Optimization-based Image & Signal Processing

Tarmizi bin Adam, Ph.D.
Faculty of Computing

December 11, 2023

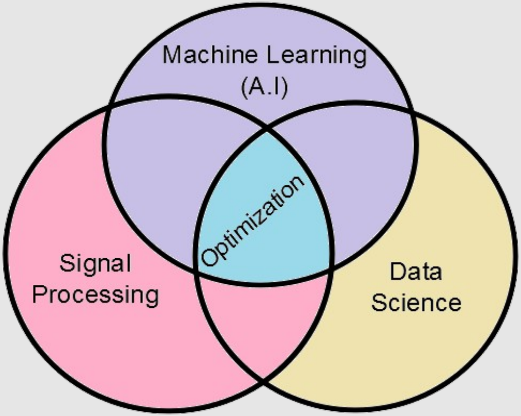
Outline

Introduction

This talk will be

- A non exhaustive overview of some algorithms.
- Avoiding **a lot** of mathematical details.
- Focusing only on signal/image processing applications.

Mathematical Optimization[1] [2] [3]



[1] L. Bottou, F. E. Curtis, and J. Nocedal, “Optimization methods for large-scale machine learning,” *Siam Review*, vol. 60, no. 2, pp. 223–311, 2018

[2] Z.-Q. Luo and W. Yu, “An introduction to convex optimization for communications and signal processing,” *IEEE Journal on selected areas in communications*, vol. 24, no. 8, pp. 1426–1438, 2006

[3] S. J. Wright and B. Recht, *Optimization for data analysis*. Cambridge University Press, 2022

Mathematical Optimization

Main problem

- In mathematical optimization we are interested in the following unconstrained optimization problem

$$\underset{\mathbf{x}\in\mathbb{R}^n}{\text{minimize}}\, F(\mathbf{x}), \tag{1}$$

  where $F : \mathbb{R}^n \to \mathbb{R}$ is a smooth convex function.
- First-order algorithms are usually the choice for solving (1) for large scale problems.
- Gradient Descent (GD)[4] and its many variants are standard choices.

[4] H. Li, C. Fang, and Z. Lin, "Accelerated first-order optimization algorithms for machine learning," *Proceedings of the IEEE*, vol. 108, no. 11, pp. 2067–2082, 2020

Gradient Descent

Augustin-Louis Cauchy

- Developed in the 19th century by Augustin-Louis Cauchy[5].
- Motivated by astronomy for calculating orbits of heavenly bodies.
- Method of choice for low or medium accuracy solutions due to low per-iteration cost[6].

[5] C. Lemaréchal, "Cauchy and the gradient method," *Doc Math Extra*, vol. 251, no. 254, p. 10, 2012

[6] V. Cevher, S. Becker, and M. Schmidt, "Convex optimization for big data: Scalable, randomized, and parallel algorithms for big data analytics," *IEEE Signal Processing Magazine*, vol. 31, no. 5, pp. 32–43, 2014

Gradient Descent

Main idea

- Consider the problem

$$\underset{\mathbf{x}\in\mathbb{R}^n}{\text{minimize}}\, F(\mathbf{x})\,. \tag{2}$$

- The GD iteration

$$\mathbf{x}_{k+1} = \mathbf{x}_k - \eta \nabla F(\mathbf{x}_k)\,, \quad k = 1, 2, \cdots \tag{3}$$

where $\eta > 0$ is the step size (learning rate) and ∇F is the gradient of the function f.

Gradient Descent

- Complete algorithm for the GD

Algorithm 1: Gradient Descent Algorithm

Initialize: $\mathbf{x}_0 \in \mathbb{R}^n$, $\eta > 0$

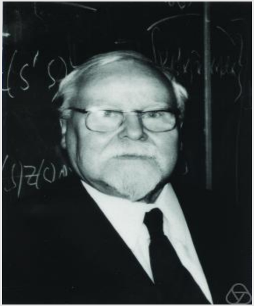
while *not converged* **do**

  $\mathbf{x}_{k+1} = \mathbf{x}_k - \eta \nabla F\left(\mathbf{x}_k\right); \quad k = 1, 2, \cdots$

  $k = k + 1$

Gradient Descent

Tikhonov regularized image denoising



Andrey Nikolayevich Tikhonov

- Image denoising as minimizing a Tikhonov functional

$$\underset{\mathbf{x}\in\mathbb{R}^n}{minimize}\, F(\mathbf{x}) = \frac{1}{2}\|\mathbf{x}-\mathbf{y}\|_2^2 + \lambda\|\mathbf{D}\mathbf{x}\|_2^2, \tag{4}$$

  where $\lambda > 0$ is regularization parameter which effects the noise filtering.
- The Tikhonov functional is smooth and convex

Gradient Descent

Tikhonov regularized image denoising

- The gradient of the Tikhonov functional,

$$\nabla F(\mathbf{x}_k) = \mathbf{x}_k - \mathbf{y} + 2\lambda \mathbf{D}^\top \mathbf{D}\mathbf{x}_k$$

- Thus, the GD step for Tikhonov image denoising,

$$\mathbf{x}_{k+1} = \mathbf{x}_k - \eta \left(\mathbf{x}_k - \mathbf{y} + 2\lambda \mathbf{D}^\top \mathbf{D}\mathbf{x}_k \right); \quad k = 1, 2, \cdots \tag{5}$$

Gradient Descent

Other Variants & Improving convergence

- We can improve the convergence of GD by *momentum* methods
  - Heavy-Ball (HB) method[7]

$$\mathbf{x}_{k+1} = \mathbf{x}_k - \eta \nabla F(\mathbf{x}_k) + \beta(\mathbf{x}_t - \mathbf{x}_{t-1}) \tag{6}$$

  - Nesterov acceleration[8]

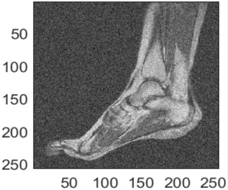
$$\begin{aligned} \mathbf{x}_{k+1} &= \mathbf{y}_k - \eta \nabla F(\mathbf{y}_k) \\ t_{k+1} &= \frac{t_k - 1}{t_k + 2} \\ \mathbf{y}_{k+1} &= \mathbf{x}_{k+1} + t_{k+1}(\mathbf{x}_{k+1} - \mathbf{x}_k) \end{aligned} \tag{7}$$

[7] B. T. Polyak, "Some methods of speeding up the convergence of iteration methods," *Ussr computational mathematics and mathematical physics*, vol. 4, no. 5, pp. 1–17, 1964

[8] Y. E. Nesterov, "A method for solving the convex programming problem with convergence rate," in *Dokl. Akad. Nauk SSSR,*, vol. 269, 1983, pp. 543–547

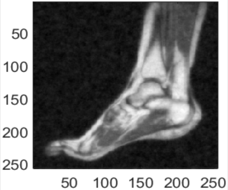
Gradient Descent

Tikhonov regularized image denoising

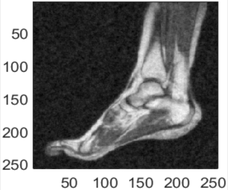


(a) Noisy (PSNR = 17.35 dB)

(b) GD (PSNR = 27.43 dB)



(c) HB (PSNR = 27.42 dB)



(d) Nesterov (PSNR = 27. 42 dB)

Figure: Noise level of $\sigma = 0.13$ with regularization parameters λ between $(1, 2.4)$,

Total Variation Denoising

Stanley Osher

- Tikhonov regularized image denoising tends to over smooth the image. We need a better model!
- A classical and very popular model is the Total Variation (TV) denoising[9] model:

$$\underset{\mathbf{x}\in\mathbb{R}^n}{\text{minimize}}\ F(\mathbf{x}) = \frac{1}{2}\|\mathbf{x}-\mathbf{y}\|_2^2 + \lambda\|\mathbf{D}\mathbf{x}\|_1, \tag{8}$$

- Widely used due to its edge preserving capabilities.

[9] L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D: Nonlinear Phenomena*, vol. 60, no. 1, pp. 259–268, 1992

Total Variation Denoising

$$\underset{\mathbf{x}\in\mathbb{R}^n}{\text{minimize}}\, F(\mathbf{x}) = \frac{1}{2}\|\mathbf{x}-\mathbf{y}\|_2^2 + \lambda\|\mathbf{D}\mathbf{x}\|_1, \qquad (9)$$

- The main problem with TV model is the non-smooth ℓ_1-norm regularization.
- The overall TV model is a smooth + non-smooth model. Thus, a non-differentiable optimization model.
- We need more involved optimization techniques for this.

Alternating Direction Method of Multipliers

(ADMM)

- The ADMM is a widely used optimization algorithm[10]. Its origins date back to the 70's[11 12].
- The composite minimization problem

$$\begin{aligned} &\underset{\mathbf{x},\mathbf{v}}{\text{minimize}} \quad f(\mathbf{x}) + g(\mathbf{v}) \\ &\text{s.t} \quad \mathbf{Ax} + \mathbf{Bv} = \mathbf{z}, \end{aligned} \tag{10}$$

  where $\mathbf{x}, \mathbf{v} \in \mathbb{R}^{n\times 1}$, $\mathbf{A}, \mathbf{B} \in \mathbb{R}^{n\times n}$ and $f(\cdot), g(\cdot) : \mathbb{R}^{n\times 1} \to \mathbb{R}$ are convex functions. The function $g(\cdot)$ is possibly non-smooth.
- ADMM is well suited for TV denoising.

[10] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Foundations and Trends® in Machine Learning*, vol. 3, no. 1, pp. 1–122, 2011

[11] D. Gabay and B. Mercier, "A dual algorithm for the solution of nonlinear variational problems via finite element approximation," *Computers & Mathematics with Applications*, vol. 2, no. 1, pp. 17–40, 1976

[12] ,R. Glowinski and A. Marroco, "Sur l'approximation, par éléments finis d'ordre un, et la résolution, par pénalisation-dualité d'une classe de problèmes de dirichlet non linéaires," *Revue française d'automatique, informatique, recherche opérationnelle. Analyse numérique*, vol. 9, no. R2, pp. 41–76, 1975

ADMM

Total Variation Denoising

- The TV in ADMM form (10)

$$\begin{aligned} &\underset{\mathbf{x}}{\text{minimize}} \quad \frac{1}{2}\|\mathbf{x}-\mathbf{y}\|_2^2 + \lambda\|\mathbf{v}\|_1, \\ &\text{s.t} \quad \mathbf{v} = \mathbf{D}\mathbf{x}, \end{aligned} \tag{11}$$

when $\mathbf{A} = \mathbf{D}$, $\mathbf{B} = -\mathbf{I}$, and $\mathbf{z} = \mathbf{0}$ in (10) thus, amenable to ADMM.

ADMM

Total Variation Denoising

- The $\mathcal{L}_A$ function for (11)

$$\mathcal{L}_{\mathcal{A}}(\mathbf{x}, \mathbf{v}, \mu) = \frac{1}{2}\|\mathbf{x}-\mathbf{y}\|_2^2 + \lambda\|\mathbf{v}\|_1 - \langle \mu, \mathbf{v} - \mathbf{D}\mathbf{x}\rangle + \frac{\rho}{2}\|\mathbf{v}-\mathbf{D}\mathbf{x}\|_2^2 \tag{12}$$

- Alternate minimization

$$\begin{cases} \mathbf{x}^{k+1} = \arg\min_{\mathbf{x}} \frac{1}{2}\|\mathbf{x}-\mathbf{y}\|_2^2 - \langle \mu, \mathbf{v}^k - \mathbf{D}\mathbf{x}\rangle + \frac{\rho}{2}\|\mathbf{v}^k - \mathbf{D}\mathbf{x}\|_2^2, \\ \mathbf{v}^{k+1} = \arg\min_{\mathbf{v}} \frac{\rho}{2}\|\mathbf{v} - \mathbf{D}\mathbf{x}^{k+1}\|_2^2 + \lambda\|\mathbf{v}\|_1 - \langle \mu, \mathbf{v} - \mathbf{D}\mathbf{x}^{k+1}\rangle, \\ \mu^{k+1} = \mu^k + \rho\left(\mathbf{v}^{k+1} - \mathbf{D}\mathbf{x}^{k+1}\right). \end{cases} \tag{13}$$

ADMM

Total Variation Denoising

- Complete ADMM total variation image denoising

Algorithm 2: ADMM for TV image denoising.

Initialize: $\mathbf{x}^0$, $\mathbf{v}^0$, μ^0, $k = 0$, $\rho > 0$, and $\lambda > 0$

while *not converged* **do**

$\mathbf{x}^{k+1} = \arg\min_{\mathbf{x}} \frac{1}{2}\|\mathbf{x} - \mathbf{y}\|_2^2 + \frac{\rho}{2}\|\mathbf{v}^k - \mathbf{D}\mathbf{x} + \frac{\mu^k}{\rho}\|_2^2$ `// solved via conjugate gradient (CG)`

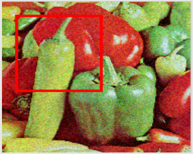
$\mathbf{v}^{k+1} = \arg\min_{\mathbf{v}} \lambda\|\mathbf{v}\|_1 + \frac{\rho}{2}\|\mathbf{v} - \mathbf{D}\mathbf{x}^{k+1} + \frac{\mu^k}{\rho}\|_2^2$

$\mu^{k+1} = \mu^k + \rho\left(\mathbf{v}^{k+1} - \mathbf{D}\mathbf{x}^{k+1}\right)$

$k = k + 1$

ADMM

Total Variation Denoising



(a) Noisy, $\sigma = 20$

(b) Zoomed

(c) ADMM denoised

(d) Noisy, $\sigma = 40$

(e) Zoomed

(f) ADMM denoised

Figure: Noise level of $\sigma = 20$ and $\sigma = 40$ with regularization parameter $\lambda = 8$ and $\lambda = 25$ respectively. Value of $\rho = 0.3$

ADMM

Total Variation Denoising

(a) Noisy, $\sigma = 80$

(b) Zoomed

(c) ADMM denoised

Figure: Noise level of $\sigma = 80$ with regularization parameter $\lambda = 50$. Value of $\rho = 0.3$

Time-series Smoothing

- Applications in fields such as climatology, macroeconomics, environmental science, and finance. In finance and economics, it is called trend filtering [13]
- Given a time-series data point $f_t,\ t = 1, \cdots, n$ assumed to consist of two components, a slow varying trend u_t and a rapidly varying random component z_t, estimate the slowly varying component u_t such that $z_t = f_t - u_t$ i.e., the error is small as possible.
- The slow varying component u_t is estimated by the j^{th} order smoothing by solving the following optimization problem

[13] R. Gençay, F. Selçuk, and B. J. Whitcher, *An introduction to wavelets and other filtering methods in finance and economics*. Elsevier, 2001

Time-series Smoothing

- The time-series smoothing problem

$$\underset{\mathbf{u}\in\mathbb{R}^n}{\text{minimize}}\ \frac{1}{2}\|\mathbf{u}-\mathbf{f}\|_2^2+\lambda\|\mathbf{D}^{j+1}\mathbf{u}\|_1, \tag{14}$$

where the scalar λ is the smoothing parameter, $\mathbf{D}^{j+1}\in\mathbb{R}^{(n-j)\times n}$ is the j^{th} order discrete difference operator.

Tseng's Alternating Minimization Algorithm

(AMA)

- For time-series smoothing, we can use Tseng's Alternating Minimization Algorithm[14] (AMA).
- One advantage, does not need to solve a linear system, unlike ADMM.
- Different from ADMM, in each iteration, AMA minimizes both the Lagrangian and the augmented Lagrangian.

[14] P. Tseng, "Applications of a splitting algorithm to decomposition in convex programming and variational inequalities," *SIAM Journal on Control and Optimization*, vol. 29, no. 1, pp. 119–138, 1991

AMA

Time-series Smoothing

- The AMA also solves optimization problems of the form (10). For the time-series smoothing, the complete AMA algorithm is

Algorithm 3: AMA (Time-series smoothing)

Initialize: $\lambda > 0$, $\tau > 0$, $\mathbf{u}_0$, $\mathbf{v}_0$, $\rho_0 > 0$, $k = 0$

while *not converged* **do**

$$\mathbf{u}_{k+1} = \underset{\mathbf{u}\in\mathbb{R}^m}{\operatorname{argmin}} \frac{1}{2}\|\mathbf{u} - \mathbf{f}\|_2^2 + \langle \rho_k, \mathbf{v}_k - \mathbf{D}^2\mathbf{u}\rangle,$$

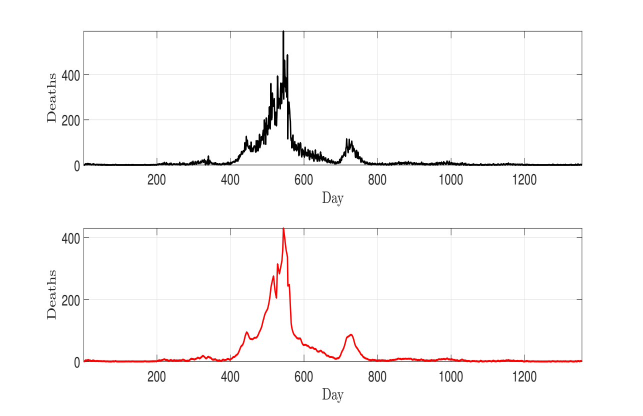
$$\mathbf{v}_{k+1} = \underset{\mathbf{v}\in\mathbb{R}^n}{\operatorname{argmin}} \lambda\|\mathbf{v}\|_1 + \langle \rho_k, \mathbf{v} - \mathbf{D}^2\mathbf{u}_{k+1}\rangle + \frac{\beta}{2}\|\mathbf{v} - \mathbf{D}^2\mathbf{u}_{k+1}\|_2^2,$$

$$\rho_{k+1} = \rho_k + \tau\left(\mathbf{v}_{k+1} - \mathbf{D}^2\mathbf{u}_{k+1}\right).$$

$$k = k + 1$$

AMA

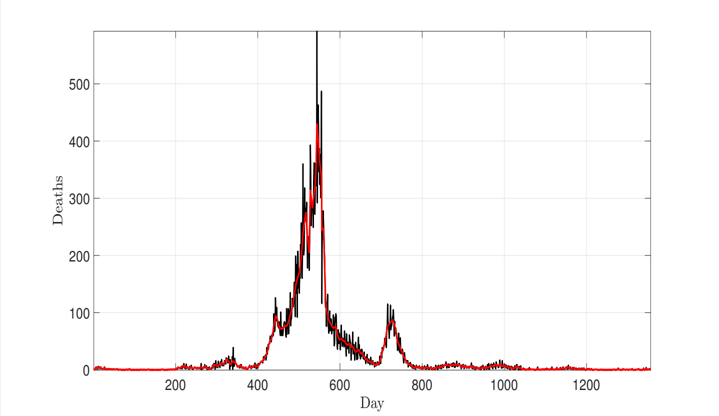
Time-series Smoothing



Malaysia Covid19 death per day (17/2/2020 - 2/12/2023)

AMA

Time-series Smoothing



Malaysia Covid19 death per day (17/2/2020 - 2/12/2023)

Summary

- Reviewed some popular mathematical optimization algorithm
  - Gradient Descent (GD) and some of its variants.
  - Alternating Direction Method of Multipliers (ADMM)
  - Tseng's Alternating Minimization Algorithm (AMA)
- Mathematical optimization is used in various applications in signal processing and image processing. Simple examples in this talk
  - Image denoising
  - Time-series smoothing

Summary

- Other mathematical optimization algorithms worth knowing (not discussed)
  - Proximal Gradient (PG) and accelerated PG methods[15].
  - Primal-Dual Hybrid Gradient (PDHG) methods[16] [17].
- Other applications (outside signal & image processing) not touched in this talk
  - Mesh processing in computer graphics [18]
  - Deep algorithm unrolling [19] [20] (provable deep learning)

---

[15] A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM journal on imaging sciences*, vol. 2, no. 1, pp. 183–202, 2009

[16] L. Condat, D. Kitahara, A. Contreras, and A. Hirabayashi, "Proximal splitting algorithms for convex optimization: A tour of recent advances, with new twists," *SIAM Review*, vol. 65, no. 2, pp. 375–435, 2023

[17] N. Komodakis and J.-C. Pesquet, "Playing with duality: An overview of recent primal-dual approaches for solving large-scale optimization problems," *IEEE Signal Processing Magazine*, vol. 32, no. 6, pp. 31–54, 2015

[18] T. Neumann, K. Varanasi, C. Theobalt, M. Magnor, and M. Wacker, "Compressed manifold modes for mesh processing," in *Computer Graphics Forum*, vol. 33, no. 5. Wiley Online Library, 2014, pp. 35–44

[19] Y. Yang, J. Sun, H. Li, and Z. Xu, "Admm-csnet: A deep learning approach for image compressive sensing," *IEEE transactions on pattern analysis and machine intelligence*, vol. 42, no. 3, pp. 521–538, 2018

[20] V. Monga, Y. Li, and Y. C. Eldar, "Algorithm unrolling: Interpretable, efficient deep learning for signal and image processing," *IEEE Signal Processing Magazine*, vol. 38, no. 2, pp. 18–44, 2021

Thank You !
Q&A

Codes to reproduce the figures can be obtained at my Github: https://github.com/tarmiziAdam2005/2023-Workshop-on-Metaverse