Math 4/896: Seminar in Mathematics
Topic: Inverse Theory

Instructor: Thomas Shores
Department of Mathematics

AvH 10

Outline

Basic Theory of SVD

Theorem

(Singular Value Decomposition) Let G be an $m \times n$ real matrix. Then there exist $m \times m$ orthogonal matrix U, $n \times n$ orthogonal matrix V and $m \times n$ diagonal matrix S with diagonal entries $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_q$, with $q = \min\{m, n\}$, such that $U^T GV = S$. Moreover, numbers $\sigma_1, \sigma_2, \ldots, \sigma_q$ are uniquely determined by G.

Definition

With notation as in the SVD Theorem, and U_p, V_p the matrices consisting of the first p columns of U, V, respectively, and S_p the first p rows and columns of S, where σ_p is the last nonzero singular value, then the **Moore-Penrose pseudoinverse** of G is

$$G^\dagger = V_p S_p^{-1} U_p^T \equiv \sum_{j=1}^{p} \frac{1}{\sigma_j} \mathbf{V}_j \mathbf{U}_j^T .$$

Matlab Knows It

Carry out these calculations in Matlab:

```
> n = 6
> G = hilb(n);
> svd(G)
>[U,S,V] = svd(G);
>U'*G*V - S
>[U,S,V] = svd(G,'econ');
> % try again with n=16 and then G=G(1:8)
> % what are the nonzero singular values of G?
```

Applications of the SVD

Use notation above and recall that the null space and column space (range) of matrix G are $N(G) = \{\mathbf{x} \in \mathbb{R}^n \mid G\mathbf{x} = \mathbf{0}\}$ and

$$R(G) = \{\mathbf{y} \in \mathbb{R}^m \mid \mathbf{y} = G\mathbf{x},\ \mathbf{x} \in \mathbb{R}^n\} = \operatorname{span}\{\mathbf{G}_1, \mathbf{G}_2, \ldots, \mathbf{G}_n\}$$

Theorem

(1) $\operatorname{rank}(G) = p$ *and* $A = \sum_{j=1}^{p} \sigma_j \mathbf{U}_j \mathbf{V}_j^T$

(2)$N(G) = \operatorname{span}\{\mathbf{V}_{p+1}, \mathbf{V}_{p+2}, \ldots, \mathbf{V}_n\}$,$R(G) = \operatorname{span}\{\mathbf{V}_1, \mathbf{V}_2, \ldots, \mathbf{V}_p\}$

(3)$N\left(G^T\right) = \operatorname{span}\{\mathbf{U}_{p+1}, \mathbf{U}_{p+2}, \ldots, \mathbf{U}_m\}$,$R(G) = \operatorname{span}\{\mathbf{U}_1, \mathbf{U}_2, \ldots, \mathbf{U}_p\}$

(4) $\mathbf{m}_\dagger = G^\dagger \mathbf{d}$ *is the least squares solution to* $G\mathbf{m} = \mathbf{d}$ *of minimum 2-norm.*

Heart of the Difficulties with Least Squares Solutions

Use the previous notation, so that G is $m \times n$ with rank p and SVD, etc as above. By **data space** we mean the vector space $\mathbb{R}^m$ and by **model space** we mean $\mathbb{R}^n$.

No Rank Deficiency:

This means that $p = m = n$. Comments:

- This means that null space of both G and G^T are trivial (both $\{0\}$).
- Then there is a perfect correspondence between vectors in data space and model space:

$$G\mathbf{m} = \mathbf{d}, \ \ \mathbf{m} = G^{-1}\mathbf{d} = G^{\dagger}\mathbf{d}.$$

- This is the ideal. But are we out of the woods?
- No, we still have to deal with data error and ill-conditioning of the coefficient matrix (remember Hilbert?).

Heart of the Difficulties with Least Squares Solutions

Use the previous notation, so that G is $m \times n$ with rank p and SVD, etc as above. By **data space** we mean the vector space $\mathbb{R}^m$ and by **model space** we mean $\mathbb{R}^n$.

No Rank Deficiency:

This means that $p = m = n$. Comments:

- This means that null space of both G and G^T are trivial (both $\{0\}$).
- Then there is a perfect correspondence between vectors in data space and model space:

$$G\mathbf{m} = \mathbf{d}, \quad \mathbf{m} = G^{-1}\mathbf{d} = G^{\dagger}\mathbf{d}.$$

- This is the ideal. But are we out of the woods?
- No, we still have to deal with data error and ill-conditioning of the coefficient matrix (remember Hilbert?).

Heart of the Difficulties with Least Squares Solutions

Use the previous notation, so that G is $m \times n$ with rank p and SVD, etc as above. By **data space** we mean the vector space $\mathbb{R}^m$ and by **model space** we mean $\mathbb{R}^n$.

No Rank Deficiency:

This means that $p = m = n$. Comments:

- This means that null space of both G and G^T are trivial (both $\{0\}$).
- Then there is a perfect correspondence between vectors in data space and model space:

$$G\mathbf{m} = \mathbf{d}, \ \mathbf{m} = G^{-1}\mathbf{d} = G^{\dagger}\mathbf{d}.$$

- This is the ideal. But are we out of the woods?
- No, we still have to deal with data error and ill-conditioning of the coefficient matrix (remember Hilbert?).

Heart of the Difficulties with Least Squares Solutions

Use the previous notation, so that G is $m \times n$ with rank p and SVD, etc as above. By **data space** we mean the vector space $\mathbb{R}^m$ and by **model space** we mean $\mathbb{R}^n$.

No Rank Deficiency:

This means that $p = m = n$. Comments:

- This means that null space of both G and G^T are trivial (both $\{0\}$).
- Then there is a perfect correspondence between vectors in data space and model space:

$$G\mathbf{m} = \mathbf{d}, \ \ \mathbf{m} = G^{-1}\mathbf{d} = G^{\dagger}\mathbf{d}.$$

- This is the ideal. But are we out of the woods?
- No, we still have to deal with data error and ill-conditioning of the coefficient matrix (remember Hilbert?).

Heart of the Difficulties with Least Squares Solutions

Use the previous notation, so that G is $m \times n$ with rank p and SVD, etc as above. By **data space** we mean the vector space $\mathbb{R}^m$ and by **model space** we mean $\mathbb{R}^n$.

No Rank Deficiency:

This means that $p = m = n$. Comments:

- This means that null space of both G and G^T are trivial (both $\{0\}$).
- Then there is a perfect correspondence between vectors in data space and model space:

$$G\mathbf{m} = \mathbf{d}, \ \mathbf{m} = G^{-1}\mathbf{d} = G^{\dagger}\mathbf{d}.$$

- This is the ideal. But are we out of the woods?
- No, we still have to deal with data error and ill-conditioning of the coefficient matrix (remember Hilbert?).

Heart of the Difficulties with Least Squares Solutions

Use the notation $\mathbf{m}_\dagger = G^\dagger \mathbf{d}$.

Row Rank Deficiency:

This means that $d = n < m$. Comments:

- This means that null space of G is trivial, but that of G^T is not.
- Here $\mathbf{m}_\dagger$ is the unique least squares solution.
- And $\mathbf{m}_\dagger$ is the exact solution to $G\mathbf{m} = \mathbf{d}$ exactly if $\mathbf{d}$ is in the range of G.
- But $\mathbf{m}$ is insensitive to any translation $\mathbf{d} + \mathbf{d}_0$ with $\mathbf{d}_0 \in N\left(G^\dagger\right)$

Heart of the Difficulties with Least Squares Solutions

Use the notation $\mathbf{m}_{\dagger} = G^{\dagger}\mathbf{d}$.

Row Rank Deficiency:

This means that $d = n < m$. Comments:

- This means that null space of G is trivial, but that of G^T is not.
- Here $\mathbf{m}_{\dagger}$ is the unique least squares solution.
- And $\mathbf{m}_{\dagger}$ is the exact solution to $G\mathbf{m} = \mathbf{d}$ exactly if $\mathbf{d}$ is in the range of G.
- But $\mathbf{m}$ is insensitive to any translation $\mathbf{d} + \mathbf{d}_0$ with $\mathbf{d}_0 \in N\left(G^{\dagger}\right)$

Heart of the Difficulties with Least Squares Solutions

Use the notation $\mathbf{m}_\dagger = G^\dagger \mathbf{d}$.

Row Rank Deficiency:

This means that $d = n < m$. Comments:

- This means that null space of G is trivial, but that of G^T is not.
- Here $\mathbf{m}_\dagger$ is the unique least squares solution.
- And $\mathbf{m}_\dagger$ is the exact solution to $G\mathbf{m} = \mathbf{d}$ exactly if $\mathbf{d}$ is in the range of G.
- But $\mathbf{m}$ is insensitive to any translation $\mathbf{d} + \mathbf{d}_0$ with $\mathbf{d}_0 \in N\left(G^\dagger\right)$

Heart of the Difficulties with Least Squares Solutions

Use the notation $\mathbf{m}_\dagger = G^\dagger \mathbf{d}$.

Row Rank Deficiency:

This means that $d = n < m$. Comments:

- This means that null space of G is trivial, but that of G^T is not.
- Here $\mathbf{m}_\dagger$ is the unique least squares solution.
- And $\mathbf{m}_\dagger$ is the exact solution to $G\mathbf{m} = \mathbf{d}$ exactly if $\mathbf{d}$ is in the range of G.
- But $\mathbf{m}$ is insensitive to any translation $\mathbf{d} + \mathbf{d}_0$ with $\mathbf{d}_0 \in N\left(G^\dagger\right)$

Heart of the Difficulties with Least Squares Solutions

Use the notation $\mathbf{m}_\dagger = G^\dagger \mathbf{d}$.

Row Rank Deficiency:

This means that $d = n < m$. Comments:

- This means that null space of G is trivial, but that of G^T is not.
- Here $\mathbf{m}_\dagger$ is the unique least squares solution.
- And $\mathbf{m}_\dagger$ is the exact solution to $G\mathbf{m} = \mathbf{d}$ exactly if $\mathbf{d}$ is in the range of G.
- But $\mathbf{m}$ is insensitive to any translation $\mathbf{d} + \mathbf{d}_0$ with $\mathbf{d}_0 \in N\left(G^\dagger\right)$

Heart of the Difficulties with Least Squares Solutions

Column Rank Deficiency:

This means $p = m < n$. Comments:

- This means that null space of G^T is trivial, but that of G is not.
- Here $\mathbf{m}_\dagger$ is a solution of minimum 2-norm.
- And $\mathbf{m}_\dagger + \mathbf{m}_0$ is also a solution to $G\mathbf{m} = \mathbf{d}$ for any $\mathbf{m}_0 \in N(G)$.
- So $\mathbf{d}$ is insensitive to any translation $\mathbf{m}_\dagger + \mathbf{m}_0$ with $\mathbf{m}_0 \in N(G)$.

Heart of the Difficulties with Least Squares Solutions

Column Rank Deficiency:

This means $p = m < n$. Comments:

- This means that null space of G^T is trivial, but that of G is not.
- Here $\mathbf{m}_\dagger$ is a solution of minimum 2-norm.
- And $\mathbf{m}_\dagger + \mathbf{m}_0$ is also a solution to $G\mathbf{m} = \mathbf{d}$ for any $\mathbf{m}_0 \in N(G)$.
- So $\mathbf{d}$ is insensitive to any translation $\mathbf{m}_\dagger + \mathbf{m}_0$ with $\mathbf{m}_0 \in N(G)$.

Heart of the Difficulties with Least Squares Solutions

Column Rank Deficiency:

This means $p = m < n$. Comments:

- This means that null space of G^T is trivial, but that of G is not.
- Here $\mathbf{m}_\dagger$ is a solution of minimum 2-norm.
- And $\mathbf{m}_\dagger + \mathbf{m}_0$ is also a solution to $G\mathbf{m} = \mathbf{d}$ for any $\mathbf{m}_0 \in N(G)$.
- So $\mathbf{d}$ is insensitive to any translation $\mathbf{m}_\dagger + \mathbf{m}_0$ with $\mathbf{m}_0 \in N(G)$.

Column Rank Deficiency:

This means $p = m < n$. Comments:

- This means that null space of G^T is trivial, but that of G is not.
- Here $\mathbf{m}_\dagger$ is a solution of minimum 2-norm.
- And $\mathbf{m}_\dagger + \mathbf{m}_0$ is also a solution to $G\mathbf{m} = \mathbf{d}$ for any $\mathbf{m}_0 \in N(G)$.
- So $\mathbf{d}$ is insensitive to any translation $\mathbf{m}_\dagger + \mathbf{m}_0$ with $\mathbf{m}_0 \in N(G)$.

Heart of the Difficulties with Least Squares Solutions

Column Rank Deficiency:

This means $p = m < n$. Comments:

- This means that null space of G^T is trivial, but that of G is not.
- Here $\mathbf{m}_\dagger$ is a solution of minimum 2-norm.
- And $\mathbf{m}_\dagger + \mathbf{m}_0$ is also a solution to $G\mathbf{m} = \mathbf{d}$ for any $\mathbf{m}_0 \in N(G)$.
- So $\mathbf{d}$ is insensitive to any translation $\mathbf{m}_\dagger + \mathbf{m}_0$ with $\mathbf{m}_0 \in N(G)$.

Heart of the Difficulties with Least Squares Solutions

Row and Column Rank Deficiency:

This means $p < \min\{m, n\}$. Comments:

- This means that null space of both G and G^T are nontrivial.
- Here $\mathbf{m}_\dagger$ is a least squares solution.
- We have trouble in both directions.

Heart of the Difficulties with Least Squares Solutions

Row and Column Rank Deficiency:

This means $p < \min\{m, n\}$. Comments:

- This means that null space of both G and G^T are nontrivial.
- Here $\mathbf{m}_\dagger$ is a least squares solution.
- We have trouble in both directions.

Heart of the Difficulties with Least Squares Solutions

Row and Column Rank Deficiency:

This means $p < \min\{m, n\}$. Comments:

- This means that null space of both G and G^T are nontrivial.
- Here $\mathbf{m}_\dagger$ is a least squares solution.
- We have trouble in both directions.

Heart of the Difficulties with Least Squares Solutions

Row and Column Rank Deficiency:

This means $p < \min\{m, n\}$. Comments:

- This means that null space of both G and G^T are nontrivial.
- Here $\mathbf{m}_\dagger$ is a least squares solution.
- We have trouble in both directions.

Outline

Covariance and Resolution

Definition

The **model resolution matrix** for the problem $G\mathbf{m} = \mathbf{d}$ is $R_\mathbf{m} = G^\dagger G$

Consequences:

- $R_\mathbf{m} = V_p V_p^T$, which is just I_n if G has full column rank.
- If $G\mathbf{m}_{\mathbf{true}} = \mathbf{d}$, then $E[\mathbf{m}_\dagger] = R_m \mathbf{m}_{true}$
- Thus, the bias in the gereralized inverse solution is $E[\mathbf{m}_\dagger] - \mathbf{m}_{true} = (R_\mathbf{m} - I)\,\mathbf{m}_{true} = -V_0 V_0^T \mathbf{m}_{true}$ with $V = [V_p V_0]$.
- Similarly, in the case of identically distributed data with variance σ^2, the covariance matrix is $\mathrm{Cov}(\mathbf{m}_\dagger) = \sigma^2 G^\dagger \left(G^\dagger\right)^T = \sigma^2 \sum_{i=1}^p \frac{\mathbf{v}_i \mathbf{v}_i^T}{s_i^2}$.
- From expected values we obtain a **resolution test**: if a diagonal entry are close to 1, we claim good resolution of that coordinate, otherwise not.

Covariance and Resolution

Definition

The **model resolution matrix** for the problem $G\mathbf{m} = \mathbf{d}$ is $R_{\mathbf{m}} = G^{\dagger}G$

Consequences:

- $R_{\mathbf{m}} = V_p V_p^T$, which is just I_n if G has full column rank.
- If $G\mathbf{m_{true}} = \mathbf{d}$, then $E[\mathbf{m}_{\dagger}] = R_m \mathbf{m}_{true}$
- Thus, the bias in the gereralized inverse solution is $E[\mathbf{m}_{\dagger}] - \mathbf{m}_{true} = (R_{\mathbf{m}} - I)\,\mathbf{m}_{true} = -V_0 V_0^T \mathbf{m}_{true}$ with $V = [V_p V_0]$.
- Similarly, in the case of identically distributed data with variance σ^2, the covariance matrix is $\mathrm{Cov}(\mathbf{m}_{\dagger}) = \sigma^2 G^{\dagger}\left(G^{\dagger}\right)^T = \sigma^2 \sum_{i=1}^{p} \frac{\mathbf{v}_i \mathbf{v}_i^T}{s_i^2}$.
- From expected values we obtain a **resolution test**: if a diagonal entry are close to 1, we claim good resolution of that coordinate, otherwise not.

Covariance and Resolution

Definition

The **model resolution matrix** for the problem $G\mathbf{m} = \mathbf{d}$ is $R_{\mathbf{m}} = G^{\dagger}G$

Consequences:

- $R_{\mathbf{m}} = V_p V_p^T$, which is just I_n if G has full column rank.
- If $G\mathbf{m}_{\mathbf{true}} = \mathbf{d}$, then $E[\mathbf{m}_{\dagger}] = R_m \mathbf{m}_{true}$
- Thus, the bias in the gereralized inverse solution is $E[\mathbf{m}_{\dagger}] - \mathbf{m}_{true} = (R_{\mathbf{m}} - I)\,\mathbf{m}_{true} = -V_0 V_0^T \mathbf{m}_{true}$ with $V = [V_p V_0]$.
- Similarly, in the case of identically distributed data with variance σ^2, the covariance matrix is $\mathrm{Cov}(\mathbf{m}_{\dagger}) = \sigma^2 G^{\dagger}\left(G^{\dagger}\right)^T = \sigma^2 \sum_{i=1}^{p} \frac{\mathbf{V}_i \mathbf{V}_i^T}{s_i^2}$.
- From expected values we obtain a **resolution test**: if a diagonal entry are close to 1, we claim good resolution of that coordinate, otherwise not.

Covariance and Resolution

Definition

The **model resolution matrix** for the problem $G\mathbf{m} = \mathbf{d}$ is $R_{\mathbf{m}} = G^{\dagger}G$

Consequences:

- $R_{\mathbf{m}} = V_p V_p^T$, which is just I_n if G has full column rank.
- If $G\mathbf{m}_{\mathbf{true}} = \mathbf{d}$, then $E[\mathbf{m}_{\dagger}] = R_m \mathbf{m}_{true}$
- Thus, the bias in the gereralized inverse solution is $E[\mathbf{m}_{\dagger}] - \mathbf{m}_{true} = (R_{\mathbf{m}} - I)\,\mathbf{m}_{true} = -V_0 V_0^T \mathbf{m}_{true}$ with $V = [V_p V_0]$.
- Similarly, in the case of identically distributed data with variance σ^2, the covariance matrix is $\mathrm{Cov}(\mathbf{m}_{\dagger}) = \sigma^2 G^{\dagger}\left(G^{\dagger}\right)^T = \sigma^2 \sum_{i=1}^{p} \frac{\mathbf{v}_i \mathbf{v}_i^T}{s_i^2}$.
- From expected values we obtain a **resolution test**: if a diagonal entry are close to 1, we claim good resolution of that coordinate, otherwise not.

Covariance and Resolution

Definition

The **model resolution matrix** for the problem $G\mathbf{m} = \mathbf{d}$ is $R_{\mathbf{m}} = G^{\dagger}G$

Consequences:

- $R_{\mathbf{m}} = V_p V_p^T$, which is just I_n if G has full column rank.
- If $G\mathbf{m}_{\mathbf{true}} = \mathbf{d}$, then $E[\mathbf{m}_{\dagger}] = R_m \mathbf{m}_{true}$
- Thus, the bias in the gereralized inverse solution is $E[\mathbf{m}_{\dagger}] - \mathbf{m}_{true} = (R_{\mathbf{m}} - I)\,\mathbf{m}_{true} = -V_0 V_0^T \mathbf{m}_{true}$ with $V = [V_p V_0]$.
- Similarly, in the case of identically distributed data with variance σ^2, the covariance matrix is $\mathrm{Cov}(\mathbf{m}_{\dagger}) = \sigma^2 G^{\dagger}\left(G^{\dagger}\right)^T = \sigma^2 \sum_{i=1}^{p} \frac{\mathbf{v}_i \mathbf{v}_i^T}{s_i^2}$.
- From expected values we obtain a **resolution test**: if a diagonal entry are close to 1, we claim good resolution of that coordinate, otherwise not.

Covariance and Resolution

Definition

The **model resolution matrix** for the problem $G\mathbf{m} = \mathbf{d}$ is $R_{\mathbf{m}} = G^{\dagger}G$

Consequences:

- $R_{\mathbf{m}} = V_p V_p^T$, which is just I_n if G has full column rank.
- If $G\mathbf{m}_{\mathbf{true}} = \mathbf{d}$, then $E\left[\mathbf{m}_{\dagger}\right] = R_m \mathbf{m}_{true}$
- Thus, the bias in the gereralized inverse solution is $E\left[\mathbf{m}_{\dagger}\right] - \mathbf{m}_{true} = \left(R_{\mathbf{m}} - I\right)\mathbf{m}_{true} = -V_0 V_0^T \mathbf{m}_{true}$ with $V = [V_p V_0]$.
- Similarly, in the case of identically distributed data with variance σ^2, the covariance matrix is $\mathrm{Cov}\left(\mathbf{m}_{\dagger}\right) = \sigma^2 G^{\dagger}\left(G^{\dagger}\right)^T = \sigma^2 \sum_{i=1}^{p} \frac{\mathbf{V}_i \mathbf{V}_i^T}{s_i^2}$.
- From expected values we obtain a **resolution test**: if a diagonal entry are close to 1, we claim good resolution of that coordinate, otherwise not.

Outline

Instability of Generalized Inverse Solution

The key results:

- For $n \times n$ square matrix G
  $\text{cond}_2(G) = \|G\|_2 \left\|G^{-1}\right\|_2 = \sigma_1/\sigma_n$.
- This inspires the definition: the condition number of an $m \times n$ matrix G is σ_1/σ_q where $q = \min\{m, n\}$.
- Note: if $\sigma_q = 0$, the condition number is infinity. Is this notion useful?
- If data $\mathbf{d}$ vector is perturbed to $\mathbf{d}'$, resulting in a perturbation of the generalized inverse solution $\mathbf{m}_\dagger$ to $\mathbf{m}'_\dagger$, then
  $\frac{\left\|\mathbf{m}'_\dagger - \mathbf{m}_\dagger\right\|_2}{\left\|\mathbf{m}_\dagger\right\|_2} \leq \text{cond}(G) \frac{\|\mathbf{d}' - \mathbf{d}\|_2}{\|\mathbf{d}\|_2}$.

Instability of Generalized Inverse Solution

The key results:

- For $n \times n$ square matrix G
  $\mathrm{cond}_2(G) = \|G\|_2 \left\|G^{-1}\right\|_2 = \sigma_1/\sigma_n$.
- This inspires the definition: the condition number of an $m \times n$ matrix G is σ_1/σ_q where $q = \min\{m, n\}$.
- Note: if $\sigma_q = 0$, the condition number is infinity. Is this notion useful?
- If data $\mathbf{d}$ vector is perturbed to $\mathbf{d}'$, resulting in a perturbation of the generalized inverse solution $\mathbf{m}_\dagger$ to $\mathbf{m}'_\dagger$, then
  $\frac{\left\|\mathbf{m}'_\dagger - \mathbf{m}_\dagger\right\|_2}{\left\|\mathbf{m}_\dagger\right\|_2} \le \mathrm{cond}(G) \frac{\|\mathbf{d}' - \mathbf{d}\|_2}{\|\mathbf{d}\|_2}$.

Instability of Generalized Inverse Solution

The key results:

- For $n \times n$ square matrix G
  $\mathrm{cond}_2(G) = \|G\|_2 \left\|G^{-1}\right\|_2 = \sigma_1/\sigma_n$.
- This inspires the definition: the condition number of an $m \times n$ matrix G is σ_1/σ_q where $q = \min\{m, n\}$.
- Note: if $\sigma_q = 0$, the condition number is infinity. Is this notion useful?
- If data $\mathbf{d}$ vector is perturbed to $\mathbf{d}'$, resulting in a perturbation of the generalized inverse solution $\mathbf{m}_\dagger$ to $\mathbf{m}'_\dagger$, then
  $\frac{\left\|\mathbf{m}'_\dagger - \mathbf{m}_\dagger\right\|_2}{\left\|\mathbf{m}_\dagger\right\|_2} \leq \mathrm{cond}(G) \frac{\|\mathbf{d}' - \mathbf{d}\|_2}{\|\mathbf{d}\|_2}$.

Instability of Generalized Inverse Solution

The key results:

- For $n \times n$ square matrix G
  $\text{cond}_2(G) = \|G\|_2 \left\|G^{-1}\right\|_2 = \sigma_1/\sigma_n$.
- This inspires the definition: the condition number of an $m \times n$ matrix G is σ_1/σ_q where $q = \min\{m, n\}$.
- Note: if $\sigma_q = 0$, the condition number is infinity. Is this notion useful?
- If data $\mathbf{d}$ vector is perturbed to $\mathbf{d}'$, resulting in a perturbation of the generalized inverse solution $\mathbf{m}_\dagger$ to $\mathbf{m}'_\dagger$, then $\frac{\left\|\mathbf{m}'_\dagger - \mathbf{m}_\dagger\right\|_2}{\left\|\mathbf{m}_\dagger\right\|_2} \leq \text{cond}(G) \frac{\|\mathbf{d}' - \mathbf{d}\|_2}{\|\mathbf{d}\|_2}$.

Instability of Generalized Inverse Solution

The key results:

- For $n \times n$ square matrix G
  $\operatorname{cond}_2(G) = \|G\|_2 \left\|G^{-1}\right\|_2 = \sigma_1/\sigma_n$.
- This inspires the definition: the condition number of an $m \times n$ matrix G is σ_1/σ_q where $q = \min\{m, n\}$.
- Note: if $\sigma_q = 0$, the condition number is infinity. Is this notion useful?
- If data $\mathbf{d}$ vector is perturbed to $\mathbf{d}'$, resulting in a perturbation of the generalized inverse solution $\mathbf{m}_\dagger$ to $\mathbf{m}'_\dagger$, then
  $\frac{\left\|\mathbf{m}'_\dagger - \mathbf{m}_\dagger\right\|_2}{\left\|\mathbf{m}_\dagger\right\|_2} \le \operatorname{cond}(G) \frac{\|\mathbf{d}' - \mathbf{d}\|_2}{\|\mathbf{d}\|_2}$.

Stability Issues

How these facts affect stability:

- If cond (G) is not too large, then the solution is stable to perturbations in data.
- If $\sigma_1 \gg \sigma_p$, there is a potential for instability. It is diminished if the data itself has small components in the direction of singular vectors corresponding to small singular values.
- If $\sigma_1 \gg \sigma_p$, and there is a clear delineation between "small" singular values and the rest, we simple discard the small singular values and treat the problem as one of smaller rank with "good" singular values.
- If $\sigma_1 \gg \sigma_p$, and there is no clear delineation between "small" singular values and the rest, we have to discard some of them, but which ones? This leads to regularization issues. In any case, any method that discards small singular values produces a **truncated SVD** (TSVD) solution.

Stability Issues

How these facts affect stability:

- If cond (G) is not too large, then the solution is stable to perturbations in data.
- If $\sigma_1 \gg \sigma_p$, there is a potential for instability. It is diminished if the data itself has small components in the direction of singular vectors corresponding to small singular values.
- If $\sigma_1 \gg \sigma_p$, and there is a clear delineation between "small" singular values and the rest, we simple discard the small singular values and treat the problem as one of smaller rank with "good" singular values.
- If $\sigma_1 \gg \sigma_p$, and there is no clear delineation between "small" singular values and the rest, we have to discard some of them, but which ones? This leads to regularization issues. In any case, any method that discards small singular values produces a **truncated SVD** (TSVD) solution.

Stability Issues

How these facts affect stability:

- If cond(G) is not too large, then the solution is stable to perturbations in data.
- If $\sigma_1 \gg \sigma_p$, there is a potential for instability. It is diminished if the data itself has small components in the direction of singular vectors corresponding to small singular values.
- If $\sigma_1 \gg \sigma_p$, and there is a clear delineation between "small" singular values and the rest, we simple discard the small singular values and treat the problem as one of smaller rank with "good" singular values.
- If $\sigma_1 \gg \sigma_p$, and there is no clear delineation between "small" singular values and the rest, we have to discard some of them, but which ones? This leads to regularization issues. In any case, any method that discards small singular values produces a **truncated SVD** (TSVD) solution.

Stability Issues

How these facts affect stability:

- If cond (G) is not too large, then the solution is stable to perturbations in data.
- If $\sigma_1 \gg \sigma_p$, there is a potential for instability. It is diminished if the data itself has small components in the direction of singular vectors corresponding to small singular values.
- If $\sigma_1 \gg \sigma_p$, and there is a clear delineation between "small" singular values and the rest, we simple discard the small singular values and treat the problem as one of smaller rank with "good" singular values.
- If $\sigma_1 \gg \sigma_p$, and there is no clear delineation between "small" singular values and the rest, we have to discard some of them, but which ones? This leads to regularization issues. In any case, any method that discards small singular values produces a **truncated SVD** (TSVD) solution.

Stability Issues

How these facts affect stability:

- If cond (G) is not too large, then the solution is stable to perturbations in data.
- If $\sigma_1 \gg \sigma_p$, there is a potential for instability. It is diminished if the data itself has small components in the direction of singular vectors corresponding to small singular values.
- If $\sigma_1 \gg \sigma_p$, and there is a clear delineation between "small" singular values and the rest, we simple discard the small singular values and treat the problem as one of smaller rank with "good" singular values.
- If $\sigma_1 \gg \sigma_p$, and there is no clear delineation between "small" singular values and the rest, we have to discard some of them, but which ones? This leads to regularization issues. In any case, any method that discards small singular values produces a **truncated SVD** (TSVD) solution.