RANDOMIZED ALGORITHMS OF OPTIMIZATION AND THEIR IMPLEMENTATION ON QUANTUM COMPUTERS

Oleg Granichin

Department of Mathematics and Mechanics
St. Petersburg State University
e-mail: oleg_granitchin@mail.ru

Contents

1. Preview Example
2. Problem Statement and Main Assumptions
3. SPSA Algorithms
   and Trial Simultaneous Perturbations
4. Rate of Convergence
5. LR Parameter Estimation
6. Prediction of the Signal (Filtering)
7. Quantum Computer and Randomized Algorithms
   - Quantum circuit model
   - Quantum circuit for the approximation of a function gradient
   - One approach of creating the intelligent system

1 The scalar “useful" signal detecting

$$y_n = \varphi_n \theta^\star + v_n, \quad n = 1, 2, \ldots$$

$\{\varphi_n\}$ — random known useful signal
with known mean valur M_φ and $\sigma_\varphi^2 > 0$
$\{\theta^\star = 1\}$ — *signal is present*
$\{\theta^\star = 0\}$ — *signal is absent*
$\{v_n\}$ — bounded measurement disturbances

Least–squares method

$$\hat{\theta}_n = \frac{\sum_{k=1}^n \varphi_k y_k}{\sum_{k=1}^n \varphi_k^2} \to \theta^\star + \frac{M_\varphi M_v}{\sigma_\varphi^2} \quad \text{a.s.}$$

$\delta = \dfrac{M_\varphi M_v}{\sigma_\varphi^2} + \dfrac{1}{2}$ — level of decisionmaking
If $\hat{\theta}_n < \delta$ then *signal is absent*
Otherwise — *signal is present*

Usual assumptions:
$M_v = 0$ and $\{v_n\}$ is a random i.i.d.

Let M_v be unknown and
$\{v_n\}$ doesn't depend on $\{\varphi_n\}$
or $\{v_n\}$ is a nonrandom *unknown but bounded*

Membership set approach
If $|v_n| \leq C_v,\ n = 1, 2, \ldots$ then at the time moment N the membership set is

$$\Theta_N = \cap_{n=1}^{N} \{\theta \in \mathrm{R} :\ |y_n - \varphi_n \theta| \leq C_v\}.$$

$$\theta^\star \in \Theta_N$$

Does the sequence $\{\Theta_N\}$ converges to $\theta^\star$???

- How to solve the problem?
- What is an appropriate result of measurements?
- Is it necessary to have many independent measurements with zero-mean?

Denote $\Delta_n = \varphi_n - M_\varphi$

$$\Delta_n y_n = \Delta_n^2 \theta^\star + \Delta_n M_\varphi \theta^\star + \Delta_n v_n$$

$$\frac{\sum_{k=1}^n \Delta_k y_k}{\sum_{k=1}^n \Delta_k^2} = \theta^\star + \frac{\sum_{k=1}^n \Delta_k (M_\varphi \theta^\star + v_k)}{\sum_{k=1}^n \Delta_k^2}$$

Randomized least–squares method

$$\hat{\theta}_n = \frac{\sum_{k=1}^n \Delta_k y_k}{\sum_{k=1}^n \Delta_k^2} \to \theta^\star \quad \text{a.s.}$$

Level of decisionmaking is $\delta = 1/2$.

Randomized algorithms:

$$\hat{\theta}_n = \hat{\theta}_{n-1} - \Delta_n Y_n$$

Δ_n is a *trial simultaneous perturbation*

Examples

$\{\varphi_n\}$ is a random i.i.d.: $\varphi_n \in [0.5, 1.5]$

$\{v_n\}$ is a nonrandom *unknown but bounded* sequence: $|v_n| \leq 2$

Let $n = 1, \ldots, 100, 350, \ldots, 500$

$\{\theta_n^\star = 0\}$ — *signal is absent*

$n = 101, \ldots, 349$

$\{\theta_n^\star = 1\}$ — *signal is present*

$$\sum_{n=1}^{250} v_n = 1, \qquad \sum_{n=250}^{500} v_n = -1$$

$$\hat{\theta}_n = \hat{\theta}_{n-1} - 0.1\varphi_n(\varphi_n\hat{\theta}_{n-1} - y_n)$$

$$\hat{\theta}_n = \hat{\theta}_{n-1} - 0.1\Delta_n(\varphi_n\hat{\theta}_{n-1} - y_n)$$

2 Problem statement and main assumption

Let $F(w,\theta):\ \mathbb{R}^p \times \mathbb{R}^d \to \mathbb{R}^1$ be diff. on θ, $x_1, x_2 \ldots$ is an observation plan,

$$y_n = F(w_n, x_n) + v_n,$$

$\{w_n\}$ is an uncontrollable i.i.d. $w_n \in \mathbb{R}^p$ with unknown distribution $\mathrm{P}_w(\cdot)$.

It is required to find the unknown vector $\theta^\star$ which minimizes a function

$$f(\theta) = \int_{\mathbb{R}^p} F(w,\theta)\mathrm{P}_w(dw)$$

by using the observations $y_1, y_2 \ldots$.

More simple observation model:

$$y_n = f(x_n) + v_n.$$

Assumptions:

(A.1) $f(\cdot)$ has a unique root in R^d at $\theta^\star$

$$\langle x-\theta^\star, \nabla f(x)\rangle \geq \mu\|x-\theta^\star)\|^2, \quad \forall x \in \mathrm{R}^d$$

with some constant $\mu > 0$.

(A.2) Lipschitz condition for the gradient

$$\|\nabla f(x) - \nabla f(\theta)\| \leq A\|x-\theta\|, \ \forall x, \theta \in \mathrm{R}^d$$

with some constant $A > \mu$.

(A.3) Function $f(\cdot) \in C^\ell$ is ℓ–times continuously differentiable and for all its partial derivatives up to the order ℓ the Holder condition of order ρ $(0 < \rho \leq 1)$, holds on R^d so that

$$|f(x) - \sum_{|\bar{l}|\leq\ell} \frac{1}{\bar{l}!} D^{\bar{l}} f(\theta)(x-\theta)^{\bar{l}}| \leq M\|x-\theta\|^\gamma,$$

where $\gamma = \ell + \rho \geq 2$, M — some constant,
$\bar{l} \in \mathbb{N}^d$ is a multi-index,
If $\gamma = 2$ then $M = A/2$.

3 SPSA algorithms and trial simultaneous perturbation

Let $\{\Delta_n\}$ be *trial simultaneous perturbation.* $\Delta_n \in \mathrm{R}^d$, its distribution function is $\mathrm{P}_n(\cdot)$.

One measurement form of SPSA method

$$\begin{cases} x_n = \hat{\theta}_{n-1} + \beta_n \Delta_n \\ y_n = F(w_n, x_n) + v_n \\ \hat{\theta}_n = \hat{\theta}_{n-1} - \frac{\alpha_n}{\beta_n} \mathcal{K}_n(\Delta_n) y_n \end{cases} \tag{1}$$

“Smoothed" versions of the Kiefer-Wolfowitz procedure

$$\begin{cases} x_{2n} = \hat{\theta}_{n-1} + \beta_n \Delta_n \\ x_{2n-1} = \hat{\theta}_{n-1} - \beta_n \Delta_n \\ \hat{\theta}_n = \hat{\theta}_{n-1} - \frac{\alpha_n}{2\beta_n} \mathcal{K}_n(\Delta_n)(y_{2n} - y_{2n-1}) \end{cases} \tag{2}$$

$$\begin{cases} x_{2n} = \hat{\theta}_{n-1} + \beta_n \Delta_n \\ \\ x_{2n-1} = \hat{\theta}_{n-1} \\ \\ \hat{\theta}_n = \hat{\theta}_{n-1} - \frac{\alpha_n}{\beta_n} \mathcal{K}_n(\Delta_n)(y_{2n} - y_{2n-1}) \end{cases} \tag{3}$$

Here $\mathcal{K}_n(\cdot) : \mathbb{R}^d \to \mathbb{R}^d$ are smoothed kernels with the finite support:

$$\int \mathcal{K}_n(x) x^{\mathrm{T}} \mathrm{P}_n(dx) = \mathrm{I}$$

$$\int \mathcal{K}_n(x) \mathrm{P}_n(dx) = 0 \tag{4}$$

$$\sup_n \int \|\mathcal{K}_n(x)\|^2 \mathrm{P}_n(dx) < \infty$$

$$\mathcal{K}_n(\Delta_n) = \Delta_n$$

$$\mathcal{K}_n(\Delta_n) = \begin{pmatrix} \frac{1}{\Delta_n^{(1)}} \\ \frac{1}{\Delta_n^{(2)}} \\ \vdots \\ \frac{1}{\Delta_n^{(d)}} \end{pmatrix}$$

Let $\{\Theta_n\}$ is a sequence of convex closed sets: $\theta_\star \in \Theta_n \subset \mathrm{R}^d$ for sufficiently large $n \geq 1$

$$\begin{cases} x_n = \hat{\theta}_{n-1} + \beta_n \Delta_n \\ \\ y_n = F(w_n, x_n) + v_n \\ \\ \hat{\theta}_n = \mathcal{P}_{\Theta_n}(\hat{\theta}_{n-1} - \frac{\alpha_n}{\beta_n}\mathcal{K}_n(\Delta_n) y_n) \end{cases} \tag{5}$$

$\mathcal{P}_{\Theta_n}$ is a projector on Θ_n.

Denote $\mathbb{W} = \mathrm{supp}(\mathrm{P}_w(\cdot))$,
$\mathcal{F}_{n-1}$ — σ-algebra $\{\hat{\theta}_0, \hat{\theta}_1, \ldots, \hat{\theta}_{n-1}\}$,

$$\bar{v}_n = v_{2n} - v_{2n-1},\ \bar{w}_n = \begin{pmatrix} w_{2n} \\ w_{2n-1} \end{pmatrix},\ d_n = 1$$

for (2) or (3)

$$\bar{v}_n = v_n,\ \bar{w}_n = w_n,\ d_n = \mathrm{diam}\Theta_n \quad \text{for} \quad (5)$$

<u>**T.**</u> **1** *If the condition* (A.1) *is held for* $f(\theta)$*;*
(A.2) *for* $F(w,\cdot)$ $\forall w \in \mathbb{W}$*;*
(4) *for* $\mathcal{K}_n(\cdot)$ *and* $\mathrm{P}_n(\cdot)$*;*
$F(\cdot,\theta)$, $\nabla_\theta F(\cdot,\theta)$ *uniformly bounded on* $\mathbb{W}$*;*
$\bar{v}_1,\ldots,\bar{v}_n,\bar{w}_1,\ldots,\bar{w}_{n-1}$ *don't depend on* $\bar{w}_n,\Delta_n$, $\bar{w}_n$ *doesn't depend on* Δ_n*;*

$$\mathrm{E}\{\bar{v}_n^2\} \le \sigma_n^2.$$

If $\sum_n \alpha_n = \infty$ *and*

$$\alpha_n \to 0,\ \beta_n \to 0,\ \alpha_n^2\beta_n^{-2}(1+d_n^2+\sigma_n^2) \to 0$$

then

$$\mathrm{E}\{\|\hat{\theta}_n - \theta^\star\|^2\} \to 0 \ \text{ as } \ n \to \infty.$$

Moreover, if $\sum_n \alpha_n\beta_n^2 < \infty$ *and*

$$\sum_n \alpha_n^2\beta_n^{-2}(1+\mathrm{E}\{\bar{v}_n^2|\mathcal{F}_{n-1}\}) < \infty \ \text{ a.s.}$$

then $\hat{\theta}_n \to \theta^\star$ *as* $n \to \infty$ *a.s.*

Newton method

$$\hat{\theta}_n = \hat{\theta}_{n-1} - [\nabla^2 f(\hat{\theta}_{n-1})]^{-1} \nabla f(\hat{\theta}_{n-1})$$

$[\nabla^2 f(\hat{\theta}_{n-1})]^{-1}$ is replaced by α_n.
Let $\nabla f(\hat{\theta}_{n-1})$ be unknown.

$$\nabla f(\hat{\theta}_{n-1})^{(i)} \approx$$

$$\approx \frac{f(\hat{\theta}_{n-1} + \beta_n e_i) - f(\hat{\theta}_{n-1} - \beta_n e_i)}{2\beta_n}$$

$2d$ measurements per iteration

$\{\Delta_n\}$ — i.i.d. random vectors:

$$\mathrm{cov}\{\Delta_n \Delta_j^{\mathrm{T}}\} = \delta_{nj}\,\mathrm{I}, \quad \mathrm{E}\{\Delta_n v_n\} = 0$$

One measurement form of SPSA method

$$\begin{cases} y_n = F(w_n, \hat{\theta}_{n-1} + \beta_n \Delta_n) + v_n, \\ \hat{\theta}_n = \hat{\theta}_{n-1} - \frac{\alpha_n}{\beta_n} \Delta_n y_n \end{cases}$$

Denote $D_n = (\hat{\theta}_n - \theta)^2$

$$\mathrm{E}\{D_n|\hat{\theta}_i, i<n\} \le D_{n-1} - \frac{\alpha_n}{\beta_n}(\hat{\theta}_{n-1}-\theta)\mathrm{E}\{\Delta_n y_n|\hat{\theta}_i, i<n\}+$$

$$+\frac{\alpha_n^2}{\beta_n^2}\mathrm{E}\{\Delta_n^2 y_n^2|\hat{\theta}_i, i<n\} = D_{n-1}-$$

$$-\frac{\alpha_n}{\beta_n}(\hat{\theta}_{n-1}-\theta)\mathrm{E}\{\Delta_n f(\theta_{n-1}+\beta_n\Delta_n)|\hat{\theta}_i, i<n\}+$$

$$+\frac{\alpha_n^2}{\beta_n^2}\mathrm{E}\{y_n^2|\hat{\theta}_i, i<n\}.$$

Taylor series

$$f(\hat{\theta}_{n-1}+\beta_n\Delta_n) = f(\hat{\theta}_{n-1})+\beta_n\Delta_n\nabla f(\hat{\theta}_{n-1})+$$

$$+\beta_n^2\zeta_n,\ \zeta_n \in [\hat{\theta}_{n-1}, \hat{\theta}_{n-1}+\beta_n\Delta_n]$$

Since $\mathrm{E}\{\Delta_n\} = 0$

$$\mathrm{E}\{D_n|\hat{\theta}_i, i<n\} \le (1+\frac{1}{2}\alpha_n\beta_n)D_{n-1}-$$

$$-\alpha_n(\hat{\theta}_{n-1}-\theta)\nabla f(\hat{\theta}_{n-1}) + \xi_n,$$

$$\xi_n = \mathrm{E}\{\alpha_n\beta_n\zeta_n^2/2 + \alpha_n^2\beta_n^{-2}y_n^2|\hat{\theta}_i, i<n\}$$

$\Longrightarrow \{D_n\}$ is almost semimartingal

$$\mathrm{E}\{D_n|\hat{\theta}_i, i<n\} \le (1-\gamma_n)D_{n-1} + \xi_n,$$

$$\gamma_n = \mu\alpha_n - \alpha_n\beta_n/2$$

4 Rate of convergence

Let $\mathrm{P}_n(\cdot) = \mathrm{P}_\Delta(\cdot)$ and

$$\mathcal{K}_n(x) = \mathcal{K}(x) = (\mathcal{K}^{(1)}(x), \ldots, \mathcal{K}^{(r)}(x))^{\mathrm{T}},$$

$$\mathcal{K}^{(i)}(x) = K_0(x^{(i)}) \prod_{j \neq i} K_1(x^{(j)}) \qquad (6)$$

T. 2 *Let be* $\alpha_n = \alpha\, n^{-1}$, $\beta_n = \beta\, n^{-\frac{1}{2\gamma}}$.
If the following conditions are fulfilled:
(A.1,3) *under* $\gamma \geq 2$, $\alpha\beta > \frac{\gamma-1}{2\mu\gamma}$ *for* $f(\theta)$;
(A.2) *for* $F(w, \cdot)$ $\forall w \in \mathbb{W}$;
$F(\cdot, \theta)$, $\nabla_\theta F(\cdot, \theta)$ *uniformly bounded on* $\mathbb{W}$;

$$\int u K_0(u) \mathrm{P}_\Delta(du) = 1, \int u^k K_0(u) \mathrm{P}_\Delta(du) = 0, \ k = 0, 2, \ldots, \ell, \qquad (7)$$

$$\int K_1(u) \mathrm{P}_\Delta(du) = 1, \int u^k K_1(u) \mathrm{P}_\Delta(du) = 0, \ k = 1, \ldots, \ell - 1;$$

$\bar{w}_n, \Delta_n$ *don't depend on* $\bar{v}_1, \ldots, \bar{v}_n, \bar{w}_1, \ldots, \bar{w}_{n-1}$,
Δ_n *doesn't depend on* $\bar{w}_n$;
$\mathrm{E}\{(v_{2n} - v_{2n-1})^2/2\} \leq \sigma_2^2$, $(\mathrm{E}\{v_n^2\} \leq \sigma_1^2)$;

$$d_n n^{-1+\frac{1}{2\gamma}} \to 0 \ \text{ as } \ n \to \infty$$

then

$$\mathrm{E}\{\|\hat{\theta}_n - \theta^\star\|^2\} = \mathcal{O}(n^{-\frac{\gamma-1}{\gamma}}) \ \text{ as } \ n \to \infty.$$

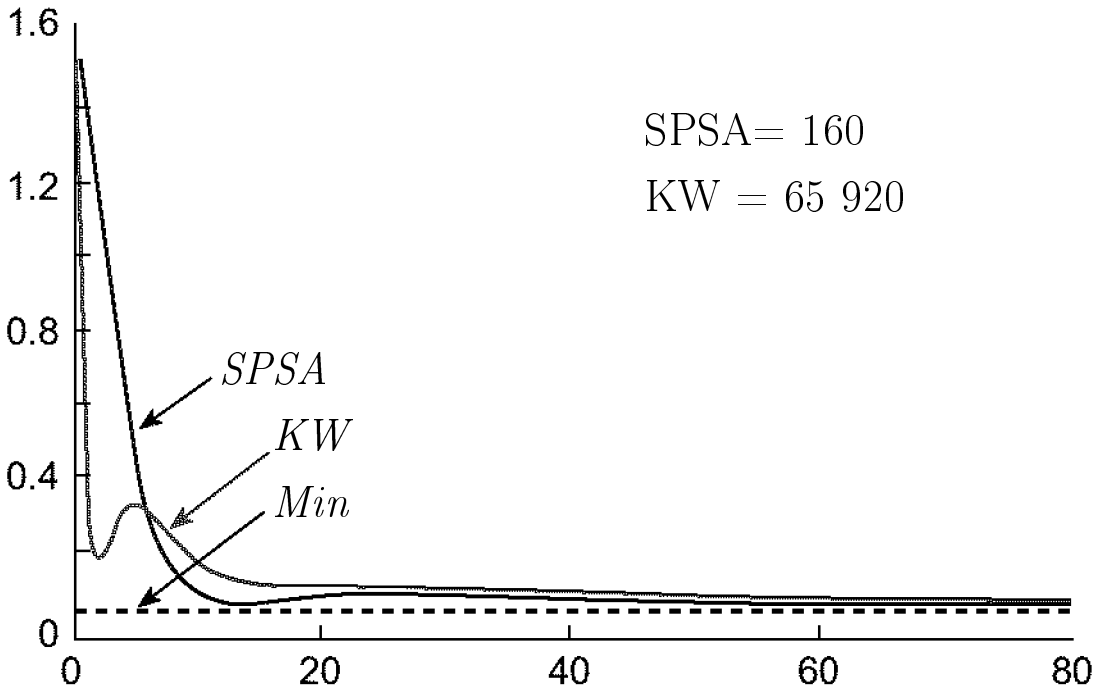
If $\{\Delta_n\}$ is uniformly distributed on $[-1/2, 1/2]$ then for $\ell = 1, 2$ (i.e. $2 \le \gamma \le 3$) we have

$K_0(u) = 12u, K_1(u) = 1,$

for $\ell = 3, 4$ (i.e. $3 < \gamma \le 5$)

$K_0(u) = 5u(15 - 84u^2),\ K_1(u) = \frac{9}{4} - 15u^2.$

Note, $K_0(u) = 0$ and $K_1(u) = 0$ as $|u| > 1/2$.



$d = 412$

5 Linear regression

$$y_n = \varphi_n^{\mathrm{T}}\theta_n^\star + v_n,\ \theta_n^\star = \theta^\star + w_n \tag{8}$$

$y_n \in \mathbb{R}^1$ — output (observation)
$\varphi_n \in \mathbb{R}^r$ — input (regressor)
$v_n \in \mathbb{R}^1,\ w_n \in \mathbb{R}^r$ — noise (disturbance)

It is required to find the unknown vector $\theta^\star$ by using the observations $y_i,\ \varphi_i,\ i \le n$.

(A) $\{\varphi_n\}$ — i.i.d., φ_n not depends with $\tilde{\mathcal{F}}_{n-1}$,
$\mathrm{E}\varphi_n$ is known: $\|\mathrm{E}\varphi_n\| \le M_\varphi < \infty$,
$\Delta_n = \varphi_n - \mathrm{E}\varphi_n$ has a symmetric distribution,

$$\mathrm{E}\Delta_n\Delta_n^T = \mathrm{B_n} > 0, \|\mathrm{B}_n\| \le \sigma_\Delta^2$$

(B) w_n doesn't depend on $\hat{\mathcal{F}}_{n-1}$ and $\mathrm{E}w_n = 0$

$(i)\quad \mathrm{E}\{\, v_n^2|\mathcal{F}_{n-1}\} \le \sigma_v^2\ \mathrm{E}\{\|w_n\|^2\} \le \sigma_w^2,$

$(ii)\quad \mathrm{E}\{v_n^2\} \le \sigma_v^2,\ \mathrm{E}\{w_n w_n^{\mathrm{T}}\} \le \mathrm{Q}_w$

$\mathcal{F}_n$ — σ-algebra
$\{\varphi_1, \ldots, \varphi_n, w_1, \ldots, w_n, v_1, \ldots, v_n\}$
$\hat{\mathcal{F}}_{n-1}\ \{\varphi_1, \ldots, \varphi_{n-1}, w_1, \ldots, w_{n-1}, v_1, \ldots, v_n\}$
$\tilde{\mathcal{F}}_{n-1}\ \{\varphi_1, \ldots, \varphi_{n-1}, w_1, \ldots, w_n, v_1, \ldots, v_n\}$
$\mathcal{F}_{n-1} \subset \hat{\mathcal{F}}_{n-1} \subset \tilde{\mathcal{F}}_{n-1} \subset \mathcal{F}_n\,.$

$$\hat{\theta}_n = \hat{\theta}_{n-1} - \alpha_n \Gamma \Delta_n (\varphi_n^{\mathrm{T}} \hat{\theta}_{n-1} - y_n) \qquad (9)$$

<u>**T.**</u> **3** *Let Assumption* (A) *be fulfilled and*

$$\alpha_n \to 0, \ \sum_{n=1}^{\infty} \alpha_n \ = \ \infty.$$

If (Bi) *and* $\sum \alpha_n^2 < \infty$
then

$$\hat{\theta}_n \to \theta^\star \ \text{ a.s.}$$

If (Bii) **then**

$$\mathrm{E}(\hat{\theta}_n - \theta^\star)(\hat{\theta}_n - \theta^\star)^{\mathrm{T}} \to 0$$

If $\Gamma = \mathrm{B}^{-1}$ then (9) = (2) and

$$F(w, x) = \frac{1}{2}(x - \theta^\star - w)^{\mathrm{T}}(x - \theta^\star - w)$$

<u>T.</u> 4 *If* (A) *and* (Bii) *are fulfilled,* $\alpha_n = n^{-1}$, $-\Gamma\mathrm{B} + \frac{1}{2}\mathrm{I}$ *is a Hurwitz matrix*

then

$$\mathrm{E}\{(\hat{\theta}_n - \theta^\star)(\hat{\theta}_n - \theta^\star)^{\mathrm{T}}\} \leq n^{-1}\,\mathrm{S} \;+\; o(n^{-1}),$$

where S *is a solution of matrix equation*

$$\Gamma\,\mathrm{B}\,\mathrm{S} \;+\; \mathrm{S}\,\mathrm{B}\,\Gamma \;-\; \mathrm{S} \;=\; \Gamma\mathrm{R}\Gamma \qquad (12)$$

$$\mathrm{R} \;=\; (\sigma_v^2(1 + M_\varphi^2\rho) + M_\varphi^2\mathrm{Tr}[\mathrm{Q}_w])\mathrm{B} + \\ + \mathrm{E}\{\Delta_n\Delta_n^{\mathrm{T}}\mathrm{Q}_w\Delta_n\Delta_n^{\mathrm{T}}\}$$

$\rho > 0$ *is any small positive constant*

If $\Gamma \;=\; \mathrm{B}^{-1}$, $\mathrm{Tr}[\mathrm{Q}_w] = 0$, $M_\varphi = 0$ then for

$$\hat{\theta}_n \;=\; \hat{\theta}_{n-1} \;-\; (n\mathrm{B})^{-1}\varphi_n(\varphi_n^{\mathrm{T}}\hat{\theta}_{n-1} - y_n)$$

we have

$$\mathrm{E}\{(\hat{\theta}_n - \theta^\star)(\hat{\theta}_n - \theta^\star)^{\mathrm{T}}\} \leq n^{-1}\sigma_v^2\mathrm{B}^{-1} + o(n^{-1})$$

$$\begin{cases} \hat{\theta}_n = \hat{\theta}_{n-1} - \Gamma_n \Delta_n (\varphi_n^{\mathrm{T}} \hat{\theta}_{n-1} - y_n) \\ \Gamma_n = \Gamma_{n-1} - \frac{\Gamma_{n-1} \Delta_n \Delta_n^{\mathrm{T}} \Gamma_{n-1}}{1 + \Delta_n^{\mathrm{T}} \Gamma_{n-1} \Delta_n} \end{cases} \tag{14}$$

<u>T.</u> 5 *Let* (A) *be fulfilled.*

If (Bi)

then

$$\hat{\theta}_n \to \theta^\star.$$

If $|v_n| \leq C_v,\ \|w_n\| \leq C_w,\ \|\Delta_n\| \leq C_\Delta$

then

$$\mathrm{E}\{(\hat{\theta}_n - \theta^\star)(\hat{\theta}_n - \theta^\star)^{\mathrm{T}}\} \to 0.$$

6 Filtering

$$y_n = \varphi_n^{\mathrm{T}} \theta_n^\star + v_n \tag{15}$$

$$\theta_{n+1}^\star = \mathrm{A}\theta_n^\star + w_{n+1} \tag{16}$$

$\|\mathrm{A}\| < 1$, w_n doesn't depend on $\mathcal{F}_{n-1}$
$\{w_n\}$ and $\{v_n\}$ satisfy conditions (Bii)

$$\mathrm{E}\|\hat{\theta}_{n+1} - \theta_{n+1}^\star\|^2 \to \min$$

$\{\varphi_n\}$ — random and (A)

Randomized least mean squares algorithm

$$\hat{\theta}_{n+1} = \mathrm{A}\hat{\theta}_n - \alpha \mathrm{A}\Gamma\Delta_n(\varphi_n^{\mathrm{T}}\hat{\theta}_n - y_n) \tag{17}$$

$$\Delta_n = \varphi_n - \mathrm{E}\{\varphi_n\}$$

$$\alpha > 0 \text{ step} - \text{size}, \ \ \Gamma > 0$$

Substitute (15) and (16) into (17):

$$\hat{\theta}_{n+1} - \theta^\star_{n+1} = \mathrm{A}(\mathrm{I} - \alpha\Gamma\Delta_n\Delta_n^{\mathrm{T}})(\hat{\theta}_n - \theta^\star_n) -$$
$$-\alpha\mathrm{A}\Gamma\Delta_n(\mathrm{E}\{\varphi_n\}^{\mathrm{T}}(\hat{\theta}_n - \theta^\star_n) - v_n) - w_{n+1}.$$

Successively take the conditional expectations

$$\mathrm{E}\{\|\hat{\theta}_{n+1} - \theta^\star_{n+1}\|^2|\tilde{\mathcal{F}}_{n-1}\} \le (1 - 2\alpha\lambda_{\min}(\mathrm{B}\Gamma) +$$
$$+\alpha^2\|\Gamma\|^2 M_4^4)\|\mathrm{A}\|^2\|\hat{\theta}_n - \theta^\star_n\|^2 +$$
$$+\alpha^2(\mathrm{E}\{\varphi_n\}^{\mathrm{T}}(\hat{\theta}_n - \theta^\star_n) - v_n)^2\|\Gamma\|^2\mathrm{Tr}[\mathrm{B}] + \mathrm{Tr}[\mathrm{Q}_w].$$

Taking the unconditional expectation we obtain with any $\rho > 0$

$$\mathrm{E}\{\|\hat{\theta}_{n+1} - \theta^\star_{n+1}\|^2\} \le b(\alpha,\rho)\mathrm{E}\{\|\hat{\theta}_n - \theta^\star_n\|^2\} +$$
$$+\alpha^2(1 + M_\varphi\rho)\|\Gamma\|^2\mathrm{Tr}[\mathrm{B}]\sigma_v^2 + \mathrm{Tr}[\mathrm{Q}_w],$$

where

$$b(\alpha,\rho) = (1 - 2\alpha\lambda_{\min}(\mathrm{B}\Gamma) + \alpha^2\|\Gamma\|^2 M_4)\|\mathrm{A}\|^2 + \tag{18}$$
$$+\alpha^2(M_\varphi + \frac{1}{\rho})M_\varphi\|\Gamma\|^2\mathrm{Tr}[\mathrm{B}].$$

<u>T.</u> 6 *If* (A) *and* (Bii) *are fulfilled*
then $\forall\rho > 0$ *and small* α*:* $b(\alpha,\rho) < 1$

$$\mathrm{E}\{\|\hat{\theta}_{n+1} - \theta^\star_{n+1}\|^2\} \le$$
$$\le \frac{\mathrm{Tr}[\mathrm{Q}_w] + \alpha^2(1 + M_\varphi\rho)\|\Gamma\|^2\mathrm{Tr}[\mathrm{B}]\sigma_v^2}{1 - b(\alpha,\rho)} +$$
$$+b(\alpha,\rho)^n\mathrm{E}\{\|\hat{\theta}_0 - \theta^\star_0\|^2\}$$

$\Gamma = \mathrm{B}^{-1}, \|\mathrm{A}\|^{-2} = 1 + \mathcal{O}(\alpha^3)$

Denote

$$r(\rho) = \frac{M_4 + (M_\varphi + 1/\rho) M_\varphi \mathrm{Tr}[\mathrm{B}]}{2\lambda^2_{\min}(\mathrm{B})}$$

For sufficiently small α

$$\mathrm{E}\{\|\hat{\theta}_{n+1} - \theta^\star_{n+1}\|^2\} \leq D(\alpha, \rho) + \mathcal{O}(\alpha^2),$$

where

$$D(\alpha, \rho) = \frac{1}{2}\mathrm{Tr}[\mathrm{Q}_w](\frac{1}{\alpha} + r(\rho) + \tag{19}$$

$$+ \left(r(\rho)^2 + \frac{(1 + M_\varphi \rho)\mathrm{Tr}[\mathrm{B}]\sigma_v^2}{\mathrm{Tr}[\mathrm{Q}_w]\lambda^2_{\min}(\mathrm{B})} \right) \alpha)$$

If $M_\varphi = 0$ then we get

$$\alpha^\star = 2\lambda^2_{\min}(\mathrm{B}) \sqrt{\frac{\mathrm{Tr}[\mathrm{Q}_w]}{M_4^2 \mathrm{Tr}[\mathrm{Q}_w] + 2\lambda^2_{\min}(\mathrm{B})\mathrm{Tr}[\mathrm{B}]\sigma_v^2}}$$

Let be $d = 1$ and $\sigma_w^2 = \mathrm{Tr}[\mathrm{Q}_w] << \sigma_v^2$

$$\mathrm{Prob}\{\varphi_n = \bar{\varphi}\} = \mathrm{Prob}\{\varphi_n = -\bar{\varphi}\} = 1/2$$

We have

$$\alpha^\star \mathrm{A}\Gamma \approx \sigma_w / |\bar{\varphi}| \sigma_v$$

$\{\varphi_n\}$ are uniformly distributed on $[0.5, 1.5]$
$\{v_n\}$: $|v_n| \leq 2$

$$\theta^\star_{n+1} = 0.9999\theta^\star_n + w_{n+1}, \theta^\star_1 = 0$$

$\{w_n\}$ are uniformly distributed on $[-\frac{1}{3}, \frac{1}{3}]$

$$\tilde{D}(\{\hat{\theta}_n\}) = \frac{1}{199}\sum_{n=1}^{199} \|\hat{\theta}_{n+1} - \theta^\star_{n+1}\|^2$$

$$\alpha^\star\Gamma = 0.2371, D(\alpha^\star, \rho^\star) = 1.3699, \rho^\star = 0.269$$

$$\Delta_n = \varphi_n - 0.1$$

$$\hat{\theta}_{n+1} = 0.9999(\hat{\theta}_n - 0.2371\Delta_n(\varphi_n\hat{\theta}_n - y_n)),$$

$$\hat{\theta}_{n+1} = 0.9999(\hat{\theta}_n - 0.2371\varphi_n(\varphi_n\hat{\theta}_n - y_n)),$$

$$\hat{\theta}_{n+1} = 0.9999\hat{\theta}_n - k_n\varphi_n(\varphi_n\hat{\theta}_n - y_n),$$

$$k_n = \frac{0.9999\gamma_{n-1}}{\frac{16}{3} + \gamma_{n-1}\phi_n^2},$$

$$\gamma_n = \gamma_{n-1}0.9999^2 - \frac{\phi_n^2\gamma_{n-1}^2}{\frac{16}{3} + \gamma_{n-1}\varphi_n^2} + \frac{2}{81}, \ \gamma_0 = 0.$$

7 Quantum computer and randomized algorithms

The Quantum Circuit Model

Classical computer treats bits $\{0, 1\}$.
Quantum computer treats the quantum bits.
Qubit is a two–state microscopic system (an atom or nuclear spin or polarized photon).
It is a unit vector in $\mathbb{C}^2$ with inner product.
Denote base vectors of this space $|0\rangle$, $|1\rangle$.

The quantum state space of k qubits systems is the complex projective space $\mathbb{C}^{2^k}$.

$$|b_1 b_2 \ldots b_k\rangle = |b_1\rangle \otimes |b_2\rangle \otimes \ldots \otimes |b_k\rangle$$

$$W = \sum_s \psi_s |s\rangle,$$

W — a superposition of $|s\rangle$,

$$\psi_s \in \mathbb{C}, \ \sum_s |\psi_s|^2 = 1,$$

$|\psi_s|$ — probabilistic amplitudes.

The projection on some V is an outcome

$$\langle V, W\rangle$$

Quantum parallelism

A black box: $\{0,1\} \to \{0,1\}$.

Computation takes 24 hours.

$$f(x) : \mathbb{Z}^2 \to \mathbb{Z}^2$$

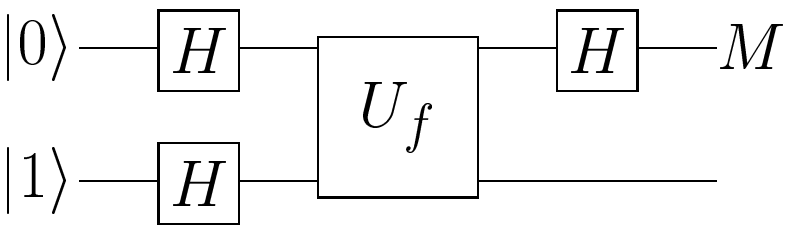
What time is needed to $M = f(0) \oplus f(1)$?

A quantum invertible black box

$$U_f : \ |x\rangle|y\rangle \to |x\rangle|y \oplus f(x)\rangle$$

Hadamard transform

$$H : |x\rangle \to \frac{1}{\sqrt{2}} \sum_y (-1)^{xy} |y\rangle, \ \ H = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$$



Quantum Fourier transform

$$QFT : \ |x\rangle \to \frac{1}{\sqrt{N}} \sum_y e^{2\pi xy/N} |y\rangle =$$

$$= \frac{1}{\sqrt{2^n}} \left(|0\rangle + e^{2\pi(.x_0)}|1\rangle\right) \left(|0\rangle + e^{2\pi(.x_1 x_0)}|1\rangle\right) \cdots \left(|0\rangle + e^{2\pi(.x_{n-1} x_{n-2} \ldots x_0)}|1\rangle\right)$$

Efficient quantum algorithms

1. Quantum Fourier transform

Factoring is an intractable problem:

- The solution can be easily verified.
- But the solution is hard to find.

If p and q are large prime numbers then

$$n = pq$$

can be computed quickly $\log_2 p \cdot \log_2 q$.
But given n, it is hard to find p and q.

The "number field sieve" algorithm requires

$$time \approx \exp\{(64/9)^{\frac{1}{3}}(\ln n)^{\frac{1}{3}}(\ln \ln n)^{\frac{2}{3}}\}$$

P. Shor $time \approx \mathcal{O}[(\ln n)^3]$

2. Quantum database search

Grover's algorithm $time \approx \frac{\pi}{4}\sqrt{N}$.

3. Quantum bit teleportation

4. Approximation of the gradient of a multivariable function (SPSA iteration)?

Approximation of a Function Gradient

$$U_f : \mathbb{C}^{2^k} \to \mathbb{C}^{2^k}, \quad k = p \times d$$

$$\forall x \in \mathrm{R} \;\; s_x = \overline{b_x^{(1)} \dots b_x^{(p)}}$$

It maps

$$|s_{x^{(1)}} \dots s_{x^{(d)}}\rangle \to |s_{f(X)} 00 \dots 0\rangle = U_f |s_{x^{(1)}} \dots s_{x^{(d)}}\rangle$$

$\mathcal{I}$ — *input*
W — *worker*
Δ — *simultaneous perturbation*

Sum U_{+R}

Turn of the first qubits $U_{R_{1,p+1,\dots,(d-1)p+1}}$

Shift by j qubits U_{S_j}

1. Send zero to $\mathcal{I}, W, \Delta$.
2. $\mathcal{I} := |s_x\rangle, \; \Delta := U_{R_{1,p+1,\dots,(d-1)p+1}} \Delta$.
3. $W := U_f U_{+\mathcal{I}} U_{S_j} U_{+\Delta} W$,
4. To measure the outcome of of calculations

$$\hat{g}^{(i)}(X) = \langle U_{S_{-(i-1)p}} \Delta, W \rangle, \; i = 1, 2, \dots, d$$

Creating the Intelligent System

$$\{T_1, T_2, \ldots, T_m\} \leftrightarrow \{D_1, D_2, \ldots, D_m\}$$
$$D_i = D_i(\theta^{(i)}),\ \theta^{(i)} \in \Theta^{(i)}$$

- $\forall P_i$ the system has D_i which is able to optimize the solving of T_i by an appropriate choosing $\theta^{(i)}$
- information from external world must be delivered to all such devices simultaneously

Performance indexes of data w processing

$$f_i = f_i(w, \theta^{(i)}),\ i = 1, 2, \ldots, m$$

$$\theta = \begin{pmatrix} \theta^{(1)} \\ \theta^{(2)} \\ \vdots \\ \theta^{(m)} \end{pmatrix},\ \theta \in \Theta = \Theta^{(1)} \otimes \cdots \otimes \Theta^{(m)}$$

$$F(w, \theta) = -\sum_{i=1}^{m} f_i(w, \theta^{(i)})$$

Informational resonance

- only one device has resonance,
- several devices have resonance,
- none of the devices has resonance.

References

[1] R.A. Fisher, *The Design of Experiments.* Oliver and Boyd, Edinburgh, 1935.

[2] O.N. Granichin, "A stochastic approximation algorithm with perturbations in the input for identification of static nonstationary plant," *Vestnik Leningr. Univ., Math.*, vol. 21, pp. 92–93, 1988.

[3] O.N. Granichin, "Unknown Function Minimum Point Estimation under Dependent Noise," *Problems Inform. Transmission*, v. 28, No. 2, pp.16–20, 1992.

[4] A.V. Goldenshluger and B.T. Polyak, "Estimation of regression parameters with arbitrary noise," *Mathematical Methods of Statistics*, vol. 2, pp. 18–29, 1993.

[5] L. Ljung and L. Guo, "The Role of Model Validation for Assessing the Size of the Unmodeled Dynamics," *IEEE Trans. Automat. Contr.*, vol. 42, pp.1230–1239, Sept. 1997.

[6] O.N. Granichin, "Estimating the parameters of linear regression in an arbitrary noise," *Automat. and Remote Contr.*, v. 63, pp. 25–35, 2002.

[7] O.N. Granichin and B.T. Polyak *Randomized Algorithms of an Estimation and Optimization Under Almost Arbitrary Noises.* M.: Nauka, 2003.

[8] O.N. Granichin, "A Stochastic Recursive Procedure with Dependent Noises in the Observation that Uses Sample Perturbations in the Input," *Vestnik Leningrad Univ., Math.*, vol. 22, No. 1(4), pp.27–31, 1989.

[9] B.T. Polyak and A.B. Tsybakov, "Optimal Orders of Accuracy for Search Algorithms of Stochastic Optimization," *Problems Inform. Transmission*, vol. 26, No. 2, pp.126–133, 1990.

[10] O.N. Granichin, "Stochastic Approximation with Sample Perturbations in the Input," *Automat. Remote Control*, vol. 53, No. 2, pp.232–237, 1992.

[11] J.C. Spall, "Multivariate Stochastic Approximation Using a Simultaneous Perturbation Gradient Approximation," *IEEE Trans. Automat. Contr.,* vol. 37, pp. 332–341, 1992.

[12] J.C. Spall, “A One-Measurement Form of Simultaneous Perturbation Stochastic Approximation," *Automatica*, vol. 33, pp. 109–112, 1997.

[13] H.J. Kushner and D.S. Clark, *Stochastic Approximation Methods for Constrain and Unconstraint System.* Berlin, Germany: Springer-Verlag, 1978.

[14] H.J Kushner. and G.G. Yin, *Stochastic Approximation Algorithms and Applications.* New York: Springer–Verlag, 1997.

[15] J. Alspector, R. Meir, A. Jayakumar and D. Lippe, "A Parallel Gradient Descent Method for Learning in Analog VLSI Neural Networks," in S.J. Hanson, J.D. Cowan and C. Lee *Advances in Neural Information Processing Systems.* San Mateo, CA: Morgan Kaufmann Publishers, Inc. pp.834–844, 1993.

[16] Y. Maeda, and Y. Kanata, “Learning Rules for Recurrent Neural Networks Using Perturbation and their Application to Neuron-control," *Trans. of IEE of Japan*, vol. 113-C, pp. 402–408, 1993.

[17] B.T. Polyak and A.B. Tsybakov, “On stochastic approximation with arbitrary noise (the KW case)," in Topics in Nonparametric Estimation, R.Z. Khasminskii ed., *Advances in Soviet Mathematics*, Amer. Math. Soc., Providence, No. 12, pp. 107–113, 1992.

[18] O.N. Granichin, "Stochastic Approximation under Dependent Noises, Detecting Signals and Adaptive Control," in *Approximation, Probability, and Related Fields*, pp. 247–271, Plenum, 1994.

[19] H.F. Chen, T.E. Duncan and B. Pasik-Duncan, “A Kiefer-Wolfowitz Algorithm with Randomized Differences," *IEEE Trans. Automat. Contr.*, vol. 44, pp. 442–453, Mar. 1999.

[20] O.N. Granichin, “Randomized algorithms for stochastic approximation under arbitrary disturbances," *Automat. and Remote Contr.*, v. 63, pp. 209–219, 2002.

[21] O.N. Granichin, “Optimal convergence rate of the randomized algorithms of stochastic approximation in arbitrary noise," *Automat. and Remote Contr.*, v. 64, pp. 252–262, 2003.

[22] Q. Zhang, “Optimal filtering of discrete-time hybrid systems," *J. Optim. Theory App.*, vol. 100, pp. 123–144, 1999.

[23] Q. Zhang, "Hybrid filtering for linear systems with non-Gaussian disturbances," *IEEE Trans. Automat. Contr.*, vol. 45, pp. 50–61, 2000.

[24] O.N. Granichin, "Nonminimax filtering in unknown irregular constrained observation noise," *Automat. and Remote Contr.*, v. 63, pp. 1482–1488, 2002.

[25] B.T. Polyak and Ya.Z. Tsypkin, "Adaptive estimation algorithms (convergence, optimality, stability)," *Automat. Remote Contr.*, vol. 40, pp. 378–389, 1979.

[26] L. Guo, L. Ljung and G-J. Wang, "Necessary and Sufficient Conditions for Stability of LMS," *IEEE Trans. on Automat. Contr.*, vol. 42, pp. 761–770, 1997.

[27] Y. Maeda, A. Nakazawa and K. Yakichi, "Hardware Implementation of a Pulse Density Neural Network Using Simultaneous Perturbation Learning Rule," *Analog Integrated Circuits and Signal Processing*, vol. 18, pp. 153–162, 1999.

[28] P.W. Shor, "Quantum computing," *Proc. of the 9th Int. Math. Congress,* Berlin, 1998, www.math.nine.edu/documenta/xvol-icm/Fields/Fields.html

[29] C.H. Bennet and P.W. Shor, "Quantum information theory," *IEEE Trans. Inform. Theory*, vol. 44, pp. 2724–2742, Oct. 1998.

[30] J. Preskill "Quantum Information and Computation." [Online]. Available: http://www.theory.caltech.edu/~ preskill/ph229

[31] H. Robbins and D. Siegmund, "A convergence theorem for nonnegative almost super-martingales and some applications," in *Optimizing Methods in Statistics*, J.S. Rustagi ed., Academic Press, New York, pp. 233–257, 1971.

[32] B.T. Polyak, *Introduction to Optimization.* New York: Optim.Software, 1987.

[33] B.T. Polyak, "Convergence and rate of convergence of recursive stochastic algorithms. Linear case," *Automat. Remote Contr.*, vol. 38, pp. 537–542, 1977.