Optimization Algorithm Design via Electric Circuits

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

Distributed convex optimization problem

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

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

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

Example: Consensus problem

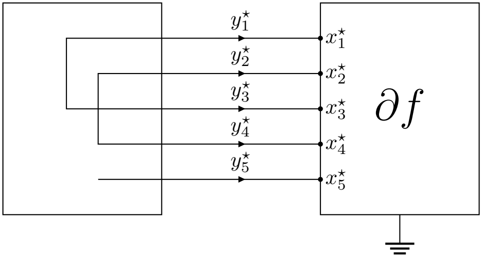
$$\begin{array}{ll} \underset{x_1,\ldots,x_N \in \mathbf{R}^{m/N}}{\text{minimize}} & f_1(x_1) + \cdots + f_N(x_N) \\ \text{subject to} & x_1 = \cdots = x_N \end{array}$$

- $x = (x_1, \ldots, x_N) \in \mathbf{R}^m$ is the decision variable
- $f(x) = f_1(x_1) + \cdots + f_N(x_N)$ is block-separable
- $E^\mathsf{T} = (I, \ldots, I) \in \mathbf{R}^{m \times m/N}$

Circuit interpretation: KKT conditions

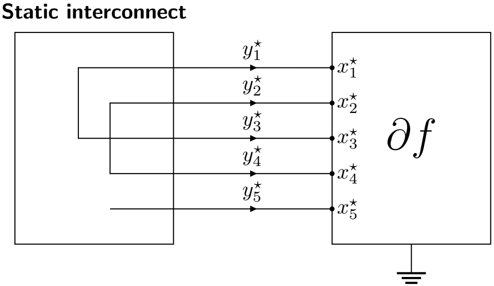
$$
\begin{array}{lcll}
y & \in & \partial f(x) & \text{(stationarity)} \\
x & \in & \mathcal{R}(E^\mathsf{T}) & \text{(primal feasibility)} \\
y & \in & \mathcal{N}(E) & \text{(dual feasibility)}
\end{array}
$$

Static interconnect



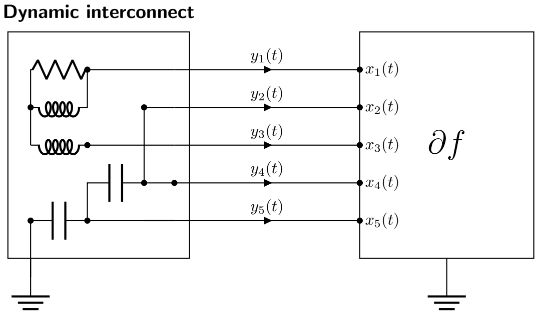
Circuit interpretation: KKT conditions

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



Circuit interpretation: Dynamic interconnect

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



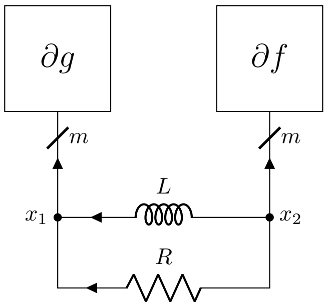
Circuits for classical algorithms: DRS

- 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

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



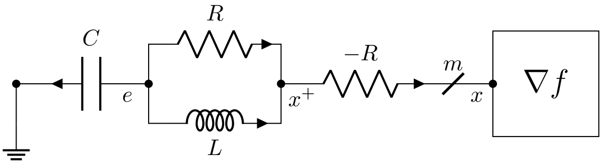
Circuits for classical algorithms: Nesterov acceleration

- V-I relations

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

- Nesterov acceleration

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



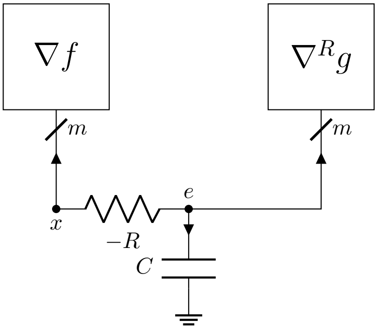
Circuits for classical algorithms: Proximal gradient

- V-I relations

$$
\begin{aligned}
i_{\mathcal{C}} &= -\nabla f(x) - \nabla^R g(e) \\
v_{\mathcal{C}} &= x - R\nabla f(x)
\end{aligned}
$$

- Proximal gradient method

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



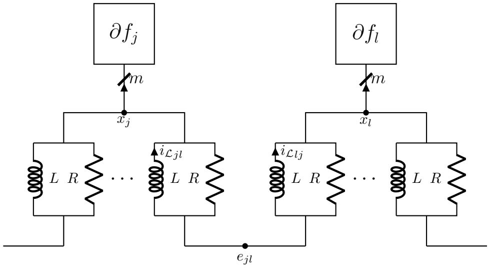
Circuits for classical algorithms: DADMM

- 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

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



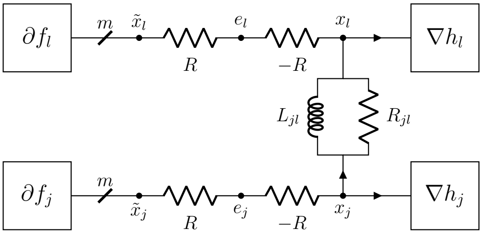
Circuits for classical algorithms: PG-EXTRA

- V-I relations

$$
\begin{aligned}
x_j &= \mathbf{prox}_{Rf_j}\left(\sum_{l=1}^{N} W_{jl}x_l - R\nabla h_j(x_j) - w_j\right) \\
\frac{d}{dt}w_j &= x_j - \sum_{l=1}^{N} W_{jl}x_l
\end{aligned}
$$

- PG-EXTRA

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



Energy dissipation

- in continuous time, energy dissipation leads to convergence (Thm 2.2)
  - $\mathcal{E}(t) = \frac{1}{2}\|v_{\mathcal{C}}(t) - v_{\mathcal{C}}^{\star}\|_{D_{\mathcal{C}}}^2 + \frac{1}{2}\|i_{\mathcal{L}}(t) - i_{\mathcal{L}}^{\star}\|_{D_{\mathcal{L}}}^2$
  - $\frac{d}{dt}\mathcal{E} \le -\langle x(t) - x^\star, y(t) - y^\star \rangle \le 0$
  - $\lim_{t\to\infty} x(t) = x^\star$
- not every discretization leads to a convergent algorithm

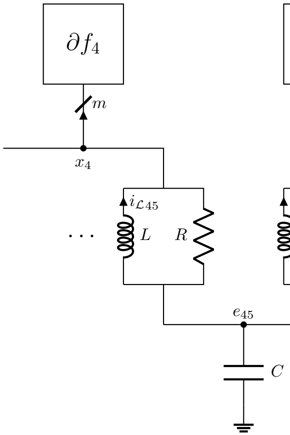
Automatic discretization

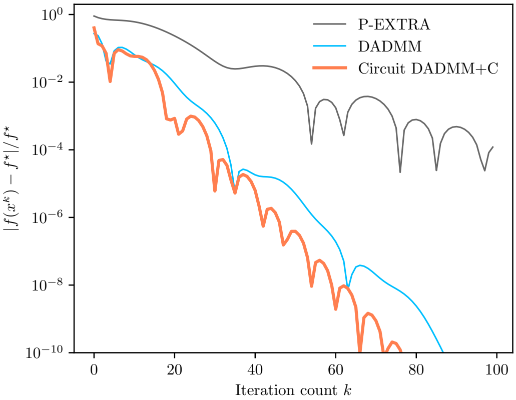
- find discretization preserving the proof structure (Lemma 4.1)
  - $\mathcal{E}_k = \frac{1}{2}\|v_{\mathcal{C}}^k - v_{\mathcal{C}}^\star\|_{D_{\mathcal{C}}}^2 + \frac{1}{2}\|i_{\mathcal{L}}^k - i_{\mathcal{L}}^\star\|_{D_{\mathcal{L}}}^2$
  - $\mathcal{E}_{k+1} - \mathcal{E}_k + \eta\langle x^k - x^\star, y^k - y^\star\rangle \le 0$ for some $\eta > 0$
  - $\lim_{k\to\infty} x^k = x^\star$
- automate using computer-assisted proof framework PEP
  - open-source package `ciropt`:
    https://github.com/cvxgrp/optimization_via_circuits

Previous discretizations

- 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

Numerical results: DADMM+C





Contributions

- introduce a framework for designing optimization algorithms via RLC circuits
  - design dynamic circuit that converges to the solution
  - discretize to obtain convergent algorithm
- 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
- PEP-based automated discretization that preserves proof structure
  - open-source package `ciropt`
    https://github.com/cvxgrp/optimization_via_circuits