A coordinate gradient descent method for nonsmooth separable minimization

Paul Tseng and Sangwoon Yun, 2009
Presented by Matt Wytock
MLD Journal Club

February 6, 2014

Problem

$$\text{minimize} \quad \underbrace{f(x)}_{\text{smooth}} + \underbrace{g(x)}_{\text{nonsmooth}}$$

Problem

$$\text{minimize} \quad \underbrace{f(x)}_{\text{smooth}} \quad + \quad \lambda\|x\|_1$$

Problem

$$\text{minimize} \quad \underbrace{f(x)}_{\text{smooth}} \; + \; \lambda \|x\|_1$$

State-of-the-art approach for sparse inverse covariance estimation, QUIC algorithm (Hsieh et al. 2011)

Newton-Lasso method

$$\text{minimize } f(x) + \lambda \|x\|_1$$

Newton-Lasso method

$$\text{minimize}\, f(x) + \lambda \|x\|_1$$

Repeat

1. Form the second-order Taylor expansion

$$\hat{f}(x + \Delta) = f(x) + \nabla f(x)^T \Delta + \frac{1}{2} \Delta^T \nabla^2 f(x) \Delta$$

Newton-Lasso method

$$\text{minimize } f(x) + \lambda \|x\|_1$$

Repeat

1. Form the second-order Taylor expansion

$$\hat{f}(x+\Delta) = f(x) + \nabla f(x)^T \Delta + \frac{1}{2}\Delta^T \nabla^2 f(x) \Delta$$

2. Solve for the regularized Newton step

$$d = \arg\min_{\Delta} \hat{f}(x+\Delta) + \lambda \|x+\Delta\|_1$$

Newton-Lasso method

$$\text{minimize } f(x) + \lambda \|x\|_1$$

Repeat

1. Form the second-order Taylor expansion

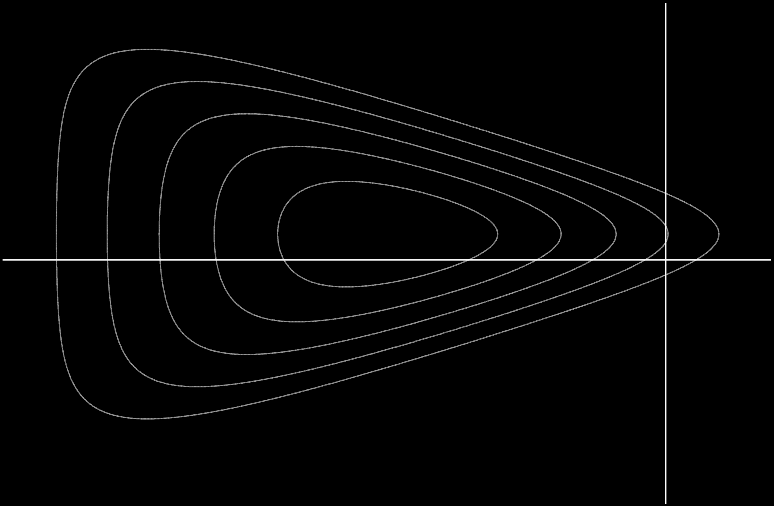
$$\hat{f}(x + \Delta) = f(x) + \nabla f(x)^T \Delta + \frac{1}{2} \Delta^T \nabla^2 f(x) \Delta$$

2. Solve for the regularized Newton step

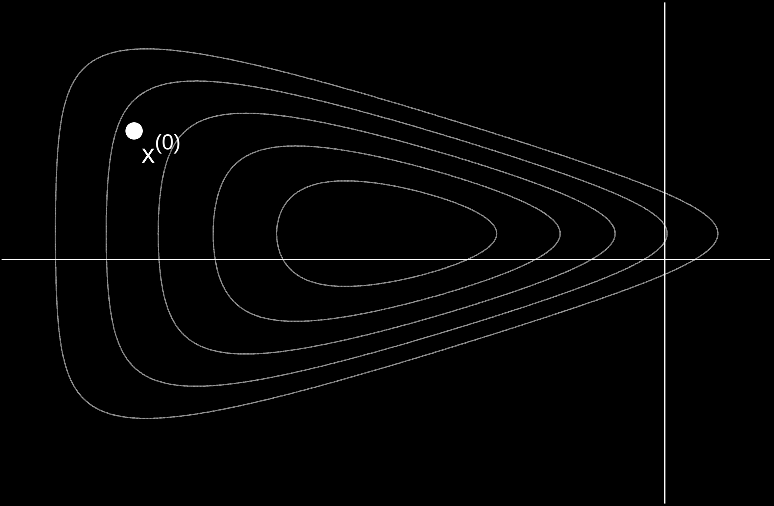
$$d = \arg\min_{\Delta} \hat{f}(x + \Delta) + \lambda \|x + \Delta\|_1$$

3. Update x using backtracking line search

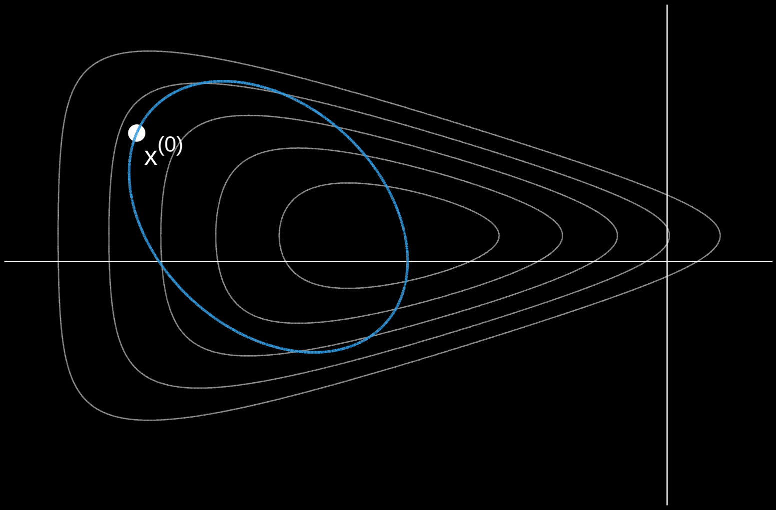
Newton-Lasso example



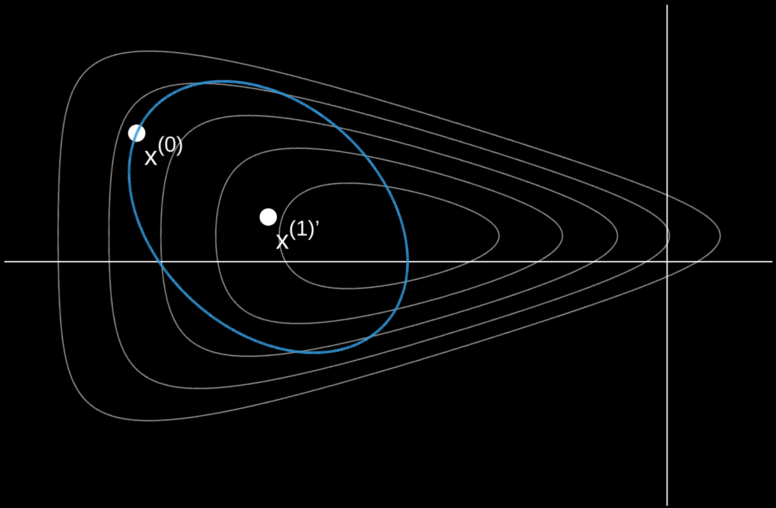
Newton-Lasso example



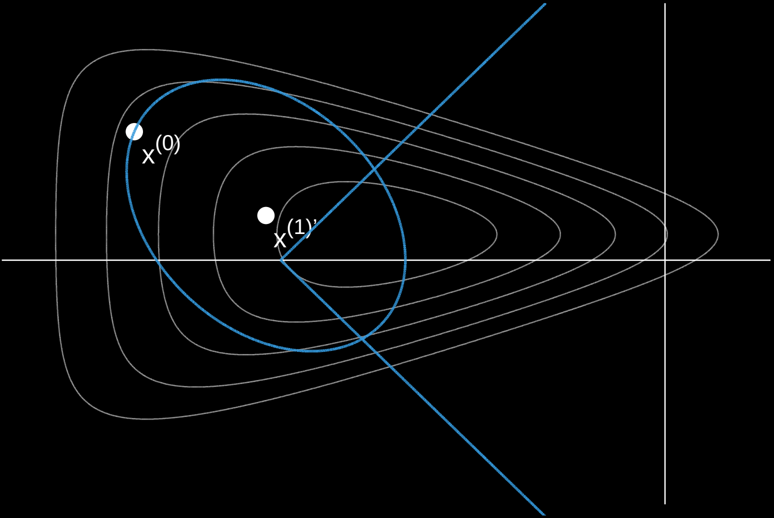
Newton-Lasso example



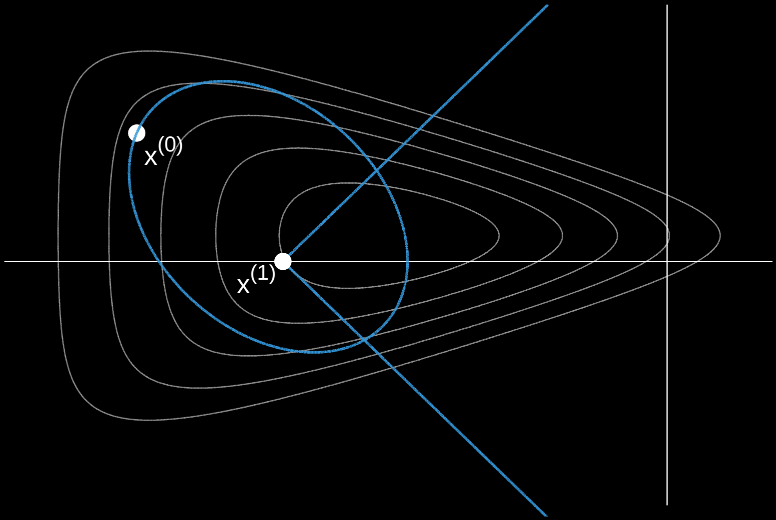
Newton-Lasso example



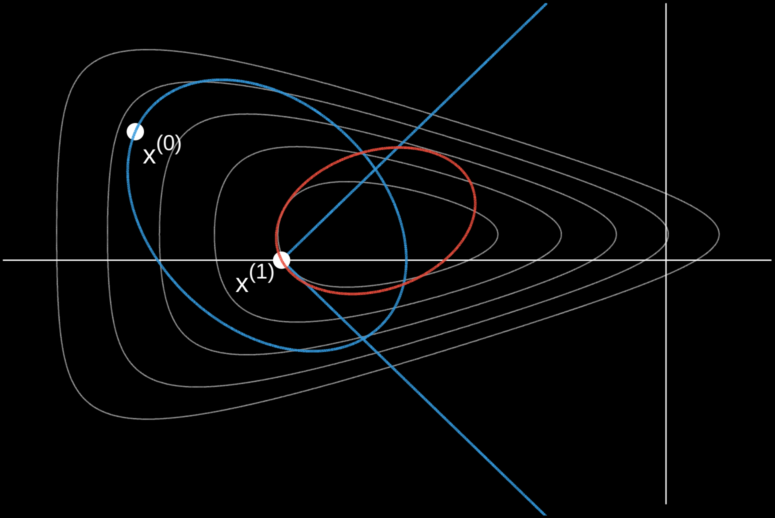
Newton-Lasso example



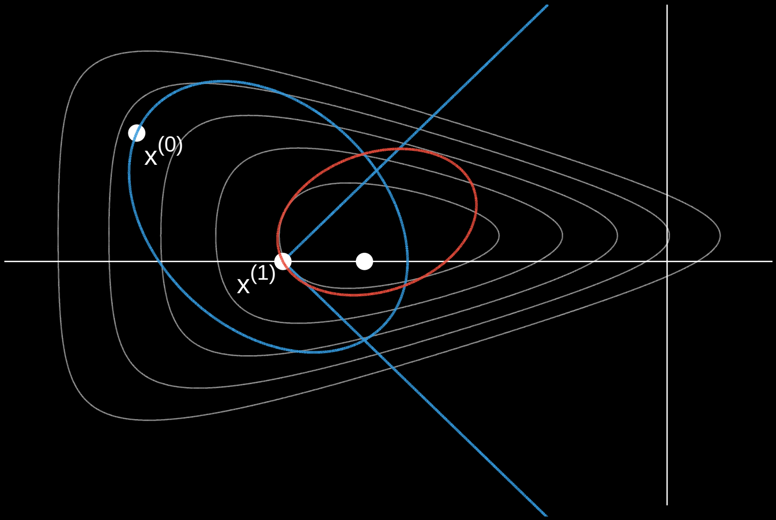
Newton-Lasso example



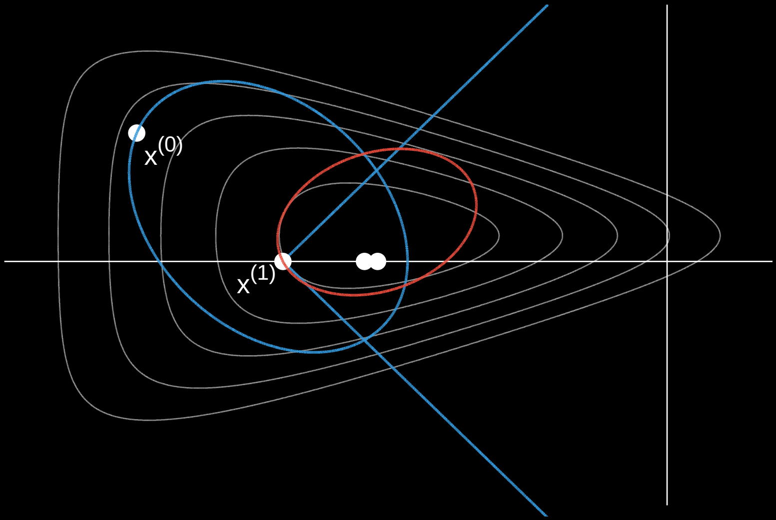
Newton-Lasso example



Newton-Lasso example



Newton-Lasso example



Computational complexity of inner loop

$$d = \arg\min_{\Delta} \hat{f}(x + \Delta) + \lambda \|x + \Delta\|_1$$

Computational complexity of inner loop

$$d = \arg\min_{\Delta} \hat{f}(x + \Delta) + \lambda \|x + \Delta\|_1$$

Exploit sparsity using a coordinate descent active set method

1. Form active set by including coordinate i if

$$|\nabla f(x)_i| > \lambda \text{ or } x_i \neq 0$$

2. Iteratively minimize, done in closed form

Theoretical analysis

Theoretical analysis

It works.

Theoretical analysis (for real)

1. The objective $F = f + g$ is nonincreasing, $F > -\infty$
2. $F(x^{k+1}) - F(x^k) \leq \sigma \alpha^k \Delta^k \leq 0$
3. $\inf_k \alpha^k > 0$
4. $\Delta^k \to 0, d^k \to 0$ as $k \to \infty$
5. $\|x^k - x^\star\| \to 0$