# A Cartoon for the Concave-Convex Procedure

Yingzhen Li

Oct 2015

This short note provides an illustration of the concave-convex procedure (CCCP) [Yuille and Rangarajan, 2003]. I also briefly discuss the majorisation-minimisation (MM) algorithm, where CCCP is a special case.

## 1 Concave-convex procedure

Consider the case that we want to fit a model parametrized by $\theta$. Often we obtain the point estimate $\theta^*$ by minimising some *energy function*

$$\theta^* = \arg\min_{\theta} E(\theta). \tag{1}$$

For convex energy functions there exists a unique global optimum, which can be easily found. However for most interesting problems $E(\theta)$ is non-convex, and it remains a challenge on developing fast optimisation algorithms that is guaranteed to converged to a local optimum.

As non-convex optimisation is challenging in general, people have tried to simplify the problem and proposed algorithms on it. A special case that has been discussed in the literature is the *difference of convex functions programming* (DC programming): the energy function is constrained to be the difference between two convex functions:

$$E(\theta) = E_{vex}(\theta) + E_{cave}(\theta), \tag{2}$$

where $E_{vex}(\theta)$ and $-E_{cave}(\theta)$ denote two arbitrary convex functions. It is a weaker constraint than convexity as in the original CCCP paper [Yuille and Rangarajan, 2003] the author proved the following algorithm using linear algebra.

**Theorem 1** *For a twice-differentiable function $E(\theta)$, if the eigenvalues of the Hessian $\nabla^2 E(\theta)$ are lower-bounded, then there exist a convex function $E_{vex}(\theta)$ and a concave function $E_{cave}(\theta)$ where $E(\theta)$ can be decomposed as eq. (2).*

Remember for a twice-differentiable function to be convex, the eigenvalues of its Hessian should be lower-bounded by zero.

Next I show a cartoon illustration of the concave-convex procedure developed in the same paper. In one sentence, it computes the gradients of the concave and convex part separately, and obtains the next updates by matching the convex part gradient to the concave part gradient:

$$\theta_{t+1} \ satisfies \ \nabla E_{vex}(\theta_{t+1}) = -\nabla E_{cave}(\theta_t). \tag{3}$$

As shown in Figure 1, this procedure always decreases the energy function: $\Delta = L_{vex}(\theta) - L_{cave}(\theta)$ denotes the distance along the function value axis between the two lines that are tangent to the curve of $E_{cave}$ ($E_{vex}$) at $\theta_t$ ($\theta_{t+1}$), respectively, and from convexity we have $E(\theta_t) \geq \Delta$ ($E_{vex}$ is convex) and $\Delta \geq E(\theta_{t+1})$ ($-E_{cave}$ is convex). Since the functions in the form of (2) are lower-bounded, the CCCP procedure is guaranteed to converge to a local optimum.

## 2 Quick link to the majorisation-minimisation algorithm

In this section I show that CCCP is a special case of the *majorisation-minimisation* (MM) algorithm, which is a surrogate type optimisation method. Figure 2 illustrate the update procedure of the MM algorithm, where at each iteration we first find a surrogate objective $E_t(\theta)$ that majorises the original objective at the current solution $\theta_t$, then apply any optimisation algorithm to the surrogate for the next update $\theta_{t+1}$. To see why this procedure also guarantees non-increasing energy, we first formulate the definition of majorisation:

**Definition 1** *A function $E'(\theta)$ is said to majorise another function $E(\theta)$ at location $\theta'$, if for all $\theta$ we have $E'(\theta) \geq E(\theta)$ and the equality is achieved at $\theta'$.*

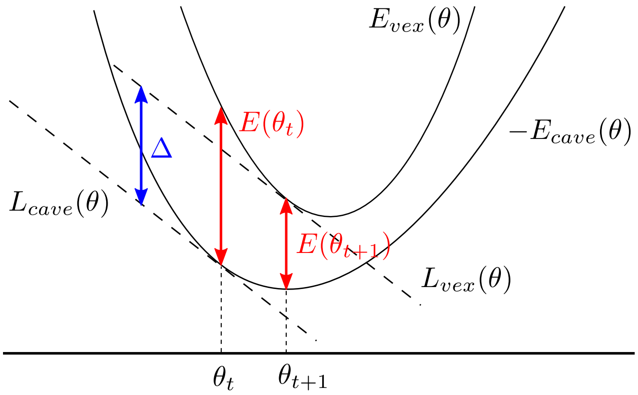


Figure 1: An illustration of the concave-convex procedure.

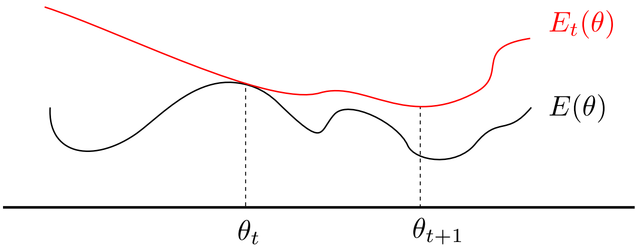


Figure 2: An illustration of the majorisation-minimisation method.

Now assume we apply some optimisation method to $E_t(\theta)$ to obtain the next update $\theta_{t+1}$. If this computation achieves $E_t(\theta_t) \geq E_t(\theta_{t+1})$, then by definition we have $E(\theta_t) = E_t(\theta_t) \geq E_t(\theta_{t+1}) \geq E(\theta_{t+1})$, thus proves the guarantee.

We now prove that CCCP is also an MM-type algorithm. At each iteration we construct a surrogate objective $E_t(\theta) = E_{vex}(\theta) - L_{cave}(\theta)$. Notice that $L_{cave}(\theta)$ depends on $E_{cave}(\theta)$ and $\theta_t$, and more importantly the slope of this linear function is given by the negative gradient of the concave part $-\nabla E_{cave}(\theta_t)$. From convexity of $-E_{cave}(\theta)$ it is straightforward to see $E_t(\theta)$ is a convex function that majorises $E(\theta)$ at $\theta_t$. Next we obtain the update by zeroing the gradient of $E_t(\theta)$, which means

$$\nabla E_t(\theta_{t+1}) = \nabla E_{vex}(\theta_{t+1}) + \nabla E_{cave}(\theta_t) = 0, \tag{4}$$

an equivalent procedure as described in eq. (3). To summarise, CCCP is a majorisation-minimisation algorithm that partially linearises the objective function and zeros the gradients on the surrogate.

Apparently there exist infinite number of functions that majorises the objective at the given location, and it is still an open question that how to design the heuristic which returns the one that provides good next-step updates and is easy to compute. CCCP is problematic in this sense because we still do not have satisfactory answers in general to 1) how to decompose the energy function and 2) how to efficiently search the point $\theta_{t+1}$ that satisfies eq. (3).

# References

[Yuille and Rangarajan, 2003] Yuille, A. L. and Rangarajan, A. (2003). The concave-convex procedure. *Neural Computation*, 15(4):915–936.