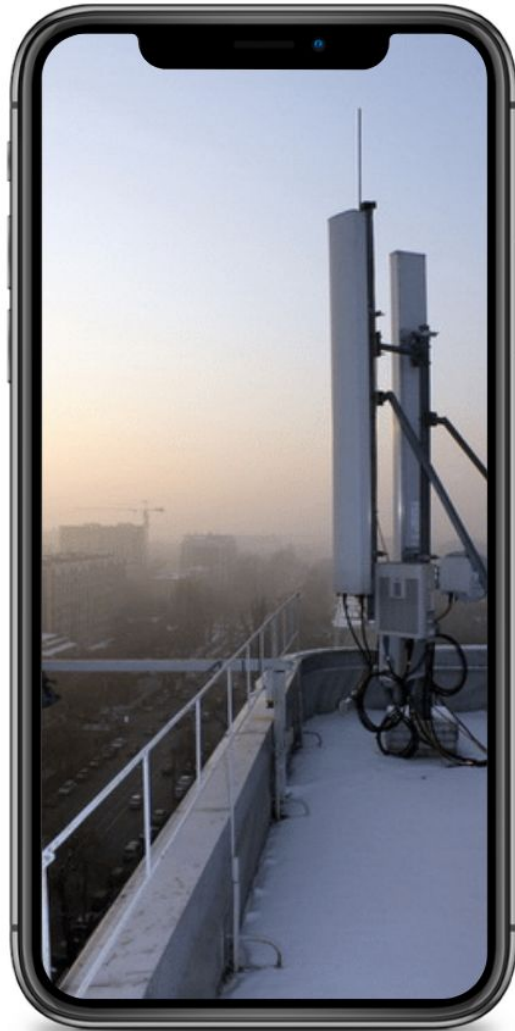




Online-learning in digital pre-distorter modelling

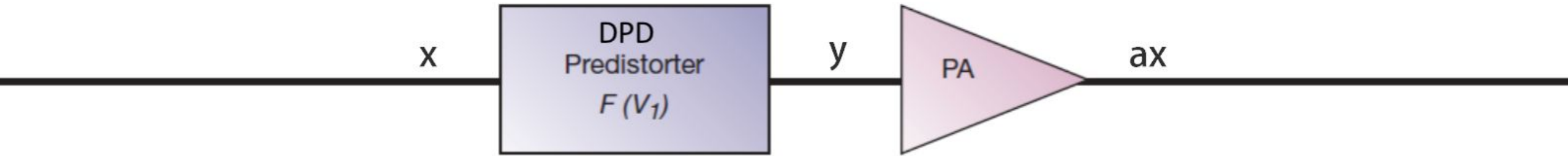




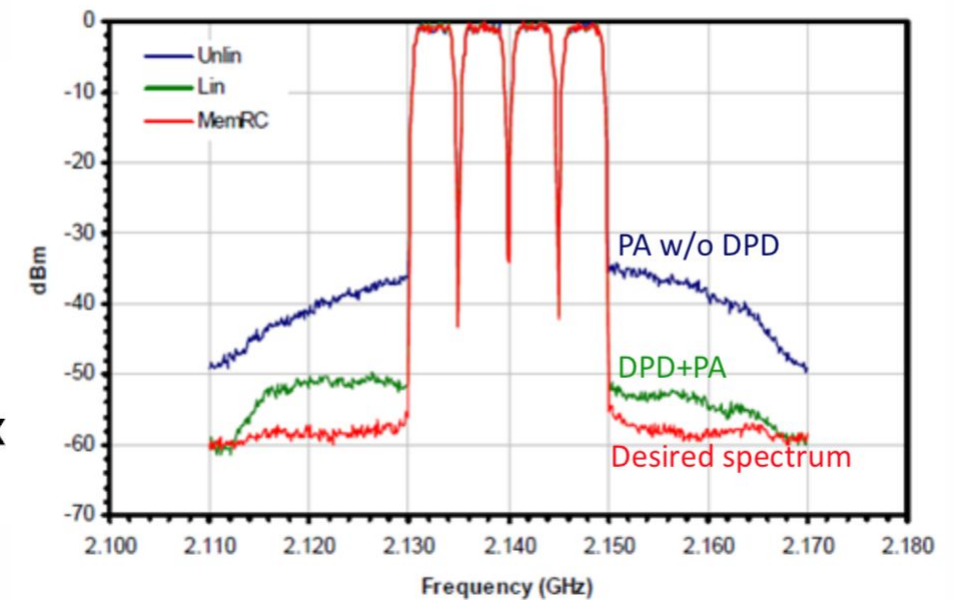
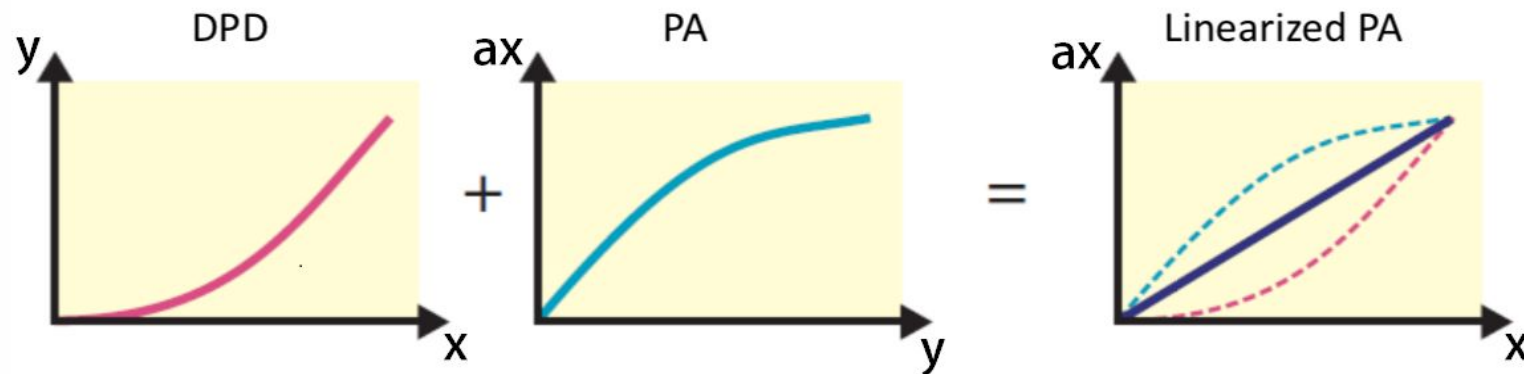
Физическая модель



БОЛЬШИЕ ВЫЗОВЫ
НАУЧНО-ТЕХНОЛОГИЧЕСКАЯ
ПРОЕКТНАЯ ПРОГРАММА



Example of Signal Spectrum with and w/o DPD





Математическая модель



БОЛЬШИЕ ВЫЗОВЫ
НАУЧНО-ТЕХНОЛОГИЧЕСКАЯ
ПРОЕКТНАЯ ПРОГРАММА

Идеальная задача

$$\mathbb{E}_{(x,y)}(y - f(x, \omega))^2 \rightarrow \min_{\omega}$$

$f(x, \omega)$ - модель

Винера-Гаммерштейна

Настоящая задача

$$\frac{1}{n} \sum_{k=1}^n (x_k - f(x_k, \omega))^2 \rightarrow \min_{\omega}$$

Метрика качества

$$NMSE(y, \bar{y}) = 10 \log_{10} \left(\frac{\sum_{k=1}^m |y(k) - \bar{y}(k)|^2}{\sum_{k=1}^m |x(k)|^2} \right) \quad \text{Db}$$



Теория оптимизации



БОЛЬШИЕ ВЫЗОВЫ
НАУЧНО-ТЕХНОЛОГИЧЕСКАЯ
ПРОЕКТНАЯ ПРОГРАММА

$$x_t = x_{t-1} - \eta \nabla f(x_t) \quad \text{Gradient Descent}$$

$$x_t = x_{t-1} - \eta \nabla f_i(x_t), i \sim \mathcal{U}\{1, \dots, n\} \quad \text{Stochastic Gradient Descent}$$

$$x_t = x_{t-1} - \eta \cdot \frac{1}{b} \sum_{i=1}^b \nabla f_i(x_t) \quad \text{Mini-batch Stochastic Gradient Descent}$$

$$x_t = x_{t-1} - \frac{\eta}{\sqrt{\sum_{i=1}^t \nabla f_i^2(x_t)}} \nabla f_i(x_t) \quad \text{AdaGrad}$$

$$x_t = x_{t-1} - \frac{\eta}{\sqrt{\sum_{i=1}^t \nabla f_i^2(x_t)}} \nabla f_i(x_t) + \beta(x_{t-1} - x_{t-2}) \quad \text{AdaGrad + momentum}$$



Метод Adam



БОЛЬШИЕ ВЫЗОВЫ
НАУЧНО-ТЕХНОЛОГИЧЕСКАЯ
ПРОЕКТНАЯ ПРОГРАММА

arXiv:2103.10552, Maslovskiy-Pasechnyuk 2021

Method	Setup	dB	
		$t = 0$ sec.	$t = 300$ sec.
ASGD	(128, 10.0)	-15.616	-28.976
Adadelta	(2048, 10.0)		-32.697
Adagrad	(2048, 0.01)		-36.608
Adam	(2048, 0.001)		-38.129
Adamax	(2048, 0.01)		-37.934
RMSprop	(2048, 0.001)		-36.499
SGD	(128, 10.0)		-34.061
FastAdaptive			-36.273

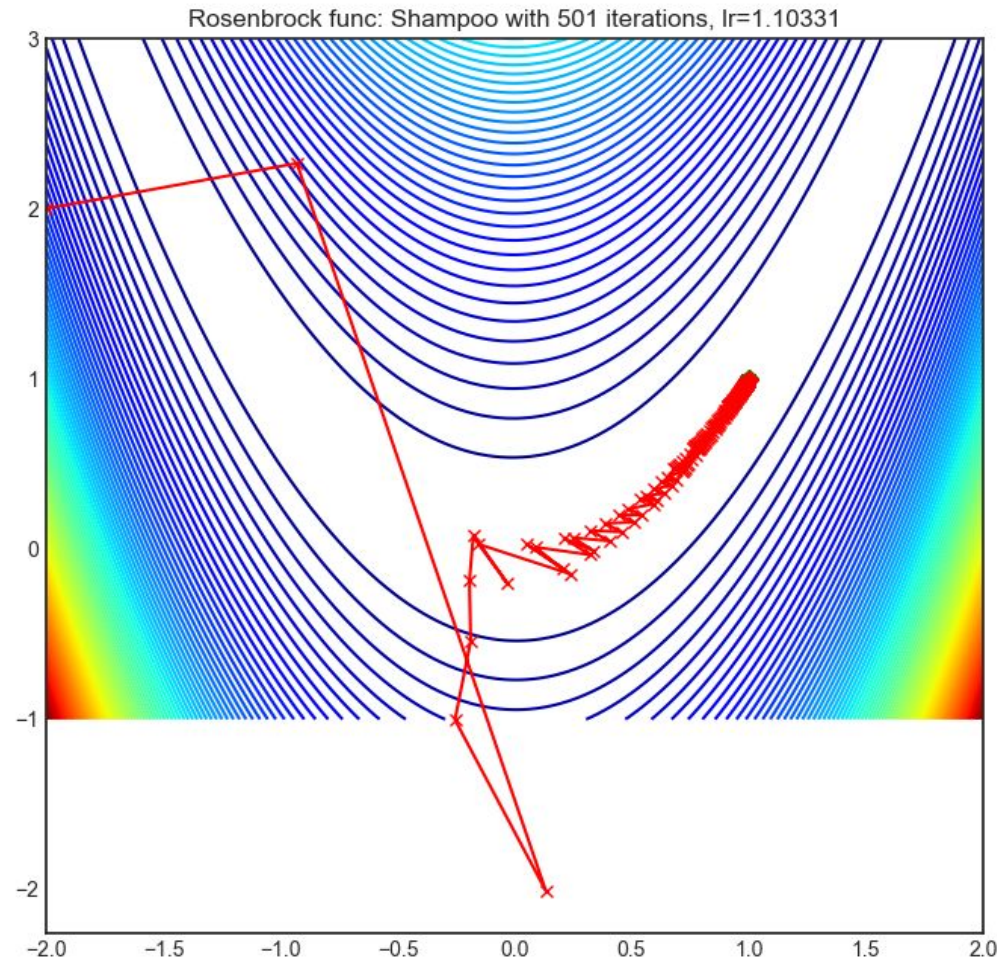
До нас в научных работах использовали Adam как наиболее эффективный метод. Мы стремились показать, что можно сделать лучше



Метод Shampoo



БОЛЬШИЕ ВЫЗОВЫ
НАУЧНО-ТЕХНОЛОГИЧЕСКАЯ
ПРОЕКТНАЯ ПРОГРАММА



arXiv:1802.09568, Gupta 2018

Shampoo - адаптивный
стохастический метод
с предобуславливанием

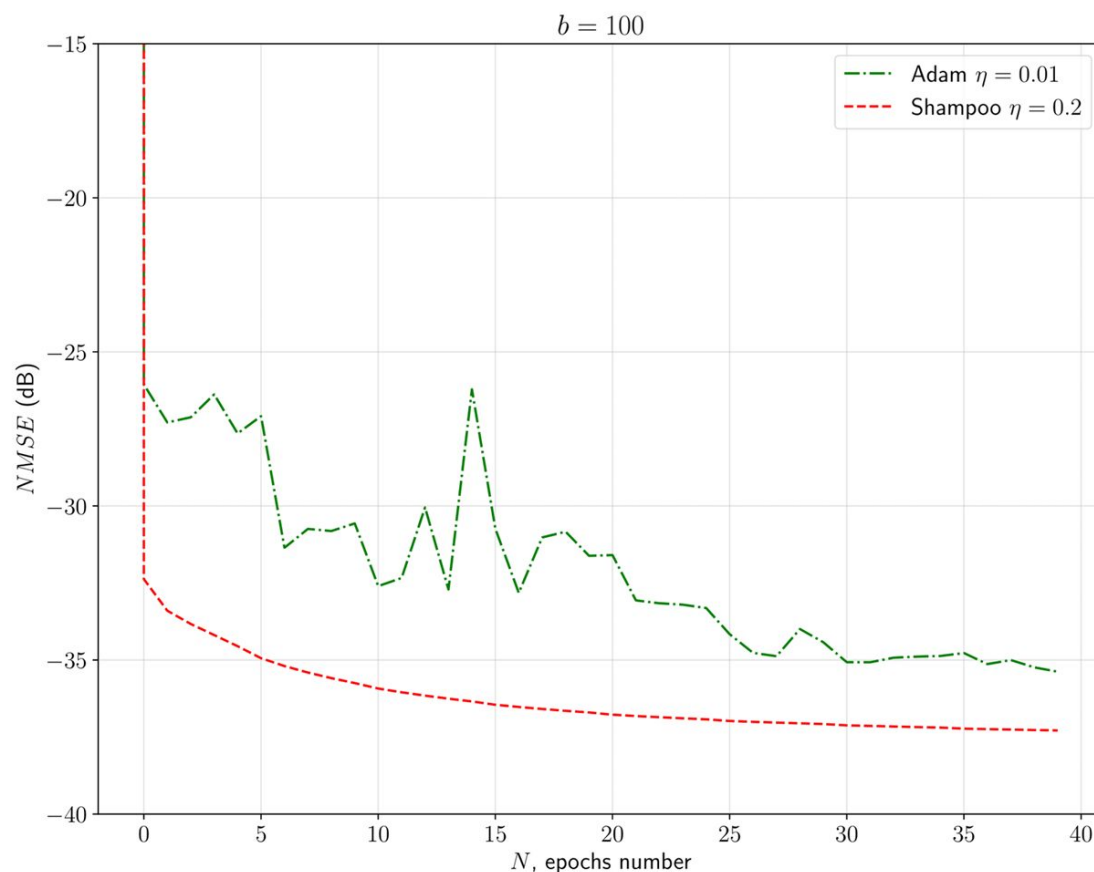


Метод Shampoo



БОЛЬШИЕ ВЫЗОВЫ

НАУЧНО-ТЕХНОЛОГИЧЕСКАЯ
ПРОЕКТНАЯ ПРОГРАММА



Вывод: Использование оптимизатора Shampoo даёт улучшение NMSE на **7%** по сравнению с Adam



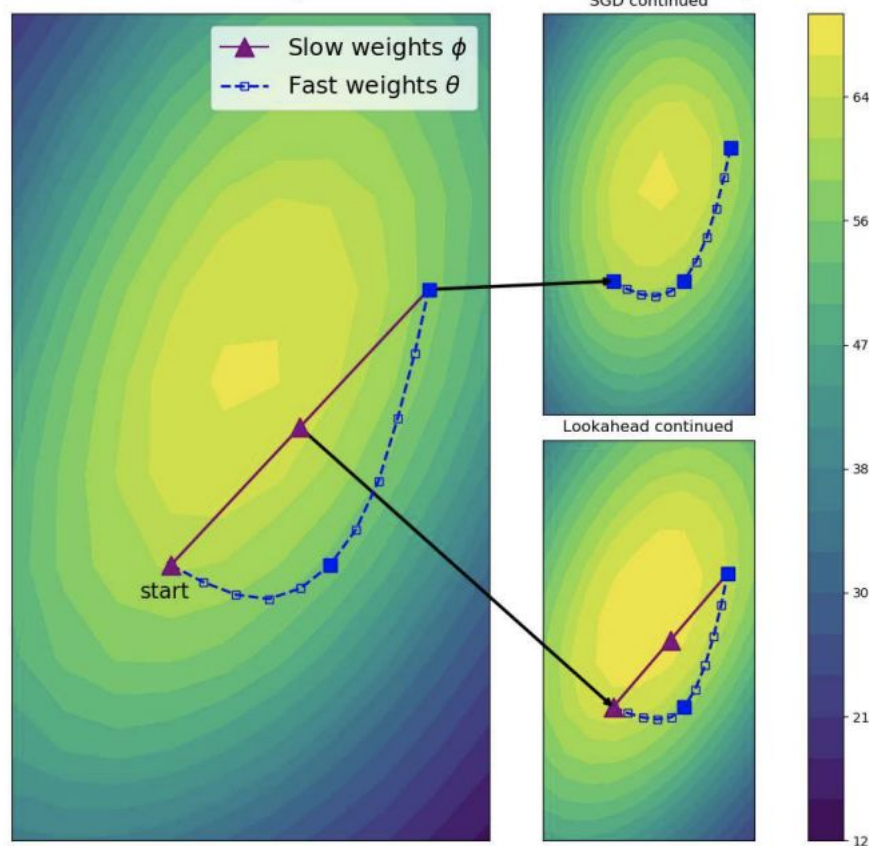
Оболочка Lookahead



БОЛЬШИЕ ВЫЗОВЫ

НАУЧНО-ТЕХНОЛОГИЧЕСКАЯ
ПРОЕКТНАЯ ПРОГРАММА

CIFAR-100 accuracy surface with Lookahead interpolation



arXiv:1907.08610, Zhang 2019

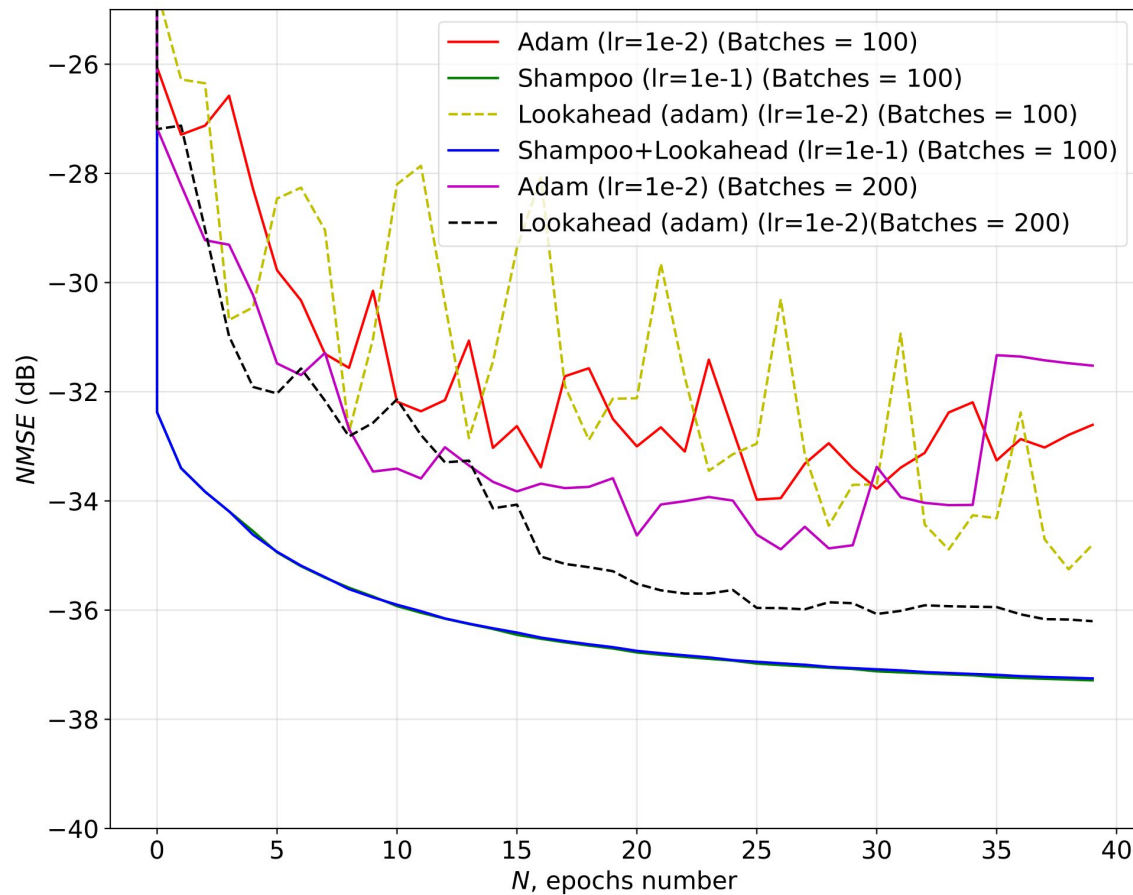
Оболочка, поддерживающая
“быстрые” и “медленные” веса,
и корректирующая “медленные”
в направлении “быстрых”



Оболочка Lookahead



БОЛЬШИЕ ВЫЗОВЫ
НАУЧНО-ТЕХНОЛОГИЧЕСКАЯ
ПРОЕКТНАЯ ПРОГРАММА



Вывод:

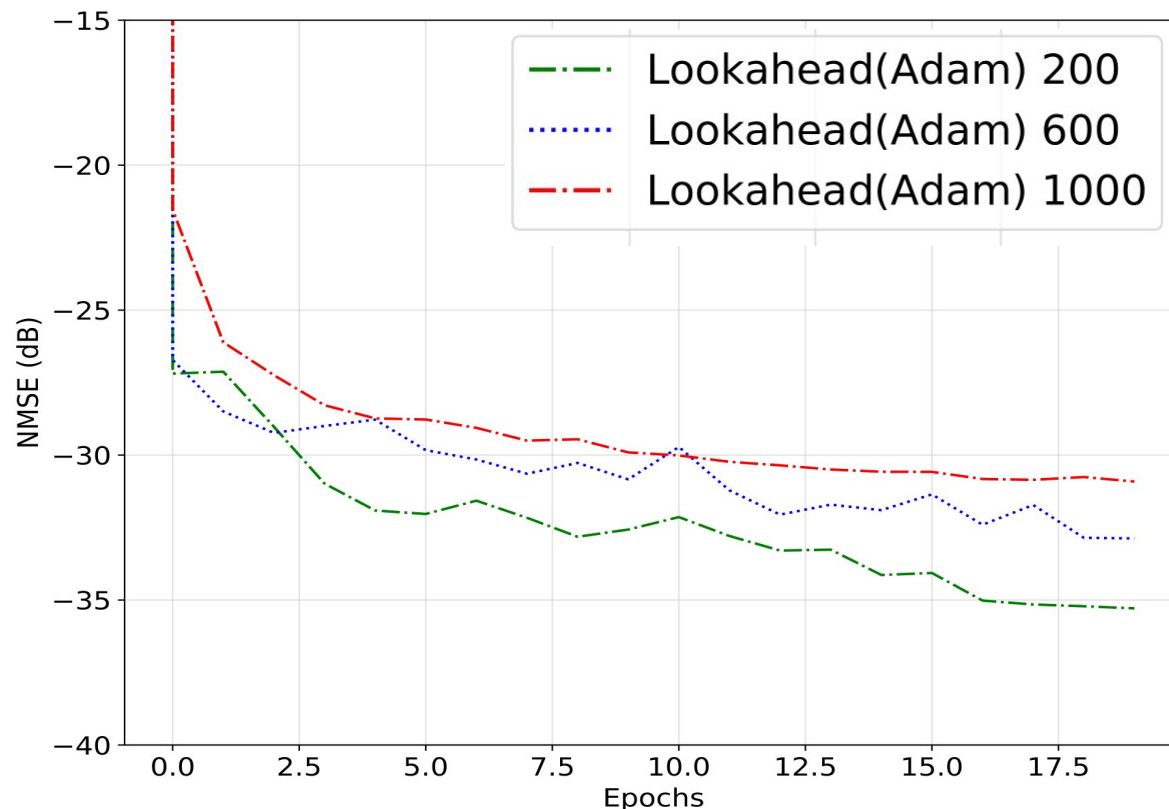
1. Оболочка Lookahead улучшает NMSE метода Adam на **3%**
2. Lookahead не улучшает Shampoo
3. Shampoo работает стабильнее



Размер батча



БОЛЬШИЕ ВЫЗОВЫ
НАУЧНО-ТЕХНОЛОГИЧЕСКАЯ
ПРОЕКТНАЯ ПРОГРАММА



Вывод:

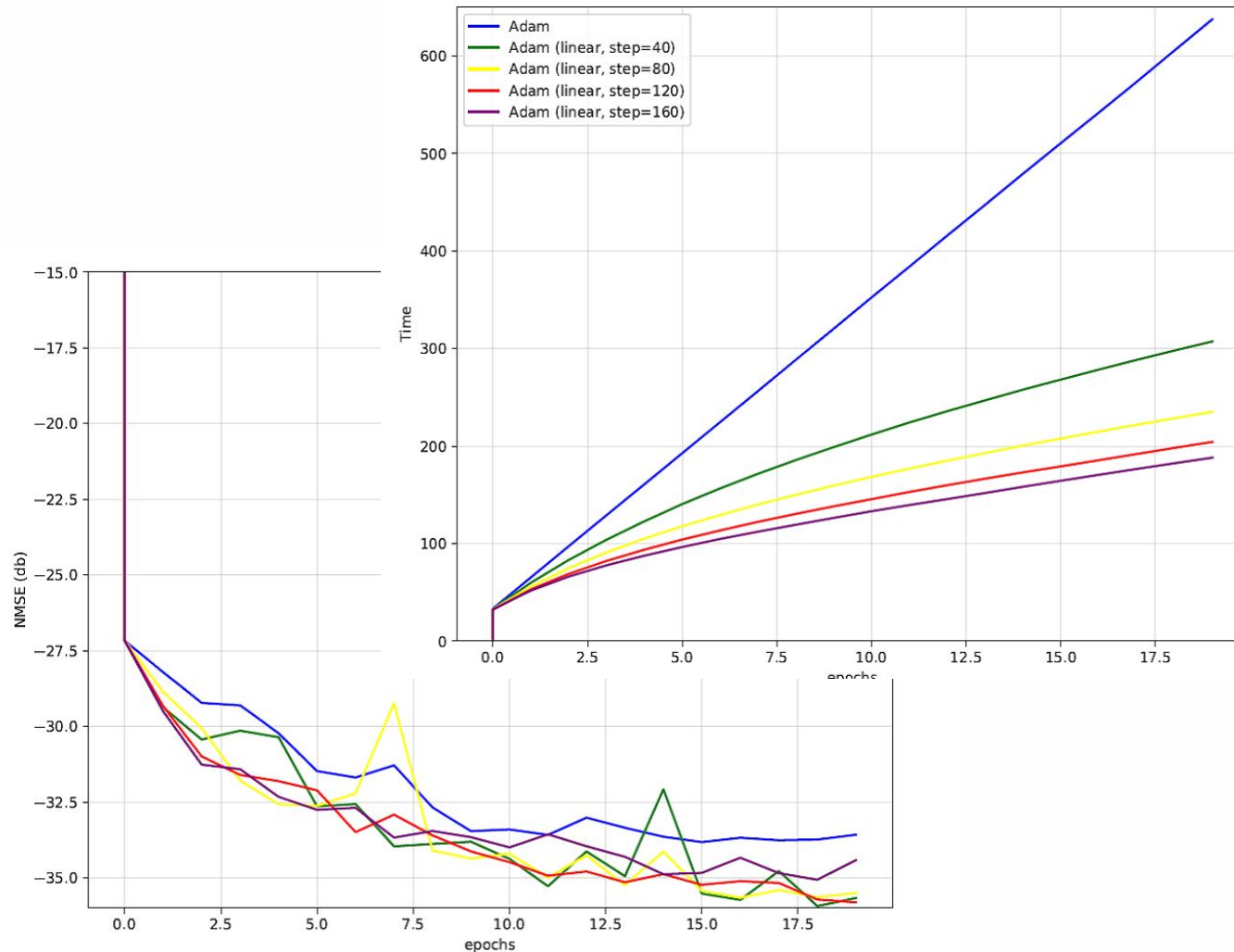
Качество тем лучше, чем меньше размер батча. Однако время обучения с уменьшением батча растёт



Размер батча $\sim t$



БОЛЬШИЕ ВЫЗОВЫ
НАУЧНО-ТЕХНОЛОГИЧЕСКАЯ
ПРОЕКТНАЯ ПРОГРАММА



Вывод:

1. Новая политика мини-батчинга улучшает NMSE на **7%**
2. Время обучения сократилось в **3** раза

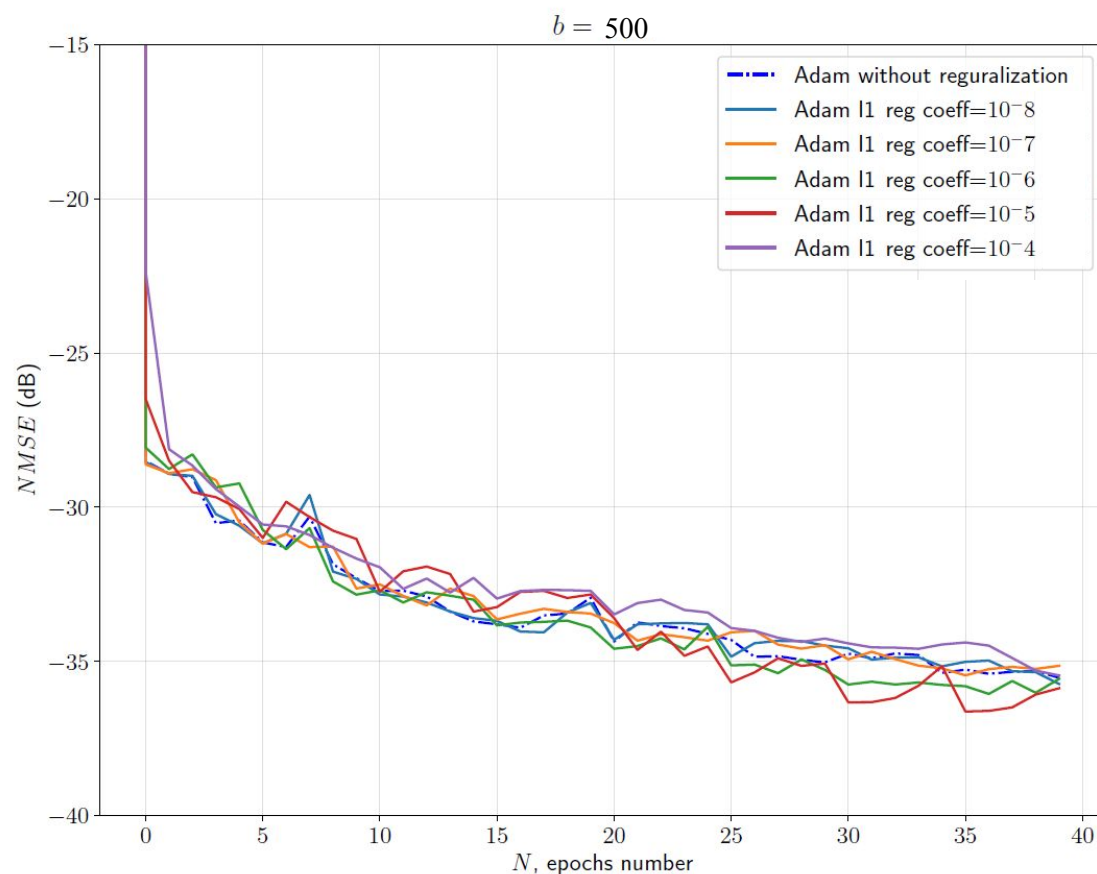


Регуляризация



БОЛЬШИЕ ВЫЗОВЫ

НАУЧНО-ТЕХНОЛОГИЧЕСКАЯ
ПРОЕКТНАЯ ПРОГРАММА



Техника регуляризации задачи:

$$f \rightarrow f + \lambda \|\omega\|$$

Вывод:

Применение ℓ_1 -регуляризации приводит к улучшению NMSE на **3%**



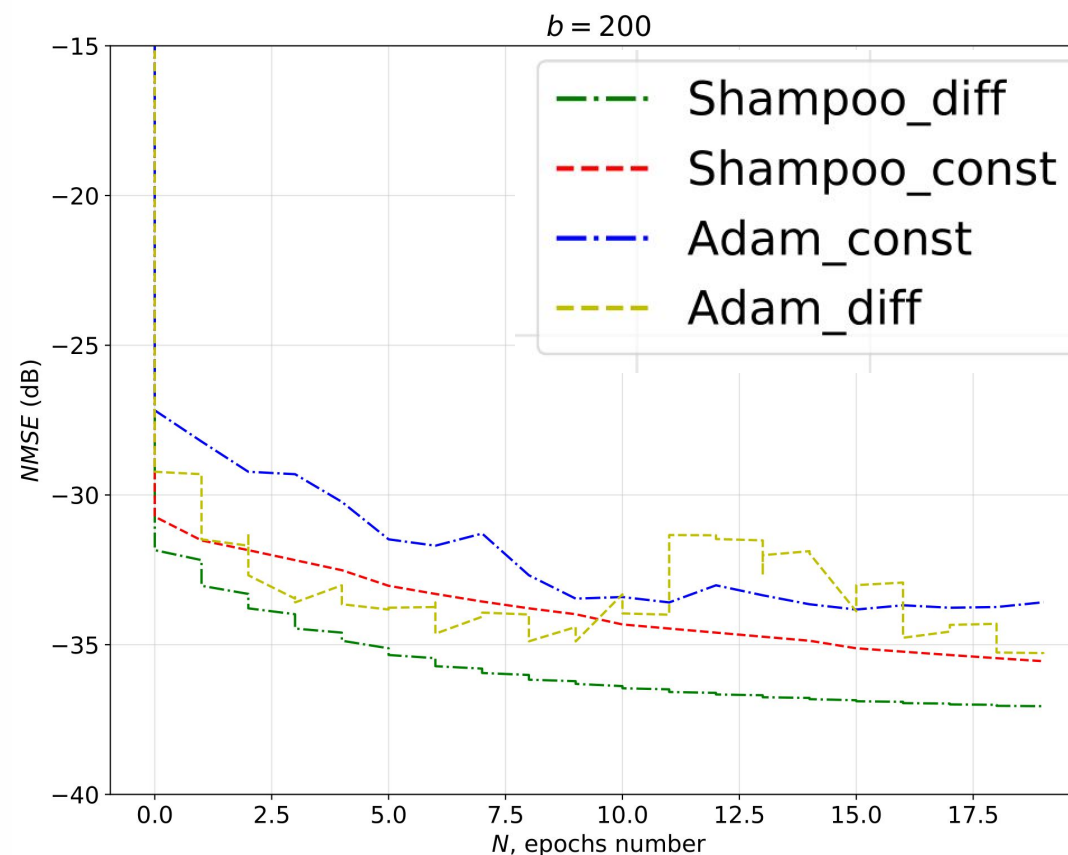
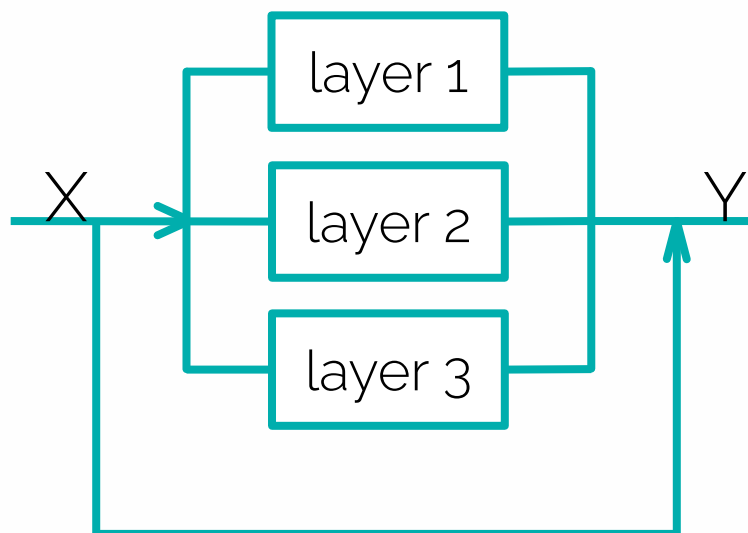
Альтернирование



БОЛЬШИЕ ВЫЗОВЫ
НАУЧНО-ТЕХНОЛОГИЧЕСКАЯ
ПРОЕКТНАЯ ПРОГРАММА

Вывод:

Послойное обучение
позволяет улучшить
NMSE на **6%**





Перепараметризация

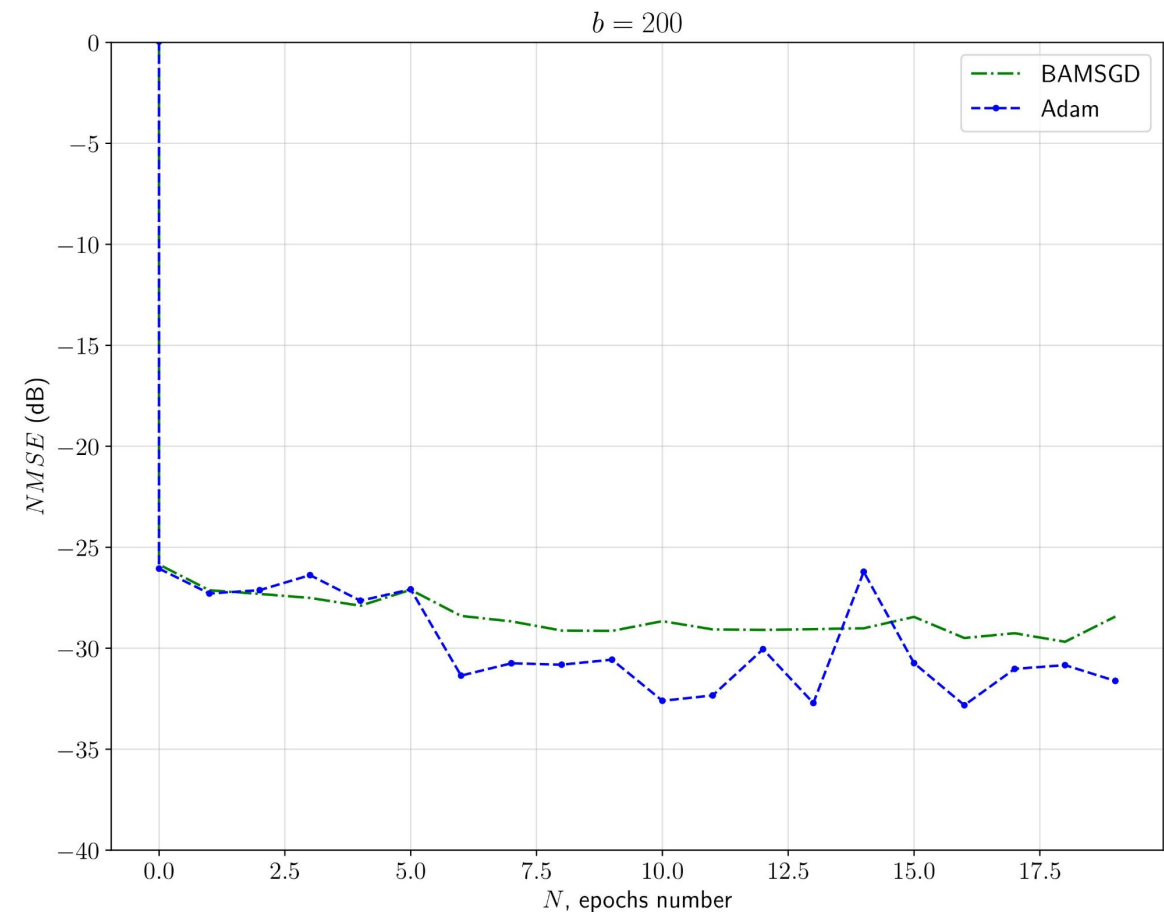


БОЛЬШИЕ ВЫЗОВЫ
НАУЧНО-ТЕХНОЛОГИЧЕСКАЯ
ПРОЕКТНАЯ ПРОГРАММА

arXiv:2106.02720, Woodworth-Srebro 2021

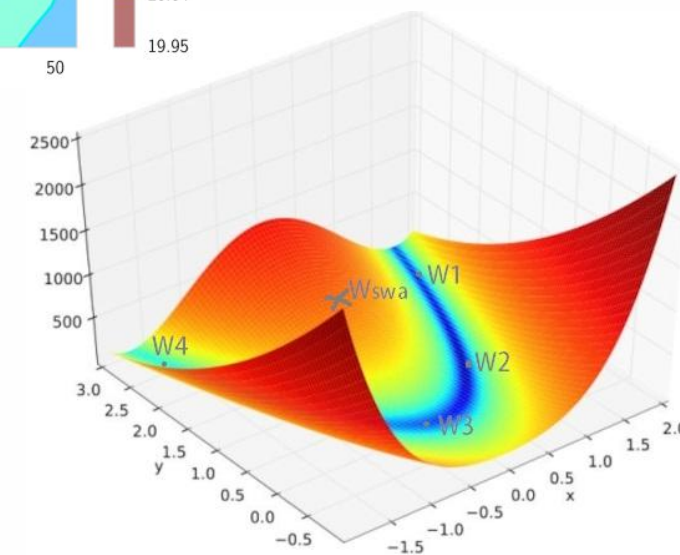
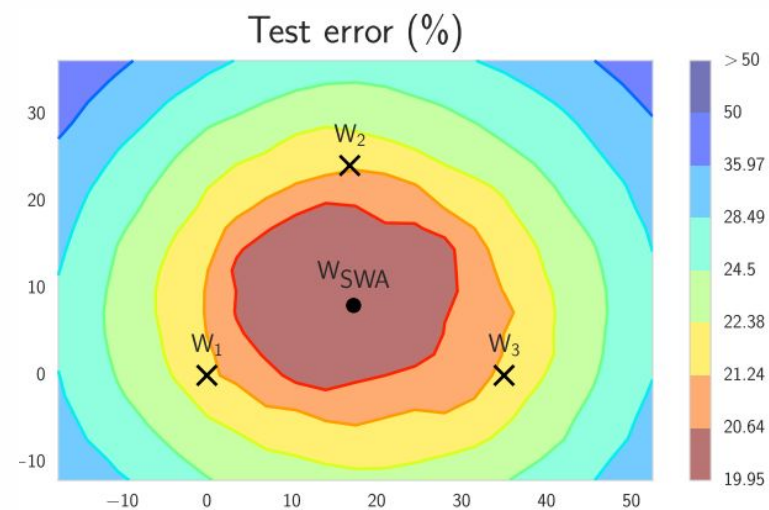
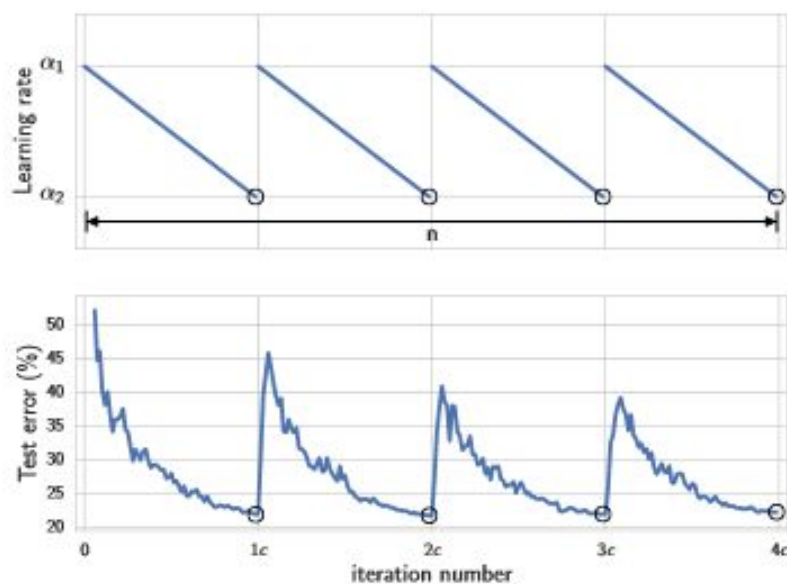
```
 $w_0^{\text{ag}} = w_0 = 0$   
for  $t = 0, 1, \dots, T - 1$  do  
     $\beta_t = 1 + \frac{t}{6}$  and  $\gamma_t = \gamma(t + 1)$   
     $w_t^{\text{md}} = \beta_t^{-1} w_t + (1 - \beta_t^{-1}) w_t^{\text{ag}}$   
     $\tilde{w}_{t+1} = w_t - \gamma_t g_t(w_t^{\text{md}})$   
     $w_{t+1} = \min\left\{1, \frac{B}{\|\tilde{w}_{t+1}\|}\right\} \tilde{w}_{t+1}$   
     $w_{t+1}^{\text{ag}} = \beta_t^{-1} w_{t+1} + (1 - \beta_t^{-1}) w_t^{\text{ag}}$   
Return:  $w_T^{\text{ag}}$ 
```

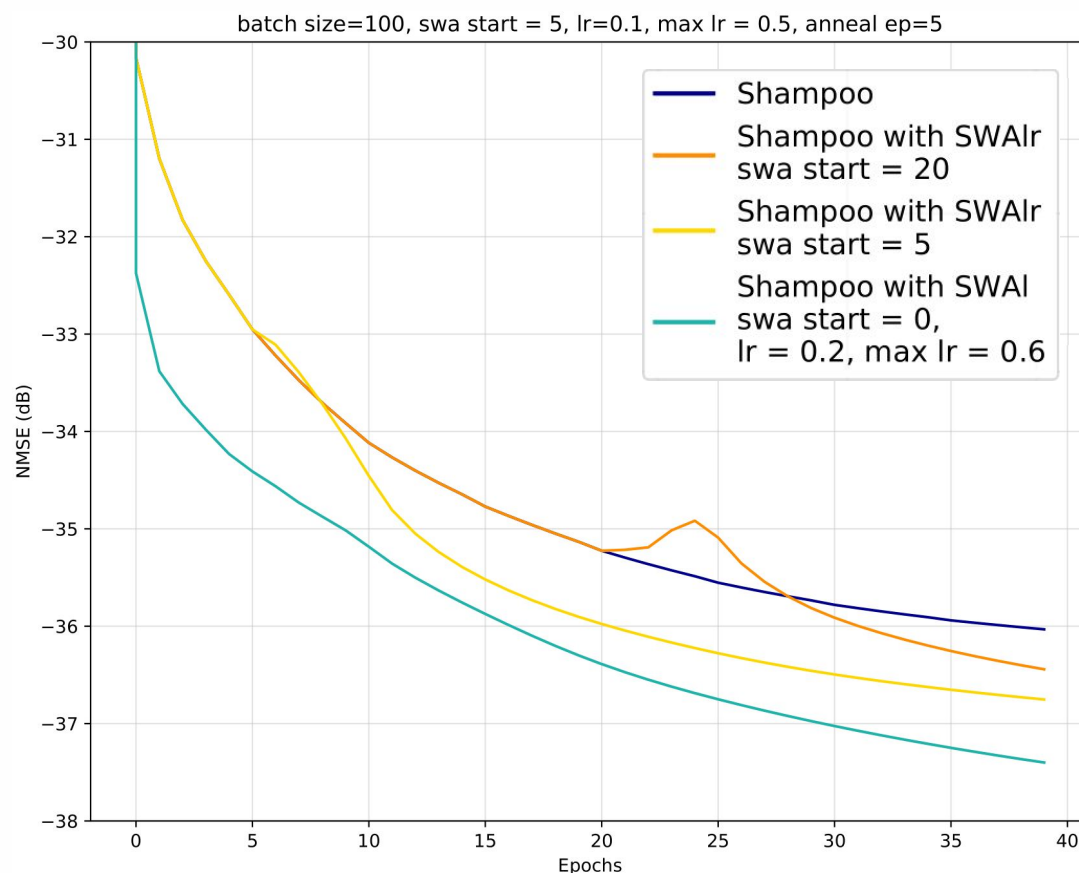
Вывод: NMSE не улучшился :(





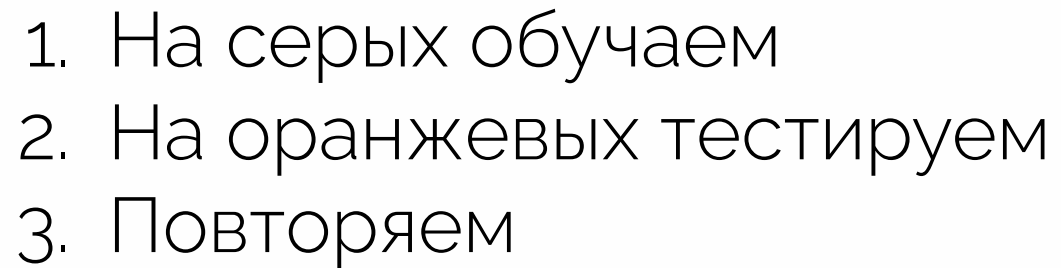
arXiv:1803.05407, Izmailov 2018

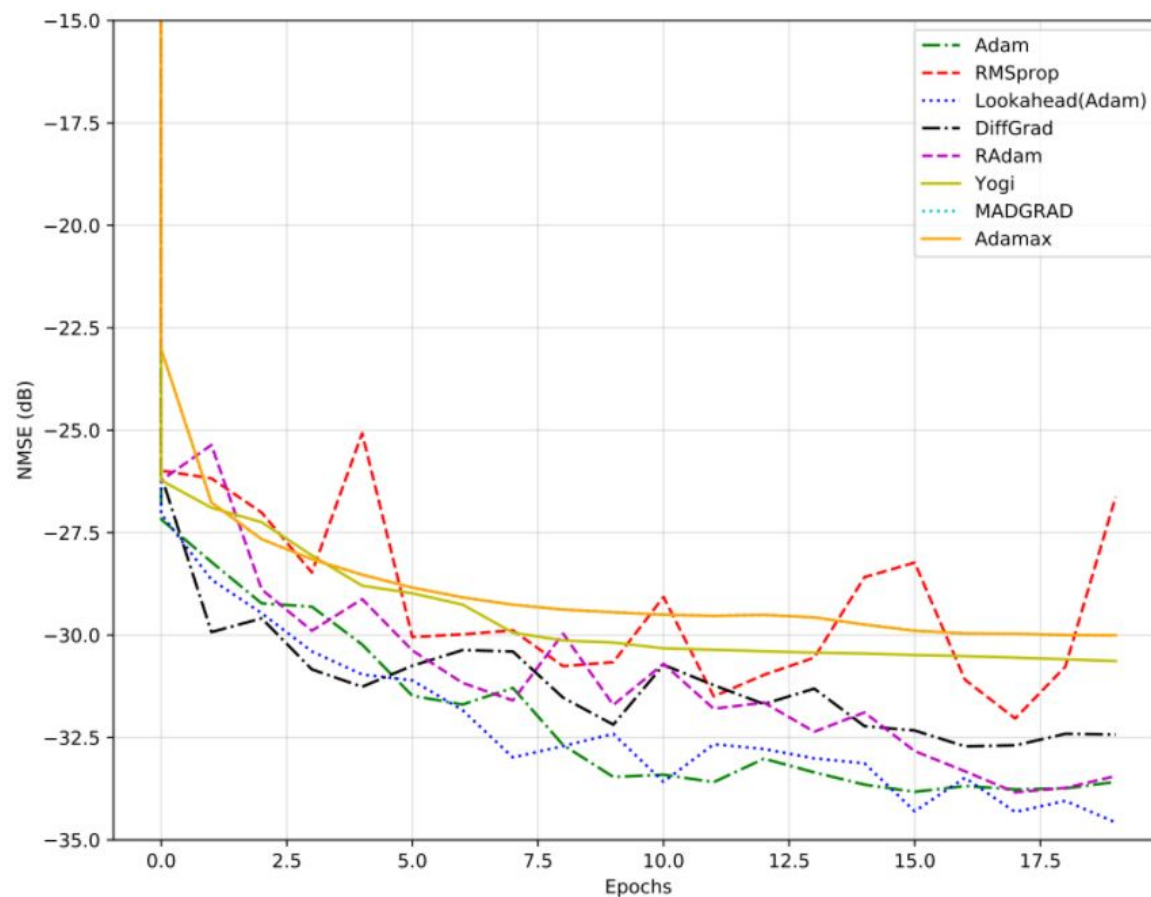




Вывод:

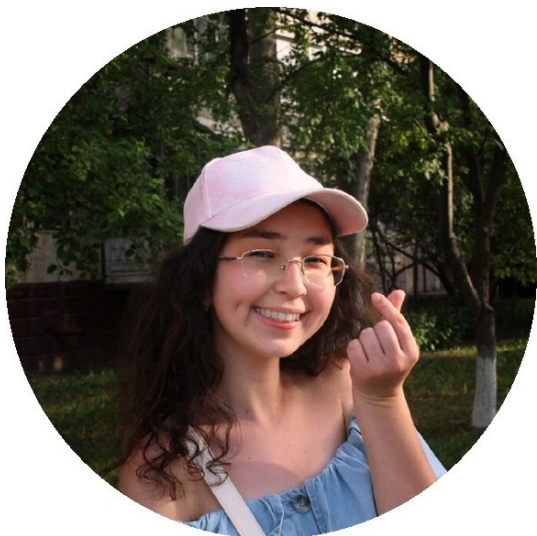
1. Усреднение неэффективно
2. Циклический learning-rate улучшает NMSE на **4%**





Вывод:

1. Прежний метод тестирования неестественен
2. Новый воспроизводит лабораторные тесты
3. Все подходы могут применяться и для Adamax !



Лилиана
Байсланова
[@kusocheck_unosti](#)



Мансур
Зайнуллин
[@Zainullin_Mansur](#)



Егор
Петерс
[@egorka_pomedorka](#)



Артём
Алпатов
[@z3av3rag3](#)

Спасибо за внимание.