Flexible Golub-Kahan Factorization for Linear Inverse Problems

Silvia Gazzola

Abstract

Discrete linear inverse problems arising in many applications in Science and Engineering are formulated as the solution of large-scale linear systems of equations of the form

$$\mathbf{A}\mathbf{x}_{\text{true}} + \mathbf{n} = \mathbf{b}\,, \qquad (1)$$

where the discretized forward operator $\mathbf{A} \in \mathbb{R}^{m\times n}$ is large-scale with ill-determined rank, and $\mathbf{n} \in \mathbb{R}^m$ are some unknown perturbations (noise) affecting the available data $\mathbf{b} \in \mathbb{R}^m$. In this setting, in order to recover a meaningful approximation of $\mathbf{x}_{\text{true}} \in \mathbb{R}^n$, one should regularize (1).

In this talk we consider variational regularization methods that compute an approximation $\mathbf{x}_{\text{reg}}$ of $\mathbf{x}_{\text{true}}$ as

$$\mathbf{x}_{\text{reg}} = \arg\min_{\mathbf{x}\in\mathbb{R}^n} \|\mathbf{R}(\mathbf{A}\mathbf{x}-\mathbf{b})\|_p^p + \lambda\|\mathbf{L}\mathbf{x}\|_q^q\,, \quad \text{where} \quad \lambda \geq 0,\ p,\ q > 0,\ \mathbf{R} \in \mathbb{R}^{m\times m}\ \mathbf{L} \in \mathbb{R}^{l\times n}. \qquad (2)$$

In the above formulation, when $p = q = 2$, many standard numerical linear algebra tools can be employed to approximate $\mathbf{x}_{\text{reg}}$: these include the SVD of $\mathbf{A}$ (when $\mathbf{A}$ has some exploitable structure and $\mathbf{L}$ is the identity), early termination of Krylov solvers for (1) (when $\lambda = 0$), and hybrid projection methods. We refer to [2] for a recent survey of these strategies. However, by properly setting p, $q \neq 2$, better approximations of $\mathbf{x}_{\text{true}}$ can be obtained in many scenarios, including: when the noise $\mathbf{n}$ is not Gaussian, nor white, and/or when wanting to enforce sparsity onto $\mathbf{L}\mathbf{x}_{\text{reg}}$ (e.g., in the compressive sensing framework, when $\mathbf{A}$ is heavily underdetermined). Although many classes of well-established optimization methods are usually employed to handle the non-smooth and possibly non-convex instances of (2), in the last decades a number of new solvers based on 'non-standard' (such as flexible [1, 4] or generalized [5]) Krylov methods have been successfully considered for this purpose; see also [3, 7]. Even though the common starting point of such 'non-standard' Krylov solvers is the reformulation of a smoothed version of (2) as an iteratively reweighted least squares problem, flexible Krylov methods for $p = 2$ are typically more efficient and stable than generalized Krylov methods, while the latter can handle also the $p \neq 2$ case and many options for $\mathbf{L}$.

This talk introduces new solvers for (2), based on a new flexible Golub-Kahan factorization of the kind

$$\widehat{\mathbf{A}}\mathbf{Z}_k = \mathbf{U}_{k+1}\bar{\mathbf{M}}_k\,, \quad \widehat{\mathbf{A}}^\top\mathbf{Y}_{k+1} = \mathbf{V}_{k+1}\mathbf{T}_{k+1}\,,$$

where: $\mathbf{U}_{k+1} \in \mathbb{R}^{m\times(k+1)}$ and $\mathbf{V}_k \in \mathbb{R}^{n\times k}$ have orthonormal columns $\mathbf{u}_i$ ($i = 1,\dots,k+1$) and $\mathbf{v}_i$ ($i = 1,\dots,k$), respectively; $\mathbf{Z}_k = [\mathbf{L}_1^\dagger\mathbf{v}_1,\dots,\mathbf{L}_k^\dagger\mathbf{v}_k]$, $\mathbf{Y}_{k+1} = [\mathbf{R}_1^\dagger\mathbf{u}_1,\dots,\mathbf{R}_{k+1}^\dagger\mathbf{u}_{k+1}]$; $\bar{\mathbf{M}}_k \in \mathbb{R}^{(k+1)\times k}$ is upper Hessenberg and $\mathbf{T}_{k+1} \in \mathbb{R}^{(k+1)\times(k+1)}$ is upper triangular; $k \ll \min\{m,n\}$. The ith approximate solution of $\mathbf{x}_{\text{reg}}$ in (2) is defined as

$$\mathbf{x}_i = \mathbf{Z}_i \arg\min_{\mathbf{s}\in\mathbb{R}^i} \|f(\mathbf{T}_{i+1},\bar{\mathbf{M}}_i)\mathbf{s} - \mathbf{c}_i\|_2^2 + \lambda_i\|\mathbf{S}_i\mathbf{s}\|_2^2\,,$$

where the regularization parameter λ_i is adaptively set, $\mathbf{S}_i \in \mathbb{R}^{i\times i}$ is a regularization matrix for the projected variable $\mathbf{s}$, $\mathbf{c}_i$ is a projected right-hand side, and f compactly denotes products and/or sums of (possibly slight modifications and transposes of) both matrices $\mathbf{T}_{i+1}$ and $\bar{\mathbf{M}}_i$; different choices of f and $\mathbf{S}_i$ define different solvers. Note that $\mathbf{R}_i^\dagger$ and $\mathbf{L}_i^\dagger$ act as variable 'preconditioners' for the constraint and solution subspaces, respectively; their role is to enforce iteration-dependent information useful for a successful regularization. Different choices of $\widehat{\mathbf{A}}$, $\mathbf{R}_i^\dagger$ and $\mathbf{L}_i^\dagger$ allow to handle different instances of (2). Namely:

(a) $\widehat{\mathbf{A}} = [\mathbf{A}^\top, \mathbf{L}^\top]^\top$, $\mathbf{R}_i^\dagger = \text{diag}(\mathbf{I}, \lambda_i \mathbf{I})$ and $\mathbf{L}_i^\dagger = \mathbf{I}$ solves Tikhonov problems in general form in the 2-norm, with adaptive regularization parameter choice strategy; this provides an alternative to the generalized Krylov method in [6].

(b) $\widehat{\mathbf{A}} = \mathbf{A}$, $\mathbf{R}_i^\dagger = \mathbf{I}$ and $\mathbf{L}_i^\dagger = \text{diag}(g_q^{-1}(\mathbf{x}_{i-1}))$ (where g_q is a function that depends on the q-norm and is applied entry-wise) solves the so-called $\ell_2 - \ell_q$ regularized problem, with adaptive regularization parameter choice strategy; this coincides with the basic version of the method in [1] (and can be reformulated to cover all the options in [1]).

(c) $\widehat{\mathbf{A}} = [\mathbf{A}^\top, \mathbf{L}^\top]^\top$, $\mathbf{R}_i^\dagger = \text{diag}(g_p(\mathbf{R}(\mathbf{A}\mathbf{x}_{i-1} - \mathbf{b})), \lambda_i g_q(\mathbf{L}\mathbf{x}_{i-1}))$ (where, similarly to g_q, g_p is a function that depends on the p-norm and is applied entry-wise) and $\mathbf{L}_i^\dagger = \mathbf{I}$ solves the so-called $\ell_p - \ell_q$ regularized problem, with adaptive regularization parameter choice strategy; this extends the methods in [1] and provides an alternative to the generalized Krylov method in [5]. As a particular case, setting $\lambda_i = 0$, $i = 1, 2, \ldots$ solves a p-norm residual minimization problem.

The new solvers are theoretically analyzed by providing optimality properties and by studying the effect of variations in $\mathbf{R}_i^\dagger$ and $\mathbf{L}_i^\dagger$ on their convergence. The new solvers can efficiently be applied to both underdetrmined and overdetermined problems, and successfully extend the current flexible Krylov solvers to handle different matrices $\mathbf{R}$ (typically the inverse square root of the noise covariance matrix), as well as regularization matrices $\mathbf{L}$ whose $\mathbf{A}$-weighted generalized pseudo-inverse cannot be cheaply computed.

Numerical experiments on inverse problems in imaging, such as deblurring and computed tomography, show that the new solvers are competitive with other state-of-the-art nonsmooth and non-convex optimization methods, as well as generalized Krylov methods.

References

[1] J. Chung, S. Gazzola. Flexible Krylov Methods for ℓ_p Regularization. *SIAM J. Sci. Comput.*, 41(5), pp. S149–S171, 2019.

[2] J. Chung, S. Gazzola. Computational Methods for Large-Scale Inverse Problems: A Survey on Hybrid Projection Methods. *SIAM Review*, 66(2), pp. 205–284, 2024.

[3] J. Cornelis, W. Vanroose. Projected Newton method for noise constrained ℓ_p regularization. *Inverse Problems*, 36 125004, 2020.

[4] S. Gazzola, J. G. Nagy, and M. Sabatè Landman. Iteratively reweighted FGMRES and FLSQR for sparse reconstruction. *SIAM J. Sci. Comput.*, 43, pp. S47–S69, 2021.

[5] G. Huang, A. Lanza, S. Morigi, L. Reichel, F. Sgallari. Majorization–minimization generalized Krylov subspace methods for $\ell_p - \ell_q$ optimization applied to image restoration. *BIT Numerical Mathematics*, 57(2), pp. 351–378, 2017.

[6] J. Lampe, L. Reichel, H. Voss. Large-scale Tikhonov regularization via reduction by orthogonal projection. *Linear Algebra Appl.*, 436(8), pp. 2845–2865, 2012.

[7] M. Sabate Landman, J. Chung. Flexible Krylov Methods for Group Sparsity Regularization. *Physica Scripta*, 2024.