Low Rank and Compressed Sensing

1. Low rank matrices. Example, $A = uv^*$.
2. Matrices with exponentially decreasing singular values (low effective rank).

   Example, Hilbert matrix $(i + j - 1)^{-1}$.
3. Completion problem. Fill in missing entries to get a low rank matrices. One approach:

   minimize the nuclear norm $\|A\|_N$ over all completions of a given partial matrix.

III.1 Changes in A^{-1} from changes in A

Basic case. $A = I_n$.

If $M = I - uv^T$ then $M^{-1} = I + \frac{uv^T}{1-v^Tu}$.

Proof 1. $(I - uv^T)(I + \frac{uv^T}{1-v^Tu})$

$$= I - uv^T - uv^T \frac{uv^T}{1-v^Tu} + \frac{uv^T}{1-v^Tu}$$

$$= I - uv^T - v^Tu \frac{uv^T}{1-v^Tu} + \frac{uv^T}{1-v^Tu}$$

$$= I. \qquad \square$$

Proof 2. Let $E = \begin{bmatrix} I_n & u \\ v^t & 1 \end{bmatrix}$. Then $\begin{bmatrix} I_n & 0 \\ -v^T & 1 \end{bmatrix} E = \begin{bmatrix} I_n & u \\ 0 & d \end{bmatrix}$

with $d = 1 - v^Tu$, and

$$E^{-1} = \begin{bmatrix} I_n & u \\ 0 & d \end{bmatrix}^{-1} \begin{bmatrix} I_n & 0 \\ -v^T & 1 \end{bmatrix} = \begin{bmatrix} I_n & -u/d \\ 0 & 1/d \end{bmatrix} \begin{bmatrix} I_n & 0 \\ -v^T & 1 \end{bmatrix}$$

$$= \begin{bmatrix} I_n + uv^T/d) & -u/d \\ -v^T/d & 1/d \end{bmatrix}.$$

Note also that $\begin{bmatrix} I_n & -u \\ 0 & 1 \end{bmatrix} E = \begin{bmatrix} I_n - uv^T & 0 \\ v^T & 1 \end{bmatrix}$. So,

$$E^{-1} = \begin{bmatrix} I_n - uv^T & 0 \\ v^T & 1 \end{bmatrix}^{-1} \begin{bmatrix} I_n & -u \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} (I_n - uv^T)^{-1} & * \\ * & * \end{bmatrix}.$$

Comparing the $(1, 1)$ blocks of E^{-1}, we get the equality. $\square$

Note that $\det(E) = 1 - v^tu$.

Extension to $I_n - UV^T$, where U, V are $n \times k$.

If $M = I_n - UV^T$ then $M^{-1} = I_n + U(1_k - V^T U)^{-1} V^T$.

Proof 1. $(I_n - UV^T)(I_n + U(I_k - V^T U)^{-1} V^T)$

$$= I_n - UV^T + (I_n - UV^T)U(I_k - V^T U)^{-1} V^T$$
$$= I_n - UV^T + U(I_k - V^T U)(I_k - V^T U)^{-1} V^T$$
$$= I_n. \qquad \square$$

Proof 2. Let $E = \begin{bmatrix} I_n & U \\ V^t & I_k \end{bmatrix}$.

We can use block Gaussian elimination to get an upper triangular block form. (Think about 2×2 matrices.)

So, $\begin{bmatrix} I_n & 0 \\ -V^T & I_k \end{bmatrix} E = \begin{bmatrix} I_n & U \\ 0 & D \end{bmatrix}$

with $D = 1_k - V^T U$, and

$$E^{-1} = \begin{bmatrix} I_n & U \\ 0 & D \end{bmatrix}^{-1} \begin{bmatrix} I_n & 0 \\ -V^T & I_k \end{bmatrix} = \begin{bmatrix} I_n & -UD^{-1} \\ 0 & D^{-1} \end{bmatrix} \begin{bmatrix} I_n & 0 \\ -V^T & I_k \end{bmatrix}$$
$$= \begin{bmatrix} I_n + UD^{-1}V^T & -UD^{-1} \\ -D^{-1}V^T & D^{-1} \end{bmatrix}.$$

Note also that we can use Gaussian elimination to get lower triangular block form so that

$$\begin{bmatrix} I_n & -U \\ 0 & I_k \end{bmatrix} E = \begin{bmatrix} I_n - UV^T & 0 \\ V^T & I_k \end{bmatrix}. \text{ So,}$$

$$E^{-1} = \begin{bmatrix} I_n - UV^T & 0 \\ V^T & I_k \end{bmatrix}^{-1} \begin{bmatrix} I_n & -U \\ 0 & I_k \end{bmatrix}$$
$$= \begin{bmatrix} (I_n - UV^T)^{-1} & * \\ * & * \end{bmatrix}.$$

Comparing the $(1,1)$ blocks of E^{-1}, we get the equality. $\square$

Note $\det(E) = \det(I_n - UV^T) = \det(I_k - V^t U)$.

We know the second equality because UV^T and VU^T have the same nonzero eigenvalues.

Example 1. If $M = I - [111]^T[111]$,

then $M^{-1} = I - [111]^T[111]/2$.
Here $u = v = [111]^T$.

Example 2. If $M = I - (E_{12} + E_{13} + E_{23})$,

then $M^{-1} = I + E_{12} + E_{23}$.
Here we use CR decomposition to figure out $U = ??, V = ??$

Most general case. If $M = A - UV^T$, then

$$
\begin{aligned}
M^{-1} &= (A - UV^T)^{-1} = [A(I - A^{-1}UV^T)]^{-1} \\
&= [(I + A^{-1}U(I_k - V^TA^{-1}U)^{-1})V^T]A^{-1} \\
&= A^{-1} + A^{-1}U(I - V^TA^{-1}U)^{-1}V^TA^{-1}.
\end{aligned}
$$

Some related identities:

$$B(I_m + AB) = (I_n + BA)B$$

$$B(I_m + AB)^{-1} = (I_n + BA)^{-1}B$$

$$A^T(AA^T + \lambda I_n)^{-1} = (A^TA + \lambda I_m)^{-1}A^T.$$

$$U(I_k + V^TU) = (I_n - UV^T)U.$$

Derivative of $f(A) = A^{-1}$

Consider $f(A) = A^{-1}$, and its change other than $A + UV$.

In general, $B^{-1} - A^{-1} = B^{-1}(A - B)A^{-1}$.

If $A(t)$ is a function of t, then

$$\frac{d}{dt}A^{-1} = \lim_{\Delta t \to 0} \frac{(A + \Delta A)^{-1} - A^{-1}}{\Delta t}.$$

Now,

$$\frac{1}{\Delta t}(A + \Delta A)^{-1} - A^{-1} = \frac{-1}{\Delta t}(A + \Delta A)^{-1}(\Delta A)A^{-1}$$

approaches

$$\frac{dA^{-1}}{dt} = -A^{-1}\frac{dA}{dt}A^{-1}.$$

When $n = 1$, we get $(1/x)' = -1/x^2$.

Applications

Updating least square To find the least square solution for $Ax = b$, one solves $A^T Ax = A^T b$, where A is $m \times n$.

If we have one more row r added to A resulting in $\tilde{A} = \begin{bmatrix} A \\ r \end{bmatrix}$ and $\tilde{A}x = \begin{bmatrix} b \\ b_{m+1} \end{bmatrix} = \tilde{b}$. Then we solve

$$\tilde{A}^T \tilde{b} = (\tilde{A}^T \tilde{A})x = (A^T A + r^T r)x.$$

If we already know $(A^T A)^{-1}$, it is easy to update

$$(A^T A + r^T r)^{-1}.$$

An extension

In some problems such as the GPS calculation, one needs to solve

$$\begin{bmatrix} A & 0 \\ -I & I \\ 0 & r \end{bmatrix} \begin{bmatrix} x_{old} \\ x_{new} \end{bmatrix} = \begin{bmatrix} b \\ v\Delta t \\ b_{m+1} \end{bmatrix},$$

where $v\Delta t$ is the change of position equal to the product of the velocity v and Δt, and the newly added (last) equation further improves the solution x_{new}.

The scheme is further improved by changing the equation to

$$A^T V^{-1} A\hat{x} = A^T V^{-1} b$$

using the variance or covariance matrix V to increase the reliability. It leads to the Kalman gain matrix K so that

$$\hat{x}_{new} = \hat{x}_{old} + K(b_{m+1} - r\hat{x}_{old}).$$

Quasi-Newton Update

In calculus, to solve $f(x) = 0$, one use the iterative formula:

$$x_{new} = x_{old} - \frac{f(x_{old})}{f'(x_{old})}.$$

In practical problem such as deep learning, one considers multiple functions $f_1, \ldots, f_n$ on multiple variables $x_1, \ldots, x_n$, we use the Jacobian matrix $J(x_{old}) = \left[\frac{\partial f_i}{\partial x_j}\right]$ and compute

$$x_{new} = x_{old} - J(x_{old})^{-1} f(x_{old}).$$

One needs to update $J(x_{new})$, which is very computationally expensive (efficiency and accuracy). So, we focus on the Quasi-newton condition

$$J_{new}(x_{new} - x_{old}) = f_{new} - f_{old}$$

and estimate J_{new}^{-1}. Here, we assume that $J_{new}\Delta x = \Delta f$.

Derivative of eigenvalues and eigenvectors

Assume that $A(t) \in M_n$ is a differentiable function depending on $t \in (-d, d)$ such that $A(t)$ has distinct eigenvalues in $(-d, d)$.

Theorem Suppose $X^{-1}(t)A(t)X(t) = \Lambda(t)$ with the $(1,1)$ entry of $\Lambda(t)$, the first column of $X(t) = x(t)$, the first row of $X^{-1}(t)$ is $y(t)^T$. Then $\quad y(t)^T x(t) = 1$,

$$A(t)x(t) = \lambda(t)x(t), \qquad y(t)^T A(t) = y(t)^T \lambda(t),$$

$$\lambda(t) = y(t)^T A(t)x(t) \quad \text{and} \quad \frac{d\lambda}{dt} = y(t)^T \frac{dA}{dt} x(t).$$

Proof. Since

$$0 = \frac{d(y(t)^T x(t))}{dt} = y(t)^T \frac{dx(t)}{dt} + \frac{dy(t)^T}{dt} x(t),$$

we see that

$$\frac{d\lambda}{dt} = \left[y(t)^T A(t) \frac{dx(t)}{dt} + \frac{dy(t)^T}{dt} A(t)x(t) \right] + y(t)^T \frac{dA}{dt} x(t)$$

$$= \lambda(t) \left[y(t)^T \frac{dx(t)}{dt} + \frac{dy(t)^T}{dt} x(t) \right] + y(t)^T \frac{dA}{dt} x(t).$$

Example If $A = \begin{bmatrix} 2t & 1 \\ 2t & 2 \end{bmatrix}$, then $\lambda(t) = 1 + t \pm \sqrt{1+t^2}$.

For $\lambda_1(0) = 2$, $y_1 = [0, 1]$ and $x_1 = [1/2, 1]^T$.

For $\lambda_2(0) = 0$, $y_2 = [1, -1/2]$ and $x_2 = [1, 0]^T$.

We have

$$\frac{d\lambda_1}{dt} = y_1^T \frac{dA}{dt} x_1 = [0, 1] \begin{bmatrix} 2 & 0 \\ 2 & 0 \end{bmatrix} \begin{bmatrix} 1/2 \\ 1 \end{bmatrix} = 1,$$

$$\frac{d\lambda_2}{dt} = y_1^T \frac{dA}{dt} x_1 = [1, -1/2] \begin{bmatrix} 2 & 0 \\ 2 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = 1,$$

Similarly, suppose $A(t) = U(t)\Sigma(t)V(t)^T$ with orthonormal matrices $U(t), V(t) \in M_n(\mathbb{R})$ and $\Sigma(t) = \operatorname{diag}(\sigma_1(t), \ldots, \sigma_n(t))$. Then

$$\frac{d\sigma_1}{dt} = u_1(t)^T A(t) v_1(t).$$

Eigenvalues inequality

Denote by $\lambda_1(A) \geq \cdots \geq \lambda_n(A)$ the eigenvalues of $A \in M_n$ satisfying $A = A^*$.

Theorem Let $A \in M_n$ and $x \in \mathbb{F}^n$. Then

$$\lambda_k(A + xx^*) \geq \lambda_k(A) \quad \text{for } k = 1, \ldots, n.$$

Proof. Suppose $B = A + xx^*$

$$A = \sum_{j=1}^n \lambda_j(A) u_j u_j^* \quad \text{and} \quad B = \sum_{j=1}^n \lambda_j(B) v_j v_j^*,$$

where $\{u_1 \cdots u_n\}$ and $\{v_1 \cdots v_n\}$ are orthonormal bases.

Then there is a nonzero vector $\mathbf{c} = [c_1 \cdots c_{n+1}]^T$ such that

$$[u_1 \cdots u_k \ v_k \cdots v_n]\mathbf{c} = 0.$$

Thus, $y = c_1 u_1 + \cdots + c_k u_k = -(c_{k+1} v_k + \cdots + c_{n+1} v_n) \neq 0$.

We may assume that $\|v\| = 1$, i.e.,

$$\sum_{j=1}^k |c_j|^2 = 1 = \sum_{j=k}^n |c_{j+1}|^2.$$

Then

$$\begin{aligned}
\lambda_k(B) &\geq \sum_{j=k}^n |c_{j+1}|^2 \lambda_j(B) \\
&= (\sum_{j=k}^n -c_{j+1} v_j)^* B (\sum_{j=k}^n -c_{j+1} v_j) \\
&= y^* B y = y(A + xx^*)y \\
&= y^* A y + y^* x x^* y \\
&\geq y^* A y \\
&= \sum_{j=1}^k |c_j|^2 \lambda_j(A) \\
&\geq \lambda_k(A).
\end{aligned}$$

□

Corollary Let $A = A^*, B = B^* \in M_n$, where B is positive semi-definite. Then for $k = 1, \ldots, n$,

$$\lambda_k(A + B) \geq \lambda_k(A).$$

Interlacing inequalities

Theorem Suppose $B = B^* \in M_n$, and $A = V^*BV \in M_{n-1}$, where $V \in M_{n,n-1}$ have orthonormal columns. Then

$$\lambda_1(B) \geq \lambda_1(A) \geq \cdots \geq \lambda_{n-1}(A) \geq \lambda_n(B).$$

Proof. We may replace B by $B - \lambda_n(B)I$ and assume that $\lambda_n(B) = 0$.

Let $\tilde{V} = [V v_n]$ be unitary, and $\tilde{B} = \tilde{V}^* B \tilde{V} = \begin{bmatrix} A & * \\ * & * \end{bmatrix}$. Then

$\tilde{B} = [B_1\ x]^*[B_1\ x]$ for some $B_1 \in M_{n,n-1}$ so that $B_1^* B_1 = A$.

Now $B, \tilde{B}$ and $[B_1\ x][B_1\ x]^* = B_1 B_1^* + xx^*$ have the same eigenvalues. Also, $B_1 B_1^* \in M_n$ and $B_1^* B_1 = A \in M_{n-1}$ have the same nonzero eigenvalues. Thus, for $k = 1, \ldots, n-1$,

$$\lambda_k(B) = \lambda_k(B_1 B_1^* + xx^*) \geq \lambda_k(B_1 B_1^*) = \lambda_k(A).$$

Applying the results to $-B$ and $-A$, we see that

$$\lambda_k(-B) \geq \lambda_k(-A) \qquad \text{for } k = 1, \ldots, n-1.$$

So, $\qquad \lambda_j(A) \geq \lambda_{j+1}(B) \qquad \text{for } j = 1, \ldots, n-1.$ □

Corollary If $B = B^* \in M_n$ and A is obtained from B by removing the ith row and ith column, then

$$\lambda_1(B) \geq \lambda_1(A) \geq \cdots \geq \lambda_{n-1}(A) \geq \lambda_n(B).$$

Weyl's inequalities

Theorem Let $A = A^*, B = B^* \in M_n$. Then for any $i, j \in \{1, \ldots, n\}$ with $i + j < n$,

$$\lambda_i(A) + \lambda_j(B) \geq \lambda_{i+j-1}(A + B).$$

In particular,

$$\lambda_1(B) \geq \lambda_i(A + B) - \lambda_i(A) \geq \lambda_n(B).$$

Proof. We prove the result by induction on n. The result is clear when $n = 1$.

Suppose $n > 1$. We may assume that $i \geq j$. Let $C = A + B$,

$$A = \sum_{j=1}^{n} \lambda_j(A) u_j u_j^* \quad \text{and} \quad C = \sum_{j=1}^{n} \lambda_j(C) v_j v_j^*,$$

where $\{u_1, \ldots, u_n\}, \{v_1, \ldots, v_n\}$ are orthonormal bases.

If $i = j = 1$, then

$$\lambda_1(C) = v_1^* C v_1 = v_1^* A v_1 + v_1^* B v_1 \leq \lambda_1(A) + \lambda_1(B).$$

So, assume that $i \geq j$ and $i > 2$. Let $X \in M_{n,n-1}$ with column space containing $v_1, \ldots, v_{i-2}, u_i, \ldots, u_n$.

Let $\tilde{A} = X^* A X, \tilde{B} = X^* B X$ and $\tilde{C} = \tilde{A} + \tilde{B}$.

Then $\lambda_\ell(A) = \lambda_{\ell-1}(\tilde{A})$ for $\ell = n, n-1, \ldots, i$.

Because $i + j - 2 \leq n - 1$, by induction assumption,

$$\lambda_{i-1}(\tilde{A}) + \lambda_j(\tilde{B}) \geq \lambda_{i+j-2}(\tilde{C}).$$

The result follows as $\lambda_i(A) = \lambda_{i-1}(\tilde{A})$, $\lambda_j(B) \geq \lambda_j(\tilde{B})$, and $\lambda_{i+j-2}(\tilde{C}) \geq \lambda_{i+j-1}(C)$.

Putting $j = 1$, we get $\lambda_1(B) \geq \lambda_i(A + B) - \lambda_i(A)$.

Applying the above to $-B$ and $-A$, we get

$$\lambda_n(B) \leq \lambda_i(A + B) - \lambda_i(A). \qquad \square$$

Liskii's inequalities

Theorem Suppose $A = A^*, B = B^* \in M_n$. Then for any $1 \le i_1 < \cdots < i_k \le n$,

$$\sum_{j=1}^{k} \lambda_{n-j+1}(B) \le \sum_{j=1}^{k} \lambda_{i_j}(A+B) - \sum_{j=1}^{k} \lambda_{i_j}(A) \le \sum_{j=1}^{k} \lambda_j(B).$$

Proof. Replace (A, B) by $(A - \lambda_k(B)I, B - \lambda_k(B)I)$. We may assume that B has k nonnegative eigenvalues. If

$$B = \sum_{j=1}^{n} \lambda_j(B) v_j v_j^* \quad \text{and} \quad B_+ = \sum_{j=1}^{k} \lambda_j(B) v_j v_j^*,$$

where $\{v_1, \ldots, v_n\}$ is an orthonormal basis, then

$$\begin{aligned}
& \sum_{j=1}^{k} (\lambda_{i_j}(A+B) - \lambda_{i_j}(A)) \\
\le\ & \sum_{j=1}^{k} (\lambda_{i_j}(A+B_+) - \lambda_{i_j}(A)) \\
& \qquad \text{as } \lambda_j(A+B) \le \lambda_j(A+B_+) \text{ for all } j \\
\le\ & \operatorname{tr}(A+B_+) - \operatorname{tr} A \\
& \qquad \text{as } \lambda_j(A) \le \lambda_j(A+B_+) \text{ for all } j \\
\le\ & \operatorname{tr}(B_+) = \sum_{j=1}^{k} \lambda_j(B).
\end{aligned}$$

□

Singular value inequalities

If $A \in M_{m,n}$ has nonzero singular values $a_1 \geq \cdots \geq a_r > 0$, then $\begin{bmatrix} 0_m & A \\ A^* & 0_n \end{bmatrix}$ has nonzero eigenvalues $\pm a_1, \ldots, \pm a_r$.

Denote by $\sigma_1(A) \geq \sigma_s(A) \geq \cdots$ the singular values of A, where $\sigma_j(A) = 0$ if j is larger than the rank of A.

If $i + j - 1 \leq \min\{m, n\}$, then

$$\sigma_{i+j-1}(A + B) \leq \sigma_i(A) + \sigma_j(B).$$

For any $1 \leq i_1 < \cdots < i_k \leq \min\{m, n\}$,

$$|\sum_{j=1}^{k} (\sigma_{i_j}(A + B) - \sigma_{i_j}(A))| \leq \sum_{j=1}^{k} \sigma_j(B).$$

A max-min characterization

Theorem *Suppose $A = A^* \in M_n$, and $\{u_1, \ldots, u_n\} \subseteq \mathbb{F}^n$ is an orthonormal basis such that*

$$A = \sum_{j=1}^{n} \lambda_j(A) u_j u_j^*.$$

If $1 \le k \le n$, then

$$\lambda_k = \max_{\dim V = k} \min_{0 \ne x \in V} \frac{x^*Ax}{x^*x}.$$

The equality holds if V is spanned by $\{v_1, \ldots, v_k\}$ and $x = v_k$.

Sketch of Proof. Suppose $\dim V = k$ and V is spanned by $\{y_1, \ldots, y_k\}$.

Show that there is a unit vector of the form

$$x = b_1 y_1 + \cdots + b_k y_k = d_k u_k + \cdots + d_n u_n.$$

Argue that $x^*Ax \le \lambda_k(A)$.

Since this is true for any V with $\dim V = k$, we see that

$$\lambda_k \ge \max_{\dim V = k} \min_{0 \ne x \in V} \frac{x^*Ax}{x^*x}.$$

If V is spanned by $\{u_1, \ldots, u_k\}$, then argue that

$$\min\{x^*Ax : x \in V, \|x\| = 1\} = \lambda_k(A).$$

The result follows. □

Theorem *Suppose $H = H^* \in M_n$ has diagonal entries $d_1 \ge \cdots \ge d_n$, and eigenvalues $\lambda_1 \ge \cdots \ge \lambda_n$. For $1 \le k < n$,*

$$\sum_{j=1}^{k} d_j \le \sum_{j=1}^{k} \lambda_j.$$

Proof. Exercise or Final Examination question.

Theorem *Suppose $H = H^* \in M_n$. Then H has p positive eigenvalues, q negative eigenvalues, and r zero eigenvalues if and only if there is an invertible $S \in M_n$ such that*

$$S(I_p \oplus -I_q \oplus 0_r)S^*.$$

Proof. Exercise or Final Examination questions. Hint / Sketch:

Let $U^*HU = \operatorname{diag}(\lambda_1, \ldots, \lambda_n)$ for a unitary U, and $\lambda_1 \ge \cdots \ge \lambda_p > 0 > \lambda_{p+1} \ge \cdots \ge \lambda_{p+q}$, and $\lambda_{p+q+1} = \cdots = \lambda_n = 0$. Then we can choose S ....

Conversely, suppose $S^*HS = I_p \oplus -I_q \oplus O_r$. Show that H has at least p positive eigenvalues, at least q negative eigenvalues, and r zero eigenvalues.

III.3 Rapidly decaying singular values

A singular values $s_1 \geq s_2 \geq \cdots$ decay exponentially if there is $C > 0$ such that $s_k \leq Ce^{-ak}$.

Theorem *Let $X \in M_n$. If there are normal matrices $A, B \in M_n$ with no common eigenvalues such that $AX - XB = C$ has low rank, then the singular values of X decay exponentially.*

Recall that $A \in M_n$ is normal if $A^*A = AA^*$.

Examples Unitary and Hermitian matrices.

Theorem *A matrix $A \in M_n$ is normal if and only if there is a unitary U such that U^*AU is a diagonal matrix.*

Proof. If $A = UDU^*$ for a diagonal matrix, then

$$AA^* = UDD^*U^* = UD^*DU = A^*A.$$

Conversely, let $A \in M_n$ such that $AA^* = A^*A$.

By the result in Examination 2, there is a unitary such that $U^*AU = T$ is in upper triangular form.

The $(1,1)$ entries of TT^* and T^*T are equal, i.e.,

$$\sum_{j=1}^{n} |t_{1j}|^2 = |t_{11}|^2.$$

We see that the $(1,1)$ entry is the only nonzero entry in the first row of T.

Repeat the argument to the $(2,2), \ldots, (n-1, n-1)$ entries of T^*T and TT^*. We see that T is a diagonal matrix. $\square$

Vandermonde matrix Let $x_1, \dots, x_n$ be distinct real or complex numbers. The Vandermonde matrix is defined by

$$V = V(x_1, \dots, x_n) = [x_i^{j-1}] \in M_n.$$

Theorem *Let $x_1, \dots, x_n \in \mathbb{F}$ be distinct numbers. For any $y_1, \dots, y_n \in \mathbb{F}$, there is a polynomial $f(x)$ of degree $n-1$ such that $f(x_j) = y_j$ for $j = 1, \dots, n$.*

Proof. Let $f(x) = a_0 + a_1 x + \cdots + a_{n-1} x^{n-1}$. Consider the system of equations

$$V \begin{bmatrix} a_0 \\ \vdots \\ a_{n-1} \end{bmatrix} = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}.$$

Note that that $\det(V) = \prod_{1 \le j < k \le n} (x_j - x_k) \neq 0$. To see this, we can show that $(x_j - x_k)$ is a factor of $\det(V)$ for any $j < k$ so that $\det(V)$ is a multiple of $\gamma \prod_{1 \le j < k \le n} (x_j - x_k)$.

One can then compare the coefficient of $x_1 x_2^2 \cdots x_n^{n-1}$ for $\det(V)$ ad $\gamma \prod_{1 \le j < k \le n} (x_j - x_k)$ and conclude $\gamma = -1$.

Consequently, we can always find $a_0, \dots, a_{n-1}$. □

Theorem If $x_1, \dots, x_n$ are distinct real numbers, then the singular values of $V(x_1, \dots, x_n)$ decay exponentially.

Proof. Let

$$A = \operatorname{diag}(x_1, \dots, x_n) \text{ and } B = E_{21} + \cdots + E_{n,n-1} - E_{1n}.$$

Then $AX - XB = \sum_{j=1}^{n} (x_j^n + 1) E_{jn}$ is rank one. □

Homework Treat other matrices including Hilbert matrices, Toeplitz matrices, Hankel matrices, Pick matrices.

Discrete Fourier transform

Data analysis / machine learning are all about

$x \to f(x)$ and $y \to f^{-1}(y)$ for some function f.

Proposition Let $w_n = e^{i2\pi/n}$ and $F = (F_{jk}) = [w^{(j-1)(k-1)}]$ be the Fourier matrix. Then $\frac{1}{\sqrt{n}}F$ is unitary.

Proof. The (j,k) entry of FF^* is

$(1, w^i, \ldots, w^{(n-1)i})(1, w^{-j}, \ldots, w^{-(n-1)j})^T$

It equals $\sum_{k=0}^{n-1}(w^(i-j))^k = [1-(w^n)^{i-j}]/[1-w^{(i-j)}] = 0$ if $i \neq j$, and equals n if $i = j$. □

So, the columns $f_1, \ldots, f_n$ of $\frac{1}{\sqrt{n}}F$ form an orthonormal basis. For any vector $v \in \mathbb{C}^n$,

$$v = c_1 f_1 + \cdots + c_n f_n,$$

where $c_j = f_j^* v$ for all j so that $F^* v = [c_1, \ldots, c_n]^T$.

Suppose $f : \mathbb{R} \to \mathbb{C}$ is periodic such that $f(x) = f(x + 2\pi)$.

If $v^T = [f(0), f(2\pi/n), \ldots, f(2(n-1)\pi/n)]$ is obtained,

then one uses the approximation

$$f(x) = c_0 + c_1 e^{ix} + \cdots + c_{n-1} e^{i(n-1)x},$$

where $F^* v = [c_0, \ldots, c_{n-1}]^T$.

More generally, one can use power series

$$f(x) = c_0 + c_1 e^{ix} + c_{-1} e^{-ix} + c_2 e^{i2x} + c_{-2} e^{-i2x} + \cdots .$$

One may even use the Fourier integral

$$f(x) = \int_{-\infty}^{\infty} \hat{f}(x) e^{ikx} dx.$$

In the real space, if $f : \mathbb{R} \to \mathbb{R}$ satisfies $f(x) = f(x + 2\pi)$,

then

$$f(x) = a_0 + a_1 \cos x + b_1 \sin x + a_2 \cos 2x + b_2 \sin 2x + \cdots .$$

For even functions $f(x) = f(-x)$, $b_j = 0$ for all j;
for odd functions $f(-x) = -f(x)$, $a_j = 0$ for all j.

Remark The singular values of F do not decay exponentially.

In applications people use nonuniform discrete Fourier transform $U = A. * F \in M_N$, where

$U = (A_{jk}F_{jk})$ with $F_{jk} = e^{-i2\pi kj/N}$, $U_{jk} = e^{-i2\pi k x_j}$,

and $A = (A_{jk}) = (F_{jk}/C_{jk})$ is a near low rank matrix.

IV.2 Circulant matrices

Theorem Let $P = E_{12} + E_{23} + \cdots + E_{n-1,n} + E_{n1}$.

- If $U = \frac{1}{\sqrt{n}}F$, then $U^*P^kU = \text{diag}\,(1, w^k, \ldots, w^{(n-1)k})$.
- If $A = a_0 I + a_1 P + \cdots + a_{n-1}P^{n-1}$, then

  $U^*AU = \text{diag}\,(\lambda_1, \ldots, \lambda_n)$ with $\lambda_k = \sum_{j=0}^{n-1} a_j w^{kj}$.

Proof. Final examination question. □

A circulant matrix is a matrix of the form $A = \sum_{j=0}^{n-1} a_j P^j$.

The product of two circulant matrices is a circulant matrix.

The set of matrices form a commutative algebra, which is isomorphic to the algebra of diagonal matrices.

Product of functions and Toeplitz matrices

Suppose $f(x) = \sum_{k=-\infty}^{\infty} f_k e^{ikx}$ and $g(x) = \sum_{m=-\infty}^{\infty} g_m e^{imx}$. Then

$$h(x) = f(x)g(x) = \sum_{k=-\infty}^{\infty} h_k e^{ikx},$$

where $h_m = \sum_{r+s=m} a_r b_s$.

Thus, $C(h) = C(f)C(g)$ if $C(f)$ is the circulant matrix with first row $[\cdots f_{-2}\ f_{-1}\ f_0\ f_1\ f_2 \cdots]$.

- Data can be stored as $f(x)$ with $x \in \mathbb{R}$ so that we can approximate data by nice functions $\tilde{f}(x)$.
- We may consider functions with finite support, periodic, etc.
- The function $f(x) = \sum f_k e^{ikx}$ is real valued if and only if $f_k = f_{-k} = a_k \in \mathbb{R}$ so that

$$f_k e^{ikx} + f_{-k} e^{-ikx} = 2a_k \cos(kx).$$

- One may consider the finite rank approximation of the infinite matrix $C[f]$ to get a Toeplitz matrix

$$A = a_0 I_n + \sum_{j=1}^{n-1} (a_j N^j + a_{-j} N^{-j}),$$

  where $N = E_{21} + \cdots + E_{n,n-1}$.
- Toeplitz matrices also arise in the study of finite difference equations, etc., and in many problems, the Toeplitz matrix is banded, i.e., $f_k = 0$ if $|k| > m$ for some m.
- Example: $A = 2I_n - N - N^T$ is tridiagonal.
- Product of Toeplitz matrices are not Toeplitz matrices; inverses of Toeplitz matrices are not Toeplitz matrices.
- The Toeplitz matrix $A = a_0 I_n + \sum_{j=1}^{n-1} (a_j N^j + a_{-j} N^{-j})$ can be embedded as the left top corner of a circulant matrix $C \in M_{2n-1}$ with first row

$$[a_0\ a_{-1} \cdots a_{1-n}\ a_{n-1} \cdots\ a_1].$$

  Then Ax with appear in $C \begin{bmatrix} x \\ 0_{n-1} \end{bmatrix}$.
- Many techniques have been developed to study Toeplitz matrices.

III.4 Algorithms for compressed sensing

Sparse solution for $Ax = b$

Minimize $\|x\|_1 = |x_1| + \cdots + |x_n|$ subject to $Ax = b$.

Lasso (in statistics)

Minimize $\frac{1}{2}\|Ax - b\|_2^2 + \lambda\|x\|_1$

or Minimize $\frac{1}{2}\|Ax - b\|_2^2$ with $\|x\|_1 \leq t$.

A set $S \subseteq \mathbb{R}^n$ is convex if for any $x, y \in S$,

$$tx + (1-t)y \in S \qquad \text{for all } t \in [0,1];$$

a function $f : \mathbb{R}^n \to \mathbb{R}$ is convex if

$$f(tx + (1-t)y) \leq tf(x) + (1-t)f(y) \qquad \text{for all } t \in [0,1].$$

More generally, for a convex set $K \subseteq \mathbb{R}^n$ and two convex functions $F_1, F_2 : K \to \mathbb{R}$, one considers

$$\text{Minimize } F_1(x) + F_2(x) \quad \text{for } x \in K.$$

There are many techniques:

- Dual decomposition.
- Augmented Lagrangian.
- Methods of multipliers.
- ADMM: Altenating direction methods of multipliers.

Basic theory

We want to find a basis $\{v_1, \ldots, v_n\}$, say, the Fourier basis, to represent a signal/a set of data b.

In statistics, sparse solution represents a reconstruction of the signal/data using sparse measurements. It also helps to avoid noise.(Sparse signals are not noisy!)

Want to find basis $\{w_1, \ldots, w_n\}$ so that $V^T W$ has small entries, say, V is Fourier, $W = I_n$. Very often, $V^T W$ is has small entries for a random matrix with columns having norm 1.

Thus, we want to solve $y = W^T f$ to reconstruct $f^* = V x^*$. The problem reduces to

$$\text{Minimize } \ \|x\|_1 \quad \text{subject to } W^T V x = y.$$

With the noise assumption, we solve the restricted isometry property with $\delta < \sqrt{1}$,

$$(1-\delta)\|x\|_2^2 \leq \|Ax\|_2^2 \leq (1+\delta)\|x\|_2^2 \ \text{ if } \ x \text{ is } S-\text{sparse},$$

x-sparse means that x has fewer than S nonzero entries.

In statistics, one solve the noisy LASSO problem

$$\text{Minimize}\|x\|_1 \quad \text{subject to } \|Ax - b\|_2 \leq \varepsilon.$$

In applications, we deal with problems with missing data, or insignificant data, or controllable data.

They reduce to completion problems of partial matrices with prescribed properties.

For example, find rank one completion (and small norm) of the following.

$$A_0 = \begin{bmatrix} 1 & 3 \\ ? & ? \end{bmatrix}, \quad B_0 = \begin{bmatrix} 1 & ? \\ ? & 5 \end{bmatrix}, \quad C_0 = \begin{bmatrix} 1 & 3 \\ 5 & ? \end{bmatrix}.$$

Find completion with minimum nuclear norm

$$\|A\|_N = \sigma_1 + \cdots + \sigma_r.$$

For A_0, $\sigma_1(A) \ge \sqrt{1+3} = \sqrt{10}$. So, ...

For B_0, $\|\mathrm{diag}\,(1,5)\|_N \le (\|B\|_N + \|DBD\|_N)/2$
with $D = \mathrm{diag}\,(1,-1)$. So, ...

For C_0, we change C_0 to $\tilde{C}_0 = \begin{bmatrix} 3 & 1 \\ ? & 5 \end{bmatrix}$. We see that the completion

$$8 = \| \begin{bmatrix} 3 & 1 \\ 1 & 5 \end{bmatrix} \|_N \le \|\mathrm{diag}\,(3,8)\|_N \le \|C\|_N$$

for any completion of $\tilde{C}_0$. **Theorem** *Let A be a partial matrix. Then the completion A_0 with minimum nuclear norm satisfies*

$$\|A_0\|_N = \min\left\{ \mathrm{tr}\,(P_1 + P_2)/2 : \begin{bmatrix} P_1 & A_0 \\ A_0^* & P_2 \end{bmatrix} \text{ is psd} \right\}.$$

Proof. Extra credit final examination question. □

One can then use positive semidefinite programming techniques to solve the problem.