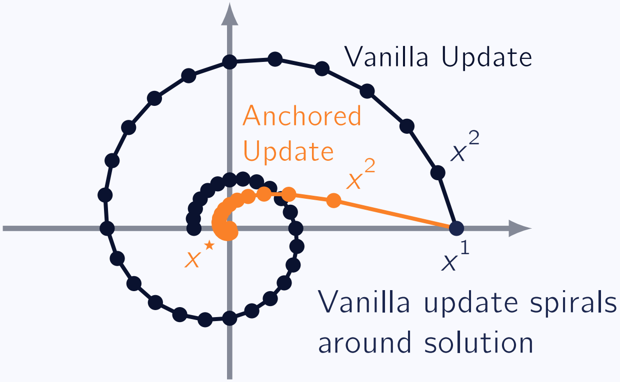
Adding an Anchor ⚓

can Speed Up Your Algorithm



Vanilla update spirals around solution

Howard Heaton

Typal Academy

Fixed Point Iteration

Many algorithms can be written using an operator, denoted here by T. Given input x^1, each update is

$$x^{k+1} = T(x^k).$$

For "nice" operators T, the sequence $\{x^k\}$ converges to a limit $x^\star$ for which $x^\star = T(x^\star)$.

Example: Under standard assumptions, the problem

$$\min_{x \in \mathcal{C}} f(x)$$

can be solved with the projected gradient operator

$$T(x) = \mathrm{proj}_{\mathcal{C}}(x - \alpha \nabla f(x)),$$

with $\alpha > 0$ a step size and $\mathrm{proj}_{\mathcal{C}}$ the projection onto $\mathcal{C}$.

Anchoring via Halpern

Given input x^1, each update for Halpern iteration is[†]

$$x^{k+1} = \frac{1}{n} \cdot x^1 + \left(1 - \frac{1}{n}\right) \cdot T(x^k).$$

Each update x^{k+1} is an average of x^1 and $T(x^k)$.

Theorem: When T has a fixed point and is "nice,"[‡] the sequence $\{x^k\}$ generated by Halpern converges to a limit $x^\star$ for which $x^\star = T(x^\star)$. Moreover,

$$\|x^{k+1} - x^k\| = \mathcal{O}(1/k^2).$$

Note: Without anchoring, we may only be able to guarantee $\mathcal{O}(1/k)$ rate of convergence.

[†] A special case of a more general formula is shown here.

[‡] That is, $\|T(x) - T(y)\| \leq \|x - y\|$ for all x and y

Example

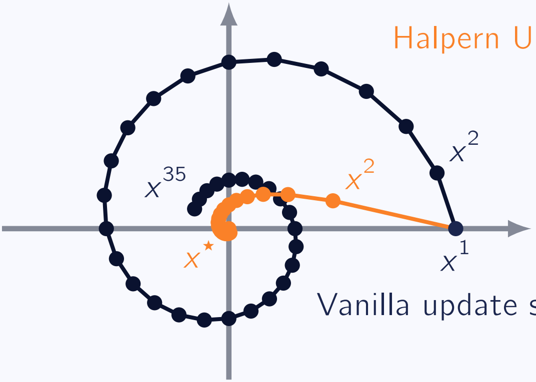
Suppose the update operator T is a rotation in the 2D plane by θ that is scaled by $0 < \lambda < 1$, *i.e.*

$$T(x) = \lambda \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}.$$

Here $x^\star$ is the origin.

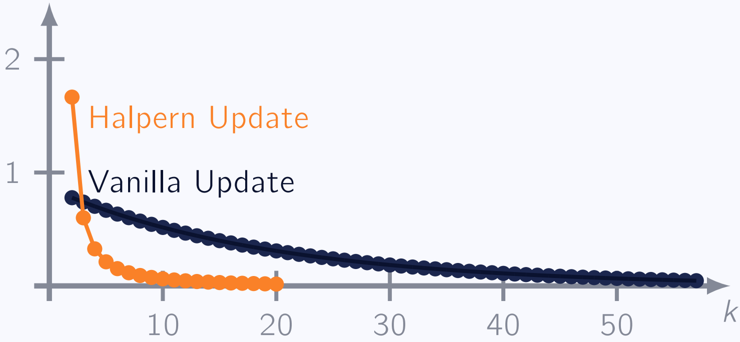
Vanilla Update: $x^{k+1} = T(x^k)$

Halpern Update

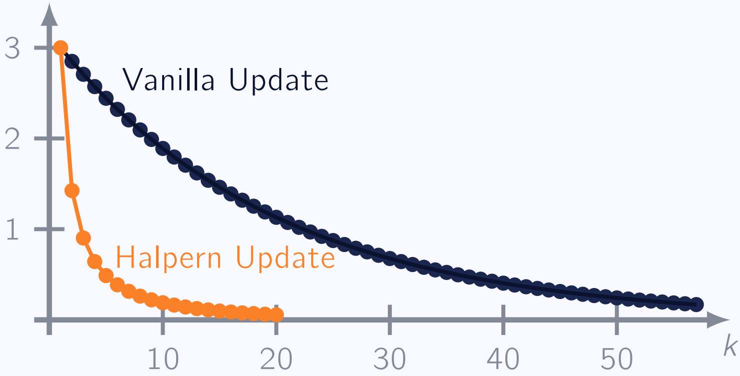


Vanilla update spirals around $x^\star$

Example Continued



Update Residual $\|x^{k+1} - x^k\|$ vs. iteration k



Estimate Error $\|x^k - x^\star\|$ vs. iteration k

Takeaways

- Anchoring is trivial to implement

  (weighted average with x^1 and $T(x^k)$)

- Anchoring can improve convergence rate

  (*i.e.* go from $\mathcal{O}(1/k)$ to $\mathcal{O}(1/k^2)$)

- Anchoring can mitigate oscillatory behavior

- Anchoring is distinct from Nesterov acceleration

Reference: Example drawn from *Exact Optimal Accelerated Complexity for Fixed-Point Iterations*

Found this useful?

- + Follow for more
- ♻ Repost to share with friends