Regularization Parameter Estimation: Stabilization of LSQR Algorithms by Iterative Reweighting for Inversion of $3D$ Gravity Data

Rosemary Renaut[1] Saeed Vatankhah[2]

1: School of Mathematical and Statistical Sciences, Arizona State University

2: University of Tehran

SIAM Imaging Sciences: May 2016

Outline

Motivation: Large Scale Gravity Inversion

Parameter estimation on the projected problem
- UPRE is a good estimator [RVA15]
- Identifying the weight parameter in the GCV [CNO08]

Identifying the optimal Subspace
- Appearance of Noise in the Subspace [HPS09]
- Minimization of the GCV for the truncated SVD [CKO15]

Simulations: Two dimensional Examples

Iteratively Reweighted Regularization [LK83]

Inversion of undersampled gravity data

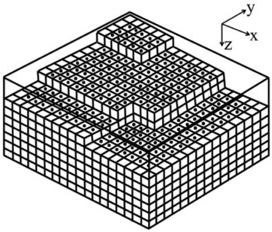
Conclusions

Motivation: $3D$ Gravity Inversion

Observation point $\mathbf{r} = (x, y, z)$

Vertical gravitational attraction $g(\mathbf{r})$

$$g(\mathbf{r}) = \Gamma \int_{d\Omega} \varrho(\mathbf{r}') \frac{\mathbf{r}' - \mathbf{r}}{|\mathbf{r}' - \mathbf{r}|^3} d\Omega'$$



Density $\varrho(\mathbf{r}')$ at $\mathbf{r}' = (x', y', z')$

Newton gravitational constant: Γ

Aim: Given surface observations g_{ij} find volume density ϱ_{ijk}

Gravity Measurements g_{ij} $\mathbf{b}$ on surface at m cells.

Density ϱ_{ijk} $\mathbf{x}$ on volume of n cells.

Projection matrix $A \in \mathbb{R}^{m \times n}$

Linear System $A\mathbf{x} \approx \mathbf{b}$:

- Severely Underdetermimed: $m \ll n$
- Noise contamination $\mathbf{b} = \mathbf{b}_{\text{true}} + \boldsymbol{\eta}$
- Ill-posed: cond(A) large
- Relatively Large: e.g. $m = 4588$, $n = 100936$

Tikhonov Regularization:

$$\mathbf{x}(\lambda) = \underset{\mathbf{x} \in \mathbb{R}^n}{\operatorname{argmin}} \{\|A\mathbf{x} - \mathbf{b}\|^2_{W_{\boldsymbol{\eta}}} + \lambda^2 \|L(\mathbf{x} - \mathbf{x}_0)\|^2_2\}$$

Mapping L defines basis for $\mathbf{x}$ with prior $\mathbf{x}_0$

Weighting $W_{\boldsymbol{\eta}} = C_{\boldsymbol{\eta}}^{-1}$, $\|\mathbf{y}\|_{W_{\boldsymbol{\eta}}} = \mathbf{y}^T W_{\boldsymbol{\eta}} \mathbf{y}$. Whitens noise in $\mathbf{b}$.

Requires automatic estimation of λ^{opt}

Large Scale Problems use Iterative Solve (Notation)

LSQR Let $\beta_1 := \|\mathbf{b}\|_2$, and $\mathbf{e}_1^{(t+1)}$ first column of I_{t+1}
Generate, lower bidiagonal $B_t \in \mathcal{R}^{(t+1)\times t}$, column orthonormal $H_{t+1} \in \mathcal{R}^{m\times(t+1)}$, $G_t \in \mathcal{R}^{n\times t}$

$$AG_t = H_{t+1}B_t, \quad \beta_1 H_{t+1}\mathbf{e}_1^{(t+1)} = \mathbf{b}.$$

Projected Problem on projected space:

$$\mathbf{w}_t(\zeta_t) = \underset{\mathbf{w}\in\mathcal{R}^t}{\operatorname{argmin}}\{\|B_t\mathbf{w} - \beta_1\mathbf{e}_1^{(t+1)}\|_2^2 + {\zeta_t}^2\|\mathbf{w}\|_2^2\}.$$

Projected Solution depends on ${\zeta_t}^{\text{opt}}$

$$\mathbf{x}_t({\zeta_t}^{\text{opt}}) = G_t\mathbf{w}_t({\zeta_t}^{\text{opt}})$$

$\boxed{\textbf{Generally: } {\zeta_t}^{\text{opt}} \neq \lambda^{\text{opt}}}$

Regularization of the LSQR solution: Questions

(i) Determine optimal t The choice of the subspace impacts the regularizing properties of the iteration: For large t noise due to numerical precision and data error enters the projected space.

(ii) Determine optimal ζ_t How do regularization parameter techniques translate to the projected problem?

(iii) Relation optimal ζ_t and optimal λ Given t how well does optimal ζ_t for projected space yield optimal λ for full space, or when is this the case?

Calculating Unbiased Predictive Risk using $\mathbf{w}_t(\lambda)$[RVA15]

Residual: $\mathbf{R}^{\text{full}}(\mathbf{x}_t) = A\mathbf{x}_t - \mathbf{b}.$

Influence Matrix $A(\lambda) = A(A^TA + \lambda^2 I)^{-1}A^T$

UPRE : Full problem

$$\lambda^{\text{opt}} = \underset{\lambda}{\operatorname{argmin}}\{\|\mathbf{R}^{\text{full}}(\mathbf{x}_t(\lambda))\|_2^2 + 2\,\mathrm{Tr}(A(\lambda)) - m\} = \underset{\lambda}{\operatorname{argmin}}\{U^{\text{full}}(\lambda)\}.$$

Using the projected solution for parameter λ and

$$\mathrm{Tr}\left((AG_t)(\lambda)\right) = \mathrm{Tr}\left(B_t(\lambda)\right)$$

$$\begin{aligned} U^{\text{full}}(\lambda) &= \|\left((AG_t)(\lambda) - I_m\right)\mathbf{b}\|_2^2 + 2\,\mathrm{Tr}\left((AG_t)(\lambda)\right) - m \\ &= \|\beta_1(B_t(\lambda) - I_{t+1})\mathbf{e}_1^{t+1}\|_2^2 + 2\,\mathrm{Tr}(B_t(\lambda)) - m \end{aligned}$$

$\boxed{\lambda^{\text{opt}} \text{ for } U^{\text{full}}(\lambda) \text{ can be estimated for projected problem}}$

Deriving UPRE for the projected problem

Is λ^{opt} relevant to ${\zeta_t}^{\text{opt}}$ for the projected problem?

Noise in the right hand side For $\mathbf{b} = \mathbf{b}^{\text{true}} + \boldsymbol{\eta}$, $\boldsymbol{\eta} \sim \mathbb{N}(0, I_m)$

$$\beta_1 \mathbf{e}_1^{t+1} = H_{t+1}^T \mathbf{b} = H_{t+1}^T \mathbf{b}^{\text{true}} + H_{t+1}^T \boldsymbol{\eta}.$$

Noise in projected right hand side $\beta_1 \mathbf{e}_1^{t+1}$, satisfies

$$H_{t+1}^T \boldsymbol{\eta} \sim \mathbb{N}(0, I_{t+1})$$

Immediately

$$\begin{aligned} U^{\text{proj}}(\zeta_t) &= \|\beta_1 (B_t(\zeta_t) - I_{t+1}) \mathbf{e}_1^{(t+1)}\|_2^2 + 2\,\mathrm{Tr}(B_t(\zeta_t)) - (t+1) \\ &= U^{\text{full}}(\zeta_t) + m - (t+1). \end{aligned}$$

Minimizer of $U^{\text{proj}}(\zeta_t)$ is minimizer of $U^{\text{full}}(\zeta_t)$

${\zeta_t}^{\mathrm{opt}}$ calculated for projected problem may not yield λ^{opt} on full problem

${\zeta_t}^{\mathrm{opt}}$ depends on t, λ^{opt} depends on $m^* =: \min(m,n)$

Trace Relations By linearity and cycling.

$$\mathrm{Tr}(A(\lambda)) = \mathrm{Tr}(A(A^TA+\lambda^2 I_n)^{-1}A^T) = m^* - \lambda^2 \sum_{i=1}^{m^*}(\sigma_i^2+\lambda^2)^{-1}$$

$$\mathrm{Tr}(B_t(\zeta_t)) = t - {\zeta_t}^2 \sum_{i=1}^{t}(\gamma_i^2+{\zeta_t}^2)^{-1}.$$

Approximate Singular Values IF $\sigma_i \approx \gamma_i$, $1 \le i \le t^* \le t$,
$\sigma_{t^*}^2/(\sigma_{t^*}^2+\lambda^2) >> \sigma_i^2/(\sigma_i^2+\lambda^2) \approx 0$, $i > t^*$,

$$\mathrm{Tr}(A(\lambda)) \approx \mathrm{Tr}(B_{t^*}(\lambda)) + \sum_{i=t^*+1}^{m^*} \sigma_i^2(\sigma_i^2+\lambda^2)^{-1} \approx \mathrm{Tr}(B_{t^*}(\lambda)).$$

If t^* approx numerical rank A, ${\zeta_t}^{\mathrm{opt}} \approx \lambda^{\mathrm{opt}}$ for $\mathcal{K}_{t^*}(A^TA, A^T\mathbf{b})$

Other Estimation Techniques for the Projected Problem

GCV: [CNO08] *weighted* GCV is introduced for $\omega > 0$.

$$G^{\text{proj}}(\zeta_t, \omega) = \frac{\|\mathbf{R}^{\text{proj}}(\mathbf{w}_t(\zeta_t))\|_2^2}{\left(\operatorname{Tr}(\omega B_t(\zeta_t) - I_{t+1})\right)^2}, \quad G(\lambda) = G^{\text{proj}}(\lambda, 1).$$

Optimal Analysing as for UPRE: $\omega = \frac{t+1}{m} < 1$.

Discrepancy Principle Seek λ such that $\|\mathbf{R}^{\text{full}}(\mathbf{x}(\lambda))\|_2^2 = \delta \approx m$. To avoid over smoothing: $\delta = \upsilon m$, $\upsilon > 1$

Discrepancy for the Projected Problem Seek ζ_t such that

$$\|\mathbf{R}^{\text{proj}}(\mathbf{w}_t(\zeta_t))\|_2^2 \approx \delta^{\text{proj}} = \upsilon(t+1).$$

We do not obtain in these cases $\zeta_t{}^{\text{opt}} \approx \lambda^{\text{opt}}$

Identifying optimal subspace size t

Noise revealing function: [HPS09] suppose θ_j and β_j on diagonal and sub diagonal of B_t

$$\rho(t) = \prod_{j=1}^{t} (\theta_j / \beta_{j+1})$$

Optimal t is given by (for user determined $t^{\min}$)

$$t^{\mathrm{opt}-\rho} = \min\{\operatorname*{argmax}_{t > t^{\min}}(\rho(t))\} + \text{step}$$

step$= 2$ is to assure that noise has entered the entries in $\rho(t)$ and hence the basis.

$t^{\min}$ is chosen based on examination of $\rho(t)$.

Only useful if discrete Picard condition holds [HPS09].

Identifying optimal subspace size t:

Minimization of the GCV for the truncated SVD of B_{t^*} [CKO15]

Projected subspace size is defined to be t^*

$$\mathcal{G}(t, t^*) = \frac{t^*}{(t^* - t)^2} \sum_{t+1}^{t^*} |\mathbf{u}_i^T \mathbf{b}|^2.$$

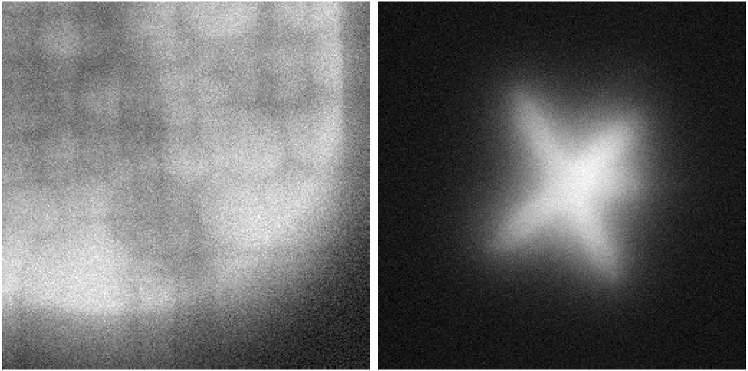
Optimal t is given by

$$t^{\text{opt}-\mathcal{G}} = \operatorname*{argmin}_{t} \mathcal{G}(t, t^*)$$

Does not require Picard condition, but $t^{\text{opt}-\mathcal{G}}$ depends on t^*

Application for Two Dimensional Examples

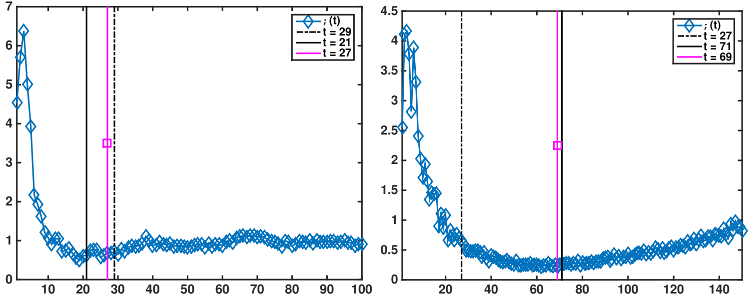
Two dimensional image deblurring [NPP04] Problem size 256×256



(a) `Data` (b) `Data`

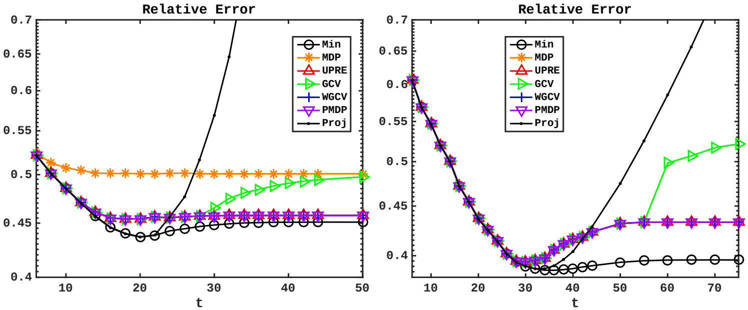
Figure: Data for grain and satellite images with blur and noise level 10%.

Noise Revealing Function $\rho(t)$: comparing $t^{\mathrm{opt}-\rho}$, $t^{\mathrm{opt}-\mathcal{G}}$, $t^{\mathrm{opt}-\min}$



(a) `grain` (b) `satellite`

Figure: $\rho(t)$ using $t^{\min} = 25$. Dashed-dot $t^{\mathrm{opt}-\rho}$, magenta $t^{\mathrm{opt}-\mathcal{G}}$ and black $t^{\mathrm{opt}-\min}$, location of minimum for $\rho(t)$ plus step.

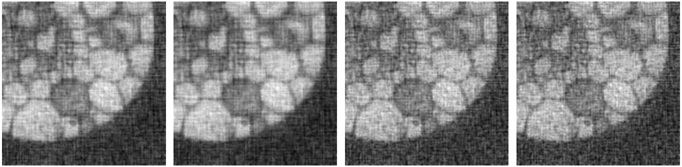


(a) `grain` RE (b) `satellite` RE

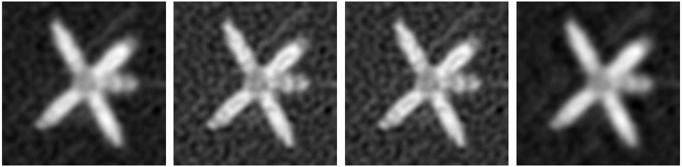
Figure: Relative error (RE) with increasing t. Solid line in each case is solution with projection and without regularization.

UPRE, WGCV and PMDP outperform GCV

Solutions for different t^{opt}: (MIN, $t^{\text{opt}-\text{min}}$, $t^{\text{opt}-\mathcal{G}}$, $t^{\text{opt}-\rho}$) Noise level 10%



(a) MIN 22 (b) 21 (c) 27 (d) 29



(e) MIN 42 (f) 71 (g) 69 (h) 27

Figure: UPRE to find ζ. Solutions obtained for $t^{\text{opt}-\rho}$, $t^{\text{opt}-\text{min}}$ and $t^{\text{opt}-\mathcal{G}}$ and MIN.

Solutions inadequate

Iteratively Reweighted Regularization [LK83]

Minimum Support Stabilizer Regularization operator $L^{(k)}$.

$$(L^{(k)})_{ii} = ((\mathbf{x}_i^{(k-1)} - \mathbf{x}_i^{(k-2)})^2 + \beta^2)^{-1/2} \quad \beta > 0$$

Parameter β ensures $L^{(k)}$ invertible

Invertibility use $(L^{(k)})^{-1}$ as right preconditioner for A

$$(L^{(k)})_{ii}^{-1} = ((\mathbf{x}_i^{(k-1)} - \mathbf{x}_i^{(k-2)})^2 + \beta^2)^{1/2} \quad \beta > 0$$

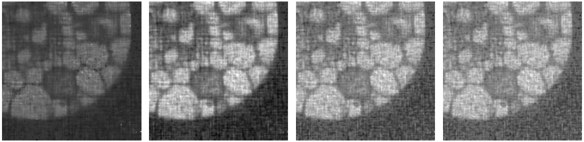
Initialization $L^{(0)} = I$, $\mathbf{x}^{(0)} = \mathbf{x}_0$. (might be $\mathbf{0}$)

Reduced System When $\beta = 0$ and $\mathbf{x}_i^{(k-1)} = \mathbf{x}_i^{(k-2)}$ remove column i, $\hat{A}$ is AL^{-1} with columns removed.

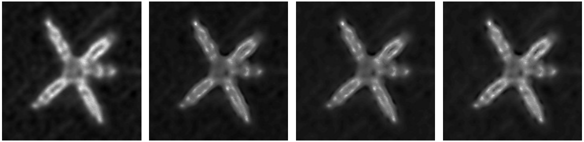
Update Equation Solve $\hat{A}\hat{\mathbf{y}} \approx \mathbf{R} = \mathbf{b} - A\mathbf{x}^{(k-1)}$. With correct indexing set $\mathbf{y}_i = \hat{\mathbf{y}}_i$ if updated, else $\mathbf{y}_i = 0$.

$$\mathbf{x}^{(k)} = \mathbf{x}^{(k-1)} + \mathbf{y}$$

Cost of $L^{(k)}$ is minimal



(a) $19\ k=3$ (b) 21 (c) 27 (d) 29

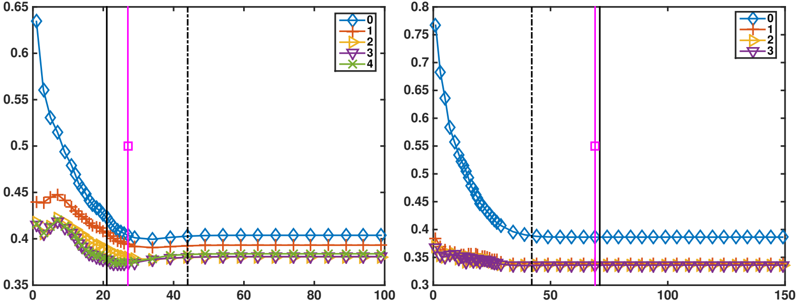


(e) 35 (f) 71 (g) 69 (h) 27

Figure: IRR $k=2$ Grain $k=2$ MIN solution is at $t^{\mathrm{opt-min}}$, show $k=3$.

Solutions are stabilized less dependent on t

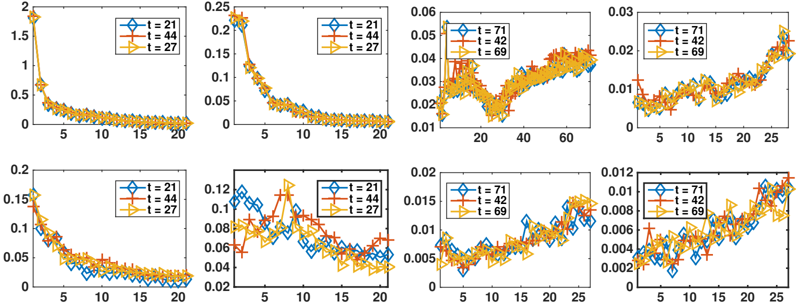
Relative error with k: 5% error using UPRE



(a) Grain RE (b) Satellite

Figure: Relative errors decrease initially with k and then increase. Dashed-dot $t^{\mathrm{opt}-\rho}$, magenta $t^{\mathrm{opt}-\mathcal{G}}$, black $t^{\mathrm{opt}-\min}$.

Noise revealing function $\rho(t)$ with k 5% error



(a) Grain $\rho(t)$ (b) Satellite

Figure: Determining t^{opt} with k for 5% noise using $\rho(t)$. Stopping Critera : Grain $k = 4$ noise enters, use $k = 2$. Satellite $k = 3$ noise enters, use $k = 1$.

Solving the Gravity Inversion problem

Undersampled Gravity inversion $m = 4900$, $n = 98000$

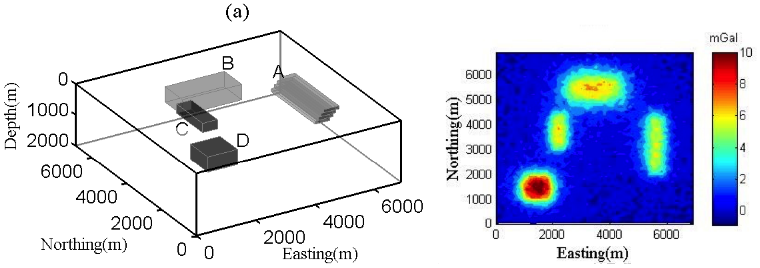


Figure: (a) The perspective view of the model. Four different bodies embedded in an homogeneous background. Densities of A and B are 0.8 g cm^{-3} and C and D are 1 g cm^{-3}; (b) The noise contaminated gravity anomaly due to the model.

Undersampled Gravity inversion

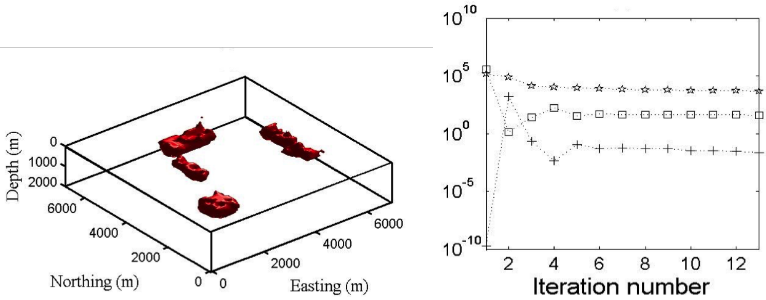


Figure: The reconstructed model with $t = 250$ and the L_1 stabilizer with $\beta^2 = 1.e{-}9$. Data misfit $\star$, the regularization term, $+$, regularization parameter $\square$ with iteration

True Data:

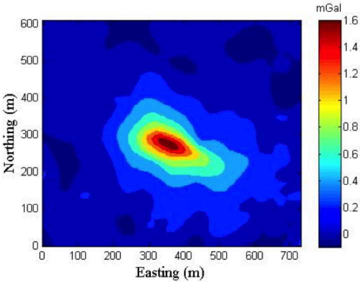


Figure: Residual Anomaly of Mobrun ore body, Noranda, Quebec, Canada.

Reconstructed Model

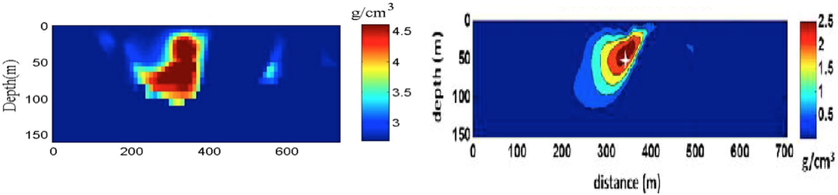


Figure: The reconstructed model with $t = 300$ and the L_1 stabilizer with $\beta^2 = 1.e{-9}$. (a) cross-section at $y = 285$ m and (b) comparison From Ialango et.al (2014)

$3-$D perspective

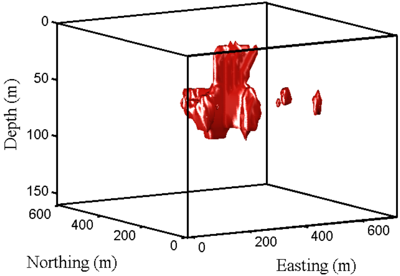


Figure: 3D view of the recovered model, the density cut off is 4 g cm^{-3}.

UPRE/WGCV regularization parameter estimation explained for projected problem.

ζ_t^{opt}, λ^{opt} related across levels

Underdetermined problems are also solved.

Iteratively Reweighted Regularization stabilizes the projected solution

Sensitivity to choice of t^{opt} reduced by IRR

t^{opt} can be estimated using $\rho(t)$, use $t^{\mathrm{opt-min}}$ as independent of other parameters

t^{opt} effectively determines a truncation of the SVD for B_t: use B_t and G_t, but truncated solution.

Some key references

Julianne M. Chung, Misha E. Kilmer, and Dianne P. O'Leary.
A framework for regularization via operator approximation.
SIAM Journal on Scientific Computing, 37(2):B332–B359, 2015.

Julianne Chung, James G Nagy, and DIANNE P O'Leary.
A weighted GCV method for Lanczos hybrid regularization.
Electronic Transactions on Numerical Analysis, 28:149–167, 2008.

Iveta Hnětynková, Martin Plešinger, and Zdeněk Strakoš.
The regularizing effect of the Golub-Kahan iterative bidiagonalization and revealing the noise level in the data.
BIT Numerical Mathematics, 49(4):669–696, 2009.

B. J. Last and K. Kubik.
Compact gravity inversion.
GEOPHYSICS, 48(6):713–721, 1983.

James G. Nagy, Katrina Palmer, and Lisa Perrone.
Iterative methods for image deblurring: A Matlab object-oriented approach.
Numerical Algorithms, 36(1):73–93, 2004.

R. A. Renaut, S. Vatankhah, and V. E. Ardestani.
Hybrid and iteratively reweighted regularization by unbiased predictive risk and weighted gcv for projected systems, 2015.
submitted and at: `http://arxiv.org/abs/1509.00096`.