ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Chair for Mathematical Information Science

Prof. Dr. H. Bölcskei
Sternwartstrasse 7
CH-8092 Zürich

Mathematics of Information

Spring semester 2022

Solutions to Problem Set 5

Problem 1 "ℓ_0-norm".

For a vector $\mathbf{x} \in \mathbb{C}^N$, we have that

$$\|2\mathbf{x}\|_0 = \|\mathbf{x}\|_0,$$

which shows that the homogeneity property of a norm is not satisfied. Therefore, $\|\cdot\|_0$ is a not a norm for $\mathbb{C}^N$. Its name and notation come from an abuse of terminology, taking $p = 0$ in the definition of the ℓ_p-norm $\|\cdot\|_p$ which is commonly defined for $p \in [1, \infty]$ as

$$\|\mathbf{x}\|_p^p = \sum_{k=1}^{N} |x_k|^p$$

for $\mathbf{x} \in \mathbb{C}^N$. Indeed, if one takes $p = 0$ in the previous equation, given that for every $k = 1, \ldots, N$, one has $|x_k|^0 = 1$ if $x_k \neq 0$ and $|x_k|^0 = 0$ if $x_k = 0$, it would amount to counting the number of nonzero elements x_k, $k = 1, \ldots, N$, in the vector $\mathbf{x}$.

Problem 2 Gram matrix.

Let $\mathbf{A} \in \mathbb{C}^{M \times N}$ be the matrix whose k-th column is $\mathbf{a}_k$. Since the entry in the k-th row and in the ℓ-th column is $\langle \mathbf{a}_\ell, \mathbf{a}_k \rangle = \mathbf{a}_k^H \mathbf{a}_\ell$, the Gram matrix can be written as $\mathbf{G} = \mathbf{A}^H \mathbf{A}$. Therefore, it holds that

$$\mathbf{G}^H = (\mathbf{A}^H \mathbf{A})^H = \mathbf{A}^H (\mathbf{A}^H)^H = \mathbf{A}^H \mathbf{A} = \mathbf{G},$$

which shows that $\mathbf{G}$ is hermitian. Moreover, for all $\mathbf{x} \in \mathbb{C}^M$, we have

$$\mathbf{x}^H \mathbf{G} \mathbf{x} = \mathbf{x}^H \mathbf{A}^H \mathbf{A} \mathbf{x} = (\mathbf{A}\mathbf{x})^H \mathbf{A}\mathbf{x} = \|\mathbf{A}\mathbf{x}\|_2^2 \geq 0.$$

Hence, $\mathbf{G}$ is positive-semidefinite. Since $\|\cdot\|_2$ is a norm, $\mathbf{x}^H \mathbf{G} \mathbf{x} = 0$ implies that $\mathbf{A}\mathbf{x} = 0$. So, if $\mathbf{A}$ has linearly independent columns, $\mathbf{A}\mathbf{x} = 0$ imposes that $\mathbf{x} = 0$, and $\mathbf{G}$ is positive-definite.

Problem 3 Summer Exam 2021 ☕☕☕

a) i) We solve the linear system $D\widehat{x} = 0$ to obtain

$$\mathcal{N}(D) = \operatorname{span}\left(\begin{pmatrix} \frac{7}{5} \\ -\frac{3}{5} \\ -\frac{3}{5} \\ 1 \end{pmatrix} \right).$$

ii) In (i) we saw that $\mathcal{N}(D) \neq \{0\}$. Therefore, the columns of D are linearly dependent which implies $\text{spark}(D) \leq 4$. As $\|x\|_0 = 3$ the condition $\|x\|_0 < \frac{\text{spark}(D)}{2}$ hence does not hold.

iii) From the proof of Theorem 3.2 in the lecture notes we know that $\text{spark}(D) \geq 1 + 1/\mu(D)$. The condition

$$\|x\|_0 < \frac{1}{2}\left(1 + \frac{1}{\mu(D)}\right)$$

together with $\text{spark}(D) \leq 4$ would hence require $\|x\|_0 < 2$. This is not satisfied as we have $\|x\|_0 = 3$.

iv) We have $\mathcal{X} = x + \mathcal{N}(D)$, where $x = (1\ 1\ 1\ 0)^T$ is the particular solution from the problem statement. Hence,

$$\mathcal{X} = \left\{ \begin{pmatrix} 1 \\ 1 \\ 1 \\ 0 \end{pmatrix} + \lambda \begin{pmatrix} \frac{7}{5} \\ -\frac{3}{5} \\ -\frac{3}{5} \\ 1 \end{pmatrix} : \lambda \in \mathbb{R} \right\}. \tag{1}$$

v) (P0) identifies the vector

$$\arg\min_{\widehat{x} \in \mathcal{X}} \|\widehat{x}\|_0,$$

where $\mathcal{X}$ denotes the solution set characterized in (1). We notice, with $\lambda = \frac{5}{3}$ in (1), that the vector

$$x' := \begin{pmatrix} 1 + \frac{7}{3} \\ 0 \\ 0 \\ \frac{5}{3} \end{pmatrix}$$

is contained in the solution set, i.e., $x' \in \mathcal{X}$. Since $\|x'\|_0 = 2 < \|x\|_0 = 3$, it follows that the solution to (P0) is not equal to x. Hence, x is not recovered through (P0).

vi) (P1) identifies the vector that minimizes

$$\begin{aligned}
\min_{\widehat{x} \in \mathcal{X}} \|\widehat{x}\|_1 &= \min_{\lambda \in \mathbb{R}} \left\| \begin{pmatrix} 1 \\ 1 \\ 1 \\ 0 \end{pmatrix} + \lambda \begin{pmatrix} \frac{7}{5} \\ -\frac{3}{5} \\ -\frac{3}{5} \\ 1 \end{pmatrix} \right\|_1 && (2) \\
&= \min_{\lambda \in \mathbb{R}} \left|1 + \frac{7}{5}\lambda\right| + 2\left|1 - \frac{3}{5}\lambda\right| + |\lambda| && (3) \\
&=: \min_{\lambda \in \mathbb{R}} f(\lambda). && (4)
\end{aligned}$$

Noting that $f(\lambda)$ has its unique minimum at $\lambda = 0$, it follows that (P1) recovers x uniquely.

In order to formally establish that $f(\lambda)$ is minimized at $\lambda = 0$, we observe that

$f(0) = 3$ and for every $\delta \in (0, \frac{5}{7})$, we have

$$f(0+\delta) = 1 + \frac{7}{5}\,\delta + 2 - \frac{6}{5}\,\delta + \delta = 3 + \frac{6}{5}\,\delta > 3$$

and

$$f(0-\delta) = 1 - \frac{7}{5}\,\delta + 2 + \frac{6}{5}\,\delta + \delta = 3 + \frac{4}{5}\,\delta > 3.$$

Hence, $\lambda = 0$ is the unique minimum in the interval $\left(-\frac{5}{7}, \frac{5}{7}\right)$. Since f is the sum of convex functions it is also the unique global minimum.

b) i) We first establish the inequality provided in the Hint.

$$|\langle u, v\rangle| = \left|\sum_{k=1}^{N} u_k \overline{v_k}\right| \leq \sum_{k=1}^{N} |u_k||v_k| \leq \max_{\ell} |v_\ell| \sum_{k=1}^{N} |u_k| = \|v\|_\infty \|u\|_1.$$

Next, we show that under (C1) for all $z \neq x$ with $Dz = y = Dx$, we have $\|z\|_1 > \|x\|_1$ as this implies that the minimization problem (P1) has the unique solution x as desired. To this end, we define $v := z - x$ and note that $Dv = D(z-x) = Dz - Dx = 0$, i.e., $v \in \mathcal{N}(D)$. We now bound

$$\begin{aligned}
\|z\|_1 = \|v + x\|_1 &= \|v_S + x_S\|_1 + \|v_{S^c}\|_1 && (5)\\
&= \|v_S + x_S\|_1 \|\mathrm{sgn}(x_S)\|_\infty + \|v_{S^c}\|_1 && (6)\\
&\geq |\langle x_S + v_S, \mathrm{sgn}(x_S)\rangle| + \|v_{S^c}\|_1 && (7)\\
&> |\langle x_S + v_S, \mathrm{sgn}(x_S)\rangle| + |\langle v_S, \mathrm{sgn}(x_S)\rangle| && (8)\\
&\geq |\langle x_S, \mathrm{sgn}(x_S)\rangle| - |\langle v_S, \mathrm{sgn}(x_S)\rangle| + |\langle v_S, \mathrm{sgn}(x_S)\rangle| && (9)\\
&= \left|\sum_{k\in S} x_k \frac{\overline{x_k}}{|x_k|}\right| = \|x_S\|_1 = \|x\|_1, && (10)
\end{aligned}$$

where in (6) we used $\|\mathrm{sgn}(x_S)\|_\infty = 1$, for $x_S \neq 0$, in (7) we applied the Hint, in (8) we used (C1), and in (9) we employed the reverse triangle inequality.

ii) First note that for all $v \in \mathcal{N}(D)\backslash\{0\}$, we have

$$0 = Dv = D_S v_S + D_{S^c} v_{S^c} \tag{11}$$

and hence $D_S v_S = -D_{S^c} v_{S^c}$. Further, we realize that $(D^H h)_S = (D_S)^H h$, $\forall h \in \mathbb{C}^m$. Next, we assume that (C2) holds and show that this implies (C1). Let $h \in \mathbb{C}^m$ be such that

$$(D^H h)_S = \mathrm{sgn}(x_S) \;\text{ and }\; \|(D^H h)_{S^c}\|_\infty < 1. \tag{12}$$

Such an $h \in \mathbb{C}^m$ exists by assumption (C2). With $(D^H h)_S = (D_S)^H h$ this implies (C1) as follows,

$$\left|\sum_{j\in S} v_j \overline{(\operatorname{sgn}(x))_j}\right| = |\langle v_S, \operatorname{sgn}(x_S)\rangle| = \left|\langle v_S, (D^H h)_S\rangle\right| \tag{13}$$
$$= |\langle v_S, (D_S)^H h\rangle| = |\langle D_S v_S, h\rangle| \tag{14}$$
$$= |\langle -D_{S^c} v_{S^c}, h\rangle| = |\langle v_{S^c}, (D_{S^c})^H h\rangle| \tag{15}$$
$$= |\langle v_{S^c}, (D^H h)_{S^c}\rangle| \leq \|v_{S^c}\|_1 \|(D^H h)_{S^c}\|_\infty \tag{16}$$
$$< \|v_{S^c}\|_1. \tag{17}$$

To see that the final inequality is, indeed, strict, we note that $v_{S^c} \neq 0$ as otherwise $v \neq 0$ would imply $v_S \neq 0$ and (11) would imply $D_S v_S = 0$. This would, however, stand in contradiction to $\mathcal{N}(D_S) = \{0\}$. Therefore, we have $\|v_{S^c}\|_1 \neq 0$. Together with $\|(D^H h)_{S^c}\|_\infty < 1$, which is by (12), this guarantees strict inequality.

iii) We assume that (C3) holds and show that this implies (C2). The first part of (C2), i.e. $\mathcal{N}(D_S) = \{0\}$, holds directly. It remains to establish that under (C3) there exists a vector h fulfilling the second part of (C2), i.e.

$$(D^H h)_S = sgn(x_S) \text{ and } \|(D^H h)_S\|_\infty < 1. \tag{18}$$

We set $h := (D_S^\dagger)^H \operatorname{sgn}(x_S)$, where we used that $(D_S)^H D_S$ is invertible since D_S has full column rank as a consequence of $\mathcal{N}(D_S) = \{0\}$. Now, we have

$$(D^H h)_S = (D_S)^H h = (D_S)^H (D_S^\dagger)^H \operatorname{sgn}(x_S) \tag{19}$$
$$= (D_S)^H D_S \,((D_S)^H D_S)^{-1} \operatorname{sgn}(x_S) \tag{20}$$
$$= \operatorname{sgn}(x_S), \tag{21}$$

where we used that the inverse of a symmetric matrix is symmetric. Hence, $(D^H h)_j = \operatorname{sgn}(x)_j,\ \forall j \in S$. Next we observe that for all $k \in S^c$,

$$|(D^H h)_k| = |\langle d_k, h\rangle| = |\langle d_k, (D_S^\dagger)^H \operatorname{sgn}(x_S)\rangle| = |\langle D_S^\dagger d_k, \operatorname{sgn}(x_S)\rangle| < 1$$

where the inequality is by condition (C3). We have thus established that h satisfies (18) as desired.

Problem 4 Exam 2019

a) For $\ell, n, \ell', n' \in \{1, \ldots, m\}$, we have

$$\langle E^{(\ell,n)}, E^{(\ell',n')}\rangle = \sum_{j,k=1}^{m} E_{jk}^{(\ell,n)} \overline{E_{jk}^{(\ell',n')}} = \begin{cases} 1, & \text{if } \ell = \ell' \text{ and } n = n' \\ 0 & \text{else} \end{cases},$$

which proves that $\mathcal{E}$ is an orthonormal system. To see that this system is complete, and hence an orthonormal basis, note that every $A \in \mathbb{C}^{m\times m}$ can be expanded as

$$A = \sum_{j,k=1}^{m} A_{j,k}\, E^{(j,k)}.$$

As $\mathbb{C}^{m\times m}$ has a basis of size $m \cdot m = m^2$, namely $\mathcal{E}$, the dimension of $\mathbb{C}^{m\times m}$ is m^2.

b) Note that D acts on vectors $\boldsymbol{v} \in \mathbb{C}^m$ as the forward cyclic rotation according to $D \cdot (v_1, v_2, \ldots, v_m)^T = (v_m, v_1, \ldots, v_{m-1})^T$, and so, for $n \in \mathbb{Z}$, D^n acts on vectors as the forward cyclic rotation by n places if $n > 0$, and as the backward cyclic rotation if $n < 0$. For $\ell, n, \ell', n' \in \{0, \ldots, m-1\}$, we have

$$\begin{aligned}
\langle G^{(\ell,n)}, G^{(\ell',n')} \rangle &= \operatorname{tr}\left((G^{(\ell',n')})^H \, G^{(\ell,n)} \right) \\
&= \frac{1}{m} \operatorname{tr}\left(D^{-n'} M^{-\ell'} M^{\ell} D^{n} \right) \\
&= \frac{1}{m} \operatorname{tr}\left(D^{n-n'} M^{\ell-\ell'} \right) \\
&= \frac{1}{m} \delta_{n,n'} \operatorname{tr}\left(M^{\ell-\ell'} \right) \\
&= \delta_{n,n'} \frac{1}{m} \sum_{k=0}^{m-1} e^{-\frac{2\pi i k(\ell-\ell')}{m}} \\
&= \delta_{n,n'} \delta_{\ell,\ell'},
\end{aligned}$$

where $\delta_{a,b}$ denotes the Kronecker delta, and we used that, for every $A \in \mathbb{C}^{m\times m}$, the matrix $D^n A$ is obtained by cycling each column of A by n places. Concretely, if $n - n' \neq 0$, the diagonal of $D^{n-n'} M^{\ell-\ell'}$ is identically zero. This establishes that $\mathcal{G}$ is an orthonormal system in $\mathbb{C}^{m\times m}$. Noting that $\#\mathcal{G} = m^2 = \dim(\mathbb{C}^{m\times m})$, i.e., $\mathcal{G}$ is a linearly independent subset of $\mathbb{C}^{m\times m}$ of the same size as the dimension of $\mathbb{C}^{m\times m}$, it follows that $\mathcal{G}$ is an orthonormal basis for $\mathbb{C}^{m\times m}$.

c) Let $\ell, n \in \{1, \ldots, m\}$ and $\ell', n' \in \{0, \ldots, m-1\}$ be arbitrary. Then $|\langle E^{(\ell,n)}, G^{(\ell',n')} \rangle| = |G^{(\ell',n')}_{\ell,n}|$. As all the nonzero entries of $G^{(\ell',n')}$ have modulus $\frac{1}{\sqrt{m}}$, we find that

$$\mu(\mathcal{E}, \mathcal{G}) = \max_{\substack{(\ell,n)\in\{1,\ldots,m\}^2 \\ (\ell',n')\in\{0,\ldots,m-1\}^2}} |\langle E^{(\ell,n)}, G^{(\ell',n')} \rangle| = \frac{1}{\sqrt{m}}.$$

d) Let $\mathcal{B}_1 = \{U_1, \ldots, U_{m^2}\}$ and $\mathcal{B}_2 = \{V_1, \ldots, V_{m^2}\}$ be arbitrary orthonormal bases for $\mathbb{C}^{m\times m}$. By way of contradiction, suppose that $\mu(\mathcal{B}_1, \mathcal{B}_2) < \frac{1}{m}$. Now, as $\mathcal{B}_2$ is an orthonormal basis for $\mathbb{C}^{m\times m}$, U_1 has the following expansion:

$$U_1 = \sum_{n=1}^{m^2} \langle U_1, V_n \rangle V_n,$$

and, moreover, we have the energy conservation relation

$$\|U_1\|^2 := \langle U_1, U_1 \rangle = \sum_{n=1}^{m^2} |\langle U_1, V_n \rangle|^2.$$

Now, using the assumption $\mu(\mathcal{B}_1, \mathcal{B}_2) < \frac{1}{m}$, we can conclude that

$$\|U_1\|^2 = \sum_{n=1}^{m^2} |\langle U_1, V_n \rangle|^2 \leq \sum_{n=1}^{m^2} (\mu(\mathcal{B}_1, \mathcal{B}_2))^2 < \sum_{n=1}^{m^2} \frac{1}{m^2} = m^2 \cdot \frac{1}{m^2} = 1.$$

But this contradicts the fact that U_1, as an element of an orthonormal basis, has unit norm. Therefore, our assumption must be wrong, and hence we deduce that $\mu(\mathcal{B}_1, \mathcal{B}_2) \geq \frac{1}{m}$.

Problem 5 Nonuniform sampling

a) Using the Cauchy-Schwarz inequality, we have

$$|\langle Y_n, x\rangle|^2 \leq \left(\sum_{k\in\mathrm{supp}(x)} \left|p_n^{-\frac{1}{2}}\langle g_k, h_n\rangle\right|^2\right)\|x\|_2^2 \leq \left(\sum_{k\in\mathrm{supp}(x)} p_n^{-1}\tilde{\mu}_n^2\right)\|x\|_2^2 \leq s p_n^{-1}\tilde{\mu}_n^2\,\|x\|_2^2\,,$$

for $n \in [N]$, as desired.

b) For $k, \ell \in [N]$, we have

$$\begin{aligned}\left(\mathbb{E}_{\boldsymbol{t}\sim\mathcal{P}}\left[Y_{\boldsymbol{t}}^{\mathsf{H}}\,Y_{\boldsymbol{t}}\right]\right)_{k\ell} &= \mathbb{E}_{\boldsymbol{t}\sim\mathcal{P}}\left[(Y_{\boldsymbol{t}})_\ell\overline{(Y_{\boldsymbol{t}})_k}\right] = \mathbb{E}_{\boldsymbol{t}\sim\mathcal{P}}\left[p_{\boldsymbol{t}}^{-\frac{1}{2}}\langle g_\ell, h_{\boldsymbol{t}}\rangle\,\overline{p_{\boldsymbol{t}}^{-\frac{1}{2}}\langle g_k, h_{\boldsymbol{t}}\rangle}\right]\\ &= \sum_{n=1}^{N} p_n\cdot p_n^{-\frac{1}{2}}\langle g_\ell, h_n\rangle\,\overline{p_n^{-\frac{1}{2}}\langle g_k, h_n\rangle} = \sum_{n=1}^{N}\langle g_\ell, h_n\rangle\,\overline{\langle g_k, h_n\rangle} = \langle g_\ell, g_k\rangle = \delta_{k\ell},\end{aligned}$$

where we used the fact that $\{g_k\}_{k=1}^N$ and $\{h_j\}_{j=1}^N$ are orthonormal bases. Therefore $\mathbb{E}_{\boldsymbol{t}\sim\mathcal{P}}\left[Y_{\boldsymbol{t}}^{\mathsf{H}}\,Y_{\boldsymbol{t}}\right] = \mathrm{Id}$.

c) We have

$$\begin{aligned}\mathbb{E}_{\boldsymbol{t}\sim\mathcal{P}}\left[|\langle Y_{\boldsymbol{t}}, x\rangle|^2\right] &= \mathbb{E}_{\boldsymbol{t}\sim\mathcal{P}}\left[\langle Y_{\boldsymbol{t}}, x\rangle\overline{\langle Y_{\boldsymbol{t}}, x\rangle}\right] = \mathbb{E}_{\boldsymbol{t}\sim\mathcal{P}}\left[\sum_{k,\ell=1}^{N}(Y_{\boldsymbol{t}})_k\overline{x_k}\,\overline{(Y_{\boldsymbol{t}})_\ell\overline{x_\ell}}\right]\\ &= \sum_{k,\ell=1}^{N}\overline{x_k}x_\ell\;\underbrace{\mathbb{E}_{\boldsymbol{t}\sim\mathcal{P}}\left[(Y_{\boldsymbol{t}})_k\overline{(Y_{\boldsymbol{t}})_\ell}\right]}_{=\delta_{k\ell}} = \sum_{k=1}^{N}\overline{x_k}x_k = \|x\|_2^2.\end{aligned}$$