Operator fixed point theory: an algorithm construction toolkit

- Nonlinear functional analysis,
- operator equations,
- operator fixed points,
- convex optimisation,
- neural networks,
- Markov chain Monte Carlo.

D. Semenovich

Operator equations

Among the topics in nonlinear functional analysis is the study of (non-linear) operators and set valued mappings. In the finite dimensional case an operator is just an arbitrary function $T : \mathbb{R}^n \to \mathbb{R}^n$ and a set-valued mapping a function $T : \mathbb{R}^n \to P(\mathbb{R}^n)$.

Central to the study of operators are operator equations $T(x) = 0$ or $T(x) \ni 0$, e.g. $Ax - b = 0$, $\nabla f(x) = 0$ or $\partial f(x) \ni 0$). Some of the questions:

- existence and uniqueness of a solution,
- stability of the solutions under perturbation or noisy measurements,
- construction of approximation methods and estimation of their convergence.

Operator fixed points

Related to operator equations is the study of operator fixed points. Point x is a fixed point of operator T if $T(x) = x$ or in the case of set-valued mappings, $T(x) \ni x$.

Some possibilities for reformulating $T(x) = 0$ as a fixed point problem include:

- basic iteration: $x = (I - T)(x)$,
- damped iteration: $x = (I - \lambda T)(x)$,
- Newton's method: $x = \Big(I - \big(DT(x)\big)^{-1} \circ T\Big)(x)$,
- operator splitting: $x = F^{-1} \circ (-T - G)(x)$, where $F + G = T$.

Fixed point iterations

Fixed point problems suggest the following "master algorithm":

$$x_{k+1} = T(x_k).$$

The question then becomes what is the largest or most useful class of operators for which the above scheme converges from an arbitrary starting point.

Consider the so called *non-expansive* operators. An operator T is non-expansive if $\|T(x) - T(y)\| \leq \|x - y\|$ for every $x, y \in \mathbb{R}^n$.

The operator $-I$ is non-expansive but the fixed point iteration clearly does not converge, even though 0 is a fixed point.

Fixed point iterations - continued

Convex optimisation

Assume $f(x) = g(x) + h(x)$ is a convex function, then the subgradient ∂f is a monotone operator. Some fixed point iterations for the operator equation $\partial f \ni 0$ include:

- subgradient descent: $x \in (I - \lambda \partial f)x$,
- proximal point method: $x = (I + \lambda \partial f)^{-1}x$, provided $h(x)$ is smooth;
- proximal gradient algorithm: $x = (I + \lambda \partial g)^{-1}(I - \lambda \nabla h)x$,
- Peaceman-Rachford splitting:

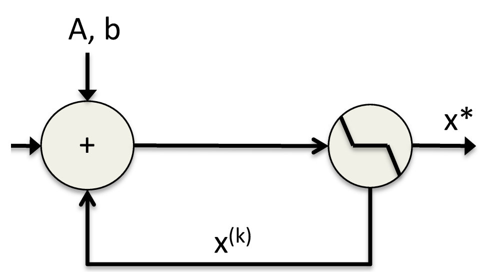
$$
\begin{aligned}
z &= \Big(2(I + \lambda \partial g)^{-1} - I\Big)\Big(2(I + \lambda \partial h)^{-1} - I\Big)z, \\
x &= (I + \lambda \partial h)^{-1}z.
\end{aligned}
$$

Neural networks

Consider the ℓ_1-norm regularised least squares regression and the corresponding proximal gradient fixed point iteration:

$$\underset{\mathbf{x}}{\text{minimise}} \quad \|b - Ax\|_2^2 + \alpha\|x\|_1 \;=\; f(x) + g(x),$$

$$\begin{aligned} x^{(k+1)} &= \Big(I + \frac{\alpha}{\lambda}\partial g\Big)^{-1}\Big(I - \frac{1}{\lambda}\nabla_x f\Big)\Big(x^{(k)}\Big) \\ &= h_{\frac{\alpha}{\lambda}}\Big((I - \frac{1}{\lambda}A^T A)x^{(k)} + A^T b\Big), \quad \text{with } [h_{\frac{\alpha}{\lambda}}(x)]_i = \text{sign}(x_i)\Big(|x_i| - \frac{\alpha}{\lambda}\Big)_+. \end{aligned}$$



Markov chain Monte Carlo

MCMC algorithms to generate samples from the target distribution $p^\star$ can be viewed as constructing an ergodic Markov chain transition operator T which has $p^\star$ as the stationary distribution:

$$p^\star = Tp^\star.$$

Popular schemes such as M-H are heuristics (i.e. the *detailed balance* condition is sufficient but not necessary) to build specific instances of such operators.
We can arbitrarily split the operator T, where $T_1, \ldots, T_n$ need not be ergodic individually:

$$p^\star = T_1 \circ T_2 \circ \ldots \circ T_n p^\star.$$

This recovers e.g. Gibbs sampling.

Conclusions

- operator notation can help to think about a large range of machine learning problems in a unified way,
- operator splitting in particular can be used to trivially generate custom algorithms that automatically provide certain (admittedly rather basic) guarantees.
- functional analysis perspective helps to find related problems in other fields.