The Unreasonable Effectiveness of Optimization in Applications: Personal Experience

Boris Polyak
boris@ipu.ru

Institute of Control Sciences, RAS, Moscow

Khachiyan prize talk
INFORMS Annual Meeting, Anaheim, CA
October 24, 2021

Motivation of the title

E.Wigner, "The Unreasonable Effectiveness of Mathematics in Natural Sciences", Richard Courant lecture, Communications on Pure and Applied Mathematics. v.13, No 1, 114, 1960.
Examples: Newton gravitational law, Maxwell equation, quantum mechanics etc. Page in Wiki.
Inspired many responses: in biology, economics, data minining etc.
Contra:
I.Gelfand: There is only one thing which is more unreasonable than the unreasonable effectiveness of mathematics in physics, and this is the unreasonable ineffectiveness of mathematics in biology.
D.Abbott "The Reasonable Ineffectiveness of Mathematics", V.101, No.10,2013, Proc. IEEE.

Effectiveness of optimization

Key point: Many classes of optimization problems and many methods were invented "purely theoretically" without orientation on real-life problems. However later practical applications appear where the methods happen to be highly effective and well fitted for.
My goal is to illustrate this point with several personal stories.

Story 1: Alternating projections

Feasibility problem: find $x \in \cap_{i=1}^{m} F_i$, $F_I \in H$ are convex closed sets.
Alternating projections $x_{k+1} = P_{i(k)} x_k$, $i(k)$ – cyclic or the remotest.

[1] *J. von Neumann*, 1933, $m = 2, F_1, F_2$ are affine subspaces.

[2] *Cheney W., Goldstein A.* Proximity maps for convex sets. Proc. AMS. 10, 3, 448 - 450, 1959.

[3] *Eremin I.* Generalization of the Motzkin-Agmon relaxation method, Usp. mat. Nauk, 20. 2, 183 - 188, 1965.

[4] *Bregman L.* Finding the common point of convex sets by the method of successive projection. DAN USSR, 162, 3, 487 -490, 1965.

[5] *Gubin L., Polyak B., Raik E.* The method of projections for finding the common point of convex sets. USSR Comp. Math. and Math. Phys. 7 (6): 124, 1967 (2–5 citations/year until 1982).

[6] *Bauschke H., Borwein J.* On projection algorithms for solving convex feasibility problems. SIAM Review. 38 (3): 367-426, 1996.

Story 1: Blow-up

1979 – Nobel prize for Computer Tomography

1982 – *D.Youla, H.Webb*, Image Restoration by the Method of Convex Projections Part I, Theory;

M.Sezan, H.Stark, Part 2, Applications and Numerical Results. IEEE Transactions on Medical Imaging, V.1, No 1, 2.

There was a reference on [5]. Since then the number of citations of [5] grew dramatically. Soon it became my most cited paper.

Of course at 1967 I had no idea on image reconstruction!

Story 2: Nonsmooth optimization

$\min f(x), x \in Q \subset H$, $f(x)$ is convex nondifferentiable, Q is convex closed.

Subgradient method $x_{k+1} = P_Q(x_k - \gamma_k \partial f(x_k))$

$\gamma_k \to 0, \sum_{k=0}^{\infty} \gamma_k = \infty$ or $\gamma_k = \frac{f(x_k)-f^*}{||\partial f(x_k)||^2}$.

[1] *N.Shor* 1962; *Ermol'ev Yu., N.Shor*, On the Minimization of Non-Differentiable Functions, Kibern., 3, 1, 1967.

[2] *B.Polyak*, A General Method of Solving Extremum Problems, Soviet Math. Dokl., 1967, **8**, 593–597.

[3] *B.Polyak*, Minimization of Nonsmooth Functionals, USSR Comp. Math. and Math. Phys., 1969, **9**, No. 3, 14–29.

No serious response till 1974.

Story 2. Applications

- Travelling salesman: *Held M., Karp R.*, 1970.
  Held M., Wolfe P., Crowder, H.P. Validation of subgradient optimization. Math. Progr. 6, 62-88, 1974.
  IIASA Workshop on Nondifferentiable Minimization, Viena, 1977.

Story 2. Applications

- Travelling salesman: *Held M., Karp R.*, 1970.
  Held M., Wolfe P., Crowder, H.P. Validation of subgradient optimization. Math. Progr. 6, 62-88, 1974.
  IIASA Workshop on Nondifferentiable Minimization, Viena, 1977.
- Sparse solutions, l_1 regularization, Lasso $\min f(x) + \alpha||x||_1$
  1986 – geophysics; 1996 – statistics.

Story 3. Heavy ball

B.Polyak, Some Methods of Speeding up the Convergence of Iteration Methods, USSR Comp. Math. and Math. Phys., 1964, **4**, No. 5, 1–17.

$\min f(x), \quad x_{k+1} = x_k - \alpha \nabla f(x_k) + \beta(x_k - x_{k-1})$

Acceleration via momentum, motivated by physical analogy with heavy ball movement in potential field. Global convergence for quadratic case, local convergence for strongly convex case. Optimal parameters:

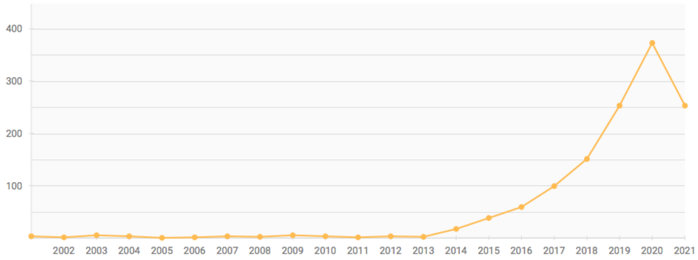
$\alpha^* = \frac{4}{(\sqrt{L}+\sqrt{l})^2}, \quad \beta^* = q^2, q = \frac{\sqrt{L}+\sqrt{l}}{\sqrt{L}+\sqrt{l}}, \quad ||x_k - x^*|| = O(q^k)$

for f being L-smooth and l-strongly convex.

Comparison with optimal gradient descent: acceleration $\sqrt{\kappa}$ times; condition number $\kappa = \frac{L}{l}$ is typically large.

Story 3. 50 years later

Citations of 1964 paper:



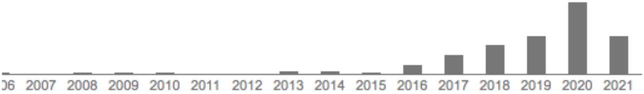
Around 2010 — breakthrough in application of accelerated first-order optimization methods for training of deep learning networks. Nesterov's accelerated gradient (1983) is the champion, but heavy ball is also exploited.

Story 4: Gradient domination condition

B.Polyak, Gradient Methods for the Minimization of Functionals, USSR Comp. Math. and Math. Phys., 1963, **3**, No. 4, 643–653.
Theorem If $f(x)$ is L-smooth, $f(x) \geq f^*$ and
$||\nabla f(x)||^2 \geq 2\mu(f(x) - f^*), \mu > 0, \quad (!)$
then gradient method converges linearly to x^*.
Condition (!) holds for nonconvex functions! Similar conditions *S. Lojasiewicz* (1963), *T. Ležanski*, Über das Minimumproblem für Funktionale in Banachschen Räumen, Math. Ann., 152 (1963), 271–274.
(!) is now known as Gradient-dominated condition or Polyak-Lojasiewicz (PL) condition.
It was completely forgotten for many years.

Story 4: 50 years later

H. Karimi, J. Nutini, M. Schmidt, Linear Convergence of Gradient and Proximal-Gradient Methods under the Polyak-Lojasiewicz Condition. JEC on Machine Learning, 2016.
Citations of 1963 paper:



- Machine learning optimization problems are mostly nonconvex. *M.Belkin*, Fit without fear: remarkable mathematical phenomena of deep learning through the prism of interpolation, 2021
- Feedback optimization problems in control are nonconvex. *I. Fatkhullin, B.Polyak*, Optimizing Static Linear Feedback: Gradient Method, SIAM J. on Control and Optimization, 2021.

Conclusions

- Many classes of optimization problems and many methods were invented "purely theoretically" without orientation on real-life problems. However later practical applications appear where the methods happen to be highly effective and well fitted for.

Conclusions

- Many classes of optimization problems and many methods were invented "purely theoretically" without orientation on real-life problems. However later practical applications appear where the methods happen to be highly effective and well fitted for.
- But to wait you have to live long

THANK YOU!