

Article

Structural Results on the HMLasso

Shin-ya Matsushita ^{1,*} and Hiromu Sasaki ²
¹ Department of Management Science and Engineering, Akita Prefectural University, Yurihonjo City 015-0055, Japan

² IT Development Center Akita, JTEKT, Akita City 010-0001, Japan; m25n009@akita-pu.ac.jp

* Correspondence: matsushita@akita-pu.ac.jp

Abstract

HMLasso (Lasso with High Missing Rate) is a useful technique for sparse regression when a high-dimensional design matrix contains a large number of missing data. To solve HMLasso, an appropriate positive semidefinite symmetric matrix must be obtained. In this paper, we present two structural results on the HMLasso problem. These results allow existing acceleration algorithms for strongly convex functions to be applied to solve the HMLasso problem.

Keywords: HMLass; smooth and strongly convex function; strongly convex FISTA

MSC: 65K05; 90C25

1. Introduction

Let X be an $n \times p$ design matrix and let y be an n -dimensional response. Consider the following standard linear regression model:

$$y = X\beta + \epsilon, \quad (1)$$

where ϵ is a noise term. The popular model focuses on the sparsity assumption of the regression vector. The Lasso [1] (Least absolute shrinkage and selection operator) is among the most popular procedures for estimating the unknown sparse regression vector in a high-dimensional linear model. Lasso is formulated as an ℓ^1 -penalized regression problem as follows:

$$\min_{\beta} \frac{1}{2n} \|y - X\beta\|_2^2 + \alpha \|\beta\|_1 \quad (2)$$

where $\alpha > 0$ is a regularization parameter and $\|\cdot\|_1$ (resp. $\|\cdot\|_2$) is the ℓ^1 (resp. ℓ^2) norm. Here, we consider the case where the design matrix X contains missing data. Missing data is prevalent and inevitable, which affects not only the representativeness and quality of data but also the results of Lasso regression. Therefore, it is crucial to develop a method applicable to process missing data. HMLasso [2] was proposed to effectively address this issue and is formulated as follows:

$$\min_{\Sigma} \frac{1}{2} \|W \odot (\Sigma - S^{\text{pair}})\|_F^2 \quad \text{s.t.} \quad \Sigma \succeq 0 \quad (3)$$

$$\min_{\beta} \frac{1}{2} \beta^\top \hat{\Sigma} \beta - \rho^{\text{pair}^\top} \beta + \alpha \|\beta\|_1, \quad (4)$$



Academic Editors: Wei-Shih Du, Hsien-Chung Wu and Manuel Arana-Jimenez

Received: 31 July 2025

Revised: 5 November 2025

Accepted: 12 November 2025

Published: 17 November 2025

Citation: Matsushita, S.-y.; Sasaki, H. Structural Results on the HMLasso. *Axioms* **2025**, *14*, 843. <https://doi.org/10.3390/axioms14110843>

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

where $\odot$ is the Hadamard product, $\|\cdot\|_F$ is the Frobenius norm, $S^{\text{pair}}, W \in \mathbb{R}^{p \times p}$, $\rho^{\text{pair}} \in \mathbb{R}^p$ is defined using X (see Section 2 for details), and $\hat{\Sigma} = \underset{\Sigma \succeq 0}{\operatorname{argmin}} \frac{1}{2} \|W \odot (\Sigma - S^{\text{pair}})\|_F^2$ (i.e., $\hat{\Sigma}$ is the solution to problem (3)). It is known that problem (4) can be equivalently written as Lasso (2), and in the literature dedicated to sparse regression, the most encountered algorithm is the proximal gradient method when dealing with a sum of two convex functions. Therefore, the key to solving HMLasso is to consider how fast and accurately problem (3) can be solved.

To deal with this challenge, we consider two structural results on the HMLasso problem. In the first, we show that the gradient of the objective function of (3) is Lipschitz continuous; in the second, we show that the objective function of (3) is strongly convex. These results guarantee that accelerated algorithms for strongly convex functions can be applied to solve (3). Finally, we conduct numerical experiments on the real data considered in [2]. The numerical results show that the accelerated algorithm for strongly convex functions is effective for solving problem (3).

2. Preliminaries

This section reviews basic definitions, facts, and notation that will be used throughout the paper.

Let $A, B \in \mathbb{R}^{n \times p}$. $\mathbb{S}_+^p$ denote the set of positive semidefinite matrices. Let $\mathcal{D} \subset \mathbb{R}^{p \times p}$. The indicator function $i_{\mathcal{D}}: \mathbb{R}^{p \times p} \rightarrow (-\infty, \infty]$ of $\mathcal{D}$ is defined by

$$i_{\mathcal{D}}(B) := \begin{cases} 0 & (B \in \mathcal{D}); \\ \infty & (\text{otherwise}). \end{cases}$$

Let $f: \mathbb{R}^{p \times p} \rightarrow (-\infty, \infty]$ be a proper and convex function. The proximal mapping $\operatorname{prox}_{\gamma f}(x)$ of f is defined by

$$\operatorname{prox}_f(\Sigma) := \underset{W \in \mathbb{R}^{p \times p}}{\operatorname{argmin}} \left\{ f(W) + \frac{1}{2} \|W - \Sigma\|_F^2 \right\}.$$

Let $\Sigma \in \mathbb{S}^p$. Then, there exist a diagonal matrix $\Lambda \in \mathbb{R}^{p \times p}$ and a $p \times p$ orthogonal matrix U such that $\Sigma = U\Lambda U^\top$. Let $P_{\mathbb{S}_+^p}$ be the matrix projection from $\mathbb{S}^p$ onto $\mathbb{S}_+^p$. Then, the following holds (see, for example, [3] (Example 29.31), [4] (Theorem 6.3)):

$$P_{\mathbb{S}_+^p}(\Sigma) = U\Lambda_+U^\top, \quad (5)$$

where Λ_+ is the diagonal matrix obtained from Λ by setting the negative entries to 0.

Let $X \in \mathbb{R}^{n \times p}$ be a design matrix. Set

$$I_{jk} := \{i : X_{ij} \text{ and } X_{ik} \text{ are observed}\}$$

and let n_{jk} be the number of elements of I_{jk} . We define matrices S^{pair} and W as follows:

$$(S^{\text{pair}})_{jk} := \begin{cases} \frac{1}{n_{jk}} \sum_{i \in I_{jk}} x_{ij}x_{ik} & (I_{jk} \neq \emptyset) \\ 0 & (I_{jk} = \emptyset) \end{cases}, \quad (W)_{jk} := \frac{n_{jk}}{n}. \quad (6)$$

We also define a vector ρ^{pair} as follows:

$$(\rho^{\text{pair}})_j := \begin{cases} \frac{1}{n_{jj}} \sum_{i \in I_{jj}} x_{ij}y_i & (I_{jj} \neq \emptyset) \\ 0 & (I_{jj} = \emptyset) \end{cases} \quad (7)$$

Let $L \in [0, \infty)$ and let $h: \mathbb{R}^{p \times p} \rightarrow \mathbb{R} \cup \{\infty\}$ be a differentiable function. The gradient ∇h of h is said to be L -Lipschitz continuous if

$$\|\nabla h(\Sigma_1) - \nabla h(\Sigma_2)\|_F \leq L \|\Sigma_1 - \Sigma_2\|_F \quad (\forall \Sigma_1, \Sigma_2 \in \mathbb{R}^{p \times p}). \quad (8)$$

This condition is often called L -smoothness in the literature. Let $h: \mathbb{R}^{p \times p} \rightarrow \mathbb{R}$ be L -smooth on $\mathbb{R}^{p \times p}$. Then, we can upper bound the function h as

$$h(\Sigma_1) \leq h(\Sigma_2) + \langle \nabla h(\Sigma_2), \Sigma_1 - \Sigma_2 \rangle + \frac{L}{2} \|\Sigma_1 - \Sigma_2\|_F^2 \quad (\forall \Sigma_1, \Sigma_2 \in \mathbb{R}^{p \times p}) \quad (9)$$

(see, for example, [5] (Theorem 5.8), [6] (Theorem A.1)).

Let $\mu \in (0, \infty)$ and let $g: \mathbb{R}^{p \times p} \rightarrow \mathbb{R} \cup \{+\infty\}$. g is said to be μ -strongly convex if for each $\Sigma_1, \Sigma_2 \in \mathbb{R}^{p \times p}$ and $\lambda \in (0, 1)$, we have

$$g(\lambda \Sigma_1 + (1 - \lambda) \Sigma_2) \leq \lambda g(\Sigma_1) + (1 - \lambda) g(\Sigma_2) - \frac{\mu}{2} \lambda (1 - \lambda) \|\Sigma_1 - \Sigma_2\|_F^2. \quad (10)$$

Suppose that g is μ -strongly convex and differentiable. Then, (10) is equivalent to the following condition:

$$\langle \nabla g(\Sigma_1) - \nabla g(\Sigma_2), \Sigma_1 - \Sigma_2 \rangle \geq \mu \|\Sigma_1 - \Sigma_2\|_F^2 \quad (\forall \Sigma_1, \Sigma_2 \in \mathbb{R}^{p \times p}). \quad (11)$$

3. Main Results

In this section, we present two structural results on the HMLasso problem (3).

Define $f(\Sigma) := \frac{1}{2} \|W \odot (\Sigma - S^{\text{pair}})\|_F^2$. In this case, the gradient ∇f of f is described as the following:

$$\nabla f(\Sigma) = \nabla \left(\frac{1}{2} \|W \odot (\Sigma - S^{\text{pair}})\|_F^2 \right) = W \odot W \odot (\Sigma - S^{\text{pair}}) \quad (12)$$

(see [2]).

3.1. Lipschitz Continuity

The first structural result deals with the gradient of the objective function in (3). We show the Lipschitz continuity of ∇f .

Lemma 1. *The gradient ∇f of f is Lipschitz continuous and its Lipschitz constant is $\|W \odot W\|_F$.*

Proof. Let $\bar{\alpha}$ be defined by $\bar{\alpha} := \sum_{j=1}^p \sum_{k=1}^p W_{jk}^4$. By (12), we obtain

$$\begin{aligned} \|\nabla f(\Sigma_1) - \nabla f(\Sigma_2)\|_F^2 &= \|W \odot W \odot (\Sigma_1 - S^{\text{pair}}) - W \odot W \odot (\Sigma_2 - S^{\text{pair}})\|_F^2 \\ &= \|W \odot W \odot (\Sigma_1 - \Sigma_2)\|_F^2 \\ &= \sum_{j=1}^p \sum_{k=1}^p \left(W_{jk}^2 (\Sigma_{1jk} - \Sigma_{2jk}) \right)^2. \end{aligned}$$

On the other hand,

$$\|W \odot W\|_F^2 \|\Sigma_1 - \Sigma_2\|_F^2 = \underbrace{\sum_{j=1}^p \sum_{k=1}^p W_{jk}^4}_{\bar{\alpha}} \cdot \sum_{j=1}^p \sum_{k=1}^p (\Sigma_{1jk} - \Sigma_{2jk})^2.$$

This implies

$$\begin{aligned} \|W \odot W\|_F^2 \|\Sigma_1 - \Sigma_2\|_F^2 - \|\nabla f(\Sigma_1) - \nabla f(\Sigma_2)\|_F^2 &= \sum_{j=1}^p \sum_{k=1}^p (\bar{\alpha} - W_{jk}^4) (\Sigma_{1jk} - \Sigma_{2jk})^2 \\ &\geq 0, \end{aligned}$$

where the last inequality follows from $\bar{\alpha} - W_{jk}^4 \geq 0$ for any j and k . Hence,

$$\|\nabla f(\Sigma_1) - \nabla f(\Sigma_2)\|_F \leq \|W \odot W\|_F \|\Sigma_1 - \Sigma_2\|_F.$$

□

3.2. Strong Convexity

We next consider the strong convexity of f . This is our second result.

Lemma 2. f is $\min_{l,m}(W \odot W)_{l,m}$ -strongly convex.

Proof. We show (11) with constant $\min_{l,m}(W \odot W)_{l,m}$. Let $\Sigma_1, \Sigma_2 \in \mathbb{R}^{p \times p}$.

$$\begin{aligned} \langle \nabla f(\Sigma_1) - \nabla f(\Sigma_2), \Sigma_1 - \Sigma_2 \rangle &= \langle W \odot W \odot (\Sigma_1 - \Sigma_2), \Sigma_1 - \Sigma_2 \rangle \\ &= \sum_{j=1}^p \sum_{k=1}^p W_{jk}^2 (\Sigma_{1jk} - \Sigma_{2jk})^2 \\ &\geq \sum_{j=1}^p \sum_{k=1}^p \min_{j,k} (W \odot W)_{j,k} (\Sigma_{1jk} - \Sigma_{2jk})^2 \\ &= \min_{l,m} (W \odot W)_{l,m} \|\Sigma_1 - \Sigma_2\|_F^2, \end{aligned}$$

where the second-to-last inequality follows from $W_{jk}^2 \geq \min_{l,m}(W \odot W)_{l,m}$ for any j and k . This implies that f is $\min_{l,m}(W \odot W)_{l,m}$ -strongly convex. □

Remark 1. From Lemmas 1 and 2, the objective function of (3) is strongly convex and has a Lipschitz gradient. It should be noted that we can guarantee a faster convergence than just a regular convex function if we know the objective function is strongly convex and smooth (see, for example, [6]).

4. Numerical Experiments

In this section, we consider a strongly convex variant of FISTA [6] (Algorithm 1) to solve problem (3).

4.1. Strongly Convex FISTA

Let $f: \mathbb{R}^{p \times p} \rightarrow \mathbb{R}$ be an L -smooth and μ -strongly convex function with $\text{dom}(f) = \mathbb{R}^{p \times p}$, and let $h: \mathbb{R}^{p \times p} \rightarrow \mathbb{R} \cup \{\infty\}$ be a convex function with $\{\Sigma : h(\Sigma) < \infty\} \neq \emptyset$. We consider the problem of minimizing the sum of f and h :

$$\min_{\Sigma \in \mathbb{R}^{p \times p}} f(\Sigma) + h(\Sigma). \quad (13)$$

In this setting, forward-backward splitting strategies were introduced as classical methods for solving (13). In the context of acceleration methods, the fast iterative shrinkage threshold algorithm (FISTA) was introduced by Beck and Teboulle [7] using the idea of forward-backward splitting. This topic is addressed in many references, and we refer to [3,5,6].

Here, we focus on the following strongly convex variant of FISTA involving backtracking investigated in [6].

Algorithm 1 Strongly convex FISTA [6]

Input: An initial point $\Sigma_0 \in \mathbb{R}^{p \times p}$, and initial estimate $L_0 > \mu$.

```

1: Initialize  $\Phi_0 = \Sigma_0$ ,  $t_0 = 0$ , and some  $\alpha > 1$ .
2: for  $k = 0, \dots$  do
3:    $L_{k+1} = L_k$ 
4:   while do
5:      $q_{k+1} = \mu / L_{k+1}$ 
6:      $t_{k+1} = \frac{2t_k + 1 + \sqrt{4t_k + 4q_{k+1}t_k^2 + 1}}{2(1 - q_{k+1})}$ 
7:     set  $\tau_k = \frac{(t_{k+1} - t_k)(1 + q_{k+1}t_k)}{t_{k+1} + 2q_{k+1}t_k t_{k+1} - q_k t_k^2}$  and  $\delta_k = \frac{t_{k+1} - t_k}{1 + q_{k+1}t_{k+1}}$ 
8:      $\Psi_k = \Sigma_k + \tau_k(\Phi_k - \Sigma_k)$ 
9:      $\Sigma_{k+1} = \text{prox}_{h/L_{k+1}}(\Psi_k - \frac{1}{L_{k+1}} \nabla f(\Psi_k))$ 
10:     $\Phi_{k+1} = (1 - q_{k+1}\delta_k)\Phi_k + q_{k+1}\delta_k\Psi_k + \delta_k(\Sigma_{k+1} - \Psi_k)$ 
11:    if  $f(\Sigma_{k+1}) \leq f(\Sigma_k) + \langle \nabla f(\Sigma_k), \Sigma_{k+1} - \Sigma_k \rangle + \frac{L_{k+1}}{2} \|\Sigma_{k+1} - \Sigma_k\|_F^2$  holds then
12:      break { $k$  will be incremented.}
13:    else
14:       $L_{k+1} = \alpha L_{k+1}$  {Recompute new  $L_{k+1}$ .}
15:    end if
16:  end while
17: end for

```

Output: An approximate solution Σ_{k+1}

Remark 2. The Lipschitz and strongly convex constants of the objective function affect the convergence of the algorithm (see [6]), but are highly dependent on the number of missing values in the data (see Lemmas 1 and 2 and (6)). The above algorithm includes a backtracking procedure with respect to the smoothness constant, which allows for robust solutions even when there are many missing values in the design matrix.

Remark 3. We demonstrate how strongly convex FISTA can be applied to (3). Set

$$f(\Sigma) := \frac{1}{2} \|W \odot (\Sigma - S^{pair})\|_F^2 \quad \text{and} \quad h(\Sigma) := i_{\mathbb{S}_+^p}(\Sigma).$$

In this case, problem (3) is the special instance of problem (12). The gradient ∇f can be computed by (12). Moreover, the proximal mapping prox_h is $P_{\mathbb{S}_+^p}$ (see, for example, [3] (Example 12.25)) and hence the computation in the algorithm is simple.

4.2. Residential Building Dataset

We compared the performance of HMLasso (3) with the strongly convex FISTA (SC-FISTA) and the alternating direction multiplier method (ADMM) used in [6]. All experiments are conducted on a PC with an Apple M1 Max CPU and 32 GB RAM. Methods are implemented with MATLAB (R2025a).

The numerical experiment uses an actual residential building dataset from the UCI Data Repository <https://archive.ics.uci.edu/ml/datasets/Residential+Building+Data+Set> (accessed on 20 May 2024). The data consisted of $n = 300$ samples and $p = 27$ variables. We set the average missing rates to 20%, 40%, 60%, and 80%. As choices for L_0 , α , and μ occurring in an algorithm, we let $L_0 := 1$, $\alpha := 1.1$, and $\mu := \min_{l,m} (W \odot W)_{l,m}$. We chose 10 random initial points $\Sigma_0^{(i)} \in \mathbb{R}^{p \times p}$ ($i = 1, 2, \dots, 10$), and all entries at the initial point are defined from a standard normal distribution. Figures demonstrate the following function:

$$D_k^{(i)} := \|\Sigma_k^{(i)} - \Sigma_{cvx}\|_F \quad \text{and} \quad D_k := (1/10) \sum_{i=1}^{10} D_k^{(i)}, \quad (14)$$

where Σ_{cvx} denotes the solution obtained by cvx and $\{\Sigma_k^{(i)}\}$ is the sequence generated by $\Sigma_0^{(i)}$ and each of SCFISTA and ADMM.

Figures 1 and 2 show the computation results of the relation between the distance to a solution and iteration number. As we can see from these figures, SCFISTA and ADMM converge to the solution as the number of iterations increases. Furthermore, we found it more effective to solve the HMLasso problem using the acceleration algorithm for strongly convex functions.

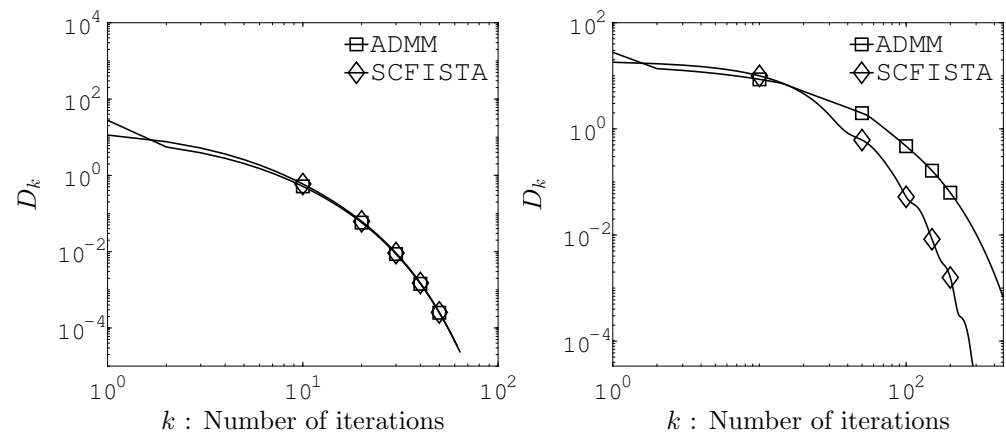


Figure 1. The numerical comparisons of SCFISTA and ADMM at average missing rates of 20% (left) and 40% (right) for the design matrix in (3).

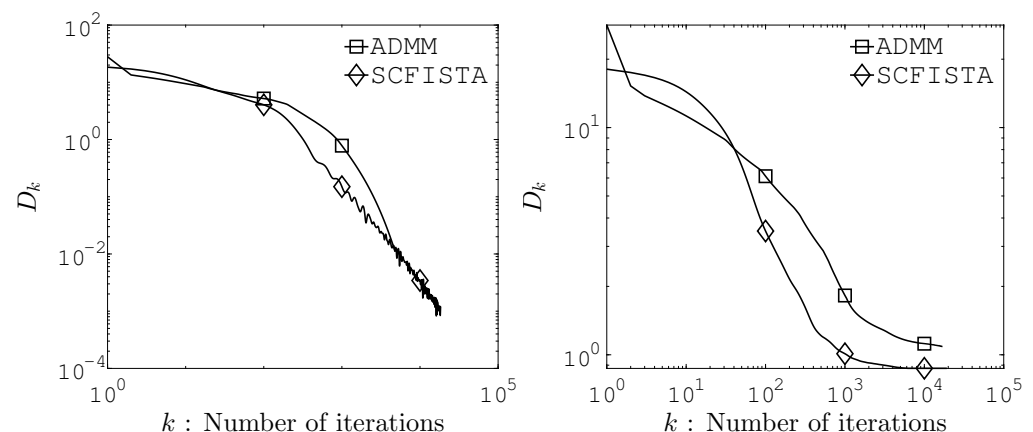


Figure 2. The numerical comparisons of SCFISTA and ADMM at average missing rates of 60% (left) and 80% (right) for the design matrix in (3).

5. Conclusions

In this paper, we present two structural results on the HMLasso problem, covering Lipschitz continuity of the gradient of the objective function and strong convexity of the objective function. These results allow existing acceleration algorithms for strongly convex functions to be applied to solve the HMLasso problem. Our numerical experiments suggest that accelerated algorithms for strongly convex functions are computationally attractive.

Author Contributions: Methodology, Writing—Original Draft Preparation, S.-y.M.; Software, H.S. All authors have read and agreed to the published version of the manuscript.

Funding: This work was partially supported by JSPS KAKENHI, Grant Number 23K03235.

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

Acknowledgments: We would like to express our sincere gratitude to Takayasu Yamaguchi of Akita Prefectural University for providing the initial inspiration for this research. This work was supported by the Research Institute for Mathematical Sciences, an International Joint Usage/Research Center located in Kyoto University.

Conflicts of Interest: Hiromu Sasaki was employed by JTEKT. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

1. Tibshirani, R. Regression shrinkage and selection via the Lasso. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **1996**, *58*, 267–288. [[CrossRef](#)]
2. Takada, M.; Fujisawa, H.; Nishikawa, T. HMLasso: Lasso with High Missing Rate. In Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence (IJCAI), Macao, 10–16 August 2019; pp. 3541–3547.
3. Bauschke, H.H.; Combettes, P.L. *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*, 2nd ed.; CMS Books in Mathematics; Springer: Berlin/Heidelberg, Germany, 2017.
4. Escalante, R.; Raydan, M. *Alternating Projection Methods*; SIAM: Philadelphia, PA, USA, 2011.
5. Beck, A. *First-Order Methods in Optimization*; MOS-SIAM Series on Optimization; SIAM: Philadelphia, PA, USA, 2017.
6. D’Aspremont, A.; Scieur, D.; Taylor, A. Acceleration Methods. *Fund. Trends Optim.* **2021**, *5*, 1–245. [[CrossRef](#)]
7. Beck, A.; Teboulle, M. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM J. Imag. Sci.* **2009**, *2*, 183–202. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.