Runge–Kutta-like scaling techniques for first-order methods in convex optimization

Federica Porta[1], Anastasia Cornelio[2] and Valeria Ruggiero[1]

[1]Università di Ferrara, [2]Università di Bologna

Abstract

It is well known that there is a strong connection between *time integration* and *convex optimization*. In this work, inspired by the equivalence between the forward Euler scheme and the gradient descent method, we broaden our analysis to the family of Runge-Kutta methods and show that they enjoy a natural interpretation as first-order optimization algorithms. The strategies intrinsically suggested by Runge-Kutta methods are exploited in order to detail novel proposal for scaling gradient-like approaches.

Framework

Constrained optimization problem

$$\min_{\boldsymbol{x}\geq \boldsymbol{0}} f(\boldsymbol{x}) \qquad (1)$$

- f is a continuously differentiable and convex function;
- the equilibrium points of
$$\frac{\mathrm{d}\boldsymbol{x}}{\mathrm{d}t} = -\nabla f(\boldsymbol{x}(t))$$
are the *unconstrained* minimizers of f.

Two–metric projection method

$$\boldsymbol{x}_{\ell+1} = \mathbb{P}_{\{\boldsymbol{x}\geq\boldsymbol{0}\}}\big(\boldsymbol{x}_\ell - \alpha_\ell \boldsymbol{M}_\ell \nabla f(\boldsymbol{x}_\ell)\big) \qquad (2)$$

- α_ℓ is chosen by means of the Armijo along-the-arc line-search strategy;
- $\boldsymbol{M}_\ell$ is symmetric, positive definite (s.d.p.) and, given $\epsilon > 0$, diagonal with respect to the subset of indexes
$$\mathcal{I}^+_\epsilon(\boldsymbol{x}_\ell) = \left\{ i \;\Big|\; 0 \leq x^{(i)}_\ell \leq \epsilon,\ \frac{\partial f(\boldsymbol{x}_\ell)}{\partial x^{(i)}_\ell} > 0 \right\}$$

Runge–Kutta methods for a special case

If applied to the *gradient flow* problem $\frac{d\boldsymbol{x}}{dt} = -(\boldsymbol{A}\boldsymbol{x}-\boldsymbol{b})$ with $\boldsymbol{A}\in\mathbb{R}^{n\times n}$ s.p.d., classical Euler methods for ODE can be written as follows:

$$\boldsymbol{x}_{\ell+1} = \boldsymbol{x}_\ell - h(\boldsymbol{A}\boldsymbol{x}_\ell - \boldsymbol{b}) \qquad \textit{Forward Euler}$$

$$\boldsymbol{x}_{\ell+1} = \boldsymbol{x}_\ell - h\big(\boldsymbol{I} - \tfrac{h}{2}\boldsymbol{A}\big)(\boldsymbol{A}\boldsymbol{x}_\ell - \boldsymbol{b}) \qquad \textit{Midpoint}$$

$$\boldsymbol{x}_{\ell+1} = \boldsymbol{x}_\ell - h\Big(\boldsymbol{I} - \tfrac{h}{2}\boldsymbol{A} + \tfrac{1}{6}h^2\boldsymbol{A}^2 - \tfrac{1}{24}h^3\boldsymbol{A}^3\Big)(\boldsymbol{A}\boldsymbol{x}_\ell - \boldsymbol{b}) \qquad \textit{Fourth-order Runge-Kutta}$$

$$\boldsymbol{x}_{\ell+1} = \boldsymbol{x}_\ell - h(\boldsymbol{I} + h\boldsymbol{A})^{-1}(\boldsymbol{A}\boldsymbol{x}_\ell - \boldsymbol{b}) \qquad \textit{Backward Euler}$$

where, for the explicit methods, h has to be chosen in the proper stability region, while for the implicit algorithms there is no limitations on h.

Remark. All the scaling matrices which appear in the previous algorithms can be viewed as approximation of the *filtered* or *regularized* inverse of $\boldsymbol{A}$, namely the matrix $(\boldsymbol{I} - e^{-h\boldsymbol{A}})\boldsymbol{A}^{-1}$. Therefore these methods are suited for *inverse problems*.

Runge–Kutta-like two–metric projection algorithms

We propose a **new idea** for selecting $\boldsymbol{M}_\ell$ in the **two–metric projection method** applied to (1) by starting from the following matrices

$$\boldsymbol{D}^{ERK}_\ell(h) = h\left(\boldsymbol{I} - \tfrac{h}{2}\nabla^2 f(\boldsymbol{x}_\ell)\right) = \left(h\boldsymbol{I} - \tfrac{h^2}{2}\nabla^2 f(\boldsymbol{x}_\ell)\right) \qquad \boldsymbol{D}^{IRK}_\ell(h) = h\left(\boldsymbol{I} + h\nabla^2 f(\boldsymbol{x}_\ell)\right)^{-1} = \left(h^{-1}\boldsymbol{I} + \nabla^2 f(\boldsymbol{x}_\ell)\right)^{-1}$$

where in the first case h is fixed such that $\boldsymbol{D}^{ERK}_\ell(h)$ is positive definite. We make them partly diagonal by introducing two diagonal matrices $\boldsymbol{E}_\ell$ and $\boldsymbol{F}_\ell$ [1] such that

$$(\boldsymbol{E}_\ell)_{ii} = \begin{cases} 1 & i \notin \mathcal{I}^+_\epsilon(\boldsymbol{x}_\ell) \\ 0 & i \in \mathcal{I}^+_\epsilon(\boldsymbol{x}_\ell) \end{cases} \qquad \text{and} \qquad \boldsymbol{F}_\ell = \boldsymbol{I} - \boldsymbol{E}_\ell\,.$$

Then possible scaling matrices diagonal with respect to $\mathcal{I}^+_\epsilon(\boldsymbol{x}_\ell)$ which can be employed for (2) are given by $\boldsymbol{M}_\ell = \boldsymbol{E}_\ell\boldsymbol{B}_\ell\boldsymbol{E}_\ell + \boldsymbol{F}_\ell$ where $\boldsymbol{B}_\ell$ can be selected as either $\boldsymbol{D}^{ERK}_\ell(h)$ or $\boldsymbol{D}^{IRK}_\ell(h)$. These choices for the scaling matrix exploit second derivatives information as in the *quasi-Newton approaches*; however the presence of the penalty terms $h\boldsymbol{I}$ or $h^{-1}\boldsymbol{I}$ makes our definitions more stable as confirmed by the **numerical results** obtained in **imaging problems** under **Poisson noise** and offered in the following.

Image formation process: an inverse problem

$$\boldsymbol{y} = H\boldsymbol{x} + \boldsymbol{b} + \boldsymbol{\eta}$$

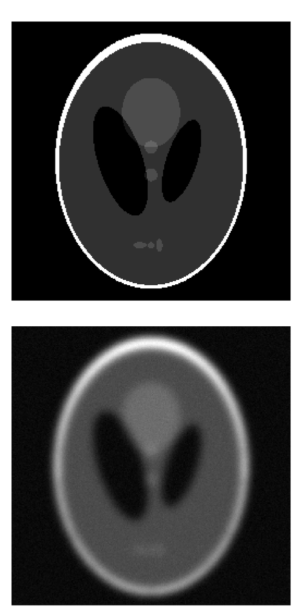
- $\boldsymbol{y}\in\mathbb{R}^n$: observed data; ▸ $\boldsymbol{x}\in\mathbb{R}^n$: image to be recovered;
- $H\in\mathbb{R}^{n\times n}$: burring matrix; ▸ $\boldsymbol{b}\in\mathbb{R}^n$: background;
- $\boldsymbol{\eta}\in\mathbb{R}^n$: Poisson noise corrupting the data.

Image restoration: optimization problem

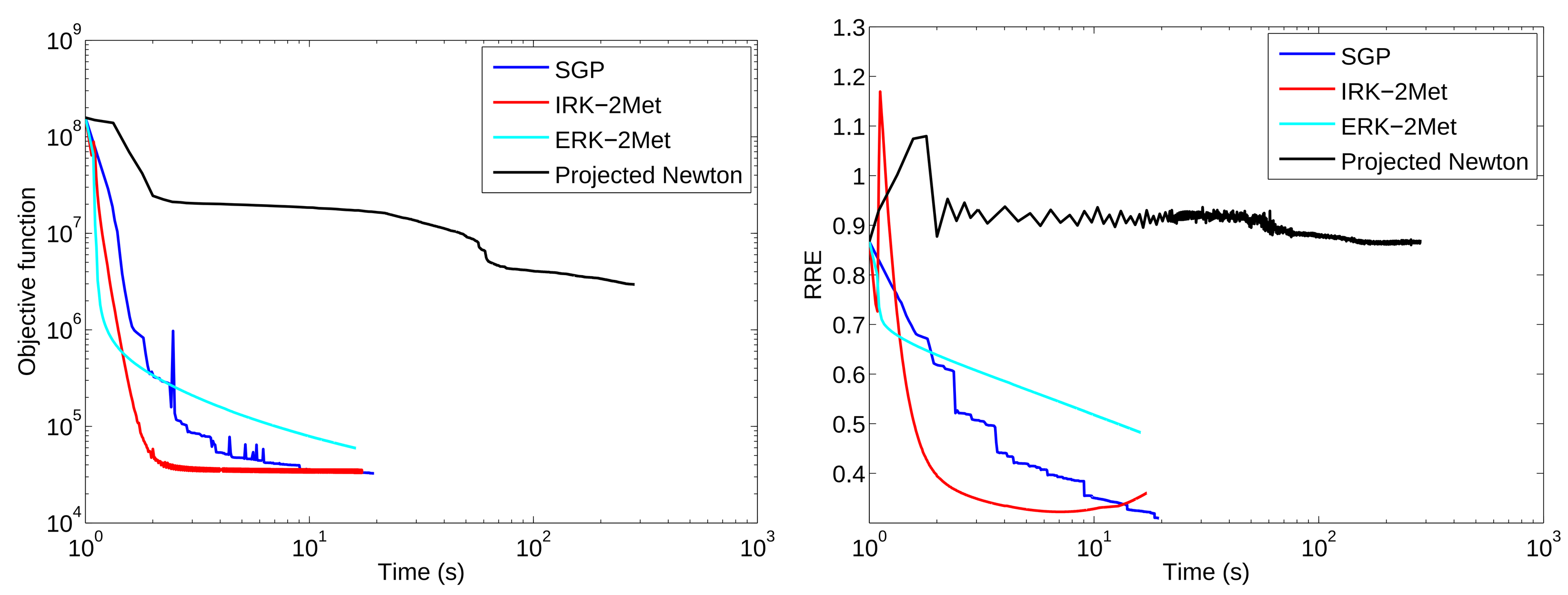
$$\min_{\boldsymbol{x}\geq 0} \sum_{i=1}^N \left\{ \boldsymbol{y}_i \ln \frac{\boldsymbol{y}_i}{(H\boldsymbol{x}+\boldsymbol{b})_i} + (H\boldsymbol{x}+\boldsymbol{b})_i - \boldsymbol{y}_i \right\}$$

Since H is a BCCB matrix, it can be factorized as $\boldsymbol{H} = \boldsymbol{F}^*\boldsymbol{\Lambda}\boldsymbol{F}$, where $\boldsymbol{F}$ is the 2D DFT, and the descent direction can be computed through the FFT.

Numerical results: comparison with SGP method [2]



Dataset



Performance of the compared algorithms

Algorithm (2) with either $\boldsymbol{D}^{ERK}_\ell$ **ERK-2Met** or $\boldsymbol{D}^{IRK}_\ell$ **IRK-2Met** shows good and stable performance.

References

[1] Vogel, C. R.: Computational Methods for Inverse Problems, SIAM, Philadelphia, 2002.

[2] Bonettini, S., Zanella, R. and Zanni, L.: A scaled gradient projection method for constrained image deblurring. Inverse Problems (2009)

OIP 2016 - Optimization Techniques for Inverse Problems III; September 19-21, 2016 - Modena, Italy