Notes on CCCP and variational algorithms

Geoff Gordon

February 11, 2002

The paper

A. L. Yuille and Anand Rangarajan. *The Concave-Convex Procedure (CCCP)*. NIPS 2001.

Algorithm recipe for optimizing “arbitrary” functions

Recipe in same sense that EM is a recipe

Significance of CCCP

Mathematically, CCCP is a special case of known recipes involving variational (aka convex duality) bounds

But, CCCP has very different intuition (geometric v. algebraic)

This new intuition can be a big help

For example, it helped develop a provably-convergent algorithm for propagating beliefs in loopy graphs

Outline

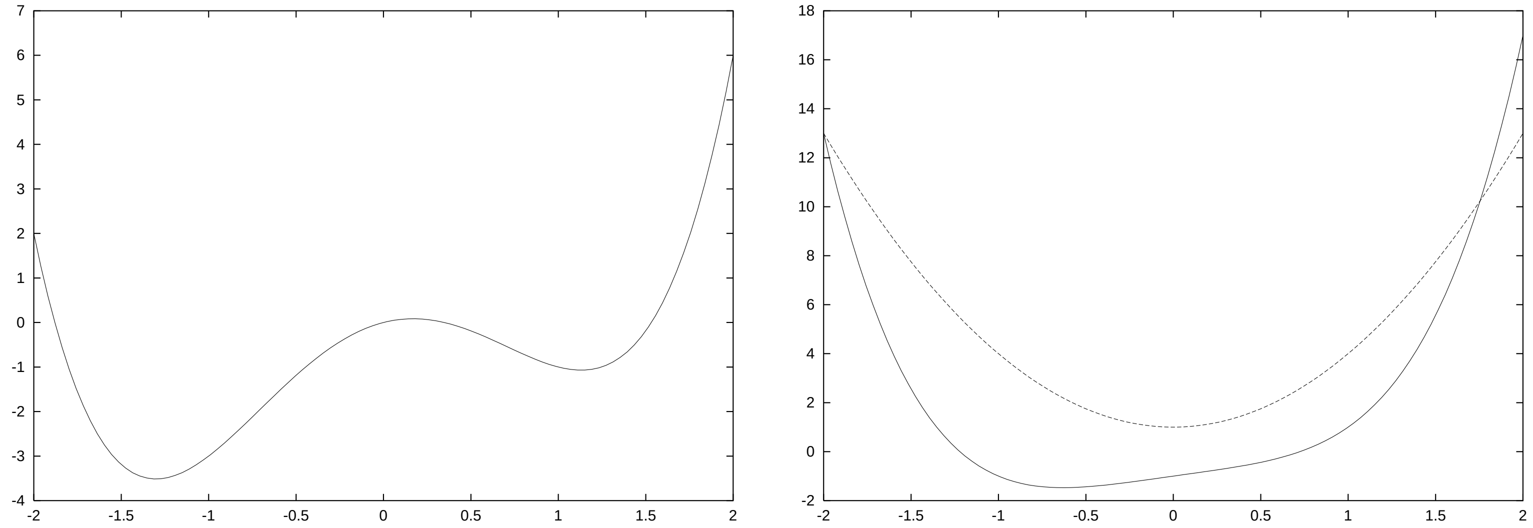
Geometric intuition

Algebraic intuition

Proofs

Discussion

CC decomposition



Any function is (convex part) + (concave part)

∞ many decompositions—hard part is finding a good one

CCCP recipe

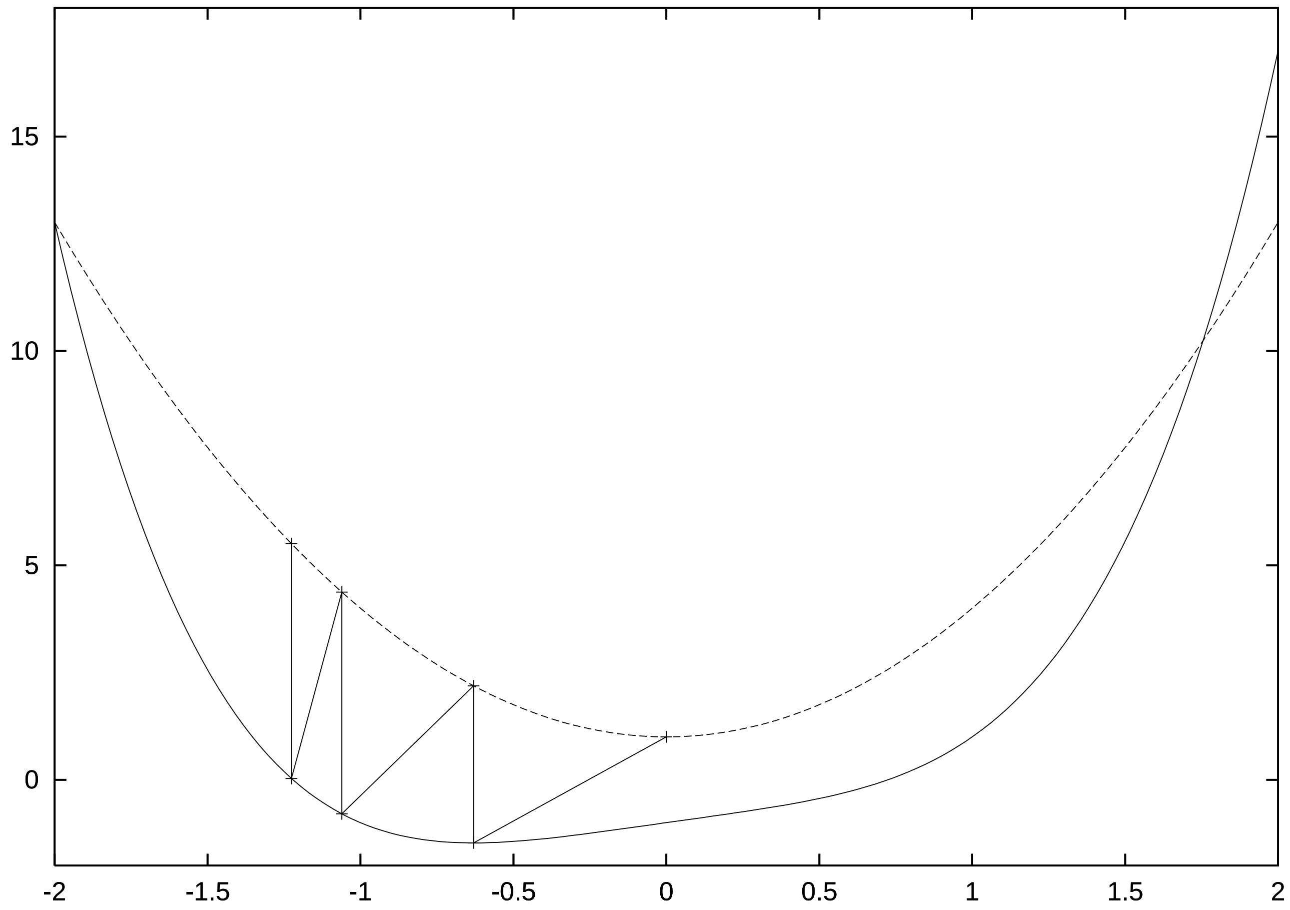
Match derivatives between concave and convex parts

To minimize $f(x) - g(x)$, repeat:

- Let $a = g'(x)$
- Find x_{new} so that $f'(x_{\text{new}}) = a$

More simply, $f'(x_{\text{new}}) = g'(x)$

CCCP picture



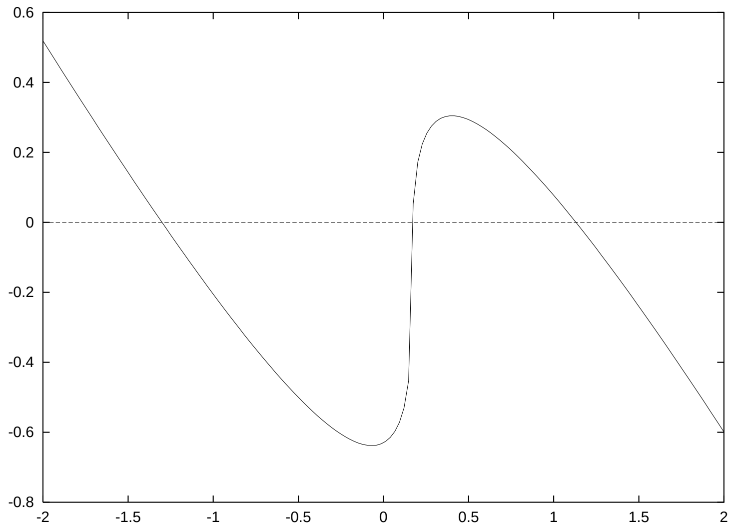
In our case

$$f(x) = x^4 + x - 1$$

$$g(x) = 3x^2 + 1$$

$$f'(x_{\text{new}}) = 4x_{\text{new}}^3 + 1 = 6x = g'(x)$$

$$x_{\text{new}} = \left(\frac{6x - 1}{4}\right)^{\frac{1}{3}}$$



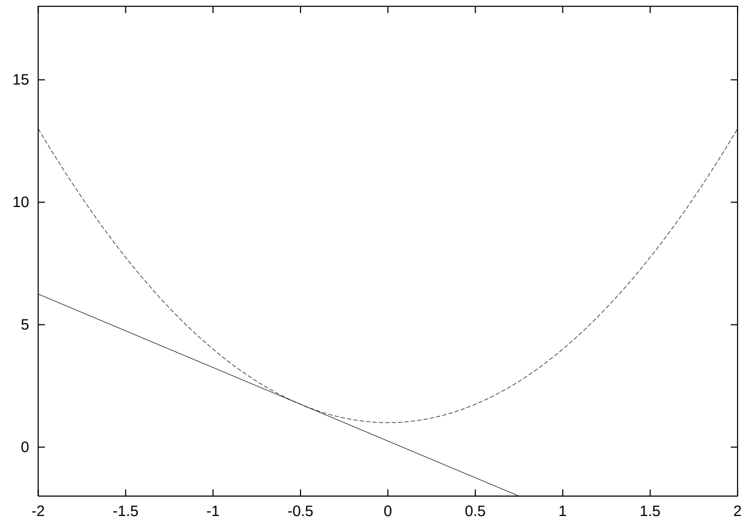
CCCP justification

If converged, $f'(x) = g'(x)$

So $(f - g)'(x) = 0$
$\Rightarrow$ min, max, or saddle

Convergence analysis below

The bounding argument



Since $g(x)$ is convex, it's lower-bounded by its tangent

$$g(x) \geq g(x_0) + (x - x_0)g'(x_0)$$

Will use this fact to develop a bound on $f - g$

Bounding math

Pick a tangent slope a, write $x_0 = (g')^{-1}(a)$

$$g(x) \geq g((g')^{-1}(a)) + (x - (g')^{-1}(a))a$$

Defining

$$g^*(a) = (g')^{-1}(a)a - g((g')^{-1}(a))$$

we have

$$g(x) \geq xa - g^*(a)$$

g^* is called *convex dual* of g

Properties of dual

g^* is convex, g^{**} is g (or convex hull of g)

Derivatives of duals are inverses:

$$a = g'(x) \Leftrightarrow x = (g^*)'(a)$$

Aside: dual exists for nondifferentiable fns as well

- Uses subgradient ∂g instead of derivative g'
- $\partial g(x)$ is set of possible tangent slopes at x
- above props true w/ gen. inverse $a \in \partial g(x) \Leftrightarrow x \in \partial g^*(a)$

Bounding algorithm

We have $g(x) \geq xa - g^*(a)$, so

$$f(x) - g(x) \leq f(x) - xa + g^*(a)$$

Minimize RHS alternately in x and a

Minimization in x is minimizing a bound on true objective

Minimization in a is adjusting bound to be as tight as possible

Convergence proof

Duality bound gives simple convergence proof

$$f(x) - g(x) \leq f(x) - xa + g^*(a)$$

Assume LHS is bounded below $\Rightarrow$ RHS is too

RHS is convex in x and in a

Minimize alternately, objective value must converge

Convergence of iterates

x and a might

- diverge
- converge to a local maximum or inflection point
- cycle at constant objective value

but most usual case is converging to a (local) minimum

Of course for more specific cases we could prove more

CCCP = bounding

In CCCP we maintain two x values, say x and x_{new}

In bounding algorithm we maintain an x value and a slope a

Minimize $f(x) - xa + g^*(a)$ wrt a:

$$\begin{aligned} 0 &= (g^*)'(a) - x \\ a &= g'(x) \end{aligned}$$

Minimize $f(x_{\text{new}}) - x_{\text{new}}a + g^*(a)$ wrt x_{new}:

$$f'(x_{\text{new}}) = a$$

Linear constraints

CCCP/bounding still works with linear constraints

Want to minimize $f(x) - g(x)$ st $Ax = b$

Write $I(x) = \begin{cases} 0 & \text{if } Ax = b \\ \infty & \text{o/w} \end{cases}$

Minimize $f(x) + I(x) - g(x)$

Linear constraints cont'd

We have $\partial I(x) = \begin{cases} \{Ay\} & \text{if } Ax = b \\ \{\} & \text{o/w} \end{cases}$

So derivative-matching turns into

$$\begin{aligned} f'(x_{\text{new}}) + Ay &= g'(x) \\ Ax_{\text{new}} &= b \end{aligned}$$

x_{new} is constrained, but derivatives don't have to match exactly

Examples of CCCP

Mean-field Potts model (= annealed mean-field Boltzmann machine)

Some EM algorithms

- paper gives example of elastic net algorithm for TSP
- it looks like you need a duality transform first (cf Neal & Hinton) before CCCP is relevant

Sinkhorn's algorithm for linear assignment problems

Generalized iterative scaling

Loopy BP

Loopy BP refs

A. L. Yuille. CCCP algorithms to minimize the Bethe and Kikuchi free energies: convergent alternatives to belief propagation. Submitted to Neural Computation.

Yedidia, Freeman, and Weiss. Generalized belief propagation. NIPS-13 (2000).

Max Welling, Yee Whye Teh. Belief Optimization for Binary Networks: A Stable Alternative to Loopy Belief Propagation. Proceedings of the Seventh Conference on Uncertainty in Artificial Intelligence (2001).

Yee Whye Teh, Max Welling. The Unified Propagation and Scaling Algorithm. NIPS-14 (2001).