Kaczmarz algorithm in Hilbert space

Ryszard Szwarc
joint work with Rainis Haller

This work was inspired by the paper by S. Kwapień and J. Mycielski (Studia Mathematica **148** (2001), 75–86).

Let $\{e_n\}_{n=0}^{\infty} \subset \mathcal{H}$, $\|e_n\| = 1$. For given x we get numbers $\{(x, e_n)\}_{n=0}^{\infty}$. We want to reconstruct x in terms of these numbers. The sequence $\{e_n\}_{n=0}^{\infty}$ should be linearly dense. Define

$$\begin{aligned} x_0 &= (x, e_0)e_0, \\ x_n &= x_{n-1} + (x - x_{n-1}, e_n)e_n. \end{aligned}$$

The formula is called the Kaczmarz algorithm. We are interested when $x_n \to x$ for any $x \in \mathcal{H}$. Sequences $\{e_n\}_{n=0}^{\infty}$ for which this holds will be called effective.

When $\{e_n\}_{n=0}^{\infty}$ is orthonormal the algorithm returns the partial sums

$$x_n = \sum_{i=0}^{n} (x, e_i) e_i.$$

Kaczmarz (1937) proved that if $\dim \mathcal{H} < +\infty$ and $\{e_n\}_{n=0}^{\infty}$ is periodic and spans $\mathcal{H}$, then $\{e_n\}_{n=0}^{\infty}$ is effective.

Let P_n be the orthogonal projection onto $e_n^{\perp}$. Then

$$\begin{aligned} x_n &= x_{n-1} + (I - P_n)(x - x_{n-1}) \\ x - x_n &= P_n(x - x_{n-1}) \\ x - x_n &= P_n P_{n-1} \ldots P_1 P_0 x. \end{aligned}$$

The sequence is effective iff $P_n P_{n-1} \ldots P_1 P_0$ tends to zero strongly.

Let $\dim \mathcal{H} < \infty$ and $\{e_n\}_{n=0}^{\infty}$ be N-periodic. For $A = P_{N-1} \dots P_1 P_0$ it suffices to show that A^n tends to zero. We claim that $\|A\| < 1$. If not there is a vector x such that $\|Ax\| = \|x\| = 1$. Then $\|P_0 x\| \geq \|Ax\| = \|x\|$, hence $P_0 x = x$. Similarly $P_1 x = x$, $\dots$, $P_{N-1} x = x$, which implies that $x \perp e_0, e_1, \dots, e_{N-1}$. Thus $x = 0$.

Assume (temporarily) linear independence of vectors $\{e_n\}_{n=0}^{\infty}$. We have $x_n \in \operatorname{span}\{e_0, \dots, e_n\}$.

$$x_n = \sum_{i=0}^{n} (x, g_i) e_i, \tag{1}$$

where $g_i \in \mathcal{H}$. It can be verified by induction that

$$g_n = e_n - \sum_{i=0}^{n-1} (e_n, e_i) g_i. \tag{2}$$

We can define g_n by (2) and then check that (1) holds. In this way we remove the linear independence assumption. The formula (2) can be written as

$$e_n = \sum_{i=0}^{n} m_{ni} g_i, \quad m_{ni} = (e_n, e_i).$$

By (2) there are numbers c_{nj} such that

$$g_n = \sum_{j=0}^{n} c_{nj} e_j, \quad c_{nn} = 1. \tag{3}$$

The coefficients $\{c_{nj}\}_{n>j}$ are crucial. We obtain these numbers by taking the inverse to the matrix $I + M$ where

$$M = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & \dots \\ m_{10} & 0 & 0 & 0 & 0 & \dots \\ m_{20} & m_{21} & 0 & 0 & 0 & \dots \\ m_{30} & m_{31} & m_{32} & 0 & 0 & \dots \\ \vdots & \vdots & \vdots & \vdots & \ddots & \ddots \end{pmatrix}$$

$$U = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & \dots \\ c_{10} & 0 & 0 & 0 & 0 & \dots \\ c_{20} & c_{21} & 0 & 0 & 0 & \dots \\ c_{30} & c_{31} & c_{32} & 0 & 0 & \dots \\ \vdots & \vdots & \vdots & \vdots & \ddots & \ddots \end{pmatrix}$$

$$(I + U)(I + M) = (I + M)(I + U) = I.$$

$$UM = MU = -U - M.$$

Proposition 1. *Let U and M be strictly lower triangular matrices such that $MU = UM = -U - M$. Then*

$$\|U\| \leq 1 \quad \textit{iff} \quad M + M^* + I \geq 0.$$

In that case there is a Hilbert space $\mathcal{H}$ and vectors $\{e_n\}_{n=0}^{\infty}$ such that $M + M^* + I$ is the Gram matrix of these vectors.

Proof. Let

$$M_n = \begin{pmatrix} 0 & & & \\ m_{10} & 0 & & \\ \vdots & \ddots & 0 & \\ m_{n0} & \cdots & m_{n,n-1} & 0 \\ 0 & \cdots & 0 & \ddots \\ \vdots & \vdots & \vdots & \vdots \end{pmatrix} \quad U_n = \begin{pmatrix} 0 & & & \\ c_{10} & 0 & & \\ \vdots & \ddots & 0 & \\ c_{n0} & \cdots & c_{n,n-1} & 0 \\ 0 & \cdots & 0 & \ddots \\ \vdots & \vdots & \vdots & \vdots \end{pmatrix}$$

Then $(I + M_n)(I + U_n) = (I + U_n)(I + M_n) = I$.

Assume $M + M^* + I \geq 0$. Hence $M_n + M_n^* + I \geq 0$ and

$$0 \leq (U_n^* + I)(M_n + M_n^* + I)(U_n + I) = I - U_n^* U_n$$

Hence $\|U_n\| \leq 1$ and $\|U\| \leq 1$.

The converse implication follows from

$$(M_n^* + I)(I - U_n^* U_n)(M_n + I) = M_n + M_n^* + I.$$

$\diamondsuit$

By construction of the algorithm we have

$$x - x_n \perp e_n$$

because $x - x_n = P_n(x - x_{n-1})$. From $x_n = \sum_{i=0}^{n} (x, g_i) e_i$ we obtain

$$x - x_{n-1} = x - x_n + (x, g_n) e_n.$$

$$\begin{aligned} \|x - x_{n-1}\|^2 &= \|x - x_n\|^2 + |(x, g_n)|^2, \quad n \geq 1, \\ \|x\|^2 &= \|x - x_0\|^2 + |(x, g_0)|^2. \end{aligned}$$

Hence

$$\|x\|^2 - \lim_n \|x - x_n\|^2 = \sum_{n=0}^{\infty} |(x, g_n)|^2.$$

Therefore effectiveness is equivalent to

$$\|x\|^2 = \sum_{n=0}^{\infty} |(x, g_n)|^2$$

for any $x \in \mathcal{H}$. Our main result is

Theorem 1. *The sequence $\{e_n\}_{n=0}^{\infty}$ is effective if and only if it is linearly dense and U is a partial isometry, i.e. U^*U is a projection.*

These two results can be interpreted as follows. We have as many effective sequences among sequences of unit vectors as partial isometries among strictly lower triangular contractions.

Kwapień and Mycielski showed that if we choose the sequence of unit vectors at random then almost surely we end up with an effective sequence. More precisely fix a Borel measure μ on the unit sphere of $\mathcal{H}$, such that the support of μ is linearly dense. Then draw consecutive vectors independently with respect to that measure.

Proof of Theorem 1. Assume effectiveness. Hence

$$(x, y) = \sum_{n=0}^{\infty} (x, g_n)(g_n, y).$$

In particular

$$m_{ij} = (e_i, e_j) = \sum_{n=0}^{\infty} (e_i, g_n)(g_n, e_j).$$

Lemma 1.

$$(g_n, e_j) = ((UM^* + M^* + I)\delta_j, \delta_n)_{\ell^2(\mathbb{N})}$$

$\diamondsuit$

Let $A = UM^* + M^* + I$. We have

$$m_{ij} = \sum_{n=0}^{\infty} (A\delta_j, \delta_n)_{\ell^2} (\delta_n, A\delta_i)_{\ell^2} = (A\delta_j, A\delta_i)_{\ell^2}.$$

Assume for simplicity that A is bounded on ℓ^2. Therefore we have

$$M + M^* + I = A^*A.$$

Taking into account relation between U and M gives

$$A^*A = MU^*UM^* - MM^* + M + M^* + I.$$

Hence

$$MM^* = MU^*UM^*$$

or

$$(M^*\delta_j, M^*\delta_i)_{\ell^2} = (UM^*\delta_j, UM^*\delta_i)_{\ell^2}. \tag{4}$$

The last formula makes sense when M is unbounded, because M^* leaves the space $\mathcal{F}(\mathbb{N}) = \operatorname{span}\{\delta_0, \delta_1, \ldots\}$ invariant. This formula can be proved directly by replacing A with $A_n = U_n M_n^* + M_n + I$ and taking the limit. (4) states that U is an isometry on

$$\mathcal{H}_0 = \overline{M^*(\mathcal{F}(\mathbb{N}))}.$$

The proof will be completed if we show that U vanishes on $\mathcal{H}_0^\perp$. We have

$$U^*, M^* : \mathcal{F}(\mathbb{N}) \to \mathcal{F}(\mathbb{N})$$

hence

$$M^*(U^* + I) = -M^* - U^* + M^* = -U^*.$$

Therefore $U^*(\mathcal{F}(\mathbb{N})) \subset M^*(\mathcal{F}(\mathbb{N})) \subset \mathcal{H}_0$ and consequently

$$\mathcal{H}_0^\perp \subset \ker U.$$

Conversely, let U be a partial isometry. Hence U is isometric on $\mathcal{H}_0 = \overline{U^*(\mathcal{F}(\mathbb{N}))}$. The formula

$$U^*(M^* + I) = -M^*$$

yields that

$$U^*(\mathcal{F}(\mathbb{N})) \supset M^*(\mathcal{F}(\mathbb{N})).$$

Hence U is an isometry on $M^*(\mathcal{F}(\mathbb{N}))$. This implies the formula

$$(M^*\delta_j, M^*\delta_i)_{\ell^2} = (UM^*\delta_j, UM^*\delta_i)_{\ell^2}.$$

Now we can track backwards the proof of the first part to obtain the conclusion. $\diamondsuit$

Particular case

Assume $\overline{M^*(\mathcal{F}(\mathbb{N}))} = \ell^2(\mathbb{N})$. Recall that $\overline{M^*(\mathcal{F}(\mathbb{N}))}$ is a carrier space of U. This case is equivalent to $U^*U = I$ and occurs when the rows of the matrix M span a dense subspace of $\ell^2(\mathbb{N})$. For example this is the case when $m_{n,n-1} \neq 0$ for any n.

$$\begin{pmatrix} 0 & & & & \\ m_{10} & 0 & & & \\ * & m_{21} & 0 & & \\ * & * & m_{32} & 0 & \\ \vdots & \vdots & & \ddots & \ddots \end{pmatrix}$$

We say that the sequence $\{e_n\}_{n=0}^{\infty}$ is strongly effective if $\{e_n\}_{n=k}^{\infty}$ is effective for each k. This is highly nonorthogonal case.

Theorem 2. *Assume* $\{e_n\}_{n=0}^{\infty}$ *is linearly dense. Then* $\{e_n\}_{n=0}^{\infty}$ *is strongly effective if and only if* $U^*U = I$.

Proof. Let $U^*U = I$ and let $M^{(k)}$ and $U^{(k)}$ denote truncated matrices when we remove the first k rows and the first k columns.

These matrices correspond to the sequence $\{e_n\}_{n=k}^{\infty}$. Also since U is lower triangular we get

$$(U^{(k)})^* U^{(k)} = I,$$

hence $U^{(k)}$ is a partial isometry. It suffices to show that $\{e_n\}_{n=k}^{\infty}$ is linearly dense. It can be shown that $U^*U = I$ implies

$$e_j = -\sum_{i=j+1}^{\infty} \overline{c_{ij}} e_i$$

which means that removing finitely many vectors doesn't spoil linear density of the system. In view of Theorem 1 the sequence $\{e_n\}_{n=k}^{\infty}$is effective.

Conversely let $\{e_n\}_{n=0}^{\infty}$ be strongly effective. Let Q_k denote the orthogonal projection onto the orthogonal complement of $\{\delta_0, \delta_1 \ldots, \delta_{k-1}\}$. Let $U_{(k)} = UP_k$. Then $U_{(k)} = 0_k \oplus U^{(k)}$. Hence $U_{(k)}$ are partial isometries as well as U. But this is possible only if U^*U and P_k commute (`Exercise`). On the other hand if U^*U commutes with P_k for any k then U^*U must be diagonal. Assume that $U^*U \neq I$. Then $U\delta_j = 0$ for some j. This implies that e_j is orthogonal to all the vectors e_i, $i \neq j$. Hence $\{e_n\}_{n=0}^{\infty}$ cannot be strongly effective. $\diamondsuit$

Theorem 3. *Assume* $\{e_n\}_{n=0}^{\infty}$ *is strongly effective. Then*

$$\sum_{i=0}^{\infty} |(e_i, e_j)|^2 = +\infty$$

for any j.

Lemma 2. $U^*U = I$ *implies* $U^*M = -M - I$.

Proof. (incorrect) Assume M is bounded. Then

$M = (U^*U)M = U^*(UM) = U^*(-U - M) = -I - U^*M.$

$\diamondsuit$

From Lemma 2 we have

$$U^*(M + M^* + I) \\ = -M - I - M^* - U^* + U^* = -(M + M^* + I).$$

Let $G = M + M^* + I$. We have $U^*G = -G$. Assume that

$$\sum_{i=0}^{\infty} |(e_i, e_j)|^2 < +\infty$$

for some j. Then $x := G\delta_j \in \ell^2$. Consequently $U^*x = -x$ and

$$(U^*)^n x = (-1)^n x.$$

But $\|U^*\| \leq 1$ and U^* is strictly upper triangular. Hence $(U^*)^n$ tends to zero strongly which implies $x = 0$. This is a contradiction, because $x_j = (e_j, e_j) = 1$. $\diamondsuit$

Stationary case

Assume

$$(e_{i+1}, e_{j+1}) = (e_i, e_j)$$

Then the matrix M is constant on diagonals.

$$M = \begin{pmatrix} 0 & & & & \\ a_1 & 0 & & & \\ a_2 & a_1 & 0 & & \\ a_3 & a_2 & a_1 & 0 & \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}$$

By the Herglotz theorem there is a measure μ on the unit circle such that

$$(e_{i+n}, e_i) = a_n = \int_{\mathbb{T}} z^n d\mu(z).$$

Kwapień and Mycielski showed that a stationary linearly dense sequence $\{e_n\}_{n=0}^{\infty}$ is effective if and only if either μ is the Lebesgue measure (orthogonal case) or it is singular with respect to the Lebesgue measure. Also U is constant on diagonals, i.e. it is a Toeplitz operator.

$$U = \begin{pmatrix} 0 & & & & \\ u_1 & 0 & & & \\ u_2 & u_1 & 0 & & \\ u_3 & u_2 & u_1 & 0 & \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}$$

U is unitarily equivalent to multiplication operator on $H^2(\mathbb{T})$

with the function

$$u(z) = \sum_{n=1}^{\infty} u_n z^n.$$

Moreover

$$\|U\| = \|u(z)\|_{H^\infty(\mathbb{T})} \leq 1.$$

Now multiplication with $u(z)$ is a partial isometry if and only if the boundary values of $|(u(z)|$ are equal 0 or 1. By F. Riesz and M. Riesz Theorem $u(z) \equiv 0$ or $|u(z)| \equiv 1$. The first case corresponds to orthogonal systems, because $M = 0$. The second case was the key point in the paper by Kwapień and Mycielski in proving that μ is singular.

Open problem

Fix $h \in L^2(\mathbb{T})$ such that $\widehat{h}(n) \neq 0$ for any n. Then the translates of h_t, where $h_t(z) = h(tz)$ span a dense subspace. Fix a sequence of numbers $t_n \in \mathbb{T}$. Kwapień and Mycielski showed that if $t_n = t^n$, where t is not a root of identity, then the sequence h_{t_n} is effective, because this sequence is stationary and its spectral measure is discrete. Also they showed that if t_n are independent random variables with uniform distribution on $\mathbb{T}$ then almost surely the sequence h_{t_n} is effective.

The problem is: for given h determine the sequences of unit numbers t_n leading to effectiveness of the sequence h_{t_n}.

Open problem by Mycielski

Let $\{e_n\}_{n=0}$ be linearly dense. By using the proof of Kaczmarz theorem it can be shown that there exists a function $\sigma : \mathbb{N} \to \mathbb{N}$ such that the sequence $\{e_{\sigma(n)}\}_{n=0}^{\infty}$ is effective. The consecutive values of such function may look as follows

$$0, 1, \ldots, 0, 1, 0, 1, 2, \ldots, 0, 1, 2, 0, 1, 2, 3, \ldots, 0, 1, 2, 3, \ldots$$

The number of repetition of the block $0, 1, 2, \ldots, k$ is set in such a way that the norm of the operator $P_k P_{k-1} \ldots, P_0$ restricted to the space spanned by $e_0, e_1, \ldots, e_k$ is less than $1/k$.

The question is if there exists a function σ which is "good" for any linearly dense sequence $\{e_n\}_{n=0}^{\infty}$, i.e. the sequence $\{e_{\sigma(n)}\}_{n=0}^{\infty}$ is always effective.