UTM
UNIVERSITI TEKNOLOGI MALAYSIA
RESEARCH UNIVERSITY

The Alternating Direction Method of Multipliers (ADMM) For Large Scale Convex Optimization Problem: Applications in Image and Signal Processing

Dr. Tarmizi bin Adam

tarmizi.adam@utm.my
https://people.utm.my/tarmizi/

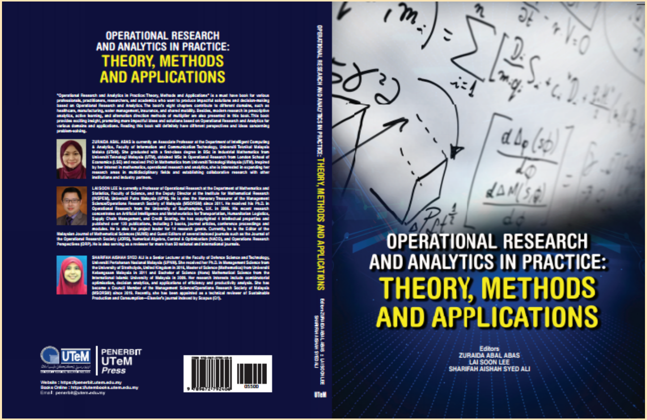
January 31, 2023

Ice Breaking

- Affiliated with ViCubeLab.
- Educational background: B.Eng (CIE) IIUM, MSc (CS) UTM, PhD (EE) UM
- Current research interest: Image and signal processing, convex & non-convex mathematical optimization.

This Talk

Based on our book chapter of the recently published book



Acknowledgements

Collaborators and authors of the chapter

Dr. Nur Syarafina Mohamed (FS)

Dr. Mohd Fikree Hassan (IUMW)

Prof. Dr. Raveendran Paramesran (UM)

Outline

- Introduction
  - Mathematical Preliminaries
  - Problem Setup
  - The ADMM
- Applications
  - Total Variation Image Denoising
  - Time-series Smoothing
- Conclusions

Introduction

The Big Data Deluge

- The amount of data collected in the digital age has been tremendously growing.

The Big Data Deluge

- The amount of data collected in the digital age has been tremendously growing.
- Some of these data are freely available and can be used without any fee.

The Big Data Deluge

- The amount of data collected in the digital age has been tremendously growing.
- Some of these data are freely available and can be used without any fee.
- *"With great power comes great responsibility"- Uncle Ben*

The Big Data Deluge

Problems and Challenges

"With big data comes big problems"

The Big Data Deluge

Problems and Challenges

- "How do we process these data?"

The Big Data Deluge

Problems and Challenges

- "How do we process these data?"
- "How do we predict future events or make the best decisions given the data at hand?"

UTM
UNIVERSITI TEKNOLOGI MALAYSIA
RESEARCH UNIVERSITY

The Big Data Deluge

Problems and Challenges

- Large data are rarely clean. The data collected contains noise and outliers.

UTM
UNIVERSITI TEKNOLOGI MALAYSIA
RESEARCH UNIVERSITY

The Big Data Deluge

Problems and Challenges

- Large data are rarely clean. The data collected contains noise and outliers.
- Incomplete data with missing values.

The Big Data Deluge

Problems and Challenges

- Large data are rarely clean. The data collected contains noise and outliers.
- Incomplete data with missing values.
- How do we extract/select only important feature in the data that will contribute to the intended task?

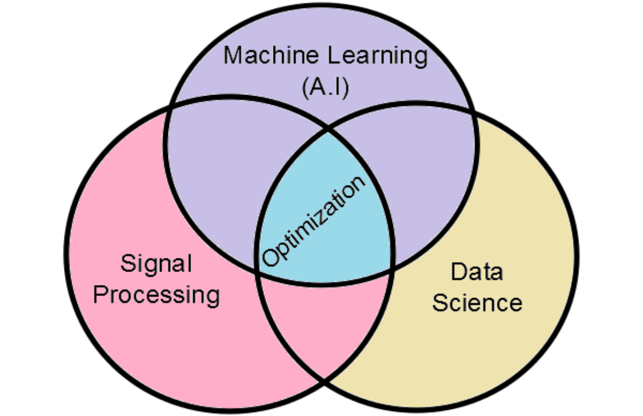
UTM
UNIVERSITI TEKNOLOGI MALAYSIA
RESEARCH UNIVERSITY

The Big Data Deluge

Problems and Challenges

The aforementioned questions and the example situations given are the driving force in Signal Processing (SP) and Machine Learning (ML).

Mathematical Optimization



Preliminaries

Mathematical Notations

- $\mathbb{R}^n :=$ Denotes the n-dimensional Euclidean space equipped with an inner product $\langle \cdot, \cdot \rangle$.

Preliminaries

Mathematical Notations

- $\mathbb{R}^n :=$ Denotes the n-dimensional Euclidean space equipped with an inner product $\langle \cdot, \cdot \rangle$
- $\mathbb{R}^{m \times n} :=$ The set of $m \times n$ matrices with real entries.

Preliminaries

Mathematical Notations

- $\mathbb{R}^n$:= Denotes the *n*-dimensional Euclidean space equipped with an inner product $\langle\cdot,\cdot\rangle$
- $\mathbb{R}^{m\times n}$:= The set of $m \times n$ matrices with real entries.
- Bold lowercase and uppercase letters denote vectors and matrices i.e., $\mathbf{u}$, $\mathbf{X}$. Hadamard product of vectors and matrices is denoted by $\odot$.

Preliminaries

Mathematical Notations

- $\mathbb{R}^n :=$ Denotes the n-dimensional Euclidean space equipped with an inner product $\langle \cdot, \cdot \rangle$
- $\mathbb{R}^{m \times n} :=$ The set of $m \times n$ matrices with real entries.
- Bold lowercase and uppercase letters denote vectors and matrices i.e., $\mathbf{u}$, $\mathbf{X}$. Hadamard product of vectors and matrices is denoted by $\odot$.
- The ℓ_1 and ℓ_2, norms are defined as

$$\|\mathbf{x}\|_1 := \sum_i |x_i|, \quad \|\mathbf{x}\|_2 := \left(\sum_i |x_i|^2 \right)^{\frac{1}{2}}$$

Preliminaries

Mathematical Notations

- The proximal operator is defined as

$$\text{prox}_{\mu,\, g}\left(\mathbf{x}\right) = \underset{\mathbf{z}}{\text{argmin}}\, \frac{1}{2\mu}\|\mathbf{z}-\mathbf{x}\|_2^2 + g\left(\mathbf{x}\right).$$

Preliminaries

Mathematical Notations

- The proximal operator is defined as

$$\text{prox}_{\mu,\, g}\left(\mathbf{x}\right) = \underset{\mathbf{z}}{\text{argmin}}\, \frac{1}{2\mu}\|\mathbf{z}-\mathbf{x}\|_2^2 + g\left(\mathbf{x}\right).$$

- When $g\left(\mathbf{x}\right) = \|\mathbf{x}\|_1$, the solution is "simple" with closed form solution [1]

$$\begin{aligned} \text{prox}_{\mu,\, \|\mathbf{x}\|_1}\left(\mathbf{x}\right) &= \underset{\mathbf{z}}{\text{argmin}}\, \frac{1}{2\mu}\|\mathbf{z}-\mathbf{x}\|_2^2 + \|\mathbf{x}\|_1, && (1) \\ &= \text{sgn}\left(\mathbf{x}\right) \odot \max\{|\mathbf{x}| - \mu,\, 0\}. && (2) \end{aligned}$$

[1] C. A. Micchelli, L. Shen, and Y. Xu, “Proximity algorithms for image models: denoising,” *Inverse Problems*, vol. 27, no. 4, p. 045009, 2011

Problem Setup

Composite Optimization Problem

What type of optimization problem ADMM solves?

- The composite minimization problem

$$\underset{\mathbf{x}\in\mathbb{R}^n}{\text{minimize}}\; f(\mathbf{x}) + \lambda g(\mathbf{x}), \tag{3}$$

where $f : \mathbb{R}^n \to \mathbb{R}$ is smooth and convex, $g : \mathbb{R}^n \to \mathbb{R}$ is convex but nonsmooth.

Problem Setup

Composite Optimization Problem

What type of optimization problem ADMM solves?

- The composite minimization problem

$$\underset{\mathbf{x}\in\mathbb{R}^n}{\text{minimize}}\ f(\mathbf{x}) + \lambda g(\mathbf{x}), \tag{4}$$

  where $f : \mathbb{R}^n \to \mathbb{R}$ is smooth and convex, $g : \mathbb{R}^n \to \mathbb{R}$ is convex but nonsmooth.
- Composite minimization problem (4) are abundant in Image Processing (IP), SP, and ML.

Problem Setup

Composite Optimization Problem

Table: Some popular optimization models that falls into the category of model (4). Regularized logistic regression (RLR) , Noisy matrix completion (NMC) , Robust principal component analysis (RPCA) , Regularized least-squares (RLS) , Discrete total variation minimization (DTVM) , and ℓ_1 Trend filtering (ℓ_1-TF).

Model	$f(\cdot)$	$g(\cdot)$
NMC [2]	$\frac{1}{2}\lVert \mathbf{Y}_\Omega - \mathbf{X}_\Omega \rVert_2^2$	$\lambda \lVert \mathbf{X} \rVert_*$
RPCA [3]	$\lVert \mathbf{L} \rVert_*$	$\lambda \lVert \mathbf{S} \rVert_1$
DTVM [4]	$\frac{1}{2}\lVert \mathbf{x} - \mathbf{y} \rVert_2^2$	$\lambda \lVert \mathbf{Dx} \rVert_1$
ℓ_1-TF [5]	$\frac{1}{2}\lVert \mathbf{x} - \mathbf{y} \rVert_2^2$	$\lambda \lVert \mathbf{D}^{k+1}\mathbf{x} \rVert_1,\ k \geq 0$

[2] K.-C. Toh and S. Yun, "An accelerated proximal gradient algorithm for nuclear norm regularized linear least squares problems," *Pacific Journal of optimization*, vol. 6, no. 615-640, p. 15, 2010

[3] E. J. Candès, X. Li, Y. Ma, and J. Wright, "Robust principal component analysis?" *Journal of the ACM (JACM)*, vol. 58, no. 3, pp. 1–37, 2011

[4] A. Chambolle and T. Pock, "An introduction to continuous optimization for imaging," *Acta Numerica*, vol. 25, pp. 161–319, 2016

[5] A. Ramdas and R. J. Tibshirani, "Fast and flexible admm algorithms for trend filtering," *Journal of Computational and Graphical Statistics*, vol. 25, no. 3, pp. 839–858, 2016

Problem Setup

Composite Optimization Problem

Table: Some popular optimization models that falls into the category of model (4). Regularized logistic regression (RLR) , Noisy matrix completion (NMC) , Robust principal component analysis (RPCA) , Regularized least-squares (RLS) , Discrete total variation minimization (DTVM) , and ℓ_1 Trend filtering (ℓ_1-TF).

Model	$f(\cdot)$	$g(\cdot)$
NMC [6]	$\frac{1}{2}\\|\mathbf{Y}_\Omega - \mathbf{X}_\Omega\\|_2^2$	$\lambda\\|\mathbf{X}\\|_*$
RPCA [7]	$\\|\mathbf{L}\\|_*$	$\lambda\\|\mathbf{S}\\|_1$
DTVM [8]	$\frac{1}{2}\\|\mathbf{x} - \mathbf{y}\\|_2^2$	$\lambda\\|\mathbf{D}\mathbf{x}\\|_1$
ℓ_1-TF [9]	$\frac{1}{2}\\|\mathbf{x} - \mathbf{y}\\|_2^2$	$\lambda\\|\mathbf{D}^{k+1}\mathbf{x}\\|_1,\ k \geq 0$

[6] K.-C. Toh and S. Yun, “An accelerated proximal gradient algorithm for nuclear norm regularized linear least squares problems,” *Pacific Journal of optimization*, vol. 6, no. 615-640, p. 15, 2010

[7] E. J. Candès, X. Li, Y. Ma, and J. Wright, “Robust principal component analysis?” *Journal of the ACM (JACM)*, vol. 58, no. 3, pp. 1–37, 2011

[8] A. Chambolle and T. Pock, “An introduction to continuous optimization for imaging,” *Acta Numerica*, vol. 25, pp. 161–319, 2016

[9] A. Ramdas and R. J. Tibshirani, “Fast and flexible admm algorithms for trend filtering,” *Journal of Computational and Graphical Statistics*, vol. 25, no. 3, pp. 839–858, 2016

Problem Setup

Composite Optimization Problem

Table: Some popular optimization models that falls into the category of model (4). Regularized logistic regression (RLR) , Noisy matrix completion (NMC) , Robust principal component analysis (RPCA) , Regularized least-squares (RLS) , Discrete total variation minimization (DTVM) , and ℓ_1 Trend filtering (ℓ_1-TF).

Model	$f(\cdot)$	$g(\cdot)$
NMC [10]	$\frac{1}{2}\|\mathbf{Y}_\Omega - \mathbf{X}_\Omega\|_2^2$	$\lambda\|\mathbf{X}\|_*$
RPCA [11]	$\|\mathbf{L}\|_*$	$\lambda\|\mathbf{S}\|_1$
DTVM [12]	$\frac{1}{2}\|\mathbf{x} - \mathbf{y}\|_2^2$	$\lambda\|\mathbf{Dx}\|_1$
ℓ_1-TF [13]	$\frac{1}{2}\|\mathbf{x} - \mathbf{y}\|_2^2$	$\lambda\|\mathbf{D}^{k+1}\mathbf{x}\|_1,\ k \geq 0$

[10] K.-C. Toh and S. Yun, "An accelerated proximal gradient algorithm for nuclear norm regularized linear least squares problems," *Pacific Journal of optimization*, vol. 6, no. 615-640, p. 15, 2010

[11] E. J. Candès, X. Li, Y. Ma, and J. Wright, "Robust principal component analysis?" *Journal of the ACM (JACM)*, vol. 58, no. 3, pp. 1–37, 2011

[12] A. Chambolle and T. Pock, "An introduction to continuous optimization for imaging," *Acta Numerica*, vol. 25, pp. 161–319, 2016

[13] A. Ramdas and R. J. Tibshirani, "Fast and flexible admm algorithms for trend filtering," *Journal of Computational and Graphical Statistics*, vol. 25, no. 3, pp. 839–858, 2016

UTM
UNIVERSITI TEKNOLOGI MALAYSIA
RESEARCH UNIVERSITY

Problem Setup

Composite Optimization Problem

Table: Some popular optimization models that falls into the category of model (4). Regularized logistic regression (RLR) , Noisy matrix completion (NMC) , Robust principal component analysis (RPCA) , Regularized least-squares (RLS) , Discrete total variation minimization (DTVM) , and ℓ_1 Trend filtering (ℓ_1-TF).

Model	$f(\cdot)$	$g(\cdot)$
NMC [14]	$\frac{1}{2}\|\mathbf{Y}_\Omega - \mathbf{X}_\Omega\|_2^2$	$\lambda\|\mathbf{X}\|_*$
RPCA [15]	$\|\mathbf{L}\|_*$	$\lambda\|\mathbf{S}\|_1$
DTVM [16]	$\frac{1}{2}\|\mathbf{x} - \mathbf{y}\|_2^2$	$\lambda\|\mathbf{Dx}\|_1$
ℓ_1-TF [17]	$\frac{1}{2}\|\mathbf{x} - \mathbf{y}\|_2^2$	$\lambda\|\mathbf{D}^{k+1}\mathbf{x}\|_1,\ k \geq 0$

[14] K.-C. Toh and S. Yun, "An accelerated proximal gradient algorithm for nuclear norm regularized linear least squares problems," *Pacific Journal of optimization*, vol. 6, no. 615-640, p. 15, 2010

[15] E. J. Candès, X. Li, Y. Ma, and J. Wright, "Robust principal component analysis?" *Journal of the ACM (JACM)*, vol. 58, no. 3, pp. 1–37, 2011

[16] A. Chambolle and T. Pock, "An introduction to continuous optimization for imaging," *Acta Numerica*, vol. 25, pp. 161–319, 2016

[17] A. Ramdas and R. J. Tibshirani, "Fast and flexible admm algorithms for trend filtering," *Journal of Computational and Graphical Statistics*, vol. 25, no. 3, pp. 839–858, 2016

The Alternating Direction Methods of Multipliers

General Problem

The ADMM solves problems of the following form,

$$\begin{aligned} \underset{\mathbf{x},\mathbf{v}}{\text{minimize}} \quad & f(\mathbf{x}) + g(\mathbf{v}) \\ \text{s.t} \quad & \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{v} = \mathbf{z}, \end{aligned} \tag{5}$$

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 Alternating Direction Methods of Multipliers

General Problem

The Augmented Lagrangian function for problem (5) is

$$\mathcal{L}_{\mathcal{A}}(\mathbf{x}, \mathbf{v}, \mu) = f(\mathbf{x}) + g(\mathbf{v}) + \langle \mu, \mathbf{Ax} + \mathbf{Bv} - \mathbf{z} \rangle + \frac{\rho}{2} \|\mathbf{Ax} + \mathbf{Bv} - \mathbf{z}\|_2^2. \tag{6}$$

The Alternating Direction Methods of Multipliers

General Problem

Goal: Find the saddle point of (6) by alternately minimizing with respect to each variable

$$\begin{cases} \mathbf{x}^{k+1} = \arg\min_{\mathbf{x}} f(\mathbf{x}) + \langle \mu, \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{v}^k - \mathbf{z} \rangle + \frac{\rho}{2} \|\mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{v}^k - \mathbf{z}\|_2^2, \\ \mathbf{v}^{k+1} = \arg\min_{\mathbf{v}} g(\mathbf{v}) + \langle \mu, \mathbf{A}\mathbf{x}^{k+1} + \mathbf{B}\mathbf{v} - \mathbf{z} \rangle + \frac{\rho}{2} \|\mathbf{A}\mathbf{x}^{k+1} + \mathbf{B}\mathbf{v} - \mathbf{z}\|_2^2, \\ \mu^{k+1} = \mu^k + \rho \left(\mathbf{A}\mathbf{x}^{k+1} + \mathbf{B}\mathbf{v}^{k+1} - \mathbf{z} \right), \end{cases} \tag{7}$$

where k is the k^{th} iteration.

The Alternating Direction Methods of Multipliers

General Problem

Algorithm 1: ADMM

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

while *not converged* **do**

$$\mathbf{x}^{k+1} = \arg\min_{\mathbf{x}} f(\mathbf{x}) + \frac{\rho}{2}\|\mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{v}^k - \mathbf{z} + \frac{\mu^k}{\rho}\|_2^2$$

$$\mathbf{v}^{k+1} = \arg\min_{\mathbf{v}} g(\mathbf{v}) + \frac{\rho}{2}\|\mathbf{A}\mathbf{x}^{k+1} + \mathbf{B}\mathbf{v} - \mathbf{z} + \frac{\mu^k}{\rho}\|_2^2$$

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

$$k = k + 1$$

Applications

Image Denoising

Degradation Model

- Assuming the following *degradation model*

$$\mathbf{y} = \mathbf{x} + \mathbf{n}, \tag{8}$$

$\mathbf{y} \in \mathbb{R}^n$ is the observed corrupted image, $\mathbf{x} \in \mathbb{R}^n$ is the clean image that is to be estimated, $\mathbf{n} \in \mathbb{R}^n$ is additive Gaussian noise.

Image Denoising

Total Variation Denoising

- A classical and very popular model is the Total Variation (TV) [18] [19] denoising model:

$$\underset{x}{\text{minimize}}\, F(x) = \int_{\Omega} (x-y)^2 + \lambda \int_{\Omega} |\nabla x|, \qquad (9)$$

[18] 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

[19] R. Chartrand and V. Staneva, "Total variation regularisation of images corrupted by non-gaussian noise using a quasi-newton method," *IET Image Processing*, vol. 2, no. 6, pp. 295–303, 2008

Image Denoising

Total Variation Denoising

- A classical and very popular model is the Total Variation (TV) denoising model:

$$\underset{x}{\text{minimize}}\, F(x) = \int_{\Omega} (x-y)^2 + \lambda \int_{\Omega} |\nabla x|, \qquad (9)$$

- Widely used due to its edge preserving capabilities.

Image Denoising

Total Variation Denoising

- The TV in vector notation

$$\underset{\mathbf{x}}{\text{minimize}}\, F(\mathbf{x}) = \frac{1}{2}\|\mathbf{x} - \mathbf{y}\|_2^2 + \lambda\|\mathbf{D}\mathbf{x}\|_1, \qquad (10)$$

Image Denoising

Total Variation Denoising

- The TV in vector notation

$$\underset{\mathbf{x}}{\text{minimize}}\, F(\mathbf{x}) = \frac{1}{2}\|\mathbf{x}-\mathbf{y}\|_2^2 + \lambda\|\mathbf{D}\mathbf{x}\|_1, \qquad (10)$$

- In the form of (5) with $\mathbf{A} = \mathbf{D}$, $\mathbf{B} = -\mathbf{I}$, and $\mathbf{z} = \mathbf{0}$ in (5) thus, amenable to ADMM.

Image Denoising

Total Variation Denoising

- The TV in vector notation

$$\underset{\mathbf{x}}{\text{minimize}}\, F(\mathbf{x}) = \frac{1}{2}\|\mathbf{x}-\mathbf{y}\|_2^2 + \lambda\|\mathbf{D}\mathbf{x}\|_1, \tag{10}$$

- In the form of (5) with $\mathbf{A} = \mathbf{D}$, $\mathbf{B} = -\mathbf{I}$, and $\mathbf{z} = \mathbf{0}$ in (5) thus, amenable to ADMM.
- The TV in ADMM form (5)

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

Image Denoising

Total Variation Denoising

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

- Next steps,
  - Construct the augmented Lagrangian function $\mathcal{L}_A$

Image Denoising

Total Variation Denoising

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

- Next steps,
  - Construct the augmented Lagrangian function $\mathcal{L}_A$
  - Alternate minimization of each subproblem

Image Denoising

Total Variation Denoising

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

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

Image Denoising

Total Variation Denoising

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

$$\mathcal{L}_{\mathcal{A}}(\mathbf{x}, \mathbf{v}, \mu) = \frac{1}{2}\|\mathbf{x}-\mathbf{y}\|_2^2 + \lambda\|\mathbf{v}\|_1 - \langle \mu, \mathbf{v} - \mathbf{D}\mathbf{x}\rangle + \frac{\rho}{2}\|\mathbf{v} - \mathbf{D}\mathbf{x}\|_2^2 \tag{13}$$

- Alternate minimization

$$\begin{cases} \mathbf{x}^{k+1} = \arg\min_{\mathbf{x}} \frac{1}{2}\|\mathbf{x}-\mathbf{y}\|_2^2 - \langle \mu, \mathbf{v}^k - \mathbf{D}\mathbf{x}\rangle + \frac{\rho}{2}\|\mathbf{v}^k - \mathbf{D}\mathbf{x}\|_2^2, \\ \mathbf{v}^{k+1} = \arg\min_{\mathbf{v}} \frac{\rho}{2}\|\mathbf{v} - \mathbf{D}\mathbf{x}^{k+1}\|_2^2 + \lambda\|\mathbf{v}\|_1 - \langle \mu, \mathbf{v} - \mathbf{D}\mathbf{x}^{k+1}\rangle, \\ \mu^{k+1} = \mu^k + \rho\left(\mathbf{v}^{k+1} - \mathbf{D}\mathbf{x}^{k+1}\right). \end{cases} \tag{14}$$

Image Denoising

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**

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

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

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

$k = k + 1$

Image Denoising

Total Variation Denoising

Some remarks on Algorithm 2

- The first-order difference matrix

$$\mathbf{D} = \begin{bmatrix} -1 & 1 & & & \\ & -1 & 1 & \mathbf{0} & \\ \mathbf{0} & & \ddots & \ddots & \\ & & & -1 & 1 \end{bmatrix}.$$

- The regularization parameter λ controls the amount of noise filtered.
- The algorithm can be terminated by several ways. The simplest is by a predefined number of iteration.

Image Denoising

Total Variation Denoising

Some remarks on Algorithm 2

- The minimization problem w.r.t $\mathbf{x}$ is a linear system. This can be solved using several iterations of *Conjugate Gradient (CG)* method.
- The minimization problem w.r.t $\mathbf{v}$ has a closed form solution via the ℓ_1-norm proximal operator [20]

$$\mathbf{v}^{k+1} = \text{sgn}\left(\mathbf{D}\mathbf{x}^{k+1} + \frac{\mu}{\rho}\right) \odot \max\{|\mathbf{D}\mathbf{x}^{k+1} + \frac{\mu}{\rho}| - \frac{\lambda}{\rho}, 0\}, \quad (15)$$

[20] C. A. Micchelli, L. Shen, and Y. Xu, “Proximity algorithms for image models: denoising,” *Inverse Problems*, vol. 27, no. 4, p. 045009, 2011

Image Denoising

Results

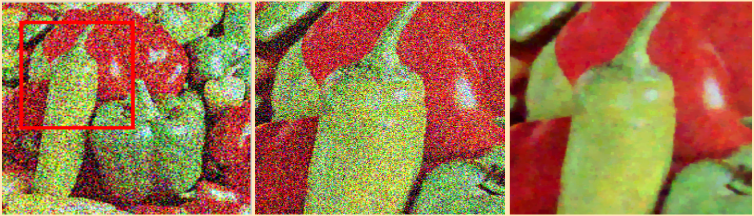
(a) Noisy, $\sigma = 20$ (b) Zoomed (c) ADMM denoised

(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 $\rho = 0.3$

Image Denoising

Results



(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

Problem Definition

- Applications in fields such as climatology, macroeconomics, environmental science, and finance. In finance and economics, it is called trend filtering [21]
- 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

[21] 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

Problem Definition

- The time-series smoothing problem posed as an ADMM optimization problem

$$\begin{aligned} & \underset{\mathbf{u}\in\mathbb{R}^n}{\text{minimize}} \ \frac{\lambda}{2}\|\mathbf{u}-\mathbf{f}\|_2^2 + \|\mathbf{D}^{j+1}\mathbf{u}\|_1 \\ & \text{s.t} \quad \mathbf{v} = \mathbf{D}^{j+1}\mathbf{u}. \end{aligned} \tag{16}$$

Time-series Smoothing

Problem Definition

- The time-series smoothing problem posed as an ADMM optimization problem

$$\begin{aligned} &\underset{\mathbf{u}\in\mathbb{R}^n}{\text{minimize}} \ \frac{\lambda}{2}\|\mathbf{u}-\mathbf{f}\|_2^2 + \|\mathbf{D}^{j+1}\mathbf{u}\|_1 \\ &\text{s.t} \quad \mathbf{v} = \mathbf{D}^{j+1}\mathbf{u}. \end{aligned} \tag{17}$$

- 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, $\mathbf{f} = [f_1, \cdots, f_n] \in \mathbb{R}^n$, and $\mathbf{u} = [u_1, \cdots, u_n] \in \mathbb{R}^n$. In this talk, $j = 1$ thus

$$\mathbf{D}^2 = \begin{bmatrix} 1 & -2 & 1 & & & & & \\ & 1 & -2 & 1 & & & & \\ & & 1 & -2 & 1 & & & \\ & & & & \ddots & \ddots & & \\ 1 & & & & & & 1 & -2 \\ -2 & 1 & & & & & & 1 \end{bmatrix}. \tag{18}$$

Time-series Smoothing

Problem Formulation

- The augmented Lagrangian $\mathcal{L}_{\mathcal{A}}$ of (17)

$$\mathcal{L}_{\mathcal{A}}(\mathbf{u},\mathbf{v},\mu)=\frac{\lambda}{2}\|\mathbf{u}-\mathbf{f}\|_2^2+\lambda\|\mathbf{v}\|_1-\langle\mu,\mathbf{v}-\mathbf{D}^2\mathbf{u}\rangle+\frac{\rho}{2}\|\mathbf{v}-\mathbf{D}^2\mathbf{u}\|_2^2, \tag{19}$$

- Alternating minimize

$$\begin{cases}\mathbf{u}^{k+1}=\arg\min_{\mathbf{u}}\frac{\lambda}{2}\|\mathbf{u}-\mathbf{f}\|_2^2-\langle\mu,\mathbf{v}-\mathbf{D}^2\mathbf{u}\rangle+\frac{\rho}{2}\|\mathbf{v}^k-\mathbf{D}^2\mathbf{u}\|_2^2,\\ \mathbf{v}^{k+1}=\arg\min_{\mathbf{v}}\frac{\rho}{2}\|\mathbf{v}-\mathbf{D}^2\mathbf{u}^{k+1}\|_2^2+\|\mathbf{v}\|_1-\langle\mu,\mathbf{v}-\mathbf{D}^2\mathbf{u}\rangle,\\ \mu^{k+1}=\mu^k+\rho\left(\mathbf{v}^{k+1}-\mathbf{D}^2\mathbf{u}^{k+1}\right).\end{cases} \tag{20}$$

Time-series Smoothing

Problem Formulation

Algorithm 3: ADMM for time-series smoothing

Initialize: $\mathbf{u}^0$, $\mathbf{f}^0$, μ^0, $k = 0$, and $\rho > 0$

while *not converged* **do**

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

```
// solved via conjugate gradient (CG)
```

$$\mathbf{v}^{k+1} = \arg\min_{\mathbf{v}} \|\mathbf{v}\|_1 + \frac{\rho}{2}\|\mathbf{v} - \left(\mathbf{D}^2\mathbf{u}^{k+1} + \frac{\mu^k}{\rho}\right)\|_2^2$$

$$\mu^{k+1} = \mu^k + \rho\left(\mathbf{v}^{k+1} - \mathbf{D}^2\mathbf{u}^{k+1}\right)$$

$$k = k + 1$$

Time-series Smoothing

Problem Formulation

Remarks on Algorithm 3

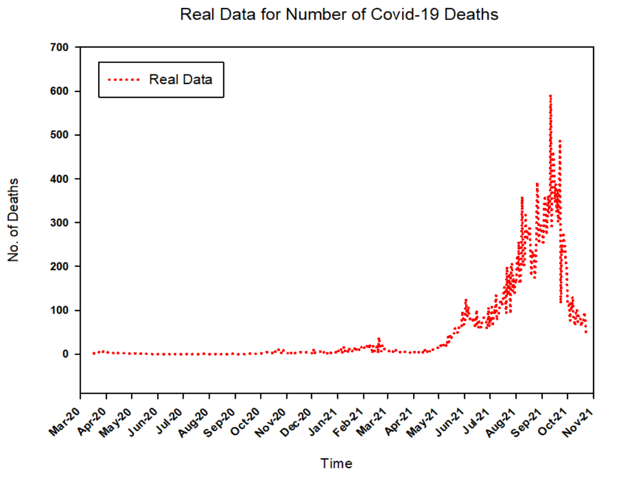
- The minimization problems w.r.t **u** and **v** are solved in a similar fashion as in the TV denoising.

UTM
UNIVERSITI TEKNOLOGI MALAYSIA
RESEARCH UNIVERSITY

Time-series Smoothing

Results

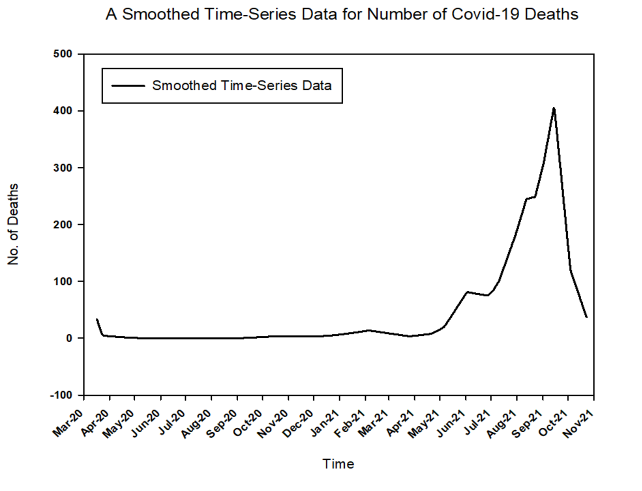
Smoothing daily death rate of COVID-19 time-series



Time-series Smoothing

Results

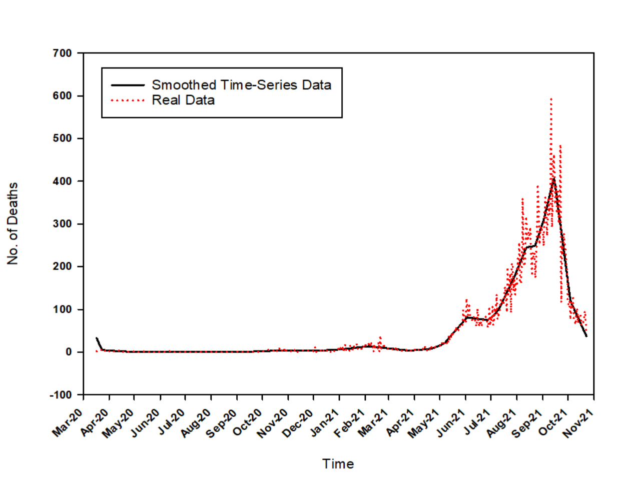
Smoothing via ADMM Algorithm 3 ($\lambda = 0.0035$ and $\rho = 1.35$)



Time-series Smoothing

Results

Real data vs Smoothed data



UTM
UNIVERSITI TEKNOLOGI MALAYSIA
RESEARCH UNIVERSITY

Conclusions

Take Home Messages

- The ADMM can be used in various applications in ML and SP. Simple examples in this talk
  - Image denoising
  - Time-series smoothing
- Applications not touched in this talk
  - Mesh processing in computer graphics [22]
  - Deep algorithm unrolling [23] [24] (provable deep learning)

---

[22] 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

[23] 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

[24] 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

UTM
UNIVERSITI TEKNOLOGI MALAYSIA
RESEARCH UNIVERSITY

Take Home Messages

- Some advantages of ADMM
  - Simple to formulate for a variety of applications. Sub minimization problems will usually have closed form solutions.
  - Implementation in code usually just takes several lines
- Steps to minimize composite optimization problems using ADMM
  1. Formulate the optimization problem into ADMM form (5)
  2. Construct the augmented Lagrangian $\mathcal{L}_{\mathcal{A}}$
  3. Alternately minimize each variables involved

Thank You !
Q&A

ADMM example codes can be obtained at my Github:
https://github.com/tarmiziAdam2005/Image-Signal-Processing