Decoupling Modal-mixed functional and structural connectome for brain disorder diagnosis with incompleteness

Author name(s) withheld EMAIL(S) WITHHELD
Address withheld

Editors: Under Review for MIDL 2023

Abstract

The brain age has been proven a phenotype with relevance to cognitive performance and brain disease. With the development of deep learning, brain age estimation accuracy has been greatly improved. However, such methods may incur over-fitting and suffer from poor generalizations, especially for insufficient brain imaging data. This paper presents a novel regularization method that penalizes the predictive distribution using knowledge distillation and introduces additional knowledge to reinforce the learning process. During knowledge distillation, we propose a gated distillation mechanism to enable the student model to attentively learn key knowledge from the teacher model, given the assumption that the teacher may not always be correct. Moreover, to enhance the capability of knowledge transfer, the hint representation similarity is also adopted to regularize the model training. We evaluate the model by a cohort of 3655 subjects from 4 public datasets, demonstrating that the proposed method improves the prediction performance over several well-established models, where the mean absolute error of the estimated ages is 2.129 years.
Keywords: List of keywords, comma separated.

1. Introduction

Neuroimging developments have been widely adopted in medical scenarios such as disease diagnosis (Zhu et al., 2021; Liu et al., 2022; Yang et al., 2022) and lesion segmentation (Falk et al., 2019; Menze et al., 2014; Zhou et al., 2018). Advanced neuroimaging candidates, i.e. functonal Magnetic Resonance Imaging (fMRI) and Diffusion Tensor Imaging (DTI), are powerful tools for brain disorder diagnosis by characterizing neural connections and information flow between brain regions (Yin et al., 2022; Li et al., 2021; Kawahara et al., 2017). Derived functional and structural connectomes are modeled as graphs by representing the activity of neurons as nodes interconnected by a set of edges, providing a more holistic view for relating abnormal discharge of neurons and brain dysfunction (Dadi et al., 2019). Analysis of brain connectome can contribute to the scientific understanding of cognitive processes and potentially aid in the diagnosis and treatment of neurological disorders (Gabrieli et al., 2015; Pu et al., 2015).

The brain networks can be catorgoized into two classes: functional connectome derived from fMRI or EEG, and structural connectome obtained from DTI or DSI. Multi-modal brain network studies offers a more constructive scene with distinctive biomarkers, and provides insights into investigating neuron activation and connection in vivo by leveraging complementary information between functional and structural networks. Despite of the promising performance achieved by multi-modal technologies, however there still remains a great challenge to collect large, diverse images with both functional and structural scans.

© 2023 Anonymous.

Accordingly, learning with insufficient samples may lead to model over-fitting and poor generalization.

One way to address the abovementioned issue is by introducing incomplete learning. Apart from multi-modal brain neuroimages, mono-modal neuroimages such as fMRI and DTI are easier to collect. In this regard, more samples, i.e. multi-modal samples as well as mono-modal samples, could be gathered for training. Nevertheless, learning with missing modalities might potentially ignore the complementary information between modalities, and bring in noise for training.

In this study, we propose to tackle the incomplete learning with missing modalities by introducing two strategies. On one hand, we propose a modality-mix data augumentation approach for synthezing samples with incompleteness into complete data for training. The idea behind this is by mix-up data augumentation approach, where samples of different classes are mixed with updated label. The modality-mix method randomly samples data with missing modalities to constitute data with complete modalities. The constructive new samples are leveraged for auxiliary training. On the other hand, to imitate the noise within the synthezed data, where the unpaired complementary information between functional and structural networks might decrease the performance, we investigate the multi-modal learning with deep supervision for decoulping inter-modal associations. A bilateral learning framework is introduced for decoulping multi-modal dependencies, where representations of each branch is fine-tuned by a deep-supervision module. The deep supervision layers are implemented to refinforce the representation learning with decoupled multi-modal associations, where the mono-modal features and multi-modal features are re-balanced for training in order to reduce the importance of unpaired complementary information for learning. Experiments on ADNI data shows the superority of our proposed approaches in improving learning performance and generalizability.

2. Related works

Multi-modal Brain Connectome study. A simple and straightforward way of multi-modal brain connectome learning is to concatenate features and feed them into a classifier such as SVM. Compared with these machine learning approaches, deep learning methods are feasible to embed high-order representations and achieve better performances. For example, (Wang et al., 2018) performed a multi-layer convolution on fMRI and DTI data simultaneously. (Dsouza et al., 2021) regularized convolution on functional connectivity with structural graph Laplacian. A triplet network with a self-attention mechanism was introduced to map high-order multi-modal representations. proposed to perform hyperedge to perform heteogenous graph convolution on multi-modal data (Zhu et al., 2022). In this study, these approaches are implemented as backbones for multi-modal brain connectome study. These approaches are proposed for multi-modal learning, however, fail to tackle incomplete learning. Learning modality interactions and complementary information from incomplete multimodal data was less unexplored by previvious multimodal machine learning research (Chen and Zhang, 2020).

Incomplete Learning. Recently, the exploration of incomplete learning with missing modalities has attracted much attention. Missing modality is a common issue in real-world multimodal scenarios (Ding et al., 2018), and the missingness can be caused by various reasons such as sensor damage, data corruption, and human mistakes in recording (Chen and Zhang, 2020). In most cases, the incomplete learning can be divided into two conditions: the test set is complete and incomplete, whereas the training samples are incomplete with missing modalities. In this study, we focus on learning with complete test set, where mono-modal data samples are implemented to improve the model performance and generalizability. Data imputation methods such like KNN and filling with zero are commonly used in most conditions. Advanced imputation methods such as adversial training with similar structure as GAN have also been proposed to deal with imputing the missing modalities (Cai et al., 2018). (Wang et al., 2020) proposed a knowledge distillation based approach to integrate the supplementary information of multiple modalities.

3. Method

3.1. Problem formulation

In this section, we introduce the problem, task and definations. In general, the data with incompleteness in our study can be divided into training sets and test sets. The training sets included complete data and incomplete data, while the test sets are complete. For a multi-modal dataset with M modalities, there are $2^M - 1$ different combinations of missing modalities. In this study, we focus on studying with functional and structural brain networks, where $M = 2$. Specially, for the training sets, the samples with missing modalities are denoted as $X^{1u} \in R^{n_{1u} \times d_{1u}}, X^{2u} \in R^{n_{2u} \times d_{2u}}$. And the complete data are represented as $X^c \in R^{n_c \times d_c}$ with $N_{train} = n_c + n_{1u} + n_{2u}$ training samples. While for the test set, the samples are complete with both two modalities $X^c \in R^{n_c \times d_c}$ with $N_{test} = n_c$.

In this study, given a collection of incomplete multi-modal data samples $\{X_i\}_{i=1}^N$ as input, where each sample consists of a set of available modalities $X_i = \{x_{i,m}\}$, our goal is to design a model to capture dependencies between modalities and fuse multi-modal data with different patterns in a architecture.

3.2. Modality-Mix

(Zhang et al., 2017) first proposed the Mixup method for image classification, where synthetic samples are generated by linearly interpolating a pair of training samples as well as their targets. Consider a pair of samples $(x^i; y^i)$ and $(x^j; y^j)$, a synthetic sample is geneated as $\hat{x}^{ij} = \lambda x^i + (1-\lambda)x^j$, $y^{ij} = \lambda y^i + (1-\lambda)y^j$, where $\lambda \in (0, 1)$ is the mixing ratio for the pair. In this study, we follow the mixup approach and propose a modality-mixup method by replacing the combination operation with concatenatation layers. In detail, the synthetic data samples are obtained by:

$$\hat{x}^{ij} = ||\{\lambda x^i, (1-\lambda)x^j\}, x^i \in X^{1u}, x^j \in X^{2u} \tag{1}$$

$$y^{ij} = \lambda y^i + (1-\lambda)y^j \tag{2}$$

where $||$ denotes the concatenatation operation. In this regard, the samples with modality 1 missing are mixed with those with modality 2 missing, and then constitute into complete data.

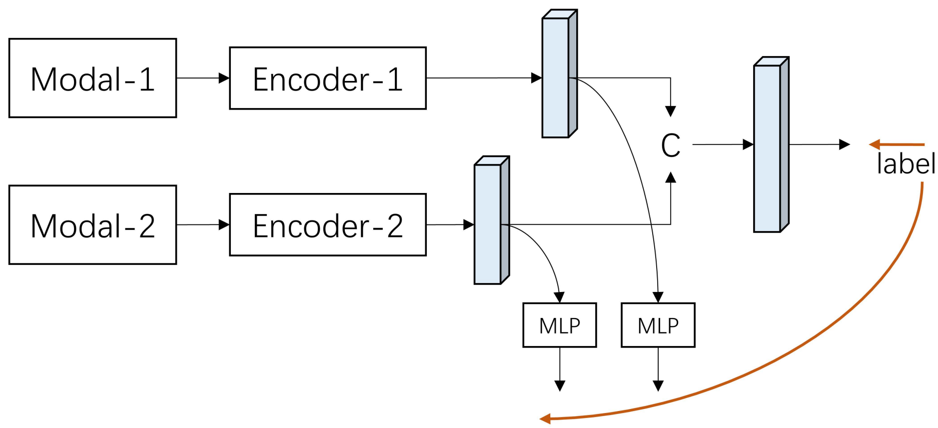

Figure 1: Illustration of the proposed Cross-GNN method including functional, structural, and cross-modal pathways. The multimodal brain networks are firstly parsed and then formulated into a correspondence matrix Φ for reasoning. The multi-modal representations are cross-embedded and cross-distilled for cross-modal representation learning.

3.3. Deep supervision

One core challenge of data imputation by modality-mixup is the unpaired inter-modal complementary information that would bring in noise and decrease the performance. In order to reduce the effect of interactions of inter-modal representations, in this study, we propose to decouple the multi-modal features. In detail, deep supervision is introduced to improve the importance of mono-modal representations in classification. Figure 1 demonstrates the scratch of the proposed framework, where representations of each modality are encoded and then fused by concatenatation to classify. Notably, a multi-layer perception is implemented to obtain the prediction output of each modal.

Since the multi-modal representations are heteogenous, the way of embedding brain networks plays a key role in classification. In this study, the encoders are developed by well-estimated backbone of brain network study, including multi-layer perception (MLP), graph embedding, and sequential models such like LSTM and GRU. In detail, the MLP layers take the vectorized brain connectome features into 32 features followed with ReLU activation and dropout. Graph embedding layers leveraged the normalized brain connectome matrix as the adjancency matrix and perform graph convolution by symmetric normalized Laplacian. The sequential models take the brain network input as a sequence with M nodes with M features.

3.4. Optimization

In the training process, the synthetic data and complete samples are fed into the framework. The objective function is constructed by a combination of cross-entropy loss from the target output and the deep supervision, as:

$$L = L_{ce}(y_m, y) + \sum_{s=1}^{m} L_{ce}(y_s, y) \tag{3}$$

where $L_{ce}(y_m, y)$ denotes the loss of prediction output of the concatenated multi-modal representations, $L_{ce}(y_s, y)$ represents the s-th deep supervision output.

4. Experiments

4.1. Datasets

In this study, two datasets are implemented for evaluation, where functional MRI and DTI images are collected.

ADNI Dataset[1]: 185 subjects were enrolled including 61 healthy controls (HC) and 63 with mild cognitive impairment (MCI), and 61 patients with Alzheimer's disease (AD). MCI is considered to be a significant stage for the preclinical diagnosis of AD. The patients were diagnosed at baseline and the HCs were healthy at their first examination. These subjects are divided into three groups: AD, MCI, and HC.

Xuanwu dataset: 53 HCs, 50 subjects with iRBD, and 85 subjects with Parkinson's Disease (PD) were recruited from the Movement Disorders Clinic of the Xuanwu Hospital of Capital Medical University. Idiopathic rapid eye movement sleep behaviour disorder (iRBD) has been increasingly recognized as the heralding features of PD and is characterized by a long incubation period (Iranzo et al., 2006; Boeve, 2010).In the dataset, the HCs were all older than 40 years, with no family history of movement disorders and no obvious cerebral lesions found in MR images. The iRBD patients were screened by the International Classification of Sleep Disorders-Third Edition (ICSD 3) diagnostic criteria and confirmed by polysomnography (Sateia, 2014). The PDs were diagnosed according to the MDS Clinical Diagnostic Criteria for Parkinson's disease.

4.2. Preprocessing

All the fMRI images were pre-processed by reference to the Configurable Pipeline for the Analysis of Connectomes (CPAC) pipeline (Craddock et al., 2013), including skull striping, slice timing correction, motion correction, global mean intensity normalization, nuisance signal regression with 24 motion parameters, and band-pass filtering (0.01-0.08Hz). The functional images were finally registered into standard anatomical space (MNI152). The mean time series for a set of regions were computed and normalized into zero mean and unit variance. The Pearson Coefficient Correlation was applied to measure functional connectivity. The DTI images were pre-processed by image denoising, head motion, eddy-current, susceptibility distortion, and field inhomogeneity correction by MRtrix 3 (Tournier et al., 2012). The streamline count was reconstructed to 5 million. The number of streamlines connecting each pair of brain regions was used to construct the structural network.

The pre-processed fMRI and DTI images were mapped by the brain template for parcellations. In this study, the images in ADNI and Xuanwu datasets were segmented by the Schaefer atlas (Schaefer et al., 2018) that identified 100 cortical parcels.

1. http://www.adni-info.org/

4.3. Implementation details

In our implementation, the number of layers of bilateral graph convolution is decided in a grid search from 1 to 4. The outputs of bilateral graph convolution layers are further fed into a 3-layer multi-layer perception classifier followed by a leaky ReLU activation function and a dropout layer. The learning rate is set as 3e-4, and the weight decay is 5e-5. All the models in this study are trained for 600 epochs and would be early stopped when the loss has not been decreased for 100 epochs. We trained the models with PyTorch on one NVIDIA 2080-Ti GPU. 10-fold cross-validation was applied for evaluation, where 10% samples are randomly selected for testing for each fold. For all experiments, we evaluated the performance in terms of the diagnosis accuracy (Acc), and sensitivity (Sen), and specificity (Spe).

4.4. Competitive baseline

In this study, we compare our proposed method with data imputation methods including training with only complete data (C), missing modality imputation by K-nearest neighbors (KNN) (Campos et al., 2015), adversial-based imputation (ADV) (Cai et al., 2018), and knowledge distillation-based imputation (KD) (Wang et al., 2020). We also compare our proposed decoulping framework with multi-modal learning frameworks including BrainNetCNN (Kawahara et al., 2017), Triplet Attention Network (Zhu et al., 2022), M-GCN (Dsouza et al., 2021), HGCN (Feng et al., 2019).

In this study, we compare our proposed Cross-GNN with baseline machine learning approaches and well-estimated graph methods. These methods include:

This is where the content of your paper goes. Some random notes[2]:

- You should use LaTeX(?).
- JMLR/PMLR uses natbib for references. For simplicity, here, `\cite` defaults to parenthetical citations, i.e. `\citep`. You can of course also use `\citet` for textual citations.
- Eprints such as arXiv papers can of course be cited (?). We recomend using a `@misc` bibtex entry for these as shown in the sample bibliography.
- You should follow the guidelines provided by the conference.
- Read through the JMLR template documentation for specific LaTeXusage questions.
- Note that the JMLR template provides many handy functionalities such as `\figureref` to refer to a figure, e.g. Figure 2, `\tableref` to refer to a table, e.g. Table 1 and `\equationref` to refer to an equation, e.g. Equation (4).

Acknowledgments

Acknowledgments withheld.

2. Random footnote are discouraged

Table 1: An Example Table

Dataset	**Result**
Data1	0.12345
Data2	0.67890
Data3	0.54321
Data4	0.09876

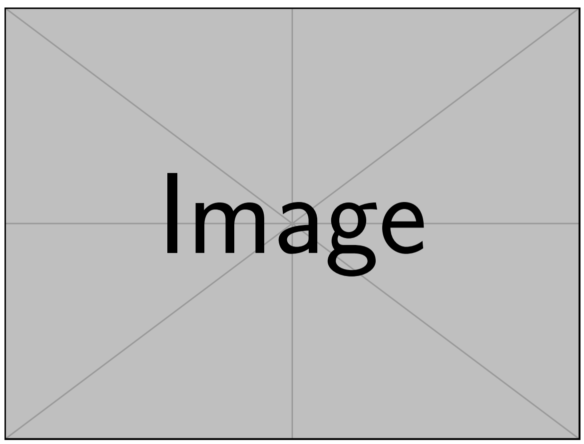

Figure 2: Example Image

Algorithm 1: Computing Net Activation

Input: $x_1, \ldots, x_n, w_1, \ldots, w_n$
Output: y, the net activation
$y \leftarrow 0$;
for $i \leftarrow 1$ **to** n **do**
    $y \leftarrow y + w_i * x_i$;
end

References

Bradley F Boeve. Rem sleep behavior disorder: updated review of the core features, the rem sleep behavior disorder-neurodegenerative disease association, evolving concepts, controversies, and future directions. *Annals of the New York Academy of Sciences*, 1184(1): 15–54, 2010.

Lei Cai, Zhengyang Wang, Hongyang Gao, Dinggang Shen, and Shuiwang Ji. Deep adversarial learning for multi-modality missing data completion. In *Proceedings of the 24th ACM SIGKDD international conference on knowledge discovery & data mining*, pages 1158–1166, 2018.

Sergio Campos, Luis Pizarro, Carlos Valle, Katherine R Gray, Daniel Rueckert, and Héctor Allende. Evaluating imputation techniques for missing data in adni: a patient classification study. In *Iberoamerican Congress on Pattern Recognition*, pages 3–10. Springer, 2015.

Jiayi Chen and Aidong Zhang. Hgmf: heterogeneous graph-based fusion for multimodal data with incompleteness. In *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 1295–1305, 2020.

Cameron Craddock, Sharad Sikka, Brian Cheung, Ranjeet Khanuja, Satrajit S Ghosh, Chaogan Yan, Qingyang Li, Daniel Lurie, Joshua Vogelstein, Randal Burns, et al. Towards automated analysis of connectomes: The configurable pipeline for the analysis of connectomes (c-pac). *Front Neuroinform*, 42:10–3389, 2013.

Kamalaker Dadi, Mehdi Rahim, Alexandre Abraham, Darya Chyzhyk, Michael Milham, Bertrand Thirion, Gaël Varoquaux, Alzheimer's Disease Neuroimaging Initiative, et al. Benchmarking functional connectome-based predictive models for resting-state fmri. *NeuroImage*, 192:115–134, 2019.

Zhengming Ding, Handong Zhao, and Yun Fu. Learning representation for multi-view data analysis: models and applications. 2018.

Niharika Shimona Dsouza, Mary Beth Nebel, Deana Crocetti, Joshua Robinson, Stewart Mostofsky, and Archana Venkataraman. M-gcn: A multimodal graph convolutional network to integrate functional and structural connectomics data to predict multidimensional phenotypic characterizations. In *Medical Imaging with Deep Learning*, pages 119–130. PMLR, 2021.

Thorsten Falk, Dominic Mai, Robert Bensch, Özgün Çiçek, Ahmed Abdulkadir, Yassine Marrakchi, Anton Böhm, Jan Deubner, Zoe Jäckel, Katharina Seiwald, et al. U-net: deep learning for cell counting, detection, and morphometry. *Nature methods*, 16(1): 67–70, 2019.

Yifan Feng, Haoxuan You, Zizhao Zhang, Rongrong Ji, and Yue Gao. Hypergraph neural networks. In *Proceedings of the AAAI conference on artificial intelligence*, volume 33, pages 3558–3565, 2019.

John DE Gabrieli, Satrajit S Ghosh, and Susan Whitfield-Gabrieli. Prediction as a humanitarian and pragmatic contribution from human cognitive neuroscience. *Neuron*, 85(1): 11–26, 2015.

Alex Iranzo, José Luis Molinuevo, Joan Santamaría, Mónica Serradell, María José Martí, Francesc Valldeoriola, and Eduard Tolosa. Rapid-eye-movement sleep behaviour disorder as an early marker for a neurodegenerative disorder: a descriptive study. *The Lancet Neurology*, 5(7):572–577, 2006.

Jeremy Kawahara, Colin J Brown, Steven P Miller, Brian G Booth, Vann Chau, Ruth E Grunau, Jill G Zwicker, and Ghassan Hamarneh. Brainnetcnn: Convolutional neural networks for brain networks; towards predicting neurodevelopment. *NeuroImage*, 146: 1038–1049, 2017.

Xiaoxiao Li, Yuan Zhou, Nicha Dvornek, Muhan Zhang, Siyuan Gao, Juntang Zhuang, Dustin Scheinost, Lawrence H Staib, Pamela Ventola, and James S Duncan. Braingnn: Interpretable brain graph neural network for fmri analysis. *Medical Image Analysis*, 74: 102233, 2021.

Yanbei Liu, Henan Li, Tao Luo, Changqing Zhang, Zhitao Xiao, Ying Wei, Yaozong Gao, Feng Shi, Fei Shan, and Dinggang Shen. Structural attention graph neural network for diagnosis and prediction of covid-19 severity. *IEEE Transactions on Medical Imaging*, 2022.

Bjoern H Menze, Andras Jakab, Stefan Bauer, Jayashree Kalpathy-Cramer, Keyvan Farahani, Justin Kirby, Yuliya Burren, Nicole Porz, Johannes Slotboom, Roland Wiest, et al. The multimodal brain tumor image segmentation benchmark (brats). *IEEE transactions on medical imaging*, 34(10):1993–2024, 2014.

Jian Pu, Jun Wang, Wenwen Yu, Zhuangming Shen, Qian Lv, Kristina Zeljic, Chencheng Zhang, Bomin Sun, Guoxiang Liu, and Zheng Wang. Discriminative structured feature engineering for macroscale brain connectomes. *IEEE Transactions on Medical Imaging*, 34(11):2333–2342, 2015.

Michael J Sateia. International classification of sleep disorders. *Chest*, 146(5):1387–1394, 2014.

Alexander Schaefer, Ru Kong, Evan M Gordon, Timothy O Laumann, Xi-Nian Zuo, Avram J Holmes, Simon B Eickhoff, and BT Thomas Yeo. Local-global parcellation of the human cerebral cortex from intrinsic functional connectivity mri. *Cerebral cortex*, 28(9):3095–3114, 2018.

J-Donald Tournier, Fernando Calamante, and Alan Connelly. Mrtrix: diffusion tractography in crossing fiber regions. *International journal of imaging systems and technology*, 22(1): 53–66, 2012.

Qi Wang, Liang Zhan, Paul Thompson, and Jiayu Zhou. Multimodal learning with incomplete modalities by knowledge distillation. In *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 1828–1838, 2020.

Yan Wang, Yanwu Yang, Xin Guo, Chenfei Ye, Na Gao, Yuan Fang, and Heather T Ma. A novel multimodal mri analysis for alzheimer's disease based on convolutional neural network. In *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 754–757. IEEE, 2018.

Yanwu Yang, Xutao Guo, Zhikai Chang, Chenfei Ye, Yang Xiang, and Ting Ma. Multimodal dynamic graph network: Coupling structural and functional connectome for disease diagnosis and classification. *arXiv preprint arXiv:2210.13721*, 2022.

Wutao Yin, Longhai Li, and Fang-Xiang Wu. Deep learning for brain disorder diagnosis based on fmri images. *Neurocomputing*, 469:332–345, 2022.

Hongyi Zhang, Moustapha Cisse, Yann N Dauphin, and David Lopez-Paz. mixup: Beyond empirical risk minimization. *arXiv preprint arXiv:1710.09412*, 2017.

Zongwei Zhou, Md Mahfuzur Rahman Siddiquee, Nima Tajbakhsh, and Jianming Liang. Unet++: A nested u-net architecture for medical image segmentation. In *Deep learning in medical image analysis and multimodal learning for clinical decision support*, pages 3–11. Springer, 2018.

Qi Zhu, Heyang Wang, Bingliang Xu, Zhiqiang Zhang, Wei Shao, and Daoqiang Zhang. Multi-modal triplet attention network for brain disease diagnosis. *IEEE Transactions on Medical Imaging*, 2022.

Wenyong Zhu, Liang Sun, Jiashuang Huang, Liangxiu Han, and Daoqiang Zhang. Dual attention multi-instance deep learning for alzheimer's disease diagnosis with structural mri. *IEEE Transactions on Medical Imaging*, 40(9):2354–2366, 2021.

Appendix A. Proof of Theorem 1

This is a boring technical proof of

$$\cos^2\theta + \sin^2\theta \equiv 1. \tag{4}$$

Appendix B. Proof of Theorem 2

This is a complete version of a proof sketched in the main text.