

Departamento de Economía
Tipo de documento: Artículo



A note on efficient minimum cost adjustment sets in causal graphical models

Autoría: Smucler, Ezequiel; Rotnitzky, Andrea

Fecha de publicación: 14/07/2022

Publicado originalmente en: Journal of Causal Inference (e-ISSN: 2193-3685)

¿Cómo citar este trabajo?

Smucler, E. & Rotnitzky, A. (2022). A note on efficient minimum cost adjustment sets in causal graphical models. Journal of Causal Inference, 10(1), 174-189. <https://doi.org/10.1515/jci-2022-0015>

El presente documento se encuentra alojado en el Repositorio Digital de la Universidad Torcuato Di Tella, bajo una [licencia Creative Commons Atribución 4.0 Internacional](#), de acuerdo a lo indicado en su fuente original.

Dirección:

<https://repositorio.utdt.edu/handle/20.500.13098/14377>

Research Article

Ezequiel Smucler* and Andrea Rotnitzky

A note on efficient minimum cost adjustment sets in causal graphical models

<https://doi.org/10.1515/jci-2022-0015>

received February 25, 2022; accepted June 23, 2022

Abstract: We study the selection of adjustment sets for estimating the interventional mean under an individualized treatment rule. We assume a non-parametric causal graphical model with, possibly, hidden variables and at least one adjustment set composed of observable variables. Moreover, we assume that observable variables have positive costs associated with them. We define the cost of an observable adjustment set as the sum of the costs of the variables that comprise it. We show that in this setting there exist adjustment sets that are minimum cost optimal, in the sense that they yield non-parametric estimators of the interventional mean with the smallest asymptotic variance among those that control for observable adjustment sets that have minimum cost. Our results are based on the construction of a special flow network associated with the original causal graph. We show that a minimum cost optimal adjustment set can be found by computing a maximum flow on the network, and then finding the set of vertices that are reachable from the source by augmenting paths. The `optimaladj` Python package implements the algorithms introduced in this article.

Keywords: adjustment set, back-door criterion, causal inference, semiparametric efficiency

MSC 2020: 62G05, 62D20

1 Introduction

This article contributes to the growing literature on graphical criteria for selecting adjustment sets that suffice to control for confounding. In a causal graphical model, a set of covariates Z is an adjustment set for the effect of a, possibly individualized, point exposure treatment rule on an outcome, if the interventional mean, i.e., the mean of the outcome in the hypothetical world in which all units in the population receive a given treatment rule, is identified by the g-formula [1] that adjusts for Z . An individualized point exposure treatment rule is a rule for assigning subjects to treatment at a single time point that depends on the subjects' covariates. This article contributes to the existing literature by incorporating the possibility that variables in the graph are associated with different costs and exploring the existence of graphical criteria for determining efficient adjustment sets under different scenarios on the budget constraint.

The literature on graphical selection of adjustment sets is particularly relevant for the task of deciding, at the stage of the design of an observational study, which variables to measure in order to control for confounding. It is well known that, under a causal graphical model, there may exist many adjustment sets, and a series of recent articles derived complete and sound graphical criteria for determining them [2–5]. A subsequent thread of articles provided graphical rules for comparing adjustment sets in causal graphical models based on efficiency criteria. Specifically, assuming a linear causal graphical model with no hidden variables and treatment effects estimated by ordinary least squares, previous studies [6] and [3] provided

* Corresponding author: Ezequiel Smucler, Glovo, Barcelona, Spain, e-mail: ezequiels.90@gmail.com

Andrea Rotnitzky: Department of Economics, Universidad Torcuato Di Tella, Buenos Aires, Argentina; Department of Biostatistics, Harvard T.H. Chan School of Public Health, Cambridge, United States, e-mail: arotnitzky@utdt.edu

criteria for comparing certain pairs of adjustment sets. Ref. [7] derived a graphical characterization of the globally most efficient adjustment set and extended the criteria of Kuroki, Cai, and Miyakawa. See also ref. [8]. Ref. [9] extended these results to non-parametric graphical models and non-parametrically adjusted estimators. All of the aforementioned articles considered only static treatment rules, i.e., “one size fits all regimes,” which assign the same treatment to all population units regardless of their covariates. Ref. [10] extended the results of ref. [9] by allowing the possibility of both individualized treatment rules and graphical models that have hidden variables. The authors also provided graphical criteria for determining optimal adjustment sets both among minimal adjustment sets and among minimum cardinality adjustment sets. Minimal adjustment sets are valid adjustment sets such that the removal of any variable from them destroys their validity. Moreover, ref. [10] provided a sufficient criterion for the existence of a globally optimal adjustment set. Ref. [11] provided a necessary and sufficient criterion for the existence of a globally optimal adjustment set in linear causal graphical models. Assuming a non-graphical online setting in which the investigator can alter the data collection mechanism adaptively, ref. [12] proposed an estimator of an optimal identifying functional among those in a predefined set.

In most realistic settings the costs associated with measuring different covariates can vary considerably. For instance, variables requiring the laboratory assessment of blood samples are usually much more expensive to measure than variables that can be obtained by clinical examination. The latter, in turn, are more expensive than those obtained from surveys. Cost considerations then give rise to the important practical problem of determining the adjustment set that yields estimators of treatment effects with minimum variance among the adjustment sets whose overall cost meets a given budget constraint. To the best of our knowledge, the literature on causal graphical models has not addressed this problem yet. The present article fills this gap. In fact, we show by means of an example that such an ideal adjustment set does not always exist, except when the variance minimization problem is restricted to observable adjustment sets with lowest overall cost. We derive a graphical criterion and a polynomial time algorithm for computing a solution to the latter optimization problem. We refer to adjustment sets that solve this problem as optimal minimum cost adjustment sets. Our results are based on building a special flow network associated with the original causal graph. We show that optimal minimum cost adjustment sets can be found by computing a maximum flow on the network, and then finding the set of vertices that are reachable from the source by augmenting paths. Flow networks had already been proposed in refs [13] and [14] as a tool to compute minimal and minimum cost adjustment sets, but with no consideration for statistical efficiency. The python package `optimaladj` available at <https://pypi.org/project/optimaladj> implements the algorithms introduced in this article.

The rest of the article is organized as follows. In Section 2, we review some necessary background on graph theory and semiparametric estimation. In Section 3, we present the non-parametric causal graphical model we will assume throughout the article. In Section 4, we provide a graphical characterization of minimum cost adjustment sets as a class of vertex separators in the undirected graph introduced in ref. [10]. In Section 5, we provide a graphical criterion for efficiency comparisons of minimum cost adjustment sets, based on the aforementioned undirected graph. Section 6 contains the main results of this article. In it, we construct a special flow network and show how min-cuts in it are related to minimum cost adjustment sets. We also show that an optimal minimum cost adjustment set always exists and provide a polynomial time algorithm to compute it. Finally, we illustrate our results in a few examples and show that the optimal budget constrained adjustment set discussed earlier does not exist in general.

2 Background

2.1 Undirected graphs

An undirected graph $\mathcal{H} = (V, E)$ is formed by a finite vertex set V and a set of undirected edges E . A weighted undirected graph is an undirected graph together with a cost function $c : V \rightarrow (0, +\infty)$.

The cost of a set of vertices Z is defined as the sum of the costs of the vertices that comprise Z . In a slight abuse of notation we write $c(Z)$ for the cost of Z .

If $U - W$ is an edge in $\mathcal{H}$, then we say that U and W are adjacent. A path between U and W is a sequence of adjacent vertices $(V_1, \dots, V_j)$ such that $V_1 = U$ and $V_j = W$. Two vertices U and W are connected in $\mathcal{H}$ if there exists a path from U to W in $\mathcal{H}$.

For Z_1, Z_2 , and Z_3 disjoint sets of vertices in $\mathcal{H}$, we write $Z_1 \perp_{\mathcal{H}} Z_2 | Z_3$ if every path in $\mathcal{H}$ between Z_1 and Z_2 intersects Z_3 . Given two vertices A and Y in $\mathcal{H}$, a set of vertices Z disjoint with A and Y is an $A - Y$ separator if $A \perp_{\mathcal{H}} Y | Z$. The set Z is a minimal $A - Y$ separator if it is an $A - Y$ separator and no proper subset of Z is an $A - Y$ separator. If $\mathcal{H}$ is a weighted undirected graph with weight function c , the set Z is a minimum cost $A - Y$ separator if it is an $A - Y$ separator that satisfies $c(Z) \leq c(Z')$ for any other $A - Y$ separator Z' .

2.2 Directed graphs

A directed graph $\mathcal{G} = (V, E)$ is formed by a finite vertex set V and a set of directed edges $E \subset V \times V$. Given $Z \subset V$ the induced subgraph $\mathcal{G}_Z = (Z, E_Z)$ is defined as the graph obtained by considering only vertices in Z and edges between vertices in Z .

Two vertices U and W are adjacent if there is an edge between them. A path between U and W in $\mathcal{G}$ is a sequence of adjacent vertices $(V_1, \dots, V_j)$ such that $V_1 = U$ and $V_j = W$. The path is directed (or causal) if $V_i \rightarrow V_{i+1}$ for all $i \in \{1, \dots, j-1\}$.

If $U \rightarrow W$, then U is a parent of W . If there is a directed path from U to W , then U is an ancestor of W and W a descendant of U . We follow the convention that every vertex is an ancestor and a descendant of itself. The sets of parents, ancestors, and descendants of W in $\mathcal{G}$ are denoted by $\text{pa}_{\mathcal{G}}(W)$, $\text{an}_{\mathcal{G}}(W)$, and $\text{de}_{\mathcal{G}}(W)$, respectively. The set of non-descendants of W is defined as $\text{nd}_{\mathcal{G}}(W) \equiv V \setminus \text{de}_{\mathcal{G}}(W)$. For a set of vertices Z we define $\text{an}_{\mathcal{G}}(Z) = \cup_{W \in Z} \text{an}_{\mathcal{G}}(W)$ and $\text{de}_{\mathcal{G}}(Z) = \cup_{W \in Z} \text{de}_{\mathcal{G}}(W)$.

A directed cycle is a directed path that begins and ends at the same vertex. A directed acyclic graph is a directed graph that does not have directed cycles.

Let $\mathcal{G}$ be a directed acyclic graph and let A and Y be vertices in $\mathcal{G}$. Let $\text{cn}(A, Y, \mathcal{G})$ be the set of vertices that lie on a directed path between A and Y and are not equal to A . Let $\text{forb}(A, Y, \mathcal{G}) \equiv \text{de}_{\mathcal{G}}(\text{cn}(A, Y, \mathcal{G})) \cup \{A\}$. We call this the set of forbidden vertices with respect to A, Y in $\mathcal{G}$.

The proper back-door graph $\mathcal{G}^{pbd}(A, Y)$ [14] is defined as the graph formed by removing from $\mathcal{G}$ the first edge of every directed path from A to Y .

The moral graph $\mathcal{G}^m$ associated with a directed acyclic graph $\mathcal{G}$ is an undirected graph with the same vertex set as $\mathcal{G}$ and an edge $U - W$ if any of the following hold in $\mathcal{G}$: $U \rightarrow W$, $W \rightarrow U$, or there exists a vertex C such that $U \rightarrow C \leftarrow W$.

2.3 Flow networks

We follow the conventions in ref. [15]. A flow network $\mathcal{D} = (V, E, k, s, t)$ is a directed graph together with a capacity function $k: E \rightarrow [0, +\infty]$ and two distinguished vertices s and t . s is called the source and t the sink of the network. $k(e)$ is called the capacity of edge e . Consider then a flow network $\mathcal{D} = (V, E, k, s, t)$. For a vertex $W \in V$, we let $\alpha(W)$ be the set of edges that point into W and $\beta(W)$ be the set of edges that point out of W . For a set of vertices $S \subset V$ we let $\bar{S} = V \setminus S$.

A flow is a function $f: E \rightarrow \mathbb{R}$ that satisfies $0 \leq f(e) \leq k(e)$ for all $e \in E$ and $\sum_{e \in \alpha(W)} f(e) = \sum_{e \in \beta(W)} f(e)$ for every vertex W not equal to the source or the sink. The total flow of f is defined as $\sum_{e \in \alpha(t)} f(e)$. A flow is called a max-flow if no other flow has a greater total flow.

A cut is a set of vertices that contains the source but not the sink of the network. If S is a set of vertices we define $(S, \bar{S})$ as the set of edges in $\mathcal{D}$ of the form $U \rightarrow W$ for some $U \in S$ and $W \notin S$. If S is a cut we define its capacity as

$$k(S) \equiv \sum_{e \in (S, \bar{S})} k(e).$$

A cut is called a min-cut if no other cut has a smaller capacity.

2.4 Semiparametric estimation

An estimator $\hat{\gamma}$ of a parameter $\gamma(P)$ based on n independent identically distributed random copies $V_1, \dots, V_n$ of V is asymptotically linear at a probability law P if there exists a random variable $\varphi_P(V)$, called the influence function of $\gamma(P)$, such that $E_P\{\varphi_P(V)\} = 0$, $\text{var}_P\{\varphi_P(V)\} < \infty$, and $n^{1/2}\{\hat{\gamma} - \gamma(P)\} = n^{-1/2}\sum_{i=1}^n \varphi_P(V_i) + o_P(1)$ under P . The central limit theorem implies that if $\hat{\gamma}$ is asymptotically linear, $n^{1/2}\{\hat{\gamma} - \gamma(P)\}$ converges in distribution to a zero mean normal distribution with variance $\text{var}_P\{\varphi_P(V_i)\}$. Given a collection of probability laws $\mathcal{P}$ for V , an estimator $\hat{\gamma}$ of $\gamma(P)$ is said to be regular at one P if its convergence to $\gamma(P)$ is locally uniform at P in $\mathcal{P}$ [16].

3 Causal graphical models

Given a directed acyclic graph $\mathcal{G}$ with vertex set V , we identify V with a random vector. The Bayesian network $\mathcal{M}(\mathcal{G})$ is the collection of laws P for V that satisfy the local Markov property:

$$W \perp\!\!\!\perp \text{nd}_{\mathcal{G}}(W) \mid \text{pa}_{\mathcal{G}}(W) \text{ under } P \text{ for all } W \in V.$$

Here $A \perp\!\!\!\perp B \mid C$ stands for conditional independence of A and B given C . Throughout, we will assume that the law P of V admits a density f with respect to some dominating measure. Then the local Markov property implies [2]

$$f(v) = \prod_{V_j \in V} f\{v_j \mid \text{pa}_{\mathcal{G}}(v_j)\}, \quad (1)$$

where $\text{pa}_{\mathcal{G}}(v_j)$ is the value taken by $\text{pa}_{\mathcal{G}}(V_j)$ when V takes the value v .

In this article, we will assume an agnostic causal graphical model [17,18] represented by a directed acyclic graph $\mathcal{G}$. This model identifies the vertex set of $\mathcal{G}$ with a factual random vector V and assumes that: (i) the law P of V satisfies $P \in \mathcal{M}(\mathcal{G})$ and (ii) for any $A \in V$, $L \subset \text{nd}_{\mathcal{G}}(A)$ and $\pi(A \mid L)$ a conditional law for A given L , the intervention density $f_{\pi}(v)$ of the variables in $\mathcal{G}$ when, possibly contrary to fact, the value of A is drawn from the law $\pi(A \mid L)$ is given by

$$f_{\pi}(v) = \pi(a \mid l) \prod_{V_j \in V \setminus \{A\}} f\{v_j \mid \text{pa}_{\mathcal{G}}(v_j)\}, \quad (2)$$

where a and l are the values taken by A and L when V takes the value v , and throughout we assume that A is a finite-valued variable and $f\{A = a \mid \text{pa}_{\mathcal{G}}(A)\} > 0$ with probability 1 under P for all a . Formula (2) is known as the g-formula [1]. The conditional law π designates a, possibly random and individualized, treatment rule. Individualized treatment rules that depend on L are also referred to as point exposure dynamic treatment regimes or L -dependent regimes in the literature. A non-random individualized treatment rule that sets $A = d(L)$ corresponds to the point mass conditional law $\pi(a \mid l) = I_{d(l)}(a)$. In particular, a constant function $d(L) = a$ corresponds to a static intervention that sets $A = a$. Throughout the article, let Y and A be the outcome and treatment of interest, respectively. Let $\chi_{\pi}(P; \mathcal{G})$ be the mean of Y under f_{π} . We refer to $\chi_{\pi}(P; \mathcal{G})$ as the interventional mean under treatment rule π . By factorizations (1) and (2), the Radon–Nikodym theorem gives

$$\chi_{\pi}(P; \mathcal{G}) = E_P \left[\frac{\pi(A \mid L)}{f\{A \mid \text{pa}_{\mathcal{G}}(A)\}} Y \right]. \quad (3)$$

Furthermore, the local Markov property implies that

$$\chi_{\pi}(P; \mathcal{G}) = E_P(E_{\pi^*}[E_P\{Y|A, \text{pa}_{\mathcal{G}}(A), L\}|\text{pa}_{\mathcal{G}}(A), L]), \quad (4)$$

where $E_P(\cdot|\cdot)$ stands for the conditional mean under P and $E_{\pi^*}(\cdot|\cdot)$ stands for the conditional mean under the conditional law of A given L and $\text{pa}_{\mathcal{G}}(A)$ defined as $\pi^*\{A|\text{pa}_{\mathcal{G}}(A), L\} \equiv \pi(A|L)$. In the special case in which the regime is deterministic and static setting A to a , the right hand side of equation (3) is equal to inverse probability weighted expression for the interventional mean $E_P[f^{-1}\{A = a|\text{pa}_{\mathcal{G}}(A)\}I_a(A)Y]$, and the right hand side of equation (4) is equal to the standardized mean expression for the interventional mean $E_P[E_P\{Y|A = a, \text{pa}_{\mathcal{G}}(A)\}]$.

We are interested in conducting inference about $\chi_{\pi}(P; \mathcal{G})$ when only a subset N of V is observable. The inferential problem is thus defined by the following assumptions: (i) $P \in \mathcal{M}(\mathcal{G})$, (ii) the available data consists of a random sample from the marginal law of N under P , (iii) the parameter of interest is $\chi_{\pi}(P; \mathcal{G})$, and (iv) at least one observable adjustment set exists. Adjustment sets will be formally defined in the following section. Throughout we will assume (i) $Y \in \text{de}_{\mathcal{G}}(A)$, (ii) $\{A, Y\} \cup L \subset N$, (iii) $L \subset \text{nd}_{\mathcal{G}}(A)$, (iv) A takes values in a finite set, and (v) $f\{A = a|\text{pa}_{\mathcal{G}}(A)\} > 0$ with probability 1 under P for all a . Moreover, we assume that each observable variable $W \in N$ has an associated positive cost $c(W)$ that could represent, for example, the cost of measuring W .

4 Minimum cost adjustment sets and their graphical characterization

Ref. [10] gave the following definitions of dynamic adjustment sets and minimal dynamic adjustment sets in graphs with hidden variables. The dynamic appellation here refers to the, possibly, dynamic nature of the interventions, meaning that the interventions can depend on the value of the variables in L .

Definition 1. A set $Z \subset V \setminus \{A, Y\}$ is an $L-N$ dynamic adjustment set with respect to A, Y in $\mathcal{G}$ if $L \subset Z \subset N$ and for all conditional laws $\pi(A|L)$ for A given L , all $P \in \mathcal{M}(\mathcal{G})$, and all $y \in \mathbb{R}$

$$E_P(E_{\pi^*}[E_P\{I_{(-\infty, y]}(Y)|A, \text{pa}_{\mathcal{G}}(A), L\}|\text{pa}_{\mathcal{G}}(A), L]) = E_P(E_{\pi_Z^*}[E_P\{I_{(-\infty, y]}(Y)|A, Z\}|Z]),$$

where $I_{(-\infty, y]}(Y)$ is the indicator variable of the event $\{Y \leq y\}$, $\pi_Z^*(A|Z) \equiv \pi(A|L)$ and, recall, $\pi^*\{A|\text{pa}_{\mathcal{G}}(A), L\} \equiv \pi(A|L)$.

An $L-N$ dynamic adjustment set Z is minimal if no proper subset of Z is an $L-N$ dynamic adjustment set.

This extends the definition of refs [4] and [19] to accommodate, possibly random, L -dependent treatment rules and graphs with hidden variables. Ref. [10] also provided a characterization of minimal $L-N$ dynamic adjustment sets as minimal $A-Y$ separators in a suitably constructed undirected graph. We will review this characterization in Section 4.1. For conciseness, in what follows we drop the dynamic appellation and simply write $L-N$ adjustment sets. Also, all $L-N$ adjustment sets are with respect to A, Y in $\mathcal{G}$.

We define the cost of an $L-N$ adjustment set Z as $\sum_{W \in Z} c(W)$, and in a slight abuse of notation we denote this cost with $c(Z)$.

Definition 2. A set Z is a minimum cost $L-N$ adjustment set if it is an $L-N$ adjustment set that satisfies $c(Z) \leq c(Z')$ for all $L-N$ adjustment sets Z' .

Consider the design of a study aimed at estimating the interventional mean under an L -dependent treatment rule π . Suppose that the investigator has postulated a causal graphical model to this end, and that due to practical or ethical reasons, she can only observe a subset N of the variables in the graph $\mathcal{G}$. Suppose further that $\mathcal{G}$ includes at least one $L-N$ adjustment set Z . This implies that A, Y , and Z suffice to identify the interventional mean $\chi_{\pi}(P; \mathcal{G})$ with the so-called g-functional [1]

$$\chi_{\pi,Z}(P; \mathcal{G}) \equiv E_P[E_{\pi_Z^*}\{E_P(Y|A, Z)|Z\}],$$

which can then be estimated non-parametrically as further explained in Section 5. For economic reasons, the investigator may then choose to use a minimum cost $L-N$ adjustment set. If several minimum cost $L-N$ adjustment sets exist, then, as we further discuss in Section 5.1, a reasonable criterion for comparing them is using the variance of the limiting distribution of the resulting non-parametric estimators of the g-functional.

Our goals in this article are to:

1. Provide a graphical characterization of minimum cost $L-N$ adjustment sets.
2. Prove the existence of an optimal minimum cost $L-N$ adjustment set that yields non-parametric estimators of the interventional mean with the smallest asymptotic variance among those that control for minimum cost $L-N$ adjustments and provide a polynomial time graphical algorithm to compute it.

To achieve this, we will leverage the undirected graph defined in ref. [10], which we review next. In what follows, we will assume that there exists at least one $L-N$ adjustment set in $\mathcal{G}$.

4.1 Minimum cost adjustment sets and undirected graphs

Let $\mathcal{H}^0 \equiv \{\mathcal{G}_{\text{an}_{\mathcal{G}}(\{A, Y\} \cup L)}^{pbd}(A, Y)\}^m$ and $\text{ignore} \equiv \{\text{an}_{\mathcal{G}}(\{A, Y\} \cup L) \setminus \{A, Y\} \cap \{[V \setminus N] \cup \text{forb}(A, Y, \mathcal{G})\}\}$. Thus, ignore is the subset of the vertices $\mathcal{H}^0$ that are not equal to A or Y and that are either not observable ($V \setminus N$) or are variables that cannot be members of any $L-N$ adjustment set ($\text{forb}(A, Y, \mathcal{G})$, see ref. [4]).

Definition 3. The non-parametric adjustment efficiency graph associated with (A, Y, L, N) in $\mathcal{G}$, denoted with $\mathcal{H}^1$, is the undirected graph constructed from $\mathcal{H}^0$ by (1) removing all vertices in ignore , (2) adding an edge between any pair of remaining vertices if they were connected in $\mathcal{H}^0$ by a path with vertices in ignore , and (3) adding an edge between A and each vertex in L and between Y and each vertex in L .

In words, $\mathcal{H}^1$ is obtained from $\mathcal{H}^0$ by first performing a latent projection on $V \setminus \text{ignore}$ and then connecting all vertices in L to both A and Y . Similar constructions were also used in refs [20] and [21]. Note that even though $\mathcal{H}^0$ and $\mathcal{H}^1$ depend on A, Y, L, N , and $\mathcal{G}$, for brevity we omit this dependence in the notation.

Proposition 2 of ref. [10] states that Z is a minimal $L-N$ adjustment set if and only if Z is a minimal $A-Y$ separator in $\mathcal{H}^1$. Note that ref. [10] uses the term cut to refer to what we here call a separator. Both terms are used in the literature, and in this article, we prefer to reserve the name cut for the concept in flow networks. Now, since all minimum cost $L-N$ adjustment sets are minimal $L-N$ adjustment sets, Proposition 2 of ref. [10] implies the following.

Lemma 1. Z is a minimum cost $L-N$ adjustment set if and only if Z is a minimum cost $A-Y$ separator in $\mathcal{H}^1$.

In what follows, for brevity, all separators in $\mathcal{H}^1$ are between A and Y . In the next section, we review the aspects of the theory of non-parametric estimation of the g-functional $\chi_{\pi,Z}(P; \mathcal{G})$ that are relevant to our derivation of the optimal minimum cost $L-N$ adjustment set.

5 Non-parametric estimation of the g-functional

Ref. [22] showed that estimators of $\chi_{\pi,Z}(P; \mathcal{G})$ that are regular and asymptotically linear at all P in a model $\mathcal{P}$ that only makes assumptions on the complexity or smoothness of $b(A, Z; P) \equiv E_P(Y|A, Z)$ and/or $f(A|Z)$ have a unique influence function $\psi_{P,\pi}(Z; \mathcal{G})$ given by

$$\psi_{P,\pi}(Z; \mathcal{G}) \equiv \frac{\pi(A|L)}{f(A|Z)}\{Y - b(A, Z; P)\} + E_{\pi_Z}\{b(A, Z; P)|Z\} - \chi_{\pi,Z}(P, \mathcal{G}).$$

Note that even though $\psi_{P,\pi}$ is also a function of A and Y , this is not reflected in the notation, for the sake of brevity. In what follows, we will assume that for all $L-N$ adjustment sets Z , it holds that $\psi_{P,\pi}(Z; \mathcal{G})$ has finite variance.

There exist multiple estimation strategies that rely on making smoothness or complexity-type assumptions on $b(A, Z; P)$ and/or $f(A|Z)$. We list a few of them next. The inverse probability weighted estimator is given by $\hat{\chi}_{\pi,IPW} = \mathbb{P}_n\{\hat{f}(A|Z)^{-1}\pi(A|L)Y\}$, where $\hat{f}(A|Z)$ is a non-parametric estimator of $f(A|Z)$ [23]. The outcome regression estimator is given by $\mathbb{P}_n[E_{\pi_Z}\{\hat{b}(A, Z)|Z\}]$, where $\hat{b}$ is a non-parametric estimator of b [24]. The doubly robust estimator [25–27], also known as augmented IPW, uses non-parametric estimators of both $f(A|Z)$ and $b(A, Z)$. Examples of non-parametric estimators of $f(A|Z)$ and $b(A, Z)$ include series or kernel-based estimators, estimators based on boosted trees, and other machine learning techniques.

We will refer to estimators that are regular and asymptotically linear with unique influence function $\psi_{P,\pi}(Z; \mathcal{G})$ as non-parametric estimators that adjust for Z . It follows from the discussion above that if $\hat{\chi}_{\pi,Z}$ is a non-parametric estimator that adjusts for Z , then $n^{1/2}\{\hat{\chi}_{\pi,Z} - \chi_{\pi}(P; \mathcal{G})\}$ converges in distribution to $N\{0, \sigma_{\pi,Z}^2(P)\}$, where $\sigma_{\pi,Z}^2(P) \equiv \text{var}_P\{\psi_{P,\pi}(Z; \mathcal{G})\}$.

5.1 Efficiency comparison of minimum cost adjustment sets

Define the following preorder on the class of $L-N$ adjustment sets:

$$Z_1 \preceq_L Z_2 \Leftrightarrow \sigma_{\pi,Z_1}^2(P) \leq \sigma_{\pi,Z_2}^2(P) \text{ for all } \pi(A|L) \text{ and all } P \in \mathcal{M}(\mathcal{G}).$$

In words, $Z_1 \preceq_L Z_2$ if adjusting for Z_1 yields more efficient non-parametric estimators of $\chi_{\pi}(P; \mathcal{G})$ than adjusting for Z_2 , uniformly over all possible treatment rules π and laws P in the Bayesian network $\mathcal{M}(\mathcal{G})$. Define the following relation between separators in $\mathcal{H}^1$:

$$Z_1 \preceq_{\mathcal{H}^1} Z_2 \Leftrightarrow Y \perp_{\mathcal{H}^1} Z_2 \setminus Z_1 | Z_1 \text{ and } A \perp_{\mathcal{H}^1} Z_1 \setminus Z_2 | Z_2.$$

Since, as stated in Lemma 1, minimum cost $L-N$ adjustment sets and minimum cost separators in $\mathcal{H}^1$ are equivalent, using Lemma 1 and Propositions 3 and 5 of ref. [10], we can deduce the following graphical criterion for comparing minimum cost $L-N$ adjustment sets.

Lemma 2. *Let Z_1 and Z_2 be minimum cost $L-N$ adjustment sets. Then*

$$Z_1 \preceq_{\mathcal{H}^1} Z_2 \Rightarrow Z_1 \preceq_L Z_2. \quad (5)$$

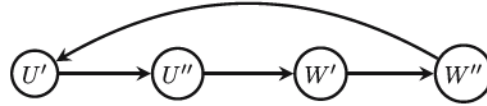
We now argue why asymptotic efficiency is a reasonable basis for comparing minimum cost adjustment sets, in the context of designing a planned study where the cost associated with each observable variable in the graph reflects the cost of measuring the variable on one subject. Consider two minimum cost adjustment sets Z_1 and Z_2 , and let $\hat{\chi}_{\pi,Z_1}$ and $\hat{\chi}_{\pi,Z_2}$ be non-parametric estimators that adjust for Z_1 and Z_2 , respectively. Suppose we know that $Z_1 \preceq_L Z_2$. If we want the length of the 95% Wald confidence interval for $\chi_{\pi}(P; \mathcal{G})$ to be bounded by M , then using $\hat{\chi}_{\pi,Z_1}$ we will need $n_1 \approx \{3.92\sigma_{\pi,Z_1}(P)M^{-1}\}^2$ samples, whereas using $\hat{\chi}_{\pi,Z_2}$ we will need $n_2 \approx \{3.92\sigma_{\pi,Z_2}(P)M^{-1}\}^2$ samples. Since $\sigma_{\pi,Z_1}^2(P) \leq \sigma_{\pi,Z_2}^2(P)$ we have that $n_1 \leq n_2$ and hence $n_1 \times c(Z_1) \leq n_2 \times c(Z_2)$. In words, for the same level of precision, the total cost of using Z_1 as an adjustment set will be lower than that of using Z_2 .

In the following section, we show that there exists a minimum cost $L-N$ adjustment set, which we denote as O_c , which satisfies that for any other minimum cost $L-N$ adjustment set Z it holds that $O_c \preceq_L Z$. We call O_c an optimal minimum cost $L-N$ adjustment set. To show this, we will make a connection between minimum cost $L-N$ adjustment sets and min-cuts in a suitably constructed flow network. This construction will also allow us to derive a polynomial time algorithm to compute O_c .

6 Optimal minimum cost adjustment sets and network flows

The following flow network construction is based on well-established ideas from the literature on networks. See Theorem 6.4 of ref. [15]. The main difference is that ref. [15] puts unit capacity on all “internal edges.”

Definition 4. Let the flow network $\mathcal{D}$ be defined as follows. For each vertex W in $\mathcal{H}^1$ add two vertices W' and W'' and the edge $W' \rightarrow W''$ to $\mathcal{D}$. We call these internal edges. If there is an edge joining U and W in $\mathcal{H}^1$ add edges $U'' \rightarrow W'$ and $W'' \rightarrow U'$. We call these external edges. Thus, an edge $U - W$ in $\mathcal{H}^1$ gives place to the following structure in $\mathcal{D}$:



The capacity of an internal edge $e = W' \rightarrow W''$ is equal to the cost of W , that is, $k(e) = c(W)$, except if W is equal to A or to Y , in which case the capacity is infinity. The capacity of external edges is infinity. We set Y'' as the source and A' as the sink of the network.

We will provide a full example of this construction in Figure 1. We show in Lemma 4 in the Appendix that there always exists a cut in $\mathcal{D}$ with finite capacity. Next, we define two mappings, $\mathfrak{d}$ and $\mathfrak{h}$. The former maps minimal separators in $\mathcal{H}^1$ to sets of vertices in $\mathcal{D}$, while the latter maps cuts with finite capacity in $\mathcal{D}$ to sets of vertices in $\mathcal{H}^1$.

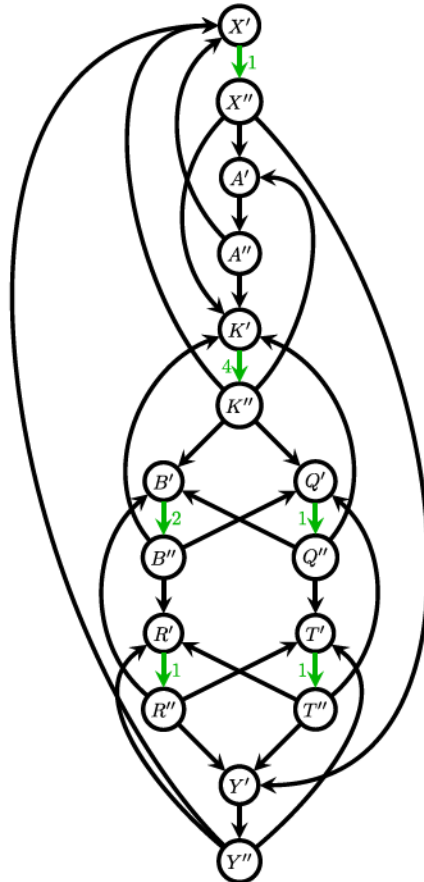


Figure 1: The flow network $\mathcal{D}$ corresponding to the directed acyclic graph in Figure 2. Edges with finite capacities are colored green, with the numbers next to the edges representing capacities. All black edges have infinite capacity.

Definition 5. Given Z a minimal separator in $\mathcal{H}^1$, we let $\partial(Z)$ be the set formed by Y'' and all vertices in $\mathcal{D}$ that lie on some directed path δ from Y'' to a vertex W' for some $W \in Z$, where δ does not intersect any other U' or U'' for $U \in Z$. Given S a cut in $\mathcal{D}$ with finite capacity we let

$$\mathfrak{h}(S) = \{W : (W', W'') \in (S, \bar{S})\}.$$

The following proposition establishes that ∂ maps minimum cost separators in $\mathcal{H}^1$ to min-cuts in $\mathcal{D}$, and $\mathfrak{h}$ maps min-cuts in $\mathcal{D}$ to minimum cost separators in $\mathcal{H}^1$.

Proposition 1.

1. Let Z be a minimal separator in $\mathcal{H}^1$. Then $\partial(Z)$ is a cut in $\mathcal{D}$ with $k\{\partial(Z)\} = c(Z)$.
2. Let S be a cut in $\mathcal{D}$ with finite capacity. Then $\mathfrak{h}(S)$ is a separator in $\mathcal{H}^1$ with $c\{\mathfrak{h}(S)\} = k(S)$.
3. Let Z be a minimum cost separator in $\mathcal{H}^1$. Then $\partial(Z)$ is a min-cut in $\mathcal{D}$.
4. Let S be a min-cut in $\mathcal{D}$. Then $\mathfrak{h}(S)$ is a minimum cost separator in $\mathcal{H}^1$.
5. Let Z be a minimum cost separator in $\mathcal{H}^1$. Then $\mathfrak{h}\{\partial(Z)\} = Z$.

Next, we establish a connection between the $\subset$ relation defined over min-cuts in $\mathcal{D}$ and the $\trianglelefteq_{\mathcal{H}^1}$ relation defined over separators in $\mathcal{H}^1$.

Proposition 2. Let S and S' be min-cuts such that $S \subset S'$. Then $\mathfrak{h}(S) \trianglelefteq_{\mathcal{H}^1} \mathfrak{h}(S')$.

Lemmas 1 and 2 together with Propositions 1 and 2 imply that if we are able to construct a min-cut S_c that is a subset of any other min-cut, then $\mathfrak{h}(S_c)$ is an optimal minimum cost $L - N$ adjustment set. We now show how such a min-cut can be constructed.

Given a flow f , we will say that a path δ connecting Y'' and W in $\mathcal{D}$ is augmenting for f if for all edges e in δ oriented from Y'' to W it holds that $f(e) < k(e)$ and for all edges e in δ oriented from W to Y'' it holds that $f(e) > 0$. Suppose that we have run a maximum flow algorithm on $\mathcal{D}$, for example, the preflow push algorithm [28], and obtained a maximum flow f^* . We are now ready to define our candidate optimal minimum cost $L - N$ adjustment set.

Definition 6. Let S_c be the set formed by Y'' and all vertices W in $\mathcal{D}$ such that there exists a path from Y'' to W that is augmenting for f^* . Let $O_c \equiv \mathfrak{h}(S_c)$.

Note that S_c could in principle depend on the computed max-flow f^* , even if this is not made explicit in the notation.

Proposition 3. S_c is a min-cut. For any other min-cut S it holds that $S_c \subset S$.

The following theorem, the main result of this article, establishes the optimality of O_c .

Theorem 1. O_c is a minimum cost $L - N$ adjustment set. For any other minimum cost $L - N$ adjustment set Z it holds that $O_c \preceq_L Z$.

We do not know if O_c is the unique optimal minimum cost $L - N$ adjustment set. Algorithm 1 summarizes the steps needed to compute O_c . The complexity of Algorithm 1 will depend on the sub-routine used to compute the maximum flow in the third step. For example, when the preflow push algorithm is used, the overall complexity of Algorithm 1 is bounded by $O(\#V^2\sqrt{\#E})$. The fourth step of Algorithm 1 can be easily implemented using a small modification of the depth first search algorithm. We provide a Python implementation of Algorithm 1 in the `optimaladj` package, available on `pip`. Our algorithm computes maximum flows using the implementation of the preflow push algorithm available in the `networkx` library [29].

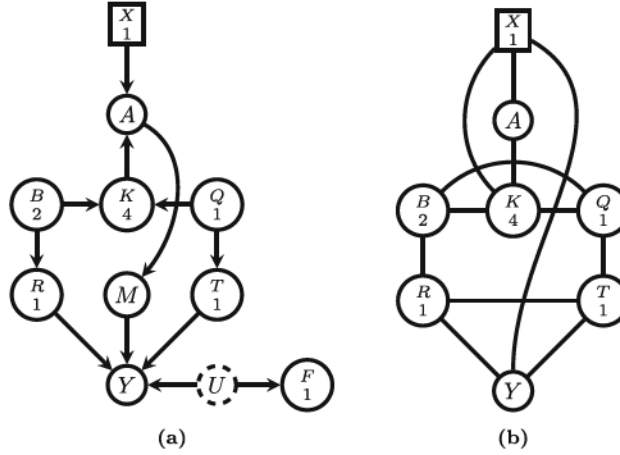


Figure 2: A directed acyclic graph $\mathcal{G}$ (a) and the undirected graph $\mathcal{H}^1$ associated with it (b).

Algorithm 1: Pseudo-algorithm to compute O_c

procedure

 construct $\mathcal{H}^1$

 construct $\mathcal{D}$

 compute a maximum flow f^* on $\mathcal{D}$

 compute S_c the set of nodes reachable from Y'' via paths that are augmenting for f^*

 compute $h(S_c)$

return $h(S_c)$

When all variables have unit costs, Algorithm 1 computes an $L - N$ adjustment set that is optimal among those of minimum cardinality. Algorithm 1 of ref. [10] does the same thing, but with $O(\#V^{3.5})$ complexity. Thus, Algorithm 1 also provides an improvement on Algorithm 1 of ref. [10] for the task of computing an optimal minimum cardinality $L - N$ adjustment set.

6.1 Examples

In the following figures we illustrate the results of this section. Dashed circles designate hidden variables and rectangles the variables that the treatment rule depends on. The numbers below the name of each vertex in $\mathcal{G}$ and $\mathcal{H}^1$ represent the cost of the variable associated with the vertex. We do not assign costs to A , Y and variables in ignore, since their costs are not relevant for the comparison of minimum-cost $L - N$ adjustment sets. Figure 1 shows the flow network $\mathcal{D}$ associated with the graphs in Figure 2. Edges with finite capacities are colored green, with the numbers next to the edges representing capacities. All black edges have infinite capacity.

For the directed acyclic graph $\mathcal{G}$ in Figure 2, let $L = \{X\}$ and $N = V \setminus \{U\}$. Then $\text{an}_{\mathcal{G}}(\{A, Y\} \cup L) = V \setminus \{F\}$, $\text{forb}(A, Y, \mathcal{G}) = \{A, Y, M\}$, and $\text{ignore} = \{U, M\}$. In $\mathcal{H}^1$, the set of all separators is given by the collection of sets Z that satisfy $X \in Z$ and at least one of the following:

- $K \in Z$.
- $B \in Z$ or $R \in Z$, and $Q \in Z$ or $T \in Z$.

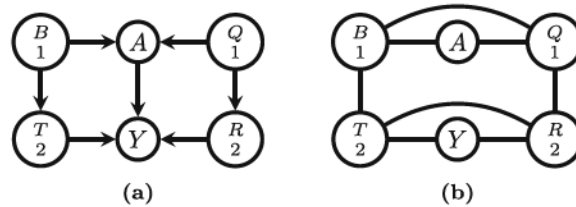


Figure 3: An example of a directed acyclic graph in which the optimal minimum cardinality $L-N$ adjustment set is more efficient than the optimal minimum cost $L-N$ adjustment set: (a) $\mathcal{G}$ and (b) $\mathcal{H}^1$.

The only minimum cost separators are $Z_1 = \{X, Q, R\}$ and $Z_2 = \{X, T, R\}$. It is easy to show that $Z_2 \trianglelefteq_{\mathcal{H}^1} Z_1$. Thus, Z_2 is an optimal minimum cost $L-N$ adjustment set. Using results from ref. [10], it is easy to show that there exists a globally optimal $L-N$ adjustment set in $\mathcal{G}$, i.e., an $L-N$ adjustment set that is more efficient than any other $L-N$ adjustment set, and that it is given by $Z_2 \cup \{F\}$. Turn now to the representation of $\mathcal{D}$ in Figure 1. The min-cut obtained by running the preflow push algorithm and then computing S_c is given by $S_c = \{Y'', X', T', R'\}$. The capacity of this cut is $k(S_c) = k\{(R', R'')\} + k\{(T', T'')\} + k\{(X', X'')\} = 3$. The optimal minimum cost $L-N$ adjustment set is then $O_c = \mathfrak{h}(S_c) = \{X, T, R\}$, matching what we obtained earlier by analyzing separators in $\mathcal{H}^1$.

It follows from Theorem 1 of ref. [10] that O_c coincides with the optimal $L-N$ adjustment set among minimal $L-N$ adjustment sets. However, it is not always the case that the optimal minimum cost and the optimal minimal $L-N$ adjustment sets are equal. For this same graph, if the cost B were 1 and the cost of R were 2, then the optimal minimal $L-N$ adjustment set would still be equal to $\{X, R, T\}$, whereas the optimal minimum cost $L-N$ adjustment set would be $\{X, B, T\}$. Since $\{X, R, T\} \trianglelefteq_{\mathcal{H}^1} \{X, B, T\}$, this is an example in which the optimal minimal $L-N$ adjustment set is more efficient than the optimal minimum cost $L-N$ adjustment set. The converse can never happen, because all minimum cost $L-N$ adjustment sets are minimal $L-N$ adjustment sets.

Going back to our original example, note that $Z_3 = \{X, K\}$ is the $L-N$ adjustment set with minimum possible cardinality. It is easy to check that $Z_2 \trianglelefteq_{\mathcal{H}^1} Z_3$ and thus in this case, the optimal minimum cost $L-N$ adjustment set is more efficient than the optimal minimum cardinality $L-N$ adjustment set. The graph in Figure 3 provides one example in which the reverse situation holds.

Indeed, for the graph $\mathcal{G}$ in Figure 3, let $L = \emptyset$ and $N = V$. Then $\text{an}_{\mathcal{G}}(\{A, Y\} \cup L) = V$, $\text{forb}(A, Y, \mathcal{G}) = \{A, Y\}$, and $\text{ignore} = \emptyset$. It is easy to check that there is only one minimum cost separator in $\mathcal{H}^1$, given by $O_c = \{B, Q\}$. However, $Z = \{T, R\}$ is a minimum cardinality separator that satisfies $Z \trianglelefteq_{\mathcal{H}^1} O_c$. Thus, in this case, the optimal minimum cardinality $L-N$ adjustment set is more efficient than the optimal minimum cost $L-N$ adjustment set.

This example also illustrates the point made in the introduction that in general there does not exist an optimal $L-N$ adjustment set among those that satisfy an upper bound on their cost. For the graph in Figure 3, if the available budget is equal to 3, the investigator has to choose between $\{B, Q\}$, $\{B, R\}$, and $\{T, Q\}$. Clearly $\{B, R\} \trianglelefteq_{\mathcal{H}^1} \{B, Q\}$ and $\{T, Q\} \trianglelefteq_{\mathcal{H}^1} \{B, Q\}$ and so by Propositions 3 and 5 of ref. [10], $\{B, R\} \trianglelefteq_L \{B, Q\}$ and $\{T, Q\} \trianglelefteq_L \{B, Q\}$. Thus, the investigator actually needs to choose between $\{B, R\}$ and $\{T, Q\}$. However, in their Example 2, ref. [9] showed that it is not possible to compare the asymptotic variances of these two adjustments based solely on the causal graph, because there exist probability laws in the Bayesian network $\mathcal{M}(\mathcal{G})$ under which $\{B, R\}$ is more efficient but also probability laws in the Bayesian network $\mathcal{M}(\mathcal{G})$ under which $\{Q, T\}$ is more efficient.

Moreover, the graph in Figure 3 can be used to show that, in general, the problem of finding an optimal efficient minimum cost adjustment set under arbitrary non-additive costs does not have a solution based on graphical criteria only. Indeed, consider the graph in Figure 3 and define the following cost structure: $c(\{B, R\}) = 1$, $c(\{T, Q\}) = 1$, and the cost of any other subset of $\{B, Q, T, R\}$ is equal to 2. Then $\{B, R\}$ and $\{T, Q\}$ are the only minimum cost adjustment sets. As we stated earlier, it is not possible to compare the asymptotic variances of these adjustment sets based solely on graphical criteria.

7 Conclusion

A reviewer suggested that, even though the problem of finding the optimal efficient adjustment set among those that satisfy a given budget constraint does not admit in general a graphical solution, an algorithm that discards provably inefficient adjustment sets among all that satisfy the budget constraint could still be useful to practitioners. We agree. A naive implementation of the suggested algorithm could go as follows. First, we use the algorithm developed in ref. [14] to list all adjustment sets in the graph. Second, we remove from this initial list all adjustment sets that have a cost greater than the budget constraint. Finally, we loop through all pairs of adjustment sets remaining in the list, and check whether the d-separation conditions in Proposition 3 from [10] hold. If they hold, then one of the adjustment sets in the pair will be inefficient and can be removed. Unfortunately, this implementation will not be computationally efficient for several reasons, including that the set of all adjustment sets can be exponentially large [14]. Whether a computationally efficient solution to this problem is possible is an interesting open problem.

Acknowledgements: The authors thank the reviewers for several useful comments and suggestions that led to improvements in the article. Andrea Rotnitzky was partially funded by the grants from the National Institutes of Health R01LM013614 and R01AI27271.

Conflict of interest: Prof. Andrea Rotnitzky is a member of the Editorial Board in the Journal of Causal Inference but was not involved in the review process of this article.

References

- [1] Robins J. A new approach to causal inference in mortality studies with a sustained exposure period-application to control of the healthy worker survivor effect. *Math Modell.* 1986;7(9–12):1393–512.
- [2] Pearl J. *Causality: models, reasoning and inference.* Cambridge, UK: Springer; 2000.
- [3] Kuroki M, Miyakawa M. Covariate selection for estimating the causal effect of control plans by using causal diagrams. *J R Statist Soc B (Statist Methodol).* 2003;65(1):209–22.
- [4] Shpitser I, VanderWeele T, Robins JM. On the validity of covariate adjustment for estimating causal effects. In: *UAI'10.* Arlington, USA: AUAI Press; 2010. p. 527–36.
- [5] Perković E, Textor J, Kalisch M, Maathuis MH. Complete graphical characterization and construction of adjustment sets in Markov equivalence classes of ancestral graphs. *J Machine Learn Res.* 2018;18(220):1–62. <http://jmlr.org/papers/v18/16-319.html>.
- [6] Kuroki M, Cai Z. Selection of identifiability criteria for total effects by using path diagrams. In: *UAI'04.* Arlington, USA: AUAI Press; 2004. p. 333–40.
- [7] Henckel L, Perković E, Maathuis MH. Graphical criteria for efficient total effect estimation via adjustment in causal linear models. 2019. [arXiv: http://arXiv.org/abs/arXiv:190702435](http://arXiv.org/abs/arXiv:190702435).
- [8] Witte J, Henckel L, Maathuis MH, Didelez V. On efficient adjustment in causal graphs. *J Machine Learn Res.* 2020;21:246.
- [9] Rotnitzky A, Smucler E. Efficient adjustment sets for population average causal treatment effect estimation in graphical models. *J Mach Learn Res.* 2020;21:1–86.
- [10] Smucler E, Sapienza F, Rotnitzky A. Efficient adjustment sets in causal graphical models with hidden variables. *Biometrika.* 2021 Mar;109:49–65. doi: 10.1093/biomet/asab018.
- [11] Runge J. Necessary and sufficient graphical conditions for optimal adjustment sets in causal graphical models with hidden variables. 2021. [arXiv: http://arXiv.org/abs/arXiv:210210324](http://arXiv.org/abs/arXiv:210210324).
- [12] Malek A, Chiappa S. Asymptotically best causal effect identification with multi-armed bandits. *Adv Neural Inform Process Sys.* 2021;34:21960–71.
- [13] Acic S, De Campos LM. An algorithm for finding minimum d-separating sets in belief networks. In: *UAI'96.* Arlington, USA: AUAI Press; 1996. p. 3–10.
- [14] van der Zander B, Liśkiewicz M, Textor J. Separators and adjustment sets in causal graphs: Complete criteria and an algorithmic framework. *Artif Intell.* 2019;270:1–40.
- [15] Even S. *Graph algorithms.* Cambridge, UK: Cambridge University Press; 2011.
- [16] Van der Vaart AW. *Asymptotic statistics.* vol. 3. Cambridge, UK: Cambridge University Press; 2000.
- [17] Spirtes P, Glymour CN, Scheines R, Heckerman D, Meek C, Cooper G, et al. *Causation, prediction, and search.* Cambridge, USA: MIT Press; 2000.

- [18] Robins JM, Richardson TS. Alternative graphical causal models and the identification of direct effects. *Causality and psychopathology: finding the determinants of disorders and their cures*. Oxford, U.K: Oxford University Press; 2010. p. 103–58.
- [19] Maathuis MH, Colombo D. A generalized back-door criterion. *Annals Statist*. 2015 June;43(3):1060–88. doi: 10.1214/14-AOS1295.
- [20] Textor J, Liskiewicz M. Adjustment criteria in causal diagrams: an algorithmic perspective. In: *UAI'11*. Arlington, USA: AUAI Press; 2011. p. 681–8.
- [21] van der Zander B, Liskiewicz M, Textor J. Constructing separators and adjustment sets in ancestral graphs. In: *UAI'14*. Arlington, USA: AUAI Press; 2014. p. 11–24.
- [22] Robins JM, Rotnitzky A. Recovery of information and adjustment for dependent censoring using surrogate markers. In: *AIDS Epidemiology*. New York, USA: Springer; 1992. p. 297–331.
- [23] Hirano K, Imbens GW, Ridder G. Efficient estimation of average treatment effects using the estimated propensity score. *Econometrica*. 2003;71(4):1161–89.
- [24] Hahn J. On the role of the propensity score in efficient semiparametric estimation of average treatment effects. *Econometrica*. 1998;66(2):315–31.
- [25] Van der Laan M, Robins JM. *Unified methods for censored longitudinal data and causality*. New York, USA: Springer Science & Business Media; 2003.
- [26] Chernozhukov V, Chetverikov D, Demirer M, Duflo E, Hansen C, Newey W, et al. Double/debiased machine learning for treatment and structural parameters. *Econometrics J*. 2018;21(1):C1–C68. <https://onlinelibrary.wiley.com/doi/abs/10.1111/ectj.12097>.
- [27] Smucler E, Rotnitzky A, Robins JM. A unifying approach for doubly-robust l_1 regularized estimation of causal contrasts. 2019. arXiv: <http://arXiv.org/abs/arXiv:190403737>.
- [28] Cheriyan J, Maheshwari S. Analysis of preflow push algorithms for maximum network flow. *SIAM J Comput*. 1989;18(6):1057–86.
- [29] Hagberg AA, Schult DA, Swart PJ. Exploring network structure, dynamics, and function using network. In: Varoquaux G, Vaught T, Millman J, editors. *Proceedings of the 7th Python in Science Conference*. Pasadena, CA USA; 2008. p. 11–15.

Appendix

This section contains the proofs of all the results in the main article, as well as preliminary technical lemmas.

We will need the following lemmas in the proofs of Propositions 1 and 2.

Lemma 3. *Let Z be a minimal separator in $\mathcal{H}^1$. Then*

$$(\partial(Z), \overline{\partial(Z)}) = \{(W', W'') : W \in Z\}.$$

Proof. We first show that $(\partial(Z), \overline{\partial(Z)}) \supset \{(W', W'') : W \in Z\}$. Since Z is a minimal separator, for any vertex $W \in Z$ there is a path connecting W and Y in $\mathcal{H}^1$ that does not intersect other vertices in Z , and such a path corresponds to a directed path in $\mathcal{D}$ from Y'' to W' that does not intersect any other vertices U' or U'' for $U \in Z$. Hence, if $W \in Z$, then $W' \in \partial(Z)$ and $W'' \notin \partial(Z)$. Thus, $(\partial(Z), \overline{\partial(Z)}) \supset \{(W', W'') : W \in Z\}$.

Next we prove that $(\partial(Z), \overline{\partial(Z)}) \subset \{(W', W'') : W \in Z\}$. Take $(B, Q) \in (\partial(Z), \overline{\partial(Z)})$, we will show that $(B, Q) \in \{(W', W'') : W \in Z\}$. We have the following four cases to analyze.

- Assume that $B = W'$ for some $W \in Z$. Due to how $\mathcal{D}$ was constructed, necessarily $Q = W''$. Thus, $(B, Q) \in \{(W', W'') : W \in Z\}$.
- Assume that $B = W'$ for some $W \notin Z$. Due to how $\mathcal{D}$ was constructed, necessarily $Q = W''$. Now, since $W' \in \partial(Z)$, there exists a directed path δ in $\mathcal{D}$ from Y'' to U' for some $U \in Z$, such that δ does not intersect X' or X'' for $X \in Z \setminus \{U\}$ and such that W' lies on δ . But δ has to go through W'' to reach U' and this implies that $W'' \in \partial(Z)$, which contradicts the assumption that $W'' = Q \in \overline{\partial(Z)}$.
- Next consider the case that $B = W''$ for some $W \in Z$. This cannot happen, since as we argued before, if $W \in Z$ then $W' \in \partial(Z)$ and $W'' \notin \partial(Z)$.
- Finally, consider the case $B = W''$ for some $W \notin Z$. Then, due to how $\mathcal{D}$ was constructed, $Q = U'$ for some U . Since $Q = U' \notin \partial(Z)$ then $U \notin Z$. Now, since $W'' \in \partial(Z)$, there exists a directed path δ in $\mathcal{D}$ from Y'' to W'' , and hence to U' , that does not intersect X' or X'' for $X \in Z$. In particular, this implies that there is a path η connecting U and Y in $\mathcal{H}^1$ that does not intersect any vertices in Z . Let ν be a path connecting U and A in $\mathcal{H}^1$. Since Z is a separator in $\mathcal{H}^1$, ν has to intersect Z . Let R be the vertex in Z that lies closest to U in ν . The sub-path of ν that goes from U to R corresponds to a directed path κ from U' to R' in $\mathcal{D}$. Joining δ and κ we obtain a directed path from Y'' to R' that does not intersect X' or X'' for $X \in Z \setminus \{R\}$. Since U' lies on that path, we obtain that $Q = U' \in \partial(Z)$, which is a contradiction.

We have thus shown that $(B, Q) \in \{(W', W'') : W \in Z\}$, finishing the proof of the lemma. $\square$

Lemma 4. *There exists a cut in $\mathcal{D}$ with finite capacity.*

Proof. Let

$$S = \{Y', Y''\} \cup \{W' : W \neq A\}.$$

This is a set of vertices of $\mathcal{D}$ that contains Y'' and does not contain A' and hence it is a cut. Its capacity is given by the sum of the capacities of all internal edges, except $Y' \rightarrow Y''$ and $A' \rightarrow A''$. Since these edges all have finite capacity, the capacity of S is finite, which is what we wanted to show. $\square$

We are now ready to prove Proposition 1.

Proof of Proposition 1. We begin with the proof of the first assertion. $\partial(Z)$ contains Y'' by definition. We will show it does not contain A' , which will prove that $\partial(Z)$ is a cut. Suppose for the sake of contradiction that $A' \in \partial(Z)$. Then there exists a directed path in $\mathcal{D}$ from Y'' to A' that does not intersect any vertices W' or W'' for $W \in Z$. This implies that there exists a path in $\mathcal{H}^1$ connecting Y to A that does not intersect Z , which

contradicts the assumption that Z is a separator in $\mathcal{H}^1$. Thus, $\partial(Z)$ is a cut. The fact that $k(\partial(Z)) = c(Z)$ follows from Lemma 3.

Next, we prove the second part of the proposition. We will first prove that $\mathfrak{h}(S)$ is a separator. If Y and A are not connected in $\mathcal{H}^1$, then $\mathfrak{h}(S)$ is trivially a separator. Suppose then that there exists a path δ that connects Y and A in $\mathcal{H}^1$. Such a path corresponds to a directed path from Y'' to A' in $\mathcal{D}$. Since S has finite capacity, any such path must contain an edge $W' \rightarrow W''$ for some $W' \in S$. This implies that δ intersects $W \in \mathfrak{h}(S)$, which is what we wanted to show. The claim that $c(\mathfrak{h}(S)) = k(S)$ follows immediately from the definition of $\mathfrak{h}(S)$.

Next, we prove part three of the proposition. Let Z be a minimum cost separator. Then it is a minimal separator, and thus by part one $\partial(Z)$ is a cut with $k(\partial(Z)) = c(Z)$. Suppose, for the sake of contradiction, that $\partial(Z)$ is not a min-cut and hence that there exists a cut S in $\mathcal{D}$ such that $k(S) < k(\partial(Z))$. Since S has finite capacity, part two of the proposition implies that $\mathfrak{h}(S)$ is a separator with $c(\mathfrak{h}(S)) = k(S)$. But then $c(\mathfrak{h}(S)) = k(S) < k(\partial(Z)) = c(Z)$, contradicting the assumption that Z was a minimum cost separator. Thus, it must be that $\partial(Z)$ is a min-cut.

Turn now to the proof of part four of the proposition. Let S be a min-cut. By Lemma 4, S has finite capacity. Then part two of the proposition implies that $\mathfrak{h}(S)$ is a separator with $c(\mathfrak{h}(S)) = k(S)$. Suppose, for the sake of contradiction, that $\mathfrak{h}(S)$ is not a minimum cost separator, and hence that there exists a minimum cost separator Z in $\mathcal{H}^1$ that satisfies $c(Z) < c(\mathfrak{h}(S))$. Since Z is a minimal separator, part one of the proposition implies that $\partial(Z)$ is a cut with $k(\partial(Z)) = c(Z)$. But then $k(\partial(Z)) = c(Z) < c(\mathfrak{h}(S)) = k(S)$, contradicting the assumption that S was a min-cut. Thus, it must be that $\mathfrak{h}(S)$ is a minimum cost separator.

Finally, we prove the fifth part of the proposition. We begin by showing that $Z \subset \mathfrak{h}(\partial(Z))$. Take $W \in Z$. We showed in Lemma 3 that $W' \in \partial(Z)$ and $W'' \notin \partial(Z)$. Thus, $W \in \mathfrak{h}(\partial(Z))$. Now we will show that $Z \supset \mathfrak{h}(\partial(Z))$. Take $W \in \mathfrak{h}(\partial(Z))$. Then $W' \in \partial(Z)$ and $W'' \notin \partial(Z)$. Assume, for the sake of contradiction, that $W \notin Z$. Since $W' \in \partial(Z)$ there exists in $\mathcal{D}$ a directed path δ from Y'' to U' for some $U' \in Z$, such that δ does not intersect any vertices X' or X'' for $X \in Z \setminus \{U\}$ and such that W' lies on δ . But since δ reaches U' , it has to go through W'' too, implying that $W'' \in \partial(Z)$, which is a contradiction. Thus, it must be that $W \in Z$. $\square$

Proof of Proposition 2. Take $W \in \mathfrak{h}(S') \setminus \mathfrak{h}(S)$ and a path δ in $\mathcal{H}^1$ connecting W to Y . We need to show that δ intersects $\mathfrak{h}(S)$. Now, in $\mathcal{D}$ there is a path corresponding to δ of the form

$$Y'' \rightarrow U_1' \rightarrow U_1'' \rightarrow \dots \rightarrow U_i' \rightarrow U_i'' \rightarrow W' \rightarrow W''.$$

Since S is a cut, $Y'' \in S$. Since S is a min-cut, by Lemma 4 it has a finite capacity, and thus it must be that $U_1' \in S$, because the edge $Y'' \rightarrow U_1'$ has infinite capacity. If $U_1'' \notin S$, then $U_1 \in \mathfrak{h}(S)$ and we are done. If $U_1'' \in S$, since S has finite capacity and the edge $U_1'' \rightarrow U_2'$ has infinite capacity it must be that $U_2' \in S$. We now repeat the same argument as before. If at some point we find that $U_j' \in S$ and $U_j'' \notin S$, then $U_j \in \mathfrak{h}(S)$ and we are done. Otherwise, all of $U_1', U_1'', \dots, U_i', U_i''$ are in S . We will show that this cannot happen. Assume it does. Since the edge $U_i'' \rightarrow W'$ has infinite capacity, it must be that $W' \in S$. If $W'' \in S$, since $S \subset S'$ we conclude that W' and W'' are both in S' , which contradicts the assumption that $W \in \mathfrak{h}(S')$. Hence, it must be that $W'' \notin S$, but this implies that $W \in \mathfrak{h}(S)$, which is a contradiction. $\square$

The following lemma is a straightforward consequence of well-known results in the theory of flow networks. We include it here for completeness sake, since we will need it in the proof of Proposition 3.

Lemma 5. Let S be a cut. Then S is a min-cut if and only if it holds that for all $e \in (S, \bar{S})$, $f^*(e) = k(e)$, and for all $e \in (\bar{S}, S)$, $f^*(e) = 0$.

Proof. We begin by noting the following. By Lemma 5.1 of ref. [15], the total flow of f^* satisfies

$$F^* = \sum_{e \in (S, \bar{S})} f^*(e) - \sum_{e \in (\bar{S}, S)} f^*(e), \quad (6)$$

and for all edges e it holds that

$$0 \leq f^*(e) \leq k(e). \quad (7)$$

Assume first that S is a min-cut. By the max-flow min-cut theorem (see Theorem 5.1 of ref. [15]), F^* satisfies

$$F^* = \sum_{e \in (S, \bar{S})} k(e). \quad (8)$$

It follows from (6), (7), and (8) that if $e \in (S, \bar{S})$, $f^*(e) = k(e)$ and if $e \in (\bar{S}, S)$, $f^*(e) = 0$, which is what we wanted to show.

Now assume that if $e \in (S, \bar{S})$, $f^*(e) = k(e)$ and if $e \in (\bar{S}, S)$, $f^*(e) = 0$. Then, by (6), the total flow of f^* satisfies (8). The max-flow min-cut theorem then implies that S is a min-cut, which is what we wanted to show. $\square$

Proof of Proposition 3. We first show that S_c is a cut. We need to show that $Y'' \in S_c$ and $A' \notin S_c$. That $Y'' \in S_c$ follows from the definition of S_c . On the other hand, since f^* is a max-flow there can be no paths from Y'' to A' that are augmenting for f^* , since if there were, the total flow of f^* could be increased. Thus, $A' \notin S_c$.

Next, we show that S_c is a min-cut. By Lemma 1, it suffices to show that if $e \in (S_c, \bar{S}_c)$ then $f^*(e) = k(e)$ and if $e \in (\bar{S}_c, S_c)$ then $f^*(e) = 0$. Take $W \in S_c$ and $U \in \bar{S}_c$. Since $W \in S_c$, there exists a path δ from Y'' to W that is augmenting for f^* . Suppose $e = (W, U)$ is an edge in $\mathcal{D}$. Then $f^*(e) = k(e)$, because if $f^*(e) < k(e)$ the path obtained by joining δ and e would be a path from Y'' to U that is augmenting for f^* , implying that $U \in S_c$, which is a contradiction. Suppose that $e = (U, W)$ is an edge in $\mathcal{D}$. Then $f^*(e) = 0$, because if $f^*(e) > 0$ the path obtained by joining δ and e would be a path from Y'' to U that is augmenting for f^* , implying that $U \in S_c$, which is a contradiction. We have thus shown that S_c is a min-cut.

Now take any other min-cut S . We will show that $S_c \subset S$. Take $U \in S_c$. We need to show that $U \in S$. Since $U \in S_c$, there exists a path δ from Y'' to U in $\mathcal{D}$ that is augmenting for f^* . Suppose the vertices in δ are $Y'', W_1, W_2, \dots, W_l, W_{l+1} = U$. Since S is a cut, we have that $Y'' \in S$. We will show that $W_i \in S$ for all $i = 1, \dots, l+1$ by induction. Let e_1 be the edge joining Y'' and W_1 in δ . Since δ is augmenting for f^* , we have that if $e_1 = (Y'', W_1)$ then $f^*(e_1) < k(e_1)$ whereas if $e_1 = (W_1, Y'')$ then $f^*(e_1) > 0$. Since S is a min-cut, Lemma 5 implies that $W_1 \in S$. Now, suppose that for some $1 \leq i < l+1$ it holds that $W_i \in S$. Let e_{i+1} be the edge joining W_i and W_{i+1} in δ . Since δ is augmenting for f^* , we have that if $e_{i+1} = (W_i, W_{i+1})$ then $f^*(e_{i+1}) < k(e_{i+1})$ whereas if $e_{i+1} = (W_{i+1}, W_i)$ then $f^*(e_{i+1}) > 0$. Since S is a min-cut, Lemma 5 implies that $W_{i+1} \in S$. This finishes the proof of the proposition. $\square$

Proof of Theorem 1. By Proposition 3, S_c is a min-cut. Thus, by part four of Proposition 1, $O_c = \mathfrak{h}(S_c)$ is a minimum cost separator in $\mathcal{H}^1$. Lemma 1 implies that O_c is a minimum cost $L-N$ adjustment set.

Now, let Z be any other minimum cost $L-N$ adjustment set. We will show that $O_c \preceq_L Z$. Lemma 1 implies that Z is a minimum cost separator in $\mathcal{H}^1$. By part three of Proposition 1, $\mathfrak{d}(Z)$ is a min-cut in $\mathcal{D}$. Thus, Proposition 3 implies that $S_c \subset \mathfrak{d}(Z)$. Proposition 2 implies that $\mathfrak{h}(S_c) \preceq_{\mathcal{H}^1} \mathfrak{h}(\mathfrak{d}(Z))$. But part five of Proposition 1 establishes that $\mathfrak{h}(\mathfrak{d}(Z)) = Z$. We have shown that $O_c \preceq_{\mathcal{H}^1} Z$, which by Lemma 2 implies that $O_c \preceq_L Z$. This finishes the proof of the theorem. $\square$