THE SINGULAR VALUE DECOMPOSITION

MARKUS GRASMAIR

1. Definition and Existence

Theorem 1. *Assume that $A \in \mathbb{R}^{m\times n}$. Then there exist orthogonal matrices $U \in \mathbb{R}^{m\times m}$ and $V \in \mathbb{R}^{n\times n}$, and values $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_p \geq 0$ with $p = \min\{m, n\}$, such that*

$$A = U\Sigma V^T,$$

where $\Sigma \in \mathbb{R}^{m\times n}$ is a diagonal matrix with diagonal entries $\sigma_1,\ldots,\sigma_p$, that is,

$$\Sigma = \left(\begin{array}{cccc|ccc} \sigma_1 & & & 0 & 0 & \ldots & 0 \\ & \sigma_2 & & & \vdots & & \vdots \\ & & \ddots & & & & \\ & & & \ddots & \vdots & & \vdots \\ 0 & & & \sigma_m & 0 & \ldots & 0 \end{array}\right)$$

in case $m \leq n$, and

$$\Sigma = \left(\begin{array}{ccc|c} \sigma_1 & & & 0 \\ & \sigma_2 & & \\ & & \ddots & \\ & & & \ddots \\ 0 & & & \sigma_n \\ \hline 0 & \ldots & & 0 \\ \vdots & & & \vdots \\ 0 & \ldots & & 0 \end{array}\right)$$

in case $m \geq n$. The values σ_k are uniquely determined by A, and are called the singular values *of A.*

Idea of proof. We assume without loss of generality that $A \neq 0$, else the assertion is trivial (we may choose $\Sigma = 0$ and any orthogonal matrices U and V). Moreover, we note that the decomposition $A = U\Sigma V^T$ is equivalent to stating that $U^T AV$ is a diagonal matrix with non-negative, decreasing diagonal entries.

We now recall that the 2-norm of the matrix A is defined as

$$\|A\|_2 := \max_{\|x\|_2=1} \|Ax\|_2. \tag{1}$$

Let $x \in \mathbb{R}^n$ be any x with $\|x\|_2 = 1$ where the maximum in (1) is attained. Define moreover

$$\sigma_1 := \|A\|_2 = \|Ax\|_2 \qquad \text{and} \qquad y := \frac{Ax}{\|Ax\|_2} = \frac{Ax}{\sigma_1}.$$

Complete x to an orthonormal basis

$$V = (x|v_2|\ldots|v_n)$$

Date: November 2016, revised in October 2017.

of $\mathbb{R}^n$, and y to an orthonormal basis

$$U = (y|u_2|\dots|u_m)$$

of $\mathbb{R}^m$. Then the product $U^T A V$ is of the form

$$U^T A V = \begin{pmatrix} \sigma_1 & w^T \\ 0 & \hat{A} \end{pmatrix} =: B$$

for some vector $w \in \mathbb{R}^{n-1}$ and a matrix $\hat{A} \in \mathbb{R}^{(m-1)\times(n-1)}$.

Now let

$$z = \begin{pmatrix} \sigma_1 \\ w \end{pmatrix}.$$

Then

$$Bz = \begin{pmatrix} \sigma_1^2 + w^T w \\ \hat{A} w \end{pmatrix}$$

and therefore

$$\|Bz\|_2^2 = (\sigma_1^2 + w^T w)^2 + \|\hat{A}w\|_2^2 \geq (\sigma_1^2 + w^T w)^2 = (\sigma_1^2 + w^T w)\|z\|_2^2 \geq \sigma_1^2 \|z\|_2^2$$

with equality holding only if $w = 0$. However, since U and V are orthogonal, we have $\|B\|_2 = \|A\|_2 = \sigma_1$ and therefore

$$\|Bz\|_2^2 \leq \sigma_1^2 \|z\|_2^2.$$

Therefore $w = 0$ and we have

$$U^T A V = \begin{pmatrix} \sigma_1 & 0 \\ 0 & \hat{A} \end{pmatrix}.$$

Using induction over p (or: applying the same idea to $\hat{A}$), we arrive at the claimed decomposition. Note here that the numbers σ_k are indeed decreasing, as $\sigma_1 = \|A\|_2 \geq \|\hat{A}\|_2 = \sigma_2$. □

Remark 2. While this Theorem contains a constructive proof of the existence of the SVD, it is not that useful if one actually wants to compute an SVD either analytically or numerically. Better methods for the analytic computation will be discussed below in Section 2, while the numerical computation will be discussed later in this class.

From the previous theorem it follows that we can write

$$\begin{array}{ccccccc} A & = & U & & \Sigma & & V^T \\ \cap & & \cap & & \cap & & \cap \\ \mathbb{R}^{m\times n} & & \mathbb{R}^{m\times m} & & \mathbb{R}^{m\times n} & & \mathbb{R}^{n\times n} \end{array}$$

However, in particular if m and n are very different, this decomposition of A contains lots zero columns or rows in Σ, which make the last columns of either U or V redundant. It can thus be an advantage to use instead a *reduced singular value decomposition* either of the form

$$A = \hat{U}\hat{\Sigma}V^T$$

in case $m > n$ with

$$\hat{\Sigma} = \begin{pmatrix} \sigma_1 & & & 0 \\ & \ddots & & \\ & & \ddots & \\ 0 & & & \sigma_n \end{pmatrix}$$

and

$$\hat{U} = (u_1|u_2|\dots|u_n) \in \mathbb{R}^{m\times n},$$

or

$$A = U\hat{\Sigma}\hat{V}^T$$

in case $m < n$ with

$$\hat{\Sigma} = \begin{pmatrix} \sigma_1 & & 0 \\ & \ddots & \\ & & \ddots \\ 0 & & \sigma_m \end{pmatrix}$$

and

$$\hat{V} = (v_1|v_2|\dots|v_m) \in \mathbb{R}^{n\times m}.$$

That is, we reduce the rectangular matrix Σ to a square matrix containing the singular values, and remove all the redundant columns from either U or V, depending on which is the bigger matrix.

In the following, we will always use the reduced singular value decomposition, and simply write this reduced decomposition as $A = U\Sigma V^T$. However, it is always necessary to keep in mind that one of the matrices U and V will be rectangular.

2. Interpretation of the SVD

Assume that $A \in \mathbb{R}^{m\times n}$ has the singular value decomposition $A = U\Sigma V^T$. As a consequence, we have

$$A^T = V\Sigma^T U^T = V\Sigma U^T,$$

which is a singular value decomposition of A^T. In particular, this implies that A and A^T have the same singular values, which in turn implies that $\|A\|_2 = \|A^T\|_2$.

Next we note that the matrices U^TU and V^TV are always the identity matrices on the respective spaces. That is, if $m \geq n$ we have $U \in \mathbb{R}^{m\times n}$, $V \in \mathbb{R}^{n\times n}$, and

$$U^TU = \mathrm{I}_{n\times n} = V^TV,$$

whereas if $n \geq m$ we have $U \in \mathbb{R}^{m\times m}$, $V \in \mathbb{R}^{n\times m}$, and

$$U^TU = \mathrm{I}_{m\times m} = V^TV.$$

Thus,

$$A^TA = V\Sigma U^TU\Sigma V^T = V\Sigma^2V^T$$

and

$$AA^T = U\Sigma V^TV\Sigma U^T = U\Sigma^2U^T.$$

This shows that the values $\sigma_1^2,\dots,\sigma_r^2$ are exactly the non-zero eigenvalues of both the matrices A^TA and AA^T with corresponding eigenvectors u_k (for A^TA) and v_k (for AA^T), respectively:

Lemma 3. *The non-zero singular values of $A \in \mathbb{R}^{m\times n}$ are precisely the non-zero eigenvalues of any of the positive semi-definite matrices A^TA and AA^T.*

For symmetric matrices, this immediately implies the following connection between singular values and eigenvalues:

Corollary 4. *If $A = A^T \in \mathbb{R}^{n\times n}$ is a symmetric matrix, then its singular values are the (non-negative) square roots of its eigenvalues. If A is SPD, then its singular values and eigenvalues are the same.*

The results above allow us to compute the singular value decomposition of a matrix A by computing eigenvalue decompositions of either A^TA or AA^T (depending on which of these is easier to compute, that is, which of these is the smaller matrix):

Example 5. Consider the matrix

$$A = \begin{pmatrix} 1 & c \\ 0 & 1 \end{pmatrix}$$

with $c \neq 0$. This matrix has a single geometric eigenvalue $\lambda_1 = 1$ with corresponding eigenvector $(1, 0)$. A possible Jordan decomposition of A reads

$$A = \begin{pmatrix} c^{-1} & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} c & 0 \\ 0 & 1 \end{pmatrix}.$$

Next we will compute a singular value decomposition of A. To that end, we will compute first the eigenvalues and eigenvectors of AA^T (doing the computations with A^TA would work fine as well). We have

$$AA^T = \begin{pmatrix} 1 + c^2 & c \\ c & 1 \end{pmatrix}$$

with eigenvalues

$$\lambda_{1,2} = 1 + \frac{c^2}{2} \pm \sqrt{c^2 + \frac{c^4}{4}}.$$

As a consequence, the singular values of A are $\sqrt{\lambda_1}$ and $\sqrt{\lambda_2}$, or

$$\sigma_1 = \sqrt{1 + \frac{c^2}{2} + \sqrt{c^2 + \frac{c^4}{4}}} \qquad \text{and} \qquad \sigma_2 = \sqrt{1 + \frac{c^2}{2} - \sqrt{c^2 + \frac{c^4}{4}}}.$$

In the particular case $c = 8/3$ (which noticeable simplifies all the calculations) we have

$$\sigma_1 = \sqrt{\lambda_1} = 3 \qquad \text{and} \qquad \sigma_2 = \sqrt{\lambda_2} = 1/3.$$

Moreover, the eigenvalues of AA^T corresponding to λ_1 and λ_2 are

$$u_1 = \frac{1}{\sqrt{10}} \begin{pmatrix} 3 \\ 1 \end{pmatrix} \qquad \text{and} \qquad u_2 = \frac{1}{\sqrt{10}} \begin{pmatrix} -1 \\ 3 \end{pmatrix}.$$

Thus we can write

$$AA^T = U\Sigma^2 U^T = \frac{1}{\sqrt{10}} \begin{pmatrix} 3 & -1 \\ 1 & 3 \end{pmatrix} \begin{pmatrix} 9 & 0 \\ 0 & 1/9 \end{pmatrix} \frac{1}{\sqrt{10}} \begin{pmatrix} 3 & 1 \\ -1 & 3 \end{pmatrix}.$$

Moreover, the matrix U in this decomposition of AA^T can be chosen to be precisely the matrix U in the singular value decomposition $A = U\Sigma V^T$ of A. Now the equation $A = U\Sigma V^T$ implies that

$$V = A^T U \Sigma^{-1} = \frac{1}{\sqrt{10}} \begin{pmatrix} 1 & -3 \\ 3 & 1 \end{pmatrix}.$$

We therefore obtain the singular value decomposition

$$\begin{pmatrix} 1 & 8/3 \\ 0 & 1 \end{pmatrix} = \frac{1}{\sqrt{10}} \begin{pmatrix} 3 & -1 \\ 1 & 3 \end{pmatrix} \begin{pmatrix} 3 & 0 \\ 0 & 1/3 \end{pmatrix} \frac{1}{\sqrt{10}} \begin{pmatrix} 1 & 3 \\ -3 & 1 \end{pmatrix}.$$

Remark 6. We note that the singular value decomposition allows for a useful geometric interpretation of linear mappings: In the case $m = n$, the mappings U and V are orthogonal, and thus either rotations or reflections. In particular, the mappings U and V leave the unit sphere unchanged. On the other hand, the mapping Σ is diagonal and therefore transforms the unit sphere to an ellipse with semi-axes of lengths $\sigma_1, \dots, \sigma_n$ parallel to the coordinate axes. In total, the mapping $A = U\Sigma V^T$ transforms the unit sphere into an ellipse with semi-axes of lengths given by the singular values, parallel to $u_1, \dots, u_n$.

In the case $n > m$, the situation is similar, although the application of V^T will be the composition of a rotation/reflection with a projection onto a lower-dimensional space, while for $n < m$ the result will be an n-dimensional ellipse embedded in $\mathbb{R}^m$.

Remark 7. Another interpretation of the singular value decomposition can be obtained by considering it as a solution of an optimisation problem (similarly as in the proof of Theorem 1). Indeed, consider the problem

$$\max \frac{1}{2}\|Ax\|_2^2 \qquad \text{subject to } \frac{1}{2}\|x\|_2^2 = \frac{1}{2}. \tag{2}$$

In order to solve this optimisation problem, we introduce its Lagrangian

$$\mathcal{L}(x,\lambda) = \frac{1}{2}\|Ax\|_2^2 - \frac{\lambda}{2}\big(\|x\|_2^2 - 1\big).$$

The possible candidates for the (primal-dual) solutions of (2) are then the pairs $(x,\lambda) \in \mathbb{R}^n \times \mathbb{R}$ satisfying the KKT conditions $\nabla_x \mathcal{L}(x,\lambda) = 0$ and $\|x\|_2 = 1$, that is, the pairs (x,λ) satisfying the equations

$$A^T Ax - \lambda x = 0 \qquad \text{and} \qquad \|x\|_2 = 1.$$

The solutions of this equations are precisely the eigenvector–eigenvalue pairs for the matrix $A^T A$. Similarly, if we consider the problem

$$\max \frac{1}{2}\|A^T x\|_2^2 \qquad \text{subject to } \frac{1}{2}\|x\|_2^2 = \frac{1}{2}, \tag{3}$$

then we obtain as solutions precisely the eigenvector–eigenvalue pairs for the matrix AA^T. Since the singular values of A are exactly the square roots of the eigenvalues of $A^T A$ and AA^T, and the singular vectors are the corresponding eigenvectors, this allows us to interpret singular values and singular vectors as KKT points (or critical points) of the optimisation problems (2) and (3).

3. Matrix Properties via the SVD

An alternative way of formulating the singular value decomposition is to write

$$A = \sum_{k=1}^{p} \sigma_k u_k v_k^T, \tag{4}$$

where $p = \min\{m,n\}$, and u_k, v_k denote the k-th column of U and V, respectively. In particular, we obtain with this notation that

$$Ax = \sum_{k=1}^{p} \sigma_k (v_k, x) u_k$$

for every $x \in \mathbb{R}^n$.

Now let

$$r := \max\big\{k : \sigma_k > 0\big\}.$$

That is, we have

$$\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_r > 0 = \sigma_{r+1} = \ldots = \sigma_p.$$

Then the decomposition (4) shortens to

$$A = \sum_{k=1}^{r} \sigma_k u_k v_k^T, \tag{5}$$

and we have

$$Ax = \sum_{k=1}^{r} \sigma_r (v_k, x) u_k.$$

Since the vectors u_k are linearly independent, we immediately obtain the following:

- We have $\operatorname{rank} A = r$.
- The range of A is

$$\operatorname{Ran} A = \operatorname{span}\{u_1, \ldots, u_r\}.$$

- The kernel of A is
$$\ker A = \operatorname{span}\{v_1, \dots, v_r\}^\perp.$$

Additionally, we have seen in the proof of the existence of the singular value decomposition that
$$\|A\|_2 = \sigma_1,$$
and it is possible to show that
$$\|A\|_F = (\sigma_1^2 + \ldots + \sigma_r^2)^{1/2}.$$
Here $\|A\|_F$ denotes the *Frobenius norm* of A given by
$$\|A\|_F = \Big(\sum_{i,j} a_{ij}^2\Big)^{1/2}.$$

Moreover, truncating the expansion (5) results in the best possible low rank approximations of A, as the following Theorem shows:

Theorem 8. *If A has the singular value decomposition $A = U\Sigma V^T$, then the matrix*
$$A_k := \sum_{j=1}^{k} \sigma_j u_j v_j^T$$
with $1 \leq k \leq p$ solves the optimisation problems
$$\min_{\operatorname{rank}(B)\leq k} \|A - B\|_2$$
and
$$\min_{\operatorname{rank}(B)\leq k} \|A - B\|_F.$$
Moreover
$$\|A - A_k\|_2 = \sigma_{k+1}$$
and
$$\|A - A_k\|_F = \Big(\sum_{j=k+1}^{r} \sigma_j^2\Big)^{1/2}.$$

In other words, the first terms of the singular value decomposition provide the best low rank approximations of the matrix A both with respect to the 2-norm and with respect to the Frobenius norm.

4. Pseudoinverses

Assume now that $A \in \mathbb{R}^{m\times n}$ has the singular value decomposition
$$A = U\Sigma V^T,$$
that $m > n$ and that A has full rank, that is, $\operatorname{rank} A = n$. Then the matrix $\Sigma \in \mathbb{R}^{n\times n}$ is invertible and we can define
$$A^\dagger := V\Sigma^{-1}U^T \in \mathbb{R}^{n\times m},$$
the *pseudo-inverse* (or *Moore–Penrose inverse*) of A.

Note that
$$A^\dagger A = V\Sigma^{-1}U^T U\Sigma V^T = VV^T = \mathrm{I}_{n\times n},$$
whereas
$$AA^\dagger = U\Sigma V^T V\Sigma^{-1}U^T = UU^T$$
is the orthogonal projection onto the range of A.

Lemma 9. *If $m > n$, $\operatorname{rank} A = n$, and $b \in \mathbb{R}^m$, then $A^\dagger b$ is the solution of the least-squares problem*

$$\min_{x \in \mathbb{R}^n} \|Ax - b\|_2^2. \tag{6}$$

Proof. Since the matrix $\Sigma V^T \in \mathbb{R}^{n \times n}$ is invertible with inverse $V\Sigma^{-1}$, the vector $x^\dagger$ solves (6), if and only if we have $x^\dagger = V\Sigma^{-1}y$, where y solves the optimisation problem

$$\min_{y \in \mathbb{R}^n} \|Uy - b\|_2^2.$$

That is, Uy is simply the orthogonal projection of b onto the range of U, which means that $Uy = UU^T b$. Since $U \in \mathbb{R}^{m \times n}$ with $n < m$ is injective, it follows that $y = U^T b$. Thus $x^\dagger = V\Sigma^{-1}y = V\Sigma^{-1}U^T b = A^\dagger b$ is the unique solution of (6). □

Now assume that $m < n$, but that $A \in \mathbb{R}^{m \times n}$ still has full rank (that is, $\operatorname{rank} A = m$). Then we can again define

$$A^\dagger := V\Sigma^{-1}U^T,$$

as $\Sigma \in \mathbb{R}^{m \times m}$ is invertible. However, in this situation we have

$$AA^\dagger = UU^T = \mathrm{I}_{m \times m},$$

whereas

$$A^\dagger A = VV^T$$

is the projection onto $\operatorname{Ran} V = (\ker A)^\perp$.

Lemma 10. *If $m < n$, $\operatorname{rank} A = m$, and $b \in \mathbb{R}^m$, then $A^\dagger b$ solves the optimisation problem*

$$\min_x \|x\|_2^2 \qquad \text{subject to } Ax = b. \tag{7}$$

Proof. First we note that $x^\dagger := A^\dagger b$ satisfies $Ax^\dagger = AA^\dagger b = b$, which means that $x^\dagger$ is indeed admissible for (7). Now assume that y is another vector satisfying $Ay = b$. Then $A^\dagger Ay = A^\dagger b = x^\dagger$ and therefore

$$\|A^\dagger Ay\|_2 = \|x^\dagger\|_2.$$

On the other hand, $A^\dagger A = VV^T$ is an orthogonal projection and therefore

$$\|A^\dagger Ay\|_2 \leq \|y\|_2$$

with equality only if $A^\dagger Ay = y$. This shows that indeed $x^\dagger$ is the unique solution of (7). □

Now we note that in both situations discussed above we could alternatively write

$$A^\dagger := \sum_{k=1}^{p} \sigma_k^{-1} v_k u_k^T = \sum_{k=1}^{r} \sigma_k^{-1} v_k u_k^T.$$

This last definition, however, is still meaningful if the matrix A does not have full rank. That is, for arbitrary $A \in \mathbb{R}^{m \times n}$ of rank r we can define the pseudoinverse

$$A^\dagger := \sum_{k=1}^{r} \sigma_k^{-1} v_k u_k^T \in \mathbb{R}^{n \times m}.$$

In this case, usually neither $AA^\dagger$ nor $A^\dagger A$ can be an identity matrix, but $A^\dagger$ still retains some semblance of an inverse of A, as the following result shows:

Lemma 11. *For every $A \in \mathbb{R}^{m \times n}$ the following identities hold:*

- $(A^\dagger)^\dagger = A$,
- $A^\dagger A A^\dagger = A^\dagger$,
- $AA^\dagger A = A$.

Moreover, the vector $A^\dagger b$ comes as close to solving the linear system $Ax = b$ as one can reasonably hope for:

Theorem 12. *If $A \in \mathbb{R}^{m\times n}$ and $b \in \mathbb{R}^m$, then $x^\dagger := A^\dagger b$ solves the bilevel optimisation problem*

$$\min_x \|x\|_2^2 \qquad \text{such that } x \text{ solves } \min_{\hat{x}} \|A\hat{x} - b\|_2^2.$$

In other words, application of the pseudoinverse $A^\dagger$ to b selects from all least squares solutions of the equation $Ax = b$ the one with the smallest norm.

5. Truncated SVD and the L-curve method

Let now $A \in \mathbb{R}^{m\times n}$ with pseudoinverse $A^\dagger \in \mathbb{R}^{n\times m}$, and assume that $b \in \mathbb{R}^m$. Denote moreover by

$$x^\dagger := A^\dagger b$$

the "true" solution of the equation $Ax = b$. In many practical applications, we have the problem that the right hand side of this equation is subject to measurement errors and that instead of the true data b we measure some noisy data

$$b^\delta = b + n^\delta$$

for some noise $n^\delta \in \mathbb{R}^n$ of size $\|n^\delta\| \approx \delta$. If we use the pseudoinverse for solving the noisy system $Ax = b^\delta$, we then obtain a noisy solution

$$x^\delta = A^\dagger b^\delta = A^\dagger(b + n^\delta) = x^\dagger + A^\dagger n^\delta,$$

that is, the error we obtain is

$$x^\delta - x^\dagger = A^\dagger n^\delta.$$

Specifically, the worst case error is

$$\max_{\|n^\delta\|_2 \le \delta} \|x^\delta - x^\dagger\|_2 = \max_{\|n^\delta\|_2 \le \delta} \|A^\dagger n^\delta\| = \frac{\delta}{\sigma_r},$$

with σ_r being the smallest positive singular value of A. In case the matrix A has small non-zero singular values, this means that the error in the solution might be several orders of magnitude larger than the error in the data.

Additionally, the mapping $b \mapsto A^\dagger b$ will be very sensitive with respect to small variations in b. That is, small changes in the measurements might lead to huge changes in the supposed solution of the system.

As a possible remedy, we may truncate the singular value decomposition of the matrix A: Given $\varepsilon > 0$ we define

$$A_\varepsilon := \sum_{k:\sigma_k \ge \varepsilon} \sigma_k u_k v_k^T$$

and

$$A_\varepsilon^\dagger = \sum_{k:\sigma_k \ge \varepsilon} \sigma_k^{-1} v_k u_k^T.$$

That is, we ignore all the singular values of A that are below the threshold ε. Using this truncated matrix for solving the noisy system $Ax = b^\delta$, we obtain a regularised solution

$$x_\varepsilon^\delta := A_\varepsilon^\dagger b^\delta.$$

In order to estimate the quality of this regularised solution, we note that the worst case error is

$$\begin{aligned}\max_{\|n^\delta\|_2\le\delta}\|x_\varepsilon^\delta - x^\dagger\|_2 &= \max_{\|n^\delta\|_2\le\delta}\|A_\varepsilon^\dagger(b+n^\delta) - x^\dagger\|_2\\ &\le \|A_\varepsilon^\dagger b - x^\dagger\|_2 + \max_{\|n^\delta\|_2\le\delta}\|A_\varepsilon^\dagger b\|_2\\ &= \|A_\varepsilon^\dagger b - x^\dagger\|_2 + \min_{k:\sigma_k\ge\varepsilon}\frac{\delta}{\sigma_k}\\ &\le \|A_\varepsilon^\dagger b - x^\dagger\|_2 + \frac{\delta}{\varepsilon}.\end{aligned}$$

That is, the error in the solution splits into a regularised data error given by δ/ε and a regularisation error given by $\|A_\varepsilon^\dagger b - x^\dagger\|_2$.

Now we note that we can write

$$x^\dagger = \sum_k (x^\dagger, v_k)v_k =: \sum_k x_k^\dagger v_k.$$

That is, $x_k^\dagger$ are the coefficients of $x^\dagger$ with respect to the orthonormal basis $v_1,\dots,v_r$ of $\operatorname{Ran} A^\dagger$. With this notation we have

$$x_\varepsilon^\dagger := A_\varepsilon^\dagger b = \sum_{k:\sigma_k\ge\varepsilon} x_k^\dagger v_k$$

and therefore

$$\|A_\varepsilon^\dagger b - x^\dagger\|_2 = \|x_\varepsilon^\dagger - x^\dagger\|_2 = \Big(\sum_{k:\sigma_k<\varepsilon}(x_k^\dagger)^2\Big)^{1/2}.$$

In other words, the regularisation error is exactly the norm of the coefficients of $x^\dagger$ (with respect to V) corresponding to singular values smaller than ε.

Apart from the actual computation of the SVD of A, a major challenge in this regularised inversion is the choice of the parameter ε. In the following we will briefly discuss one heuristic method, which is called the *L-curve method*. As a first step, we note that, instead of choosing the parameter ε, we may as well choose the number of singular values that are used in the truncated SVD of A. That is, we denote for $k = 1,\dots,r$

$$A_k^\dagger := \sum_{j=1}^k \sigma_j^{-1} v_j u_j^T$$

and

$$x_\delta^{(k)} := A_k^\dagger b^\delta = \sum_{j=1}^k \sigma_j^{-1}\langle u_j, b^\delta\rangle v_j.$$

Then

$$\|x_\delta^{(k)}\|_2^2 = \sum_{j=1}^k \sigma_j^{-1}\langle u_j, b^\delta\rangle^2,$$

which implies that the mapping

$$k \mapsto \|x_\delta^{(k)}\|_2$$

is increasing in k. At the same time,

$$Ax_\delta^{(k)} - b^\delta = \sum_{j=k+1}^n \langle u_j, b^\delta\rangle u_j,$$

and therefore the mapping

$$k \mapsto \|Ax_\delta^{(k)} - b^\delta\|_2$$

is decreasing. Now an often observed behaviour of these mappings is that for small k the data fit $\|Ax_\delta^{(k)} - b^\delta\|_2$ significantly improves with k for relatively small increases of $\|x_\delta^{(k)}\|_2$. However, once the true solution $x^\dagger$ is sufficiently well reconstructed, further increases in k will mainly reconstruct the noise. Therefore the norm $\|x_\delta^{(k)}\|_2$ will increase significantly with a comparably small decrease of the residual $\|Ax_\delta^{(k)} - b^\delta\|_2$. The best reconstructions should therefore be obtained for parameters k for which neither $\|x_\delta^{(k)}\|_2$ and $\|Ax_\delta^{(k)} - b^\delta\|_2$ show large variations in k.

Now consider a log-log-plot of these two mappings, that is, plot the "curve" given by the points

$$\bigl(\log(\|x_\delta^{(k)}\|_2), \log(\|Ax_\delta^{(k)}\|_2)\bigr) \in \mathbb{R}^2$$

for $k = 1, \dots, r$. Following the considerations above, one would expect that these points follow an L-shape, and the corner of the L marks those parameters k for which the most plausible solutions can be expected to be obtained.

Example 13 (Discrete deconvolution)**.** Let $k\colon \mathbb{R} \to \mathbb{R}$ be some bounded and integrable function. Given a function $g\colon [0,1] \to \mathbb{R}$, we want to find a function $f\colon [0,1] \to \mathbb{R}$ such that $k * f = g$, that is,

$$\int_0^1 f(y)k(x-y)\,dy = g(x)$$

for all $x \in [0,1]$.

We discretise this equation by using the midpoint rule (the choice of the quadrature rule does not fundamentally change the results) and obtain the system of equations

$$\frac{1}{n}\sum_{j=1}^{n} k(x_i - x_j)f_j = g_i$$

with $x_j = (j - 1/2)/n$, $j = 1, \dots, n$, and $f_j \approx f(x_j)$, $g_i := g(x_i)$.

Consider in particular the case $k(x) = e^{-x^2/2}$. That is, the function k is, up to scaling, the standard Gaussian kernel of variance 1. In this case, it turns out that already with $n = 100$, the resulting system of equations is sufficiently ill-posed as to yield useless results even if the only error comes from rounding errors due to computations with "only" double precision.

We assume in the following that the true solution is $f^\dagger(x) = x^2(1-x^2)$ and that we are given exact (that is, exact up to machine precision) data $g = k * f^\dagger$. As can be seen in Figure 1, the unregularised solution of the equation $k * f = g$ is essentially useless. However, using a truncated singular value decomposition, one may obtain an almost perfect reconstruction of $f^\dagger$. In Figure 1 the reconstructions $f^{(k)}$ using $k = 8$, $k = 33$, and $k = 50$ singular values are shown. For $k = 8$ and $k = 33$, the reconstructions capture the actual shape of $f^\dagger$ reasonably well, and for values of k between 9 and 32, the reconstructions become visually indistinguishable from $f^\dagger$.

In the case of noisy data $g^\delta = k * f^\dagger + n^\delta$ the situtation is even worse, and it is only possible to obtain reasonable reconstructions with a very small number of singular values. As an example, consider the situation shown in Figure 2. Here, using $k = 9$ singular values for the reconstruction provides a reasonable result, in which the error is only slightly larger than the measurement error. For $k = 15$, however, the solution is dominated by large oscillations and essentially useless.

In Figure 3, the L-curves both for the noise-free and the noisy case are shown. The former predicts that the best reconstructions in the noise-free case can be found with $k \approx 30$, whereas the latter predicts good reconstructions for $k \approx 9$. In both cases, these predictions match reality surprisingly well.

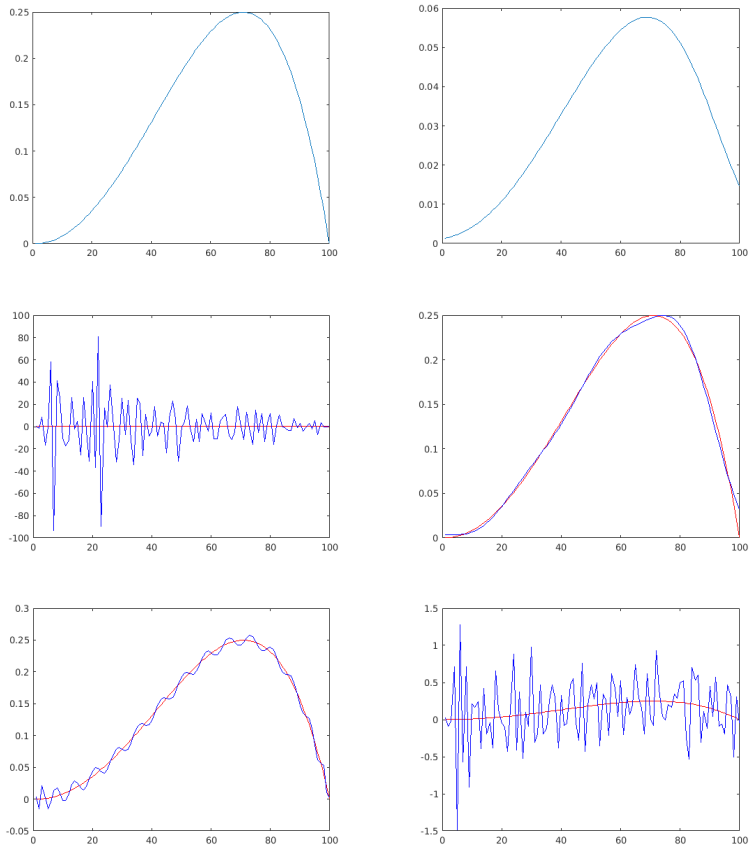

FIGURE 1. *First row, left:* true solution $f^\dagger(x) = x^2(1-x^2)$ for $0 \leq x \leq 1$; *First row, right:* given noise-free data $g = k * f^\dagger$ with k being a Gaussian kernel of variance 1; *Second row, left:* solution of the discretised equation $k * f = g$ without any regularisation; *Second row, right:* regularised solution using TSVD with 8 singular values; *Last row, left:* regularised solution using TSVD with 33 singular values; *Last row, right:* regularised solution using TSVD with 50 singular values. In the second and last row, the red curve shows always the true solution $f^\dagger$, while the blue curve shows the reconstruction $f^{(k)}$.

6. Regularisation

In the following, we will briefly discuss an alternative interpretation of the truncated SVD, which allows for the definition of more general families of regularisation methods.

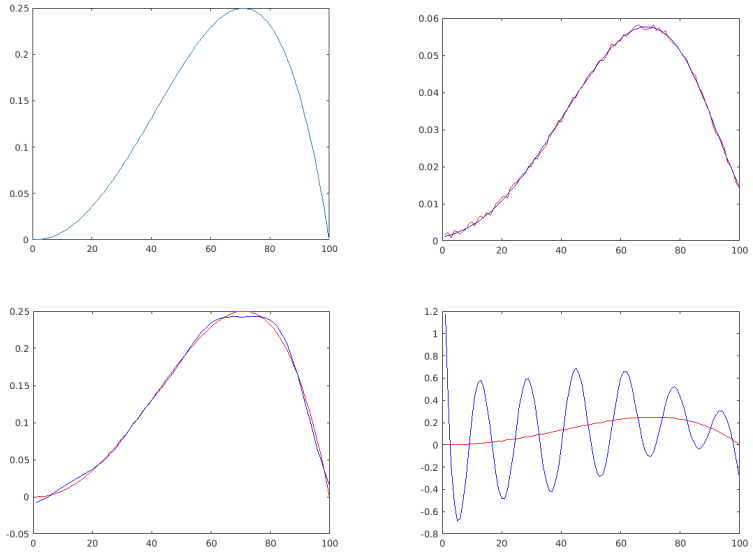

FIGURE 2. *First row, left:* true solution $f^\dagger(x) = x^2(1-x^2)$ for $0 \leq x \leq 1$; *First row, right:* the blue curve shows the noise-free data $g = k * f^\dagger$, and the red curve shows noisy data $g^\delta = g + n^\delta$ with n^δ being a realisation of an i.i.d. Gaussian random variable with standard deviation $\delta = 0.002$; *Last row, left:* regularised solution from noisy data using TSVD with 9 singular values; *Last row, right:* regularised solution from noisy data using TSVD with 15 singular values. The red curve shows the true solution $f^\dagger$, while the blue curve shows the reconstruction $f_\delta^{(k)}$.

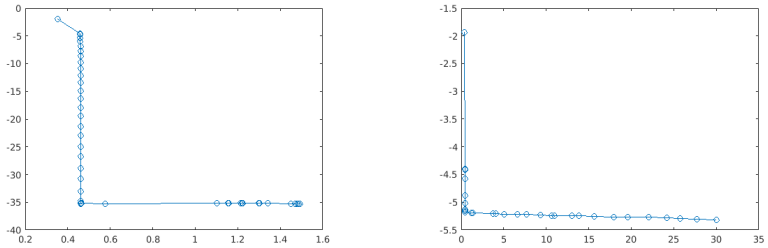

FIGURE 3. L-curves for the numerical examples depicted in Figures 1 and 2. The left figure shows the L-curve in the noise-free case. Here the corner of the L is formed by the reconstructions with 28–31 singular values. The left figure shows the L-curve in the noisy case. Here the corner of the L is formed by the reconstruction with 9 singular values.

We start by writing

$$A^\dagger = V \begin{pmatrix} \sigma_1^{-1} & & & & & & 0 \\ & \ddots & & & & & \\ & & \ddots & & & & \\ & & & \sigma_r^{-1} & & & \\ & & & & 0 & & \\ & & & & & \ddots & \\ 0 & & & & & & 0 \end{pmatrix} U^T.$$

Define now the function

$$f\colon \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}, \qquad f(s) = \begin{cases} 1/s & \text{if } s > 0, \\ 0 & \text{if } s = 0, \end{cases}$$

and let

$$f(\Sigma) = \begin{pmatrix} f(\sigma_1) & & & 0 \\ & \ddots & & \\ & & \ddots & \\ & & & f(\sigma_p) \end{pmatrix}.$$

Then we can equivalently write

$$A^\dagger = V f(\Sigma) U^T.$$

Moreover, setting

$$f_\varepsilon(s) := \begin{cases} 1/s & \text{if } s \geq \varepsilon, \\ 0 & \text{if } s < \varepsilon, \end{cases}$$

we obtain that

$$A_\varepsilon^\dagger = V f_\varepsilon(\Sigma) U^T.$$

That is, the approximation $A_\varepsilon^\dagger$ to the pseudoinverse of A can be obtained by approximating the function f (which essentially is the mapping $s \mapsto 1/s$) by means of the bounded function f_ε. Because of the boundedness of f_ε the approximation is more stable than $A^\dagger$ itself, but the increased stability comes at the cost of a potential decrease in accuracy in case of a low noise level δ.

This interpretation of the truncated singular value decomposition allows us to consider alternative regularisation methods by using different approximations of f.

Amongst the most important are:

- *Lavrentiev regularisation* defined by the function

  $$g_\lambda(s) := \begin{cases} \frac{1}{\lambda + s} & \text{if } s > 0, \\ 0 & \text{if } s = 0. \end{cases}$$

  Instead of truncating the function $1/s$, we shift it to the left and thus get rid of the singularity at zero. This method is particularly interesting if $A \in \mathbb{R}^{n \times n}$ is a positive definite square matrix, in which case

  $$V g_\lambda(\Sigma) U^T = (\lambda \operatorname{I} + A)^{-1}.$$

  That is, the regularised solution x_λ of $Ax = b$ can be found by solving the system

  $$(\lambda \operatorname{I} + A)x = b.$$

- *Tikhonov regularisation* defined by the function

  $$h_\alpha(s) := \frac{s}{s^2 + \alpha}.$$

  In this case,

  $$V h_\alpha(\Sigma) U^T = (\alpha \operatorname{I}_{n \times n} + A^T A)^{-1} A^T.$$

  That is, the regularised solution x_α can be found by solving the system

  $$(\alpha \operatorname{I} + A^T A)x = A^T b.$$

Department of Mathematical Sciences, Norwegian University of Science and Technology, 7491 Trondheim, Norway

E-mail address: markus.grasmair@ntnu.no