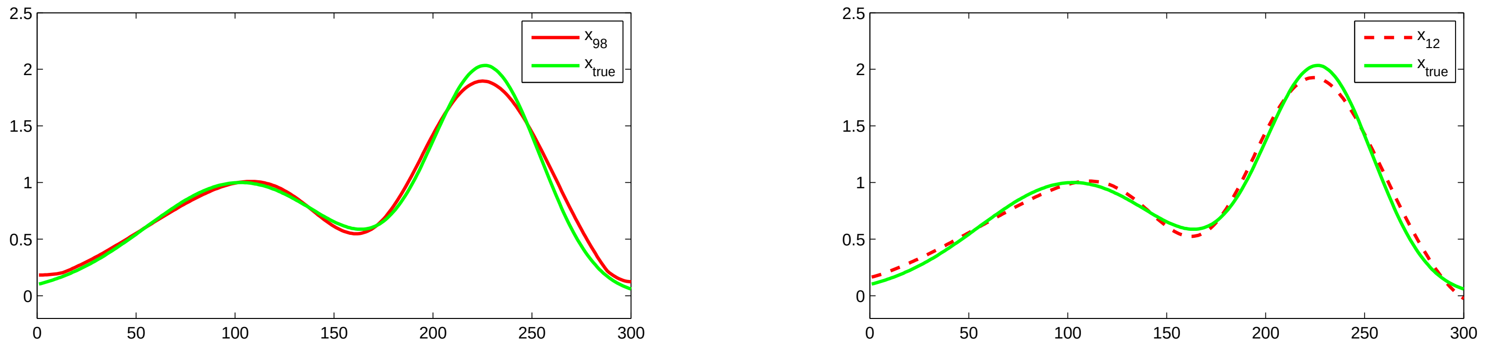


Figure 3: Computed solutions at convergence for standard Tikhonov (left), and nonstationary iterated Tikhonov (right).

Transformation to standard form:

Let $A \in \mathbf{R}^{m\times n}$ and $L \in \mathbf{R}^{p\times n}$. Consider

$$\min_{h\in\mathbf{R}^n} \{\|Ah - r_0\|_2^2 + \mu\|Lh\|_2^2\}. \qquad (2)$$

Define the A-weighted generalized inverse of L:

$$L_A^\dagger = (I - (A(I - L^\dagger L))^\dagger A)L^\dagger.$$

Proposition (Buccini): Let A and L be square, of the same size, and commute. Then $L_A^\dagger = L^\dagger$.

Assume that the conditions of the propositionns hold.
Define

$$\begin{aligned} \bar{h} &= Lh, \\ h^{(0)} &= (A(I - L^{\dagger}L))^{\dagger} r_0, \\ \bar{r}_0 &= r_0 - Ah^{(0)}. \end{aligned}$$

Then (2) can be expressed in standard form

$$\min_{\bar{h}\in\mathbf{R}^n}\{\|AL^\dagger\bar{h}-\bar{r}_0\|_2^2+\mu\|\bar{h}\|_2^2\}.$$

Denote the solution by $\bar{h}_\mu$.

The solution of the minimization problem in general form is

$$h_\mu = L^\dagger\bar{h}_\mu + h^{(0)}.$$

Stationary iterated Tikhonov with general penalty term:

Algorithm 2: Let $\mu > 0$ and $x_0 \in \mathbf{R}^n$.
Compute

$$
\begin{aligned}
&\textbf{for } k = 0, 1, \ldots \\
&\quad r_k = b - Ax_k \\
&\quad \textbf{if } \|r_k\|_2 < \eta\delta \textbf{ exit} \\
&\quad x_{k+1} = x_k + \left(A^T A + \mu L^T L\right)^{-1} A^T r_k \\
&\textbf{end}
\end{aligned}
$$

Convergene analysis for square matrices A and L:

Define the splitting

$$\mathbf{R}^n = \mathcal{N}(L) \oplus \mathcal{N}(L)^{\perp}.$$

We will study convergence in these subspaces separately.

The minimization problem with $\mu > 0$

$$\min_{h \in \mathbf{R}^n} \{ \|Ah - r_k\|_2^2 + \mu \|Lh\|_2^2 \}$$

has the solution

$$h_k = (A^T A + \mu L^T L)^{-1} A^T r_k.$$

Transformation to standard form with $\bar{A} = AL^\dagger$ yields

$$h_k = h_k^\perp + h_k^{(0)},$$

where

$$\begin{aligned} h_k^{(0)} &= (A(I - L^\dagger L))^\dagger r_k, \\ \bar{r}_k &= r_k - Ah_k^{(0)}, \\ h_k^\perp &= L^\dagger(\bar{A}^T \bar{A} + \mu I)^{-1} \bar{A}^T \bar{r}_k. \end{aligned}$$

Lemma:

$$h_k^{\perp} \in \mathcal{N}(L)^{\perp}, \qquad h_k^{(0)} \in \mathcal{N}(L), \quad k = 0, 1, \ldots \ .$$

Consider the splitting

$$x_k = x_k^{(0)} + x_k^{\perp},$$

where

$$\begin{aligned} x_k^{(0)} &= x_0^{(0)} + \sum_{j=0}^{k-1} h_j^{(0)} \in \mathcal{N}(L), \\ x_k^{\perp} &= x_0^{\perp} + \sum_{j=0}^{k-1} h_j^{\perp} \in \mathcal{N}(L)^{\perp}. \end{aligned}$$

Proposition: Let $x_0 = 0$. Then

$$\begin{aligned} x_k^{\perp} &\to P_{\mathcal{N}(L)^{\perp}}(A^{\dagger}b) \text{ as } k \to \infty. \\ x_k^{(0)} &= P_{\mathcal{N}(L)}(A^{\dagger}b), \qquad k = 1, 2, \ldots \ . \end{aligned}$$

Convergence result of interest for error-free problems:

Theorem 1: Let the matrices A and L be square and of the same size, and let their nullspaces intersect trivially. Let $x_0 = 0$. Then the iterates determined by Algorithm 2 converge to the minimum norm solution $A^{\dagger}b$ of the linear system of equations $Ax = b$.

Convergence result of interest for error contaminated problems:

Theorem 2: Let the assumptions of the above theorem hold. Then Algorithm 2 terminates after finitely many, $k = k_\delta$, steps and

$$\limsup_{\delta \searrow 0} \|x_{\text{true}} - x_{k_\delta}\|_2 = 0.$$

Extensions to rectangular matrices A and L:

- A rectangular: make square by zero-padding.
- $L \in \mathbf{R}^{p \times n}$ rectangular:
  - $p < n$: make square by zero-padding,
  - $p > n$: compute QR factorization $L = QR$, $R \in \mathbf{R}^{n \times n}$. Use R instead of L.

Nonstationary iterated Tikhonov with general A and L:

Consider the iterations

$$x_{k+1} = x_k + (A^T A + \mu_k L^T L)^{-1} r_k, \quad r_k = b - A x_k, \quad k = 0, 1, \ldots$$

and assume

$$\sum_{k=0}^{\infty} \mu_k^{-1} = \infty.$$

Algorithm 3: Let $\mu > 0$ and let $x_0 \in \mathbf{R}^n$ be an available initial approximation of x_{true}. Compute

$$
\begin{array}{l}
\textbf{for } k = 0, 1, \ldots \\
\quad r_k = b - Ax_k \\
\quad \textbf{if } \|r_k\|_2 < \eta\delta \textbf{ exit} \\
\quad x_{k+1} = x_k + (A^T A + \mu_k L^T L)^{-1} A^T r_k \\
\textbf{end}
\end{array}
$$

Theorem 3: Let the conditions of Theorem 1 hold and let the regularization parameters satisfy $\sum_{k=0}^{\infty} \mu_k^{-1} = \infty$. Then the iterates determined by Algorithm 3 converge to the minimum norm solution $A^\dagger b$ of the linear system of equations $Ax = b$.

Theorem 4: Let the assumptions of the above theorem hold. Then Algorithm 3 terminates after finitely many, $k = k_\delta$, steps and

$$\limsup_{\delta \searrow 0} \|x_{\text{true}} - x_{k_\delta}\|_2 = 0.$$

Computed examples:

Example 1: Problem baart from Regularization Tools. Discretize integral equation of the first kind,

$$\int_0^\pi \exp(s\cos(t))x(t)dt = 2\frac{\sinh(s)}{s}, \qquad 0 < s \leq \frac{\pi}{2}.$$

Gives $A \in \mathbf{R}^{1000\times 1000}$ and $b_{\text{true}} \in \mathbf{R}^{1000}$. Add 1% noise to b_{true} to obtain error-contaminated right-hand side b.

Use regularization matrices $L = I$ or

$$L_1 = \begin{bmatrix} -1 & 1 & & \\ & \ddots & \ddots & \\ & & -1 & 1 \\ & & & 0 \end{bmatrix}, \quad L_2 = \begin{bmatrix} 0 & 0 & & & \\ -1 & 2 & -1 & & \\ & \ddots & \ddots & \ddots & \\ & & -1 & 2 & -1 \\ & & & 0 & 0 \end{bmatrix}.$$

Then

$$\begin{aligned} \mathcal{N}(L_1) &= \operatorname{span}([1, 1, \ldots, 1]^T), \\ \mathcal{N}(L_2) &= \operatorname{span}([1, 1, \ldots, 1]^T, [1, 2, \ldots, 1000]^T). \end{aligned}$$

We report the relative error $\|x_k - x_{\text{true}}\|_2/\|x_{\text{true}}\|_2$ in the approximate solution x_k computed with Algorithm 3.

$$\mu_k = \mu_0 q^k, \quad q = 0.8, \quad k = 1, 2, \dots \ .$$

L	μ_0	$\|x_k - x_{\text{true}}\|_2/\|x_{\text{true}}\|_2$	# iterations
I	10^{-2}	0.17131	4
L_1	10^2	0.12331	3
L_2	10^6	0.04290	2

Example 2: Problem gravity from Regularization Tools: Discretize integral equation of the first kind,

$$\int_0^1 \frac{d}{(d^2 + (s-t)^2)^{3/2}} x(t)dt = g(s), \qquad 0 \le s \le 1,$$

with $d = 1/4$ and g chosen so that

$$x(t) = \sin(\pi t) + \frac{1}{2}\sin(2\pi t).$$

Gives $A \in \mathbf{R}^{1000 \times 1000}$ and $b_{\text{true}} \in \mathbf{R}^{1000}$. Add 1% Gaussian noise to b_{true} to obtain error-contaminated right-hand side b.

L	μ_0	$\|x_k - x_{\text{true}}\|_2 / \|x_{\text{true}}\|_2$	# iterations
I	10^{-2}	0.17001	2
L_1	10^{2}	0.10165	2
L_2	10^{6}	0.08148	2

Example 3: Image restoration problem. Define

$$L_1^c = \begin{bmatrix} -1 & 1 & & & \\ & \ddots & \ddots & & \\ & & -1 & 1 & \\ 1 & & & -1 & \end{bmatrix}, \quad L_2^c = \begin{bmatrix} 2 & -1 & & & -1 \\ -1 & 2 & -1 & & \\ & \ddots & \ddots & \ddots & \\ & & -1 & 2 & -1 \\ -1 & & & -1 & 2 \end{bmatrix}.$$

as well as

$$L_1 = L_1^c \otimes I + I \otimes L_1^c, \quad L_2 = L_2^c \otimes I + I \otimes L_2^c,$$

where $\otimes$ denotes Kronecker product.

(a) (b)

Figure 4: (a) Uncontaminated image (512×512 pixels), (b) blur- and noise-contaminated image. Error 3%.

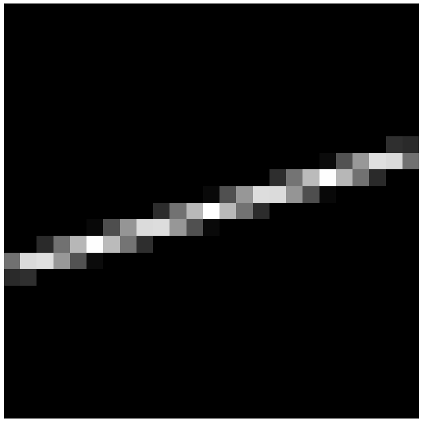
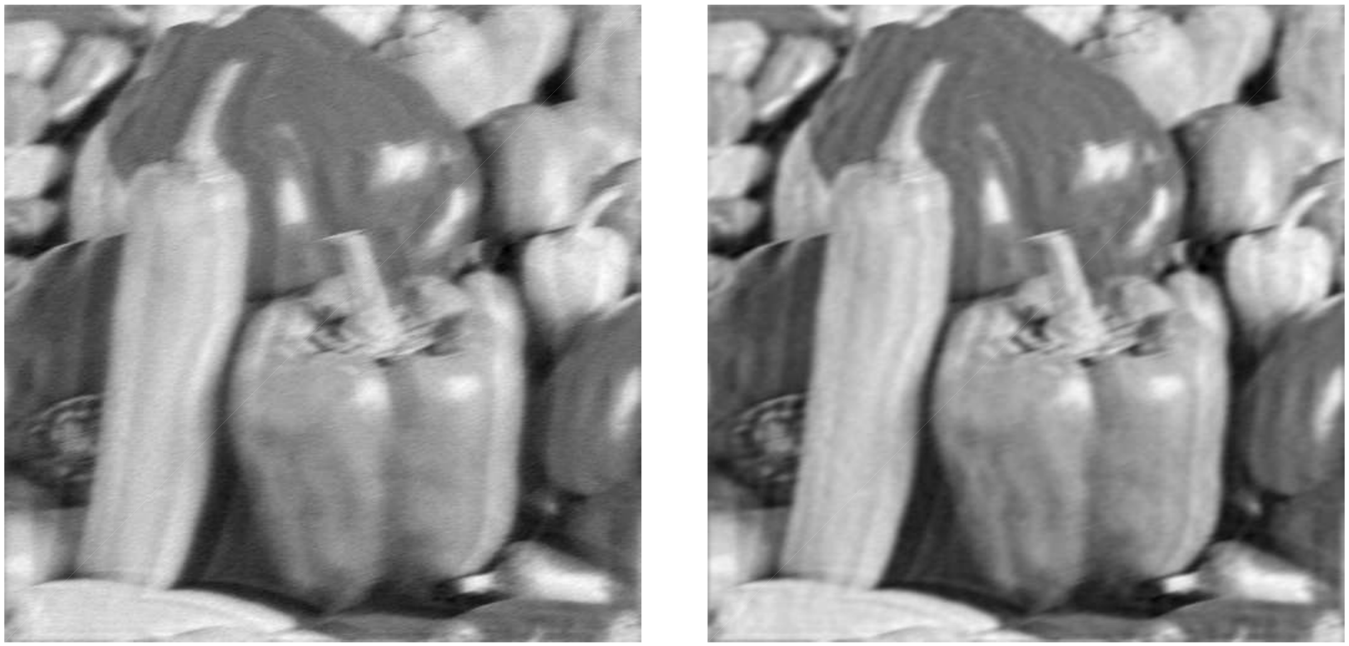


Figure 5: PSF (25×25 pixels) models motion blur.

Restoration of “peppers” by Algorithm 3, $\mu_0 = 1$.
Matrix-vector products can be evaluated quickly with the aid of the FFT.

L	$\|x_k - x_{\text{true}}\|_2/\|x_{\text{true}}\|_2$	# iterations
I	0.10743	7
L_1	0.09368	4
L_2	0.08516	3



(a) (b)

Figure 6: Restorations determined by Algorithm 3 with (a) $L = I$, (b) $L = L_1$.

Figure 7: Restorations determined by Algorithm 3 with $L = L_2$.

Preconditioning

Tikhonov regularization is closely related to preconditiing. Let the regularization matrix L be square and nonsingular. Then Tikhonov regularization in general form

$$\min_{x \in \mathbf{R}^n} \{\|Ax - b\|_2^2 + \mu \|Lx\|_2^2\}$$

easily can be transformed to standard form by letting $x = Ly$:

$$\min_{y \in \mathbf{R}^n} \{\|AL^{-1}y - b\|_2^2 + \mu \|y\|_2^2\}$$

The matrix A is right-preconditioned by L^{-1}.

When L is not square, we can replace L^{-1} above by the A-weighted generalized inverse of L:

$$L_A^\dagger = (I - (A(I - L^\dagger L))^\dagger A)L^\dagger.$$

Thus, we solve the minimization problem

$$\min_{y \in \mathbf{R}^n} \{\|AL_A^\dagger y - b\|_2^2 + \mu\|y\|_2^2\}$$

The matrix A is right-preconditioned by $L_A^\dagger$.

Note: The "preconditioner" should not be an accurate approximation of $A^\dagger$, because this would result in a large propagetd error (stemming from the error in b) in the computed solution.

We conclude that the "preconditioner"

- should approximate A well enough to make it possible to determine an accurate approximation of x_{exact} in a solution subspace of low dimension,
- should not approximate A well enough to cause propagation and amplification of the error in b into the computed approximation of x_{exact}.

We describe a method by Donatelli and Hanke that achieves these goals.

The method by Donatelli and Hanke

We would like to compute an approximate solution of the discrete ill-posed problem

$$Ax = b,$$

where the singular values of $A \in \mathbf{R}^{n\times n}$ "cluster" at the origin.

The normal equations for Tikhonov regularization

$$(A^T A + \mu I)x = A^T b,$$

determine the approximation x_μ of x_{exact}, where $\mu > 0$ is a regularization parameter.

Assume the normal equations are expensive to solve. Let $C \in \mathbf{R}^{n\times n}$ approximate A and be such that

$$(C^T C + \mu I)h = C^T b,$$

is easier to solve.

Donatelli and Hanke proposed the method: Let $x^{(0)} \in \mathbf{R}^n$ and repeat

for $k = 0, 1, 2, \ldots$ until discrepancy principle satisfied:

1. $r^{(k)} = b - Ax^{(k)}$
2. $h^{(k)} = C^T(CC^T + \mu_k I)^{-1} r^{(k)}$
3. $x^{(k+1)} = x^{(k)} + h^{(k)}$

end

Some observations:

- When $C = A$ and

$$\mu_k = \alpha q^k, \quad \alpha > 0, \quad 0 < q < 1, \quad k = 0, 1, 2, \dots \;,$$

  the iterations are identical with nonstationary iterated Tikhonov regularization.

- Iterated Tikhonov is an iterative refinement procedure. We terminate early due to the discrepancy principle. This introduces an error. Replacing A by C introduces another error.

- Convergence analysis of the method requires that for some $0 < \rho < 1/2$,

$$\|(C - A)x\|_2 \leq \rho \|Ax\|_2 \quad \forall x \in \mathbf{R}^n$$

This inequality may be difficult to verify. It leads to that

$$\|r_k - C(x_{\text{exact}} - x_k)\|_2 < (1 - \rho)\|r_k\|_2.$$

The latter inequality has been verified in image restoration applications.

In image restoration applications with a space invariant point spread function

- A is a block-Toeplitz-Toeplitz-block matrix, except for the boundary conditions that may destroy some of the structure,
- C is a block-circulant-circulant-block matrix.

This allows fast evaluation of $Ax^{(k)}$ and $C^T(CC^T + \mu_k)^{-1}r^{(k)}$ with the FFT.