Optimization Algorithm Design via Electric Circuits

Stephen P. Boyd
Stanford University

Tetiana Parshakova*
Stanford University

Ernest K. Ryu
UCLA

Jaewook J. Suh*
Rice University

Abstract

We present a novel methodology for convex optimization algorithm design using ideas from electric RLC circuits. Given an optimization problem, the first stage of the methodology is to design an appropriate electric circuit whose continuous-time dynamics converge to the solution of the optimization problem at hand. Then, the second stage is an automated, computer-assisted discretization of the continuous-time dynamics, yielding a provably convergent discrete-time algorithm. Our methodology recovers many classical (distributed) optimization algorithms and enables users to quickly design and explore a wide range of new algorithms with convergence guarantees.

1 Introduction

In the classical literature of optimization theory, optimization algorithms are designed with the goal of establishing fast worst-case convergence guarantees. However, these methods, designed with the pessimistic framework of worst-case analysis, often exhibit slow practical performance. In the modern machine learning literature, optimizers are designed with the goal of obtaining fast empirical performance on a set of practical problems of interest. However, these methods, designed without consideration of the feasibility of a convergence analysis, tend to be much more difficult to analyze theoretically, and such methods sometimes even fail to converge under nice idealized assumptions such as convexity [92, 131].

In this work, we present a novel methodology for convex optimization algorithm design using ideas from electric RLC (resistor-inductor-capacitor) circuits and the performance estimation problem [54, 150]. (To clarify, our proposal does not involve building a physical circuit.) Specifically, our methodology provides a quick and systematic recipe for designing new, provably convergent optimization algorithms, including distributed optimization algorithms. The ease of the methodology enables users to quickly explore a wide range of algorithms with convergence guarantees.

Optimization problem formulation. We consider the standard-form optimization problem

$$\begin{aligned} \text{minimize} \quad & f(x) \\ \text{subject to} \quad & x \in \mathcal{R}(E^\mathsf{T}), \end{aligned} \tag{1}$$

where $x \in \mathbf{R}^m$ is the optimization variable, $f \colon \mathbf{R}^m \to \mathbf{R} \cup \{\infty\}$ is closed, convex, and proper, and $E \in \mathbf{R}^{n \times m}$. Assume we have n nets $N_1, \ldots, N_n$ forming a partition of $\{1, \ldots, m\}$. More specifically, we let $E \in \mathbf{R}^{n \times m}$ be a *selection matrix* defined as

$$E_{ij} = \begin{cases} +1 & \text{if } j \in N_i \\ 0 & \text{otherwise.} \end{cases} \tag{2}$$

Our goal is to find a primal-dual solution satisfying the KKT conditions [126, Theorem 28.3]

$$y \in \partial f(x), \quad x \in \mathcal{R}(E^\mathsf{T}), \quad y \in \mathcal{N}(E). \tag{3}$$

As we show through examples, this standard-form problem (1) conveniently models many optimization problem setups of practical interest.

---

*Lead authors (author list ordered alphabetically)

38th Conference on Neural Information Processing Systems (NeurIPS 2024).

In the analysis and design of optimization algorithms, a standard approach is to consider a continuous-time model of a given algorithm, corresponding to the limit of small stepsizes [124, 79, 6, 156, 143, 93, 86, 137]. Our work is based on the key observation that such continuous-time models can be interpreted as RLC circuits connected to the subdifferential operator ∂f, which we interpret as a nonlinear resistor. We expand on this observation and propose a general methodology for designing optimization algorithms by designing RLC circuits that relax to the nets defined by E.

Example. Problem (1) represents a general form of distributed optimization, where the constraints enforce consensus among the primal variables. An example is the so-called consensus problem [30, §7]

$$\begin{array}{ll} \underset{x_1,\ldots,x_N \in \mathbf{R}^{m/N}}{\text{minimize}} & f_1(x_1) + \cdots + f_N(x_N) \\ \text{subject to} & x_1 = \cdots = x_N, \end{array}$$

where $x = (x_1, \ldots, x_N)$ is the decision variable, the objective function $f(x) = f_1(x_1) + \cdots + f_N(x_N)$ is block-separable, and $E^{\mathsf{T}} = (I, \ldots, I) \in \mathbf{R}^{m \times m/N}$. Refer to sections §E and §F, for an overview of classical splitting methods and decentralized methods for solving (1).

1.1 Preliminaries and Notation

We generally follow the standard definitions and notations of convex optimization [31, 118, 23, 119, 129]. Consider the extended-valued function $f \colon \mathbf{R}^n \to \mathbf{R} \cup \{\infty\}$. We say f is closed if its epigraph is closed set in $\mathbf{R}^{n+1}$ and proper if its value is finite somewhere. We say f is CCP if it is closed, convex, and proper. For $R > 0$, we say f is R-smooth if f is finite and differentiable everywhere and $\|\nabla f(x) - \nabla f(y)\| \leq R\|x - y\|$ for all $x, y \in \mathbf{R}^n$. For $\mu > 0$, we say f is μ-strongly convex if $f(x) - (\mu/2)\|x\|^2$ is convex. Let $f^*(y) = \sup_{x \in \mathbf{R}^n}\{\langle y, x \rangle - f(x)\}$ denote the Fenchel conjugate of f. For $R > 0$ and a CCP f, define the R-Moreau envelope of f as $^R f(x) = \inf_{z \in \mathbf{R}^n} \left\{ f(z) + \frac{1}{2R}\|z - x\|^2 \right\}$. One can show [23, Proposition 13.24] that the R-Moreau envelope is given by $^R f = \left(f^* + \frac{R}{2}\| \cdot \|^2 \right)^*$. If f is $1/R$-smooth, we can define [23, Theorem 18.15] the R-pre-Moreau envelope of f as

$$\tilde{f} = \left(f^* - \tfrac{R}{2}\| \cdot \|^2 \right)^*,$$

which is defined such that $^R(\tilde{f}) = f$.

Due to the limited space, we defer the review of prior works to §A of the appendix.

1.2 Contributions

Our work presents two technical novelties, one in continuous time and the other in discrete time. The first is the observation that many standard optimization algorithms can be interpreted as discretizations of electric RLC circuits connected to the subdifferential operator ∂f. The second is the use of the performance estimation problem to obtain an automated recipe for discretizing convergent continuous-time dynamics into convergent discrete-time algorithms, and we provide code implementing our automatic discretization methodology.

By combining these two insights, we provide a quick and systematic methodology for designing new, provably convergent optimization algorithms, including distributed optimization algorithms. We provide an open-source package that implements automatic discretization of our circuits:

```
https://github.com/cvxgrp/optimization_via_circuits
```

2 Continuous-time optimization with circuits

2.1 Interconnects

We now describe two types of electric circuits that we call *static* and *dynamic interconnects*. Both interconnects have m terminals, and we will later connect them to the m inputs of ∂f.

2

Static interconnect. The static interconnect is a set of (ideal) wires connecting m terminals and forming n nets. See Figure 1 for an example. Let $x \in \mathbf{R}^m$ be a vector of terminal potentials and $y \in \mathbf{R}^m$ be a vector of currents leaving the terminals. Using matrix $E \in \mathbf{R}^{n \times m}$ as defined in (2), we can express Kirchhoff's voltage law (KVL) as $x \in \mathcal{R}(E^\mathsf{T})$ and Kirchhoff's current law (KCL) as $y \in \mathcal{N}(E)$. In other words, the static interconnect enforces the V-I relationship

$$(x, y) \in \mathcal{R}(E^\mathsf{T}) \times \mathcal{N}(E). \tag{4}$$

Dynamic interconnect. The dynamic interconnect is an RLC circuit with m terminals and 1 ground node. We assume all inductances and capacitances have values in $(0, \infty)$ while the resistances have values in $[0, \infty)$. (A 0-ohm resistor is an ideal wire. We do not permit ideal wire loop.) Each RLC component has two (scalar-valued) terminals: the $+$ and $-$ terminals.

Static interconnect

Figure 1: Example of a static interconnect, $m = 5$, $N_1 = \{1, 3\}$, $N_2 = \{2, 4\}$, $N_3 = \{5\}$.

Denote the number of nodes in the RLC circuit by τ. Connect nodes $1, 2, \ldots, m$ to terminals $1, 2, \ldots, m$, and let the last node, node τ, be the ground node. (This implies $\tau \geq m + 1$.) Denote the number of RLC components by σ. We describe the topology with a reduced node incidence matrix (with the bottom row corresponding to the ground node removed) $A \in \mathbf{R}^{(\tau-1) \times \sigma}$ defined as

$$A_{ij} = \begin{cases} +1 & \text{if node } i \text{ connects to } + \text{ terminal of component } j \\ -1 & \text{if node } i \text{ connects to } - \text{ terminal of component } j \\ 0 & \text{otherwise.} \end{cases}$$

See Figure 2 for an example.

Dynamic interconnect

$$A = \begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ 7 \end{array} \begin{array}{c} \begin{array}{ccccccc} R_1 & R_2 & R_3 & L_1 & L_2 & C_1 & C_2 \end{array} \\ \left[\begin{array}{ccccccc} +1 & 0 & 0 & -1 & 0 & 0 & 0 \\ 0 & +1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & +1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -1 & +1 \\ -1 & 0 & 0 & +1 & +1 & 0 & 0 \\ 0 & -1 & -1 & 0 & 0 & +1 & 0 \end{array} \right] \end{array}$$

Figure 2: Example of a dynamic interconnect with $\tau = 8$ nodes, $\sigma = 7$ RLC components, $m = 5$ terminals, and 1 ground node. Reduced node incidence matrix A is provided. (R_2 and R_3 are 0-ohm resistors.) This dynamic interconnect is admissible with respect to the static interconnect of Figure 1.

The ground node is designated to have 0 potential, and the *potential* of any node is the potential relative to ground. The *voltage* across a component is the difference of potentials between the $+$ and $-$ terminals. The *current* through a component is defined as the current flowing from the $+$ terminal to the $-$ terminal.

Let $x \in \mathbf{R}^m$ be the potentials at the m terminals, which are connected to nodes $1, \ldots, m$, and $y \in \mathbf{R}^m$ be the currents leaving the terminals. Denote the node potential vector with the ground node excluded (since the potential at ground is 0) by

$$\begin{bmatrix} x \\ e \end{bmatrix} \in \mathbf{R}^{\tau-1}.$$

So, $e \in \mathbf{R}^{\tau-1-m}$ denotes the potentials at the non-terminal nodes. Denote the vector of voltages by $v \in \mathbf{R}^\sigma$ and the vector of currents by $i \in \mathbf{R}^\sigma$. Then, the currents and voltages of the dynamic

3

interconnect satisfy the following V-I relations

$$\text{(i)} \; Ai = \begin{bmatrix} -y \\ 0 \end{bmatrix} \quad \text{(KCL)} \qquad \text{(ii)} \; v = A^{\mathsf{T}} \begin{bmatrix} x \\ e \end{bmatrix} \quad \text{(KVL)}$$

$$\text{(iii)} \; v_{\mathcal{R}} = D_{\mathcal{R}} i_{\mathcal{R}} \quad \text{(Resistor)} \qquad \text{(iv)} \; v_{\mathcal{L}} = D_{\mathcal{L}} \frac{d}{dt} i_{\mathcal{L}} \quad \text{(Inductor)} \qquad \text{(v)} \; i_{\mathcal{C}} = D_{\mathcal{C}} \frac{d}{dt} v_{\mathcal{C}} \quad \text{(Capacitor)}$$

where $D_{\mathcal{R}}$, $D_{\mathcal{L}}$, and $D_{\mathcal{C}}$ are diagonal matrices respectively with resistances, inductances, and capacitances values in the diagonals.

Admissibility. When an RLC circuit reaches equilibrium, voltages across inductors and currents through capacitors are 0. We say a dynamic interconnect is *admissible* if it relaxes to the static interconnect at equilibrium. Mathematically, this condition is expressed as

$$\left\{ (x, y) \, \middle| \, Ai = \begin{bmatrix} -y \\ 0 \end{bmatrix}, v = A^{\mathsf{T}} \begin{bmatrix} x \\ e \end{bmatrix}, v_{\mathcal{R}} = D_{\mathcal{R}} i_{\mathcal{R}}, v_{\mathcal{L}} = 0, i_{\mathcal{C}} = 0 \right\} = \mathcal{R}(E^{\mathsf{T}}) \times \mathcal{N}(E).$$

As an example, the dynamic interconnect of Figure 2 is admissible with respect to the static interconnect of Figure 1.

2.2 Composing interconnects with ∂f

We view the subdifferential operator ∂f as an m-terminal electric device that is also grounded. Let $x \in \mathbf{R}^m$ be the potentials at the m terminals (excluding ground) and $y \in \mathbf{R}^m$ be the currents flowing into the m terminals. The ∂f operator enforces the V-I relation

$$y \in \partial f(x).$$

We connect the m terminals of ∂f to the m terminals of the static and dynamic interconnects. Immediately, connecting the static interconnect with ∂f enforces the V-I relations (4) and $y \in \partial f(x)$, which combine to be the optimality condition (3). Therefore, the potentials at the m terminals as a vector in $\mathbf{R}^m$ is an optimal $x^\star \in \mathbf{R}^m$ solving (1). To clarify, connecting the static interconnect with ∂f leads to a *static* circuit in the sense that the potential x and current y do not depend on time.

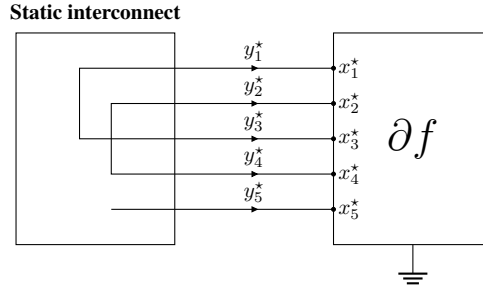



Figure 3: The static interconnect of Figure 1 connected with ∂f. The potentials at the m terminals is an optimal $x^\star \in \mathbf{R}^m$ solving (1).

Next, we compose (connect) the dynamic interconnect with ∂f. Due to capacitors and inductors, this circuit is *dynamic* in the sense that the voltages $v(t)$ and $x(t)$ and currents $i(t)$ and $y(t)$ depend on time, although we often omit explicitly writing the t-dependence for notational convenience. Then, the V-I relations of the dynamic interconnect combined with $y \in \partial f(x)$ leads to the V-I relation

$$\left\{ (v, i) \, \middle| \, y \in \partial f(x), \; Ai = \begin{bmatrix} -y \\ 0 \end{bmatrix}, \; v = A^{\mathsf{T}} \begin{bmatrix} x \\ e \end{bmatrix}, \right. \tag{5}$$

$$\left. v_{\mathcal{R}} = D_{\mathcal{R}} i_{\mathcal{R}}, \; v_{\mathcal{L}} = D_{\mathcal{L}} \frac{d}{dt} i_{\mathcal{L}}, \; i_{\mathcal{C}} = D_{\mathcal{C}} \frac{d}{dt} v_{\mathcal{C}}, \; t \in (0, \infty) \right\},$$

where $v(t) = (v_{\mathcal{R}}(t), v_{\mathcal{L}}(t), v_{\mathcal{C}}(t)) \in \mathbf{R}^\sigma$, $i(t) = (i_{\mathcal{R}}(t), i_{\mathcal{L}}(t), i_{\mathcal{C}}(t)) \in \mathbf{R}^\sigma$, $e(t) \in \mathbf{R}^{\tau-m-1}$, $x(t) \in \mathbf{R}^m$, and $y(t) \in \mathbf{R}^m$ for $t \in [0, \infty)$.

4

Dynamic interconnect

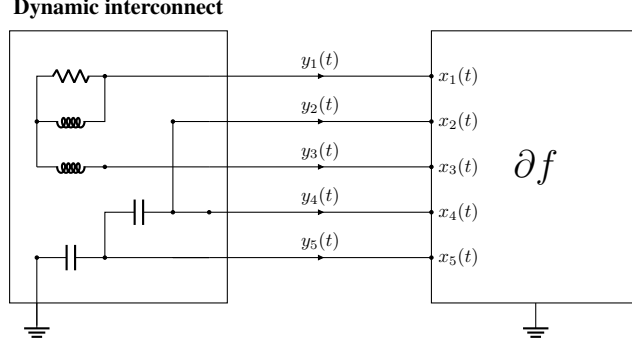



Figure 4: The dynamic interconnect of Figure 2 connected with ∂f. The potentials at the m terminals satisfy $x(t) \to x^\star$ for an optimal $x^\star \in \mathbf{R}^m$ solving (1) under the conditions of Theorem 2.2.

Under appropriate conditions, the dynamics (5) is mathematically well-posed in the sense that there exist unique Lipschitz-continuous curves $v(t)$, $i(t)$, $x(t)$, and $y(t)$ satisfying the V-I relation (5) as formalized in the following Theorem 2.1. The proof, which utilizes the machinery of monotone operator theory [21, 23, 129], is provided in §B of the appendix.

Theorem 2.1. *Assume f is μ-strongly convex and M-smooth. Suppose (v^0, i^0, x^0, y^0) satisfy*

$$Ai^0 = \begin{bmatrix} -y^0 \\ 0 \end{bmatrix}, \quad v^0 = A^\mathsf{T} \begin{bmatrix} x^0 \\ e \end{bmatrix}, \quad v_{\mathcal{R}}^0 = D_{\mathcal{R}} i_{\mathcal{R}}^0, \quad y^0 = \nabla f(x^0).$$

Then there is a unique Lipschitz continuous curve $(v, i, x, y)\colon [0, \infty) \to \mathbf{R}^\sigma \times \mathbf{R}^\sigma \times \mathbf{R}^m \times \mathbf{R}^m$ satisfying the conditions in (5) *and the initial condition $(v(0), i(0), x(0), y(0)) = (v^0, i^0, x^0, y^0)$.*

Equillibrium yields a primal-dual solution. With the dynamic interconnect composed with ∂f, we generically expect the circuit state $(v(t), i(t), x(t), y(t))$ to converge (relax) to an equilibrium state. The admissibility condition ensures that at such an equilibrium, (x, y) will be a primal-dual solution. We formally state this fact as Theorem C.2 of the Appendix.

2.3 Energy dissipation

Let $(v^\star, i^\star, x^\star, y^\star)$ be an equilibrium of an admissible dynamic interconnect composed with ∂f. Since the voltages across resistors and inductors and the currents through capacitors are zero under equilibrium, we have

$$v^\star = (v_{\mathcal{R}}^\star, v_{\mathcal{L}}^\star, v_{\mathcal{C}}^\star) = (0, 0, v_{\mathcal{C}}^\star), \qquad i^\star = (i_{\mathcal{R}}^\star, i_{\mathcal{L}}^\star, i_{\mathcal{C}}^\star) = (0, i_{\mathcal{L}}^\star, 0).$$

(We formally show this in Theorem C.2 of the appendix.) Define the energy of the circuit at time t as

$$\mathcal{E}(t) = \frac{1}{2}\|v_{\mathcal{C}}(t) - v_{\mathcal{C}}^\star\|_{D_{\mathcal{C}}}^2 + \frac{1}{2}\|i_{\mathcal{L}}(t) - i_{\mathcal{L}}^\star\|_{D_{\mathcal{L}}}^2, \tag{6}$$

which is a dissipative (non-increasing) quantity:

$$\frac{d}{dt}\mathcal{E} = \langle v_{\mathcal{C}} - v_{\mathcal{C}}^\star, i_{\mathcal{C}} - i_{\mathcal{C}}^\star \rangle + \langle i_{\mathcal{L}} - i_{\mathcal{L}}^\star, v_{\mathcal{L}} - v_{\mathcal{L}}^\star \rangle$$
$$= -\|i_{\mathcal{R}}\|_{D_{\mathcal{R}}}^2 - \langle x - x^\star, y - y^\star \rangle \le 0. \tag{7}$$

Here, we use $i_{\mathcal{C}}^\star = 0$ and $v_{\mathcal{L}}^\star = 0$ and the fact that the power dissipated by the resistors and ∂f must come from the energy stored in the capacitors and inductors. This dissipativity property leads to the following continuous-time convergence.

Theorem 2.2. *Assume $f\colon \mathbf{R}^m \to \mathbf{R}$ is strongly convex and smooth. Assume the dynamic interconnect is admissible, and let $(x^\star, y^\star)$ be a primal-dual solution pair. Let $(v(t), i(t), x(t), y(t))$ be a curve satisfying* (5). *Then,*

$$\lim_{t \to \infty} (x(t), y(t)) = (x^\star, y^\star).$$

Theorem 2.2 largely follows as a corollary of Theorem 2.1. The formal proof is provided in §D of the appendix. In §4, we present a systematic framework for finding discretized versions of Theorem 2.2 the corresponding discretized algorithms.

5

3 Circuits for classical algorithms

In this section, we present circuits recovering the classical Nesterov acceleration, decentralized ADMM, and PG-EXTRA. For additional examples and detailed derivations, refer to §E and §F of the appendix, where we provide circuits and analyses of classical algorithms such as gradient descent [35], proximal point method [125], proximal gradient method [42], primal decomposition [72, 140], dual decomposition [59, 96, 63, 142], Douglas–Rachford splitting [123, 53, 101], Davis–Yin splitting [46], decentralized gradient descent [116, 171], and diffusion [33, 34].

Multi-wire notation. We start by quickly introducing the multi-wire notation depicted in Figure 5. When optimizing $f \colon \mathbf{R}^m \to \mathbf{R}$ and using the m-terminal device ∂f, we will often use dynamic interconnects that have the same RLC circuit across each net, *i.e.*, the dynamic interconnect consists of m identical copies of the same RLC circuit for the m coordinates of $x \in \mathbf{R}^m$. In this case, we use the diagonal-line notation depicted in Figure 5.

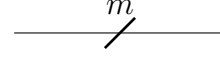
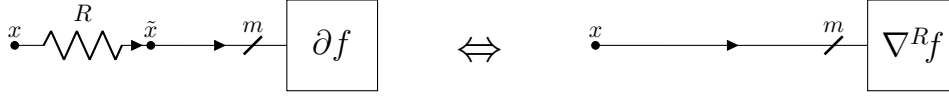



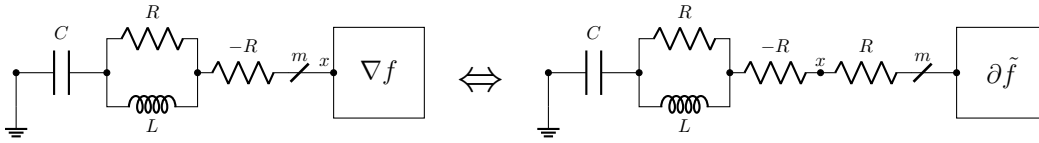
Figure 5: Multi-wire notation.

Moreau envelope. We use the following simple identity throughout this work: ∂f composed with a resistor is equivalent to $\nabla^R f(x)$.

To clarify, the equivalence means the two circuits impose the same V-I relation on the m pins of x. To see this, note $\big[\partial f(\tilde{x}) = \frac{1}{R}(x - \tilde{x})\big] \Leftrightarrow \big[\tilde{x} = \mathbf{prox}_{Rf}(x)\big]$ and use the identity for the gradient of the Moreau envelope to conclude

$$\nabla^R f(x) = \frac{1}{R}(x - \mathbf{prox}_{Rf}(x)) = \frac{1}{R}(x - \tilde{x}).$$

See §E.1 of the appendix for further details.

3.1 Nesterov acceleration

Let $f \colon \mathbf{R}^m \to \mathbf{R}$ be a $1/R$-smooth convex function. Then, the circuit corresponding to the classical Nesterov acceleration is given below.

The use of a *negative* resistor $-R$ may seem unconventional, but the fact that this circuit is stable is easier to see if we consider the equivalent circuit with the pre-Moreau envelope $\tilde{f}$, *i.e.*, $\tilde{f}$ is the convex function such that ${}^R\tilde{f} = f$. To clarify, negative resistors satisfy the same V-I relations of the standard resistors but with a negative slope. Negative resistors have also been considered in [153].

The V-I relations of this circuit lead to the ODE

$$\frac{d^2}{dt^2}x + \frac{R}{L}\frac{d}{dt}x + \left(\frac{1}{C} - \frac{R^2}{L}\right)\frac{d}{dt}\nabla f(x) + \frac{R}{LC}\nabla f(x) = 0.$$

If we set $R = \sqrt{L/C}$, which can be interpreted as an instance of critical damping [164, 174, 40], $L = \frac{1}{8\mu\sqrt{\mu}}$, and $C = 2\sqrt{\mu}$, we recover the Nesterov ODE [162]

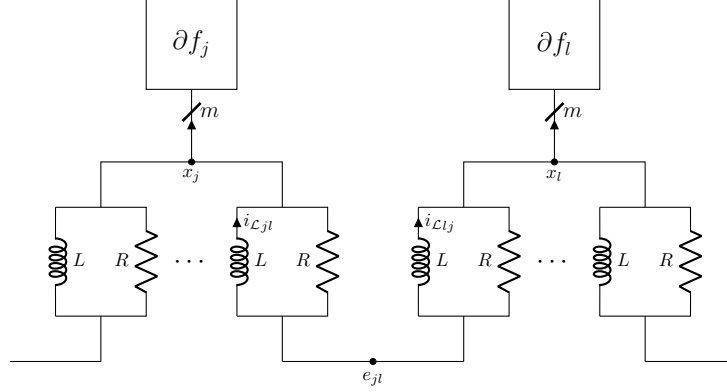
$$\frac{d^2}{dt^2}x + 2\sqrt{\mu}\frac{d}{dt}x + \nabla f(x) = 0.$$

We also quickly point out that other choices of parameters lead to the high-resolution ODE introduced in [136]. See §E.3 of the appendix for further details.

6

3.2 Decentralized ADMM

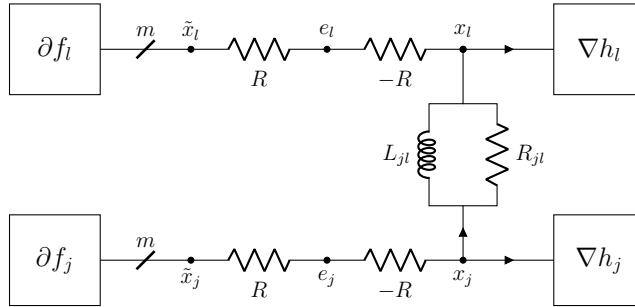
Let $f_1, \ldots, f_N \colon \mathbf{R}^m \to \mathbf{R} \cup \{\infty\}$ be CCP functions. Consider a decentralized optimization setup with graph G. We provide the full description of the decentralized setup and notations in §F of the appendix. Define Γ_j to be the neighbors of j in graph G. For simplicity, we only illustrate the circuit related to nodes j and l, where j and l are directly connected through an edge in the graph G.

The circuit corresponding to decentralized ADMM [74, 71, 70, 157, 139] is given below.

In the following, the left column presents the dynamics of the continuous-time circuit and the right column presents the discretization with stepsize L/R, recovering the standard decentralized ADMM:

$$
\begin{aligned}
a_j &= \frac{1}{|\Gamma_j|} \sum_{l \in \Gamma_j} (R i_{\mathcal{L}jl} + e_{jl}) & a_j^{k+1} &= \frac{1}{|\Gamma_j|} \sum_{l \in \Gamma_j} (R i_{\mathcal{L}j,l}^k + e_{jl}^k) \\
x_j &= \mathbf{prox}_{(R/|\Gamma_j|)f_j}(a_j) & x_j^{k+1} &= \mathbf{prox}_{(R/|\Gamma_j|)f_j}\left(a_j^{k+1}\right) \\
e_{jl} &= \frac{1}{2}(x_j + x_l) & e_{jl}^{k+1} &= \frac{1}{2}(x_j^{k+1} + x_l^{k+1}) \\
\frac{d}{dt} i_{\mathcal{L}jl} &= \frac{1}{L}(e_{jl} - x_j) & i_{\mathcal{L}\ jl}^{k+1} &= i_{\mathcal{L}jl}^k + \frac{1}{R}(e_{jl}^{k+1} - x_j^{k+1})
\end{aligned}
$$

for every node $j = 1, \ldots, N$ and every edge (j, l) in graph G.

3.3 PG-EXTRA

Let $f_1, \ldots, f_N \colon \mathbf{R}^m \to \mathbf{R} \cup \{\infty\}$ be CCP functions and $h_1, \ldots, h_N \colon \mathbf{R}^m \to \mathbf{R}$ be convex M-smooth functions. Consider a decentralized optimization setup with graph G. The circuit corresponding to PG-EXTRA [138] is given below.

Define the mixing matrix $W \in \mathbf{R}^{N \times N}$ with

$$
W_{jl} = \begin{cases} 1 - \sum_{l \in \Gamma_j} \frac{R}{R_{jl}} & \text{if } j = l \\ \frac{R}{R_{jl}} & \text{if } j \neq l, \quad l \in \Gamma_j \\ 0 & \text{otherwise.} \end{cases}
$$

7

In the following, the left column presents the V-I relations for the continuous-time circuit and the right column presents the discretization with stepsize $\frac{1}{2}$, recovering the standard PG-EXTRA:

$$x_j = \mathbf{prox}_{Rf_j} \left(\sum_{l=1}^{N} W_{jl} x_l - R\nabla h_j(x_j) - w_j \right) \quad \bigg| \quad x_j^{k+1} = \mathbf{prox}_{Rf_j} \left(\sum_{l=1}^{N} W_{jl} x_l^k - R\nabla h_j(x_j^k) - w_j^k \right)$$

$$\frac{d}{dt} w_j = x_j - \sum_{l=1}^{N} W_{jl} x_l \quad \bigg| \quad w_j^{k+1} = w_j^k + \frac{1}{2}(x_j^k - \sum_{l=1}^{N} W_{jl} x_l^k)$$

for every node $j = 1, \ldots, N$ and every edge (j, l) in graph G.

4 Automatic discretization

We discretize the continuous-time dynamics given by the circuit with an admissible dynamic interconnect using a two-stage Runge–Kutta method with parameters α, β and stepsize $h > 0$. The explicit form of the discretization is stated in §G of the appendix. Let $\{(v^k, i^k, x^k, y^k)\}_{k=1}^{\infty}$ be the iterates generated by the discretized algorithm. Then the energy stored in the circuit at time $t = kh$ is

$$\mathcal{E}_k = \frac{1}{2}\|v_{\mathcal{C}}^k - v_{\mathcal{C}}^\star\|_{D_C}^2 + \frac{1}{2}\|i_{\mathcal{L}}^k - i_{\mathcal{L}}^\star\|_{D_{\mathcal{L}}}^2.$$

To guarantee convergence of the discretized algorithm, we search for discretization parameters that ensure the $\mathcal{E}_1, \mathcal{E}_2, \ldots$ sequence is dissipative in the following sense. Specifically, we say the algorithm or the discretization is *sufficiently dissipative* if there is an $\eta > 0$ such that

$$\mathcal{E}_{k+1} - \mathcal{E}_k + \eta\langle x^k - x^\star, y^k - y^\star \rangle \leq 0, \tag{8}$$

holds for all $k = 1, 2, \ldots$. This requirement is analogous to the "sufficient decrease" conditions in optimization [31, 121]. The following Lemma 4.1, which proof we provide in §G of the appendix, states that sufficient dissipativity ensures convergence under suitable conditions.

Lemma 4.1. *Assume $f \colon \mathbf{R}^m \to \mathbf{R} \cup \{\infty\}$ is a strictly convex function and the dynamic interconnect is admissible. If the two-stage Runge–Kutta discretization, as explicitly stated in §G of the appendix, generates a discrete-time sequence $\{(v^k, i^k, x^k, y^k)\}_{k=1}^{\infty}$ satisfying the sufficient dissipativity condition* (8)*, then x^k converges to a primal solution.*

We find such a discretization with the following automated methodology. Given a discretization characterized by (α, β, h), the dissipativity condition (8) for a given $\eta > 0$ is implied if the optimal value of the following optimization problem is non-positive:

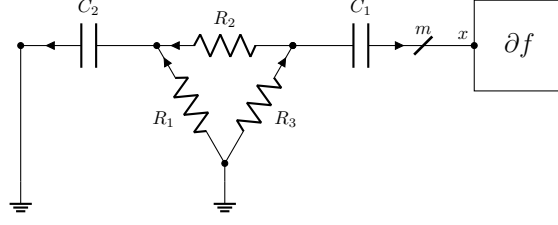
$$\begin{aligned}
\text{maximize} \quad & \mathcal{E}_2 - \mathcal{E}_1 + \eta\langle x^1 - x^\star, y^1 - y^\star \rangle \\
\text{subject to} \quad & \mathcal{E}_s = \frac{1}{2}\|v_{\mathcal{C}}^s - v_{\mathcal{C}}^\star\|_{D_C}^2 + \frac{1}{2}\|i_{\mathcal{L}}^s - i_{\mathcal{L}}^\star\|_{D_{\mathcal{L}}}^2, \quad s \in \{1, 2\} \\
& (v^1, i^1, x^1, y^1) \text{ is feasible initial point} \\
& (v^2, i^2, x^2, y^2) \text{ is generated by discrete optimization method from initial point} \\
& f \in \mathcal{F},
\end{aligned} \tag{9}$$

where $f, v^1, i^1, x^1, y^1, v^\star, i^\star, x^\star, y^\star$ are the decision variables and $\mathcal{F}$ is a family of functions (*e.g.*, L-smooth convex) that the algorithm is to be applied to. Here, we are using the fact that (8) is homogeneous with respect to k (*i.e.*, (8) essentially has no k-dependence), and therefore it is sufficient to verify the condition for $k = 1$ but for all feasible initial points (v^1, i^1, x^1, y^1). It turns out that (9) can be solved exactly as a semidefinite program (SDP) for many commonly considered function classes $\mathcal{F}$. This technique was initially proposed as the performance estimation problem (PEP) [54, 150], a computer-aided methodology for constructing convergence proofs of first-order optimization methods. See, *e.g.*, PEPit [76] package that implements PEP in Python.

Further, (9) can be posed as a nonconvex quadratically constrained quadratic problem (QCQP) with only a few tens of variables and such problems can be solved exactly with spatial branch-and-bound algorithms [2, 102, 80, 98, 45].

In conclusion, we can solve a non-convex QCQP to find a provably convergent discretization of the continuous-time circuit with an admissible dynamic interconnect. We use the Ipopt [155, 9] solver. Further details are provided in §G of the appendix.

8

Example. Consider the following example circuit for the minimization of a convex function f. Let $R_1 = R_2 = R_3 = 1$, and $C_1 = C_2 = 10$.

With our automatic discretization methodology, we find the sufficiently dissipative parameters

$$\eta = 6.66, \qquad h = 6.66, \quad \alpha = 0, \quad \beta = 1.$$

The resulting provably convergent algorithm is

$$
\begin{aligned}
x^k &= \mathbf{prox}_{(1/2)f}(z^k), \quad y^k = 2(z^k - x^k) \\
w^{k+1} &= w^k - 0.33(y^k + 3w^k) \\
z^{k+1} &= z^k - 0.16(5y^k + 3w^k).
\end{aligned}
$$

is provably convergent[2] under the condition that f is strictly convex, see §H for details.

5 Experiments

In this section, we use our methodology to obtain a new algorithm and experiment with it on a specific problem instance. Consider a decentralized optimization problem with a communication graph G with $N = 6$ nodes and 7 edges, as shown in Figure 8. Specifically, we consider the optimization problem

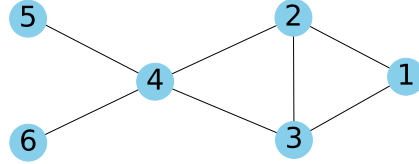



Figure 8: Underlying graph G.

$$\underset{x \in \mathbf{R}^{100}}{\text{minimize}} \quad \sum_{i \in \{4,5\}} \left(\|x - b_i\|_2 + \|x - b_i\|_2^2 \right) + \sum_{i \notin \{4,5\}} \|x - b_i\|_2,$$

where each agent $i \in \{1, \ldots, 6\}$ holds the vector $b_i \in \mathbf{R}^{100}$. To leverage the strong convexity of f_4 and f_5, we propose a modification to the DADMM circuit described in §F.3. Given that a circuit with a capacitor and inductor corresponds to a momentum method (see §3.1), and momentum is known to accelerate convergence for strongly convex functions [124], we add a capacitor to e_{45} to DADMM circuit as shown in the left column of Figure 9. We then discretize the circuit and refer the the resulting algorithm DADMM+C. We apply DADMM+C to the decentralized optimization problem and observe a speedup as shown in the right columns of Figure 9. The relative error for DADMM+C decreases to 10^{-10} in 66 iterations, for DADMM in 87 iterations and for P-EXTRA in 294 iterations. For further details, see §I.1 of the appendix.

---

[2]Our pipeline has a final verification stage that numerically checks whether point returned by the Ipopt solver is indeed feasible for the small QCQP. Strictly speaking, our theoretical convergence guarantee relies on the correctness of this numerical verification of feasibility.

9

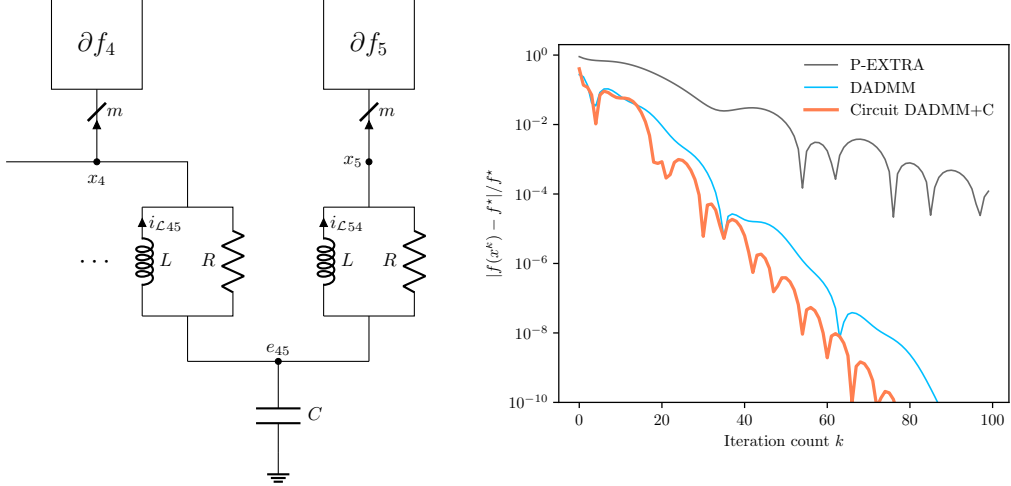



Figure 9: (Left) Circuit of DADMM+C. Compared to the DADMM circuit of §F.3, the DADMM+C circuit has an additional capacitor. (Right) Relative error $\left| f(x^k) - f^\star \right| / f^\star$ vs. iteration count.

Further, we define a general version of the DADMM+C method for any connected graph and establish a general convergence proof in Lemma I.1 of in §I.1.1 of the appendix. This convergence analysis demonstrates how to use our methodology to discover a new family of methods with a classical convergence proof. Finally, we provide another set of similar experiments in §I.2 of the appendix.

6 Conclusion

In this work, we present a novel approach to optimization algorithm design using ideas from electric RLC circuits. The continuous-time RLC circuit models combined with the automatic discretization method provide a foundation for designing algorithms that inherently possess convergence guarantees. Further, we provide code implementing the automatic discretization. Our framework opens the door to future research by applying this methodology to a broader range of optimization problems and extending the problem to other setups, such as the stochastic optimization setup.

Acknowledgments and Disclosure of Funding

This work was supported by the Samsung Science and Technology Foundation (Project Number SSTF-BA2101-02), the National Research Foundation of Korea (NRF) grant funded by the Korean government (No.RS-2024-00421203, RS-2024-00406127), and the Oliger Memorial Fellowship. We thank Hangjun Cho for the helpful discussions on the continuous-time analysis. We also thank anonymous reviewers for the highly constructive feedback.

References

[1] B. Abbas and H. Attouch. Dynamical systems and forward–backward algorithms associated with the sum of a convex subdifferential and a monotone cocoercive operator. *Optimization. A Journal of Mathematical Programming and Operations Research*, 64(10):2223–2252, 2015.

[2] T. Achterberg and E. Towle. Non-Convex Quadratic Optimization: Gurobi 9.0. 2020. `https://www.gurobi.com/resource/non-convex-quadratic-optimization/`.

[3] S. Adly and H. Attouch. Finite convergence of proximal-gradient inertial algorithms combining dry friction with Hessian-driven damping. *SIAM Journal on Optimization*, 30(3):2134–2162, 2020.

[4] A. Agarwal, C. Fiscko, S. Kar, L. Pileggi, and B. Sinopoli. An equivalent circuit workflow for unconstrained optimization. *arXiv preprint arXiv:2305.14061*, 2023.

10

[5] A. Agarwal and L. Pileggi. An equivalent circuit approach to distributed optimization. *arXiv preprint arXiv:2305.14607*, 2023.

[6] F. Alvarez. On the minimizing property of a second order dissipative system in hilbert spaces. *SIAM Journal on Control and Optimization*, 38(4):1102–1119, 2000.

[7] F. Alvarez and H. Attouch. An inertial proximal method for maximal monotone operators via discretization of a nonlinear oscillator with damping. *Set-Valued Analysis*, 9(1):3–11, 2001.

[8] F. Alvarez, H. Attouch, J. Bolte, and P. Redont. A second-order gradient-like dissipative dynamical system with Hessian-driven damping : Application to optimization and mechanics. *Journal de Mathématiques Pures et Appliquées*, 81(8):747–779, 2002.

[9] J. Andersson, J. Gillis, G. Horn, J. B. Rawlings, and M. Diehl. CasADi: a software framework for nonlinear optimization and optimal control. *Mathematical Programming Computation*, 11:1–36, 2019.

[10] V. Apidopoulos, J.-F. Aujol, and C. Dossal. The differential inclusion modeling FISTA algorithm and optimality of convergence rate in the case $b \leq 3$. *SIAM Journal on Optimization*, 28(1):551–574, 2018.

[11] H. Attouch and F. Alvarez. The heavy ball with friction dynamical system for convex constrained minimization problems. *Belgian-French-German Conference on Optimization*, 1998.

[12] H. Attouch, Z. Chbani, J. Fadili, and H. Riahi. First-order optimization algorithms via inertial systems with Hessian driven damping. *Mathematical Programming*, 2020.

[13] H. Attouch, Z. Chbani, J. Peypouquet, and P. Redont. Fast convergence of inertial dynamics and algorithms with asymptotic vanishing viscosity. *Mathematical Programming*, 168(1):123–175, 2018.

[14] H. Attouch, Z. Chbani, and H. Riahi. Fast proximal methods via time scaling of damped inertial dynamics. *SIAM Journal on Optimization*, 29(3):2227–2256, 2019.

[15] H. Attouch, Z. Chbani, and H. Riahi. Rate of convergence of the Nesterov accelerated gradient method in the subcritical case $\alpha \leq 3$. *ESAIM: Control, Optimisation and Calculus of Variations*, 25:2, 2019.

[16] H. Attouch and M.-O. Czarnecki. Asymptotic control and stabilization of nonlinear oscillators with non-isolated equilibria. *Journal of Differential Equations*, 179(1):278–310, 2002.

[17] H. Attouch, X. Goudou, and P. Redont. The heavy ball with friction method, I. The continuous dynamical system: Global exploration of the local minima of a real-valued function by asymptotic analysis of a dissipative dynamical system. *Communications in Contemporary Mathematics*, 02(01):1–34, 2000.

[18] H. Attouch and S. C. László. Newton-like inertial dynamics and proximal algorithms governed by maximally monotone operators. *SIAM Journal on Optimization*, 30(4):3252–3283, 2020.

[19] H. Attouch and J. Peypouquet. Convergence of inertial dynamics and proximal algorithms governed by maximally monotone operators. *Mathematical Programming*, 174(1):391–432, 2019.

[20] H. Attouch, J. Peypouquet, and P. Redont. A dynamical approach to an inertial forward-backward algorithm for convex minimization. *SIAM Journal on Optimization*, 24(1):232–256, 2014.

[21] J.-P. Aubin and A. Cellina. *Differential Inclusions: Set-Valued Maps and Viability Theory*, volume 264. Springer Science & Business Media, 2012.

[22] M. Barré, A. B. Taylor, and F. Bach. Principled analyses and design of first-order methods with inexact proximal operators. *Mathematical Programming*, 201(1):185–230, 2023.

[23] H. Bauschke and P. Combettes. *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*. Springer International Publishing, second edition, 2017.

11

[24] M. Betancourt, M. I. Jordan, and A. C. Wilson. On symplectic optimization. *arXiv:1802.03653*, 2018.

[25] R. I. Boţ and E. R. Csetnek. Convergence rates for forward–backward dynamical systems associated with strongly monotone inclusions. *Journal of Mathematical Analysis and Applications*, 457(2):1135–1152, 2018.

[26] R. I. Boţ, E. R. Csetnek, and D.-K. Nguyen. Fast Optimistic Gradient Descent Ascent (OGDA) method in continuous and discrete time. *Foundations of Computational Mathematics*, 2023.

[27] R. I. Boţ and D. A. Hulett. Second order splitting dynamics with vanishing damping for additively structured monotone inclusions. *Journal of Dynamics and Differential Equations*, 2022.

[28] S. Boyd. Distributed optimization: Analysis and synthesis via circuits. Lecture Note EE364b, Stanford University, 2010.

[29] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends in Machine Learning*, 3(1):1–122, 2011.

[30] S. Boyd, N. Parikh, E. Chu, B. Peleato, J. Eckstein, et al. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends® in Machine learning*, 3(1):1–122, 2011.

[31] S. P. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004.

[32] R. E. Bruck. Asymptotic convergence of nonlinear contraction semigroups in Hilbert space. *Journal of Functional Analysis*, 18(1):15–26, 1975.

[33] F. S. Cattivelli, C. G. Lopes, and A. H. Sayed. A diffusion rls scheme for distributed estimation over adaptive networks. In *Signal Processing Advances in Wireless Communications*, 2007.

[34] F. S. Cattivelli and A. H. Sayed. Diffusion LMS strategies for distributed estimation. *IEEE Transactions on Signal Processing*, 58(3):1035–1048, 2010.

[35] A.-L. Cauchy. Méthode générale pour la résolution des systémes d'équations simultanées. *Comptes Rendus Hebdomadaires des Séances de l'Académie des Sciences*, 25:536–538, 1847.

[36] T. Chaffey, S. Banert, P. Giselsson, and R. Pates. Circuit analysis using monotone+skew splitting. *European Journal of Control*, 74:100854, 2023.

[37] T. Chaffey and A. Padoan. Circuit model reduction with scaled relative graphs. *Conference on Decision and Control (CDC)*, pages 6530–6535, 2022.

[38] T. Chaffey and R. Sepulchre. Monotone RLC circuits. *European Control Conference*, 2021.

[39] T. Chaffey and R. Sepulchre. Monotone one-port circuits. *IEEE Transactions on Automatic Control*, 69(2):783–796, 2024.

[40] S. Chen, B. Shi, and Y.-X. Yuan. On underdamped Nesterov's acceleration. *arXiv:2304.14642*, 2023.

[41] L. Chua and G.-N. Lin. Nonlinear programming without computation. *IEEE Transactions on Circuits and Systems*, 31(2):182–188, 1984.

[42] P. L. Combettes and V. R. Wajs. Signal recovery by proximal forward-backward splitting. *Multiscale Modeling and Simulation*, 4(4):1168–1200, 2005.

[43] E. R. Csetnek, Y. Malitsky, and M. K. Tam. Shadow Douglas–Rachford splitting for monotone inclusions. *Applied Mathematics & Optimization*, 80(3):665–678, 2019.

[44] S. Cyrus, B. Hu, B. Van Scoy, and L. Lessard. A robust accelerated optimization algorithm for strongly convex functions. *American Control Conference (ACC)*, pages 1376–1381, 2018.

12

[45] S. Das Gupta, B. P. G. Van Parys, and E. K. Ryu. Branch-and-bound performance estimation programming: A unified methodology for constructing optimal optimization methods. *Mathematical Programming*, 2023.

[46] D. Davis and W. Yin. A three-operator splitting scheme and its optimization applications. *Set-Valued and Variational Analysis*, 25(4):829–858, 2017.

[47] G. B. De Luca and E. Silverstein. Born-infeld (BI) for AI: Energy-conserving descent (ECD) for optimization. *International Conference on Machine Learning*, 162, 2022.

[48] J. B. Dennis. *Mathematical Programming and Electrical Networks*. PhD thesis, Massachusetts Institute of Technology, 1959.

[49] C. A. Desoer and J. Katzenelson. Nonlinear RLC networks. *Bell System Technical Journal*, 44(1):161–198, 1965.

[50] C. A. Desoer and E. S. Kuh. *Basic Circuit Theory*. Electronic Engineering. McGraw-Hill, 1969.

[51] C. A. Desoer and F. F. Wu. Nonlinear monotone networks. *SIAM Journal on Applied Mathematics*, 26(2):315–333, 1974.

[52] J. Diakonikolas and M. I. Jordan. Generalized momentum-based methods: A Hamiltonian perspective. *SIAM Journal on Optimization*, 31(1):915–944, 2021.

[53] J. Douglas and H. H. Rachford. On the numerical solution of heat conduction problems in two and three space variables. *Transactions of the American Mathematical Society*, 82(2):421–439, 1956.

[54] Y. Drori and M. Teboulle. Performance of first-order methods for smooth convex minimization: A novel approach. *Mathematical Programming*, 145(1):451–482, 2014.

[55] R. J. Duffin. Nonlinear networks. I. *Bulletin of the American Mathematical Society*, 52(10):833–838, 1946.

[56] J. Eckstein and D. P. Bertsekas. On the Douglas—Rachford splitting method and the proximal point algorithm for maximal monotone operators. *Mathematical Programming*, 55(1):293–318, 1992.

[57] L. Euler. *Institutiones Calculi Differentialis*. Petropolis, 1755.

[58] M. Even, R. Berthier, F. Bach, N. Flammarion, H. Hendrikx, P. Gaillard, L. Massoulié, and A. Taylor. Continuized accelerations of deterministic and stochastic gradient descents, and of gossip algorithms. *Neural Information Processing Systems*, 2021.

[59] H. Everett. Generalized Lagrange multiplier method for solving problems of optimum allocation of resources. *Operations Research*, 11(3):399–417, 1963.

[60] M. Fazlyab, A. Ribeiro, M. Morari, and V. M. Preciado. Analysis of optimization algorithms via integral quadratic constraints: Nonstrongly convex problems. *SIAM Journal on Optimization*, 28(3):2654–2689, 2018.

[61] K. Feng. On difference schemes and symplectic geometry. *Proceedings of the 5th International Symposium on Differential Geometry and Differential Equations*, pages 42–58, 1984.

[62] S. Fiori. Quasi-geodesic neural learning algorithms over the orthogonal group: A tutorial. *Journal of Machine Learning Research*, 6(26):743–781, 2005.

[63] M. L. Fisher. The Lagrangian relaxation method for solving integer programming problems. *Management Science*, 50(12_supplement):1861–1871, 2004.

[64] G. França, M. I. Jordan, and R. Vidal. On dissipative symplectic integration with applications to gradient-based optimization. *Journal of Statistical Mechanics: Theory and Experiment*, 2021(4):043402, 2021.

13

[65] G. França, D. Robinson, and R. Vidal. ADMM and accelerated ADMM as continuous dynamical systems. *International Conference on Machine Learning*, 2018.

[66] G. França, D. P. Robinson, and R. Vidal. Gradient flows and proximal splitting methods: A unified view on accelerated and stochastic optimization. *Physical Review E: Statistical Physics, Plasmas, Fluids, and Related Interdisciplinary Topics*, 103(5):053304, 2021.

[67] G. França, D. P. Robinson, and R. Vidal. A nonsmooth dynamical systems perspective on accelerated extensions of ADMM. *IEEE Transactions on Automatic Control*, 68(5):2966–2978, 2023.

[68] G. França, J. Sulam, D. Robinson, and R. Vidal. Conformal symplectic and relativistic optimization. *Neural Information Processing Systems*, 2020.

[69] G. França, J. Sulam, D. P. Robinson, and R. Vidal. Conformal symplectic and relativistic optimization. *Journal of Statistical Mechanics: Theory and Experiment*, 2020(12):124008, 2020.

[70] D. Gabay. Chapter IX applications of the method of multipliers to variational inequalities. In M. Fortin and R. Glowinski, editors, *Augmented Lagrangian Methods: Applications to the Numerical Solution of Boundary-Value Problems*, volume 15 of *Studies in Mathematics and Its Applications*, pages 299–331. Elsevier, 1983.

[71] D. Gabay and B. Mercier. A dual algorithm for the solution of nonlinear variational problems via finite element approximation. *Computers and Mathematics with Applications*, 2(1):17–40, 1976.

[72] A. M. Geoffrion. Primal resource-directive approaches for optimizing nonlinear decomposable systems. *Operations Research*, 18(3):375–403, 1970.

[73] B. Gharesifard and J. Cortés. Distributed continuous-time convex optimization on weight-balanced digraphs. *IEEE Transactions on Automatic Control*, 59(3):781–786, 2014.

[74] R. Glowinski and A. Marroco. Sur l'approximation, par éléments finis d'ordre un, et la résolution, par pénalisation-dualité d'une classe de problèmes de Dirichlet non linéaires. *Revue Française d'Automatique, Informatique, Recherche Opérationnelle. Analyse Numérique*, 9(2):41–76, 1975.

[75] E. Gorbunov, N. Loizou, and G. Gidel. Extragradient method: $O(1/K)$ last-iterate convergence for monotone variational inequalities and connections with cocoercivity. *International Conference on Artificial Intelligence and Statistics*, 2022.

[76] B. Goujaud, C. Moucer, F. Glineur, J. M. Hendrickx, A. B. Taylor, and A. Dieuleveut. PEPit: Computer-assisted worst-case analyses of first-order optimization methods in Python. *Mathematical Programming Computation*, 16(3):337–367, 2024.

[77] E. Hairer, C. Lubich, and W. Gerhard. *Geometric Numerical Integration: Structure-Preserving Algorithms for Ordinary Differential Equations*. Springer, 2 edition, 2006.

[78] S. Hassan-Moghaddam and M. R. Jovanović. Proximal gradient flow and Douglas–Rachford splitting dynamics: Global exponential stability via integral quadratic constraints. *Automatica*, 123:109311, 2021.

[79] U. Helmke and J. Moore. Optimization and dynamical systems. *Proceedings of the IEEE*, 84(6):907–, 1996.

[80] R. Horst and H. Tuy. *Global Optimization: Deterministic Approaches*. Springer Science & Business Media, 2013.

[81] B. Hu and L. Lessard. Dissipativity theory for Nesterov's accelerated method. *International Conference on Machine Learning*, 2017.

[82] T. H. Hughes. Passivity and electric circuits: A behavioral approach. *IFAC-PapersOnLine*, 50(1):15500–15505, 2017.

14

[83] A. Iserles. *A First Course in the Numerical Analysis of Differential Equations*. Cambridge University Press, 2009.

[84] U. Jang, S. D. Gupta, and E. K. Ryu. Computer-assisted design of accelerated composite optimization methods: OptISTA. *arXiv:2305.15704*, 2023.

[85] H. K. Khalil. *Nonlinear Systems*. Pearson Education. Prentice Hall, 2002.

[86] S. S. Kia, J. Cortés, and S. Martínez. Distributed convex optimization via continuous-time coordination algorithms with discrete-time communication. *Automatica*, 55:254–264, 2015.

[87] D. Kim. Accelerated proximal point method for maximally monotone operators. *Mathematical Programming*, 190(1–2):57–87, 2021.

[88] D. Kim and J. A. Fessler. Optimized first-order methods for smooth convex minimization. *Mathematical Programming*, 159(1-2):81–107, 2016.

[89] D. Kim and J. A. Fessler. Optimizing the efficiency of first-order methods for decreasing the gradient of smooth convex functions. *Journal of Optimization Theory and Applications*, 188(1):192–219, 2021.

[90] J. Kim and I. Yang. Convergence analysis of ODE models for accelerated first-order methods via positive semidefinite kernels. *Neural Information Processing Systems*, 2023.

[91] J. Kim and I. Yang. Unifying Nesterov's accelerated gradient methods for convex and strongly convex objective functions. *International Conference on Machine Learning*, 2023.

[92] D. P. Kingma and J. Ba. Adam: A method for stochastic optimization. *International Conference on Learning Representations*, 2015.

[93] W. Krichene, A. Bayen, and P. L. Bartlett. Accelerated mirror descent in continuous and discrete time. *Neural Information Processing Systems*, 2015.

[94] W. Kutta. *Beitrag Zur Näherungsweisen Integration Totaler Differentialgleichungen*. Teubner, 1901.

[95] S. Lee and D. Kim. Fast extra gradient methods for smooth structured nonconvex-nonconcave minimax problems. *Neural Information Processing Systems*, 2021.

[96] C. Lemaréchal. Lagrangian relaxation. In M. Jünger and D. Naddef, editors, *Computational Combinatorial Optimization: Optimal or Provably Near-Optimal Solutions*, Lecture Notes in Computer Science, pages 112–156. Springer, Berlin, Heidelberg, 2001.

[97] L. Lessard, B. Recht, and A. Packard. Analysis and design of optimization algorithms via integral quadratic constraints. *SIAM Journal on Optimization*, 26(1):57–95, 2016.

[98] L. Liberti. Introduction to global optimization. *Ecole Polytechnique*, 2008.

[99] F. Lieder. On the convergence rate of the Halpern-iteration. *Optimization Letters*, 15(2):405–418, 2021.

[100] P. Lin, W. Ren, and J. A. Farrell. Distributed continuous-time optimization: Nonuniform gradient gains, finite-time convergence, and convex constraint set. *IEEE Transactions on Automatic Control*, 62(5):2239–2253, 2017.

[101] P. L. Lions and B. Mercier. Splitting algorithms for the sum of two nonlinear operators. *SIAM Journal on Numerical Analysis*, 16(6):964–979, 1979.

[102] M. Locatelli and F. Schoen. *Global Optimization: Theory, Algorithms, and Applications*. SIAM, 2013.

[103] H. Lu. An $O(s^r)$-resolution ODE framework for understanding discrete-time algorithms and applications to the linear convergence of minimax problems. *Mathematical Programming*, 194(1):1061–1112, 2022.

15

[104] J. Lu and C. Y. Tang. Zero-gradient-sum algorithms for distributed convex optimization: The continuous-time case. *IEEE Transactions on Automatic Control*, 57(9):2348–2354, 2012.

[105] C. J. Maddison, D. Paulin, Y. W. Teh, B. O'Donoghue, and A. Doucet. Hamiltonian descent methods. *arXiv:1809.05042*, 2018.

[106] J. C. Maxwell. *A Treatise on Electricity and Magnetism*, volume 1. Oxford: Clarendon Press, 1873.

[107] R. McLachlan and M. Perlmutter. Conformal Hamiltonian systems. *Journal of Geometry and Physics*, 39(4):276–300, 2001.

[108] A. Megretski and A. Rantzer. System analysis via integral quadratic constraints. *IEEE Transactions on Automatic Control*, 42(6):819–830, 1997.

[109] W. Millar. CXVI. Some general theorems for non-linear systems possessing resistance. *Philosophical Magazine and Journal of Science*, 42(333):1150–1160, 1951.

[110] G. Minty. Solving steady-state nonlinear networks of 'monotone' elements. *IRE Transactions on Circuit Theory*, 8(2):99–104, 1961.

[111] G. J. Minty. Monotone networks. *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences*, 257(1289):194–212, 1960.

[112] G. J. Minty. On the maximal domain of a "monotone" function. *The Michigan Mathematical Journal*, 8(2):135–137, 1961.

[113] C. Moucer, A. Taylor, and F. Bach. A systematic approach to Lyapunov analyses of continuous-time models in convex optimization. *SIAM Journal on Optimization*, 33(3):1558–1586, 2023.

[114] M. Muehlebach and M. Jordan. A dynamical systems perspective on Nesterov acceleration. *International Conference on Machine Learning*, 2019-06-09/2019-06-15.

[115] M. Muehlebach and M. I. Jordan. Optimization with momentum: Dynamical, control-theoretic, and symplectic perspectives. *Journal of Machine Learning Research*, 22(73):1–50, 2021.

[116] A. Nedic and A. Ozdaglar. Distributed Subgradient Methods for Multi-Agent Optimization. *IEEE Transactions on Automatic Control*, 54(1):48–61, 2009.

[117] Y. Nesterov. A method of solving a convex programming problem with convergence rate $O(1/k^2)$. *Doklady Akademii Nauk SSSR*, 269(3):543–547, 1983.

[118] Y. Nesterov. *Introductory Lectures on Convex Optimization: A Basic Course*. Springer, 2004.

[119] Y. Nesterov. *Lectures on Convex Optimization*. Springer, 2 edition, 2018.

[120] R. Nishihara, L. Lessard, B. Recht, A. Packard, and M. Jordan. A general analysis of the convergence of ADMM. *International Conference on Machine Learning*, 2015.

[121] J. Nocedal and S. J. Wright. *Numerical Optimization*. Springer, 1999.

[122] J. Park and E. K. Ryu. Exact optimal accelerated complexity for fixed-point iterations. *International Conference on Machine Learning*, 2022.

[123] D. W. Peaceman and J. Rachford, H. H. The numerical solution of parabolic and elliptic differential equations. *Journal of the Society for Industrial and Applied Mathematics*, 3(1):28–41, 1955.

[124] B. T. Polyak. Some methods of speeding up the convergence of iteration methods. *USSR Computational Mathematics and Mathematical Physics*, 4(5):1–17, 1964.

[125] R. T. Rockafellar. Monotone operators and the proximal point algorithm. *SIAM Journal on Control and Optimization*, 14(5):877–898, 1976.

[126] T. Rockafellar. *Convex analysis*, volume 11. Princeton University Press, 1997.

16

[127] C. Runge. Ueber die numerische Auflösung von Differentialgleichungen. *Mathematische Annalen*, 46(2):167–178, 1895.

[128] R. D. Ruth. A canonical integration technique. *IEEE Transactions on Nuclear Science*, 30(4):2669–2671, 1983.

[129] E. Ryu and W. Yin. *Large-Scale Convex Optimization via Monotone Operators*. Cambridge University Press, 2022.

[130] E. K. Ryu, A. B. Taylor, C. Bergeling, and P. Giselsson. Operator splitting performance estimation: Tight contraction factors and optimal parameter selection. *SIAM Journal on Optimization*, 30(3):2251–2271, 2020.

[131] S. K. Sashank J. Reddi, Satyen Kale. On the convergence of Adam and beyond. *International Conference on Learning Representations*, 2018.

[132] K. Sawant, D. Nguyen, A. Liu, J. Poon, and S. Dhople. A hybrid-computing solution to nonlinear optimization problems. *IEEE Transactions on Circuits and Systems I: Regular Papers*, 71(12):6555–6568, 2024.

[133] J. Schropp and I. Singer. A dynamical systems approach to constrained minimization. *Numerical Functional Analysis and Optimization*, 21(3-4):537–551, 2000.

[134] D. Scieur, V. Roulet, F. Bach, and A. d'Aspremont. Integration methods and optimization algorithms. *Neural Information Processing Systems*, 2017.

[135] S. Seshu and M. B. Reed. *Linear Graphs and Electrical Networks*. Addison-Wesley Series in Behavioral Science: Quantitative Methods. Addison-Wesley Publishing Company, 1961.

[136] B. Shi, S. Du, W. Su, and M. Jordan. Acceleration via symplectic discretization of high-resolution differential equations. *Neural Information Processing Systems*, 2019.

[137] B. Shi, S. S. Du, M. I. Jordan, and W. J. Su. Understanding the acceleration phenomenon via high-resolution differential equations. *Mathematical Programming*, 2021.

[138] W. Shi, Q. Ling, G. Wu, and W. Yin. A proximal gradient algorithm for decentralized composite optimization. *IEEE Transactions on Signal Processing*, 63(22):6013–6023, 2015.

[139] W. Shi, Q. Ling, K. Yuan, G. Wu, and W. Yin. On the linear convergence of the ADMM in decentralized consensus optimization. *IEEE Transactions on Signal Processing*, 62(7):1750–1761, 2014.

[140] G. J. Silverman. Primal decomposition of mathematical programs by resource allocation: I. Basic theory and a direction-finding procedure. *Operations Research*, 20(1):58–74, 1972.

[141] C. S. Simon Michalowsky and C. Ebenbauer. Robust and structure exploiting optimisation algorithms: An integral quadratic constraint approach. *International Journal of Control*, 94(11):2956–2979, 2021.

[142] D. Sontag, A. Globerson, and T. Jaakkola. Introduction to dual decomposition for inference. In S. Sra, S. Nowozin, and S. J. Wright, editors, *Optimization for Machine Learning*, pages 219–254. The Massachusetts Institute of Technology Press, 2011.

[143] W. Su, S. Boyd, and E. J. Candès. A differential equation for modeling Nesterov's accelerated gradient method: Theory and insights. *Neural Information Processing Systems*, 2014.

[144] W. Su, S. Boyd, and E. J. Candès. A differential equation for modeling Nesterov's accelerated gradient method: Theory and insights. *Journal of Machine Learning Research*, 17(153):1–43, 2016.

[145] J. J. Suh, J. Park, and E. K. Ryu. Continuous-time analysis of anchor acceleration. *Neural Information Processing Systems*, 2023.

[146] J. J. Suh, G. Roh, and E. K. Ryu. Continuous-time analysis of AGM via conservation laws in dilated coordinate systems. *International Conference on Machine Learning*, 2022.

17

[147] A. Sundararajan, B. Van Scoy, and L. Lessard. Analysis and design of first-order distributed optimization algorithms over time-varying graphs. *IEEE Transactions on Control of Network Systems*, 7(4):1597–1608, 2020.

[148] A. Taylor and F. Bach. Stochastic first-order methods: Non-asymptotic and computer-aided analyses via potential functions. *Conference on Learning Theory*, 2019-06-25/2019-06-28.

[149] A. Taylor and Y. Drori. An optimal gradient method for smooth strongly convex minimization. *Mathematical Programming*, 199(1-2):557–594, 2023.

[150] A. B. Taylor, J. M. Hendrickx, and F. Glineur. Smooth strongly convex interpolation and exact worst-case performance of first-order methods. *Mathematical Programming*, 161(1):307–345, 2017.

[151] K. Ushiyama, S. Sato, and T. Matsuo. A unified discretization framework for differential equation approach with Lyapunov arguments for convex optimization. *Neural Information Processing Systems*, 2023.

[152] B. Van Scoy, R. A. Freeman, and K. M. Lynch. The fastest known globally convergent first-order method for minimizing strongly convex functions. *IEEE Control Systems Letters*, 2(1):49–54, 2018.

[153] S. Vichik and F. Borrelli. Solving linear and quadratic programs with an analog circuit. *Computers & Chemical Engineering*, 70:160–171, 2014.

[154] R. D. Vogelaere. Methods of integration which preserve the contact transformation property of the hamilton equations. 1956.

[155] A. Wächter and L. T. Biegler. On the implementation of an interior-point filter line-search algorithm for large-scale nonlinear programming. *Mathematical Programming*, 106:25–57, 2006.

[156] J. Wang and N. Elia. Control approach to distributed optimization. *Annual Allerton Conference on Communication, Control, and Computing*, 2010.

[157] E. Wei and A. Ozdaglar. On the $O(1/k)$ convergence of asynchronous distributed alternating direction method of multipliers. *Global Conference on Signal and Information Processing*, 2013.

[158] A. Wibisono, A. C. Wilson, and M. I. Jordan. A variational perspective on accelerated methods in optimization. *Proceedings of the National Academy of Sciences*, 113(47):E7351–E7358, 2016.

[159] J. C. Willems. The Generation of Lyapunov Functions for Input-Output Stable Systems. *SIAM Journal on Control*, 9(1):105–134, 1971.

[160] J. C. Willems. Dissipative dynamical systems part I: General theory. *Archive for Rational Mechanics and Analysis*, 45(5):321–351, 1972.

[161] A. C. Wilson, L. Mackey, and A. Wibisono. Accelerating rescaled gradient descent: Fast optimization of smooth functions. *Neural Information Processing Systems*, 2019.

[162] A. C. Wilson, B. Recht, and M. I. Jordan. A Lyapunov analysis of accelerated methods in optimization. *Journal of Machine Learning Research*, 22(113):1–34, 2021.

[163] G. Wilson. Quadratic programming analogs. *IEEE Transactions on Circuits and Systems*, 33(9):907–911, 1986.

[164] L. Yang, R. Arora, V. braverman, and T. Zhao. The physical systems behind optimization algorithms. *Neural Information Processing Systems*, 31, 2018.

[165] T. Yang, X. Yi, J. Wu, Y. Yuan, D. Wu, Z. Meng, Y. Hong, H. Wang, Z. Lin, and K. H. Johansson. A survey of distributed optimization. *Annual Reviews in Control*, 47:278–305, 2019.

18

[166] T. Yoon, J. Kim, J. J. Suh, and E. K. Ryu. Optimal acceleration for minimax and fixed-point problems is not unique. *International Conference on Machine Learning*, 2024.

[167] T. Yoon and E. K. Ryu. Accelerated algorithms for smooth convex-concave minimax problems with $\mathcal{O}(1/k^2)$ rate on squared gradient norm. *International Conference on Machine Learning*, 2021.

[168] Y. Yu and B. Açıkmeşe. RC circuits based distributed conditional gradient method. *arXiv:2003.06949*, 2020.

[169] Y. Yu and B. Açıkmeşe. RLC circuits-based distributed mirror descent method. *IEEE Control Systems Letters*, 4(3):548–553, 2020.

[170] H. Yuan, Y. Zhou, C. J. Li, and Q. Sun. Differential inclusions for modeling nonsmooth ADMM variants: A continuous limit theory. *International Conference on Machine Learning*, 2019.

[171] K. Yuan, Q. Ling, and W. Yin. On the convergence of decentralized gradient descent. *SIAM Journal on Optimization*, 26(3):1835–1854, 2016.

[172] J. Zhang, A. Mokhtari, S. Sra, and A. Jadbabaie. Direct Runge–Kutta discretization achieves acceleration. *Neural Information Processing Systems*, 2018.

[173] J. Zhang, S. Sra, and A. Jadbabaie. Acceleration in first order quasi-strongly convex optimization by ODE discretization. *Conference on Decision and Control*, 2019.

[174] P. Zhang, A. Orvieto, and H. Daneshmand. Rethinking the variational interpretation of accelerated optimization methods. *Neural Information Processing Systems*, 2021.

19

A Prior works

Distributed optimization as RLC circuits. This work started as a lecture for the Stanford University EE 364b class given in 2010 [28]. The lecture proposed the idea of relating distributed optimization algorithms to the dynamics of RLC circuits. Different from the prior studies [48, 163, 41, 153, 132], that consider solving specific optimization problems through implementing physical circuits, our focus is on using insights from circuit theory to design new algorithms, without any consideration of implementing physical circuits. The follow-up works [168, 169, 4, 5], have built upon this setup [28].

Optimization algorithms from continuous-time dynamics. Relating continuous-time dynamics described by ordinary differential equation (ODE) with optimization algorithm is a technique with a long history [32, 79, 6, 133, 62]. The continuous-time dynamics related to Polyak's heavy ball method [124] were studied by [11, 17, 8, 16]. The ODE model for Nesterov acceleration [117] was introduced by [143, 144], analyses for generalized cases were followed by [10, 13, 15], and the ODE model for Nesterov acceleration for strongly convex function (NAG-SC) was introduced in [162]. Together with [93], the studies by [143, 144] initiated continuous-time analyses of accelerated first-order methods and inspired much follow-up works such as [158, 19, 115, 52, 58, 162, 18, 27, 146, 91, 145, 26]. As a further refined continuous-time model preserving more information from the discretization, the high-resolution ODE for NAG-SC was introduced in [137], and was further developed by [103].

In addition to accelerated methods, various topics and methods in optimization have been studied in a continuous-time framework. Continuous-time dynamics related to splitting methods were studied by [20, 1, 25, 43, 66, 78]. [65] studied continuous-time dynamics of ADMM [74, 71, 70, 56, 29], and provided an accelerated ADMM by discretizing the ODE model combined with [143]. The analyses were furthermore generalized to differential inclusions by [170, 67]. There are numerous works of continuous-time analyses for distributed optimization, [156, 104, 73, 86, 100] to name a few, and we refer the readers to the survey paper [165] for a comprehensive overview.

Computer-assisted analysis of optimization algorithms. There has been lines of work automating the analysis of optimization methods using semidefinite programs (SDP). One line of work is performance estimation problems (PEP) introduced by [54], which provides a systematic way to obtain worst-case performance guarantees of a given fixed-step first-order method. The range and technique of utilizing PEP have been further developed by [150, 148, 130, 113, 90], and many efficient algorithms with tight analyses utilizing PEP are discovered [88, 99, 87, 89, 167, 95, 122, 75, 149, 84, 22, 166].

Another line of work is an approach adapting integral quadratic constraints (IQC) [108]. IQCs are a powerful analysis method in control theory for analyzing interconnected dynamical systems with nonlinear feedback. This approach was first adapted for analyzing first-order optimization algorithms by [97] and followed by [60]. Analyses based on IQC have lead to tight bounds for well-known algorithms [120, 81]. IQC has also been utilized to develop new fast algorithms with tight convergence rates [152, 44, 147, 141].

Recently, an extension of PEP to leveraging quadratic constrained quadratic programs (QCQP) was introduced by [45]. Treating the step-sizes as optimization variables, this work furthermore provides systematic computer-assisted methodology to optimize the step-sizes. Our work adapts this approach to finding appropriate discretizations. To the best of our knowledge, our proposal is the first instance of using computer-assisted methodologies to find discretizations of continuous-time dynamics.

Physics-bases approaches to designing optimization algorithms. Optimization methods obtained by discretizing conformal Hamiltonian dynamics [107] were considered by [105]. Studying structure-preserving discretizations for conformal (dissipative) Hamiltonian systems, [68, 69] analyzed symplectic structure of Nesterov and heavy ball, and introduced Relativistic Gradient Descent (RGD) by adopting ideas from special relativity. Based on relativistic Born-Infeld (BI) dynamics, [47] considered a class of frictionless, energy-conserving system and introduced Bouncing BI (BBI) algorithm as a discretization.

Our work is based on nonlinear resistive electric circuits, the study of which dates back to [55]. The stationary condition for nonlinear networks were considered by [109], generalizing theorems of Maxwell [106] for linear networks. The study of nonlinear resistive networks influenced the

20

refinement of the concept of maximal monotonicity [111, 112, 110], which is now a fundamental concept in convex optimization. Well-posedness of the solutions for nonlinear networks was studied by [49, 51], but only for one-descent nonlinear resistors. Recently, the study of nonlinear electrical circuits was revisited by [38, 37, 36, 39] using contemporary methods of convex optimization. However, their main focus was on circuits, not on designing new optimization algorithms. To the best of our knowledge, our work is the first to introduce a generalized framework for designing optimization methods based on electric circuits.

Discretization.   Continuous-time analyses of optimization algorithms must eventually contend with the issue of discretizing the dynamics into a discrete-time algorithm. Discretization of differential equations is a subject of numerical analysis, and it has a long history, even dating back to Euler [57]. Standard discretization schemes such as Euler, Runge–Kutta [127, 94] and symplectic integrators [154, 128, 61], have a rich body of research analyzing their convergence [77, 83] for example. However, these theories in numerical analysis primarily focus on the convergence of the discretized sequence to the trajectory of the solution flow in differential equations throughout a finite time-interval, which differs from the focus of optimization. Therefore, directly applying standard discretization schemes from numerical analysis does not ensure convergence to the optimality criteria of interest in optimization, such as function value or optimal point convergence.

In optimization, the study of discretization can broadly be divided into two categories. One involves applying standard discretization schemes or their variants, and the other provides special rules tailored to the specific dynamics of interest. As previously discussed, the former cases can only guarantee the convergence involving certain errors [24, 64], or introduce specific and limited cases they can cover [134, 172, 136, 114, 173, 146, 151]. The latter type of works do provide discretization rules with analytic proofs for certain families of ODEs [7, 144, 158, 14, 161, 3, 12, 52, 26], but cannot be applied to general cases. Of course, both approaches have brought significant advances in obtaining new methods from continuous-time dynamics, however, it is still true that previous approaches cannot immediately applied the new ODEs that emerge from our framework. To the best of our knowledge, our work is the first to propose to automate the process of finding a discretized method from ODE using computer-assisted tools.

21

B   Proof of Theorem 2.1

To prove Theorem 2.1, it is sufficient to consider the cases without 0-ohm resistors and furthermore all resistor, inductance, capacitance values are 1. We first state the theorem for such cases, which implies Theorem 2.1.

Theorem B.1. *Let* $f\colon \mathbf{R}^m \to \mathbf{R}^m$ *be a* μ*-strongly convex and* M*-smooth function and* $B\colon \mathbf{R}^{\mathcal{J}} \to \mathbf{R}^{\mathcal{K}}$ *be a matrix. Suppose* (v^0, i^0, x^0, y^0) *satisfy*

$$\begin{bmatrix} i^0 \\ y^0 \end{bmatrix} \in \mathcal{N}(B), \quad \begin{bmatrix} v^0 \\ x^0 \end{bmatrix} \in \mathcal{R}(B^\mathsf{T}), \quad v_\mathcal{R}^0 = i_\mathcal{R}^0, \quad y^0 = \nabla f(x^0). \tag{10}$$

Then there is a uniquely determined Lipschitz continuous curve $(v, i, x, y)\colon [0, \infty) \to \mathbf{R}^{2\mathcal{K}}$ *satisfies*

$$\begin{bmatrix} i \\ y \end{bmatrix} \in \mathcal{N}(B), \quad \begin{bmatrix} v \\ x \end{bmatrix} \in \mathcal{R}(B^\mathsf{T}), \quad y = \nabla f(x), \quad v_\mathcal{R} = i_\mathcal{R}, \quad v_\mathcal{L} = \frac{d}{dt} i_\mathcal{L}, \quad i_\mathcal{C} = \frac{d}{dt} v_\mathcal{C}, \tag{11}$$

for all $t \in (0, \infty)$ *and the initial condition* $(v(0), i(0), x(0), y(0)) = (v^0, i^0, x^0, y^0)$.

Lemma B.2. *Theorem B.1 implies Theorem 2.1.*

Proof. (i) KCL, KVL and V-I relations for equivalent dynamics without 0*-ohm resistors.*
We first consider the equivalent dynamic interconnect without 0-ohm resistors. As 0-ohm resistors are ideal wires, from basic circuit theory we know the nodes connected by 0-ohm resistors can be considered as a single node. We find the expression for KCL, KVL and V-I relations for the equivalent dynamic interconnect composed with ∂f. The equivalent expression for KCL and KVL can be considered as consequence of Tellegen's theorem in [50, §10.2.3], however, we write the detail here to make it self-contained.

Observe, KCL and KVL can be equivalently written as

$$\begin{bmatrix} A & | & I_m \\ & & 0 \end{bmatrix} \begin{bmatrix} i \\ y \end{bmatrix} = 0, \qquad \begin{bmatrix} v \\ x \end{bmatrix} = \begin{bmatrix} A & | & I_m \\ & & 0 \end{bmatrix}^\mathsf{T} \begin{bmatrix} x \\ e \end{bmatrix}.$$

We furthermore restrict the values to satisfy Ohm's law for 0-ohm resistors, *i.e.*, the potential values of two nodes connected to a 0-ohm resistor is identical.

Let's first focus on KCL, the left equation. Suppose node j and l are connected with 0-ohm resistor named as $\mathcal{R}_{jl}$. Suppose the k'th column of A corresponds to $\mathcal{R}_{jl}$. Eliminating $\mathcal{R}_{jl}$ corresponds to eliminating the k'th column of A and eliminating $i_{\mathcal{R}_{jl}}$ from i. However, if we just directly eliminate them, as $i_{\mathcal{R}_{jl}}$ may not be zero, the equation will no longer be satisfied. We need to keep the information that currents flowing into node j (except for $-i_{\mathcal{R}_{jl}}$) flows to node l. As we do not permit ideal wire loop, without loss of generality we may assume node j is not the ground node.

To preserve the information, when node l is not the ground node, we add the j'th row of $\begin{bmatrix} A & | & I_m \\ & & 0 \end{bmatrix}$ to the l'th row. Then k'th component of the l'th row becomes 0, thus the equation corresponding to the l'th row will still be satisfied after eliminating the k'th column and $\mathcal{R}_{jl}$. When node l is the ground node, skip the row addition. Now eliminate the j'th row. Note that column is eliminated only from A.

We now move on to KVL. Eliminating a column of A and a component in i corresponds to eliminating a row of A^T and a component in v. This conserves the validity of the equation. Next, the row operation for $\begin{bmatrix} A & | & I_m \\ & & 0 \end{bmatrix}$ corresponds to column operation for $\begin{bmatrix} A & | & I_m \\ & & 0 \end{bmatrix}^\mathsf{T} = \begin{bmatrix} A^\mathsf{T} \\ \hline I_m\ 0 \end{bmatrix}$. Recall we've restricted the potential values of the nodes connected with 0-ohm resistor to be same, values in $\begin{bmatrix} x \\ e \end{bmatrix}$ corresponding to column j and l coincide. Thus when node l is not the ground node, adding j'th column to the l'th column and eliminating j'th component in $\begin{bmatrix} x \\ e \end{bmatrix}$, will not change the values on the left hand side. When node l is the ground node, the same argument holds by skipping the column addition.

Repeat this process until there is no 0-ohm resistors. Name the reduced matrix as $\tilde{B}$ and reduced current as $\tilde{i}$. Then KCL reduces to $\tilde{B} \begin{bmatrix} \tilde{i} \\ y \end{bmatrix} = 0$ and KVL reduces to $\begin{bmatrix} v \\ x \end{bmatrix} = \tilde{B}^\mathsf{T} \begin{bmatrix} \tilde{x} \\ \tilde{e} \end{bmatrix}$, or equivalently $\begin{bmatrix} \tilde{v} \\ x \end{bmatrix} \in \mathcal{R}(\tilde{B}^\mathsf{T})$.

22

Now name the reduced diagonal matrices $\tilde{D}_\mathcal{R}$, $\tilde{D}_\mathcal{L}$ and $\tilde{D}_\mathcal{C}$ as the reduced matrices that without the entries corresponding to eliminated components. Note $\tilde{D}_\mathcal{R}$ has no zero diagonal entries. Then KCL, KVL and V-I relations for the equivalent dynamic interconnect composed with ∂f become as follows

$$\begin{bmatrix} \tilde{i} \\ y \end{bmatrix} \in \mathcal{N}(\tilde{B}), \quad \begin{bmatrix} \tilde{v} \\ x \end{bmatrix} \in \mathcal{R}(\tilde{B}^\mathsf{T}), \quad y = \nabla f(x), \quad \tilde{v}_\mathcal{R} = \tilde{D}_\mathcal{R} \tilde{i}_\mathcal{R}, \quad \tilde{v}_\mathcal{L} = \tilde{D}_\mathcal{L} \frac{d}{dt} \tilde{i}_\mathcal{L}, \quad \tilde{i}_\mathcal{C} = \tilde{D}_\mathcal{C} \frac{d}{dt} \tilde{v}_\mathcal{C}.$$

(12)

As an equivalent dynamics, it is enough to prove the curve that satisfies (12) and the initial condition $(\tilde{v}(0), \tilde{i}(0), x(0), y(0)) = (\tilde{v}^0, \tilde{i}^0, x^0, y^0)$ with condition

$$\begin{bmatrix} \tilde{i}^0 \\ y^0 \end{bmatrix} \in \mathcal{N}(\tilde{B}), \quad \begin{bmatrix} \tilde{v}^0 \\ x^0 \end{bmatrix} \in \mathcal{R}(\tilde{B}^\mathsf{T}), \quad \tilde{v}_\mathcal{R}^0 = \tilde{D}_\mathcal{R} \tilde{i}_\mathcal{R}^0, \quad y^0 = \nabla f(x^0)$$

(13)

is unique and Lispchitz continuous.

(ii) Sufficient to consider only the cases with $\tilde{D}_\mathcal{R}$, $\tilde{D}_\mathcal{L}$ and $\tilde{D}_\mathcal{C}$ are identity matrices.
For a dynamic interconnect composed with ∂f, consider the equivalent dynamics without 0-ohm resistors. Let $\tilde{B}$ be the matrix in (12) for the dynamics, and let $\mathcal{K}$ be the number of columns of $\tilde{B}$. Suppose $(\tilde{v}^0, \tilde{i}^0, x^0, y^0)$ satisfy (13). Define the diagonal matrix

$$P = \mathbf{diag}\left(\sqrt{\tilde{D}_\mathcal{R}^{-1}}, \sqrt{\tilde{D}_\mathcal{L}^{-1}}, \sqrt{\tilde{D}_\mathcal{C}}, I_m \right),$$

and define $B = \tilde{B}P$. Define i^0 and v^0 to satisfy $\begin{bmatrix} i^0 \\ y^0 \end{bmatrix} = P^{-1} \begin{bmatrix} \tilde{i}^0 \\ y^0 \end{bmatrix}$ and $\begin{bmatrix} v^0 \\ x^0 \end{bmatrix} = P \begin{bmatrix} \tilde{v}^0 \\ x^0 \end{bmatrix}$. Then (v^0, i^0, x^0, y^0) satisfies (10) since

$$\tilde{B} \begin{bmatrix} \tilde{i}^0 \\ y^0 \end{bmatrix} = 0 \quad \Longleftrightarrow \quad B \begin{bmatrix} i^0 \\ y^0 \end{bmatrix} = (\tilde{B}P)\left(P^{-1} \begin{bmatrix} \tilde{i}^0 \\ y^0 \end{bmatrix} \right) = 0,$$

$$\exists z^0, \begin{bmatrix} \tilde{v}^0 \\ x^0 \end{bmatrix} = \tilde{B}^\mathsf{T} z^0 \quad \Longleftrightarrow \quad \exists z^0, \begin{bmatrix} v^0 \\ x^0 \end{bmatrix} = P \begin{bmatrix} \tilde{v}^0 \\ x^0 \end{bmatrix} = P\tilde{B}^\mathsf{T} z = Bz^0,$$

and

$$\tilde{v}_\mathcal{R}^0 = \tilde{D}_\mathcal{R} \tilde{i}_\mathcal{R}^0 \quad \Longleftrightarrow \quad \sqrt{\tilde{D}_\mathcal{R}} \tilde{v}_\mathcal{R}^0 = \sqrt{\tilde{D}_\mathcal{R}^{-1}} \tilde{i}_\mathcal{R}^0 \quad \Longleftrightarrow \quad v_\mathcal{R}^0 = i_\mathcal{R}^0.$$

Then by Theorem B.1, there is a Lipschitz continuous curve $(v, i, x, y)\colon [0, \infty) \to \mathbf{R}^{2\mathcal{K}}$ that satisfies (11) for all $t \in (0, \infty)$ and the initial condition $(v(0), i(0), x(0), y(0)) = (v^0, i^0, x^0, y^0)$. Define $\tilde{i}$ and $\tilde{v}$ to satisfy $\begin{bmatrix} \tilde{i} \\ y \end{bmatrix} = P \begin{bmatrix} i \\ x \end{bmatrix}$ and $\begin{bmatrix} \tilde{v} \\ y \end{bmatrix} = P^{-1} \begin{bmatrix} v \\ y \end{bmatrix}$. Then $\tilde{i}$ and $\tilde{v}$ are Lipschitz continuous as well, as they are composition of linear operators and Lipschitz continuous curves. Furthermore, we can check (12) and the initial condition $(\tilde{v}(0), \tilde{i}(0), x(0), y(0)) = (\tilde{v}^0, \tilde{i}^0, x^0, y^0)$ is satisfied, with the similar argument above.

Reversing the arguments, the uniqueness can be obtained since P is invertible and thus $(v, i) \mapsto (\tilde{v}, \tilde{i})$ is bijective. This concludes the proof.

$\square$

By Lemma B.2, our goal has reduced to Theorem B.1. We will establish the well-posedness for $v_\mathcal{C}$ and $i_\mathcal{L}$ first, then extended them to whole curve. The well-posedness of $v_\mathcal{C}$, $i_\mathcal{L}$ can be obtained by reducing the dynamics to a differential inclusion with a maximal monotone operator. We first restate the theorem in [21] and its immediate implication as a remark, which we use in the proof.

Theorem B.3. *[21, Thm 3.2.1] Let $\mathbf{M}\colon \mathbf{R}^n \rightrightarrows \mathbf{R}^n$ be a maximal monotone operator, consider the differential inclusion*

$$\dot{X}(t) \in -\mathbf{M}(X(t)),$$

(14)

with initial condition $X(0) = X_0 \in \mathbf{dom}\,\mathbf{M}$. Then there is a unique solution $X\colon [0, \infty) \to \mathbf{R}^n$ that is absolutely continuous and satisfies (14) for almost all t. Moreover, if we denote $\mathcal{T} = \{t \in [0, \infty) \mid X \text{ is differentiable at } t\}$, then followings are true.

23

(i) *Let $X(\cdot), Y(\cdot)$ are the solutions issued from $X_0, Y_0 \in \mathbf{dom}\,\mathbb{M}$ respectively. Then $\|X(t) - Y(t)\| \leq \|X_0 - Y_0\|$ for all $t \geq 0$.*

(ii) *For all $t \geq 0$, $\dot{X}_+(t) := \lim_{h \to 0+} \frac{X(t+h) - X(t)}{h}$ is well-defined and continuous from the right. Note, $\dot{X}(t) = \dot{X}_+(t)$ for all $t \in \mathcal{T}$.*

(iii) *$t \mapsto \left\|\dot{X}_+(t)\right\|$ is nonincreasing.*

(iv) *$\dot{X}_+(t) = -m(\mathbb{M}(X(t)))$ holds for all $t \geq 0$. Here $m(K)$ is the element of $K \subset \mathbf{R}^n$ with minimal norm, that is, $m(K) = \Pi_K(0) = \operatorname*{argmin}_{k \in K} \|k\|$. Therefore $\dot{X}(t) = -m(\mathbb{M}(X(t)))$ holds for all $t \in \mathcal{T}$, and so (14) is satisfied almost everywhere.*

Remark. *From (iii) we have $\left\|\dot{X}_+(t)\right\| \leq \left\|\dot{X}_+(0)\right\| = \|m(\mathbb{M}(X_0))\|$ for all $t \geq 0$, thus for $t_1, t_2 \geq 0$ we have*

$$\|X(t_1) - X(t_2)\| = \left\|\int_{t_2}^{t_1} \dot{X}_+(s) ds\right\| \leq \int_{t_2}^{t_1} \left\|\dot{X}_+(s)\right\| ds$$

$$\leq \int_{t_2}^{t_1} \|m(\mathbb{M}(X_0))\| \, ds = |t_1 - t_2| \, \|m(\mathbb{M}(X_0))\| \,.$$

Therefore the theorem implies that X is Lipschitz-continuous, in particular with parameter $\|m(\mathbb{M}(X_0))\|$.

Thus our first goal is to prove the condition (11) can be equivalently written as

$$\frac{d}{dt}\begin{bmatrix} v_{\mathcal{C}} \\ i_{\mathcal{L}} \end{bmatrix} \in -\mathbb{A} \begin{bmatrix} v_{\mathcal{C}} \\ i_{\mathcal{L}} \end{bmatrix}$$

for some maximal monotone operator $\mathbb{A}\colon \mathbf{R}^{|\mathcal{C}|+|\mathcal{L}|} \rightrightarrows \mathbf{R}^{|\mathcal{C}|+|\mathcal{L}|}$. We first establish an efficient reformulation of KCL and KVL.

Lemma B.4. *There is a skew-symmetric matrix $\hat{H}\colon \mathbf{R}^{\sigma+m} \to \mathbf{R}^{\sigma+m}$ and a corresponding diagonal matrix $J\colon \mathbf{R}^{\sigma+m} \to \mathbf{R}^{\sigma+m}$ with entries 0 of 1 that satisfies*

$$\begin{bmatrix} i \\ y \end{bmatrix} \in \mathcal{N}(B), \quad \begin{bmatrix} v \\ x \end{bmatrix} \in \mathcal{R}(B^{\intercal}) \quad \Longleftrightarrow \quad \hat{u} = \hat{H}\hat{w},$$

where $\hat{u}$ and $\hat{w}$ are defined as

$$\hat{w} = [J \quad I_{\sigma+m} - J]\begin{bmatrix} v \\ x \\ i \\ y \end{bmatrix}, \quad \hat{u} = [I_{\sigma+m} - J \quad J]\begin{bmatrix} v \\ x \\ i \\ y \end{bmatrix}.$$

Moreover, let $Q\colon \mathbf{R}^{\sigma+m} \to \mathbf{R}^{\sigma+m}$ be a permutation matrix, define $w = Q\hat{w}$, $u = Q\hat{u}$. Then there is a skew-symmetric matrix H that satisfies

$$\begin{bmatrix} i \\ y \end{bmatrix} \in \mathcal{N}(B), \quad \begin{bmatrix} v \\ x \end{bmatrix} \in \mathcal{R}(B^{\intercal}) \quad \Longleftrightarrow \quad u = Hw.$$

Remark. The diagonal matrix J determines whether to select voltage or current for each component, to construct $\hat{w}$. To clarify, $w, u \in \mathbf{R}^{\sigma+m}$ are the vectors that $\{w_l, u_l\}$ becomes a current and voltage pair of a component for $l = 1, 2, \ldots, \sigma + m$. Such partitions of current, voltages values w, u and skew-symmetric matrix H were also considered in [82] with different notation. However, we introduce our method of constructing them here, as we will consider H with a special property in Corollary B.4.1 that plays a key role in the proof.

Proof. Define N and $\tilde{B}$ be matrices consisted with basis of $\mathcal{N}(B)$ and $\text{Row}(B)$ respectively. Then KCL and KVL can be shortly rewritten as

$$\begin{bmatrix} N & 0 \\ 0 & \tilde{B} \end{bmatrix}\begin{bmatrix} v \\ x \\ i \\ y \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

24

We now show there is a diagonal matrix $J\colon \mathbf{R}^{\sigma+m} \to \mathbf{R}^{\sigma+m}$ with entries 0 or 1, that makes the below square matrix invertible

$$G = \begin{bmatrix} N & 0 \\ 0 & \tilde{B} \\ J & I_{\sigma+m} - J \end{bmatrix} \in \mathbf{R}^{2(\sigma+m)\times 2(\sigma+m)}.$$

Name $N_0 = \begin{bmatrix} N \\ 0 \end{bmatrix}$ and $\tilde{B}_0 = \begin{bmatrix} 0 \\ \tilde{B} \end{bmatrix}$. We will attach the standard basis vectors or 0 below, and increase the index with attached number of rows. We proceed induction on the index, until the index becomes $\sigma + m$.

Suppose, for $0 \le k \le \sigma + m - 1$, N_k and $\tilde{B}_k$ satisfy the form

$$N_k = \begin{bmatrix} N \\ 0 \\ j_1\mathbf{e}_1 \\ \vdots \\ j_k\mathbf{e}_k \end{bmatrix}, \quad \tilde{B}_k = \begin{bmatrix} 0 \\ \tilde{B} \\ (1 - j_1)\mathbf{e}_1 \\ \vdots \\ (1 - j_k)\mathbf{e}_k \end{bmatrix}, \tag{15}$$

where $j_l \in \{0, 1\}$ and $\mathbf{e}_l \in \mathbf{R}^{\sigma+m}$ is a standard basis (row) vector for $1 \le l \le k$. We claim either $\mathbf{e}_{k+1} \notin \mathrm{Row}(N_k)$ or $\mathbf{e}_{k+1} \notin \mathrm{Row}(\tilde{B}_k)$ is true.

Proof by contradiction. Suppose not. That is, suppose $\mathbf{e}_{k+1} \in \mathrm{Row}(N_k)$ and $\mathbf{e}_{k+1} \in \mathrm{Row}(\tilde{B}_k)$. Then there are $\mathbf{n} = (n_1, \ldots, n_{\sigma+m}) \in \mathrm{Row}(N)$, $\mathbf{r} = (r_1, \ldots, r_{\sigma+m}) \in \mathrm{Row}(\tilde{B})$ and coefficients a_l, b_l such that

$$\mathbf{e}_{k+1} = \mathbf{n} + \sum_{l=1}^{k} a_l j_l \mathbf{e}_l = \mathbf{r} + \sum_{l=1}^{k} b_l(1 - j_l)\mathbf{e}_l.$$

Taking inner product with $\mathbf{e}_p$, $1 \le p \le \sigma + m$, we have

$$n_p = \begin{cases} -a_p & \text{if } 1 \le p \le k, \ j_p = 1 \\ 0 & \text{if } 1 \le p \le k, \ j_p = 0 \\ 1 & \text{if } p = k + 1 \\ 0 & \text{if } k + 1 < p \le \sigma + m, \end{cases} \qquad r_p = \begin{cases} 0 & \text{if } 1 \le p \le k, \ j_p = 1 \\ -b_p & \text{if } 1 \le p \le k, \ j_p = 0 \\ 1 & \text{if } p = k + 1 \\ 0 & \text{if } k + 1 < p \le \sigma + m. \end{cases}$$

Therefore

$$\langle \mathbf{n}, \mathbf{r} \rangle = \sum_{p=1}^{\sigma+m} n_p r_p = n_{k+1} r_{k+1} = 1.$$

By the way, since $\mathbf{n} \in \mathrm{Row}(N) = N(B)$, $\mathbf{r} \in \mathrm{Row}(\tilde{B}) = R(N^{\mathsf{T}})$ we have $\mathbf{n} \perp \mathbf{r}$ and so $\langle \mathbf{n}, \mathbf{r} \rangle = 0$. A contradiction, we conclude either $\mathbf{e}_{k+1} \notin \mathrm{Row}(N_k)$ or $\mathbf{e}_{k+1} \notin \mathrm{Row}(\tilde{B}_k)$ is true.

From the proved claim, we can extend N_0, $\tilde{B}_0$ to $N_{\sigma+m}$, $\tilde{B}_{\sigma+m}$ with keeping the form of (15) by repeating the process below. Recall, the desired form of the matrix was

$$G = \begin{bmatrix} N & 0 \\ 0 & \tilde{B} \\ J & I_{\sigma+m} - J \end{bmatrix}$$

with diagonal matrix $J \in \mathbf{R}^{(\sigma+m)\times(\sigma+m)}$ with entries 0 or 1. By the construction, we see matrix $\begin{bmatrix} N_{\sigma+m} & \tilde{B}_{\sigma+m} \end{bmatrix}$ satisfies the desired form. Moreover, we know the nonzero rows of $N_{\sigma+m}$ and $\tilde{B}_{\sigma+m}$ are linearly independent respectively, by their construction. By the form of G we see if l-th row of $N_{\sigma+m}$ is nonzero then l-th row of $\tilde{B}_{\sigma+m}$ is zero and vice-versa, we conclude the rows of G are linearly independent. Therefore, G is invertible.

Observe

$$G \begin{bmatrix} v \\ x \\ i \\ y \end{bmatrix} = \begin{bmatrix} N & 0 \\ 0 & \tilde{B} \\ J & I_{\sigma+m} - J \end{bmatrix} \begin{bmatrix} v \\ x \\ i \\ y \end{bmatrix} = \begin{bmatrix} 0 \\ \hat{w} \end{bmatrix} \implies \begin{bmatrix} v \\ x \\ i \\ y \end{bmatrix} = G^{-1} \begin{bmatrix} 0 \\ \hat{w} \end{bmatrix}. \tag{16}$$

25

We know above equation holds for arbitrarily chosen (v, x), (i, y) that satisfies KVL and KCL respectively. Observe $\mathbf{dom}(G) = R(B^\mathsf{T}) \times N(B)$ and from dimension theorem we know

$$\dim(R(B^\mathsf{T}) \times N(B)) = \dim(R(B^\mathsf{T})) + \dim(N(B)) = \sigma + m.$$

As G is invertible, we have $\dim(R(G)) = \dim(\mathbf{dom}(G)) = \sigma + m$. Therefore the values of the components of $\hat{w}$ can be arbitrary values in $\mathbf{R}$.

Rearranging the rows of G^{-1}, from (16) we obtain $\tilde{H} \in \mathbf{R}^{2(\sigma+m) \times 2(\sigma+m)}$ that satisfies

$$\begin{bmatrix} \hat{w} \\ \hat{u} \end{bmatrix} = \tilde{H} \begin{bmatrix} 0 \\ \hat{w} \end{bmatrix} = \begin{bmatrix} \tilde{H}_0^w & \tilde{H}_w^w \\ \tilde{H}_0^u & \tilde{H}_w^u \end{bmatrix} \begin{bmatrix} 0 \\ \hat{w} \end{bmatrix},$$

where the block matrices are in $\mathbf{R}^{(\sigma+m) \times (\sigma+m)}$. Now, naming $\hat{H} = \tilde{H}_w^u$ we get

$$\hat{u} = \hat{H}\hat{w}.$$

Now to show H is skew-symmetric, recall from $(v, x) \in R(B^\mathsf{T})$ and $(i, y) \in N(B)$ we have $\langle (v, x), (i, y) \rangle = 0$. Thus for all $\hat{w} \in \mathbf{R}^{\sigma+m}$, we have

$$\left\langle \hat{w}, \hat{H}\hat{w} \right\rangle = \langle \hat{w}, \hat{u} \rangle = \langle (v, x), (i, y) \rangle = 0.$$

Therefore $\hat{H}$ is skew-symmetric.

Finally, let $Q\colon \mathbf{R}^{\sigma+m} \to \mathbf{R}^{\sigma+m}$ be a permutation matrix. Define $H = Q\hat{H}Q^\mathsf{T}$. Since

$$H^\mathsf{T} = Q\hat{H}^\mathsf{T}Q^\mathsf{T} = Q(-\hat{H})Q^\mathsf{T} = -H,$$

H is skew-symmetric. And from $Q^\mathsf{T}Q = I_{\sigma+m}$ we have

$$\hat{u} = \hat{H}\hat{w} \quad \Longleftrightarrow \quad u = Q\hat{u} = Q\hat{H}\hat{w} = Q\hat{H}Q^\mathsf{T}Q\hat{w} = Hw,$$

we conclude the proof.

$\square$

Corollary B.4.1. *Recall $\hat{w}$ is composed with voltage or current values of each component. Integrate the values of resistors and denote as r, integrate $v_\mathcal{C}, i_\mathcal{L}$ as p and integrate $i_\mathcal{C}, v_\mathcal{L}$ as p_*. Then we may rearrange the elements of $\hat{w}$ with certain permutation matrix Q, that $w = Q\hat{w}$ can be decomposed as following order*

$$w = \begin{bmatrix} w_p \\ w_{p_*} \\ w_r \end{bmatrix}, \qquad where \quad w_p = \begin{bmatrix} w_{v_\mathcal{C}} \\ w_{i_\mathcal{L}} \end{bmatrix}, \quad w_{p_*} = \begin{bmatrix} w_{i_\mathcal{C}} \\ w_{v_\mathcal{L}} \end{bmatrix}, \quad w_r = \begin{bmatrix} w_{v_r} \\ w_{i_r} \end{bmatrix} = \begin{bmatrix} w_{v_\mathcal{R}} \\ w_x \\ w_{i_\mathcal{R}} \\ w_y \end{bmatrix}.$$

Consider rewriting $u = Hw$ in the decomposed way as

$$\begin{bmatrix} u_{p_*} \\ u_p \\ u_r \end{bmatrix} = \begin{bmatrix} H_p^{p_*} & H_{p_*}^{p_*} & H_r^{p_*} \\ H_p^p & H_{p_*}^p & H_r^p \\ H_p^r & H_{p_*}^r & H_r^r \end{bmatrix} \begin{bmatrix} w_p \\ w_{p_*} \\ w_r \end{bmatrix}. \tag{17}$$

Then there is a diagonal matrix J satisfies the properties considered in Lemma B.4, that corresponding H satisfies

$$H_{p_*}^r = 0, \quad H_{p_*}^p = 0, \quad H_r^{p_*} = 0.$$

Proof. Name the indices as $\mathcal{C}_l, \mathcal{L}_k \in \{1, \dots, \sigma + m\}$ for $l \in \{1, \dots, |\mathcal{C}|\}$, $k \in \{1, \dots, |\mathcal{L}|\}$ that satisfy

$$\mathbf{e}_{\mathcal{C}_l} \begin{bmatrix} v \\ x \end{bmatrix} = v_{\mathcal{C}_l}, \qquad \mathbf{e}_{\mathcal{L}_k} \begin{bmatrix} i \\ y \end{bmatrix} = i_{\mathcal{L}_k}.$$

First we put $\mathbf{e}_{\mathcal{C}_l}$'s and $\mathbf{e}_{\mathcal{L}_k}$'s in J as many as possible. That is, determine the values of $j_{\mathcal{C}_l}$'s and $j_{\mathcal{L}_k}$'s to satisfy

- $\{\mathbf{e}_{\mathcal{C}_l} \mid j_{\mathcal{C}_l} = 1\}$ is linearly independent to $\mathrm{Row}(N)$.

26

- $\{\mathbf{e}_{\mathcal{C}_l} \mid j_{\mathcal{C}_l} = 1\} \cup \{\mathbf{e}_{\mathcal{C}_{l'}}\}$ is linearly dependent to $\text{Row}(N)$ for any $l' \in \{j_{\mathcal{C}_l} \neq 1\}$.

- $\{\mathbf{e}_{\mathcal{L}_s} \mid j_{\mathcal{L}_k} = 0\}$ is linearly independent to $\text{Row}(\tilde{B})$.

- $\{\mathbf{e}_{\mathcal{L}_s} \mid j_{\mathcal{L}_k} = 0\} \cup \{\mathbf{e}_{\mathcal{L}_{s'}}\}$ is linearly dependent to $\text{Row}(\tilde{B})$ for any $s' \in \{j_{\mathcal{L}_s} \neq 0\}$.

Next, fill the remaining j's as we've done in Lemma B.4.

Since the proof can be applied using the same argument to other cases, we will focus on a specific case. Focusing on the last row of (17), we can furthermore decompose and write as following

$$u_r = \begin{bmatrix} H_p^r & H_{p_*}^r & H_r^r \end{bmatrix} \begin{bmatrix} w_p \\ w_{p_*} \\ w_r \end{bmatrix} \iff \begin{bmatrix} u_{i_r} \\ u_{v_r} \end{bmatrix} = \begin{bmatrix} 0 & H_{i\mathcal{L}}^{i_r} & H_{i\mathcal{C}}^{i_r} & 0 & 0 & H_{i_r}^{i_r} \\ H_{v\mathcal{C}}^{v_r} & 0 & 0 & H_{v\mathcal{L}}^{v_r} & H_{v_r}^{v_r} & 0 \end{bmatrix} \begin{bmatrix} w_{v_{\mathcal{C}}} \\ w_{i_{\mathcal{L}}} \\ w_{i_{\mathcal{C}}} \\ w_{v_{\mathcal{L}}} \\ w_{v_r} \\ w_{i_r} \end{bmatrix}.$$

Note, since above equations origin from KCL and KVL (which are linear equations only with current values or voltage values), $H_\alpha^\beta = 0$ if α is current and β is voltage, and vice-versa. Refer [135, Theorem 6.3].

Observe $H_{p_*}^r = \begin{bmatrix} H_{i\mathcal{C}}^{i_r} & 0 \\ 0 & H_{v\mathcal{L}}^{v_r} \end{bmatrix}$, here we show $H_{i\mathcal{C}}^{i_r} = 0$. Focusing on arbitrary k'th row of $H_{i\mathcal{C}}^{i_r}$, from above equality we get

$$u_{i_{r_k}} = \begin{bmatrix} H_{i\mathcal{L}}^{i_r} & H_{i\mathcal{C}}^{i_r} & H_{i_r}^{i_r} \end{bmatrix}_k \begin{bmatrix} w_{i_{\mathcal{L}}} \\ w_{i_{\mathcal{C}}} \\ w_{i_r} \end{bmatrix} \iff 0 = \begin{bmatrix} H_{i\mathcal{L}}^{i_r} & H_{i\mathcal{C}}^{i_r} & H_{i_r}^{i_r} \end{bmatrix}_k \begin{bmatrix} w_{i_{\mathcal{L}}} \\ w_{i_{\mathcal{C}}} \\ w_{i_r} \end{bmatrix} - u_{i_{r_k}},$$

where the subscript k means the k'th row of the block matrix. As this is a linear equation of current values, it origins from KCL, thus there is a vector $\mathbf{r} \in \text{Row}(\tilde{B})$ corresponding to this equation, *i.e.*

$$\mathbf{r} \begin{bmatrix} i \\ y \end{bmatrix} = \begin{bmatrix} H_{i\mathcal{L}}^{i_r} & H_{i\mathcal{C}}^{i_r} & H_{i_r}^{i_r} \end{bmatrix}_k \begin{bmatrix} w_{i_{\mathcal{L}}} \\ w_{i_{\mathcal{C}}} \\ w_{i_r} \end{bmatrix} - u_{i_{r_k}}.$$

On the other hand, as $w_{i_{\mathcal{L}}}, w_{i_{\mathcal{C}}}, w_{i_r}$ are consisted with the components of i, y that corresponds to $j_l = 0$, there are coefficient vectors $a \in \mathbf{R}^{|\mathcal{L}|}, b \in \mathbf{R}^{|\mathcal{C}|}, c \in \mathbf{R}^{|\mathcal{R}|+m}$ that satisfies

$$\begin{bmatrix} H_{i\mathcal{L}}^{i_r} & H_{i\mathcal{C}}^{i_r} & H_{i_r}^{i_r} \end{bmatrix}_k \begin{bmatrix} w_{i_{\mathcal{L}}} \\ w_{i_{\mathcal{C}}} \\ w_{i_r} \end{bmatrix} - u_{i_{r_k}} = \left(\sum_{s \in \{j_{\mathcal{L}_s} = 0\}} a_s \mathbf{e}_{\mathcal{L}_s} + \sum_{l \in \{j_{\mathcal{C}_l} = 0\}} b_l \mathbf{e}_{\mathcal{C}_l} + \sum_{q \in \{j_{r_q} = 0\}} c_q \mathbf{e}_{r_q} - \mathbf{e}_{r_k} \right) \begin{bmatrix} i \\ y \end{bmatrix}.$$

Note b_l's correspond to components of $\begin{bmatrix} H_{i\mathcal{C}}^{i_r} \end{bmatrix}_k$. Organizing, we have

$$\mathbf{r} = \sum_{s \in \{j_{\mathcal{L}_s} = 0\}} a_s \mathbf{e}_{\mathcal{L}_s} + \sum_{l \in \{j_{\mathcal{C}_l} = 0\}} b_l \mathbf{e}_{\mathcal{C}_l} + \sum_{q \in \{j_{r_q} = 0\}} c_q \mathbf{e}_{r_q} - \mathbf{e}_{r_k}.$$

Observe that from right hand side, we can see $\mathbf{r}$ is orthogonal to $\{\mathbf{e}_{\mathcal{C}_{l'}} \mid j_{\mathcal{C}_{l'}} = 1\}$.

By the way, as $\{\mathbf{e}_{\mathcal{C}_{l'}} \mid j_{\mathcal{C}_{l'}} = 1\} \cup \{\mathbf{e}_{\mathcal{C}_l}\}$ is linearly dependent to $\text{Row}(N)$ for all $l \in \{j_{\mathcal{C}_l} = 0\}$, we see

$$\sum_{l \in \{j_{\mathcal{C}_l} = 0\}} b_l \mathbf{e}_{\mathcal{C}_l} \in \mathbf{span}\left(\{\mathbf{e}_{\mathcal{C}_{l'}} \mid j_{\mathcal{C}_{l'}} = 1\} \cup \text{Row}(N) \right),$$

so there is some coefficient vector $d \in \mathbf{R}^{|\mathcal{C}|}$ and $\mathbf{n} \in \text{Row}(N)$ that satisfies

$$\sum_{l \in \{j_{\mathcal{C}_l} = 0\}} b_l \mathbf{e}_{\mathcal{C}_l} = \sum_{l' \in \{j_{\mathcal{C}_{l'}} = 1\}} d_{l'} \mathbf{e}_{\mathcal{C}_{l'}} + \mathbf{n}.$$

However, as $\mathbf{r} \in \text{Row}(\tilde{B})$ and $\text{Row}(\tilde{B}) \perp \text{Row}(N)$, we have $\langle \mathbf{r}, \mathbf{n} \rangle = 0$. Moreover, as $\mathbf{r}$ is orthogonal to $\{\mathbf{e}_{\mathcal{C}_{l'}} \mid j_{\mathcal{C}_{l'}} = 1\}$, we conclude

$$0 = \left\langle \mathbf{r}, \sum_{l' \in \{j_{\mathcal{C}_{l'}} = 1\}} d_{l'} \mathbf{e}_{\mathcal{C}_{l'}} + \mathbf{n} \right\rangle = \left\langle \mathbf{r}, \sum_{l \in \{j_{\mathcal{C}_l} = 0\}} b_l \mathbf{e}_{\mathcal{C}_l} \right\rangle = \left\| \sum_{l \in \{j_{\mathcal{C}_l} = 0\}} b_l \mathbf{e}_{\mathcal{C}_l} \right\|^2 = \sum_{l \in \{j_{\mathcal{C}_l} = 0\}} b_l^2.$$

27

Therefore, as b_l's corresponds to components of $\left[H_{i_C}^{i_r}\right]_k$, we conclude $\left[H_{i_C}^{i_r}\right]_k = 0$. As k was arbitrary, we get $H_{i_C}^{i_r} = 0$. Similarly we can show $H_{v_{\mathcal{L}}}^{v_r} = 0$, and thus $H_{p_*}^r = 0$. Repeating the same argument, we can show $H_{p_*}^p = 0$. Finally, as H is skew-symmetric, we have $H_r^{p_*} = -(H_{p_*}^r)^{\mathsf{T}} = 0$. $\qquad\square$

We now move on to V-I relations of resistors. To express V-I relations in terms of w and u, we adopt partial inverse.

Definition. [23, Definition 20.42] Let $\mathbb{M}\colon \mathbf{R}^d \rightrightarrows \mathbf{R}^d$ be a set-valued operator and let K be a closed linear subspace of $\mathbf{R}^d$. Denote $\Pi_K\colon \mathbf{R}^d \to \mathbf{R}^d$ the projection onto K as

$$\Pi_K(z) = \operatorname*{argmin}_{k \in K} \|z - k\|.$$

The *partial inverse* of $\mathbb{M}$ with respect to K is the operator $\mathbb{M}_K\colon \mathbf{R}^d \rightrightarrows \mathbf{R}^d$ defined by

$$\operatorname{gra}\mathbb{M}_K = \{(\Pi_K\mathbf{x} + \Pi_{K^\perp}\mathbf{y}, \Pi_K\mathbf{y} + \Pi_{K^\perp}\mathbf{x}) \mid (\mathbf{x}, \mathbf{y}) \in \operatorname{gra}\mathbb{M}\},$$

i.e.,

$$u \in \mathbb{M}_K(w) \iff \exists \mathbf{x}, \mathbf{y} \text{ such that } \mathbf{y} \in \mathbb{M}\mathbf{x} \text{ and } (w, u) = (\Pi_K\mathbf{x} + \Pi_{K^\perp}\mathbf{y}, \Pi_K\mathbf{y} + \Pi_{K^\perp}\mathbf{x}).$$

We then prove important properties of the function related to V-I relations for resistors.

Lemma B.5. *Suppose f is μ-strongly convex and M-smooth function. Let $Q_r, H_r^r, J_r\colon \mathbf{R}^{|\mathcal{R}|+m} \to \mathbf{R}^{|\mathcal{R}|+m}$ be a permutation matrix, a skew-symmetric matrix, a diagonal matrix with entries 1 or 0 respectively and let $K = \mathcal{R}(J_r)$. Define $F\colon \mathbf{R}^{|\mathcal{R}|+m} \to \mathbf{R}$ as*

$$F(v_\mathcal{R}, x) = \frac{1}{2}\|v_\mathcal{R}\|^2 + f(x).$$

Then the following holds.

   (i) $\operatorname{\mathbf{dom}}(Q_r(\nabla F)_K Q_r^{\mathsf{T}} - H_r^r)^{-1} = \mathbf{R}^{|\mathcal{R}|+m}$.

   (ii) $(Q_r(\nabla F)_K Q_r^{\mathsf{T}} - H_r^r)^{-1}$ *is Lipschitz continuous monotone operator.*

Proof. Take $(w_r^l, u_r^l) \in \mathbf{R}^{|\mathcal{R}|+m}$ for $l \in \{1, 2\}$, such that $u_r^l \in (Q_r(\nabla F)_K Q_r^{\mathsf{T}})w_r^l$. As Q_r is permutation matrix, we know $(Q_r)^{-1} = (Q_r)^{\mathsf{T}}$, and thus $Q_r^{\mathsf{T}} u_r^l \in (\nabla F)_K(Q_r^{\mathsf{T}} w_r^l)$. Then there are $\begin{bmatrix} v_\mathcal{R}^l \\ x^l \end{bmatrix}, \begin{bmatrix} i_\mathcal{R}^l \\ y^l \end{bmatrix} \in \mathbf{R}^{|\mathcal{R}|+m}$ such that

$$\begin{bmatrix} i_\mathcal{R}^l \\ y^l \end{bmatrix} = \nabla F \begin{bmatrix} v_\mathcal{R}^l \\ x^l \end{bmatrix}, \quad (Q_r^{\mathsf{T}} w_r^l, Q_r^{\mathsf{T}} u_r^l) = \left(\Pi_K \begin{bmatrix} v_\mathcal{R}^l \\ x^l \end{bmatrix} + \Pi_{K^\perp} \begin{bmatrix} i_\mathcal{R}^l \\ y^l \end{bmatrix}, \Pi_K \begin{bmatrix} i_\mathcal{R}^l \\ y^l \end{bmatrix} + \Pi_{K^\perp} \begin{bmatrix} v_\mathcal{R}^l \\ x^l \end{bmatrix}\right).$$
$$\tag{18}$$

By [23, Proposition 20.44, (iii)], we have

$$\left\langle Q_r^{\mathsf{T}}(w_r^1 - w_r^2), Q_r^{\mathsf{T}}(u_r^1 - u_r^2) \right\rangle = \left\langle \begin{bmatrix} v_\mathcal{R}^1 \\ x^1 \end{bmatrix} - \begin{bmatrix} v_\mathcal{R}^2 \\ x^2 \end{bmatrix}, \begin{bmatrix} i_\mathcal{R}^1 \\ y^1 \end{bmatrix} - \begin{bmatrix} i_\mathcal{R}^2 \\ y^2 \end{bmatrix} \right\rangle.$$

Moreover, we can check

$$\left\| \begin{bmatrix} v_\mathcal{R}^1 \\ x^1 \end{bmatrix} - \begin{bmatrix} v_\mathcal{R}^2 \\ x^2 \end{bmatrix} \right\|^2 + \left\| \begin{bmatrix} i_\mathcal{R}^1 \\ y^1 \end{bmatrix} - \begin{bmatrix} i_\mathcal{R}^2 \\ y^2 \end{bmatrix} \right\|^2$$

$$= \left\| \Pi_K \left(\begin{bmatrix} v_\mathcal{R}^1 \\ x^1 \end{bmatrix} - \begin{bmatrix} v_\mathcal{R}^2 \\ x^2 \end{bmatrix} \right) \right\|^2 + \left\| \Pi_{K^\perp} \left(\begin{bmatrix} v_\mathcal{R}^1 \\ x^1 \end{bmatrix} - \begin{bmatrix} v_\mathcal{R}^2 \\ x^2 \end{bmatrix} \right) \right\|^2$$

$$+ \left\| \Pi_K \left(\begin{bmatrix} i_\mathcal{R}^1 \\ y^1 \end{bmatrix} - \begin{bmatrix} i_\mathcal{R}^2 \\ y^2 \end{bmatrix} \right) \right\|^2 + \left\| \Pi_{K^\perp} \left(\begin{bmatrix} i_\mathcal{R}^1 \\ y^1 \end{bmatrix} - \begin{bmatrix} i_\mathcal{R}^2 \\ y^2 \end{bmatrix} \right) \right\|^2 = \|w_r^1 - w_r^2\|^2 + \|u_r^1 - u_r^2\|^2.$$

Define $\mu_{\min} = \{\mu, 1\}$ and $M_{\min} = \{M, 1\}$. Then we can check ∇F is $\mu_{\min}$-strongly convex and $M_{\min}$-smooth, we see

$$\left\langle (i_\mathcal{R}^1, y^1) - (i_\mathcal{R}^2, y^2), (v_\mathcal{R}^1, x^1) - (v_\mathcal{R}^2, x^2) \right\rangle \geq \mu_{\min} \left\| (v_\mathcal{R}^1, x^1) - (v_\mathcal{R}^2, x^2) \right\|^2,$$

$$\left\langle (i_\mathcal{R}^1, y^1) - (i_\mathcal{R}^2, y^2), (v_\mathcal{R}^1, x^1) - (v_\mathcal{R}^2, x^2) \right\rangle \geq M_{\min} \left\| (i_\mathcal{R}^1, y^1) - (i_\mathcal{R}^2, y^2) \right\|^2.$$

28

Thus

$$
\begin{aligned}
\left\langle w_r^1 - w_r^2, u_r^1 - u_r^2 \right\rangle &= \left\langle Q_r^{\intercal}(w_r^1 - w_r^2), Q_r^{\intercal}(u_r^1 - u_r^2) \right\rangle = \left\langle \begin{bmatrix} v_{\mathcal{R}}^1 \\ x^1 \end{bmatrix} - \begin{bmatrix} v_{\mathcal{R}}^2 \\ x^2 \end{bmatrix}, \begin{bmatrix} i_{\mathcal{R}}^1 \\ y^1 \end{bmatrix} - \begin{bmatrix} i_{\mathcal{R}}^2 \\ y^2 \end{bmatrix} \right\rangle \\
&\geq \frac{\mu_{\min} + M_{\min}}{2} \left(\left\| \begin{bmatrix} v_{\mathcal{R}}^1 \\ x^1 \end{bmatrix} - \begin{bmatrix} v_{\mathcal{R}}^2 \\ x^2 \end{bmatrix} \right\|^2 + \left\| \begin{bmatrix} i_{\mathcal{R}}^1 \\ y^1 \end{bmatrix} - \begin{bmatrix} i_{\mathcal{R}}^2 \\ y^2 \end{bmatrix} \right\|^2 \right) \\
&= \frac{\mu_{\min} + M_{\min}}{2} \left(\left\| w_r^1 - w_r^2 \right\|^2 + \left\| u_r^1 - u_r^2 \right\|^2 \right) \\
&\geq \frac{\mu_{\min} + M_{\min}}{2} \left\| w_r^1 - w_r^2 \right\|^2 ,
\end{aligned}
$$

we see $Q_r(\nabla F)_K Q_r^{\intercal}$ is $\frac{\mu_{\min} + M_{\min}}{2}$-strongly monotone. Note we can check $(\nabla F)_K$ is also strongly monotone, by considering the special case $Q_r = I_{|r|+m}$. Lastly, since H_r^r is skew-symmetric, we know $\langle H_r^r z, z \rangle = 0$ for all $z \in \mathbf{R}^{|\mathcal{R}|+m}$. Therefore for arbitrary $(\tilde{w}_r^l, \tilde{u}_r^l) \in \mathbf{R}^{|\mathcal{R}|+m}$ with $l \in \{1, 2\}$ such that $\tilde{u}_r^l \in (Q_r(\nabla F)_K Q_r^{\intercal} - H_r^r) \tilde{w}_r^l$, since $u_r^l = \tilde{u}_r^l + H_r^r \tilde{w}_r^l \in (Q_r(\nabla F)_K Q_r^{\intercal}) \tilde{w}_r^l$,

$$
\begin{aligned}
\left\langle \tilde{w}_r^1 - \tilde{w}_r^2, \tilde{u}_r^1 - \tilde{u}_r^2 \right\rangle &= \left\langle \tilde{w}_r^1 - \tilde{w}_r^2, u_r^1 - u_r^2 - H_r^r(\tilde{w}_r^1 - \tilde{w}_r^2) \right\rangle \\
&= \left\langle \tilde{w}_r^1 - \tilde{w}_r^2, u_r^1 - u_r^2 \right\rangle \\
&\geq \frac{\mu_{\min} + M_{\min}}{2} \left\| w_r^1 - w_r^2 \right\|^2
\end{aligned}
$$

thus $(Q_r(\nabla F)_K Q_r^{\intercal} - H_r^r)$ is also $\frac{\mu_{\min} + M_{\min}}{2}$-strongly monotone.

Now since ∇F is maximal monotone, $(\nabla F)_K$ is maximal monotone by [23, Proposition 20.44, (v)]. Since $(\nabla F)_K$ is strongly monotone, we have $\mathbf{dom}(\nabla F)_K = \mathbf{R}^{|\mathcal{R}|+m}$ by [23, Proposition 22.11], and thus $Q_r(\nabla F)_K Q_r^{\intercal}$ is maximal monotone by [129, Theorem 12]. Moreover, since both $Q_r(\nabla F)_K Q_r^{\intercal}$ and $-H_r^r$ have full domain, $(Q_r(\nabla F)_K Q_r^{\intercal} - H_r^r)$ is maximal monotone by [129, Theorem 10].

Organizing, $(Q_r(\nabla F)_K Q_r^{\intercal} - H_r^r)$ is maximal monotone and strongly monotone. Therefore by [23, Proposition 22.11], we conclude (ii). Finally, observe

$$
\begin{aligned}
(Q_r(\nabla F)_K Q_r^{\intercal} - H_r^r)^{-1} &= \left(\underbrace{Q_r(\nabla F)_K Q_r^{\intercal} - H_r^r - \frac{\mu_{\min} + M_{\min}}{2} I_{|\mathcal{R}|+m}}_{=: \mathbb{M}} + \frac{\mu_{\min} + M_{\min}}{2} I_{|\mathcal{R}|+m} \right)^{-1} \\
&= \mathbb{J}_{\frac{2}{\mu_{\min} + M_{\min}} \mathbb{M}} \circ \frac{2}{\mu_{\min} + M_{\min}} I_{|\mathcal{R}|+m}.
\end{aligned}
$$

Since $\mathbb{M}$ is monotone, $\frac{2}{\mu_{\min} + M_{\min}} \mathbb{M}$ is also monotone, by [23, Corollary 23.9] we know $\mathbb{J}_{\frac{2}{\mu_{\min} + M_{\min}} \mathbb{M}}$ is 1-Lipschitz continuous. Therefore $(Q_r(\nabla F)_K Q_r^{\intercal} - H_r^r)^{-1}$ is $\frac{2}{\mu_{\min} + M_{\min}}$-Lipschitz continuous. Finally it is monotone as it is an inverse of a monotone operator, we conclude (ii). $\qquad\square$

Finally, we are ready to prove Theorem 2.1.

Proof of Theorem 2.1. By Lemma B.2 it is suffices to show Theorem B.1.

(i) Well-posedness and Lipschitz continuity of $v_{\mathcal{C}}, i_{\mathcal{L}}$. Existence of the whole curve (v, x, i, y).
Define an operator $\mathbb{A} \colon \mathbf{R}^{|\mathcal{L}|+|\mathcal{C}|} \rightrightarrows \mathbf{R}^{|\mathcal{L}|+|\mathcal{C}|}$ as

$$
\mathbb{A} = \left\{ \left(\begin{bmatrix} v_{\mathcal{C}} \\ i_{\mathcal{L}} \end{bmatrix}, -\begin{bmatrix} i_{\mathcal{C}} \\ v_{\mathcal{L}} \end{bmatrix} \right) \;\middle|\; \exists\, v = (v_{\mathcal{R}}, v_{\mathcal{L}}, v_{\mathcal{C}}),\; i = (i_{\mathcal{R}}, i_{\mathcal{L}}, i_{\mathcal{C}}),\; (x, y) \right. \tag{19}
$$

$$
\left. \text{such that } \begin{bmatrix} i \\ y \end{bmatrix} \in \mathcal{N}(B),\; \begin{bmatrix} v \\ x \end{bmatrix} \in \mathcal{R}(B^{\intercal}),\; y = \nabla f(x),\; v_{\mathcal{R}} = i_{\mathcal{R}} \right\}.
$$

We prove $\mathbb{A}$ is maximal monotone by providing an explicit expression of $\mathbb{A}$ and apply Theorem B.3.

29

From Corollary B.4.1, we know there is a diagonal matrix $J\colon \mathbf{R}^{\sigma+m} \to \mathbf{R}^{\sigma+m}$, a permutation matrix $Q\colon \mathbf{R}^{\sigma+m} \to \mathbf{R}^{\sigma+m}$ and a corresponding skew-symmetric matrix $H\colon \mathbf{R}^{\sigma+m} \to \mathbf{R}^{\sigma+m}$ that satisfies

$$\begin{bmatrix} i \\ y \end{bmatrix} \in \mathcal{N}(B), \ \begin{bmatrix} v \\ x \end{bmatrix} \in \mathcal{R}(B^{\mathsf{T}}) \iff \begin{bmatrix} u_{p*} \\ u_p \\ u_r \end{bmatrix} = \begin{bmatrix} H_p^{p*} & H_{p*}^{p*} & H_r^{p*} \\ H_p^p & 0 & 0 \\ H_p^r & 0 & H_r^r \end{bmatrix} \begin{bmatrix} w_p \\ w_{p*} \\ w_r \end{bmatrix} \qquad (20)$$

where w and u are defined as in Lemma B.4.

Define $F\colon \mathbf{R}^{|\mathcal{R}|+m} \to \mathbf{R}$ as $F(v_{\mathcal{R}}, x) = \frac{1}{2}\|v_{\mathcal{R}}\|^2 + f(x)$. Then it is straight forward that

$$y = \nabla f(x), \ v_{\mathcal{R}} = i_{\mathcal{R}} \iff \begin{bmatrix} i_{\mathcal{R}} \\ y \end{bmatrix} = \nabla F \begin{bmatrix} v_{\mathcal{R}} \\ x \end{bmatrix}.$$

By the construction of w, u, there is a diagonal matrix $J_r\colon \mathbf{R}^{|\mathcal{R}|+m} \to \mathbf{R}^{|\mathcal{R}|+m}$ with entries 1 or 0 and a permutation matrix $Q_r\colon \mathbf{R}^{|\mathcal{R}|+m} \to \mathbf{R}^{|\mathcal{R}|+m}$ that satisfies

$$\begin{bmatrix} Q_r^{-1} w_r \\ Q_r^{-1} u_r \end{bmatrix} = \begin{bmatrix} Q_r^{\mathsf{T}} w_r \\ Q_r^{\mathsf{T}} u_r \end{bmatrix} = \begin{bmatrix} J_r & I_{|\mathcal{R}|+m} \\ I_{|\mathcal{R}|+m} & J_r \end{bmatrix} \begin{bmatrix} v_{\mathcal{R}} \\ x \\ i_{\mathcal{R}} \\ y \end{bmatrix}.$$

Define $K_r = \mathcal{R}(J_r)$. Note, we can check $\Pi_{K_r}(z) = J_r z$ and $\Pi_{K_r^{\perp}}(z) = (I_{|\mathcal{R}|+m} - J_r)z$ for all $z \in \mathbf{R}^{|\mathcal{R}|+m}$. Recalling (18) in Lemma B.5 we see

$$\begin{bmatrix} i_{\mathcal{R}} \\ y \end{bmatrix} = \nabla F \begin{bmatrix} v_{\mathcal{R}} \\ x \end{bmatrix} \iff u_r = (Q_r(\nabla F)_{K_r} Q_r^{\mathsf{T}})w_r.$$

From

$$(Q_r(\nabla F)_{K_r} Q_r^{\mathsf{T}})w_r = u_r = H_p^r w_p + H_r^r w_r,$$

we get expression for w_r in terms of w_p

$$w_r = (Q_r(\nabla F)_{K_r} Q_r^{\mathsf{T}} - H_r^r)^{-1} H_p^r w_p. \qquad (21)$$

Now focusing on the expression for $\begin{bmatrix} u_{p*} \\ u_p \end{bmatrix}$, since $H_p^r = -(H_r^{p*})^{\mathsf{T}}$ as H is skew-symmetric, eliminating w_r by applying (21) we see

$$\begin{bmatrix} u_{p*} \\ u_p \end{bmatrix} = \begin{bmatrix} H_p^{p*} & H_{p*}^{p*} & H_r^{p*} \\ H_p^p & 0 & 0 \end{bmatrix} \begin{bmatrix} w_p \\ w_{p*} \\ w_r \end{bmatrix} \qquad (22)$$

$$= -\underbrace{\left(-\begin{bmatrix} H_p^{p*} & H_{p*}^{p*} \\ H_p^p & 0 \end{bmatrix} + \begin{bmatrix} H_p^r & 0 \end{bmatrix}^{\mathsf{T}} (Q_r(\nabla F)_{K_r} Q_r^{\mathsf{T}} - H_r^r)^{-1} \begin{bmatrix} H_p^r & 0 \end{bmatrix} \right)}_{\mathbb{B}} \begin{bmatrix} w_p \\ w_{p*} \end{bmatrix}.$$

Since H is skew-symmetric, its principal minors $-\begin{bmatrix} H_p^{p*} & H_{p*}^{p*} \\ H_p^p & 0 \end{bmatrix}$ and $-H_r^r$ are skew-symmetric and so maximal monotone. Furthermore, by Lemma B.5 we have $(Q_r(\nabla F)_{K_r} Q_r^{\mathsf{T}} - H_r^r)^{-1}$ is maximal monotone and $\mathbf{dom}((\nabla F)^{1,-1} - H_r^r)^{-1} = \mathbf{R}^{|\mathcal{R}|+m}$. Invoking [129, Theorem 11, 12], we conclude $\mathbb{B}$ is maximal monotone.

Organizing, we see

$$\begin{bmatrix} i_{\mathcal{C}} \\ v_{\mathcal{L}} \end{bmatrix} \in -\mathbb{A} \begin{bmatrix} v_{\mathcal{C}} \\ i_{\mathcal{L}} \end{bmatrix} \iff \begin{bmatrix} i \\ y \end{bmatrix} \in \mathcal{N}(B), \ \begin{bmatrix} v \\ x \end{bmatrix} \in \mathcal{R}(B^{\mathsf{T}}), \ y = \nabla f(x), \ v_{\mathcal{R}} = i_{\mathcal{R}}$$

$$\iff \begin{bmatrix} u_{p*} \\ u_p \end{bmatrix} = -\mathbb{B} \begin{bmatrix} w_p \\ w_{p*} \end{bmatrix}, \ \text{where} \ \begin{bmatrix} w \\ u \end{bmatrix} = \begin{bmatrix} Q & 0 \\ 0 & Q \end{bmatrix} \begin{bmatrix} J & I_{\sigma+m} - J \\ I_{\sigma+m} - J & J \end{bmatrix} \begin{bmatrix} v \\ x \\ i \\ y \end{bmatrix}.$$

Therefore there is a diagonal matrix $J_{p,p*}\colon \mathbf{R}^{|\mathcal{L}|+|\mathcal{C}|} \to \mathbf{R}^{|\mathcal{L}|+|\mathcal{C}|}$ with entries 1 or 0 and a permutation matrix $Q_{p,p*}\colon \mathbf{R}^{|\mathcal{L}|+|\mathcal{C}|} \to \mathbf{R}^{|\mathcal{L}|+|\mathcal{C}|}$ that satisfies

$$\begin{bmatrix} i_{\mathcal{C}} \\ v_{\mathcal{L}} \end{bmatrix} \in -\mathbb{A} \begin{bmatrix} v_{\mathcal{C}} \\ i_{\mathcal{L}} \end{bmatrix} \iff \begin{bmatrix} u_{p*} \\ u_p \end{bmatrix} = -\mathbb{B} \begin{bmatrix} w_p \\ w_{p*} \end{bmatrix},$$

30

where

$$
\begin{bmatrix} Q_{p,p^*}^\mathsf{T} & 0 \\ 0 & Q_{p,p^*}^\mathsf{T} \end{bmatrix} \begin{bmatrix} w_p \\ w_{p^*} \\ u_{p^*} \\ u_p \end{bmatrix} = \begin{bmatrix} J_{p,p^*} & I_{|\mathcal{L}|+|\mathcal{C}|} - J_{p,p^*} \\ I_{|\mathcal{L}|+|\mathcal{C}|} - J_{p,p^*} & J_{p,p^*} \end{bmatrix} \begin{bmatrix} v_\mathcal{C} \\ i_\mathcal{L} \\ i_\mathcal{C} \\ v_\mathcal{L} \end{bmatrix}.
$$

Therefore, for $K_{p,p^*} = \mathcal{R}(J_{p,p^*})$ we have

$$
\mathbb{A}_{K_{p,p^*}} = Q_{p,p^*}^\mathsf{T}\, \mathbb{B}\, Q_{p,p^*}.
$$

Since $\mathbb{B}$ is a maximal monotone operator with $\mathbf{dom}\,\mathbb{B} = \mathbf{R}^{|\mathcal{L}|+|\mathcal{C}|}$, we have $Q_{p,p^*}^\mathsf{T}\, \mathbb{B}\, Q_{p,p^*}$ is maximal monotone. Finally from [23, Proposition 20.44, (v)], we conclude $\mathbb{A}$ is maximal monotone.

By applying Theorem B.3 and its remark, we know there is a unique Lipschitz continuous curve $(v_\mathcal{C}, i_\mathcal{L})\colon [0,\infty) \to \mathbf{R}^{|\mathcal{L}|} \times \mathbf{R}^{|\mathcal{C}|}$ that satisfies

$$
\frac{d}{dt}\begin{bmatrix} v_\mathcal{C}(t) \\ i_\mathcal{L}(t) \end{bmatrix} = -m\mathbb{A}\begin{bmatrix} v_\mathcal{C}(t) \\ i_\mathcal{L}(t) \end{bmatrix} \in -\mathbb{A}\begin{bmatrix} v_\mathcal{C}(t) \\ i_\mathcal{L}(t) \end{bmatrix}
$$

for almost all $t \in [0,\infty)$, where $m\mathbb{A}$ is the minimum-norm selection of $\mathbb{A}$. Let $\begin{bmatrix} i_\mathcal{C}(t) \\ v_\mathcal{L}(t) \end{bmatrix} = -m\mathbb{A}\begin{bmatrix} v_\mathcal{C}(t) \\ i_\mathcal{L}(t) \end{bmatrix}$. Moreover, the definition of $\mathbb{A}$ implies the existence of accompanying curves $v_\mathcal{R}$, $i_\mathcal{R}$, x and y that satisfy KCL, KVL and V-I relations. This concludes the existence of the curve.

(ii) The whole flow (v, x, i, y) is well-posed and Lipschitz continuous.
Finally, we show other curves besides $(v_\mathcal{L}, i_\mathcal{C})$ are defined uniquely and Lipschitz continuous. To do so, we prove there is a Lipschitz continuous function $\mathcal{G}\colon \mathbf{R}^{|\mathcal{C}|} \times \mathbf{R}^{|\mathcal{L}|} \to \mathbf{R}^\sigma \times \mathbf{R}^\sigma \times \mathbf{R}^m \times \mathbf{R}^m$ that satisfies

$$
\mathcal{G}(v_\mathcal{L}, i_\mathcal{C}) = (v, i, x, y).
$$

We prove the claim by finding the explicit expression of the component functions of G. We first show one key equation

$$
w_{p^*} = -(H_{p^*}^{p^*})^\mathsf{T} u_{p^*}. \tag{23}
$$

From (20) we have $H_{p^*}^{p^*} w_{p^*} = u_{p^*}$. And as H is skew-symmetric, we have $H_p^p = -(H_{p^*}^{p^*})^\mathsf{T}$. The core information we additionally use here, is the V-I relations $\frac{d}{dt}w_p = u_{p^*}$. As differentiation is a linear operation, we get (23) by following

$$
-\left(H_{p^*}^{p^*}\right)^\mathsf{T} u_{p^*} = H_p^p u_{p^*} = H_p^p\left(\frac{d}{dt}w_p\right) = \frac{d}{dt}\left(H_p^p w_p\right) = \frac{d}{dt}u_p = w_{p^*}.
$$

Recall, from (22) we have

$$
u_{p^*} = \left(H_p^{p^*} + H_r^{p^*}\left(Q_r(\nabla F)_{K_r}Q_r^\mathsf{T} - H_r^r\right)^{-1} H_p^r\right) w_p + H_{p^*}^{p^*} w_{p^*}.
$$

Moving the last term of right hand side to left hand side, multiplying both sides by $-(H_{p^*}^{p^*})^\mathsf{T}$ and using (23), we have

$$
\left(\mathbb{I} + (H_{p^*}^{p^*})^\mathsf{T} H_{p^*}^{p^*}\right) w_{p^*} = -(H_{p^*}^{p^*})^\mathsf{T}\left(H_p^{p^*} + H_r^{p^*}\left(Q_r(\nabla F)_{K_r}Q_r^\mathsf{T} - H_r^r\right)^{-1} H_p^r\right) w_p.
$$

Since $\mathbb{I} + (H_{p^*}^{p^*})^\mathsf{T} H_{p^*}^{p^*} \succ 0$ its inverse exists, we conclude

$$
w_{p^*} = \underbrace{-\left(\mathbb{I} + (H_{p^*}^{p^*})^\mathsf{T} H_{p^*}^{p^*}\right)^{-1} (H_{p^*}^{p^*})^\mathsf{T}\left(H_p^{p^*} + H_r^{p^*}\left(Q_r(\nabla F)_{K_r}Q_r^\mathsf{T} - H_r^r\right)^{-1} H_p^r\right)}_{=:\mathbb{C}} w_p.
$$

Organizing, we have

$$
\begin{bmatrix} w_p \\ w_{p^*} \\ w_r \end{bmatrix} = \left(\begin{bmatrix} \mathbb{I} \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ \mathbb{C} \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ (Q_r(\nabla F)_{K_r}Q_r^\mathsf{T} - H_r^r)^{-1} H_p^r \end{bmatrix}\right) w_p.
$$

From Lemma B.5 we know $\left(Q_r(\nabla F)_{K_r}Q_r^\mathsf{T} - H_r^r\right)^{-1}$ is Lipschitz continuous, and clearly linear operators are Lipschitz continuous, so $\mathbb{C}$ is Lipschitz continuous as it is composition and sum of

31

Lipschitz continuous functions. Therefore $w_p \mapsto w$ is Lipschitz continuous. Finally since $u = Hw$ and H is indeed Lipschitz continuous as a linear operator, mapping $w_p \mapsto u$ is also Lipschitz continuous. As (w, u) is rearrangement of (v, i, x, y) and $w_{v_{\mathcal{C}}}, w_{i_{\mathcal{L}}}$ are component functions of $v_{\mathcal{C}}$ and $i_{\mathcal{L}}$, we get the desired result.

For (v, i, x, y) that satisfies (11) with proper initial value, we know $(v_{\mathcal{C}}, i_{\mathcal{L}})$ is uniquely defined by previous observation and $(v, i, x, y) = \mathcal{G}(v_{\mathcal{C}}, i_{\mathcal{L}})$ should hold, we conclude (v, i, x, y) is uniquely determined since $\mathcal{G}$ is single valued. Furthermore, as $v_{\mathcal{C}}(t)$ and $i_{\mathcal{L}}(t)$ are Lipschitz continuous with respect to t, we have $(v(t), i(t), x(t), y(t)) = \mathcal{G}(v_{\mathcal{C}}(t), i_{\mathcal{L}}(t))$ is also Lipschitz continuous as it is composition of Lipschitz continuous functions. This concludes the proof.

$\square$

32

C Equilibrium condition

Define the set of voltages and currents in the equilibrium of the interconnect

$$D_{x,y} = \left\{ (v,i) \mid Ai = \begin{bmatrix} -y \\ 0 \end{bmatrix}, v = A^\mathsf{T} \begin{bmatrix} x \\ e \end{bmatrix}, v_\mathcal{R} = D_\mathcal{R} i_\mathcal{R}, v_\mathcal{L} = 0, i_\mathcal{C} = 0 \right\}.$$

Lemma C.1. *Assume the dynamic interconnect is admissible. Then for all $(v,i) \in D_{x,y}$ we have*

$$v_\mathcal{R} = i_\mathcal{R} = 0.$$

Proof. First, note that for all (v,i) in the dynamic interconnect

$$\langle v, i \rangle = \langle e, Ai \rangle = \left\langle \begin{bmatrix} x \\ e \end{bmatrix}, \begin{bmatrix} -y \\ 0 \end{bmatrix} \right\rangle = -\langle x, y \rangle. \tag{24}$$

Now suppose $(v,i) \in D_{x,y}$, then we have

$$\langle v, i \rangle = \langle v_\mathcal{R}, i_\mathcal{R} \rangle = \|i_\mathcal{R}\|_{D_\mathcal{R}}^2 = -\langle x, y \rangle.$$

From the admissibility assumption we have $x \in \mathcal{R}(E^\mathsf{T})$ and $y \in \mathcal{N}(E)$. Thus

$$\|i_\mathcal{R}\|_{D_\mathcal{R}}^2 = -\langle x, y \rangle = 0,$$

which concludes the proof. $\qquad\square$

Theorem C.2. *Assume the dynamic interconnect is admissible. If $(x^\star, y^\star)$ is a primal-dual solution pair (with zero duality gap) for the optimization problem, then there exist $v_\mathcal{C} \in \mathbf{R}^{|C|}$ and $i_\mathcal{L} \in \mathbf{R}^{|L|}$ such that*

$$((0, 0, v_\mathcal{C}), (0, i_\mathcal{L}, 0)) \in D_{x^\star, y^\star}.$$

Conversely, if $y \in \partial f(x)$ and

$$((v_\mathcal{R}, v_\mathcal{L}, v_\mathcal{C}), (i_\mathcal{R}, i_\mathcal{L}, i_\mathcal{C})) \in D_{x,y},$$

then $v_\mathcal{R} = i_\mathcal{R} = 0$ and (x, y) is a primal-dual solution pair (with zero-duality) for the optimization problem.

Proof. First, observe the admissibility assumption can be rewritten as

$$\{(x,y) \mid \exists (v,i) \text{ such that } (v,i) \in D_{x,y}\} = \mathcal{R}(E^\mathsf{T}) \times \mathcal{N}(E).$$

Now suppose $(x^\star, y^\star) \in X^\star \times Y^\star$. Then by Karush-Kuhn-Tucker (KKT) optimality conditions, we have $(x^\star, y^\star) \in \mathcal{R}(E^\mathsf{T}) \times \mathcal{N}(E)$. Thus there exists $(v^\star, i^\star)$ such that

$$(v^\star, i^\star) = ((v_\mathcal{R}^\star, 0, v_\mathcal{C}^\star), (i_\mathcal{R}^\star, i_\mathcal{L}^\star, 0)) \in D_{x^\star, y^\star}.$$

Furthermore from Lemma C.1 we have $v_\mathcal{R}^\star = i_\mathcal{R}^\star = 0$. Therefore $v_\mathcal{C}^\star, i_\mathcal{L}^\star$ are the vectors that satisfiy the desired statement.

Conversely, suppose

$$(v, i) = ((v_\mathcal{R}, v_\mathcal{L}, v_\mathcal{C}), (i_\mathcal{R}, i_\mathcal{L}, i_\mathcal{C})) \in D_{x,y}.$$

From Lemma C.1, we have $v_\mathcal{R} = i_\mathcal{R} = 0$. Moreover, since there exists (v,i) such that $(v,i) \in D_{x,y}$, by admissibility assumption we have $(x,y) \in \mathcal{R}(E^\mathsf{T}) \times \mathcal{N}(E)$. Finally, given the assumption that $y \in \partial f(x)$, by using KKT optimality conditions, we conclude $(x,y) \in X^\star \times Y^\star$.

$\qquad\square$

33

D   Energy dissipation analysis

In this section, we provide the proof of Theorem 2.2. The energy function (6) used in the proof is related to the Lyapunov function considered in [159]. However, the dissipativity theory presented in [159, 160] does not directly apply to our setup. In our setup, we allow cases where v_C, i_L oscillate, for example, a circuit with a disconnected $L - C$ loop. The proof is obtained by combining Barbalat's lemma [85, Lemma 8.2] with Theorem 2.1.

Proof of Theorem 2.2. Let $(x^\star, y^\star)$ be a primal-dual solution pair. Then by Theorem C.2, there is $(v^\star, i^\star) \in D_{x^\star, y^\star}$ that satisfies

$$(v^\star, i^\star) = ((0, 0, v_{\mathcal{C}}^\star), (0, i_{\mathcal{L}}^\star, 0)). \tag{25}$$

In particular, $i_{\mathcal{C}}^\star = 0$ and $v_{\mathcal{L}}^\star = 0$. Define the total energy at time t as

$$\mathcal{E}(t) = \frac{1}{2}\|v_{\mathcal{C}} - v_{\mathcal{C}}^\star\|_{D_{\mathcal{C}}}^2 + \frac{1}{2}\|i_{\mathcal{L}} - i_{\mathcal{L}}^\star\|_{D_{\mathcal{L}}}^2.$$

Then the power at time t is given by

$$
\begin{aligned}
\frac{d}{dt}\mathcal{E}(t) &= \langle v_{\mathcal{C}} - v_{\mathcal{C}}^\star, D_{\mathcal{C}}\dot{v}_{\mathcal{C}}\rangle + \langle i_{\mathcal{L}} - i_{\mathcal{L}}^\star, D_{\mathcal{L}}\dot{i}_{\mathcal{L}}\rangle \\
&= \langle v_{\mathcal{C}} - v_{\mathcal{C}}^\star, i_{\mathcal{C}} - i_{\mathcal{C}}^\star\rangle + \langle i_{\mathcal{L}} - i_{\mathcal{L}}^\star, v_{\mathcal{L}} - v_{\mathcal{L}}^\star\rangle \\
&= -\langle v_{\mathcal{R}} - \overset{0}{y_{\mathcal{R}}^\star}, i_{\mathcal{R}} - \overset{0}{i_{\mathcal{R}}^\star}\rangle - \langle x - x^\star, y - y^\star\rangle \\
&= -\|i_{\mathcal{R}}\|_{D_{\mathcal{R}}}^2 - \underbrace{\langle x - x^\star, y - y^\star\rangle}_{\geq 0} \\
&\leq 0.
\end{aligned}
\tag{26}
$$

where we used (24) and the monotonicity of ∂f. Therefore $\mathcal{E}(\infty) = \lim_{t \to \infty} \mathcal{E}(t)$ exists. Now, integrating from 0 to ∞ we have

$$0 \leq \int_0^\infty \langle x(t) - x^\star, y(t) - y^\star\rangle dt \leq \mathcal{E}(0) - \mathcal{E}(\infty) < \infty.$$

From Theorem 2.1 we know the integrand is Lipschitz continuous, by Barbalat's lemma we get

$$\lim_{t \to \infty} \langle x(t) - x^\star, y(t) - y^\star\rangle = 0.$$

Since f is μ-strongly convex and M-smooth, ∇f and $(\nabla f)^{-1}$ are strictly monotone, we conclude

$$\lim_{t \to \infty} x(t) = x^\star, \quad \lim_{t \to \infty} y(t) = y^\star,$$

which is our desired result.

$\square$

34

E Centralized classical algorithms

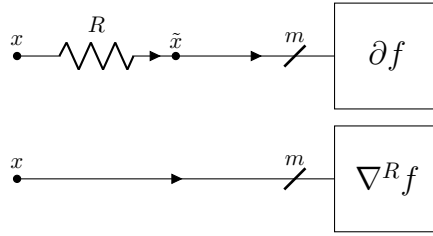
E.1 Resistors and Moreau envelope

For $R > 0$, define the Moreau envelope of $f \colon \mathbf{R}^m \to \mathbf{R}$ of parameter R as

$$^R f(x) = \inf_{z \in \mathbf{R}^m} \left(f(z) + \frac{1}{2R} \|z - x\|_2^2 \right).$$

Then $^R f$ is $1/R$-smooth with gradient given by

$$\nabla^R f(x) = \frac{1}{R}(x - \mathbf{prox}_{Rf}(x)). \tag{27}$$

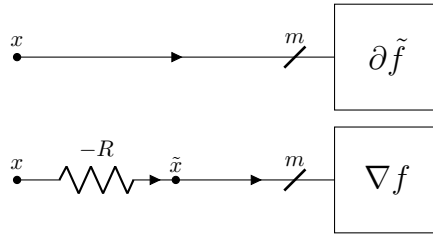
In this section we show that composing linear resistors with ∂f is equivalent to taking a Moreau envelope of f. See two circuits below.

By KCL and Ohm's law for the first circuit, we have

$$\frac{1}{R}(x - \tilde{x}) = i \in \partial f(\tilde{x}),$$

which is equivalent to $\tilde{x} = \mathbf{prox}_{Rf}(x)$. Using identity for the gradient of the Moreau envelope, we get

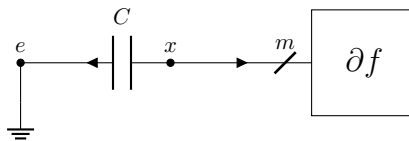
$$\nabla^R f(x) = \frac{1}{R}(x - \mathbf{prox}_{Rf}(x)) = \frac{1}{R}(x - \tilde{x}) = i.$$

Therefore, the V-I relation on m pins of x in both circuits is identical.

As a consequence, consider f to be $1/R$-smooth. Let $\tilde{f}$ be pre-Moreau envelope of f, *i.e.*, $^R \tilde{f} = f$. Note that $\tilde{f}$ is a convex function. Then from the series connection of the resistors ($-R$ in series with R is the same as 0-ohm resistor), we get the equivalence of the two circuits below.

Note that for this circuit $x = \tilde{x} - R \nabla f(\tilde{x})$.

E.2 Gradient flow

Let $f \colon \mathbf{R}^m \to \mathbf{R}$ be a convex function. Consider the circuit below.

35

Let x be the potentials at m pins of ∂f, and y be the current entering those pins. Applying KCL and the V-I relations of the capacitor we get

$$D_{\mathcal{C}}\frac{d}{dt}v_{\mathcal{C}} = i_{\mathcal{C}} = -y \in -\partial f(x).$$

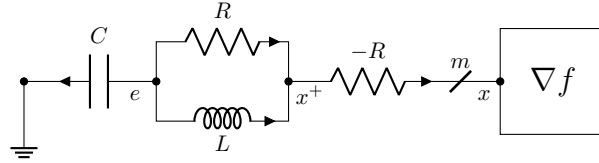
Since e is connected to ground, we have $v_{\mathcal{C}} = x - e = 0$. The resulting differential inclusion is

$$\frac{d}{dt}x \in -D_{\mathcal{C}}^{-1}\partial f(x).$$

For the automatic discretization of the continuous-time dynamics of this circuit, we define the energy at time k as

$$\mathcal{E}_k = \frac{1}{2}\|x^k - x^\star\|_{D_{\mathcal{C}}}^2.$$

E.3 Nesterov acceleration

Let $f\colon \mathbf{R}^m \to \mathbf{R}$ be a $1/R$-smooth convex function. Consider the circuit below.

Observe, by Ohm's law and $y = \nabla f(x)$ we have

$$x^+ = x - Ry = x - R\nabla f(x).$$

From KCL, KVL, and V-I relations we have

$$\frac{d}{dt}i_{\mathcal{L}} = D_{\mathcal{L}}^{-1}(v_{\mathcal{C}} - x^+)$$

$$\frac{d}{dt}v_{\mathcal{C}} = -D_{\mathcal{C}}^{-1}\nabla f(x).$$

Applying KCL and Ohm's law at x, it follows

$$\nabla f(x) = i_{\mathcal{L}} + \frac{1}{R}(v_{\mathcal{C}} - x^+),$$

which implies that

$$x = v_{\mathcal{C}} + Ri_{\mathcal{L}}.$$

Differentiating the above equality twice and plugging in the V-I relations, we get

$$\frac{d^2}{dt^2}x = -R(D_{\mathcal{L}}D_{\mathcal{C}})^{-1}\nabla f(x) - RD_{\mathcal{L}}^{-1}\frac{d}{dt}x - (D_{\mathcal{C}}^{-1} - R^2 D_{\mathcal{L}}^{-1})\frac{d}{dt}\nabla f(x).$$

Reorganizing, we conclude

$$\frac{d^2}{dt^2}x + RD_{\mathcal{L}}^{-1}\frac{d}{dt}x + (D_{\mathcal{C}}^{-1} - R^2 D_{\mathcal{L}}^{-1})\frac{d}{dt}\nabla f(x) + R(D_{\mathcal{L}}D_{\mathcal{C}})^{-1}\nabla f(x) = 0. \qquad (28)$$

Under the proper selection of parameters for μ-strongly convex and L-smooth function f, (28) corresponds to the high-resolution ODE for NAG-SC introduced in [136]

$$\frac{d^2}{dt^2}x + 2\sqrt{\mu}\frac{d}{dt}x + \sqrt{s}\frac{d}{dt}\nabla f(x) + (1 + \sqrt{\mu s})\nabla f(x) = 0.$$

As an immediate consequence, if we set $R = \frac{1}{4\mu}$, $L_i = \frac{1}{8\mu\sqrt{\mu}}$, $C_i = 2\sqrt{\mu}$, we recover the low-resolution ODE of NAG-SC

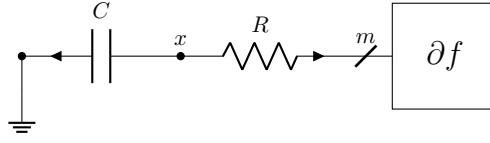
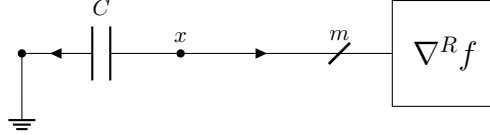
$$\frac{d^2}{dt^2}x + 2\sqrt{\mu}\frac{d}{dt}x + \nabla f(x) = 0.$$

For the automatic discretization of the continuous-time dynamics of this circuit, we define the energy at time k as

$$\mathcal{E}_k = \frac{1}{2}\|v_{\mathcal{C}}^k - x^\star\|_{D_{\mathcal{C}}}^2 + \frac{1}{2}\|i_{\mathcal{L}}^k - y^\star\|_{D_{\mathcal{L}}}^2.$$

36

E.4 Proximal point method

Consider the circuit below.

Then from the discussion in §E.1, the above circuit is equivalent to the circuit below.

According to §E.2, the ODE for the above circuit is

$$\frac{d}{dt}x = -D_C^{-1}\nabla^R f(x). \qquad (29)$$

Since from (27) we have

$$\mathbf{prox}_{Rf}(x) = x - R\nabla^R f(x),$$

this circuit gives a continuous model for the proximal point method.

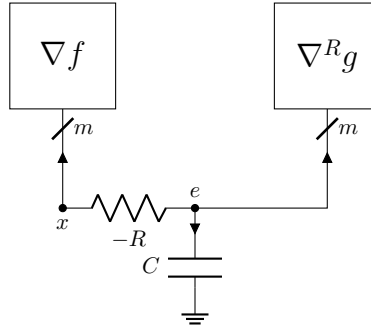
Applying Euler discretization to (29) with a stepsize of $C_i R$ for each ith coordinate, we recover proximal point method

$$x^{k+1} = x^k - R\nabla^R f(x^k) = x^k - (x^k - \mathbf{prox}_{Rf}(x^k)) = \mathbf{prox}_{Rf}(x^k).$$

For the automatic discretization of the continuous-time dynamics of this circuit, we define the energy at time k as

$$\mathcal{E}_k = \frac{1}{2}\|x^k - x^\star\|_{D_C}^2.$$

E.5 Proximal gradient method

Let $f\colon \mathbf{R}^m \to \mathbf{R}$ be $1/R$-smooth convex function, and $g\colon \mathbf{R}^m \to \mathbf{R}$ be convex function. Consider the circuit below.

Observe, by the Ohm's law $e = x - R\nabla f(x)$. Using KCL at x and KVL at e, we get

$$
\begin{aligned}
i_C &= -\nabla f(x) - \nabla^R g(e) \\
v_C &= x - R\nabla f(x).
\end{aligned}
$$

Applying V-I relation for the capacitor and eliminating e gives

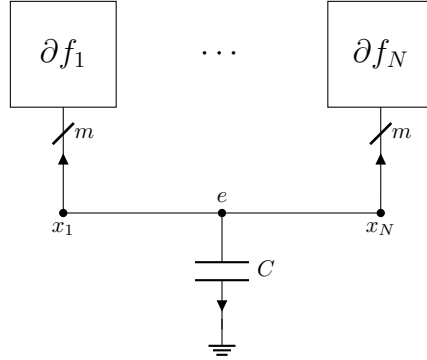
$$\frac{d}{dt}x - R\frac{d}{dt}\nabla f(x) = \frac{d}{dt}v_C = -\frac{1}{C}\left(\nabla f(x) + \nabla^R g(x - R\nabla f(x))\right).$$

37

Organizing
$$\frac{d}{dt}x = -\frac{1}{C}\left(\nabla^R g(I - R\nabla f) + \nabla f\right)(x) + R\frac{d}{dt}\nabla f(x).$$
We can show that $R\left\|\frac{d}{dt}\nabla f(x)\right\| < M$ for some $M > 0$, thus
$$\frac{d}{dt}x = -\frac{1}{CR}\left(\left(R\nabla^R g(I - R\nabla f) + R\nabla f\right)(x) + O\left(MCR\right)\right). \tag{30}$$
Applying Euler discretization with stepsize CR we have
$$\frac{x^{k+1} - x^k}{CR} = -\frac{1}{CR}\left(\left(R\nabla^R g(I - R\nabla f) + R\nabla f\right)(x^k) + O\left(MCR\right)\right).$$
Multiplying CR on both sides and reorganizing gives
$$\begin{aligned}
x^{k+1} &= x^k - \left(R\nabla^R g(I - R\nabla f) + R\nabla f\right)(x^k) + O(MCR) \\
&= \left(\mathbf{prox}_{Rg} - I\right)(I - R\nabla f)(x^k) + (I - R\nabla f)(x^k) + O(MCR) \\
&= \mathbf{prox}_{Rg}(I - R\nabla f)(x^k) + O(MCR).
\end{aligned}$$
If we set $C \ll R$, we recover the proximal gradient method.

For the automatic discretization of the continuous-time dynamics of this circuit, we define the energy at time k as
$$\mathcal{E}_k = \frac{C}{2}\|e^k - e^\star\|_2^2,$$
where $e^\star = x^\star - R\nabla f(x^\star)$.

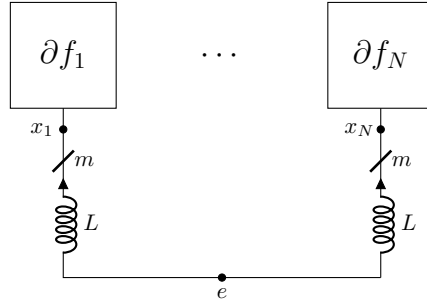
E.6   Primal decomposition

Let $f_1, \ldots, f_N \colon \mathbf{R}^m \to \mathbf{R}$ be convex functions. Consider the circuit below.

Let $x_1, \ldots, x_N \in \mathbf{R}^m$ be vectors of potentials at pins of $\partial f_1, \ldots, \partial f_N$ respectively. From KVL, we have
$$e = x_1 = \cdots = x_N = v_C.$$
Using KCL at e and the V-I relation of nonlinear resistors we get $\sum_{j=1}^N y_j + i_C = 0$, where $y_j \in \partial f_j(x_j)$. Using the V-I relation for capacitor we have
$$\frac{d}{dt}e = -\frac{1}{C}\sum_{j=1}^N y_j.$$
Discretizing above V-I relations, we recover primal decomposition
$$\begin{aligned}
y_j^k &\in \partial f_j(x_j^k) \\
e^{k+1} &= e^k - \frac{h}{C}\sum_{j=1}^N y_j^k.
\end{aligned}$$

For the automatic discretization of the continuous-time dynamics of this circuit, we define the energy at time k as
$$\mathcal{E}_k = \frac{C}{2}\|e^k - x^\star\|_2^2.$$

38

E.7 Dual decomposition

Let $f_1, \ldots, f_N \colon \mathbf{R}^m \to \mathbf{R}$ be convex functions. Consider the circuit below.

Using KCL at x_j and V-I relation for nonlinear resistors we get

$$x_j \in \partial f_j^*(y_j) = \partial f_j^*(i_{\mathcal{L}_j}).$$

Using KCL at e yields $\sum_{j=1}^N y_j = 0$. Using KVL and V-I relation for inductors we get

$$e - x_j = v_{\mathcal{L}_j} = L \frac{d}{dt} i_{\mathcal{L}_j}.$$

Summing over $j = 1, \ldots, N$ gives

$$Ne - \sum_{j=1}^N x_j = L \sum_{j=1}^N \frac{d}{dt} i_{\mathcal{L}_j} = L \frac{d}{dt} \sum_{j=1}^N y_j = 0,$$

leading to $e = (1/N) \sum_{j=1}^N x_j$. Discretizing above V-I relations, we recover dual decomposition

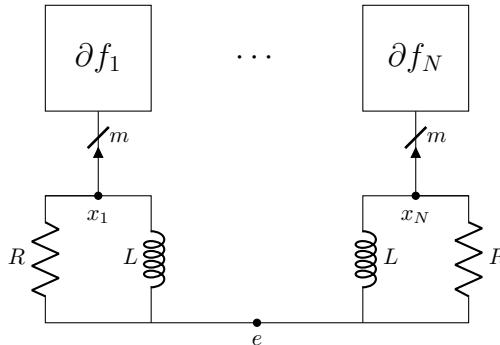
$$
\begin{aligned}
x_j^k &\in& \partial f_j^*(i_{\mathcal{L}_j}^k) \\
e^k &=& \frac{1}{N} \sum_{j=1}^N x_j^k \\
i_{\mathcal{L}_j}^{k+1} &=& i_{\mathcal{L}_j}^k - \frac{h}{L}(e^k - x_j^k).
\end{aligned}
$$

For the automatic discretization of the continuous-time dynamics of this circuit, we define the energy at time k as

$$\mathcal{E}_k = \sum_{j=1}^N \frac{L}{2} \|i_{\mathcal{L}_j}^k - y_j^\star\|_2^2.$$

E.8 Proximal decomposition

Let $f_1, \ldots, f_N \colon \mathbf{R}^m \to \mathbf{R}$ be convex functions. Consider the circuit below.

39

Let $x_1, \ldots, x_N \in \mathbf{R}^m$ be vectors of potentials at pins of $\partial f_1, \ldots, \partial f_N$ respectively. Define $e \in \mathbf{R}^m$ to be a vector of potentials on the bottom of the circuit. Observe, by Ohm's law and KCL we have

$$y_j = i_{\mathcal{L}_j} + \frac{1}{R}(e - x_j) \in \partial f_j(x_j).$$

It implies that $x_i = \mathbf{prox}_{Rf_i}(e + Ri_{\mathcal{L}_i})$. The V-I relations for inductors are given by

$$\frac{d}{dt}i_{\mathcal{L}} = v_{\mathcal{L}}/L = (E^\mathsf{T}e - x)/L,$$

where $E^\mathsf{T} = (I, \ldots, I) \in \mathbf{R}^{Nm \times m}$.

Further, note that by KCL $Ey = \sum_{j=1}^N y_j = 0$, therefore $E\frac{d}{dt}y = \frac{d}{dt}Ey = 0$. Using the above V-I relations for $g = Ne - Ex$ we get the following ODE

$$\dot{g} = \frac{d}{dt}RE(y - i_{\mathcal{L}}) = -RE\frac{d}{dt}i_{\mathcal{L}} = -\frac{R}{L}E(E^\mathsf{T}e - x) = -\frac{R}{L}g. \tag{31}$$

We initialize circuit with $Ei_{\mathcal{L}}(0) = 0$ and $EE^\mathsf{T} = NI$ gives

$$0 = Ey(0) = E(i_{\mathcal{L}}(0) + (E^\mathsf{T}e(0) - x(0))/R) = -\frac{1}{R}g(0).$$

Thus the solution to an ODE (31) is $g = 0$ and we conclude $e = \frac{1}{N}Ex$.

The V-I relations for the circuit are

$$\begin{aligned}
x_j &= \mathbf{prox}_{Rf_j}(e + Ri_{\mathcal{L}_j}), \quad j = 1, \ldots, N \\
e &= \frac{1}{N}Ex \\
\frac{d}{dt}i_{\mathcal{L}} &= (E^\mathsf{T}e - x)/L.
\end{aligned}$$

Discretizing above V-I relations we recover proximal decomposition

$$\begin{aligned}
x_j^{k+1} &= \mathbf{prox}_{Rf_j}(e^k + Ri_{\mathcal{L}_j}^k), \quad j = 1, \ldots, N \\
e^{k+1} &= \frac{1}{N}Ex^k \\
i_{\mathcal{L}}^{k+1} &= i_{\mathcal{L}}^k + \frac{h}{L}(E^\mathsf{T}e^{k+1} - x^{k+1}).
\end{aligned}$$

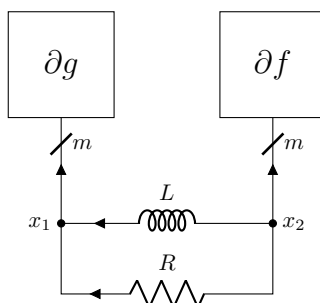
For the automatic discretization of the continuous-time dynamics of this circuit, we define the energy at time k as

$$\mathcal{E}_k = \sum_{j=1}^N \frac{L}{2}\|i_{\mathcal{L}_j}^k - y_j^\star\|_2^2 + \gamma\|e^k - x^\star\|_2^2,$$

where γ is a parameter that is being optimized, see §G.

E.9 Douglas–Rachford splitting

Let $f, g \colon \mathbf{R}^m \to \mathbf{R}$ be convex functions. Consider the circuit below.

40

Using KCL at x_1 and Ohm's law we get

$$\frac{1}{R}(x_2 - x_1) + i_{\mathcal{L}} \in \partial g(x_1),$$

which implies $x_1 = \mathbf{prox}_{Rg}(x_2 + Ri_{\mathcal{L}})$. Similarly, using KCL at x_2 and Ohm's law we get

$$\frac{1}{R}(x_1 - x_2) - i_{\mathcal{L}} \in \partial f(x_2),$$

which implies $x_2 = \mathbf{prox}_{Rf}(x_1 - Ri_{\mathcal{L}})$. From KVL and V-I relation for inductors we have

$$\frac{d}{dt}i_{\mathcal{L}} = \frac{1}{L}(x_2 - x_1).$$

Discretizing above V-I relations with $R = L = 1$ and stepsize $h = 1$, we recover Douglas–Rachford splitting

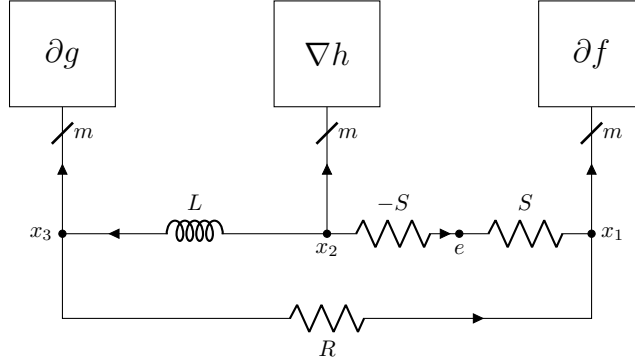
$$
\begin{aligned}
x_1^{k+1} &= \mathbf{prox}_{Rg}(x_2^k + Ri_{\mathcal{L}}^k) \\
x_2^{k+1} &= \mathbf{prox}_{Rf}(x_1^{k+1} - Ri_{\mathcal{L}}^k) \\
i_{\mathcal{L}}^{k+1} &= i_{\mathcal{L}}^k + \frac{h}{L}(x_2^{k+1} - x_1^{k+1}).
\end{aligned}
$$

For the automatic discretization of the continuous-time dynamics of this circuit, we define the energy at time k as

$$\mathcal{E}_k = \frac{L}{2}\|i_{\mathcal{L}}^k - y_1^\star\|_2^2 + \gamma\|x_2^k - x^\star\|_2^2,$$

where γ is a parameter that is being optimized, see §G, and $y_1^\star \in \partial g(x^\star)$.

E.10 Davis-Yin splitting

Let $f, g, h\colon \mathbf{R}^m \to \mathbf{R}$ be convex functions, with h also being $1/S$-smooth. Consider the circuit below.

Using KCL at e and Ohm's law we have

$$x_1 = x_2 - Si_{-S} + Si_S = x_2.$$

Applying KCL at x_2, we get

$$i_{\mathcal{L}} = \frac{e - x_2}{-S} - \nabla h(x_2). \tag{32}$$

Using KCL at x_3 and (32) it follows

$$\frac{x_1 - x_3}{R} + i_{\mathcal{L}} = \frac{x_1 - x_3}{R} + \frac{x_2 - e}{S} - \nabla h(x_2) \in \partial g(x_3).$$

Using KCL at x_1, we get

$$\frac{x_3 - x_1}{R} + \frac{e - x_1}{S} \in \partial f(x_1).$$

41

Organizing, and applying V-I relation for inductor

$$x_3 = \mathbf{prox}_{Rg}\left(\left(1+\frac{R}{S}\right)x_1 - \frac{R}{S}e - R\nabla h(x_1)\right)$$

$$x_1 = \mathbf{prox}_{Rf}\left(x_3 + \frac{R}{S}(e - x_1)\right)$$

$$i_{\mathcal{L}} = \frac{e - x_1}{-S} - \nabla h(x_1)$$

$$\frac{d}{dt}i_{\mathcal{L}} = \frac{1}{L}(x_1 - x_3).$$

Now we eliminate the term $i_{\mathcal{L}}$. Differentiating (32), applying $L\frac{d}{dt}i_{\mathcal{L}} = x_1 - x_3$ we get

$$\frac{1}{L}(x_1 - x_3) = \frac{d}{dt}i_{\mathcal{L}} = \frac{d}{dt}\frac{x_1 - e}{S} - \frac{d}{dt}\nabla h(x_1).$$

In other words,

$$\frac{d}{dt}e = \frac{S}{L}(x_3 - x_1) + \frac{d}{dt}x_1 - S\frac{d}{dt}\nabla h(x_1).$$

Using "alternating update" and Euler discretization of e and x_1, we have

$$x_3^{k+1} = \mathbf{prox}_{Rg}\left(\left(1+\frac{R}{S}\right)x_1^k - \frac{R}{S}e^k - R\nabla h(x_1^k)\right)$$

$$x_1^{k+1} = \mathbf{prox}_{Rf}\left(x_3^{k+1} + \frac{R}{S}(e^k - x_1^k)\right)$$

$$e^{k+1} = e^k + \frac{Sh}{L}(x_3^{k+1} - x_1^{k+1}) + x_1^{k+1} - x_1^k - Sh\frac{d}{dt}\nabla h(x_1^k).$$

Set $R = S = h = \alpha$ and $L = \alpha^2$, then the above can be rewritten as

$$x_3^{k+1} = \mathbf{prox}_{\alpha g}\left(2x_1^k - e^k - \alpha\nabla h(x_1^k)\right)$$

$$x_1^{k+1} = \mathbf{prox}_{\alpha f}\left(e^k + x_3^{k+1} - x_1^k\right)$$

$$e^{k+1} = e^k + x_3^{k+1} - x_1^k - \alpha^2\frac{d}{dt}\nabla h(x_1^k).$$

When $\left\|\frac{d}{dt}\nabla h(x^k)\right\|$ is bounded, then $\alpha^2\frac{d}{dt}\nabla h(x^k) = O(\alpha^2)$. For small α we may ignore this term and recover DYS.

For the automatic discretization of the continuous-time dynamics of this circuit, we define the energy at time k as

$$\mathcal{E}_k = \frac{L}{2}\|i_{\mathcal{L}}^k - y_1^\star\|_2^2 + \gamma\|e_1^k - e^\star\|_2^2,$$

where γ is a parameter that is being optimized, see §G, and $e^\star = x^\star - R(y_1^\star + y_3^\star)$, $y_1^\star \in \partial f(x^\star)$, $y_3^\star \in \partial g(x^\star)$.

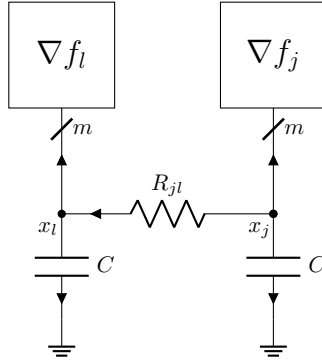
42

F   Decentralized classical algorithms

In a decentralized optimization setup, we are given a graph $G = (V, A)$ which defines the communication pattern between agents. This means that each agent is constrained to communicate only to its direct neighbors for the edges of the graph.

We define Γ_j as the neighbors of j in graph G. For simplicity, in each example we only illustrate the circuit between components indexed by j and l, where j and l are connected through an edge in the graph G.

F.1   Decentralized gradient descent

Let $f_1, \ldots, f_N \colon \mathbf{R}^m \to \mathbf{R}$ be differentiable convex functions. Decentralized gradient descent (DGD) is derived as a gradient descent of the appropriate penalty formulation of a decentralized problem. Similarly, to construct a DGD circuit we apply the gradient flow circuit of §E.2 to appropriate nonlinear resistors and arrive at the following circuit.

The right side of the circuit contains the graph with resistors R_{jl} connecting vectors of potentials $x_j \in \mathbf{R}^m$ and $x_l \in \mathbf{R}^m$ for every neighbors j and l in the given graph G.

Using the KCL at x_j we get

$$0 = \nabla f_j(x_j) + \sum_{l \in \Gamma_j} \frac{x_j - x_l}{R_{jl}} + i_{\mathcal{C}_j}.$$

Applying the V-I relation for the capacitors we get

$$\frac{d}{dt}x_j = \frac{d}{dt}v_{\mathcal{C}_j} = -\frac{1}{C}\left(\nabla f_j(x_j) + \sum_{l \in \Gamma_j}(x_j - x_l)/R_{jl}\right).$$

Euler discretization recovers the DGD

$$x_j^{k+1} = \left(1 - \sum_{l \in \Gamma_j} \frac{h}{CR_{jl}}\right)x_j^k + \sum_{l \in \Gamma_j} \frac{h}{CR_{jl}}x_l^k - \frac{h}{C}\nabla f_j(x_j^k),$$

with gradient stepsize h/C and the mixing matrix

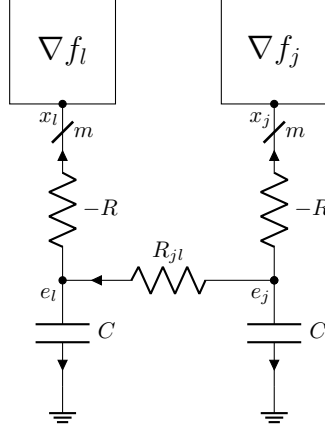
$$W_{jl} = \begin{cases} 1 - \sum_{l \in \Gamma_j} \frac{h}{CR_{jl}} & \text{if } j = l \\ \frac{h}{CR_{jl}} & \text{if } j \neq l, \quad l \in \Gamma_j \\ 0 & \text{otherwise.} \end{cases}$$

For the automatic discretization of the continuous-time dynamics of this circuit, we define the energy at time k as

$$\mathcal{E}_k = \sum_{j=1}^{N} \frac{C}{2}\|x_j^k - x_j^\star\|_2^2.$$

43

F.2 Diffusion

Let $f_1, \ldots, f_N \colon \mathbf{R}^m \to \mathbf{R}$ be $1/R$-smooth convex functions. Decentralized gradient descent is derived as a forward-backward splitting fixed point iteration of the appropriate penalty formulation of a decentralized problem. Similarly, to construct a diffusion circuit we apply the proximal gradient circuit of §E.5 to appropriate nonlinear resistors and arrive at the following circuit.

The right side of the circuit is the graph with linear resistors R_{jl} connecting vectors of potentials $e_j \in \mathbf{R}^m$ and $e_l \in \mathbf{R}^m$ for every neighbors j and l in the given graph G.

By Ohm's law we have $v_{\mathcal{C}_j} = e_j = x_j - R\nabla f_j(x_j)$. Using the KCL at e_j we get

$$\nabla f_j(x_j) + \sum_{l \in \Gamma_j} \frac{e_j - e_l}{R_{jl}} + i_{\mathcal{C}_j} = 0.$$

Applying the V-I relation for the capacitors we get

$$\begin{aligned}
\frac{d}{dt} e_j &= \frac{d}{dt} x_j - R \frac{d}{dt} \nabla f_j(x_j) \\
&= -\frac{1}{C} \left(\nabla f_j(x_j) + \sum_{l \in \Gamma_j} \frac{(x_j - R\nabla f_j(x_j)) - (x_l - R\nabla f_l(x_l))}{R_{jl}} \right).
\end{aligned}$$

We can show that $R \left\| \frac{d}{dt} \nabla f_j(x_j) \right\| < M$ for some $M > 0$, thus

$$\frac{d}{dt} x_j = -\frac{1}{C} \left(\nabla f_j(x_j) + \sum_{l \in \Gamma_j} \frac{(x_j - R\nabla f_j(x_j)) - (x_l - R\nabla f_l(x_l))}{R_{jl}} + O(MC) \right).$$

Applying Euler discretization with stepsize CR gives

$$\begin{aligned}
x_j^{k+1} &= \left(1 - \sum_{l \in \Gamma_j} \frac{R}{R_{jl}} \right) (x_j^k - R\nabla f_j(x_j^k)) \\
&\quad + \sum_{l \in \Gamma_j} \frac{R}{R_{jl}} (x_l^k - R\nabla f_l(x_l^k)) + O(MCR),
\end{aligned}$$

with gradient stepsize R and the mixing matrix

$$W_{jl} = \begin{cases} 1 - \sum_{l \in \Gamma_j} \frac{R}{R_{jl}} & \text{if } j = l \\ \frac{R}{R_{jl}} & \text{if } j \neq l, \quad l \in \Gamma_j \\ 0 & \text{otherwise.} \end{cases}$$

If we set $C \ll R$, we recover the diffusion method.

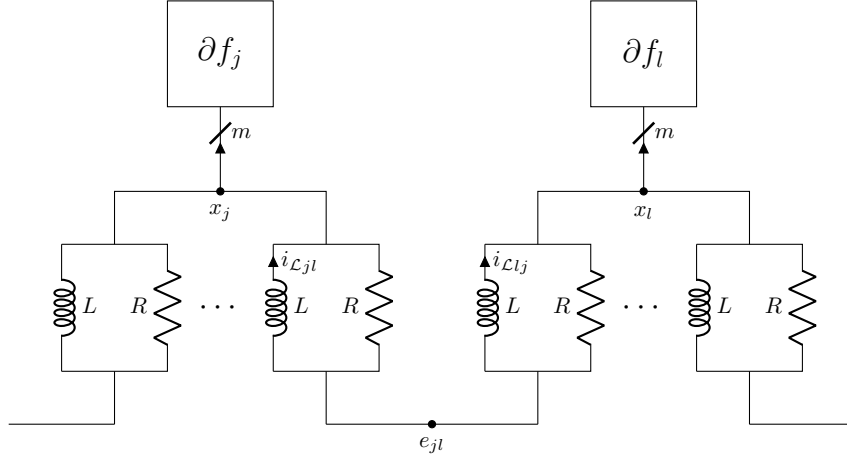
44

For the automatic discretization of the continuous-time dynamics of this circuit, we define the energy at time k as

$$\mathcal{E}_k = \sum_{j=1}^{N} \frac{C}{2} \|e_j^k - e_j^\star\|_2^2,$$

where $e_j^\star = x_j^\star - R\nabla f_j(x_j^\star)$ for all $j = 1, \ldots, N$.

F.3   Decentralized ADMM

Let $f_1, \ldots, f_N \colon \mathbf{R}^m \to \mathbf{R}$ be convex functions. Then the decentralized ADMM circuit is given below.

Note that this circuit is similar to the one in proximal decomposition in §E.8 with the difference that instead of a single net e we now have a net e_{jl} for each edge (j, l) in graph G. Denote currents on inductors to be $i_{\mathcal{L}jl} \in \mathbf{R}^m$ and $i_{\mathcal{L}lj} \in \mathbf{R}^m$.

Using KCL at x_j we get

$$\sum_{l \in \Gamma_j} \left(i_{\mathcal{L}jl} + \frac{e_{jl} - x_j}{R} \right) \in \partial f_j(x_j). \tag{33}$$

We initialize the circuit such that $i_{\mathcal{L}jl}(0) + i_{\mathcal{L}lj}(0) = 0$ for each edge (j, l) in graph G. Now consider KCL at e_{jl}

$$i_{\mathcal{L}jl} + i_{\mathcal{L}lj} = -\frac{(e_{jl} - x_j)}{R} - \frac{(e_{jl} - x_l)}{R}. \tag{34}$$

Using V-I relation for inductor we also have

$$\frac{d}{dt} i_{\mathcal{L}jl} + \frac{d}{dt} i_{\mathcal{L}lj} = \frac{1}{L}(2e_{jl} - x_j - x_l).$$

Combining the two equalities above we get an ODE

$$\frac{d}{dt} \left(i_{\mathcal{L}jl} + i_{\mathcal{L}lj} \right) = -\frac{R}{L} \left(i_{\mathcal{L}jl} + i_{\mathcal{L}lj} \right).$$

Using initial conditions the solution of an ODE is $i_{\mathcal{L}jl} + i_{\mathcal{L}lj} = 0$. From (34) we conclude that $e_{jl} = \frac{1}{2}(x_j + x_l)$.

Using (33), we get the V-I relations for the circuit

$$
\begin{aligned}
x_j &= \mathbf{prox}_{(R/|\Gamma_j|)f_j} \left(\frac{1}{|\Gamma_j|} \sum_{l \in \Gamma_j} (Ri_{\mathcal{L}jl} + e_{jl}) \right) \\
e_{jl} &= \frac{1}{2}(x_j + x_l) \\
\frac{d}{dt} i_{\mathcal{L}jl} &= \frac{1}{L}(e_{jl} - x_j),
\end{aligned}
$$

45

for every $j = 1, \ldots, N$ and every edge (j, l) in graph G. Discretizing the V-I relations with stepsize L/R, we recover decentralized ADMM,

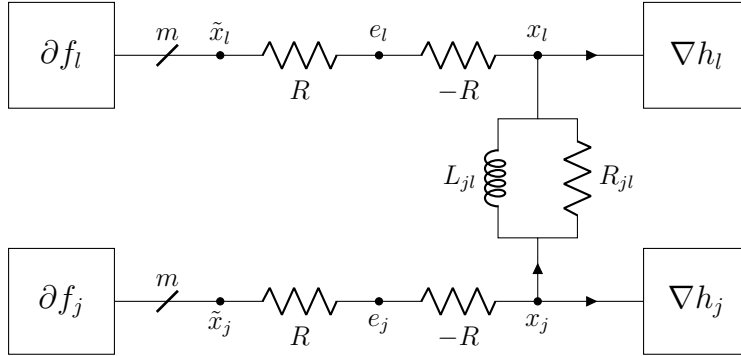
$$
\begin{aligned}
x_j^{k+1} &= \mathbf{prox}_{(R/|\Gamma_j|)f_j} \left(\frac{1}{|\Gamma_j|} \sum_{l \in \Gamma_j} (R i_{\mathcal{L}jl}^k + e_{jl}^k) \right) \\
e_{jl}^{k+1} &= \frac{1}{2}(x_j^{k+1} + x_l^{k+1}) \\
i_{\mathcal{L}jl}^{k+1} &= i_{\mathcal{L}jl}^k + \frac{1}{R}(e_{jl}^{k+1} - x_j^{k+1}).
\end{aligned}
$$

For the automatic discretization of the continuous-time dynamics of this circuit, we define the energy at time k as

$$
\mathcal{E}_k = \sum_{\text{edge } \{j,l\}} \left(\frac{L}{2} \|i_{\mathcal{L}jl}^k - i_{\mathcal{L}jl}^\star\|_2^2 + \frac{L}{2} \|i_{\mathcal{L}lj}^k - i_{\mathcal{L}lj}^\star\|_2^2 + \gamma \|e_{jl}^k - x^\star\|_2^2 \right),
$$

where γ is a parameter that is being optimized, see §G, and $i_{\mathcal{L}jl}^\star$ is the current through inductor at equilibrium.

F.4 PG-EXTRA

Let $f_1, \ldots, f_N \colon \mathbf{R}^m \to \mathbf{R}$ be convex functions, and $h_1, \ldots, h_N \colon \mathbf{R}^m \to \mathbf{R}$ be convex M-smooth functions. Then the PG-EXTRA circuit is given below. Denote current on inductor going from x_j to

x_l to be $i_{\mathcal{L}jl} \in \mathbf{R}^m$.

Recall §E.1 and apply Ohm's law to get

$$
\frac{e_j - \tilde{x}_j}{R} = \frac{e_j - \mathbf{prox}_{Rf_j}(e_j)}{R} = \nabla^R f_j(e_j) = \frac{x_j - e_j}{-R}.
$$

This yields $x_j = e_j - R \nabla^R f_j(e_j) = \mathbf{prox}_{Rf_j}(e_j)$. Using KCL at x_j we get

$$
\frac{e_j - x_j}{-R} = \nabla h_j(x_j) + \sum_{l \in \Gamma_j} \left(i_{\mathcal{L}jl} + \frac{x_j - x_l}{R_{jl}} \right). \tag{35}
$$

Define the mixing matrix

$$
W_{jl} = \begin{cases} 1 - \sum_{l \in \Gamma_j} \frac{R}{R_{jl}} & \text{if } j = l \\ \frac{R}{R_{jl}} & \text{if } j \neq l, \quad l \in \Gamma_j \\ 0 & \text{otherwise.} \end{cases}
$$

46

Rearranging the terms in (35) we get

$$\begin{aligned}
e_j &= x_j - R\nabla h_j(x_j) - \sum_{l\in\Gamma_j}\left(Ri_{\mathcal{L}jl} + \frac{x_j - x_l}{R_{jl}/R}\right) \\
&= x_j - \sum_{l=1}^{N} W_{jl}(x_j - x_l) - R\nabla h_j(x_j) - \sum_{l\in\Gamma_j} Ri_{\mathcal{L}jl} \\
&= \sum_{l=1}^{N} W_{jl}x_l - R\nabla h_j(x_j) - \sum_{l\in\Gamma_j} Ri_{\mathcal{L}jl}.
\end{aligned}$$

Using the V-I relation for inductor we also have

$$\frac{d}{dt}i_{\mathcal{L}jl} = \frac{1}{L_{jl}}(x_j - x_l).$$

Set $L_{jl} = R_{jl}$ for every edge (j, l) in graph G. Define $w_j = \sum_{l\in\Gamma_j} Ri_{\mathcal{L}jl}$, then

$$\begin{aligned}
\frac{d}{dt}w_j &= \sum_{l\in\Gamma_j}\frac{R}{L_{jl}}(x_j - x_l) \\
&= x_j - \sum_{l=1}^{N} W_{jl}x_l.
\end{aligned}$$

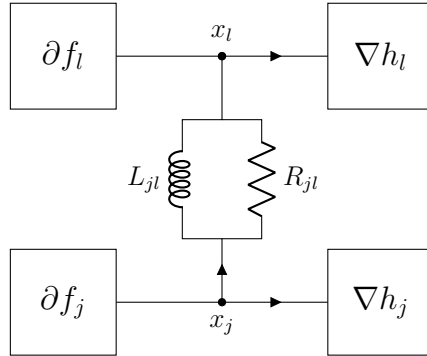
Combining the above, we get the V-I relations for the circuit

$$x_j = \mathbf{prox}_{Rf_j}\left(\sum_{l=1}^{N} W_{jl}x_l - R\nabla h_j(x_j) - w_j\right) \tag{36}$$

$$\frac{d}{dt}w_j = x_j - \sum_{l=1}^{N} W_{jl}x_l,$$

for every $j = 1, \ldots, N$ and every edge (j, l) in graph G. Discretizing the above V-I relations with stepsize $1/2$, and following the decentralized notation of [129, §11.3], we recover PG-EXTRA,

$$\begin{aligned}
x^{k+1} &= \mathbf{prox}_{Rf}\left(Wx^k - R\nabla h(x^k) - w^k\right) \tag{37} \\
w^{k+1} &= w^k + \frac{1}{2}(I - W)x^k.
\end{aligned}$$

We can simplify the circuit by eliminating potentials e_j as shown below.

For the automatic discretization of the continuous-time dynamics of this circuit, we define the energy at time k as

$$\mathcal{E}_k = \sum_{\text{edge }\{j,l\}}\frac{L_{jl}}{2}\|i_{\mathcal{L}jl}^{k+1} - i_{\mathcal{L}jl}^\star\|_2^2 + \sum_{j=1}^{N}\gamma\|x_j^k - x^\star\|_2^2,$$

where γ is a parameter that is being optimized, see §G, and $i_{\mathcal{L}jl}^\star$ is the current through inductor at equilibrium.

47

G   Automatic discretization

In this paper, we discretize admissible dynamic interconnects corresponding to the following decentralizes setup with graph consensus

$$
\begin{array}{ll}
\underset{x_1,\ldots,x_N \in \mathbf{R}^{m/N}}{\text{minimize}} & f_1(x_1) + \cdots + f_N(x_N) \\
\text{subject to} & x_j = x_l, \quad j = 1,\ldots,N, \quad l \in \Gamma_j,
\end{array}
\tag{38}
$$

where Γ_j contains the neighbors of agent j in the communication graph, see §F. We assume that the communication graph is connected, ensuring that all agents can communicate with each other [129, §11.2]. The static interconnect for this problem corresponds to the consensus problem

$$
\begin{array}{ll}
\underset{x_1,\ldots,x_N \in \mathbf{R}^{m/N}}{\text{minimize}} & f_1(x_1) + \cdots + f_N(x_N) \\
\text{subject to} & x_1 = \cdots = x_N,
\end{array}
$$

which is a special case of (1) where $E^{\mathsf{T}} = (I,\ldots,I) \in \mathbf{R}^{m \times m/N}$. Therefore, we have $n = m/N$ nets each of size N with $x_j \in \mathbf{R}^n$ for all $j = 1,\ldots,N$. This setup generalizes the setup of the classical methods discussed in §E and §F.

For automatic discretization, we focus on dynamic interconnects that have the same RLC circuit across each net, *i.e.*, the dynamic interconnects represented with the multi-wire notation.

Runge–Kutta method.   The capacitor and inductor ODEs are of the form

$$
\frac{d}{dt}x(t) = F(x(t)).
$$

We discretize ODEs using the two-stage Runge–Kutta method, with coefficients α, β, and stepsize h:

$$
\begin{aligned}
x^{k+1/2} &= x^k + \alpha h F(x^k) \\
x^{k+1} &= x^k + \beta h F(x^k) + (1-\beta)h F(x^{k+1/2}).
\end{aligned}
$$

We clarify that simpler one-stage discretization schemes can also be used. We chose two-stage Runge–Kutta to demonstrate that multi-stage discretization schemes are compatible with our automatic discretization methodology.

Energy descent.   Let a discrete-time optimization algorithm generate a sequence $\{(v^k, i^k, x^k, y^k)\}_{k=1}^{\infty}$ with $v^k, i^k \in \mathbf{R}^{\sigma}$ (voltages across and currents through the branches of interconnect) and $x^k, y^k \in \mathbf{R}^m$ (potentials at terminals and currents leaving terminals). Let the subscripts $\mathcal{R}$, $\mathcal{L}$, and $\mathcal{C}$ denote the components related to resistors, inductors, and capacitors, respectively. Then the energy stored in the circuit is given by

$$
\mathcal{E}_k = \frac{1}{2}\|v_{\mathcal{C}}^k - v_{\mathcal{C}}^{\star}\|_{D_{\mathcal{C}}}^2 + \frac{1}{2}\|i_{\mathcal{L}}^k - i_{\mathcal{L}}^{\star}\|_{D_{\mathcal{L}}}^2.
$$

Lemma G.1. *Assume $f \colon \mathbf{R}^m \to \mathbf{R} \cup \{\infty\}$ is a strictly convex function and the dynamic interconnect is admissible. Let a discrete-time optimization algorithm generate a sequence $\{(v^k, i^k, x^k, y^k)\}_{k=1}^{\infty}$. If there exists $\eta > 0$ such that for all $k = 1, 2, \ldots$ the energy descent*

$$
D_k = \big(\mathcal{E}_{k+1} + \eta\langle x^k - x^{\star}, y^k - y^{\star}\rangle\big) - \mathcal{E}_k \leq 0
\tag{39}
$$

holds, then x^k converges to a primal solution.

Proof. Suppose there exists $\eta > 0$ for which (39) holds. Then we have

$$
\begin{aligned}
0 &\leq \mathcal{E}_{K+1} \\
&\leq \mathcal{E}_K - \eta\langle x^K - x^{\star}, y^K - y^{\star}\rangle \\
&\leq \mathcal{E}_0 - \sum_{k=0}^{K} \eta\langle x^k - x^{\star}, y^k - y^{\star}\rangle.
\end{aligned}
$$

48

By monotonicity of subdifferential operator ∂f, we have

$$\langle x^k - x^\star, y^k - y^\star \rangle \geq 0, \quad y^k \in \partial f(x^k).$$

Thus rearranging the terms we get

$$0 \leq \sum_{k=0}^{K} \eta \langle x^k - x^\star, y^k - y^\star \rangle \leq \mathcal{E}_0.$$

Sending K to infinity, by the summability argument it follows that

$$\langle x^k - x^\star, y^k - y^\star \rangle \to 0. \tag{40}$$

Define the Lagrangian function

$$L(x, z, y) = f(x) - y^T(x - E^\mathsf{T} z).$$

Since $x^\star \in \mathcal{R}(E^\mathsf{T})$, there exists some $z^\star$ such that $x^\star = E^\mathsf{T} z^\star$. Then for fixed $z^\star$ and $y^\star$, function $L(x, z^\star, y^\star)$ is strictly convex. Its subgradient is given by $(y - y^\star) \in \partial_x L(x, z^\star, y^\star)$ for $y \in \partial f(x)$, therefore, $0 \in L(x^\star, z^\star, y^\star)$. Together with strict convexity this implies that $L(x, z^\star, y^\star)$ achieves a unique global minimum at $x^\star$ with $L(x^\star, z^\star, y^\star) = f(x^\star)$. By the subgradient inequality, we have

$$\langle x^k - x^\star, y^k - y^\star \rangle \geq L(x^k, z^\star, y^\star) - f(x^\star) \geq 0.$$

Then the condition (40) implies $L(x^k, z^\star, y^\star) \to f(x^\star)$. Therefore, $x^k \to x^\star$ which concludes the proof. $\qquad\square$

By the descent lemma G.1, the discretization is dissipative if there exist value $\eta > 0$ such that

$$D_k = \left(\mathcal{E}_{k+1} + \eta \langle x^k - x^\star, y^k - y^\star \rangle \right) - \mathcal{E}_k \leq 0$$

for all $k = 1, 2, \ldots$ Since the descent D_k is defined using a one-step transition, without loss of generality, it suffices to consider $k = 1$.

Solver dissipative term. To provide more flexibility with the Ipopt [155, 9] solver, we also incorporate dissipation from the linear resistors as in the continuous-time energy dissipation (7), *i.e.*, we try to establish

$$D_k = \left(\mathcal{E}_{k+1} + \eta \langle x^k - x^\star, y^k - y^\star \rangle + \rho R \|i_\mathcal{R}^k\|_2^2 \right) - \mathcal{E}_k \tag{41}$$

with $\eta > 0$ and $\rho \geq 0$. Also see §D. If there exist values $\eta > 0$ and $\rho \geq 0$ such that $D_k \leq 0$ holds, then the discretization is sufficiently dissipative and Lemma G.1 applies.

G.1 Dissipative discretization

In this section, we fix α, β, h, η, and ρ and describe a convex optimization problem that checks whether the discretization is dissipative. We focus on problem (38).

Worst-case optimization problem. To verify if the dissipativity condition $D_k \leq 0$ (41) is satisfied for a given discretization, we can alternatively solve a worst-case problem. Specifically, this entails determining if the optimal value of the following optimization problem is non-positive:

$$
\begin{aligned}
\text{maximize} \quad & \mathcal{E}_2 - \mathcal{E}_1 + \eta \langle x^1 - x^\star, y^1 - y^\star \rangle + \rho R \|i_\mathcal{R}^1\|_2^2 \\
\text{subject to} \quad & \mathcal{E}_s = \tfrac{1}{2} \|v_\mathcal{C}^s - v_\mathcal{C}^\star\|_{D_\mathcal{C}}^2 + \tfrac{1}{2} \|i_\mathcal{L}^s - i_\mathcal{L}^\star\|_{D_\mathcal{L}}^2, \quad s \in \{1, 2\} \\
& (v^1, i^1, x^1, y^1) \text{ is feasible initial point} \\
& (v^2, i^2, x^2, y^2) \text{ is generated by discrete optimization method from initial point} \\
& f \in \mathcal{F},
\end{aligned}
\tag{42}
$$

where $f, v^k, i^k, x^k, y^k, v^\star, i^\star, x^\star, y^\star$ are the decision variables and $\mathcal{F}$ is a family of functions (*e.g.*, L-smooth convex) that the algorithm is to be applied to.

49

Reformulated worst-case optimization problem. Recall that we assume that the RLC circuit across each net is the same. Thus we can define an operator $\text{mat}(z)$ that reshapes vector $z \in \mathbf{R}^\sigma$ into a matrix of size $n \times \sigma/n$, where each row contains information (voltage or current) of the electric components that belong to the same net. Define index sets

$$
\begin{aligned}
I_K &= \{1, 1.5, 2, \star\}, \\
I_N &= \{1, \ldots, N\}, \\
I_K \times I_N &= \{(k, l) \mid l \in I_N, k \in I_K\},
\end{aligned}
$$

and matrices

$$
\begin{aligned}
H &= \begin{bmatrix} \text{mat}(v^1) & \text{mat}(i^1) & \left[y_l^k\right]_{(k,l) \in I_K \times I_N} \end{bmatrix} \in \mathbf{R}^{n \times (2\sigma/n + |I_K|N)}, \\
G &= H^T H \in \mathbf{S}_+^{2\sigma/n + |I_K|N}, \\
F &= \left[f_l^k\right]_{(k,l) \in I_K \times I_N} \in \mathbf{R}^{|I_K|N},
\end{aligned}
$$

where $y_l^k \in \partial f_l(x^k)$ and $f_l^k = f_l(x^k)$ for all $l \in I_N$. Note that we have

$$
|F| = |I_K|N, \qquad |G| = (2\sigma/n + |I_K|N)^2,
$$

and for $|I_K| = 4$ this simplifies to

$$
|F| = 4N, \qquad |G| = (2\sigma/n + 4N)^2.
$$

Recall the circuit ODEs are discretized with the two-stage Runge–Kutta method, leading to the variables $\text{mat}(v^k)$, $\text{mat}(i^k)$, x^k, y^k that are linear combinations of columns in H. The coefficients of these linear combinations are polynomials in α, β, and h. In other words, there exist matrices $\mathbf{v}^k, \mathbf{i}^k, \mathbf{x}^k, \mathbf{y}_l^k$ such that

$$
\text{mat}(v^k) = H\mathbf{v}^k, \quad \text{mat}(i^k) = H\mathbf{i}^k, \quad x^k = H\mathbf{x}^k, \quad y_l^k = H\mathbf{y}_l^k
$$

for all $k \in I_K$, $l \in I_N$. Similarly, we can find $\mathbf{f}_l^k$ such that $f_l^k = F\mathbf{f}_l^k$.

For fixed parameters α, β, h, η, and ρ, the problem (42) can be reformulated as

$$
\begin{aligned}
\underset{f_1, \ldots, f_N, H}{\text{maximize}} \quad & \mathcal{E}_2 - \mathcal{E}_1 + \eta\langle x^1 - x^\star, y^1 - y^\star\rangle + \rho R\|i_\mathcal{R}^1\|_2^2 \\
\text{subject to} \quad & \mathcal{E}_s = \tfrac{1}{2}\|v_\mathcal{C}^s - v_\mathcal{C}^\star\|_{D_\mathcal{C}}^2 + \tfrac{1}{2}\|i_\mathcal{L}^s - i_\mathcal{L}^\star\|_{D_\mathcal{L}}^2, \quad s \in \{1, 2\} \\
& \text{mat}(v^k) = H\mathbf{v}^k, \quad k \in I_K \\
& \text{mat}(i^k) = H\mathbf{i}^k, \quad k \in I_K \\
& x^k = H\mathbf{x}^k, \quad k \in I_K \\
& y_l^k = H\mathbf{y}_l^k, \quad k \in I_K, \quad l \in I_N \\
& f_l \in \mathcal{F}_{\mu_l, M_l}(\mathbf{R}^n), \quad l \in I_N.
\end{aligned}
$$

By the interpolation lemma ([149], Theorem 2), $f_l \in \mathcal{F}_{\mu_l, M_l}(\mathbf{R}^n)$ if and only if

$$
\begin{aligned}
0 \geq \ & f_l^j - f_l^i + \langle g_l^j, x^i - x^j\rangle + \frac{1}{2M_l}\|g_l^i - g_l^j\|_2^2 \\
& + \frac{\mu_l}{2(1 - \mu_l/M_l)}\|x^i - x^j - 1/M_l(g_l^i - g_l^j)\|_2^2, \quad i, j \in I_K.
\end{aligned}
$$

Therefore, we can replace infinite dimensional decision variable $f_l \in \mathcal{F}_{\mu_l, M_l}(\mathbf{R}^n)$ with $|I_K|(|I_K|-1)$ inequalities.

Grammian formulation. Now using Grammian formulation, the problem of finding the worst-case energy difference over a given family of functions reduces to solving an SDP, similar to [150]. This SDP can be presented compactly as

$$
\begin{aligned}
\underset{G, F}{\text{maximize}} \quad & [F^T \ \text{vec}(G)^T]Dp \\
\text{subject to} \quad & [F^T \ \text{vec}(G)^T]S_{lij}p \leq 0, \quad l = 1, \ldots, N, \ i, j \in I_K \\
& G \succeq 0,
\end{aligned} \tag{43}
$$

where p is a vector with dummy variables that encode the monomials of $\alpha, \beta, h, \eta, \rho$, and $D \in \mathbf{R}^{(|F|+|G|) \times |p|}$ and $S \in \mathbf{R}^{(|I_K|N) \times (|F|+|G|) \times |p|}$ are some matrices with constant coefficients.

50

Dualization. Define variables for the energy descent as

$$V_D = [F^T \ \text{vec}(G)^T]Dp,$$

and for interpolating inequality indexed by lij as

$$(V_S)_{lij} = [F^T \ \text{vec}(G)^T]S_{lij}p.$$

Let Z and λ_{lij} for all $i, j \in I_K$, $l = 1, \ldots, N$ be the dual variables for problem (43). Vertically stack λ_{lij} and $(V_S)_{lij}$ to form vectors λ and V_S respectively. The Lagrangian that generates primal problem (43) is

$$L(G, f, Z, \lambda) = V_D - \lambda^T V_S + \mathbf{Tr}(GZ),$$

and the dual problem is given by

$$
\begin{aligned}
\underset{Z,\lambda}{\text{minimize}} \quad & 0 \\
\text{subject to} \quad & D_F p - \lambda^T S_F p = 0 \\
& D_G p - \lambda^T S_G p + Z = 0 \\
& Z \succeq 0 \\
& \lambda \geq 0.
\end{aligned}
\tag{44}
$$

Algebraic proof. Using a weak duality we have $p^\star \leq d^\star$. Let $Z^\star$ and $\lambda^\star$ be optimal dual variables with $d^\star = 0$, then for all $G \in \mathbf{S}_+^{(|\mathcal{C}|+|\mathcal{L}|)/n+|I_K|N}$ and $F \in \mathbf{R}^{|I_K|N}$ it follows that

$$
\begin{aligned}
L(G, F, Z^\star, \lambda^\star) \quad &= \quad F^T \underbrace{\left(D_F p - (\lambda^\star)^T S_F p\right)}_{=0} + \mathbf{Tr}\left(G \underbrace{\left(D_G p - (\lambda^\star)^T S_G p + Z^\star\right)}_{=0}\right) \\
&= \quad 0.
\end{aligned}
$$

Therefore, having $Z^\star$ and $\lambda^\star$ gives us an algebraic proof for the worst-case one step energy difference

$$V_D = \sum_{l,i,j} \underbrace{\lambda^\star_{lij}}_{\geq 0} \underbrace{(V_S)_{lij}}_{\leq 0} - \underbrace{\mathbf{Tr}(GZ^\star)}_{\geq 0} \leq 0,$$

where $G \succeq 0$ because G is a Gram matrix and $(V_S)_{lij} \leq 0$ for all $f_l \in \mathcal{F}_{\mu_l, L_l}$.

G.2 Optimizing over discretizations

In this section we also optimize over the parameters α, β, h, η, and ρ.

G.2.1 QCQP formulation

We can formulate the dual problem (44) as QCQP following [45],

$$
\begin{aligned}
\underset{p,\lambda,P}{\text{minimize}} \quad & 0 \\
\text{subject to} \quad & D_F p - \lambda^T S_F p = 0 \\
& D_G p - \lambda^T S_G p + PP^T = 0 \\
& p^T Q_e p + a_e^T p = 0, \quad e = 1, \ldots, |p| \\
& P \text{ is lower triangular} \\
& \mathbf{diag}(P) \geq 0 \\
& \lambda \geq 0,
\end{aligned}
$$

where relations for dummy variables are specified using quadratic or bilinear constraints $p^T Q_e p + a_e^T p = 0$.

51

G.2.2 Lifted nonconvex SDP

Alternatively, we can formulate the dual problem (44) as lifted nonconvex semidefinite problem with respect to a variable $w = (p, \lambda) \in \mathbf{R}^{|p|+|\lambda|}$. Specifically, we have

$$
\begin{aligned}
\underset{w,W,Z}{\text{minimize}} \quad & 0 \\
\text{subject to} \quad & \overline{D}_F(i,:)w - \mathbf{Tr}\left(\overline{S}_F(:,i,:)W\right) = 0, \quad i = 1, \ldots, |F| \\
& \overline{D}_G(i,:)w - \mathbf{Tr}\left(\overline{S}_G(:,i,:)W\right) + Z(i,i) = 0, \quad i = 1, \ldots, |G| \\
& \mathbf{Tr}\left(\overline{Q}_e W\right) + \overline{a}_e^T w = 0 \\
& W = ww^T \\
& Z \succeq 0 \\
& \lambda \geq 0,
\end{aligned}
\tag{45}
$$

where $\overline{S}_F = \begin{bmatrix} 0 & \frac{1}{2}S_F^T \\ \frac{1}{2}S_F & 0 \end{bmatrix}$, $\overline{D}_F = [\, D_F \quad 0 \,]$, $\overline{S}_G = \begin{bmatrix} 0 & \frac{1}{2}S_G^T \\ \frac{1}{2}S_G & 0 \end{bmatrix}$, $\overline{D}_G = [\, D_G \quad 0 \,]$,

$\overline{Q}_e = \begin{bmatrix} Q_e & 0 \\ 0 & 0 \end{bmatrix}$ and $\overline{a}_e = (a_e, 0)$. In the above the transpose for the third order tensors S_F and S_G is obtained by transposing the first and third dimensions. Note that with the exception of the rank-1 constraint $W = ww^T$, the constraints define convex sets.

G.2.3 SDP relaxation

To find globally optimal solutions to the nonconvex optimization problem, methods like spacial branch-and-bound require good initial bounds on the variables. Following [45], an SDP relaxation of (45) is given by

$$
\begin{aligned}
\underset{w,W,Z}{\text{minimize}} \quad & 0 \\
\text{subject to} \quad & \overline{D}_F(i,:)w - \mathbf{Tr}\left(\overline{S}_F(:,i,:)W\right) = 0, \quad i = 1, \ldots, |F| \\
& \overline{D}_G(i,:)w - \mathbf{Tr}\left(\overline{S}_G(:,i,:)W\right) + Z(i,i) = 0, \quad i = 1, \ldots, |G| \\
& \mathbf{Tr}\left(\overline{Q}_e W\right) + \overline{a}_e^T w = 0 \\
& \begin{bmatrix} W & w \\ w^T & 1 \end{bmatrix} \succeq 0 \\
& Z \succeq 0 \\
& \lambda \geq 0.
\end{aligned}
\tag{46}
$$

Problem (46) is now a convex optimization problem, since the rank-1 constraint $W = ww^T$ has been relaxed to $W \succeq ww^T$. This constraint in turn can be represented equivalently using the Schur complement.

52

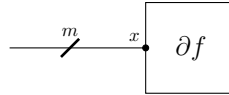
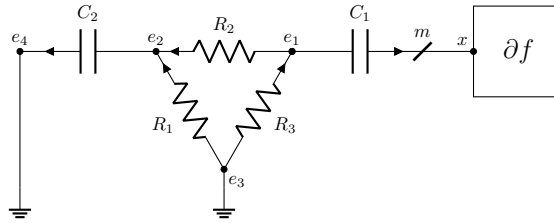
H Package `ciropt`

In this section, we present a simple problem instance to demonstrate the step-by-step process of obtaining a discretized algorithm with our methodology.

Optimization problem. Consider a problem

$$\text{minimize} \quad f(x),$$

where f is a convex function.

Determine the static interconnect. Static interconnect is determined from the optimality conditions. The optimality condition for this problem is to find an x such that $0 \in \partial f(x)$. The corresponding static interconnect for this condition provided below.

Admissible dynamic interconnect. An admissible dynamic interconnect with RLC components relaxes to the static interconnect in equilibrium. The following provides an example of such a dynamic interconnect.

The V-I relations for the circuit (left column) and convergent discretized method found by our method (right) are displayed below.

$$
\begin{aligned}
x &= \textbf{prox}_{(R/2)f}(z) \\
y &= \frac{2}{R}(z - x) \\
\frac{d}{dt}e_2 &= -\frac{1}{2CR}(Ry + 3e_2) \\
\frac{d}{dt}z &= -\frac{1}{4CR}(5Ry + 3e_2)
\end{aligned}
\quad \Bigg| \quad
\begin{aligned}
x^k &= \textbf{prox}_{(R/2)f}(z^k) \\
y^k &= \frac{2}{R}(z^k - x^k) \\
e_2^{k+1} &= e_2^k - \frac{h}{2CR}(Ry^k + 3e_2^k) \\
z^{k+1} &= z^k - \frac{h}{4CR}(5Ry^k + 3e_2^k).
\end{aligned}
$$

Automatic discretization. Now we find a discretization parameters for this dynamic interconnect that guarantee algorithm convergence using `ciropt` package.

Step 1. Define a problem.

```
import ciropt as co
problem = co.CircuitOpt()
```

Step 2. Define function class, in this example f is convex and nondifferentiable, *i.e.*, $\mu = 0$ and $M = \infty$.

```
f = co.def_function(problem, mu=0, M=np.inf)
```

53

Step 3. Define the optimal points.

```
x_star, y_star, f_star = f.stationary_point(
    return_gradient_and_function_value=True)
```

Step 4. Define values for the RLC components and discretization parameters, here for simplicity we take $\alpha = 0$ and $\beta = 1$.

```
R, C = 1, 10
h, eta = problem.h, problem.eta
```

Step 5. Define the one step transition in the discretized V-I relations.

```
z_1 = problem.set_initial_point()
e2_1 = problem.set_initial_point()
x_1 = co.proximal_step(z_1, f, R/2)[0]
y_1 = (2 / R) * (z_1 - x_1)
e1_1 = (e2_1 - R * y_1) / 2
v_C1_1 = e2_1 / 2 - z_1
v_C2_1 = e2_1

e2_2 = e2_1  -  h / (2 * R * C) * (R * y_1 + 3 * e2_1)
z_2 = z_1  -  h / (4 * R * C) * (5 * R * y_1 + 3 * e2_1)
x_2 = co.proximal_step(z_2, f, R/2)[0]
y_2 = (2 / R) * (z_2 - x_2)
v_C1_2 = e2_2 / 2 - z_2
v_C2_2 = e2_2
```

Step 6. Define the dissipative term
$$\mathcal{E}_2 - \mathcal{E}_1 + \eta\langle x^1 - x^\star, y^1 - y^\star\rangle.$$
Solve the final problem.

```
E_1 = (C/2) * (v_C1_1 + x_star)**2 + (C/2) * (v_C2_1)**2
E_2 = (C/2) * (v_C1_2 + x_star)**2 + (C/2) * (v_C2_2)**2
Delta_1 = eta * (x_1 - x_star) * (y_1 - y_star)

problem.set_performance_metric(E_2 - (E_1 - Delta_1))
params = problem.solve()[:1]
```

This gives the disretization parameters
$$b = 6.66, \qquad h = 6.66.$$
The resulting provably convergent algorithm is
$$
\begin{aligned}
x^k &= \mathbf{prox}_{(1/2)f}(z^k) \\
y^k &= 2(z^k - x^k) \\
w^{k+1} &= w^k - 0.33(y^k + 3w^k) \\
z^{k+1} &= z^k - 0.16(5y^k + 3w^k).
\end{aligned}
$$

New algorithm. Solve your problem using new algorithm. Consider Huber penalty function $\phi : \mathbf{R} \to \mathbf{R}$
$$
\phi(x) = \begin{cases} x^2 & |x| \leq 1 \\ 2x - 1 & |x| > 1. \end{cases}
$$

54

We consider the primal problem

$$\begin{aligned}\text{minimize} \quad & f(x) = \textstyle\sum_i \phi(x_i - c_i) \\ \text{subject to} \quad & Ax = b,\end{aligned}$$

where $A \in \mathbf{R}^{m \times n}$, $b \in \mathbf{R}^m$ and $c \in \mathbf{R}^n$, and solve the dual problem

$$\text{maximize} \quad g(y) = -f^*(-A^\mathsf{T} y) - b^\mathsf{T} y.$$

We apply our algorithm to solve the dual problem. Note that the proximal operator $\mathbf{prox}_{\alpha g}(\tilde{y})$ is equivalent to

$$x = \operatorname*{argmin}_x \left(f(x) + (\alpha/2)\|Ax - b\|_2^2 + \tilde{y}^\mathsf{T}(Ax - b) \right), \quad y = \tilde{y} + \alpha(Ax - b).$$

Since f is CCP and 2-smooth (as a Huber loss), f^* is $1/2$-strongly convex. We take $m = 30$, $n = 100$ and sample entries of A, c and b from i.i.d. Gaussian distribution. Finally we rescale the entries of A by $\lambda_{\min}(AA^\mathsf{T})$ to have g that is $1/2$-strongly convex. The following Figure 10 presents the results of the algorithm applied to a random problem instance.

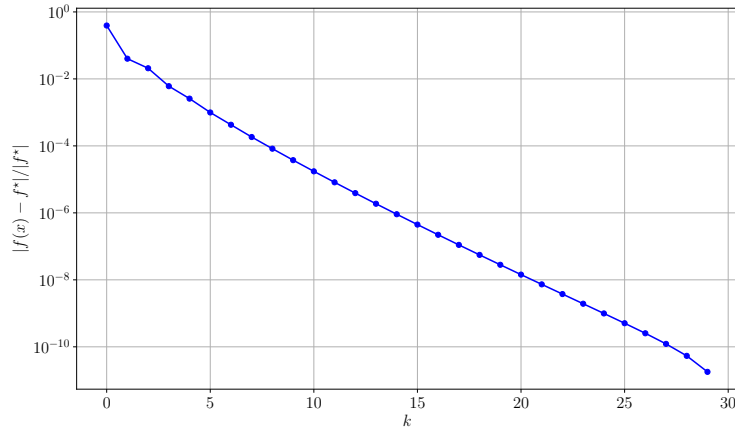



Figure 10: Relative error across iterations when applying the new algorithm.

55

I   Numerical experiments

I.1   Decentralized ADMM+C

Consider a decentralized optimization problem

$$\underset{x\in\mathbf{R}^m}{\text{minimize}}\quad \sum_{i=1}^{N} f_i(x),$$

where $f_1,\ldots,f_N$ are CCP. Suppose furthermore we know some of the functions are strongly convex, that is, suppose there is a subset $S \subset \{1,2,\ldots,N\}$ such that f_j are strongly convex for $j \in S$. We wish to find an efficient algorithm that fully exploits the additional information for f_j's.

To solve the problem in a decentralized manner, define a primal variable $x_i \in \mathbf{R}^m$ for each agent function f_i. To leverage the strong convexity of f_j for each $j \in S$, we could consider implementing a specialized update rule for x_j that is more effective for strongly convex functions.

We consider a modification of the DADMM circuit in §F.3. Recall from §3.1, a circuit with a capacitor and inductor corresponds to a method with momentum. It is known [124] that momentum accelerates the convergence of methods for strongly convex functions. Therefore, we propose to attach capacitors to the circuit in §F.3, on the nets that are directly related to x_j's in $j \in S$. We anticipate that the method derived by discretization of a new circuit (using our automatic discretization methodology) will outperform the DADMM.

Consider a modified decentralized geometric median problem from [138]. Suppose each agent $i \in \{1,\ldots,N\}$ holds vector $b_i \in \mathbf{R}^m$, and consider the minimization problem

$$\underset{x\in\mathbf{R}^m}{\text{minimize}}\quad \sum_{i\in S}\left(\|x-b_i\|_2 + \|x-b_i\|_2^2\right) + \sum_{i\notin S}\|x-b_i\|_2. \tag{47}$$

The minimization subproblem has an explicit solution, *i.e.*,

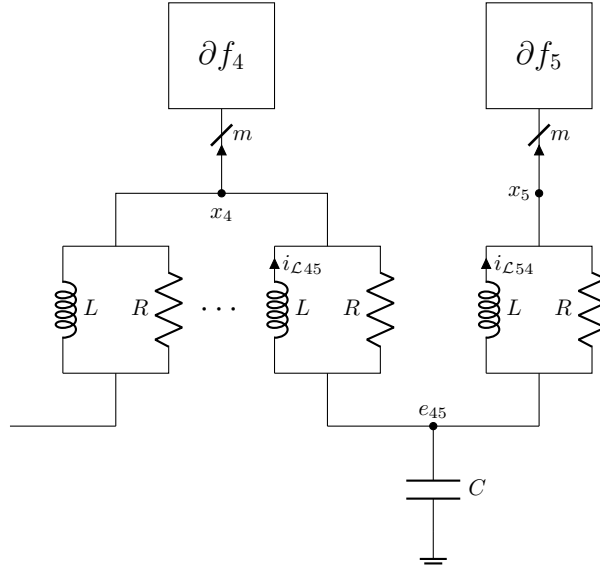
$$\mathbf{prox}_{\rho f_i}(z) = b_i - \frac{b_i - \tilde{z}}{\|b_i - \tilde{z}\|_2}(\|b_i - \tilde{z}\|_2 - \tilde{\rho})_+,$$

where

$$\tilde{z} = \begin{cases} z & i\notin S \\ \frac{1}{1+2\rho}(z+2\rho b_i) & i\in S, \end{cases} \qquad \tilde{\rho} = \begin{cases} \rho & i\notin S \\ \frac{\rho}{1+2\rho} & i\in S. \end{cases}$$

We set $m = 100$, $N = 6$, $S = \{4,5\}$, and sample vectors $b_i \in \mathbf{R}^{100}$ from the uniform distribution over $[-100,100]^{100}$. We use graph G provided in Figure 9. We initialize iterates to $x_i^0 = b_i$ for all i.

We use a modified DADMM circuit §F.3 for the graph in Figure 9. This modified version includes an extra capacitor connected at e_{45}, to which we refer as DADMM+C.

56

Note when $N = 1$, the DADMM+C circuit corresponds to the Nesterov acceleration circuit §3.1. Using KCL at e_{45}, we have

$$i_{\mathcal{L}45} + i_{\mathcal{L}54} = -\frac{(e_{45} - x_4)}{R} - \frac{(e_{45} - x_5)}{R} - C\frac{d}{dt}e_{45}.$$

Thus for $\{j, l\} \neq \{4, 5\}$ the update rule of e_{jl} is given by (34), while for e_{45} we get

$$\frac{d}{dt}e_{45} = -\frac{1}{C}\left(i_{\mathcal{L}45} + i_{\mathcal{L}54} + \frac{1}{R}(2e_{45} - x_4 - x_5)\right).$$

Other V-I relations remain unchanged as in §F.3. The resulting algorithm becomes

$$
\begin{aligned}
x_j^{k+1} &= \mathbf{prox}_{(R/|\Gamma_j|)f_j}\left(\frac{1}{|\Gamma_j|}\sum_{l\in\Gamma_j}(Ri_{\mathcal{L}jl}^k + e_{jl}^k)\right) \\
e_{jl}^{k+1} &= \begin{cases} e_{45}^k - \frac{h}{CR}\left(R(i_{\mathcal{L}45}^k + i_{\mathcal{L}54}^k) + 2e_{45}^k - x_4^{k+1} - x_5^{k+1}\right) & \{j,l\} = \{4,5\} \\ \frac{1}{2}(x_j^{k+1} + x_l^{k+1}) & \text{otherwise} \end{cases} \\
i_{\mathcal{L}jl}^{k+1} &= i_{\mathcal{L}jl}^k + \frac{h}{L}(e_{jl}^{k+1} - x_j^{k+1}).
\end{aligned}
$$

We consider the circuit with $R = 0.8$, $L = 2$ and $C = 15$. To discretize the circuit, we take advantage of the fact that the strong convexity of f_i is 2 for $i \in S$ (47). Specifically, we apply our automatic discretization methodology to convex functions, setting $\mu = 0$ for f_i with $i \notin S$ and $\mu = 2$ for f_i with $i \in S$, and using smoothness $M = 100$. The sufficiently dissipative parameters we find are

$$\eta = 3.70, \quad h = 3.52, \quad \rho = 0, \quad \alpha = 0, \quad \beta = 1, \quad \gamma = 4.48.$$

We compare DADMM+C with DADMM and P-EXTRA. Based on grid search, we set $R = 0.6$ for DADMM in §F.3, and $R = 1$ and $h_1 = \cdots = h_N = 0$ for PG-EXTRA in §F.4 to get P-EXTRA. Note that the parameters of the proximal operators for DADMM are scaled by $1/|\Gamma_j|$, in contrast to P-EXTRA, where $|\Gamma_j|$ is generally not equal to 1. We use Metropolis mixing matrix for P-EXTRA,

$$W_{ij} = \begin{cases} \frac{1}{\max\{|\Gamma_i|, |\Gamma_j|\}+1} & \text{if } i \in \Gamma_j \\ 1 - \sum_{j\in\Gamma_j} W_{ij} & \text{if } i = j \\ 0 & \text{otherwise.} \end{cases}$$

The numerical results are illustrated in Figure 11. The relative error for DADMM+C decreases to 10^{-10} in 66 iterations, for DADMM in 87 iterations and for P-EXTRA in 294 iterations.

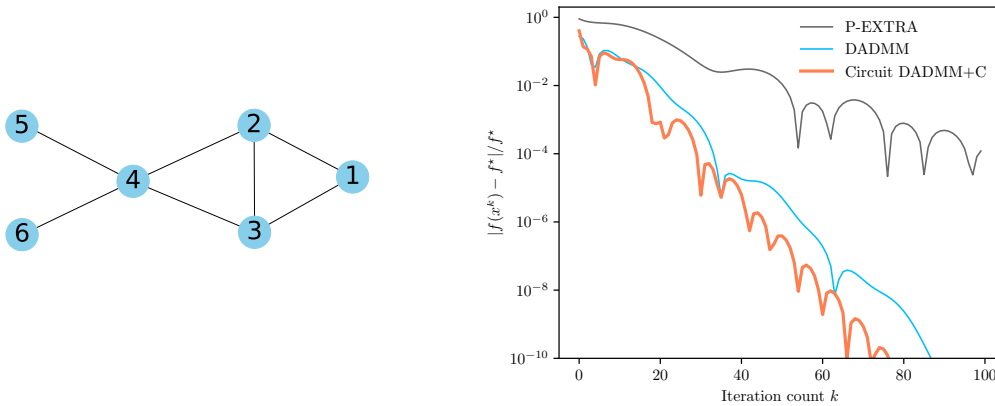



Figure 11: (Left) Underlying graph G. (Right) Relative error $\left|f(x^k) - f^\star\right|/f^\star$ vs. k.

57

I.1.1 Convergence proof of decentralized ADMM+C

We first review the meaning of the numerical values in the previous section. Suppose $(x^\star, y^\star)$ be a primal-dual solution pair. Then by Theorem C.2, there is $(v^\star, i^\star) \in D_{x^\star, y^\star}$ that satisfies $(v^\star, i^\star) = ((0, 0, v^\star_{\mathcal{C}}), (0, i^\star_{\mathcal{L}}, 0))$. The numerical values imply, for the energy function

$$\mathcal{E}_k = \sum_{(j,l) \in A} \left\| i^k_{\mathcal{L}jl} - i^\star_{\mathcal{L}jl} \right\|^2 + \sum_{j<l, \{j,l\} \subset S} \frac{15}{2} \left\| e^k_{jl} - x^\star_j \right\|^2 + 3.52 \sum_{j<l, \{j,l\} \not\subset S} \left\| e^k_{jl} - x^\star_j \right\|^2,$$

following inequality is true up to certain numerical precision

$$\left(\mathcal{E}_{k+1} + 3.7 \langle x^{k+1} - x^\star, y^{k+1} - y^\star \rangle \right) - \mathcal{E}_k \leq 0.$$

This inequality guarantees the convergence of the method we've used in the experiment. Inspired form the numerical results, we could obtain an analytic proof for generalized cases as well. To clarify, A is the set of edges introduced in §F, and each edge is counted twice in the sum $\sum_{(j,l) \in A}$.

Lemma I.1. *Let $f_j : \mathbf{R}^m \to \mathbf{R} \cup \{\infty\}$ are CCP functions for $j \in \{1, \ldots, N\}$ and $S \subset \{1, 2, \ldots, N\}$. Consider the generalized DADMM+C*

$$x^{k+1}_j = \mathbf{prox}_{(R/|\Gamma_j|)f_j} \left(\frac{1}{|\Gamma_j|} \sum_{l \in \Gamma_j} (Ri^k_{\mathcal{L}jl} + e^k_{jl}) \right)$$

$$e^{k+1}_{jl} = \begin{cases} e^k_{jl} - \frac{h}{CR} \left(R(i^k_{\mathcal{L}jl} + i^k_{\mathcal{L}lj}) + 2e^k_{jl} - x^{k+1}_j - x^{k+1}_l \right) & \{j,l\} \subset S \\ \frac{1}{2}(x^{k+1}_j + x^{k+1}_l) & \text{otherwise} \end{cases}$$

$$i^{k+1}_{\mathcal{L}jl} = i^k_{\mathcal{L}jl} + \frac{h}{L}(e^{k+1}_{jl} - x^{k+1}_j),$$

with initilzation $i^0_{\mathcal{L}jl} = i^0_{\mathcal{L}lj}$ for all edge (j,l) in G. Let $(x^\star, y^\star)$ be a primal-dual solution pair and $(v^\star, i^\star) \in D_{x^\star, y^\star}$. Define the energy function as

$$\mathcal{E}_k = \sum_{(j,l) \in A} \frac{L}{2} \left\| i^k_{\mathcal{L}jl} - i^\star_{\mathcal{L}jl} \right\|^2 + \sum_{j<l, \{j,l\} \subset S} \frac{C}{2} \left\| e^k_{jl} - x^\star_j \right\|^2 + \sum_{j<l, \{j,l\} \not\subset S} \frac{h}{2R} \left\| e^k_{jl} - x^\star_j \right\|^2.$$

Then for all $R, L, C, h, \tau > 0$ that satisfy

$$\max \left\{ 1, \frac{2h}{CR} \right\} \leq \tau^2 \leq 2 - \frac{hR}{L},$$

following inequality is true

$$\left(\mathcal{E}_{k+1} + \frac{h}{2R} \left(2 - \frac{hR}{L} - \tau^2 \right) \sum_{(j,l) \in A} \left\| e^{k+1}_{jl} - x^{k+1}_j \right\|^2 + h \langle x^{k+1} - x^\star, y^{k+1} - y^\star \rangle \right) - \mathcal{E}_k \leq 0.$$

Proof. For notation simplicity, define

$$y^{k+1}_{jl} = i^k_{\mathcal{L}jl} + \frac{1}{R} \left(e^k_{jl} - x^{k+1}_j \right).$$

Note, from the first line of the algorithm we have $x^{k+1}_j + \frac{R}{|\Gamma_j|} y^{k+1}_j = \frac{1}{|\Gamma_j|} \sum_{l \in \Gamma_j} \left(Ri^k_{\mathcal{L}jl} + e^k_{jl} \right)$, and therefore

$$y^{k+1}_j = \sum_{l \in \Gamma_j} \left(i^k_{\mathcal{L}jl} + \frac{1}{R} \left(e^k_{jl} - x^{k+1}_j \right) \right) = \sum_{l \in \Gamma_j} y^{k+1}_{jl}.$$

(i) Difference of $\sum_{(j,l) \in A} \frac{L}{2} \left\| i^k_{\mathcal{L}jl} - i^\star_{\mathcal{L}jl} \right\|^2$.

Name

$$\Delta_L = \sum_{(j,l) \in A} \frac{L}{2} \left\| i^{k+1}_{\mathcal{L}jl} - i^\star_{\mathcal{L}jl} \right\|^2 - \sum_{(j,l) \in A} \frac{L}{2} \left\| i^k_{\mathcal{L}jl} - i^\star_{\mathcal{L}jl} \right\|^2.$$

58

Observe

$$
\begin{aligned}
\frac{\Delta_L}{h} &= \frac{1}{h}\left(\sum_{(j,l)\in A}\frac{L}{2}\left\|i^k_{\mathcal{L}_{jl}} - i^\star_{\mathcal{L}_{jl}} + \frac{h}{L}(e^{k+1}_{jl} - x^{k+1}_j)\right\|^2 - \sum_{(j,l)\in A}\frac{L}{2}\left\|i^k_{\mathcal{L}_{jl}} - i^\star_{\mathcal{L}_{jl}}\right\|^2\right)\\
&= \sum_{(j,l)\in A}\left\langle e^{k+1}_{jl} - x^{k+1}_j, i^k_{\mathcal{L}_{jl}} - i^\star_{\mathcal{L}_{jl}}\right\rangle + \sum_{(j,l)\in A}\frac{h}{2L}\left\|e^{k+1}_{jl} - x^{k+1}_j\right\|^2\\
&= \sum_{(j,l)\in A}\left\langle e^{k+1}_{jl} - x^{k+1}_j, y^{k+1}_{jl} - \frac{1}{R}\left(e^k_{jl} - x^{k+1}_j\right) - i^\star_{\mathcal{L}_{jl}}\right\rangle + \sum_{(j,l)\in A}\frac{h}{2L}\left\|e^{k+1}_{jl} - x^{k+1}_j\right\|^2\\
&= -\frac{1}{R}\sum_{(j,l)\in A}\left\langle e^{k+1}_{jl} - x^{k+1}_j, e^k_{jl} - x^{k+1}_j\right\rangle + \sum_{(j,l)\in A}\left\langle e^{k+1}_{jl} - x^{k+1}_j, y^{k+1}_{jl} - i^\star_{\mathcal{L}_{jl}}\right\rangle\\
&\quad + \sum_{(j,l)\in A}\frac{h}{2L}\left\|e^{k+1}_{jl} - x^{k+1}_j\right\|^2\\
&= -\left(\frac{1}{R} - \frac{h}{2L}\right)\sum_{(j,l)\in A}\left\|e^{k+1}_{jl} - x^{k+1}_j\right\|^2 - \frac{1}{R}\sum_{(j,l)\in A}\left\langle e^{k+1}_{jl} - x^{k+1}_j, e^k_{jl} - e^{k+1}_{jl}\right\rangle\\
&\quad - \sum_{(j,l)\in A}\left\langle x^{k+1}_j - x^\star_j, y^{k+1}_{jl} - i^\star_{\mathcal{L}_{jl}}\right\rangle + \sum_{(j,l)\in A}\left\langle e^{k+1}_{jl} - x^\star_j, y^{k+1}_{jl} - i^\star_{\mathcal{L}_{jl}}\right\rangle.
\end{aligned}
$$

On the other hand, from Theorem C.2 we know $i^\star_{\mathcal{R}} = 0$, by KCL at x_j we have $y^\star_j = \sum_{l\in\Gamma_j} i^\star_{\mathcal{L}_{jl}}$. Therefore

$$
\sum_{(j,l)\in A}\left\langle x^{k+1}_j - x^\star_j, y^{k+1}_{jl} - i^\star_{\mathcal{L}_{jl}}\right\rangle = \sum_{j=1}^N\sum_{l\in\Gamma_j}\left\langle x^{k+1}_j - x^\star_j, y^{k+1}_{jl} - i^\star_{\mathcal{L}_{jl}}\right\rangle = \sum_{j=1}^N\left\langle x^{k+1}_j - x^\star_j, y^{k+1}_j - y^\star_j\right\rangle.
$$

Moreover, $i^\star_{\mathcal{L}_{jl}} = -i^\star_{\mathcal{L}_{lj}}$, $e^{k+1}_{jl} = e^{k+1}_{lj}$ holds by their definition, and $x^\star_j = x^\star_l$ as $x^\star$ is the solution. Therefore we see

$$
\sum_{(j,l)\in A}\left\langle e^{k+1}_{jl} - x^\star_j, i^\star_{\mathcal{L}_{jl}}\right\rangle = \sum_{j<l,(j,l)\in A}\left\langle e^{k+1}_{jl} - x^\star_j, i^\star_{\mathcal{L}_{jl}} + i^\star_{\mathcal{L}_{lj}}\right\rangle = 0.
$$

Lastly, following equality is true for $\tau\in(0,\infty)$

$$
\begin{aligned}
\left\langle e^{k+1}_{jl} - x^{k+1}_j, e^k_{jl} - e^{k+1}_{jl}\right\rangle &= \frac{1}{2}\left\|\tau(e^{k+1}_{jl} - x^{k+1}_j) + \frac{1}{\tau}(e^k_{jl} - e^{k+1}_{jl})\right\|^2\\
&\quad - \frac{\tau^2}{2}\left\|e^{k+1}_{jl} - x^{k+1}_j\right\|^2 - \frac{1}{2\tau^2}\left\|e^k_{jl} - e^{k+1}_{jl}\right\|^2.
\end{aligned}
$$

Finally, applying above observations we have

$$
\begin{aligned}
\frac{\Delta_L}{h} &= -\left(\frac{1}{R} - \frac{h}{2L} - \frac{\tau^2}{2R}\right)\sum_{(j,l)\in A}\left\|e^{k+1}_{jl} - x^{k+1}_j\right\|^2\\
&\quad + \frac{1}{2R\tau^2}\sum_{(j,l)\in A}\left\|e^k_{jl} - e^{k+1}_{jl}\right\|^2 - \frac{1}{2R}\sum_{(j,l)\in A}\left\|\tau(e^{k+1}_{jl} - x^{k+1}_j) + \frac{1}{\tau}(e^k_{jl} - e^{k+1}_{jl})\right\|^2\\
&\quad - \sum_{j=1}^N\left\langle x^{k+1}_j - x^\star_j, y^{k+1}_j - y^\star_j\right\rangle + \sum_{(j,l)\in A}\left\langle e^{k+1}_{jl} - x^\star_j, y^{k+1}_{jl}\right\rangle.
\end{aligned}
$$

(ii) Difference of $\sum_{j<l,\{j,l\}\subset S}\frac{C}{2}\left\|e^k_{jl} - x^\star_j\right\|^2$.

Name

$$
\Delta_C = \sum_{j<l,\{j,l\}\subset S}\frac{C}{2}\left\|e^{k+1}_{jl} - x^\star_j\right\|^2 - \sum_{j<l,\{j,l\}\subset S}\frac{C}{2}\left\|e^k_{jl} - x^\star_j\right\|^2.
$$

59

Plugging the definition of the method, we have

$$
\begin{aligned}
\frac{\Delta_C}{h} &= \frac{1}{h}\left(\frac{1}{2}\sum_{\{j,l\}\subset S}\frac{C}{2}\left\|e_{jl}^{k+1}-x_j^\star\right\|^2 - \frac{1}{2}\sum_{\{j,l\}\subset S}\frac{C}{2}\left\|e_{jl}^{k+1}-x_j^\star-\left(e_{jl}^{k+1}-e_{jl}^k\right)\right\|^2\right) \\
&= \frac{C}{2h}\sum_{\{j,l\}\subset S}\left\langle e_{jl}^{k+1}-e_{jl}^k, e_{jl}^{k+1}-x_j^\star\right\rangle - \frac{C}{4h}\sum_{\{j,l\}\subset S}\left\|e_{jl}^{k+1}-e_{jl}^k\right\|^2 \\
&= -\frac{1}{2}\sum_{\{j,l\}\subset S}\left\langle y_{jl}^{k+1}+y_{lj}^{k+1}, e_{jl}^{k+1}-x_j^\star\right\rangle - \frac{C}{4h}\sum_{\{j,l\}\subset S}\left\|e_{jl}^{k+1}-e_{jl}^k\right\|^2 \\
&= -\sum_{\{j,l\}\subset S}\left\langle y_{jl}^{k+1}, e_{jl}^{k+1}-x_j^\star\right\rangle - \frac{C}{4h}\sum_{\{j,l\}\subset S}\left\|e_{jl}^{k+1}-e_{jl}^k\right\|^2.
\end{aligned}
$$

(iii) Difference of $\sum_{j<l,\{j,l\}\not\subset S}\frac{h}{2R}\left\|e_{jl}^k-x_j^\star\right\|^2$.

Name
$$
\Delta_\gamma = \sum_{j<l,\{j,l\}\not\subset S}\frac{h}{2R}\left\|e_{jl}^{k+1}-x_j^\star\right\|^2 - \sum_{j<l,\{j,l\}\not\subset S}\frac{h}{2R}\left\|e_{jl}^k-x_j^\star\right\|^2.
$$
For $\{j,l\}\not\subset S$, from the initialization $i_{\mathcal{L}_{jl}}^0 = -i_{\mathcal{L}_{lj}}^0$ and from $e_{jl}^{k+1}=\frac{1}{2}\left(x_j^{k+1}+x_l^{k+1}\right)$, inductively we can check
$$
i_{\mathcal{L}_{jl}}^{k+1} = i_{\mathcal{L}_{jl}}^k + \frac{h}{L}\left(e_{jl}^{k+1}-x_j^{k+1}\right) = -i_{\mathcal{L}_{lj}}^k - \frac{h}{L}\left(e_{jl}^{k+1}-x_l^{k+1}\right) = -i_{\mathcal{L}_{lj}}^{k+1}.
$$
Therefore $i_{\mathcal{L}_{jl}}^k = -i_{\mathcal{L}_{lj}}^k$ for all k. And from the definition of y_{lj}^{k+1}, we have
$$
y_{lj}^{k+1} = -i_{\mathcal{L}_{jl}}^k - \frac{1}{R}\left(e_{jl}^k-x_j^{k+1}\right) + \frac{2}{R}\left(e_{jl}^k-e_{jl}^{k+1}\right) = -y_{jl}^{k+1} + \frac{2}{R}\left(e_{jl}^k-e_{jl}^{k+1}\right).
$$
And thus
$$
e_{jl}^{k+1}-e_{jl}^k = -\frac{R}{2}\left(y_{lj}^{k+1}+y_{lj}^{k+1}\right).
$$
Now proceeding the similar calculation and argument for Δ_C, we have
$$
\frac{\Delta_\gamma}{h} = -\sum_{\{j,l\}\not\subset S}\left\langle y_{jl}^{k+1}, e_{jl}^{k+1}-x_j^\star\right\rangle - \sum_{j<l,\{j,l\}\not\subset S}\frac{1}{2R}\left\|e_{jl}^{k+1}-e_{jl}^k\right\|^2.
$$

Finally, summing the calculations in (i), (ii), (iii), we have
$$
\frac{1}{h}\left(\mathcal{E}_{k+1}-\mathcal{E}_k\right) + \sum_{j=1}^N\left\langle x_j^{k+1}-x_j^\star, y_j^{k+1}-y_j^\star\right\rangle + \left(\frac{1}{R}-\frac{h}{2L}-\frac{\tau^2}{2R}\right)\sum_{(j,l)\in A}\left\|e_{jl}^{k+1}-x_j^{k+1}\right\|^2
$$
$$
= \frac{1}{h}\left(\Delta_L+\Delta_C+\Delta_\gamma\right) + \sum_{j=1}^N\left\langle x_j^{k+1}-x_j^\star, y_j^{k+1}-y_j^\star\right\rangle + \frac{1}{2R}\left(2-\frac{hR}{L}-\tau^2\right)\sum_{(j,l)\in A}\left\|e_{jl}^{k+1}-x_j^{k+1}\right\|^2
$$
$$
= -\frac{1}{R}\sum_{(j,l)\in A}\left\|\tau(e_{jl}^{k+1}-x_j^{k+1})+\frac{1}{\tau}(e_{jl}^k-e_{jl}^{k+1})\right\|^2
$$
$$
- \frac{1}{2R}\left(\frac{CR}{2h}-\frac{1}{\tau^2}\right)\sum_{\{j,l\}\subset S}\left\|e_{jl}^{k+1}-e_{jl}^k\right\|^2 - \frac{1}{2R}\left(1-\frac{1}{\tau^2}\right)\sum_{\{j,l\}\not\subset S}\left\|e_{jl}^{k+1}-e_{jl}^k\right\|^2.
$$
Therefore, for all $R, L, C, h, \tau > 0$ that satisfy
$$
2-\frac{hR}{L}-\tau^2\geq 0, \quad \frac{CR}{2h}-\frac{1}{\tau^2}\geq 0, \quad 1-\frac{1}{\tau^2}\geq 0
$$
or equivalently,
$$
\max\left\{1,\frac{2h}{CR}\right\}\leq \tau^2 \leq 2-\frac{hR}{L},
$$
we conclude the desired inequality.

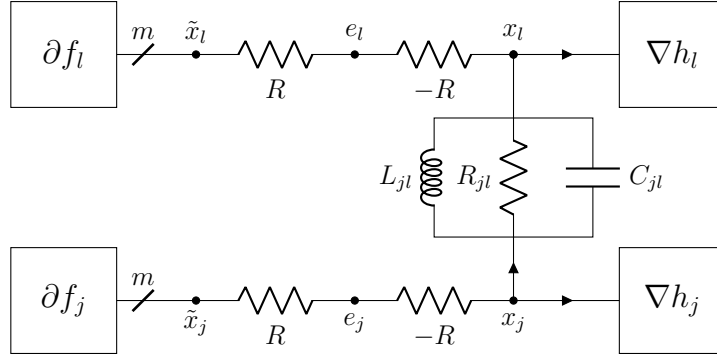
$\square$

60

I.2   PG-EXTRA + Parallel C

In this section, we introduce an additional pipeline of designing new optimization algorithm via circuit. The previous automatized discretization pipeline has the advantage of guaranteeing convergence; however, it may provide a conservative step size since it considers all worst-case scenarios. As a result, it may eliminate the possibility of finding efficient step size that works for certain optimization problem in practice.

Our circuit-based approach has the advantage of designing a variant of the prior method quickly, that is likely to converge and possibly works better based on physical intuition. Furthermore, the variant method provides greater freedom in selecting parameters to tune. We provide an example of new optimization method obtained with exploiting these advantages, that outperforms PG-EXTRA for the problem considered in the paper introduced PG-EXTRA [138].

We use a modified PG-EXTRA circuit §F.4, that includes extra capacitors connected parallel to inductors.

Recalling (6), we know the energy for this circuit is defined as below

$$\mathcal{E}(t) = \sum_{j<l,(j,l)\in A} \frac{C_{jl}}{2}\left\|v_{\mathcal{C}_{jl}}(t) - v^\star_{\mathcal{C}_{jl}}\right\|^2 + \sum_{j<l,(j,l)\in A} \frac{L_{jl}}{2}\left\|i_{\mathcal{L}_{jl}}(t) - i^\star_{\mathcal{L}_{jl}}\right\|^2.$$

Note, compared to the energy of PG-EXTRA, we have additional energy terms for capacitors. Observe

$$v_{\mathcal{C}_{jl}}(t) - v^\star_{\mathcal{C}_{jl}} = x_j(t) - x_l(t) - (x^\star_j - x^\star_l) = x_j(t) - x_l(t).$$

Previously, the energy dissipation by the resistors only reduced the values of $\left\|i_{\mathcal{L}_{jl}}(t) - i^\star_{\mathcal{L}_{jl}}\right\|^2$, but now it also reduces the values of $\|x_j(t) - x_l(t)\|^2$ for $(j,l) \in A$. Intuitively, we may hope that this dissipation accelerates the convergence $\lim_{t\to\infty}(x_j(t) - x_l(t)) = 0$, thus eventually speed up the convergence to the optimal. This observation provides informal motivation for the method.

Following the arguments of §F.4, and additionally defining $u_j = \sum_{l\in\Gamma_j} Ri_{\mathcal{C}_{jl}}$ and setting $C_{jl} = \frac{C}{R_{jl}}$ for $(j,l) \in A$ with some constant $C > 0$, we derive the following method

$$
\begin{aligned}
x^{k+1} &= \mathbf{prox}_{Rf}\left(Wx^k - R\nabla h(x^k) - w^k - u^k\right) \\
w^{k+1} &= w^k + s(I - W)x^k \\
u^{k+1} &= \frac{C}{s}(I - W)(x^{k+1} - x^k).
\end{aligned}
\tag{48}
$$

We now consider the decentralized quadratic programming from [138]. Suppose each agent $j \in \{1,\ldots,N\}$ holds a symmetric positive semidefinite matrix $Q_j \in \mathbf{R}^{m\times m}$, vectors $a_j, p_j \in \mathbf{R}^m$ and scalars $b_j \in \mathbf{R}$. Consider the minimization problem

$$
\begin{aligned}
\underset{x\in\mathbf{R}^m}{\text{minimize}} \quad & \tfrac{1}{N}\sum_{j=1}^N \left(x^\mathsf{T}Q_j x + p_j^\mathsf{T}x\right), \\
\text{subject to} \quad & a_j^\mathsf{T}x \le b_j, \quad j = 1,\ldots,N.
\end{aligned}
$$

Set $f_j(x) = \delta_{\{z|a_j^\mathsf{T}z\le b_j\}}(x)$ and $h_j(x) = x^\mathsf{T}Q_j x + p_j^\mathsf{T}x$, where

$$\delta_{\{z|a_j^\mathsf{T}z\le b_j\}}(x) = \begin{cases} 0 & \text{if } a_j^\mathsf{T}x \le b_j \\ \infty & \text{otherwise} \end{cases}$$

61

is the indicator function. Then the given optimization problem recasts to

$$\underset{x\in\mathbf{R}^m}{\text{minimize}} \quad \tfrac{1}{N}\textstyle\sum_{j=1}^N \left(f_j(x) + h_j(x)\right).$$

Since the minimization subproblem has an explicit solution

$$\mathbf{prox}_{Rf_j}(z) = \begin{cases} z & \text{if } a_j^\mathsf{T}z \le b_j \\ z + \frac{b_j - a_j^\mathsf{T}z}{\|a_j\|_2^2} a_j & \text{otherwise,} \end{cases}$$

this problem can be solved by PG-EXTRA.

We follow the same setting of [138]. We set $m = 50$. Each Q_j is generated by taking the product of $\tilde{Q}_j$ and its transpose, where $\tilde{Q}_j \in \mathbf{R}^{m\times m}$ is a matrix with elements that follow an i.i.d. Gaussian distribution. Each p_j is generated to follow an i.i.d. Gaussian distribution. Vectors a_j and b_j are also randomly generated, however, we conducted the experiment for the case that the solution of the constrained problem differs from that of the unconstrained problem. We use Metropolis mixing matrix as in §I.1.

The numerical results are illustrated in Figure 12. We compare PG-EXTRA and the variant method (48) obtained from the modified circuit with additional parallel capacitors. We use $R = 0.05$, $R = 0.07$ for PG-EXTRA and $R = 0.07$, $C = 0.3$ and $s = 0.8$ for (48). The parameters for PG-EXTRA were obtained through a grid search. The parameters for (48) are hand-optimized starting from $C = 0$ and $s = 0.5$, the parameter selection that makes (48) to coincide with PG-EXTRA when $u^0 = 0$. The relative error for (48) decreases to 10^{-8} in 147 iterations, while for PG-EXTRA with $R = 0.05$ in 214 iterations.

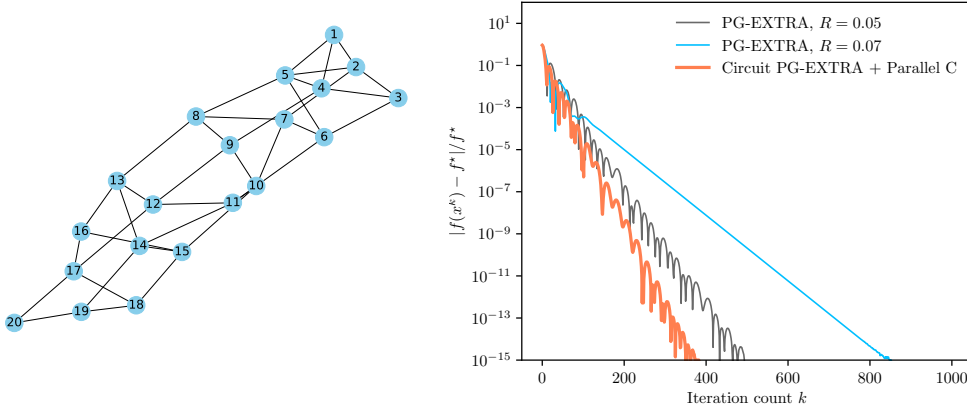



Figure 12: (Left) Underlying graph G. (Right) Relative error $\left|f(x^k) - f^\star\right|/f^\star$ vs. k.

62

NeurIPS Paper Checklist

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: We provide the methodology to design an appropriate electric circuit whose continuous-time dynamics converge to the solution of the optimization problem in §2, §3, and introduce the automated, computer-assisted discretization methodology in §4, §5.

   Guidelines:

   - The answer NA means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A No or NA answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
   - It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [NA]

   Justification: This work proposes a new methodology for optimization algorithm design. The theoretical nature makes the discussion of such limitations not strongly relevant.

   Guidelines:

   - The answer NA means that the paper has no limitation while the answer No means that the paper has limitations, but those are not discussed in the paper.
   - The authors are encouraged to create a separate "Limitations" section in their paper.
   - The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
   - The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
   - The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
   - The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
   - If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
   - While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory Assumptions and Proofs**

   Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

63

Answer: [Yes]

Justification: Our paper provides the full set of assumptions and complete proofs for all theoretical results, with theorems and lemmas properly numbered and cross-referenced.

Guidelines:

- The answer NA means that the paper does not include theoretical results.
- All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
- All assumptions should be clearly stated or referenced in the statement of any theorems.
- The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
- Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
- Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental Result Reproducibility**

Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

Answer: [Yes]

Justification: We provide the code and thoroughly describe the details to ensure reproducibility of the experimental results.

Guidelines:

- The answer NA means that the paper does not include experiments.
- If the paper includes experiments, a No answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
- If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
- Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
- While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
  (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
  (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
  (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).
  (d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. **Open access to data and code**

64

Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

Answer: [Yes]

Justification: We provide our codes and the instructions for using them in the `README.md` file.

Guidelines:

- The answer NA means that paper does not include experiments requiring code.
- Please see the NeurIPS code and data submission guidelines (`https://nips.cc/public/guides/CodeSubmissionPolicy`) for more details.
- While we encourage the release of code and data, we understand that this might not be possible, so "No" is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
- The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (`https://nips.cc/public/guides/CodeSubmissionPolicy`) for more details.
- The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
- The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
- At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
- Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental Setting/Details**

Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer, etc.) necessary to understand the results?

Answer: [Yes]

Justification: We provide the necessary details and code to understand and reproduce the experiments.

Guidelines:

- The answer NA means that the paper does not include experiments.
- The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
- The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment Statistical Significance**

Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

Answer: [NA]

Justification: Our work is theoretical and includes non-statistical simple experiments with synthetic data. Therefore, the notion of statistical significance is not applicable to our work.

Guidelines:

- The answer NA means that the paper does not include experiments.
- The authors should answer "Yes" if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.

65

- The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
- The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
- The assumptions made should be given (e.g., Normally distributed errors).
- It should be clear whether the error bar is the standard deviation or the standard error of the mean.
- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g. negative error rates).
- If error bars are reported in tables or plots, The authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments Compute Resources**

Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

Answer: [Yes]

Justification: We use minimal CPU computation for toy experiments.

Guidelines:

- The answer NA means that the paper does not include experiments.
- The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
- The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
- The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code Of Ethics**

Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics `https://neurips.cc/public/EthicsGuidelines`?

Answer: [Yes]

Justification: We have reviewed the NeurIPS Code of Ethics and our paper conforms to it in every respect.

Guidelines:

- The answer NA means that the authors have not reviewed the NeurIPS Code of Ethics.
- If the authors answer No, they should explain the special circumstances that require a deviation from the Code of Ethics.
- The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader Impacts**

Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

Answer: [NA]

Justification: We conduct theoretical research on convex optimization and do not anticipate any negative social impacts from our results.

Guidelines:

- The answer NA means that there is no societal impact of the work performed.

66

- If the authors answer NA or No, they should explain why their work has no societal impact or why the paper does not address societal impact.
- Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
- The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
- The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
- If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pretrained language models, image generators, or scraped datasets)?

Answer: [NA]

Justification: Our theoretical research on convex optimization does not involve releasing data or models, hence no safeguards are needed.

Guidelines:

- The answer NA means that the paper poses no such risks.
- Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
- Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
- We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

Answer: [Yes]

Justification: We explicitly mention the solvers and the packages we are referring to in the paper and the code.

Guidelines:

- The answer NA means that the paper does not use existing assets.
- The authors should cite the original paper that produced the code package or dataset.
- The authors should state which version of the asset is used and, if possible, include a URL.
- The name of the license (e.g., CC-BY 4.0) should be included for each asset.

67

- For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
- If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, `paperswithcode.com/datasets` has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
- For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
- If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New Assets**

    Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

    Answer: [NA]

    Justification: This paper does not release new datasets or assets.

    Guidelines:

    - The answer NA means that the paper does not release new assets.
    - Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
    - The paper should discuss whether and how consent was obtained from people whose asset is used.
    - At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and Research with Human Subjects**

    Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

    Answer: [NA]

    Justification: This paper does not involve crowdsourcing nor research with human subjects.

    Guidelines:

    - The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
    - Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
    - According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional Review Board (IRB) Approvals or Equivalent for Research with Human Subjects**

    Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

    Answer: [NA]

    Justification: This paper does not involve crowdsourcing nor research with human subjects.

    Guidelines:

    - The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.

68

- Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
- We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
- For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

69