Рассматриваемтся задача выпуклой композитной стохастической оптимизации:

$$\min_{x \in X} \left[F(x) = \mathbb{E}\{\Phi(x, \omega)\} + \psi(x) \right], \tag{1}$$

где X- компактное выпуклое подмножество конечномерного действительного векторного пространства с нормой $||\cdot||$, ω случайная величина на пространстве Ω с распределением P, функция ψ выпукла и непрерывна, и функция $\Phi : X \times \Omega \to R$. Также предполагаем, что мат ожидание

$$\phi(x) := \mathbb{E}\{\Phi(x, \omega)\} = \int_\Omega \Phi(x, \omega) dP(\omega)$$

конечное для всех $x \in X$, и выпукла и дифференцируема как функция по x. Тогда задача имеет решение $F_* = \min_{x \in X} F(x)$. Также предполагаем, что есть оракул, который по точкам $(x, \omega) \in X \times \Omega$ возвращает стохастический грагиент, такой что вектор $G(x, \omega)$

$$\mathbb{E}\{G(x, \omega)\} = \nabla \phi(x), \qquad \mathbb{E}\{||G(x, \omega) - \nabla \phi(x)||_*^2\} \leq \sigma^2, \forall x \in X$$

Определение 1. *Дифференцируемая функция $\varphi : \mathbb{R}^d \to \mathbb{R}$ является сильно выпуклой функцией с константой μ, если для любых $x, y \in \mathbb{R}^d$ выполнено*

$$\varphi(x) \geq \varphi(y) + \langle \nabla \varphi(y), x - y \rangle + \frac{\mu}{2} \|x - y\|^2. \tag{2}$$

1 Робастный Стохастический Зеркальный Спуск (RSMD)

Проксимальный оператор имеет вид

$$\begin{aligned} Prox_{\beta,x}(\xi) &:= \arg\min_{z \in X} \{\langle \xi, z \rangle + \psi(z) + \beta V_x(z)\} = \\ &= \arg\min_{z \in X} \{\langle \xi - \beta \nu'(x), z \rangle + \psi(z) + \beta \nu(z)\}, \end{aligned}$$

где $\beta > 0$ - настраиваемый параметр.

Метод RSMD имеет вид

$$x_i = Prox_{\beta_{i-1}, x_{i-1}}(y_i), x_0 \in X,$$

$$y_i = \begin{cases} G(x_{i-1}, \omega_i), & \text{если } ||G(x_{i-1}, \omega_i) - g(\overline{x})||_* \leq L||\overline{x} - x_{i-1}|| + \lambda + u\sigma, \\ g(\overline{x}), & \text{иначе} \end{cases}$$

Обозначим аппроксимацию решения как

$$\hat{x}_N = \left[\sum_{i=1}^N \beta_{i-1}^{-1} \right]^{-1} \sum_{i=1}^N \beta_{i-1}^{-1} x_i$$

Лемма 1. *Пусть функции ψ и ϕ удовлетворяют предположениям и пусть $x_0, \ldots, x_N$ - произвольные точки множества X. Определим*

$$\varepsilon_{i+1}(z) := \langle \nabla \phi(x_i), x_{i+1} - z \rangle + \langle \psi'(x_{i+1}), x_{i+1} - z \rangle + L V_{x_i}(x_{i+1})$$

Тогда для любого $z \in X$ справедливо

$$F(x_{i+1}) - F(z) \leq \varepsilon_{i+1}(z),$$

Если $\hat{x}_N = \frac{1}{N}\sum_{i=1}^{N} x_i$, то

$$F(\hat{x}_N) - F(z) \leq N^{-1}\sum_{i=1}^{N}[F(x_i) - F(z)] \leq N^{-1}\sum_{i=0}^{N-1}\varepsilon_{i+1}(z)$$

Доказательство. Используя свойства $V_x(z) \geq \frac{1}{2}||x-z||^2$, выпуклость функций ϕ и ψ и условие Липшица на $\nabla\phi$, получаем для любого $z \in X$

$$\begin{aligned} F(x_i) - F(z) &= [\phi(x_{i+1}) - \phi(z)] + [\psi(x_{i+1}) - \psi(z)] = \\ &= [\phi(x_{i+1}) - \phi(x_i)] = [\phi(x_i) - \phi(z)] + [\psi(x_{i+1}) - \psi(z)] \leq \\ &\leq \langle \nabla\phi(x_i), x_{i+1} - z\rangle + \langle \psi'(x_{i+1}), x_{i+1} - z\rangle + LV_{x_i}(x_{i+1}) = \varepsilon_{i+1}(z) \end{aligned}$$

Суммируя по всем i от 0 до $N-1$ и используя неравенство Йенсена, получаем результат леммы. $\square$

Лемма 2. *Пусть предположения выше выполнены и пусть x_i, y_i удовлетворяют RSMD рекурсии. Тогда*

$$(a)||\xi_i||_* \leq 2(M + u\sigma) + \lambda,$$

$$(b)||\mathbb{E}_{x_{i-1}}\{\xi_i\}||_* \leq (M + u\sigma)\left(\frac{\sigma}{\lambda}\right)^2 + \frac{\sigma^2}{\lambda},$$

$$(c)(\mathbb{E}_{x_{i-1}}\{||\xi_i||_*^2\})^{1/2} \leq \sigma + (M + u\sigma)\frac{\sigma}{\lambda},$$

Доказательство. Обозначим $\chi_i = 1_{||G(x_{i-1},\omega_i) - g(\overline{x})||_* > L||x_{i-1} - \overline{x}|| + \lambda + u\sigma}$. Добавим еще обозначение $\chi_i \leq \eta_i := 1_{||G(x_{i-1},\omega_i) - g(\overline{x})||_* > \lambda}$ Имеем

$$\begin{aligned} \xi_i &= y_i - \nabla\phi(x_{i-1}) = [G(x_{i-1},\omega_i) - \nabla\phi(x_{i-1})](1 - \chi_i) + [g(\overline{x}) - \nabla\phi(x_{i-1})]\chi_i = \\ &= [G(x_{i-1},\omega_i) - g(\overline{x})](1 - \chi_i) + [g(\overline{x}) - \nabla\phi(x_{i-1})] = \\ &= [G(x_{i-1},\omega_i) - \nabla\phi(x_{i-1})] + [g(\overline{x}) - G(x_{i-1},\omega_i)]\chi_i \end{aligned}$$

Тогда получаем, что

$$||\xi_i||_* \leq ||[G(x_{i-1},\omega_i) - g(\overline{x})](1 - \chi_i)||_* + ||g(\overline{x}) - \nabla\phi(x_{i-1})||_* \leq 2(M + u\sigma) + \lambda$$

Так как $\mathbb{E}_{x_{i-1}}\{G(x_{i-1},\omega_i)\} = \nabla\phi(x_{i-1})$, то

$$\begin{aligned} ||\mathbb{E}_{x_{i-1}}\{\xi_i\}||_* &= ||\mathbb{E}_{x_{i-1}}\{[G(x_{i-1},\omega_i) - \nabla\phi(x_{i-1})) - (g(\overline{x}) - \nabla\phi(x_{i-1}))]\chi_i]\}||_* \leq \\ &\leq \mathbb{E}_{x_{i-1}}\{||[G(x_{i-1},\omega_i) - \nabla\phi(x_{i-1}))||_* + ||(g(\overline{x}) - \nabla\phi(x_{i-1}))]\chi_i]\}||_* \leq \\ &\leq \mathbb{E}_{x_{i-1}}\{||[G(x_{i-1},\omega_i) - \nabla\phi(x_{i-1}))||_*\eta_i\} + (M + u\sigma)\mathbb{E}_{x_{i-1}}\{\eta_i\} \leq \\ &\leq \frac{\sigma^2}{\lambda} + (M + u\sigma)\left(\frac{\sigma}{\lambda}\right)^2. \end{aligned}$$

Теперь заметим, что

$$||\xi_i||_* \leq ||G(x_{i-1},\omega_i) - \nabla\phi(x_{i-1})||_*(1 - \chi_i) + ||g(\overline{x}) - \nabla\phi(x_{i-1})||_*\chi_i,$$

и

$$\begin{aligned}\mathbb{E}_{x_{i-1}}\{||\xi_i||_*^2\}^{1/2} &\leq \mathbb{E}_{x_{i-1}}\{||G(x_{i-1},\omega_i)-\nabla\phi(x_{i-1})||_*^2\}^{1/2}+\mathbb{E}_{x_{i-1}}\{||g(\overline{x})-\nabla\phi(x_{i-1})||_*^2\chi_i\}^{1/2}\leq\\ &\leq \sigma+(M+u\sigma)\mathbb{E}_{x_{i-1}}\{\chi_i\}^{1/2}\leq\sigma+(M+u\sigma)\mathbb{E}_{x_{i-1}}\{\eta_i\}^{1/2}\leq\\ &\leq \sigma+(M+u\sigma)\frac{\sigma}{\lambda}.\end{aligned}$$

□

Лемма 3. *Пусть выполнены предположения выше и x_i, y_i удовлетворяют рекурсии RSMD. (i) Если $\tau\leq N/u^2$ и $\lambda=\max\{\sigma\sqrt{\frac{N}{\tau}},M\}+u\sigma$ тогда, для любого $z\in X$,*

$$\mathsf{P}\{\sum_{i=1}^{N}\langle\xi_i,z-x_{i-1}\rangle\geq 16R\max\{\sigma\sqrt{N\tau},M\tau\}\}\leq e^{-\tau},$$

и

$$\mathsf{P}\{\sum_{i=1}^{N}||\xi_i||_*^2\geq 40\max\{N\sigma^2,M^2\tau\}\}\leq e^{-\tau}.$$

(ii) Если $N\geq u^2$ и $\lambda=\max\{\sigma\sqrt{N},M\}+u\sigma$ тогда для любой $z\in X$,

$$\mathsf{P}\{\sum_{i=1}^{N}\langle\xi_i,z-x_{i-1}\rangle\geq 8(1+\tau)R\max\{\sigma\sqrt{N},M\}\}\leq e^{-\tau},$$

и

$$\mathsf{P}\{\sum_{i=1}^{N}||\xi||_*^2\geq 8(2+3\tau)\max\{N\sigma^2,M^2\}\}\leq e^{-\tau}.$$

Теорема 1. *Пусть $\beta_i\geq 2L$ для всех $i=0,1,\ldots,$ и пусть $\hat{x}_N$ определен выше, где x_i получен RSMD для произвольного y_i. Тогда для любого $z\in X$ имеем*

$$\begin{aligned}\left[\sum_{i=1}^{N}\beta_{i-1}^{-1}\right][F(\hat{x}_N)-F(z)] &\leq \sum_{i=1}^{N}\beta_{i-1}^{-1}[F(x_i)-F(z)]\leq\varepsilon(x^N,z)\\ &\leq V_{x_0}(z)+\sum_{i=1}^{N}\left[\frac{\langle\xi_i,z-x_{i-1}\rangle}{\beta_{i-1}}+\frac{||\xi_i||_*^2}{\beta_{i-1}^2}\right]\\ &\leq 2V_{x_0}(z)+\sum_{i=1}^{N}\left[\frac{\langle\xi_i,z_{i-1}-x_{i-1}\rangle}{\beta_{i-1}}+\frac{3}{2}\frac{||\xi_i||_*^2}{\beta_{i-1}^2}\right],\end{aligned}$$

где z_i - это случайный вектор из X, зависящий от $x_0,\xi_1,\ldots,\xi_i$

Доказательство.

$$\langle y_{i+1}+\psi'(x_{i+1})+\beta_i[\nu'(x_{i+1})-\nu'(x_i)],z-x_{i+1}\rangle\geq 0,\forall z\in X,$$

или в другом виде

$$\begin{aligned}\langle y_i+\psi'(x_{i+1}),x_{i+1}-z\rangle &\leq \beta_i\rangle[\nu'(x_{i+1})-\nu'(x_i)],z-x_{i+1}\rangle=\langle\beta_iV'_{x_i}(x_{i+1}),z-x_{i+1}\rangle=\\ &= \beta_i[V_[x_i(z)-V_{x_{i+1}}(z)-V_{x_i}(x_{i+1})],\forall z\in X,\end{aligned}$$

Используя определение $\xi_i=y_i-\nabla\phi(x_{i-1})$ получаем

$$\langle\nabla\phi(x_i),x_{i+1}-z\rangle+\langle\psi'(x_{i+1}),x_{i+1}-z\rangle\leq\beta_i[V_[x_i(z)-V_{x_{i+1}}(z)-V_{x_i}(x_{i+1})]-\langle\xi_{i+1},x_{i+1}-z\rangle$$

Используя Лемму 1 и условие $\beta_i \geq 2L$, получаем

$$F(x_{i+1}) - F(z) \leq \varepsilon_{i+1}(z) \leq \langle \nabla\phi(x_i), x_{i+1} - z\rangle + \langle \psi'(x_{i+1}), x_{i+1} - z\rangle + \frac{\beta_i}{2} V_{x_i}(x_{i+1})$$

Заметим, что

$$\varepsilon_{i+1}(z) \leq \beta_i [V_[x_i(z) - V_{x_{i+1}}(z) - \frac{1}{2} V_{x_i}(x_{i+1})]] - \langle \xi_{i+1}, x_{i+1} - z\rangle$$

С другой стороны, в силу сильной выпуклости $V_x()$

$$\begin{aligned}\langle \xi_{i+1}, z - x_{i+1}\rangle - \frac{\beta_i}{2} V_{x_i}(x_{i+1}) &= \langle \xi_i, z - x_i\rangle + \langle \xi_{i+1}, x_i - x_{i+1}\rangle - \frac{\beta_i}{2} V_{x_i}(x_{i+1}) \\ &\leq \langle \xi_{i+1}, z - x_i\rangle + \frac{||\xi_{i+1}||_*^2}{\beta_i}\end{aligned}$$

В итоге получаем, что

$$F(x_{i+1}) - F(z) \leq \varepsilon_{i+1}(z) \leq \beta_i [V_{x_i}(z) - V_{x_{i+1}}(z)] - \langle \xi_{i+1}, x_i - z\rangle + \frac{||\xi_{i+1}||_*^2}{\beta_i}$$

Разделив на β_i и просуммировав для всех i от 0 до $N-1$ получаем основной результат утверждения. □

Теорема 2. *Пусть $\beta_i = \overline{\beta} \geq 2L$ для всех $i = 0, 1, \ldots,$ и пусть $1 \leq \tau \leq N/u^2$,*

$$\lambda = \max\{\sigma\sqrt{\frac{N}{\tau}}, M\} + u\sigma$$

Пусть $\hat{x}_N$ аппроксимация решения, где x_i получен из RSMD. Тогда есть случайное событие $\mathcal{A}_N \subset \Omega^{\otimes N}$ с вероятностью по крайней мере $1 - 2e^{-\tau}$ такое, что для всех $\omega^N \in \mathcal{A}_N$ справедливо следующее неравенство

$$\begin{aligned}F(\hat{x}_N) - F_* &\leq \frac{\overline{\beta}}{N} \sup_{z \in X} \varepsilon(x^N, z) \leq \\ &\leq \frac{C}{N}\left(\overline{\beta} R^2 \Theta + R \max\{\sigma\sqrt{\tau N}, M\tau\} + \overline{\beta}^{-1} \max\{N\sigma^2, M^2\tau\}\right)\end{aligned}$$

В частности, выбирая $\overline{\beta} = \max\{2L, \frac{\sigma\sqrt{N}}{R\sqrt{\Theta}}$, имеем для всех $\omega^N \in \mathcal{A}_N$,

$$F(\hat{x}_N) - F_* \leq \max\{C_1 \frac{LR^2[r \vee \Theta]}{N}, C_2 \sqrt{\frac{\tau \vee \theta}{N}}$$

Доказательство. С условием $\tau \leq N/u^2$ имеем, что с вероятностью по крайней мере $1 - 2e^{-\tau}$,

$$\begin{aligned}\sum_{i=1}^{N} \langle \xi_i, z_{i-1} - x_{i-1}\rangle &\leq 16R \max\{\sigma\sqrt{N\tau}, M\tau\}, \\ \sum_{i=1}^{N} ||\xi_i||_*^2 &\leq 40 \max\{N\sigma^2, M^2\tau\}\end{aligned}$$

Подставляя эту оценку в утверждение 1 получаем, что с вероятностью по крайней мере $1 - 2e^{-\tau}$ имеет место следующее

$$\begin{aligned}\overline{\beta} \sup_{z \in X} \varepsilon(x^N, z) &\leq 2\overline{\beta} V_{x_0}(z) + \sum_{i=1}^{N} [\langle \xi_i, z_{i-1} - x_{i-1}\rangle + \frac{3}{2}\overline{\beta}^{-1} ||\xi_i||_*^2] \leq \\ &\leq 2\overline{\beta} R^2 \Theta + 16R \max\{\sigma\sqrt{N\tau}, M\tau\} + 60\overline{\beta}^{-1} \max\{N\sigma^2, M^2\tau\}.\end{aligned}$$

Далее, выбрав $\overline{\beta} = \max\{2L, \frac{\sigma}{R}\sqrt{\frac{N}{\Theta}}\}$ получаем

$$
\begin{aligned}
N[F(\hat{x}_N) - F(z)] &\leq \max\{4LR^2\Theta, 2\sigma R\sqrt{N\Theta}\} + 16R\max\{\sigma\sqrt{N\tau}, M\tau\} + 60\max\{LR^2\tau/2, \sigma R\sqrt{N\Theta}\} \leq \\
&\leq \max\{46LR^2\tau, 4LR^2\Theta, 62\sigma R\sqrt{N\Theta}, 16\sigma R\sqrt{N\tau}\}
\end{aligned}
$$

$\square$