A Generalized Matrix Krylov Subspace Method for TV regularization

El Guide Mohamed : UM6P, Ben Guerir

Joint work with :
A. H. Bentbib : FSTG Marrakech.
K. Jbilou : Littoral University, France.

03 July, NASCA 2018, Kalamata

Outline

We consider the solution of the following matrix equation

$$B = H_2 X H_1^T, \tag{1}$$

where,

- $B \in \mathbb{R}^{m \times n}$ is generally contaminated by noise.
- $H_1 \in \mathbb{R}^{n \times m}$ and $H_2 \in \mathbb{R}^{m \times n}$ are matrices of ill-determined rank.

Discrete ill-posed problems of the form (1) arise, for instance, from :

Fredholm integral equations of the first kind in two space dimensions

$$\int\int_{\Omega} K(x,y,s,t)f(s,t)dsdt = g(x,y), \qquad (x,y) \in \Omega',$$

where,

- Ω and Ω' are rectangles in $\mathbb{R}^2$
- The kernel is separable

$$K(x,y,s,t) = k_1(x,s)k_2(y,t), \quad (x,y) \in \Omega', \quad (s,t) \in \Omega,$$

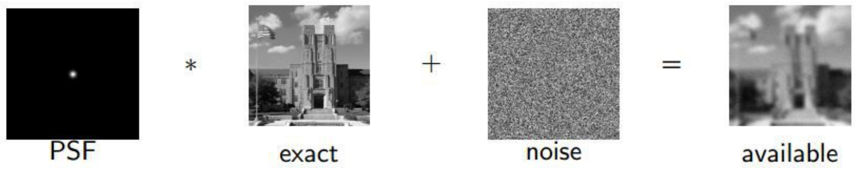
Image restoration with a separable kernel

The blurring process of an image can be formulated as a Fredholm integral equation of the first kind which has the following classic form :

$$g(x,y) = \int\int_{\Omega} K(x,y,s,t) f(s,t) ds dt,$$

where

- f represents the true object,
- g the observed image,
- K is a given PSF



Multi-channel images

The within-channel blurring process of a digital RGB color image can be formulated as a 3 Fredholm integral equation of the first kind which have the following classic form :

$$g^{(k)}(x,y) = \int\int_{\Omega} K(x,y,s,t) f^{(k)}(s,t)\,dsdt, \quad k \in \{r,g,b\}$$

where

- $f^{(k)}$ represents the true k channel,
- $g^{(k)}$ the blurred k channel,
- K is a given PSF.

Simple Example : Blurred Image Restoration

- We consider $\widehat{X}$ the popular test image `peppers.png`, in its original size 256×256 pixels
- We consider the Gaussian PSF $\frac{1}{2\pi\sigma^2}\exp(-\frac{x^2+y^2}{2\sigma^2})$
- We use Hansen's function blur.m (Hansen, SIAM 99) to build the blurring matrices H_1 and H_2,
- We consider a noise level $\nu = \frac{\|E\|_F}{\|\widehat{B}\|_F} = 0.01$
- We construct the corrupted image, in the matrix form, as $H_2 X H_1^T = \widehat{B} + E$

Sensitivity to noise in the data : How much damage can a little noise really do ?

Solution depends on the conditioning of H_1 and H_2, here $4.0692e+06$

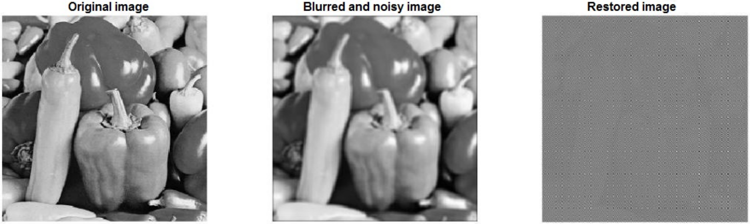


Figure – Inverse crime $X = H_2^{-1}(\widehat{B} + E)H_1^{-T}$

- Introduce the linear operator

$$\begin{aligned} \mathcal{A} : \mathbb{R}^{m\times n} &\rightarrow \mathbb{R}^{m\times n} \\ \mathcal{A}(X) &= H_2 X H_1^T. \end{aligned}$$

  Its transpose is given by $\mathcal{A}^T(X) = H_2^T X H_1$. Define the symmetric linear operator

$$\tilde{\mathcal{A}}(X) = (\mathcal{A}^T \circ \mathcal{A})(X),$$

- Using Tikhonov regularization (C. W. Groetsch, 1984) is equivalent to following matrix

$$(\tilde{\mathcal{A}} + \mu I)(X) = \mathcal{A}^T(B)$$

  where μ is the regularization parameter

Solution methods for Tikhonov regularization

- Tikhonov regularization with a solution norm constraint (A.H. Bentbib, M. El Guide, K. Jbilou and L. Reichel, JCAM 2016) :
  - Global Lanczos algorithm
  - Gauss quadrature rules
- Tikhonov regularization with error norm constraint (A.H. Bentbib, M. El Guide, K. Jbilou and L. Reichel, JCAM 2017) :
  - Global Golub-Kahan bidiagonalization
  - Gauss quadrature rules
- The block Lanczos algorithm for linear ill-posed problems (A.H. Bentbib, M. El Guide, K. Jbilou, CALCOLO 2017)
- Block algorithms for Tikhonov regularization (A.H. Bentbib, M. El Guide, K. Jbilou, E. Onunwor and L. Reichel, BIT 2018) :
  - Block Golub-Kahan bidiagonalization
  - Block Gauss quadrature rules

Total Variation (TV) regularization

- The TV model (Rudin, Osher and Fatemi, Physica D,1992) is a well known regularization method in preserving sharp edges
- For the additive noise we consider the following TV/L2 problem

$$\min_X \left(\|\mathcal{A}(X) - B\|_F^2 + \mu \mathrm{TV}(X) \right).$$

- For the impulsive noise we consider the following TV/L1 problem

$$\min_X \left(\|\mathcal{A}(X) - B\|_{1,1} + \mu \mathrm{TV}(X) \right) \quad .$$

  where $\|.\|_F$ is the Frobenius norm, $\|.\|_{1,1}$ is the ℓ_1 and μ is a regularization parameter.

- $$\mathrm{TV}(X) = \sum_{i=1}^{m} \sum_{j=1}^{n} \sqrt{\left((D_{1,n}X)_{ij}^2 + (D_{1,m}X)_{ij}^2 \right)}$$

- $D_{1,m}$ and $D_{1,n}$ are defined as follows

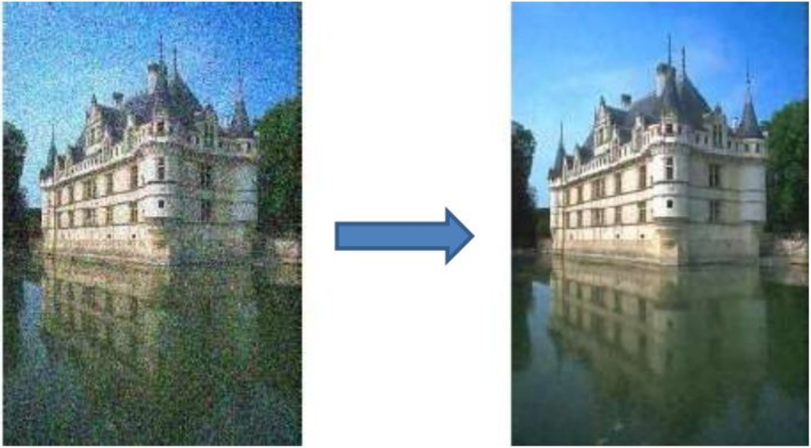
$$\begin{pmatrix} D_{1,n} \\ D_{1,m} \end{pmatrix} X = \begin{pmatrix} CX \\ XC^T \end{pmatrix},$$

where

$$C := \begin{bmatrix} -1 & 1 & & \\ & \ddots & \ddots & \\ & & -1 & 1 \end{bmatrix} \in \mathbb{R}^{d-1 \times d},$$

Example of applications

- Deblurring and denoising of color images



- TV/L2 and TV/L1 models are very difficult to solve directly due to the non-differentiability and nonlinearity of the terms

How to solve this problem ?

- Convex Optimization Method for ℓ_1-minimization
  1. Augmented Lagrangian Methods (ALM) (M. R. Hestenes, JOTA 1969)
  2. Alternating Direction Method of Multipliers (ADMM) (D. Gabay and B. Mercier, JCAM 1976)
- Generalized Krylov subspaces (J. Lampey, L. Reichel and H. Voss, JLAA 2012))

Augmented Lagrangian Method for TV/L1

- The idea of ALM is to transform the unconstrained TV/L1 minimization into an equivalent constrained optimization problem
- The TV/L1 the problem can be rewritten as

$$\begin{aligned} &\min \quad F(R) + G(Y), \\ &\text{subject to} \quad DX = Y, \quad \mathcal{A}(X) = R \end{aligned}$$

where,

$$F(R) = \|R - B\|_{1,1}, \quad G(Y) = \mu \sum_{i=1}^{m} \sum_{j=1}^{n} \|M_{i,j}\|_2, \quad D = \begin{pmatrix} D_{1,n} \\ D_{1,m} \end{pmatrix},$$

$$Y = \begin{pmatrix} M^{(n)} \\ M^{(m)} \end{pmatrix}, M_{i,j} = \left[(D_{1,n}X)_{ij}, (D_{1,m}X)_{ij} \right],$$

$$M_{i,j}^{(n)} = (D_{1,n}X)_{ij}, \quad M_{i,j}^{(m)} = (D_{1,m}X)_{ij}.$$

Augmented Lagrangian Method

- The augmented Lagrangian function of our constrained is defined as follows

$$\begin{aligned}\mathcal{L}(X,R,Y,Z,W) &= F(R)+G(Y)+\frac{\beta}{2}\|DX-Y\|_F^2+\langle DX-Y,Z\rangle_F \\ &+ \frac{\rho}{2}\|\mathcal{A}(X)-R\|_F^2+\langle \mathcal{A}(X)-R,W\rangle_F\end{aligned}$$

- $Z \in \mathbb{R}^{2m\times n}$ and $W \in \mathbb{R}^{m\times n}$ are the Lagrange multipliers of the linear constraint $DX = Y$ and $R = \mathcal{A}(X)$, respectively
- The parameters $\beta > 0$ and $\rho > 0$ are the penalty parameters for the violation of the linear constraint.

Alternating Direction Method of Multipliers

- The idea of this method is to apply an alternating minimization iterative procedure, namely, for $k = 0, 1, \ldots,$ we solve

$$(X_k, R_k, Y_k) = \underset{X,R,Y}{\arg\min}\ \mathcal{L}(X, R, Y, Z_k, W_k).$$

- The Lagrange multipliers are updated by

$$\begin{aligned} Z_{k+1} &= Z_k + \beta\left(DX_k - Y_k\right). \\ W_{k+1} &= W_k + \rho\left(\mathcal{A}(X_k) - R_k\right). \end{aligned}$$

Solving the R-problem

- Given X, the iterate R_k can be obtained by solving the minimization problem

$$\min_{R} \|R - B\|_{1,1} + \frac{\rho}{2} \|\mathcal{A}(X) - R\|_F^2 + \langle \mathcal{A}(X) - R, W_k \rangle_F .$$

- by using the following well-known one-dimensional Shrinkage formula (C. Li, Ph.D. thesis 2009)

$$\textbf{Shrink}(y, \gamma, \delta) = \max\left\{ \left| y + \frac{\gamma}{\delta} \right| - \frac{1}{\delta}, 0 \right\} .\text{sign}\left(y + \frac{\gamma}{\delta} \right),$$

- The minimizer is then given by

$$\max\left\{ \left| \mathcal{A}(X) - B + \frac{1}{\rho} W \right| - \frac{1}{\rho}, 0 \right\} .\text{sign}\left(\mathcal{A}(X) - B + \frac{1}{\rho} W \right)$$

Solving the Y-problem

- Given X and R, we compute the iterates Y_k can be obtained by solving the problem

$$\min_{Y} \; G(Y) + \frac{\beta}{2}\|DX - Y\|_F^2 + \langle DX - Y, Z_k \rangle_F$$

- We use following well-known two dimensional shrinkage formula (C. Li, Ph.D. thesis 2009)

$$\textbf{Shrink}(y, \gamma, \delta) = \max\left\{ \left\| y + \frac{\gamma}{\delta} \right\|_2 - \frac{1}{\delta}, 0 \right\} \frac{y + \gamma/\delta}{\|y + \gamma/\delta\|_2},$$

- The solution is then given by

$$M_{i,j} = \max\left\{ \|T_{i,j}\|_2 - \frac{\mu}{\beta}, 0 \right\} \frac{T_{i,j}}{\|T_{i,j}\|_2},$$

where $T_{i,j} = \left[(D_{1,n}X_k)_{i,j} + \frac{1}{\beta}\left(Z_k^{(1)}\right)_{i,j}, (D_{1,m}X_k)_{i,j} + \frac{1}{\beta}\left(Z_k^{(2)}\right)_{i,j} \right].$

Solving the X-problem

- Given Y and R, X_k can be obtained by solving the minimization problem

$$\min_X \frac{\beta}{2}\|DX - Y\|_F^2 + \langle DX - Y, Z_k\rangle_F + \frac{\rho}{2}\|\mathcal{A}(X) - R\|_F^2 + \langle \mathcal{A}(X) - R, W_k\rangle_F$$

- The problem be solved by considering the following normal equation

$$\rho H_1^T H_1 X H_2^T H_2 + \beta D^T D X = H_1^T\left(\rho R - W_k\right) H_2 + D^T\left(\beta Y - Z_k\right).$$

- The linear matrix equation can be rewritten in the following form

$$A_1 X A_2 + A_3 X A_4 = E_k,$$

where $A_1 = \rho H_1^T H_1$, $A_2 = H_2^T H_2$, $A_3 = \beta D^T D$, $A_4 = I$ and $E_k = H_1^T\left(\rho R - W_k\right) H_2 + D^T\left(\beta Y - Z_k\right)$.

Solving the X-problem

- Let us first introduce the following linear matrix operator

$$\begin{aligned}\mathcal{H}:\mathbb{R}^{m\times n} &\rightarrow \mathbb{R}^{m\times n}\\ \mathcal{H}(X) &:= A_1XA_2 + A_3XA_4.\end{aligned}$$

- The X problems can be then expressed as follows

$$\mathcal{H}(X) = E_k, \quad k = 0, 1, \ldots$$

- We start with the solution X_1 of the following linear matrix equation

$$\mathcal{H}(X) = E_0$$

- We search for an approximation of the solution by solving the following minimization problem,

$$\min_X \|\mathcal{H}(X) - E_0\|_F$$

Solving the X-problem

- Let X_0 be an initial guess of X_1 and $P_0 = \mathcal{H}(X_0) - E_0$ the corresponding residual. We use the modified global Arnoldi algorithm to construct an F-orthonormal basis $\mathcal{V}_m = [V_1, V_2, ..., V_m]$ of the following matrix Krylov subspace (K. Jbilou, A. Messaoudi and H. Sadok, Appl. Numer. Math, 1999)

$$\mathcal{K}_m(\mathcal{H}, P_0) = \text{span}\left\{P_0, \mathcal{H}(P_0), ..., \mathcal{H}^{m-1}(P_0)\right\}.$$

- This gives the following relation

$$\mathcal{H}(\mathcal{V}_m) = \mathcal{V}_{m+1}(H_m \otimes I_n),$$

  where $H_m \in \mathbb{R}^{(m+1)\times m}$ is an upper Hessenberg matrix.
- We search for an approximated solution X_1^m of X_1 belonging to $X_0 + \mathcal{K}_m(\mathcal{H}, P_0)$. This shows that X_1^m can be obtained as follows

$$X_1^m = X_0 + \mathcal{V}_m(y_m \otimes I_n),$$

Solving the X-problem

- y_m is the solution of the following reduced minimization problem

$$\min_{y\in\mathbb{R}^m} \left\| H_m y - \|P_0\|_F e_1 \right\|_2 ,$$

where e_1 denotes the first unit vector of $\mathbb{R}^{m+1}$.

- Now we turn to the solutions of

$$\mathcal{H}(X) = E_k, \quad k = 1, 2, \ldots$$

- For example, in the beginning of solving $\mathcal{H}(X) = E_1$, we reuse the F-orthonormal vectors $\mathcal{V}_m$ and we expand it to $\mathcal{V}_{m+1} = [\mathcal{V}_m, V_{\text{new}}]$, where V_{new} is obtained normalizing the residual as follows

$$V_{\text{new}} = \frac{P_1}{\|P_1\|_F}, \quad P_1 = \mathcal{H}(X_1) - E_1$$

Solving the X-problem

- We then solve the following minimization problem

$$\min_{X \in \text{span}(\mathcal{V}_{m+1})} \|P_k - \mathcal{H}(X)\|_F ,$$

by the updated version of the global QR decomposition (R. Bouyouli, K. Jbilou, R. Sadaka and H. Sadok, J. Comput. Appl. Math, 2006)

- We can then continue with $\mathcal{H}(X) = E_k$, $k = 2, 3, \ldots$ in a similar manner

Theorem

Assume that (X_*, R_*, Y_*) *is a solution of constrained* TV/L1 *problem. The sequence* (X_k, R_k, Y_k) *generated by ALM satisfies*

1. $\lim_{k \to +\infty} F(R_k) + G(Y_k) = F(R_*) + G(Y_*)$,
2. $\lim_{k \to +\infty} \|DX_k - Y_k\|_F = 0$,
3. $\lim_{k \to +\infty} \|\mathcal{A}(X_k) - R_k\|_F = 0$.

Numerical experiments : TV/L2 model for Gaussian blur and additive noise.

- To determine the effectiveness of our solution methods, we evaluate the Signal-to-Noise Ratio (SNR) defined by

$$\text{SNR}(X_k) = 10\log_{10}\frac{\|\widehat{X} - E(\widehat{X})\|_F^2}{\|X_k - \widehat{X}\|_F^2}$$

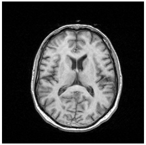
  where $E(\widehat{X})$ denotes the mean gray-level of the uncontaminated image $\widehat{X}$.

- We aslo evaluate the relative error

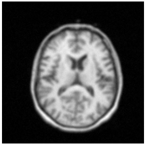
$$\text{Re} = \frac{||\widehat{X} - X_k||_F}{||\widehat{X}||_F}$$

Numerical experiments : TV/L2 model for Gaussian blur and additive noise.

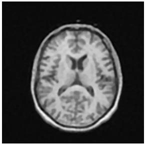
	Parameters		TV/L2		
Noise %	μ	β	Iter	SNR	time
0.001	0.0001	0.1	53	18.32	9.30
0.01	0.001	30	20	15.70	2.65



(a) Original image



(b) Blurred and noisy

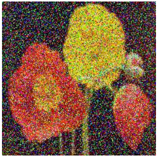


(c) Rest(SNR=18.32)

Numerical experiments : TV/L1 model for Gaussian blur and salt-and-pepper noise.

	Parameters			TV/L1		
Noise %	μ	β	ρ	Iter	SNR	time
10	0.1	80	5	13	24.66	9.01
20	0.125	80	5	17	23.00	12.64
30	0.125	80	5	19	20.90	13.35

(d) Original image



(e) Blurred and noisy

(f) Restored(SNR=21)

Numerical experiments : TV/L2 model for phillips problem

Consider the Fredholm integral equation

$$\int\int_{\Omega} K(x,y,s,t)f(s,t)dsdt = g(x,y), \qquad (x,y) \in \Omega,$$

where $\Omega = [-6,6] \times [-6,6]$ and. Let the kernel be given by

$$K(x,y,s,t) = k_1(x,s)k_1(y,t), \quad (x,y) \in \Omega, \quad (s,t) \in \Omega,$$

and define

$$g(x,y) = g_1(x)g_1(y), \quad f(s,t) = f_1(s)f_1(t)$$

where

$$f_1(s) := \begin{cases} 1 + \cos(\frac{\pi}{3}s), & |s| \leq \frac{\pi}{3}, \\ 0, & \text{otherwise.} \end{cases}$$

$$k_1(s,x) := f_1(s-x)$$

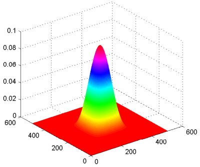
$$g_1(s) := (6 - |s|)\left(1 + \frac{1}{2}\cos\left(\frac{\pi}{3}s\right)\right) + \frac{9}{2\pi}\sin\left(\frac{\pi}{3}|s|\right).$$

Numerical experiments : phillips problem

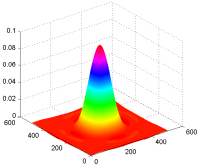
1. We use the Matlab code `phillips` to discretize and to obtain H_1 and H_2 both of size 1500×1500 of ill-determined rank.
2. $\kappa(H_1) = \kappa(H_2) = 2 \times 10^{18}$
3. We add an error-matrix E to obtain a specific noise level $\nu = \frac{||E||_F}{||\hat{B}||_F}$
4. The linear discrete ill-posed problem is now of the form

$$H_1 X H_2^T = \widehat{B} + E$$

	Parameters		TV/$L2$		
Noise %	μ	β	Iter	Re	time
0.001	0.0001	0.1	12	4.01×10^{-2}	9.05
0.01	0.001	30	13	3.99×10^{-2}	9.63
0.1	0.1	40	15	4.07×10^{-2}	10.94



(g) Exact solution



(h) Approximated solution

Thank you for your attention