NOTES ON STOCHASTIC FEJÉR MONOTONICITY

NICHOLAS PISCHKE[a], THOMAS POWELL[b]

[a] Department of Mathematics, Technische Universität Darmstadt,
Schlossgartenstraße 7, 64289 Darmstadt, Germany,
[b] Department of Computer Science, University of Bath,
Claverton Down, Bath, BA2 7AY, United Kingdom,

E-mails: trjp20@bath.ac.uk, pischke@mathematik.tu-darmstadt.de

Abstract.

Keywords:
MSC2020 Classification:

1. Very slight reformulation of Nicholas' Discord notes

Lemma 1.1. *For $p \in P$ let $(U_{p,n})$ be a nonnegative supermartingale and suppose furthermore that Ψ is a function satisfying*

$$\exists n \leqslant \Psi(\varepsilon)\, \exists p\, (\mathbb{E}[U_{p,n}] < \varepsilon)$$

for all $\varepsilon > 0$. Let (X_n) and $(Y_{n,m})$ be nonnegative stochastic processes.

(a) If π_1 is such that
$$\mathbb{E}[U_{p,n}] < \pi_1(\varepsilon) \to \mathbb{E}[X_n] < \varepsilon$$
then $\mathbb{E}[X_n] \to 0$ with rate $\phi_1(\varepsilon) := \Psi(\pi_1(\varepsilon))$.

(b) If π_2 is such that almost surely
$$U_{p,n} < \pi_2(\varepsilon) \to X_n < \varepsilon$$
then $X_n \to 0$ almost surely with rate $\phi_2(\lambda, \varepsilon) := \Psi(\lambda \cdot \pi_2(\varepsilon))$.

(c) If π_3 is such that
$$\mathbb{E}[U_{p,n}], \mathbb{E}[U_{p,m}] < \pi_3(\varepsilon) \to \mathbb{E}[Y_{n,m}] < \varepsilon$$
then $Y_{m,n} \to 0$ with rate $\phi_3(\varepsilon) := \Psi(\pi_3(\varepsilon))$.

(d) If π_4 is such that almost surely
$$U_{p,n}, U_{p,m} < \pi_4(\varepsilon) \to Y_{n,m} < \varepsilon$$
then $Y_{n,m} \to 0$ almost surely with rate $\phi_4(\lambda, \varepsilon) := \Psi(\lambda \cdot \pi_4(\varepsilon)/2)$.

Proof. For (a), fix $\varepsilon > 0$ and let $N \leqslant \Psi(\pi_1(\varepsilon))$ and $p \in P$ be such that $\mathbb{E}[U_{p,N}] < \pi_1(\varepsilon)$. For $n \geqslant \Psi(\pi_1(\varepsilon))$ we have $\mathbb{E}[U_{p,n}] \leqslant \mathbb{E}[U_{p,N}] < \pi_1(\varepsilon)$ and thus $\mathbb{E}[X_n] < \varepsilon$. Part (c) is proven similarly.

For (d) we use Ville's inequality: Let $N \leqslant \Psi(\lambda \cdot \pi_4(\varepsilon)/2)$ and $p \in P$ be such that $\mathbb{E}[U_{p,n}] < \lambda \cdot \pi_4(\varepsilon)/2$. Then

$$\mathbb{P}\left(\exists n \geqslant \phi_4(\lambda, \varepsilon)(U_{p,n} \geqslant \pi_4(\varepsilon))\right) \leqslant \frac{\mathbb{E}[U_{p,N}]}{\pi_4(\varepsilon)} < \frac{\lambda}{2}$$

and thus

$$\mathbb{P}(\exists m, n \geqslant \phi_4(\lambda, \varepsilon)(Y_{m,n} \geqslant \varepsilon)) \leqslant 2\mathbb{P}\left(\exists n \geqslant \phi_4(\lambda, \varepsilon)(U_{p,n} \geqslant \pi_4(\varepsilon))\right) < \lambda$$

Part (b) is proven similarly. □

Date: April 17, 2025.

Theorem 1.2. *Let (X, d) be a metric space, $F : X \to \overline{\mathbb{R}}$ be such that $\mathrm{zer}(F) \neq \varnothing$, and $\phi : X \times X \to \overline{\mathbb{R}}$ be some a distance function. Let (x_n) be a sequence of functions $\Omega \to X$ such that $\phi(x_n, p)$, $d(x_n, x_m)$ and $F(x_n)$ are real-valued (nonnegative) random variables. We suppose that (x_n) is quasi-Fejér monotone in the sense that*

$$\mathbb{E}[\phi(x_{n+1}, p) \mid \mathcal{F}_n] \leqslant (1 + a_n)\phi(x_n, p) + c_n$$

for all $p \in \mathrm{zer}(F)$ and $n \in \mathbb{N}$, where $\phi(x_n, p)$, a_n and c_n are adapted to $(\mathcal{F}_n)$, and $K > 0$ and χ satisfy

$$\prod_{i=0}^{\infty}(1 + \mathbb{E}[a_i]) < K \quad \text{and} \quad \sum_{i=\xi(\varepsilon)}^{\infty} \mathbb{E}[c_i] < \varepsilon$$

for all $\varepsilon > 0$. Finally, suppose that

- *$F(x_n)$ has liminf modulus Φ in the sense that*
$$\forall \varepsilon > 0\, \forall N\, \exists n \in [N; \Phi(\varepsilon, N)](\mathbb{E}[F(x_n)] < \varepsilon)$$
- *F possesses a modulus of ϕ-regularity in the sense that*
$$\forall \varepsilon > 0\, \forall n\ (\mathbb{E}[F(x_n)] < \rho(\varepsilon) \to \exists p \in \mathrm{zer}(F)\ (\mathbb{E}[\phi(x_n, p)] < \varepsilon))$$
- *ϕ is uniformly triangular with modulus in the sense that*
$$\forall \varepsilon > 0\, \forall p \in \mathrm{zer}(F)\, \forall m, n\ (\phi(x_n, p), \phi(x_m, p) < \theta(\varepsilon) \to d(x_n, x_m) < \varepsilon)$$

Then $\mathbb{E}\left[\mathrm{dist}\,\phi(x_n, \mathrm{zer}(F))\right] \to 0$ with rate

$$\phi(\varepsilon) := \Psi\left(\varepsilon/K\right)$$

also $\mathrm{dist}\,\phi(x_n, \mathrm{zer}(F)) \to 0$ almost surely with rate

$$\phi'(\lambda, \varepsilon) := \Psi\left(\lambda\varepsilon/K\right)$$

and finally (x_n) is almost surely Cauchy with rate

$$\psi(\lambda, \varepsilon) := \Psi\left(\frac{\lambda \cdot \theta(\varepsilon)}{2K}\right)$$

where

$$\Psi(\varepsilon) := \Phi(\rho(\varepsilon/2), \xi(\varepsilon/2))$$

Proof. Apply Lemma 1.1 with

$$U_{p,n} := \frac{\phi(x_n, p)}{\tilde{a}_{n-1}} + \mathbb{E}\left[\sum_{i=n}^{\infty} \frac{c_i}{\tilde{a}_i} \mid \mathcal{F}_n\right] \quad \text{for} \quad \tilde{a}_i := \prod_{i=0}^{j}(1 + a_j)$$

This is a supermartingale by the quasi Fejér monotoncity property, more precisely:

$$\begin{aligned}
\mathbb{E}[U_{p,n+1} \mid \mathcal{F}_n] &= \mathbb{E}\left[\frac{\phi(x_{n+1}, p)}{\tilde{a}_n} + \mathbb{E}\left[\sum_{i=n+1}^{\infty} \frac{c_i}{\tilde{a}_i} \mid \mathcal{F}_{n+1}\right] \mid \mathcal{F}_n\right] \\
&\leqslant \frac{\mathbb{E}[\phi(x_{n+1}, p) \mid \mathcal{F}_n]}{\tilde{a}_n} + \mathbb{E}\left[\left[\sum_{i=n+1}^{\infty} \frac{c_i}{\tilde{a}_i} \mid \mathcal{F}_{n+1}\right] \mid \mathcal{F}_n\right] \\
&= \frac{\mathbb{E}[\phi(x_{n+1}, p) \mid \mathcal{F}_n]}{\tilde{a}_n} + \mathbb{E}\left[\sum_{i=n+1}^{\infty} \frac{c_i}{\tilde{a}_i} \mid \mathcal{F}_n\right] \\
&\leqslant \frac{\phi(x_n, p)}{\tilde{a}_{n-1}} + \frac{c_n}{\tilde{a}_n} + \mathbb{E}\left[\sum_{i=n+1}^{\infty} \frac{c_i}{\tilde{a}_i} \mid \mathcal{F}_n\right] = U_{p,n}
\end{aligned}$$

Now defining

$$\Psi(\varepsilon) := \Phi(\rho(\varepsilon/2), \xi(\varepsilon/2))$$

there exists some $n \in [\xi(\varepsilon/2); \Psi(\varepsilon)]$ such that $\mathbb{E}[F(x_n)] < \rho(\varepsilon/2)$ and thus $\mathbb{E}[\phi(x_n, p)] < \varepsilon/2$ for some $p \in \mathrm{zer}(F)$, so that (using $\tilde{a}_i \geqslant 1$)

$$\mathbb{E}[U_{p,n}] \leqslant \mathbb{E}[\phi(x_n,p)] + \mathbb{E}\left[\mathbb{E}\left[\sum_{i=n}^{\infty} c_i \mid \mathcal{F}_n\right]\right] = \mathbb{E}[\phi(x_n,p)] + \sum_{i=n}^{\infty}\mathbb{E}[c_i] < \varepsilon$$

Applying Lemma 1.1 (a) and (b) to $X_n := \mathrm{dist}\,\phi(x_n, \mathrm{zer}(F))$, observing that for any $p \in \mathrm{zer}(F)$

$$\frac{X_n}{K} \leqslant \frac{\phi(x_n,p)}{K} \leqslant \frac{\phi(x_n,p)}{\tilde{a}_{n-1}} \leqslant U_{p,n}$$

and so setting $\pi_1(\varepsilon) := \varepsilon/K$, we get the first two rates. Applying Lemma 1.1 (d) to $Y_{n,m} := d(x_n, x_m)$, noting that

$$U_{p,n}, U_{p,m} < \frac{\theta(\varepsilon)}{K} \to \phi(x_n,p), \phi(x_m,p) < \theta(\varepsilon) \to d(x_n,x_m) < \varepsilon$$

thus setting $\pi_4(\varepsilon) := \theta(\varepsilon)/K$, we get the final rate. □

2. A stochastic Mann scheme in Hilbert spaces

This is based on a very rough study of Combettes.

Suppose we are now working in some Hilbert space, and let $T : X \to X$ be a nonexpansive operator such that $\mathrm{fix}(T) \neq \varnothing$ with modulus of regularity in expectation ρ satisfying

$$\forall \varepsilon > 0\, \forall n\ (\mathbb{E}[\|Tx - x\|] < \rho(\varepsilon) \to \exists p \in \mathrm{fix}(T)(\mathbb{E}[\|x-p\|] < \varepsilon))$$

Suppose that x_0 and e_n are random variables and define

$$x_{n+1} = (1-\lambda_n)x_n + \lambda_n(Tx_n + e_n)$$

where $\sum \lambda_i(1-\lambda_i) = \infty$. We also suppose that $(\mathcal{X}_n)$ is a filtration such that x_n is $\mathcal{X}_n$-measurable. We first observe that (x_n) is quasi-Fejér monotone w.r.t $\mathrm{fix}(T)$ for $a_n := 0$ and $c_n := \lambda_n\mathbb{E}[e_n \mid \mathcal{X}_n]$ as for any $p \in \mathrm{fix}(T)$ we have

$$\begin{aligned}\mathbb{E}[\|x_{n+1}-p\| \mid \mathcal{X}_n] &\leqslant \mathbb{E}[(1-\lambda_n)\|x_n - p\| + \lambda_n\|Tx_n - p\| + \lambda_n\|e_n\| \mid \mathcal{X}_n]\\ &\leqslant \|x_n - p\| + \lambda_n\mathbb{E}[\|e_n\| \mid \mathcal{X}_n]\end{aligned}$$

where we note that both $(\|x_n - p\|)$ and $(\mathbb{E}[\|e_n\| \mid \mathcal{X}_n])$ are adapted to $(\mathcal{X}_n)$. Next we observe that (using $\|x+y\|^2 \leqslant \|x\|^2 + \langle x \mid y\rangle + \|y\|^2$ and $\|\alpha x + (1-\alpha)y\|^2 = \alpha\|x\|^2 + (1-\alpha)\|y\|^2 - \alpha(1-\alpha)\|x-y\|^2$)

$$\begin{aligned}&\|x_{n+1}-p\|^2\\ &= \|(1-\lambda_n)(x_n-p) + \lambda_n(Tx_n - p) + \lambda_n e_n\|^2\\ &= \|(1-\lambda_n)(x_n-p) + \lambda_n(Tx_n-p)\|^2 + 2\langle(1-\lambda_n)(x_n-p) + \lambda_n(Tx_n-p), \lambda_n e_n\rangle + \lambda_n^2\|e_n\|^2\\ &\leqslant \|(1-\lambda_n)(x_n-p) + \lambda_n(Tx_n-p)\|^2 + 2\lambda_n\|x_n-p\|\,\|e_n\| + \lambda_n^2\|e_n\|^2\\ &= \|x_n - p\| - \lambda_n(1-\lambda_n)\|Tx_n - x_n\| + 2\lambda_n\|x_n-p\|\,\|e_n\| + \lambda_n^2\|e_n\|^2\end{aligned}$$

and therefore taking expectations

$$\begin{aligned}\lambda_n(1-\lambda_n)\mathbb{E}[\|Tx_n - x_n\|] &\leqslant \mathbb{E}[\|x_n-p\|] - \mathbb{E}[\|x_{n+1}-p\|] + 2\lambda_n\mathbb{E}[\|x_n-p\|\,\|e_n\|] + \lambda_n^2\mathbb{E}[\|e_n\|^2]\\ &\leqslant \mathbb{E}[\|x_n-p\|] - \mathbb{E}[\|x_{n+1}-p\|] + 2\lambda_n\sqrt{\mathbb{E}[\|x_n-p\|^2]\mathbb{E}[\|e_n\|^2]} + \lambda_n^2\mathbb{E}[\|e_n\|^2]\end{aligned}$$

Now suppose that

- $K > 0$ is such that $\mathbb{E}[\|x_n - p\|^2] < K$ for all $n \in \mathbb{N}$ and **some** $p \in \text{fix}(T)$.
- $L > 0$ is such that $\sum \lambda_n \sqrt{\mathbb{E}[\|e_n\|^2]} < L$
- ξ is a rate of convergence for $\sum \lambda_n \sqrt{\mathbb{E}[\|e_n\|^2]}$
- r is a rate of divergence for $\sum \lambda_n(1-\lambda_n)$ in the sense that

$$\forall N \in \mathbb{N}, x > 0 \left(\sum_{i=N}^{r(N,x)} \lambda_n(1-\lambda_n) > x \right)$$

We also assume that $\mathbb{E}[\|e_n\|^2] \leqslant 1$ for all $n \in \mathbb{N}$ so that $\lambda_n^2 \mathbb{E}[\|e_n\|^2] \leqslant \lambda_n \sqrt{\mathbb{E}[\|e_n\|^2]}$. Then for this particular p we have

$$\begin{aligned}\lambda_n(1-\lambda_n)\mathbb{E}[\|Tx_n - x_n\|] &\leqslant \mathbb{E}[\|x_n - p\|] - \mathbb{E}[\|x_{n+1} - p\|] + 2\lambda_n \sqrt{\mathbb{E}[\|x_n - p\|^2]\mathbb{E}[\|e_n\|^2]} + \lambda_n^2 \mathbb{E}[\|e_n\|^2] \\ &\leqslant \mathbb{E}[\|x_n - p\|] - \mathbb{E}[\|x_{n+1} - p\|] + (2\sqrt{K}+1)\lambda_n \sqrt{\mathbb{E}[\|e_n\|^2]}\end{aligned}$$

and thus

$$\begin{aligned}\sum_{n=N}^{k} \lambda_n(1-\lambda_n)\mathbb{E}[\|Tx_n - x_n\|] &\leqslant \mathbb{E}[\|x_N - p\|] - \mathbb{E}[\|x_{k+1} - p\|] + (2\sqrt{K}+1) \sum_{n=N}^{k} \lambda_n \sqrt{\mathbb{E}[\|e_n\|^2]} \\ &\leqslant \sqrt{K} + (2\sqrt{K}+1)L\end{aligned}$$

Now suppose that $\varepsilon > 0$ and that $\mathbb{E}[\|Tx_n - x_n\|] \geqslant \varepsilon$ for all $n \geqslant N$. Then we have

$$\varepsilon \sum_{n=N}^{k} \lambda_n(1-\lambda_n) \leqslant \sum_{n=N}^{k} \lambda_n(1-\lambda_n)\mathbb{E}[\|Tx_n - x_n\|] \leqslant \sqrt{K} + (2\sqrt{K}+1)L$$

for all $k \in \mathbb{N}$, which is a contradiction for $k := r\left(N, \frac{\sqrt{K}+(2\sqrt{K}+1)L}{\varepsilon}\right) =: \Phi_{r,K,L}(\varepsilon, N)$ and so $\Phi_{K,L}$ as just defined is a liminf modulus in the sense that

$$\forall \varepsilon > 0 \, \forall N \, \exists n \in [N; \Phi_{r,K,L}(\varepsilon, N)](\mathbb{E}[\|Tx_n - x_n\|] < \varepsilon)$$

We have established the following:

Theorem 2.1. *Let X be a Hilbert space and $T : X \to X$ a nonexpansive operator with modulus of regularity in expectation ρ in the sense that*

$$\forall \varepsilon > 0 \, \forall n \, (\mathbb{E}[\|Tx - x\|] < \rho(\varepsilon) \to \exists p \in \text{fix}(T)(\mathbb{E}[\|x - p\|] < \varepsilon))$$

Suppose that x_0 and (e_n) are X-valued random variables and define

$$x_{n+1} = (1-\lambda_n)x_n + \lambda_n(Tx_n + e_n)$$

where $\sum \lambda_n(1-\lambda_n) = \infty$ with rate of divergence r. Suppose that $K > 0$ is a uniform upper bound on $(\mathbb{E}[\|x_n - p\|^2])$ for some $p \in \text{fix}(T)$, and $L > 0$ is such that

$$\sum \lambda_n \sqrt{\mathbb{E}[\|e_n\|^2]} < L$$

with rate χ. Then $\mathbb{E}[\text{dist}\,\|x_n - \text{fix}(T)\|] \to 0$ with rate Ψ, also $\text{dist}\,\|x_n - \text{fix}(T)\| \to 0$ almost surely with rate $\phi(\lambda, \varepsilon) := \Psi(\lambda\varepsilon)$ and finally (x_n) is almost surely Cauchy with rate $\psi(\lambda, \varepsilon) := \Psi(\lambda\varepsilon/4)$ where

$$\Psi(\varepsilon) := r\left(\xi(\varepsilon/2), \frac{\sqrt{K} + (2\sqrt{K}+1)L}{\rho(\varepsilon/2)}\right)$$

Proof. From Lemma 1.2, the above discussion, and noting a few things like $\mathbb{E}[\|e_n\|] \leqslant \sqrt{\mathbb{E}[\|e_n\|^2]}$ so that ξ is also a rate of convergence for $\sum \lambda_n \mathbb{E}[\|e_n\|]$. Note that

$$U_{p,n} := \|x_n - p\| + \sum_{i=n}^{\infty} \lambda_n \mathbb{E}[\|e_n\| \mid \sigma(x_0, \ldots, x_n)]$$

is the underlying supermartingale that governs almost sure convergence. □

References