Gradient flow 1

Derivation of gradient, proximal point, extra-gradient and Runge-Kutta update step

Andersen Ang

Mathématique et recherche opérationnelle
UMONS, Belgium

manshun.ang@umons.ac.be Homepage: angms.science

First draft : July 22, 2020
Last update : July 22, 2020

ODE and Euler method

- Consider an ordinary differential equation

$$\frac{dx(t)}{dt} = Y\Big(t, x(t)\Big).$$

with initial value $Y(0, x(0))$. Here $\frac{dx(t)}{dt}$ is the rate of change of $x(t)$ with respect to time, and the function $Y\Big(t, x(t)\Big)$ describes such rate of change.

- In general ODE is hard to solve. A numerical method to solve such ODE is the Euler method, the idea is to approximate the term $\frac{dx(t)}{dt}$ by the approximate

$$\frac{\Delta x}{\Delta t} = \frac{x(t+h) - x(t)}{h}.$$

Euler method

- Let $x_k = x(k\Delta t)$ and $h = \Delta t$, based on the setup of Δx, we have
  - Forward Euler (FE): $\dfrac{dx(t)}{dt} \approx \dfrac{x_{k+1} - x_k}{h}$.
  - Backward Euler (BE): $\dfrac{dx(t)}{dt} \approx \dfrac{x_k - x_{k-1}}{h}$.
- The numerical scheme that solves the ODE becomes
  - FE: $\dfrac{x_{k+1} - x_k}{h} = Y\Big(t, x_k\Big)$.
  - BE: $\dfrac{x_{k+1} - x_k}{h} = Y\Big(t, x_{k+1}\Big)$.

  Rearrange terms give
  - FE: $x_{k+1} = x_k + hY\Big(t, x_k\Big)$.
  - BE: $x_{k+1} = x_k + hY\Big(t, x_{k+1}\Big)$.

Gradient flow

- The gradient flow is an ODE in the form

$$\frac{dx(t)}{dt} = -\nabla f\Big(x(t)\Big),$$

  i.e., $Y\Big(t, x(t)\Big) = -\nabla f\Big(x(t)\Big)$.

- The Euler methods give
  - FE: $x_{k+1} = x_k - h\nabla f(x_k)$.
  - BE: $x_{k+1} = x_k - h\nabla f(x_{k+1})$.

- We see immediately that FE is the gradient step.

- The BE is in fact the proximal point step

BE is the proximal point step

- For the equation $x_{k+1} = x_k - h\nabla f(x_{k+1})$, rearrange gives $x_{k+1} + h\nabla f(x_{k+1}) = x_k$. Consider a "factorization" on the expression $x_{k+1} + h\nabla f(x_{k+1})$:

$$x_{k+1} + h\nabla f(x_{k+1}) = (I + h\nabla f)x_{k+1}.$$

  Then we have

$$x_{k+1} = (I + h\nabla f)^{-1} x_k.$$

- Formally, the notation of $(I + h\nabla f)^{-1}$ is $(I + h\partial f)^{-1}$ and it is called the resolvent operator.
- Using proximal operator notation, the expression is

$$x_{k+1} = \mathrm{prox}_{hf}(x_k).$$

Coarse Runge – Kutta method

- A coarse Runge – Kutta method for solving the ODE gives the so-called extra-gradient update

$$x_{k+\frac{1}{2}} = x_k - h\nabla f(x_k), \qquad x_{k+1} = x_k - h\nabla f(x_{k+\frac{1}{2}}).$$

  The two-step update combined into one-step

$$x_{k+1} = x_k - h\nabla f\Big(x_k - h\nabla f(x_k)\Big).$$

Fourth-order Runge – Kutta method

- The 4th-order Runge – Kutta method for solving the ODE gives

$$\mathbf{x}_{k+1} = \mathbf{x}_k + \frac{h}{6}\Big(s_1 + 2s_2 + 2s_3 + s_4\Big),$$

  where

$$\begin{aligned} s_1 &= \nabla f(x_k), \\ s_2 &= \nabla f(x_k + hs_1/2), \\ s_3 &= \nabla f(x_k + hs_2/2), \\ s_4 &= \nabla f(x_k + hs_3), \end{aligned}$$

- To illustrate the update of RK4, we consider quadratic programming problem

$$\min_{\mathbf{x}} \frac{1}{2}\langle \mathbf{x}, \mathbf{Q}\mathbf{x} \rangle.$$

  with $\nabla f(\mathbf{x}) = \mathbf{Q}\mathbf{x}$.

RK4 on QP

- The RK4 update

$$\mathbf{x}_{k+1} = \mathbf{x}_k + \frac{h}{6}\Big(s_1 + 2s_2 + 2s_3 + s_4\Big),$$

where

$$\begin{aligned}
s_1 &= -\mathbf{Q}\mathbf{x}_k, \\
s_2 &= -\mathbf{Q}(\mathbf{x}_k + hs_1/2) &&= (-\mathbf{I} + \tfrac{h}{2}\mathbf{Q})\mathbf{Q}\mathbf{x}_k, \\
s_3 &= -\mathbf{Q}(\mathbf{x}_k + hs_2/2) &&= (-\mathbf{I} + \tfrac{h}{2}\mathbf{Q} - \tfrac{h^2}{4}\mathbf{Q}^2)\mathbf{Q}\mathbf{x}_k, \\
s_4 &= -\mathbf{Q}(\mathbf{x}_k + hs_3) &&= (-\mathbf{I} + h\mathbf{Q} - \tfrac{h^2}{2}\mathbf{Q}^2 + \tfrac{h^3}{4}\mathbf{Q}^3)\mathbf{Q}\mathbf{x}_k
\end{aligned}$$

- After some algebra, the RK4 update step for solving QP is thus

$$\mathbf{x}_{k+1} = \mathbf{x}_k - h\left(\mathbf{I} - \frac{h}{2}\mathbf{Q} + \frac{h^2}{6}\mathbf{Q}^2 - \frac{h^3}{24}\mathbf{Q}^3\right)\mathbf{Q}\mathbf{x}_k.$$

- A closer look at the bracket term, it is the first four term in the sum

$$\sum_{k=0}^{\infty} \frac{(-h\mathbf{Q})^k}{(k+1)!}.$$

RK4 and GD update

- RK4 update

$$\mathbf{x}_{k+1} = \mathbf{x}_k - h\left(\mathbf{I} - \frac{h}{2}\mathbf{Q} + \frac{h^2}{6}\mathbf{Q}^2 - \frac{h^3}{24}\mathbf{Q}^3\right)\mathbf{Q}\mathbf{x}_k.$$

- GD update

$$\mathbf{x}_{k+1} = \mathbf{x}_k - h\mathbf{Q}\mathbf{x}_k.$$

- Scaled GD update

$$\mathbf{x}_{k+1} = \mathbf{x}_k - h\mathbf{D}\mathbf{Q}\mathbf{x}_k.$$

- RK4 can be interpreted as a scaled gradient update, with

$$\mathbf{D} = \left(\mathbf{I} - \frac{h}{2}\mathbf{Q} + \frac{h^2}{6}\mathbf{Q}^2 - \frac{h^3}{24}\mathbf{Q}^3\right).$$

Stepsize of RK4 update

- We now derive the stepsize h selection rule in the RK4 update. First some algebra

$$\begin{aligned}
\mathbf{x}_{k+1} &= \mathbf{x}_k - h\Big(\mathbf{I} - \tfrac{h}{2}\mathbf{Q} + \tfrac{h^2}{6}\mathbf{Q}^2 - \tfrac{h^3}{24}\mathbf{Q}^3\Big)\mathbf{Q}\mathbf{x}_k \\
&= \mathbf{x}_k - h\mathbf{Q}\mathbf{x}_k + \tfrac{h^2}{2}\mathbf{Q}^2\mathbf{x}_k - \tfrac{h^3}{6}\mathbf{Q}^3\mathbf{x}_k + \tfrac{h^4}{24}\mathbf{Q}^4\mathbf{x}_k \\
&= \Big(\mathbf{I} - h\mathbf{Q} + \tfrac{h^2}{2}\mathbf{Q}^2 - \tfrac{h^3}{6}\mathbf{Q}^3 + \tfrac{h^4}{24}\mathbf{Q}^4\Big)\mathbf{x}_k
\end{aligned}$$

- Let $\mathbf{Q} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^{-1}$ be the eigendecomposition of $\mathbf{Q}$, then

$$\begin{aligned}
&\mathbf{V}^{-1}\Big(\mathbf{I} - h\mathbf{Q} + \tfrac{h^2}{2}\mathbf{Q}^2 - \tfrac{h^3}{6}\mathbf{Q}^3 + \tfrac{h^4}{24}\mathbf{Q}^4\Big) \\
&= \mathbf{V}^{-1} - h\mathbf{\Lambda}\mathbf{V}^{-1} + \tfrac{h^2}{2}\mathbf{\Lambda}^2\mathbf{V}^{-1} - \tfrac{h^3}{6}\mathbf{\Lambda}^3\mathbf{V}^{-1} + \tfrac{h^4}{24}\mathbf{\Lambda}^4\mathbf{V}^{-1} \\
&= \Big(\mathbf{I} - h\mathbf{\Lambda} + \tfrac{h^2}{2}\mathbf{\Lambda}^2 - \tfrac{h^3}{6}\mathbf{\Lambda}^3 + \tfrac{h^4}{24}\mathbf{\Lambda}^4\Big)\mathbf{V}^{-1}
\end{aligned}$$

- Now let $\mathbf{y}_k = \mathbf{V}^{-1}\mathbf{x}_{k+1}$

$$\begin{aligned}
\mathbf{V}^{-1}\mathbf{x}_{k+1} &= \Big(\mathbf{I} - h\mathbf{\Lambda} + \tfrac{h^2}{2}\mathbf{\Lambda}^2 - \tfrac{h^3}{6}\mathbf{\Lambda}^3 + \tfrac{h^4}{24}\mathbf{\Lambda}^4\Big)\mathbf{V}^{-1}\mathbf{x}_k \\
\mathbf{y}_{k+1} &= \Big(\mathbf{I} - h\mathbf{\Lambda} + \tfrac{h^2}{2}\mathbf{\Lambda}^2 - \tfrac{h^3}{6}\mathbf{\Lambda}^3 + \tfrac{h^4}{24}\mathbf{\Lambda}^4\Big)\mathbf{y}_k
\end{aligned}$$

We have now decoupled the elements in $\mathbf{y}_k$.

Stepsize of RK4 update

- The ith element in $\mathbf{y}_k$ is independent to other elements in $\mathbf{y}_k$, since $\mathbf{\Lambda}$ is a diagonal matrix.

$$y_{k+1}[i] = \left(1 - h\lambda_i + \frac{h^2}{2}\lambda_i^2 - \frac{h^3}{6}\lambda_i^3 + \frac{h^4}{24}\lambda_i^4\right) y_k[i]$$

- In the QP example, we know the minimizer is $\mathbf{x} = 0$. Hence the stepsize h for the RK4 update has to make the $y_{k+1}[i]$ smaller than $y_k[i]$, which translate to the condition

$$\left|1 - \lambda_i h + \frac{\lambda_i^2}{2}h^2 - \frac{\lambda_i^3}{6}h^3 + \frac{\lambda_i^4}{24}h^4\right| < 1,$$

that is, to select stepsize for RK4 update, h has to be inside the set

$$\left\{ h \in \mathbb{R} \ : \ \left|1 - \lambda_i h + \frac{\lambda_i^2}{2}h^2 - \frac{\lambda_i^3}{6}h^3 + \frac{\lambda_i^4}{24}h^4\right| < 1, \right\}$$

for $i = 1, 2, \dots, \operatorname{rank}(\mathbf{Q})$. That is, the stepsize has to fulfill the "characteristic inequality" above for all the eigenvalue of $\mathbf{Q}$.

Last page - summary

- ODE and gradient flow
- Gradient update comes from applying forward Euler on approximating the gradient flow
- Proximal point update comes from applying backward Euler on approximating the gradient flow
- Extra-gradient update comes from applying a coarse Runge-Kutta method on approximating the gradient flow
- The 4th-order Runge-Kutta method on approximating the gradient flow
- Stepsize selection for the 4th-order Runge-Kutta update on a special quadratic programming problem

End of document