Computational Issues Relating to Inversion of Practical Data: Where is the Uncertainty? Can we solve $A\mathbf{x} = \mathbf{b}$?

Rosemary Renaut
http://math.asu.edu/~rosie

Bridging the Gap?
Oct 2, 2012

- Yuen: Solve large scale problems efficiently - find solutions and features - edge resolution
- Reusable kernels -effective libraries - better utilize HPC/GPU environments
- How can you “differentiate” your data
- Solve for solution and features together
- Does this all mesh together successfully?
- How does analysis of computation fit?

Discussion

- Yuen: Solve large scale problems efficiently - find solutions and features - edge resolution
- Reusable kernels -effective libraries - better utilize HPC/GPU environments
- How can you “differentiate” your data
- Solve for solution and features together
- Does this all mesh together successfully?
- How does analysis of computation fit?

Discussion

- Yuen: Solve large scale problems efficiently - find solutions and features - edge resolution
- Reusable kernels -effective libraries - better utilize HPC/GPU environments
- How can you “differentiate” your data
- Solve for solution and features together
- Does this all mesh together successfully?
- How does analysis of computation fit?

Discussion

- Yuen: Solve large scale problems efficiently - find solutions and features - edge resolution
- Reusable kernels -effective libraries - better utilize HPC/GPU environments
- How can you “differentiate” your data
- Solve for solution and features together
- Does this all mesh together successfully?
- How does analysis of computation fit?

Discussion

- Yuen: Solve large scale problems efficiently - find solutions and features - edge resolution
- Reusable kernels -effective libraries - better utilize HPC/GPU environments
- How can you “differentiate” your data
- Solve for solution and features together
- Does this all mesh together successfully?
- How does analysis of computation fit?

Discussion

- Yuen: Solve large scale problems efficiently - find solutions and features - edge resolution
- Reusable kernels -effective libraries - better utilize HPC/GPU environments
- How can you “differentiate” your data
- Solve for solution and features together
- Does this all mesh together successfully?
- How does analysis of computation fit?

Goals

- Numerical instability?
- Impact on solving ill-posed problems?
- What problem did we actually solve?
- Example - linear case least squares?
- Little mathematics - many images
- Nonlinear LS use multiple LLS cases

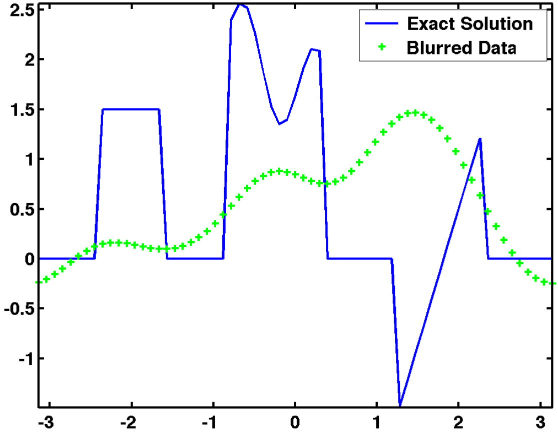
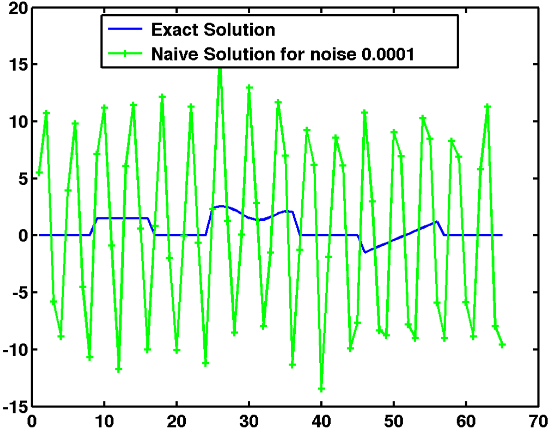


Figure: Standard blurred signal, desire to find signal and its features

Inverse Problem given model A, Condition $1.8679e+05$data b find x, noise $.0001$

$x = A^{-1}b$ (inverse crime) - need an alternative feasible solution



Tikhonov Regularized Solutions $\mathbf{x}(\lambda)$ and derivative $L\mathbf{x}$ for changing λ

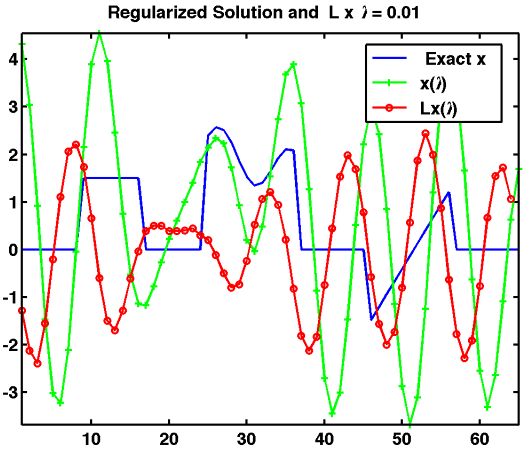
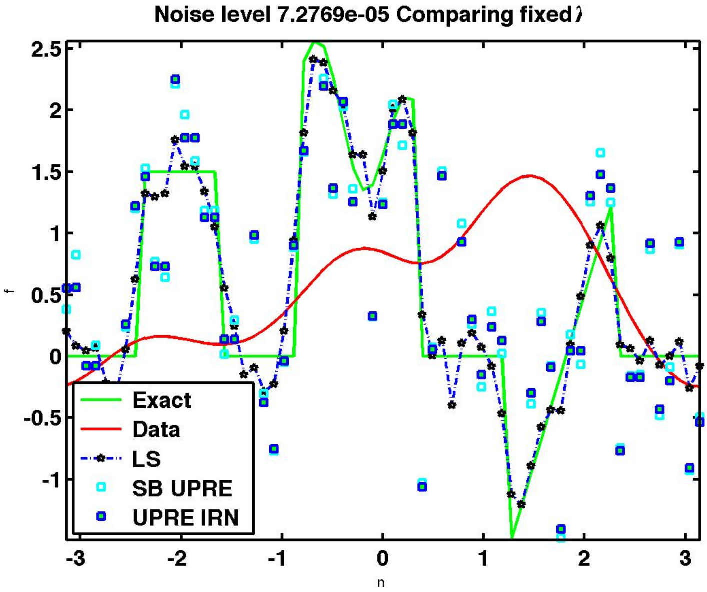
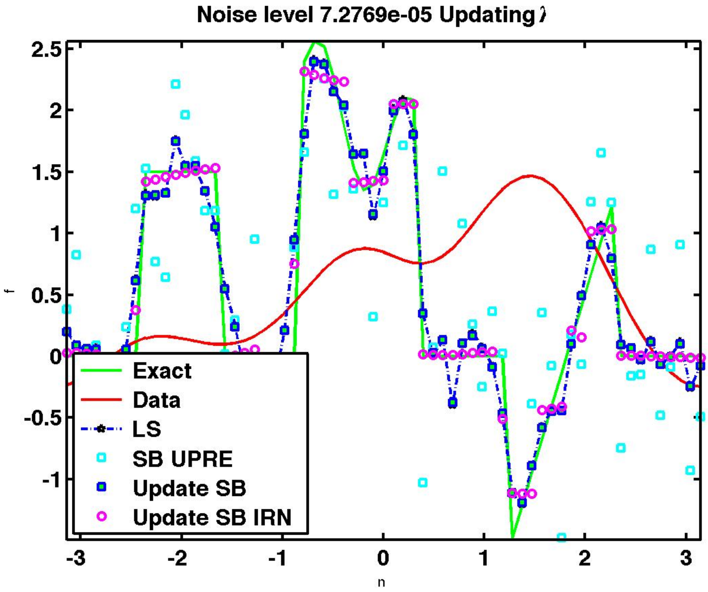


Figure: We cannot capture $L\mathbf{x}$ (red) from the solution (green): Notice that $\|L\mathbf{x}\|$ decreases as λ increases

Example TV Solution: 1D: No Updates for the parameters



Example TV Solution: 1D Updates for the parameters



TV Solutions - Solve for both $L\mathbf{x}$ and $\mathbf{x}$ concurrently: Split Bregman Formulation (Goldstein and Osher, 2009)

Introduce $\mathbf{d} \approx L\mathbf{x}$ and let $R(\mathbf{x}) = \frac{\lambda^2}{2}\|\mathbf{d} - L\mathbf{x}\|_2^2 + \mu\|\mathbf{d}\|_1$

$$(\mathbf{x}, \mathbf{d})(\lambda, \mu) = \arg\min_{\mathbf{x},\mathbf{d}}\{\frac{1}{2}\|A\mathbf{x} - \mathbf{b}\|_2^2 + \frac{\lambda^2}{2}\|\mathbf{d} - L\mathbf{x}\|_2^2 + \mu\|\mathbf{d}\|_1\}$$

Alternating minimization separates steps for $\mathbf{d}$ from $\mathbf{x}$

Various versions of the iteration can be defined. Fundamentally:

$$\mathrm{S1}: \mathbf{x}^{(k+1)} = \arg\min_{\mathbf{x}}\{\frac{1}{2}\|A\mathbf{x} - \mathbf{b}\|_2^2 + \frac{\lambda^2}{2}\|L\mathbf{x} - (\mathbf{d}^{(k+1)} - \mathbf{g}^{(k)})\|_2^2\}$$

$$\mathrm{S2}: \mathbf{d}^{(k+1)} = \arg\min_{\mathbf{d}}\{\frac{\lambda^2}{2}\|\mathbf{d} - (L\mathbf{x}^{(k+1)} + \mathbf{g}^{(k)})\|_2^2 + \mu\|\mathbf{d}\|_1\}$$

$$\mathrm{S3}: \mathbf{g}^{(k+1)} = \mathbf{g}^{(k)} + L\mathbf{x}^{(k+1)} - \mathbf{d}^{(k+1)}.$$

So how does this go?

1. Inverse problem we need regularization
2. For feature extraction we need more than Tikhonov Regularization - e.g. TV
3. The TV iterates over many Tikhonov solutions
4. Both techniques are parameter dependent
5. Moreover the parameters are needed
6. We need to **fully** understand the Tikhonov and ill-posed problems
7. Can we do blackbox solvers?
8. Be careful

Decompositions: SVD $A = U\Sigma V^T$, GSVD $A = UGZ^T$, $L = VMZ^T$

1. A (full column rank): $\mathbf{u}_i$, $\mathbf{v}_i$ left and right singular vectors, σ_i spectral values of A

$$\mathbf{x} = \sum_{i=1}^{n} \frac{\mathbf{u}_i^T \mathbf{b}}{\sigma_i} \mathbf{v}_i$$

   A weighted linear combination of the basis vectors $\mathbf{v}_i$

2. Tikhonov Regularization I is a spectral filtering

$$\mathbf{x}_{\text{filt}} = \sum_{i=1}^{n} \gamma_i \left(\frac{\mathbf{u}_i^T \mathbf{b}}{\sigma_i}\right) \mathbf{v}_i$$

3. Generalized Tikhonov: Generalized SVD expansion

$$\mathbf{x}^{(k+1)} = \sum_{i=1}^{p} \left(\frac{\nu_i \mathbf{u}_i^T \mathbf{b}}{\nu_i^2 + \lambda^2 \mu_i^2} + \frac{\lambda^2 \mu_i \mathbf{v}_i^T \mathbf{h}^{(k)}}{\nu_i^2 + \lambda^2 \mu_i^2} \right) \mathbf{z}_i + \sum_{i=p+1}^{n} (\mathbf{u}_i^T \mathbf{b}) \mathbf{z}_i$$

An example: $n = 32$ Left Singular Vectors and Basis Depend on A

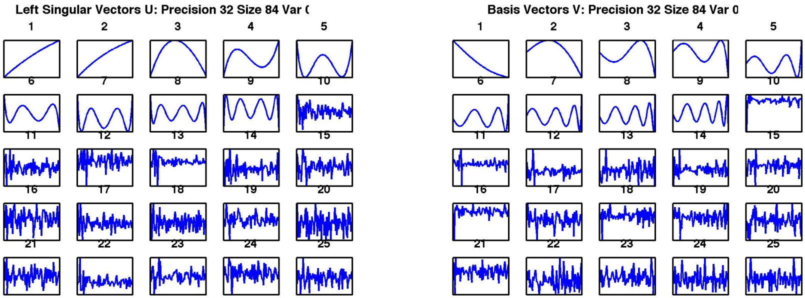


Figure: The first few left singular vectors $\mathbf{u}_i$ and basis vectors $\mathbf{v}_i$. Can we use these basis vectors

Second example: $n = 84$ Left Singular Vectors and Basis Depend on A

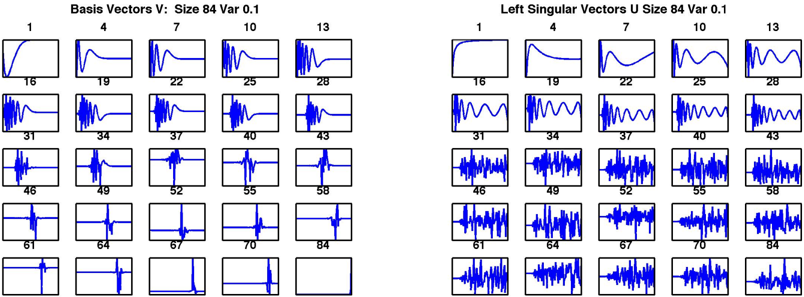


Figure: The first few left singular vectors $\mathbf{u}_i$ and basis vectors $\mathbf{v}_i$. Can we use these basis vectors

Not Always so Bad a case with $n = 64$

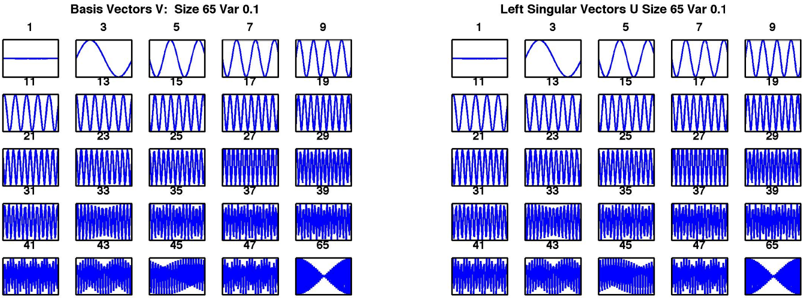


Figure: Left singular vectors $\mathbf{u}_i$ and basis vectors $\mathbf{v}_i$. Can we use these basis vectors

Solutions expressed with respect to a basis

- Given A is the basis a good representation for “true” basis of A?
- Mathematical ”backward” stability - for SVD the basis is ”not too far” from appropriate ”orthogonal” manifold
- But although the basis is ”orthogonal” - it eventually contributes “ noise”
- When we look at residual for $\mathbf{r} = A\mathbf{x} - \mathbf{b}$ is $\|\mathbf{r}\|$ small sufficient? What is “small”
- Need to start looking at the noise entering the residual.
- Need to extend statistical techniques to examining the stability in a new context?

Cumulative Periodogram for the left / right singular vectors

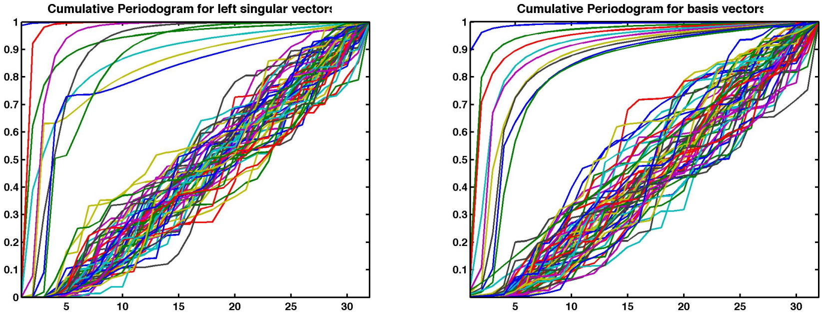


Figure: On left the left singular vectors and on the right the basis vectors $\mathbf{v}$. Low frequency vectors lie above the diagonal and high frequency below the diagonal. White noise follows the diagonal

Second Example Cumulative Periodogram for the left / right singular vectors

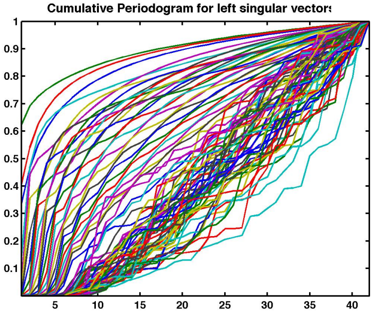


Figure: On left the left singular vectors and on the right the basis vectors $\mathbf{v}$. Low frequency vectors lie above the diagonal and high frequency below the diagonal. White noise follows the diagonal

Cumulative Periodogram for the left / right singular vectors

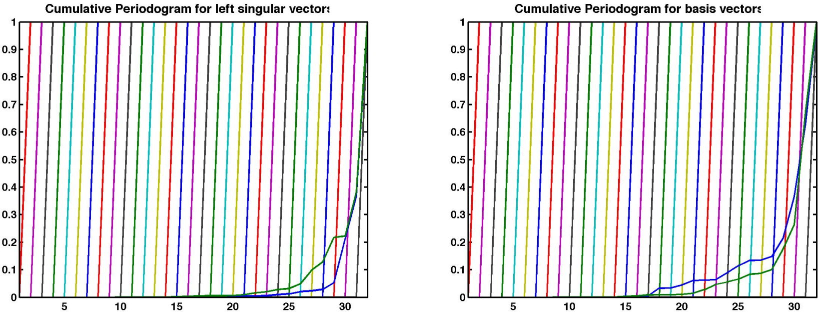


Figure: On left the left singular vectors and on the right the basis vectors $\mathbf{v}$. Low frequency vectors lie above the diagonal and high frequency below the diagonal. White noise follows the diagonal

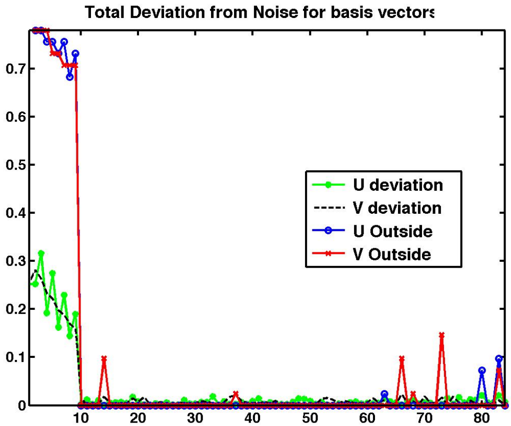


Figure: Calculate the cumulative periodogram, measure deviation from "white noise" line, assess proportion of vector outside Kolmogorov Smirnov test at a 5% confidence level.

Cannot expect to use more than 9 vectors in the expansion for $\mathbf{x}$. Additional terms are contaminated by noise - **independent** of noise in $\mathbf{b}$

Measure Deviation from Straight Line: testing for white noise

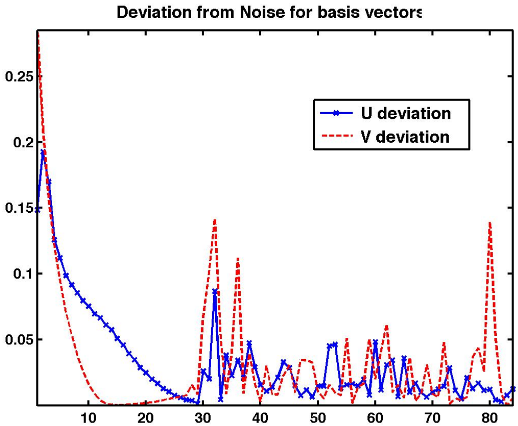


Figure: Calculate the cumulative periodogram, measure deviation from “white noise” line, assess proportion of vector outside Kolmogorov Smirnov test at a 5% confidence level.

Cannot expect to use more than 12 vectors in the expansion for $\mathbf{x}$. Additional terms are contaminated by noise - **independent** of noise in $\mathbf{b}$

Measure Deviation from Straight Line: testing for white noise

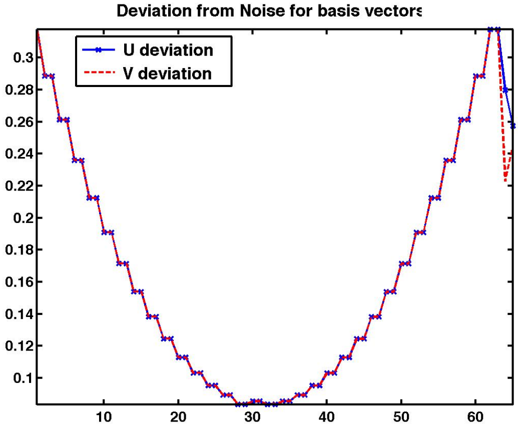


Figure: Calculate the cumulative periodogram, measure deviation from “white noise” line, assess proportion of vector outside Kolmogorov Smirnov test at a 5% confidence level.

Use all but last 2 vectors in the expansion for $\mathbf{x}$. Additional terms are contaminated by noise - **independent** of noise in $\mathbf{b}$

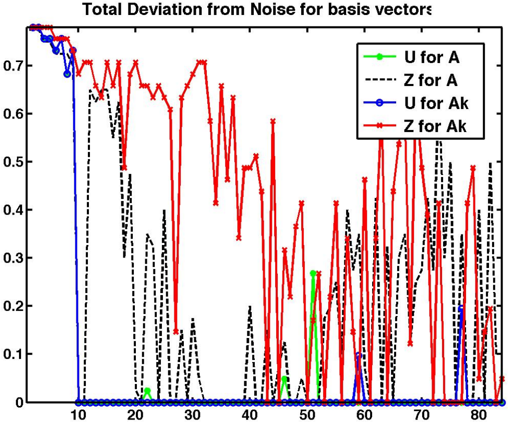


Figure: First Example: It pays to truncate the SVD before finding the GSVD to give the basis for $\mathbf{x}$

Testing for the GSVD Basis

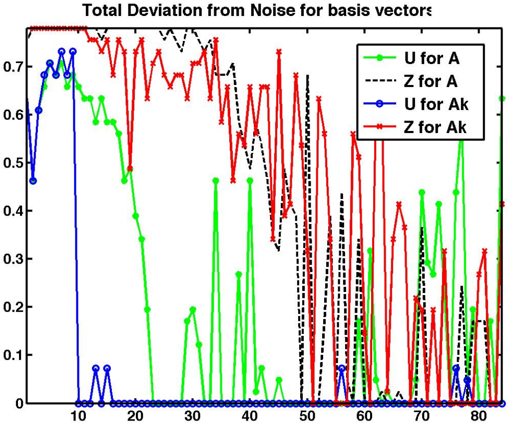


Figure: Second Example: It pays to truncate the SVD before finding the GSVD to give the basis for $\mathbf{x}$

- Libraries need to include additional methods for assessing reliability of the residual
- Even when committing the **inverse crime** we will not achieve the solution if we cannot approximate the basis correctly.
- We need all basis vectors which contain the high frequency terms in order to approximate a solution with high frequency components - e.g. edges.
- Reminder - this is **independent** of the data.
- But is an indication of an ill-posed problem. In this case the data that is modified exhibits in the matrix A decomposition.

- Basis vectors are subject to noise and contaminate the solution **independent** the data.
- Nonlinear least squares use repeated solutions of least squares problems - SVD/GSVD analysis is relevant
- Solutions are obtained in a contaminated basis
- if we do not recognize the noise from the basis- how can we estimate uncertainty due to noise in data
- Analysis here in terms of SVD/GSVD - but equivalent results apply when using Krylov methods, need to examine the basis
- We have to truncate the basis but implies that we will not see high frequency in the solutions
- This is work in progress - lots to discuss

Power Spectrum for detecting white noise : a time series analysis technique

Suppose for a given vector $\mathbf{y}$ that it is a time series indexed by position, i.e. index i.

Diagnostic 1 Does the histogram of entries of $\mathbf{y}$ generate histogram consistent with $\mathbf{y} \sim \mathbb{N}(0, 1)$? (i.e. independent normally distributed with mean 0 and variance 1) Not practical to automatically look at a histogram and make an assessment

Diagnostic 2 Test the expectation that $\mathbf{y}_i$ are selected from a white noise time series. Take the Fourier transform of $\mathbf{y}$ and form cumulative periodogram $\mathbf{z}$ from power spectrum $\mathbf{c}$

$$\mathbf{c}_j = |(\mathrm{dft}(\mathbf{y})_j|^2, \quad \mathbf{z}_j = \frac{\sum_{i=1}^{j} \mathbf{c}_j}{\sum_{i=1}^{q} \mathbf{c}_i}, \quad j = 1, \dots, q,$$

Automatic: Test is the line $(z_j, j/q)$ close to a straight line with slope 1 and length $\sqrt{5}/2$?

Advantages of the SB formulation

Update for $\mathbf{g}$: updates the Lagrange multiplier $\mathbf{g}$

$$\mathrm{S3}: \mathbf{g}^{(k+1)} = \mathbf{g}^{(k)} + L\mathbf{x}^{(k+1)} - \mathbf{d}^{(k+1)}.$$

This is just - a *vector update*

Update for $\mathbf{d}$:

$$\begin{aligned}\mathrm{S2}: \mathbf{d} &= \arg\min_{\mathbf{d}}\{\mu\|\mathbf{d}\|_1 + \frac{\lambda^2}{2}\|\mathbf{d}-\mathbf{c}\|_2^2\}, \quad \mathbf{c} = L\mathbf{x}+\mathbf{g} \\ &= \arg\min_{\mathbf{d}}\{\|\mathbf{d}\|_1 + \frac{\gamma}{2}\|\mathbf{d}-\mathbf{c}\|_2^2\}, \quad \gamma = \frac{\lambda^2}{\mu}.\end{aligned}$$

This is achieved using *soft* thresholding. - in place operation

The Tikhonov Step of the Algorithm

$$S1: \mathbf{x}^{(k+1)} = \arg\min_{\mathbf{x}} \{ \frac{1}{2} \| A\mathbf{x} - \mathbf{b} \|_2^2 + \frac{\lambda^2}{2} \| L\mathbf{x} - (\mathbf{d}^{(k+1)} - \mathbf{g}^{(k)}) \|_2^2 \}$$

Standard least squares update using a Tikhonov regularizer.

$$\boxed{\mathbf{x} = \arg\min_{\mathbf{x}} \{ \frac{1}{2} \| A\mathbf{x} - \mathbf{b} \|_2^2 + \frac{\lambda^2}{2} \| L\mathbf{x} - \mathbf{h} \|_2^2 \}, \quad \mathbf{h} = \mathbf{d} - \mathbf{g}}$$

Disadvantages of the formulation

update for $\mathbf{x}$: A Tikhonov LS update each step

Right hand side deepndent on k

Regularization parameter λ - dependent on k?

Threshold parameter μ - dependent on k?