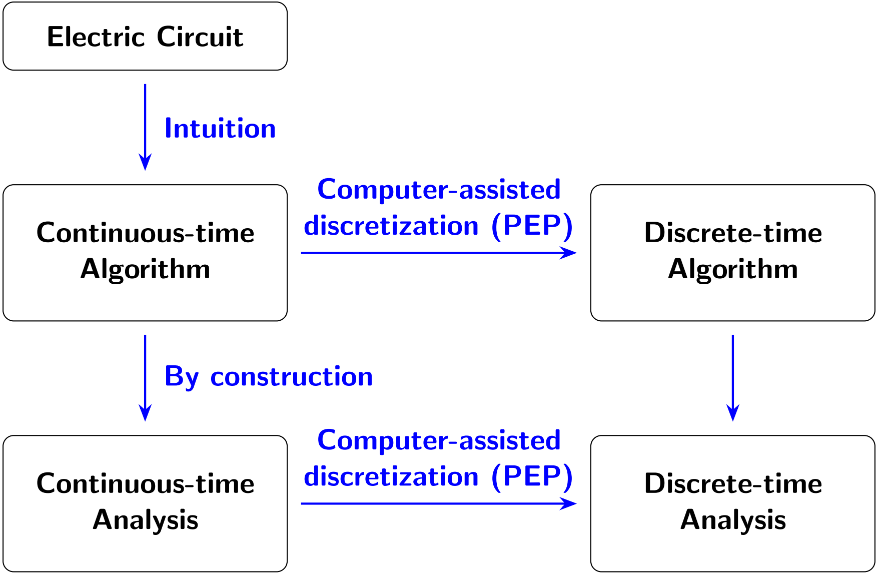
Optimization Algorithm Design via Electric Circuits[1]

Stephen P. Boyd, Tetiana Parshakova*, Ernest K. Ryu, **Jaewook J. Suh***

2025 Spring Quantitative Methods Seminar 03.14.

[1]Neural Information Processing Systems (spotlight), 2024.
*Lead authors.

Our goal: a novel framework to design a new algorithm



Outline

Outline

Outline

V-I relations: Resistor, Capacitor, Inductor

- Resistor: Ohm's law

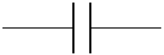
$$v(t) = Ri(t)$$

- Inductor

$$v(t) = L\frac{d}{dt}i(t)$$

- Capacitor

$$i(t) = C\frac{d}{dt}v(t)$$

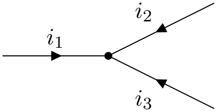


Mathematical meaning: generate an ODE/dynamics

Kirchhoff's laws[2]

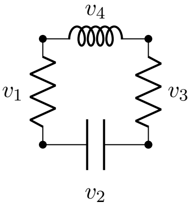
- Kirchhoff's current law (KCL)

$$\sum_k i_k = 0$$



- Kirchhoff's voltage law (KVL)

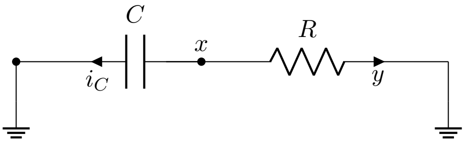
$$\sum_k v_k = 0$$



Mathematical meaning: linear equations for v and i

[2] C. A. Desoer and E. S. Kuh. *Basic Circuit Theory*. Electronic Engineering. McGraw-Hill, 1969, § 1.

R-C circuit



- Ohm's Law: $x(t) = Ry(t)$. Dividing both sides by R this is equivalent to

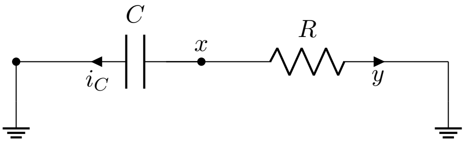
$$y(t) = \frac{1}{R}x(t)$$

- Capacitor: $i_C(t) = C\frac{d}{dt}v_C(t)$
- From KVL: $v_C(t) - x(t) = 0$
- From KCL: $i_C(t) + y(t) = 0$

Combining all together, we conclude

$$\frac{d}{dt}x(t) = \frac{d}{dt}v_C(t) = \frac{1}{C}i_C(t) = -\frac{1}{C}y(t) = -\frac{1}{CR}x(t).$$

R-C circuit



- Ohm's Law: $x(t) = Ry(t)$. Dividing both sides by R this is equivalent to

$$y(t) = \frac{1}{R}x(t)$$

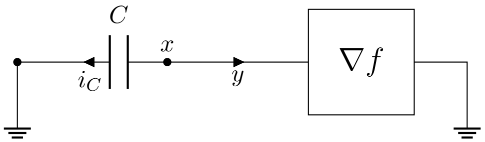
- Capacitor: $i_{\mathcal{C}}(t) = C\frac{d}{dt}v_{\mathcal{C}}(t)$
- From KVL: $v_{\mathcal{C}}(t) - x(t) = 0$
- From KCL: $i_{\mathcal{C}}(t) + y(t) = 0$

Combining all together, we conclude

$$\frac{d}{dt}x(t) = -\frac{1}{CR}x(t).$$

Thus $x(t) = e^{-\frac{t}{CR}}x(0)$.

∇f-C circuit: Gradient flow



- Consider a device with nonlinear V-I relation[3]

$$y(t) = \nabla f(x(t))$$

- Capacitor: $i_C(t) = C \frac{d}{dt} v_C(t)$
- From KVL: $v_C(t) - x(t) = 0$
- From KCL: $i_C(t) + y(t) = 0$

Combining all together, we obtain gradient flow:

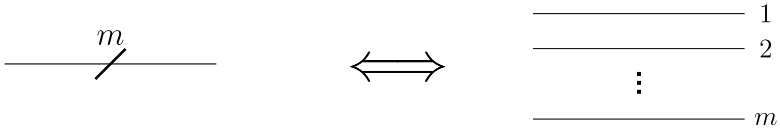
$$\frac{d}{dt} x(t) = -\frac{1}{C} \nabla f(x(t))$$

Linear resistor can be considered as special case $f(x) = \frac{1}{R} \|x\|^2$.

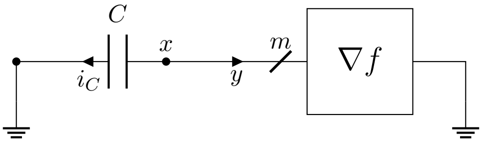
[3] Jack Bonnell Dennis. "Mathematical Programming and Electrical Networks". PhD thesis. Massachusetts Institute of Technology, 1959.

Multi-wire notation

To consider multi-dimensional $x \in \mathbf{R}^m$, we adopt notation:



∇f-C circuit: Gradient flow



- Consider a device with V-I relation

$$y(t) = \nabla f(x(t))$$

- Capacitor: $i_{\mathcal{C}}(t) = D_{\mathcal{C}} \frac{d}{dt} v_{\mathcal{C}}(t)$
- From KCL: $v_{\mathcal{C}}(t) - x(t) = 0$
- From KVL: $i_{\mathcal{C}}(t) + y(t) = 0$

Combining all together, we obtain gradient flow:

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

Here $D_{\mathcal{C}} = \mathbf{diag}(C_1, \ldots, C_m)$.

Convex function: Important property

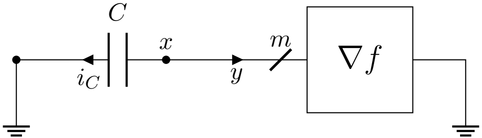
Following property is important in our framework.

Fact. Suppose $f\colon \mathbf{R}^m \to \mathbf{R}$ is a convex differentiable function. Then for all $x, z \in \mathbf{R}^m$,

$$\langle \nabla f(x) - \nabla f(z), x - z \rangle \geq 0.$$

We call this property as *incremental passivity*.

∇f-C circuit: Energy dissipation



Recall, from circuit laws we have $x(t) = v_{\mathcal{C}}(t)$, $y(t) = -i_{\mathcal{C}}(t)$, $i_{\mathcal{C}}(t) = D_{\mathcal{C}}\frac{d}{dt}v_{\mathcal{C}}(t)$ and

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

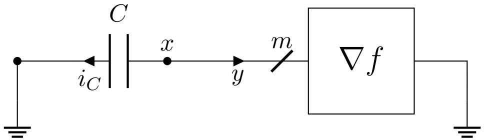
Let $v_{\mathcal{C}}^\star = x^\star = \operatorname{argmin}_x f(x)$. Denote $\|u\|_{D_{\mathcal{C}}}^2 := \langle D_{\mathcal{C}}u, u\rangle$. Define the energy as

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

Denote $y^\star = \nabla f(x^\star) = 0$. Differentiating, from *incremental passivity* we have

$$\begin{aligned}\frac{d}{dt}\mathcal{E}(t) = \left\langle D_{\mathcal{C}}\frac{d}{dt}v_{\mathcal{C}}(t), v_{\mathcal{C}}(t) - v_{\mathcal{C}}^\star\right\rangle &= \langle i_{\mathcal{C}}(t), v_{\mathcal{C}}(t) - v_{\mathcal{C}}^\star\rangle \\ &= -\langle y(t) - y^\star, x(t) - x^\star\rangle \le 0.\end{aligned}$$

∇f-C circuit: Energy dissipation (since $x = v$ and $y = \nabla f(x)$)



Recall, from circuit laws we have $x(t) = v_{\mathcal{C}}(t)$, $y(t) = -i_{\mathcal{C}}(t)$, $i_{\mathcal{C}}(t) = D_{\mathcal{C}} \frac{d}{dt} v_{\mathcal{C}}(t)$ and

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

Let $v_{\mathcal{C}}^\star = x^\star = \mathrm{argmin}_x\, f(x)$. Denote $\|u\|_{D_{\mathcal{C}}}^2 := \langle D_{\mathcal{C}} u, u \rangle$. Define the energy as

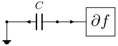
$$\mathcal{E}(t) = \frac{1}{2} \|x(t) - x^\star\|_{D_{\mathcal{C}}}^2 .$$

Denote $y^\star = \nabla f(x^\star) = 0$. Differentiating, from *incremental passivity* we have

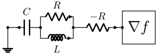
$$\frac{d}{dt} \mathcal{E}(t) = -\langle y(t) - y^\star, x(t) - x^\star \rangle = -\langle \nabla f(x), x(t) - x^\star \rangle \le 0.$$

Zoo of Electric Circuits for Optimization Algorithms

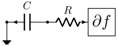
Gradient Descent



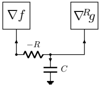
Nesterov acceleration with $R = \sqrt{L/C}$



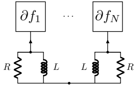
Proximal point method



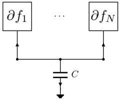
Prox-grad method



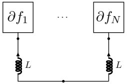
Prox Decomposition



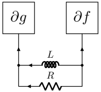
Primal Decomposition



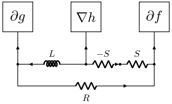
Dual Decomposition



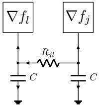
Douglas–Rachford splitting



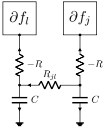
Davis–Yin splitting



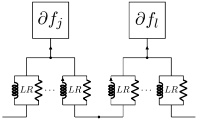
Decentralized GD



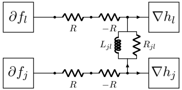
Adapt-then-combine



Decentralized ADMM



PG-EXTRA



Outline

Problem setup[4]

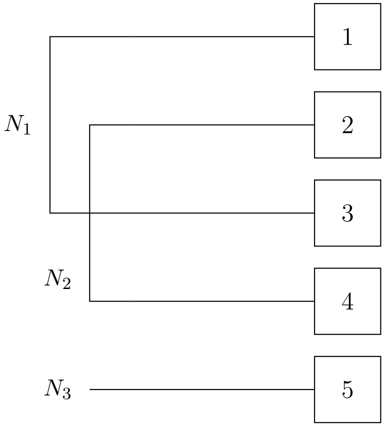
Consider the problem

$$\begin{array}{ll} \underset{x}{\text{minimize}} & f(x) \\ \text{subject to} & x \in \mathcal{R}(E^\intercal) \end{array}$$

- $f : \mathbf{R}^m \to \mathbf{R} \cup \{\infty\}$ is convex and differentiable
- n nets $N_1, \ldots, N_n$ forming a partition of $\{1, \ldots, m\}$
- $E \in \mathbf{R}^{n \times m}$ is a selection matrix

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

- $x \in \mathcal{R}(E^\intercal) \iff x_i = x_j$ when i, j are in the same net



[4] The paper considers closed and proper f, but here we assume differentiability for simplicity

Lagrangian dual problem

The problem setup is equivalent to

$$\begin{array}{ll} \underset{x,z}{\text{minimize}} & f(x) \\ \text{subject to} & E^\intercal z - x = 0 \end{array}$$

Define the Lagrangian $\mathcal{L}\colon (\mathbf{R}^m \times \mathbf{R}^n) \times \mathbf{R}^m \to \mathbf{R}$ as

$$\mathcal{L}(x, z, y) = f(x) + \langle y, E^\intercal z - x \rangle = f(x) - \langle y, x \rangle + \langle Ey, z \rangle \,.$$

Lagrangian dual function $d\colon \mathbf{R}^m \to \mathbf{R}$ becomes[5]

$$d(y) = \inf_{x,z} \mathcal{L}(x, z, y) = \begin{cases} -f^*(y) & \text{if } Ey = 0 \\ -\infty & \text{otherwise.} \end{cases}$$

Thus $\mathbf{dom}\, d = \{y \mid Ey = 0\} = \mathcal{N}(E)$.

[5] $f^*(y) := \sup_x \{\langle y, x \rangle - f(x)\}$

Lagrangian dual problem

Let $x^\star$ be an optimal point of the primal problem. Then there is $z^\star$ s.t. $E^\intercal z^\star = x^\star$. Note

$$d(y) = \inf_{x,z} \mathcal{L}(x, z, y) \leq \mathcal{L}(x^\star, z^\star, y) = f(x^\star) \tag{1}$$

This motivates following *Lagrangian dual problem*

$$\begin{array}{ll} \underset{y}{\text{maximize}} & d(y) \\ \text{subject to} & Ey = 0 \end{array}$$

We call y as dual variable. Suppose $y^\star$ is optimal solution of above problem. From (1),

$$d(y^\star) \leq f(x^\star).$$

When $d(y^\star) = f(x^\star)$, we say that *strong duality* holds.[6]

[6] For more details, refer § 5 Boyd & Vandenberghe. *Convex Optimization*. [1]

KKT optimality conditions

Suppose $x^\star$, $y^\star$ are primal, dual optimal points and strong duality holds. Then they satisfy following *Karush-Kuhn-Tucker(KKT) optimality conditions*:

$$\begin{array}{rcll} y^\star & = & \nabla f(x^\star) & \text{(stationarity)} \\ x^\star & \in & \mathcal{R}(E^\mathsf{T}) & \text{(primal feasibility)} \\ y^\star & \in & \mathcal{N}(E) & \text{(dual feasibility)} \end{array}$$

As f is convex, this is also a necessary condition[7].

[7] S. P. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004, § 5.5.3.

KKT optimality conditions: Example $E = I$

Consider a special case $E = I$. Then $\mathcal{R}(E^\intercal) = \mathbf{R}^m$ and $\mathcal{N}(E) = \{0\}$.
Primal problem becomes the unconstrained problem

$$\underset{x \in \mathbf{R}^m}{\text{minimize}} \quad f(x).$$

Since f is convex, if minimizer exists, $x^\star = \operatorname{argmin}_x f(x) \iff \nabla f(x^\star) = 0$.
On the other hand, KKT optimality condition becomes

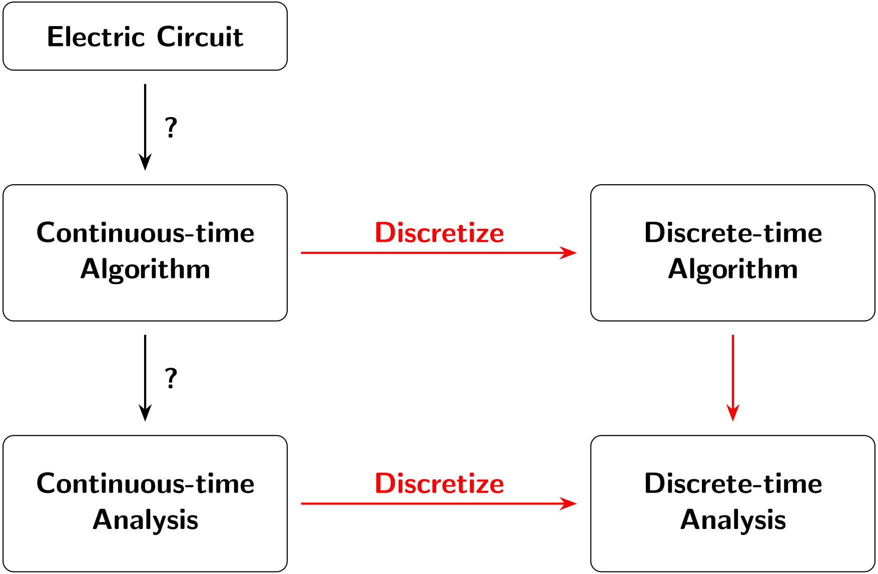
$$\begin{array}{rcll} y^\star & = & \nabla f(x^\star) & \text{(stationarity)} \\ x^\star & \in & \mathbf{R}^m & \text{(primal feasibility)} \\ y^\star & \in & \{0\} & \text{(dual feasibility).} \end{array}$$

This implies $\nabla f(x^\star) = y^\star = 0$.

Outline

Outline

Pipeline diagram



Review: Problem setup

Consider the problem

$$\begin{array}{ll} \underset{x}{\text{minimize}} & f(x) \\ \text{subject to} & x \in \mathcal{R}(E^\mathsf{T}) \end{array}$$

- $f : \mathbf{R}^m \to \mathbf{R} \cup \{\infty\}$ is convex and differentiable
- n nets $N_1, \ldots, N_n$ forming a partition of $\{1, \ldots, m\}$
- $E \in \mathbf{R}^{n \times m}$ is a selection matrix

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

Consensus convex optimization problem ($x_i \in \mathbf{R}$)

Consider the problem

$$\begin{array}{ll} \underset{x}{\text{minimize}} & f(x) \\ \text{subject to} & x \in \mathcal{R}(E^\intercal) \end{array}$$

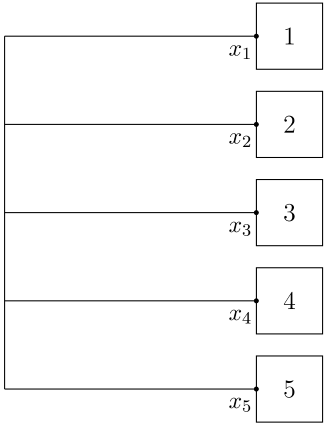
- $f : \mathbf{R}^m \to \mathbf{R} \cup \{\infty\}$ is convex and differentiable

$$f(x) = \sum_{i=1}^{m} f_i(x_i)$$

- $E \in \mathbf{R}^{1 \times m}$ for right graph corresponds to

$$E = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \end{bmatrix}$$

- Clearly, $x \in \mathcal{R}(E^\intercal) \iff x_1 = \cdots = x_m$



Consensus convex optimization problem ($x_i \in \mathbf{R}^n$)

Consider the problem

$$\begin{array}{ll} \underset{x}{\text{minimize}} & f(x) \\ \text{subject to} & x \in \mathcal{R}(E^\intercal) \end{array}$$

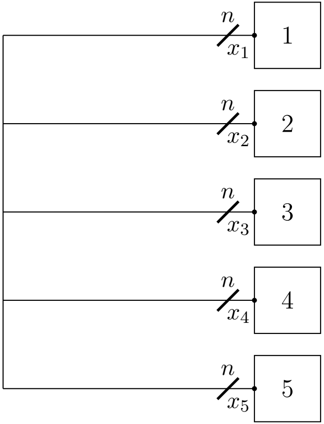
- $f : \mathbf{R}^m \to \mathbf{R} \cup \{\infty\}$ is convex and differentiable

$$f(x) = \sum_{i=1}^{m/n} f_i(x_i)$$

- $E \in \mathbf{R}^{n \times m}$ for right graph corresponds to

$$E = \begin{bmatrix} I & I & I & I & I \end{bmatrix}$$

- Clearly, $x \in \mathcal{R}(E^\intercal) \iff x_1 = \cdots = x_m$
- For simplicity, we consider an example with $n = 1$



KKT optimality conditions

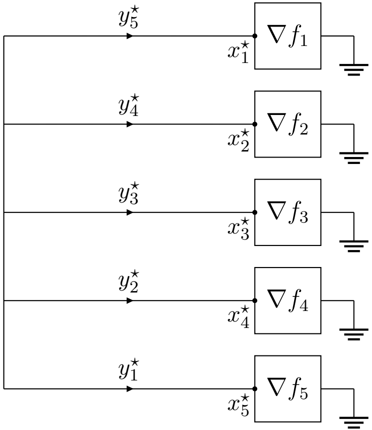
$$\begin{aligned} y^\star &= \nabla f(x^\star) \quad \text{(stationarity)} \\ x^\star &\in \mathcal{R}(E^\mathsf{T}) \quad \text{(primal feasibility)} \\ y^\star &\in \mathcal{N}(E) \quad \text{(dual feasibility)} \end{aligned}$$

$$x^\star \in \mathcal{R}(E^\mathsf{T}) \iff x_1^\star = \cdots = x_m^\star$$

$$y^\star \in \mathcal{N}(E) \iff \sum_{j=1}^{m} y_j^\star = 0$$

Circuit interpretation: KKT optimality conditions

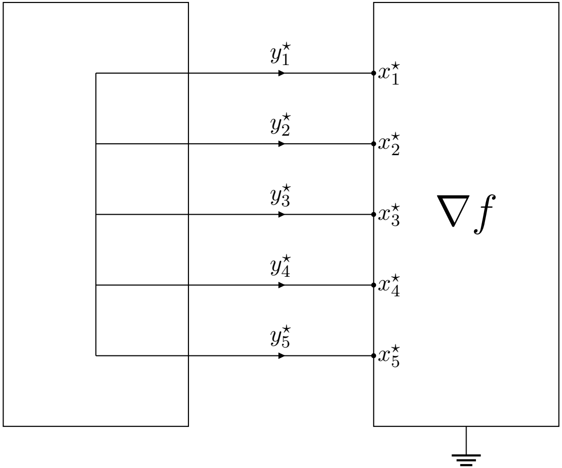
$$\begin{aligned} y^\star &= \nabla f(x^\star) \quad \text{(nonlinear resistor)} \\ x^\star &\in \mathcal{R}(E^\mathsf{T}) \quad \text{(KVL)} \\ y^\star &\in \mathcal{N}(E) \quad \text{(KCL)} \end{aligned}$$



Circuit interpretation: KKT optimality conditions

$$\begin{aligned} y^\star &= \nabla f(x^\star) \quad \text{(nonlinear resistor)} \\ x^\star &\in \mathcal{R}(E^\mathsf{T}) \quad \text{(KVL)} \\ y^\star &\in \mathcal{N}(E) \quad \text{(KCL)} \end{aligned}$$

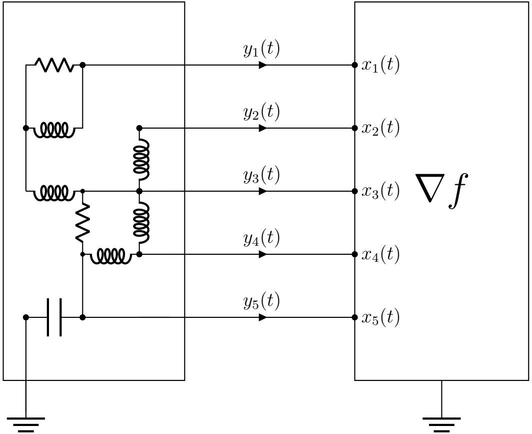
Static interconnect



Circuit interpretation: Dynamic interconnect[8]

$$
\begin{aligned}
y(t) &= \nabla f(x(t)) \quad \text{(nonlinear resistor)}\\
v(t) &= A^\mathsf{T} \begin{bmatrix} x(t) \\ e(t) \end{bmatrix} \quad \text{(KVL)}\\
Ai(t) &= \begin{bmatrix} -y(t) \\ 0 \end{bmatrix} \quad \text{(KCL)}\\
v_\mathcal{R}(t) &= D_\mathcal{R} i_\mathcal{R}(t) \quad \text{(resistor)}\\
v_\mathcal{L}(t) &= D_\mathcal{L} \frac{d}{dt} i_\mathcal{L}(t) \quad \text{(inductor)}\\
i_\mathcal{C}(t) &= D_\mathcal{C} \frac{d}{dt} v_\mathcal{C}(t) \quad \text{(capacitor)}
\end{aligned}
$$

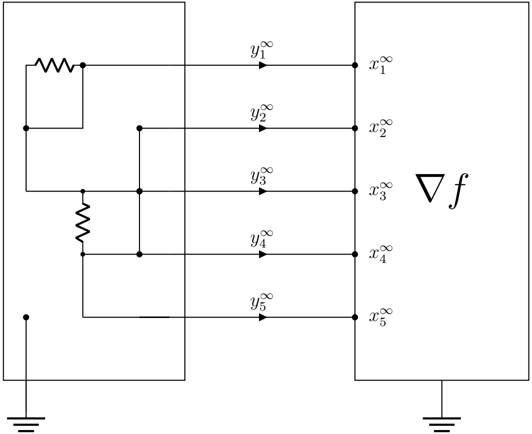
Dynamic interconnect



[8] For definition of A, see §2.1 of our paper. The reformulation of KVL and KCL can be thought of as a consequence of Tellegen's theorem. [Basic Circuit Theory, §10.2.3].

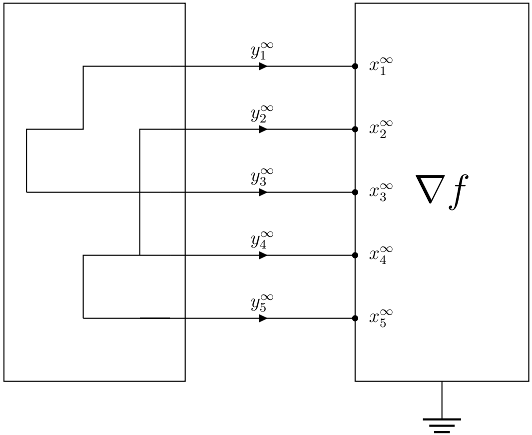
Point: Equilibrium satisfies KKT optimality conditions

$$\begin{aligned} v_{\mathcal{R}}(t) &= D_{\mathcal{R}} i_{\mathcal{R}}(t) \to 0 \\ v_{\mathcal{L}}(t) &= D_{\mathcal{L}} \frac{d}{dt} i_{\mathcal{L}}(t) \to 0 \\ i_{\mathcal{C}}(t) &= D_{\mathcal{C}} \frac{d}{dt} v_{\mathcal{C}}(t) \to 0 \end{aligned}$$



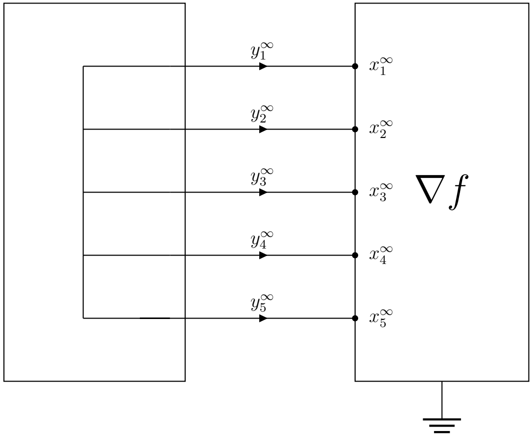
Point: Equilibrium satisfies KKT optimality conditions

$$y^\infty = \nabla f(x^\infty) \quad \text{(nonlinear resistor)}$$
$$x^\infty \in \mathcal{R}(E^\intercal) \quad \text{(KVL)}$$
$$y^\infty \in \mathcal{N}(E) \quad \text{(KCL)}$$



Point: Equilibrium satisfies KKT optimality conditions

$$
\begin{aligned}
y^\infty &= \nabla f(x^\infty) \quad \text{(nonlinear resistor)}\\
x^\infty &\in \mathcal{R}(E^\mathsf{T}) \quad \text{(KVL)}\\
y^\infty &\in \mathcal{N}(E) \quad \text{(KCL)}
\end{aligned}
$$



Continuous-time analysis

Theorem 2.2. (Informal)
Let $(v(t), i(t), x(t), y(t))$ satisfies the ODE for dynamic circuit, which its equilibrium corresponds to the static circuit. Let $(x^\star, y^\star)$ be a primal-dual optimal solution, then

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

Proof outline.
Define the energy of the circuit at time t as

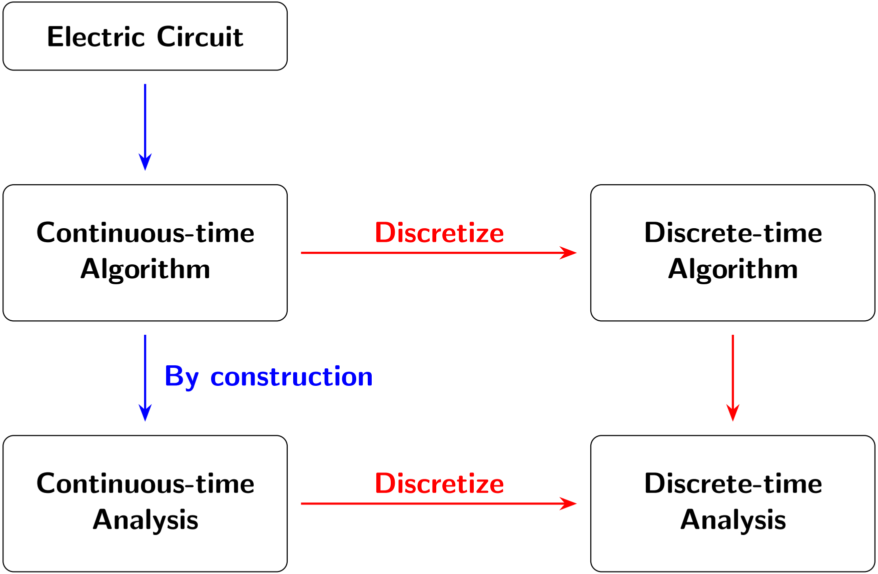
$$\mathcal{E}(t) = \frac{1}{2}\|v_\mathcal{C}(t) - v_\mathcal{C}^\star\|_{D_\mathcal{C}}^2 + \frac{1}{2}\|i_\mathcal{L}(t) - i_\mathcal{L}^\star\|_{D_\mathcal{L}}^2.$$

Consider energy dissipation

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

From integrability, we have $\lim_{t\to\infty}\langle x - x^\star, y - y^\star\rangle = 0$, concluding the result. □

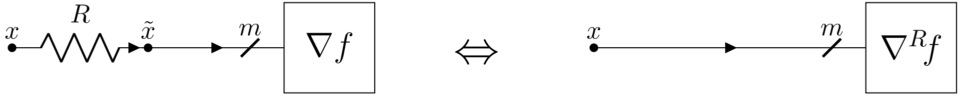
Where are we


Electric Circuit
Continuous-time Algorithm
Discretize
Discrete-time Algorithm
By construction
Continuous-time Analysis
Discretize
Discrete-time Analysis

Outline

Resistor: Moreau envelope

We can consider equivalent resistance of ∇f and R.



Denote Moreau envelope of a convex function f of parameter $R > 0$ as ${}^R f(x)$.
Then $\nabla^R f$ is $\frac{1}{R}$-Lipschitz continuous and given by

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

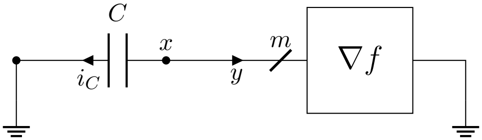
Moreau envelope is a famous smoothing technique. By KCL and Ohm's law,

$$\frac{1}{R}(x - \tilde{x}) = y = \nabla f(\tilde{x}) \quad \Longrightarrow \quad \tilde{x} = \mathbf{prox}_{Rf}(x)$$

Thus $\frac{1}{R}(x - \tilde{x}) = \nabla^R f$, currents for two circuits coincide.

Capacitor: Primal variable update

We've already observed this point. Recall:



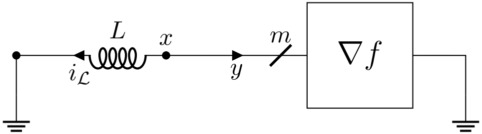
- V-I relation for nonlinear resistor: $y(t) = \nabla f(x(t))$
- V-I relation capacitor: $i_C(t) = D_C \frac{d}{dt} v_C(t)$
- From KCL: $v_C(t) - x(t) = 0$
- From KVL: $i_C(t) + y(t) = 0$

For simplicity, let $D_{\mathcal{C}} = CI$. Combining all together, we obtain:

$$\frac{d}{dt}x(t) = -\frac{1}{C}\nabla f(x(t)) \quad \longrightarrow \quad x^{k+1} = x^k - \frac{h}{C}\nabla f(x^k).$$

Therefore this circuit **updates primal variable x.**

Inductor: Dual variable update



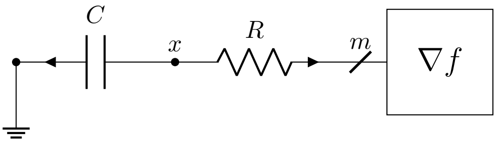
- V-I relation for nonlinear resistor: $y(t) = \nabla f(x(t))$
- V-I relation inductor: $v_{\mathcal{L}}(t) = D_{\mathcal{L}} \frac{d}{dt} i_{\mathcal{L}}(t)$
- From KCL: $v_{\mathcal{L}}(t) - x(t) = 0$
- From KVL: $i_{\mathcal{L}}(t) + y(t) = 0$

For simplicity, let $D_{\mathcal{L}} = LI$ and $(\nabla f)^{-1}$ exists. Combining all together, we obtain:

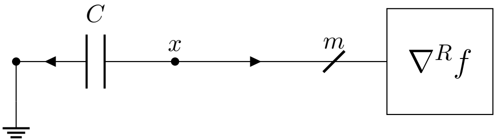
$$\frac{d}{dt} y(t) = -\frac{1}{L} x(t) = (\nabla f)^{-1}(y(t)) \quad \longrightarrow \quad y^{k+1} = y^k - \frac{h}{L} (\nabla f)^{-1}(y^k).$$

Therefore this circuit **updates dual variable** y**.**

Circuits for classical algorithms: Proximal Point Method



Considering the equivalent resistor, above circuit is equivalent to



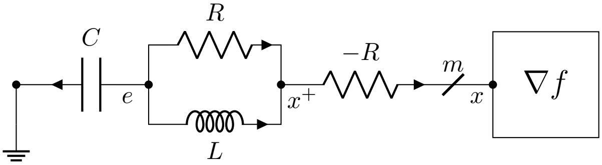
Thus the ODE becomes gradient flow of Moreau envelope ${}^{R}f$

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

Applying Euler discretization with a stepsize of CR:

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

Circuits for classical algorithms: Nesterov acceleration



Ohm's law for negative resistor, we have $x^+ = x - R\nabla f(x)$. Other V-I relations are:

$$\begin{aligned} \frac{d}{dt} i_{\mathcal{L}} &= D_{\mathcal{L}}^{-1}(v_{\mathcal{C}} - x^+) \\ \frac{d}{dt} v_{\mathcal{C}} &= -D_{\mathcal{C}}^{-1} \nabla f(x). \end{aligned}$$

Combining, we get high-resolution ODE[9] of Nesterov acceleration:

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

[9] Bin Shi et al. "Understanding the Acceleration Phenomenon via High-Resolution Differential Equations". In: *Mathematical Programming* (2021).

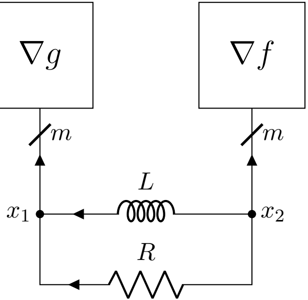
Circuits for classical algorithms: Douglas–Rachford splitting (DRS)

- V-I relations

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

- Douglas–Rachford splitting

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



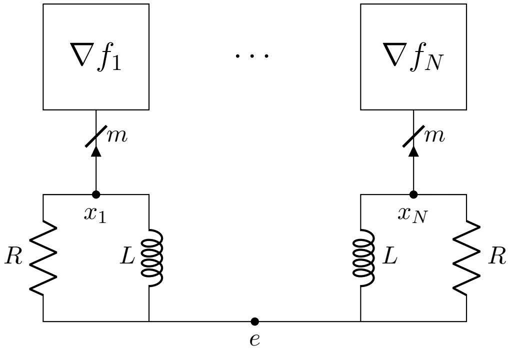
Circuits for classical algorithms: Proximal decomposition

- V-I relations

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

- Proximal decomposition

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



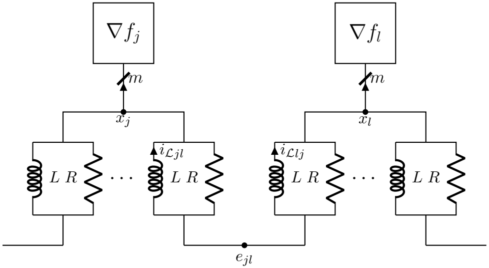
Circuits for classical algorithms: DADMM

- V-I relations

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

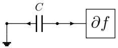
- Decentralized ADMM (DADMM)

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

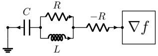


Zoo of Electric Circuits for Optimization Algorithms

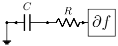
Gradient Descent



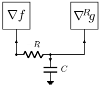
Nesterov acceleration with $R = \sqrt{L/C}$



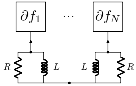
Proximal point method



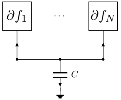
Prox-grad method



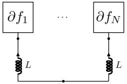
Prox Decomposition



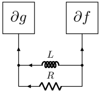
Primal Decomposition



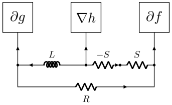
Dual Decomposition



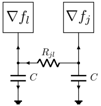
Douglas–Rachford splitting



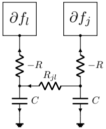
Davis–Yin splitting



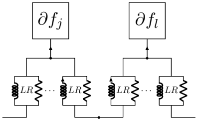
Decentralized GD



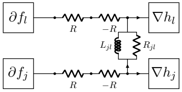
Adapt-then-combine



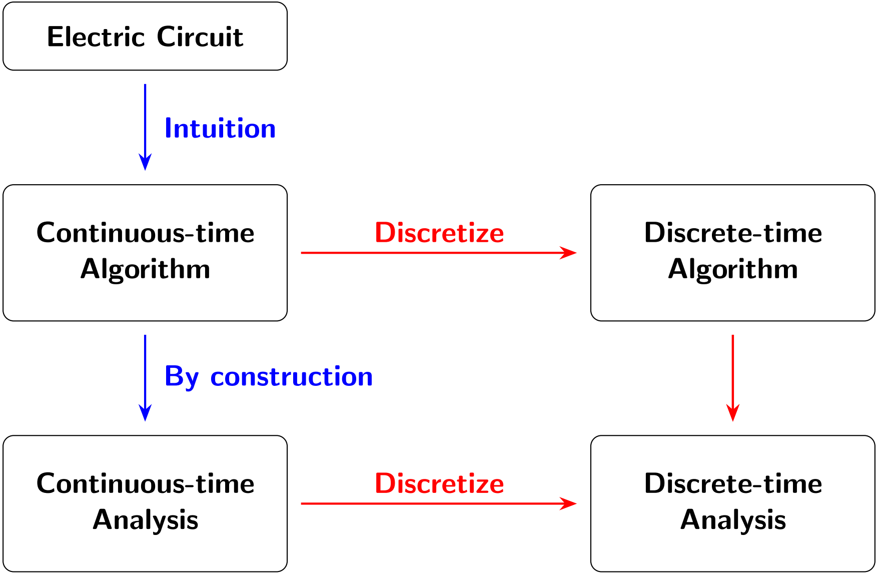
Decentralized ADMM



PG-EXTRA



Outline diagram



Outline

Challenging point on discretization

Some might wonder:

"Isn't discretization already a solved problem?"

However:

- Discretization studies in numerical analysis, focus on convergence of x^k to the solution **curve** $x(t)$ as step size goes to zero
- But this is not the interest in optimization;
  we are interested in convergences like $\lim_{k\to\infty} x^k = x^\star$, $\lim_{k\to\infty} f(x^k) = f(x^\star)$
- We want to find discretized algorithm that guarantees such convergence.
  In this context, **finding discretization involves finding convergence proof.**
- We aim to find discretization that **preserves the proof structure.**

Continuous-time analysis (Review)

Theorem 2.2. (Informal)
Let $(v(t), i(t), x(t), y(t))$ be a curve satisfying the ODE. Let $(x^\star, y^\star)$ be a primal-dual optimal solution pair. Then

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

Proof outline.
Define the energy of the circuit at time t as

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

Consider energy dissipation

$$\frac{d}{dt}\mathcal{E} + \langle x - x^\star, y - y^\star\rangle \leq 0.$$

From integrability, we have $\lim_{t\to\infty} \langle x - x^\star, y - y^\star\rangle = 0$, concluding the result. $\square$

Discrete-time analysis (Goal)

Lemma 4.1. (Informal)
Let $\{(v^k, i^k, x^k, y^k)\}_{k\in\mathbb{N}}$ be a sequence that is *sufficient dissipative*, *i.e.*, there is $\eta > 0$ that satisfies (2). Let $(x^\star, y^\star)$ be a primal-dual optimal solution pair. Then

$$\lim_{k\to\infty} x^k = x^\star.$$

Proof outline.
Define the energy as

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

From *sufficient dissipativity* condition, there is $\eta > 0$ such that

$$\mathcal{E}_{k+1} - \mathcal{E}_k + \eta \left\langle x^k - x^\star, y^k - y^\star \right\rangle \leq 0. \tag{2}$$

From summability, we have $\lim_{k\to\infty} \langle x^k - x^\star, y^k - y^\star\rangle = 0$, concluding the result. $\square$

Computer-assisted discretization (informal)

(i) Considering 2-stage Runge-Kutta method, express the algorithm as

$$(v^{k+1}, i^{k+1}, x^{k+1}, y^{k+1}) = T_{\alpha,\beta,h}(v^k, i^k, x^k, y^k) \tag{3}$$

with some linear function $T_{\alpha,\beta,h}$. We want to find step size parameters α, β, h.

(ii) Consider the optimization problem

$$\begin{aligned} \underset{\eta,\alpha,\beta,h}{\text{maximize}} \quad & \eta \\ \text{subject to} \quad & \mathcal{E}_{k+1} - \mathcal{E}_k + \eta\langle x^k - x^\star, y^k - y^\star\rangle \le 0 \\ & \left\{(v^k, i^k, x^k, y^k)\right\}_{k\in\mathbb{N}} \text{ is generated by (3)} \end{aligned}$$

(iii) Leverage formulation from computer-assisted proof technique known as PEP[10]. **Reformulate to a problem can be solved by numerical solver**.

(iv) By running the solver, find α, β, h, η with $\eta > 0$

[10] Performance Estimation Problem, refer [5, 7, 2]

Computer-assisted discretization (informal)

(iii) Leverage formulation from computer-assisted proof technique, PEP[11][12][13].
Reformulate to a problem can be solved by numerical solver.

(iv) By running the solver, find α, β, h, η with $\eta > 0$

Plugging the numerical values, under certain numerical precision, we obtain a new discretized algorithm that guaranteed to converge!

---

[11] Yoel Drori and Marc Teboulle. "Performance of first-order methods for smooth convex minimization: A novel approach". In: *Mathematical Programming* 145.1-2 (2014), pp. 451–482.

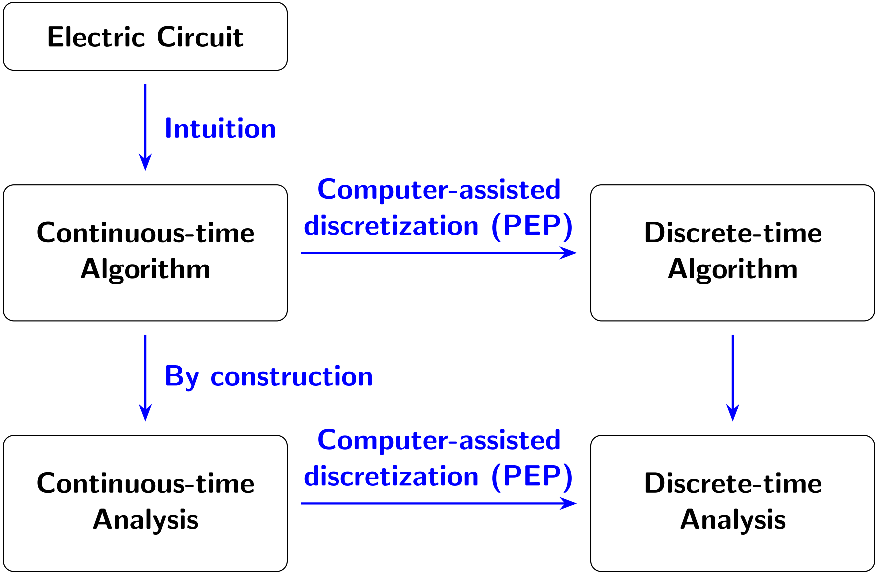
[12] A. Taylor, J. Hendrickx, and F. Glineur. "Smooth strongly convex interpolation and exact worst-case performance of first-order methods". In: *Mathematical Programming* 161 (2017), pp. 307–345.

[13] Shuvomoy Das Gupta, Bart P. G. Van Parys, and Ernest K. Ryu. "Branch-and-Bound Performance Estimation Programming: A Unified Methodology for Constructing Optimal Optimization Methods". In: *Mathematical Programming* (2023).

Novelty of our discretization

- Previous discretization studies can be divided into two categories
  - special rules tailored to the specific dynamics
  - apply standard discretization schemes or their variants
- Our discretization methodology is novel
  - aim to find parameters that preserve the proof structure
  - find such parameters automatically by leveraging PEP
- Automate using computer-assisted proof technique PEP
  - open-source package `ciropt`:
    https://github.com/cvxgrp/optimization_via_circuits

This completes our framework



Outline

Material 1: DADMM

- V-I relations

$$x_j = \mathbf{prox}_{(R/|\Gamma_j|)f_j}\left(\frac{1}{|\Gamma_j|}\sum_{l\in\Gamma_j}(Ri_{\mathcal{L}jl} + e_{jl})\right)$$

$$e_{jl} = \frac{1}{2}(x_j + x_l)$$

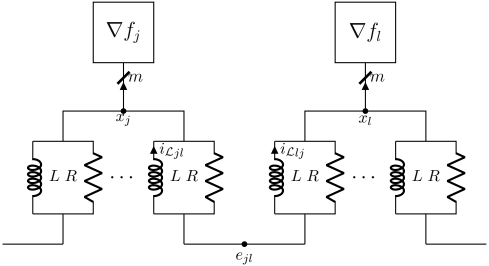
$$\frac{d}{dt}i_{\mathcal{L}jl} = \frac{1}{L}(e_{jl} - x_j)$$

- Decentralized ADMM (DADMM)

$$x_j^{k+1} = \mathbf{prox}_{(R/|\Gamma_j|)f_j}\left(\frac{1}{|\Gamma_j|}\sum_{l\in\Gamma_j}(Ri_{\mathcal{L}jl}^k + e_{jl}^k)\right)$$

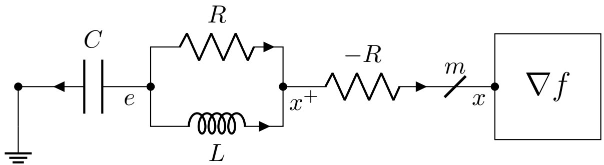
$$e_{jl}^{k+1} = \frac{1}{2}(x_j^{k+1} + x_l^{k+1})$$

$$i_{\mathcal{L}jl}^{k+1} = i_{\mathcal{L}jl}^k + \frac{1}{R}(e_{jl}^{k+1} - x_j^{k+1})$$



Q. What if only two functions, say f_4 and f_5, are known to be strongly convex?

Material 2: Nesterov acceleration

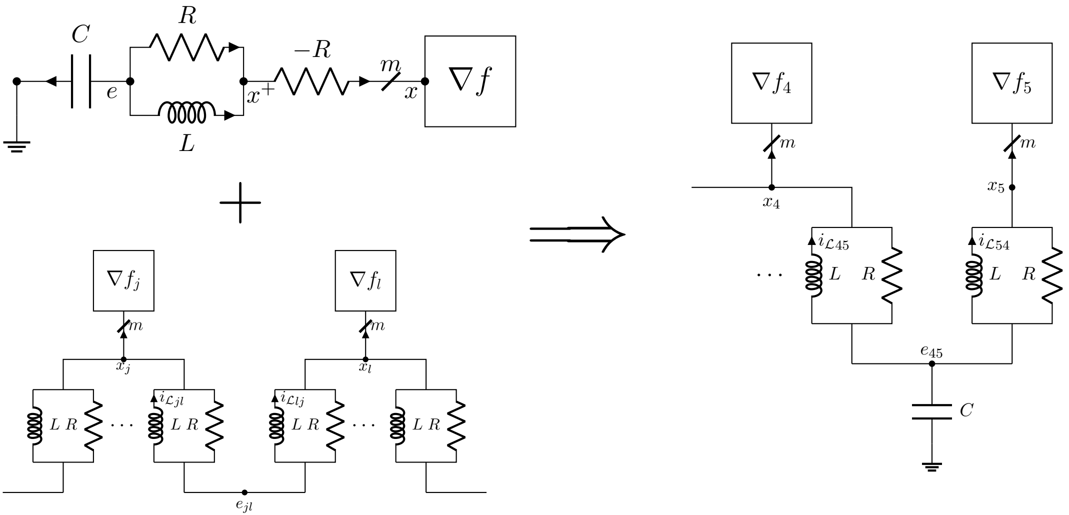


Recall, above circuit corresponds to high-resolution Nesterov ODE:

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

Nesterov is known to work well for strongly convex functions.

DADMM+C: DADMM + Nesterov



DADMM+C

We attach capacitor on e_{45}:

$$\begin{aligned} x_j^{k+1} &= \mathbf{prox}_{(R/|\Gamma_j|)f_j}\left(\frac{1}{|\Gamma_j|}\sum_{l\in\Gamma_j}(Ri_{\mathcal{L}jl}^k + e_{jl}^k)\right) \\ e_{jl}^{k+1} &= \begin{cases} e_{45}^k - \frac{h}{CR}\left(R(i_{\mathcal{L}45}^k + i_{\mathcal{L}54}^k) + 2e_{45}^k - x_4^{k+1} - x_5^{k+1}\right) & \{j,l\} = \{4,5\} \\ \frac{1}{2}(x_j^{k+1} + x_l^{k+1}) & \text{otherwise} \end{cases} \\ i_{\mathcal{L}jl}^{k+1} &= i_{\mathcal{L}jl}^k + \frac{h}{L}(e_{jl}^{k+1} - x_j^{k+1}). \end{aligned}$$

We consider $R = 0.8$, $L = 2$, $C = 15$. Running our package `ciropt`, we obtain:

```
{'eta': 3.7032579140049147,
 'gamma': 4.477637915336295,
 'h': 3.515416990114162,
 'rho': 9.889087616006122e-10}
```

This suggest us $h = 3.52$.

DADMM+C

Recall:

```
{'eta': 3.7032579140049147,
 'gamma': 4.477637915336295,
 'h': 3.515416990114162,
 'rho': 9.889087616006122e-10}
```

The numerical values imply, for the energy function

$$\mathcal{E}_k = \sum_{(j,l)\in A} \left\| i^k_{\mathcal{L}_{jl}} - i^\star_{\mathcal{L}_{jl}} \right\|^2 + \sum_{j<l,\{j,l\}\subset S} \frac{15}{2} \left\| e^k_{jl} - x^\star_j \right\|^2 + 3.52 \sum_{j<l,\{j,l\}\not\subset S} \left\| e^k_{jl} - x^\star_j \right\|^2,$$

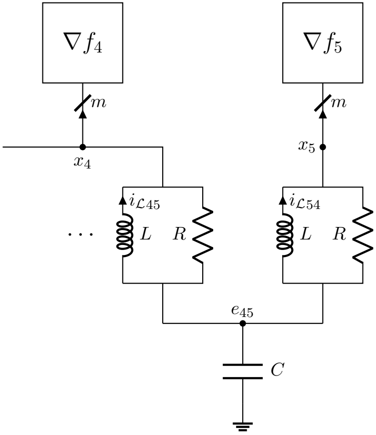
following inequality is true up to certain numerical precision

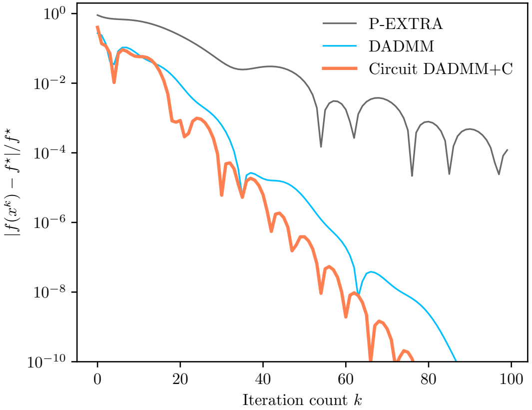
$$\left(\mathcal{E}_{k+1} + 3.7\langle x^{k+1} - x^\star, y^{k+1} - y^\star \rangle \right) - \mathcal{E}_k \leq 0.$$

This inequality guarantees the convergence.

Numerical results: DADMM+C

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





Furthermore: Genearlized DADMM+C

Previous instance further motivates generalized version of algorithm and proof:

Lemma I.1. (Informal)
Let $f_j\colon \mathbf{R}^m \to \mathbf{R}$ are convex functions for $j \in \{1,\ldots,N\}$ and $S \subset \{1,\ldots,N\}$.
Consider the generalized DADMM+C

$$
\begin{aligned}
x_j^{k+1} &= \mathbf{prox}_{(R/|\Gamma_j|)f_j}\left(\frac{1}{|\Gamma_j|}\sum_{l\in\Gamma_j}(Ri_{\mathcal{L}jl}^k + e_{jl}^k)\right) \\
e_{jl}^{k+1} &= \begin{cases} e_{jl}^k - \frac{h}{CR}\left(R(i_{\mathcal{L}jl}^k + i_{\mathcal{L}lj}^k) + 2e_{jl}^k - x_j^{k+1} - x_l^{k+1}\right) & \{j,l\}\subset S \\ \frac{1}{2}(x_j^{k+1} + x_l^{k+1}) & \text{otherwise} \end{cases} \\
i_{\mathcal{L}jl}^{k+1} &= i_{\mathcal{L}jl}^k + \frac{h}{L}(e_{jl}^{k+1} - x_j^{k+1}).
\end{aligned}
$$

Furthermore: Genearlized DADMM+C (continue)

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

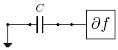
$$\max\left\{1, \frac{2h}{CR}\right\} \le \tau^2 \le 2 - \frac{hR}{L},$$

following inequality is true

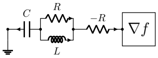
$$\left(\mathcal{E}_{k+1} + \frac{h}{2R}\left(2 - \frac{hR}{L} - \tau^2\right) \sum_{(j,l)\in A} \left\| e^{k+1}_{jl} - x^{k+1}_j \right\|^2 + h\langle x^{k+1} - x^\star, y^{k+1} - y^\star\rangle\right) - \mathcal{E}_k \le 0.$$

Zoo of Electric Circuits for Optimization Algorithms

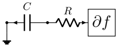
Gradient Descent



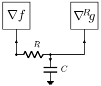
Nesterov acceleration with $R = \sqrt{L/C}$



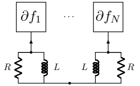
Proximal point method



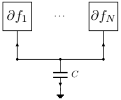
Prox-grad method



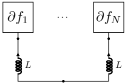
Prox Decomposition



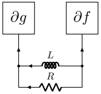
Primal Decomposition



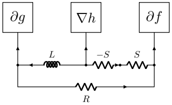
Dual Decomposition



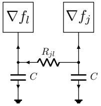
Douglas–Rachford splitting



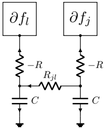
Davis–Yin splitting



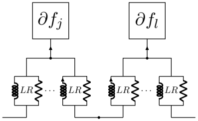
Decentralized GD



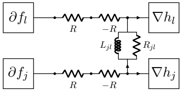
Adapt-then-combine



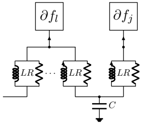
Decentralized ADMM



PG-EXTRA

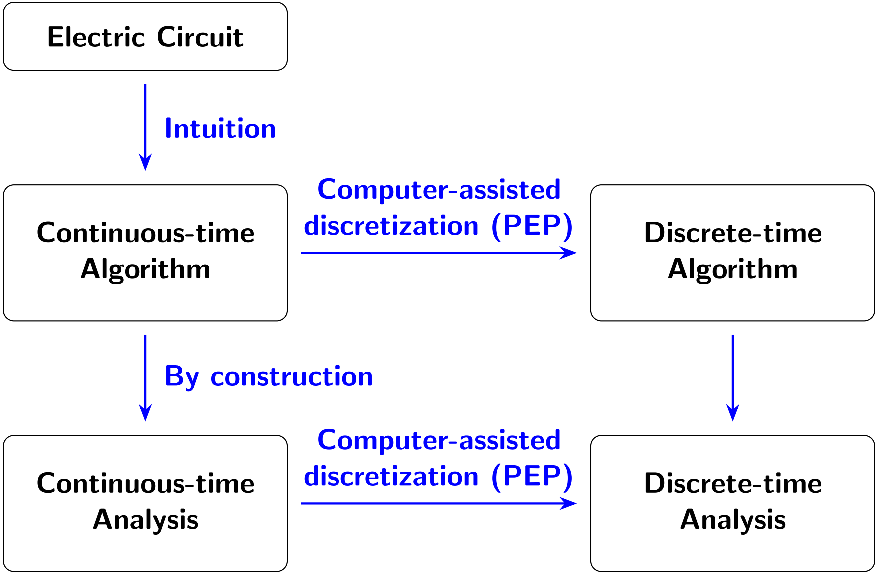


DADMM+C (new!)

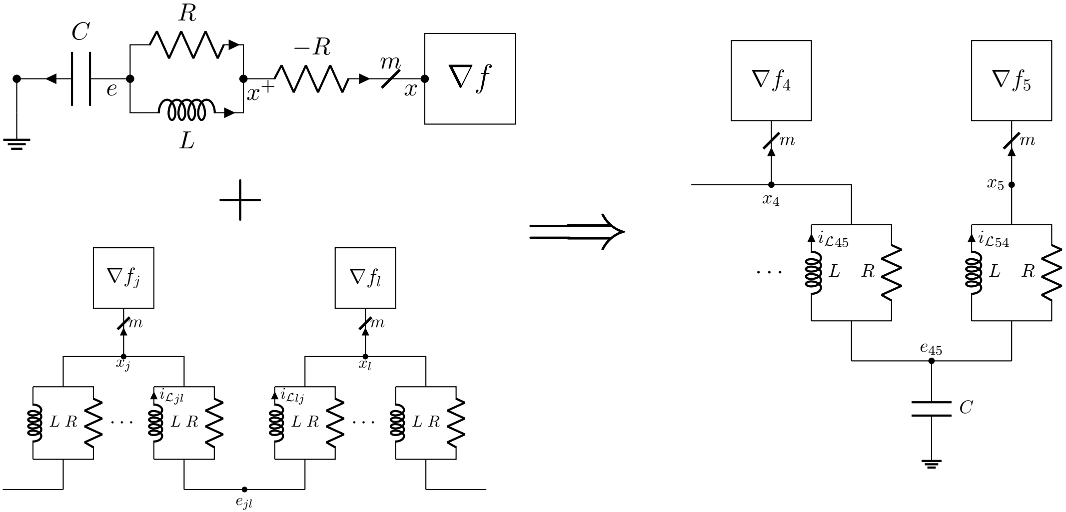


Outline

Summary: a novel framework to design a new algorithm

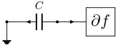


A new algorithm leveraging the framework: DADMM+C

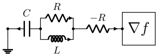


We expect follow-up works will verify potential of our framework!

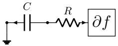
Gradient Descent



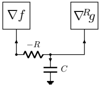
Nesterov acceleration with $R = \sqrt{L/C}$



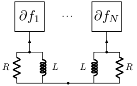
Proximal point method



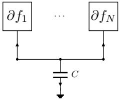
Prox-grad method



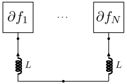
Prox Decomposition



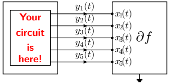
Primal Decomposition



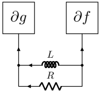
Dual Decomposition



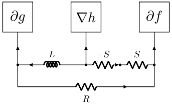
Your new algorithm



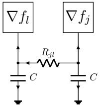
Douglas–Rachford splitting



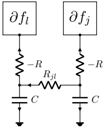
Davis–Yin splitting



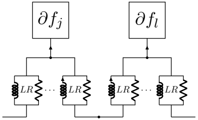
Decentralized GD



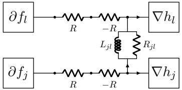
Adapt-then-combine



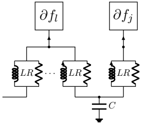
Decentralized ADMM



PG-EXTRA



DADMM+C (new!)



Conclusion

- Introduce a framework for designing optimization algorithms via RLC circuits
  - design dynamic circuit that converges to the solution
  - discretize to obtain convergent algorithm
- Provide electric circuits for standard methods:
  - Nesterov acceleration, proximal point method, prox-gradient, primal decomposition, dual decomposition, DYS, DRS, decentralized gradient descent, diffusion, DADMM and PG-EXTRA
- Convergence proof of circuit dynamics based on energy dissipation
- Complete the framework by automating discretization with PEP
  - find parameters preserve the proof structure
  - open-source package `ciropt`
    https://github.com/cvxgrp/optimization_via_circuits

References I

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

[2] Shuvomoy Das Gupta, Bart P. G. Van Parys, and Ernest K. Ryu. "Branch-and-Bound Performance Estimation Programming: A Unified Methodology for Constructing Optimal Optimization Methods". In: *Mathematical Programming* (2023).

[3] Jack Bonnell Dennis. "Mathematical Programming and Electrical Networks". PhD thesis. Massachusetts Institute of Technology, 1959.

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

[5] Yoel Drori and Marc Teboulle. "Performance of first-order methods for smooth convex minimization: A novel approach". In: *Mathematical Programming* 145.1-2 (2014), pp. 451–482.

References II

[6] Bin Shi et al. "Understanding the Acceleration Phenomenon via High-Resolution Differential Equations". In: *Mathematical Programming* (2021).

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