Ill-posed Questions

These slides are available here: http://illposed.net/illposed_ksu_nov_2013.pdf
The example R program is here: http://goo.gl/dnUeVW

Bryan W. Lewis

blewis@illposed.net

Linear ill-posed problems

$$g(s) = \int k(s,t) f(t) dt, \qquad k \quad \text{smooth},$$

Can be discretized to yield a linear algebraic system of equations:

$$\tilde{b} = Ax$$

$A \in \mathbf{R}^{n \times n}$, $x, \tilde{b} \in \mathbf{R}^n$ (let's assume square for now). The matrix A is usually severely ill-conditioned and of ill-defined numerical rank.

n can be huge in many real-world problems.

The right hand side b may be contaminated by errors that we may know little about:

$$b = \tilde{b} + d, \qquad d \in \mathbf{R}^n.$$

Unfortunately,

$$\begin{aligned} A^{-1}b &= A^{-1}\tilde{b} + A^{-1}d \\ &= x + A^{-1}d \end{aligned}$$

is usually a very poor approximation of x.

Trivial example

$$\begin{bmatrix} 1/2 & 0 \\ 0 & 10^{-6} \end{bmatrix} \begin{bmatrix} 2 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

Let's say that instead of the exact data, we observe a slightly contaminated right-hand side:

$$\begin{bmatrix} 1.01 \\ 0.01 \end{bmatrix}$$

Notice that the contaminated solution is really different from the uncontaminated one:

$$\begin{bmatrix} 1/2 & 0 \\ 0 & 10^{-6} \end{bmatrix}^{-1} \begin{bmatrix} 1.01 \\ 0.01 \end{bmatrix} = \begin{bmatrix} 2.02 \\ 10000 \end{bmatrix}$$

Regularization

Replace the problem $Ax = b$ with a related problem that reduces the influence of the noise on the solution.

For example, we might think of shifting the matrix eigenvalues a little bit:

$$\left(\begin{bmatrix} 1/2 & 0 \\ 0 & 10^{-6} \end{bmatrix} + \begin{bmatrix} 0.1 & 0 \\ 0 & 0.1 \end{bmatrix} \right)^{-1} \begin{bmatrix} 1.01 \\ 0.01 \end{bmatrix} = \begin{bmatrix} 1.683\ldots \\ 0.099\ldots \end{bmatrix}$$

(This gets something much closer to the exact solution.)

Or projecting the matrix into a lower-dimensional subspace:

$$\left(\begin{bmatrix} 1/2 & 0 \\ 0 & 0 \end{bmatrix}\right)^{\dagger} \begin{bmatrix} 1.01 \\ 0.01 \end{bmatrix} = \begin{bmatrix} 2.02 \\ 0 \end{bmatrix}$$

(The dagger means pseudoinverse. This approach gets something much closer to the exact solution too.)

SVD and Truncated SVD

Express A using the singular value decomposition:

$$A = \sum_{j=1}^{n} \sigma_j u_j v_j^T, \qquad v_j^T v_k = u_j^T u_k = \begin{cases} 1 & \text{if } j = k, \\ 0 & \text{o.w.}, \end{cases}$$

$u_j \in \mathbf{R}^n$, $v_j \in \mathbf{R}^n$, $j = 1, 2, \ldots, n$, and $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_n \geq 0$.

Truncated SVD methods constrain their solutions:

$$x_m^{TSVD} \in \text{span}\{v_1, v_2, \ldots, v_m\}, \quad m = 1, 2, \ldots, n.$$

Krylov subspaces

Let $A \in \mathbf{R}^{n \times n}$.

$$\mathcal{K}_m(A, b) = \text{span}\{b, Ab, \ldots, A^{m-1}b\}, \;\; m = 1, 2, 3, \ldots, n\,.$$

Krylov methods constrain their solutions x_m to lie in $\mathcal{K}_m(A, b)$.

How good are these subspaces for ill-posed problems?

That is, how good a choice is a subspace in which to seek a solution to a discrete ill-posed problem?

We can compare the relative error norm,

$$\|x - \mathcal{P}x\| / \|x\|,$$

of the exact solution and its orthogonal projection $\mathcal{P}x$ in the subspace for some famous test problems from Per Christian Hansen's Regularization Toolbox.

Inverse Laplace transform example

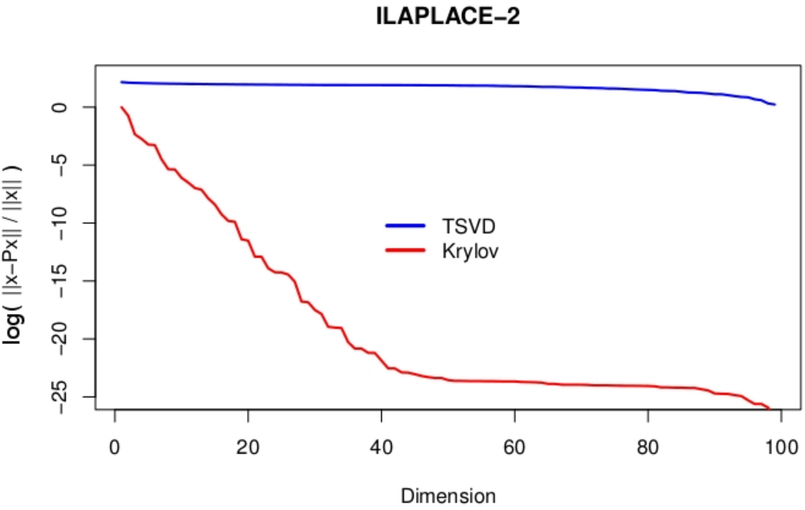
Let $s, t \in [0, \infty)$. Let

$$g(s) = \int \exp(-st) f(t) dt, \text{ where,}$$

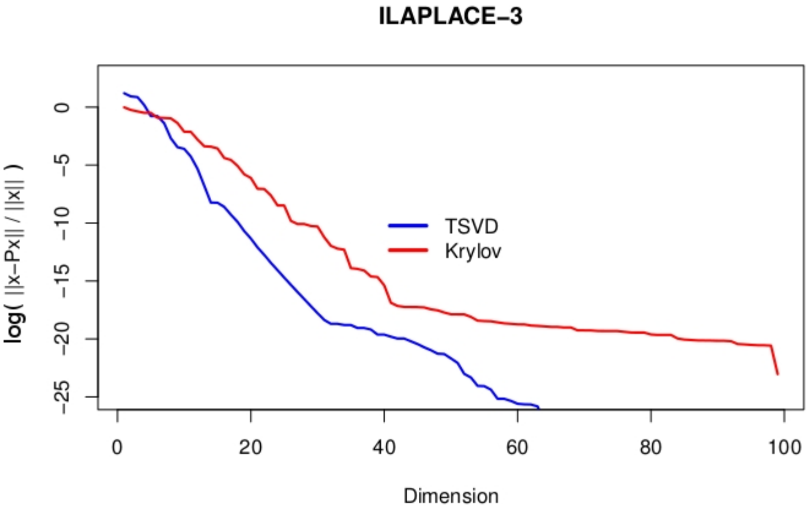
ILAPLACE-2: $g(s) = 1/(s+0.5)$, $f(t) = \exp(-t/2)$.
ILAPLACE-3: $g(s) = 2/(s+0.5)^3$, $f(t) = t^2 \exp(-t/2)$.

We discretize to form the 100×100 system $\tilde{b} = Ax$ and let $b = b + 0.01\|\tilde{b}\|d$, where d is a vector of normally distributed entries of zero mean and uniform variance scaled such that $\|d\| = 1$.

Inverse Laplace transform example



Inverse Laplace transform example



Other Toolbox examples

Problem	Krylov subspaces	TSVD subspaces
baart	wins	
ilaplace-1		wins
ilaplace-2	wins	
ilaplace-3		wins
ilaplace-4	wins	
phillips	wins	
deriv2-1	wins	
heat		wins
shaw	tie	tie
foxgood	tie	tie

Here is a question: is there a good systematic way to pick a subspace projection method rich in solution components for a given problem?

Solution methods

We often just solve a projected ordinary least-squares problem in the subspaces.

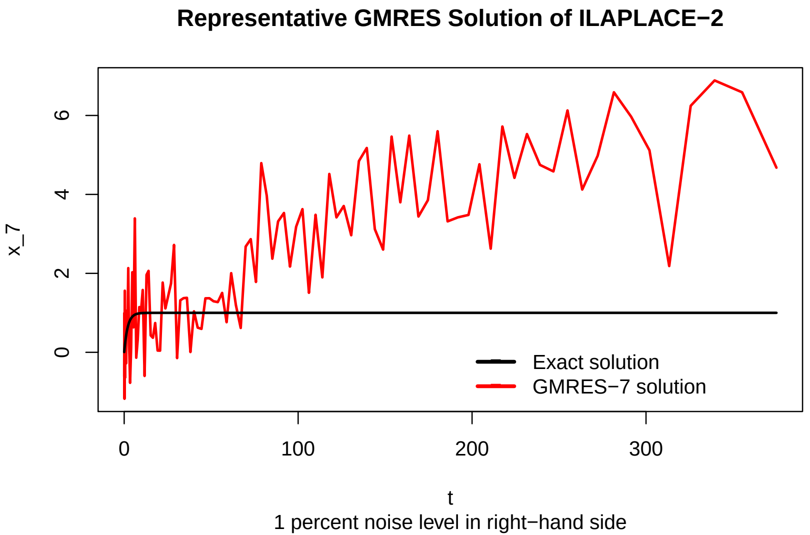
For example, in Krylov subspaces this is called the GMRES method:

$$\min_{x \in \mathcal{K}_m(A,b)} \|b - Ax\|,$$

or, for truncated SVD:

$$\min_{x \in \text{span}\{v_1, v_2, \dots, v_m\}} \|b - Ax\|.$$

But projected OLS solutions are not necessarily good!



We know Krylov subspaces are often good candidate spaces in which to seek solutions. This feels like a little bit like data mining problem to me!

Here is something to think about: alternatives to OLS in these subspaces. For example, use inner regularization methods? Explore solutions in other norms like the T.V. norm or 1-norm? Something else?

Part 2: Another context for regularization

Let's say we have m observations of multivariate normal data of dimension n with mean μ and true covariance Σ.

Arrange the data into an $m \times n$ matrix A.

Examples

- Matrix rows represent time and the columns stock returns. Typically $m >> n$ in this case.
- Rows represent biological samples and the columns expressions of genes, ($n >> m$ typically).

One computation of the empirical sample covariance is:

$$S := \frac{1}{m-1}\left(A - \frac{1}{m}uu^T A\right)^T\left(A - \frac{1}{m}uu^T A\right),$$

where $u = (1, 1, \ldots, 1)^T$.

In many applications (finance, genomics), S can be very different from Σ due to noise in S, incomplete data, and other reasons.

Sometimes folks are really interested in the matrix Σ^{-1} (sometimes called the precision matrix).

$(\Sigma^{-1})_{i,j} = 0$ implies that the corresponding observations i and j are conditionally independent given the other observations.

One kind of regularization problem tries to modify S to better approximate the true (but unknown) Σ.

One approach is to replace S with $S + \lambda L$, where L might be the identity matrix or another regularization operator. This yields Tikhonov-like regularization schemes.

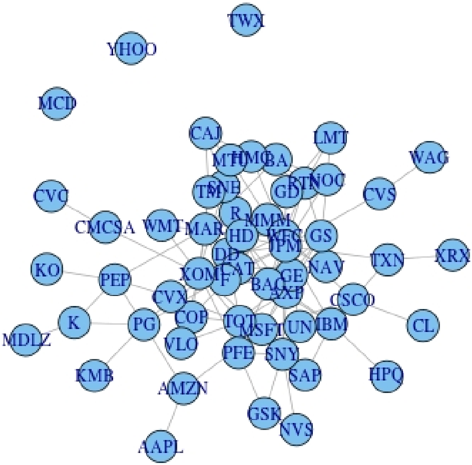
Fun Demo

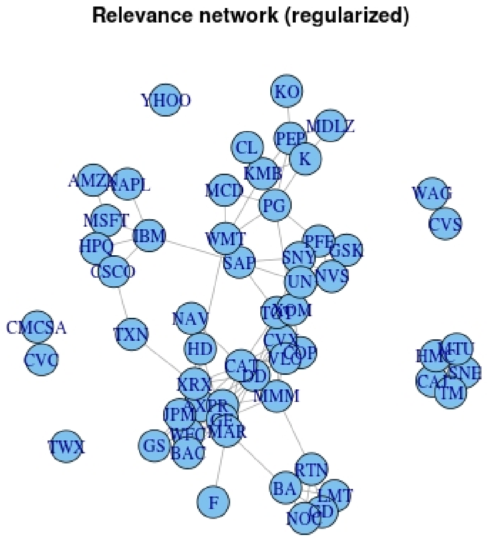
- Let's grab some stock price data from Yahoo finance.
- Convert the prices to log returns.
- Compute the regularized sample correlation matrix S.
- Estimate Σ^{-1} by S^{-1}.
- Threshold S^{-1} to estimate a relevance graph.

R program for the following plots available here:

http://goo.gl/dnUeVW

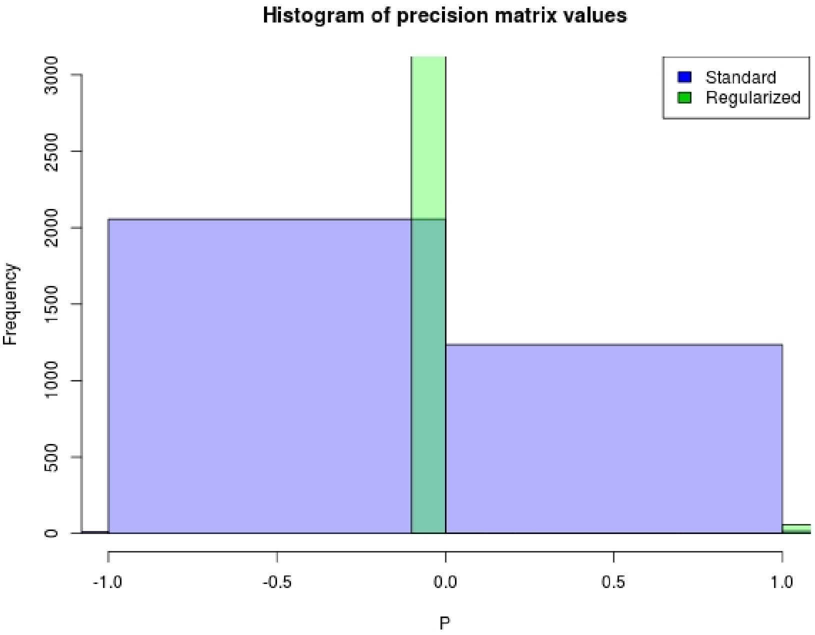
Relevance network (unregularized)





Uncanny clusters of stocks are apparent in the graph!

Effect of regularization on precision matrix entries



Some questions to think about...

What about just projecting into subspaces instead of Tikhonov?

In this example, we seek a sparse S^{-1}. Maybe use a “sparsified” SVD-like decomposition like PMD?

Notes

The stock market example was adapted from
http://scikit-learn.org/stable/auto_examples/applications/plot_stock_market.html

Although, instead of the excellent LASSO method, we use a variation of the widely used (in finance) Lediot-Wolf regularization method from the R package: http://cran.r-project.org/web/packages/corpcor/index.html.

Misha Kilmer and colleagues have written about inner regularization methods.

Curt Vogel and others have written alot about using the total variation norm in regularization problems.

Per Christian Hansen has an excellent introductory book on numerical solution of ill-posed problems.

Daniela Witten came up with a really cool sparse SVD-like matrix decomposition method called PMD.