Newton's method for FastICA algorithm

Tomas Zeman*

May 2000

Abstract

Here we present in more details the Newton's method modified for performing ICA - so called FastICA algorithm introduced by Aapo Hyvarinen in [1, 2, 3, 4], overview of the methods in [5]. It is a batch algorithm which optimizes some nonlinear function used as a measure of statistcal independence. Commonly used are e.g. cummulants (but these are sensitive to outliers), approximation of negentropy (tanh, pow3) or sigmoidal function in infomax principle (Bell & Sejnowski [6, 7]). The optimization by means of Newton's method is very effective and a slight modification makes the approach more robust.

FastICA algorithm derivation

Maxima of approximation of negentropy are reached at certain optima of $\mathrm{E}\left\{G(\mathbf{w}^T\mathbf{x})\right\}$. According to Kuhn-Tucker conditions, the optima of $\mathrm{E}\left\{G(\mathbf{w}^T\mathbf{x})\right\}$ under the constraint $\|\mathbf{w}\| = 1$ are obtained at points where

$$\mathrm{E}\left\{\mathbf{x}g(\mathbf{w}^T\mathbf{x})\right\} - \beta\mathbf{w} = 0, \tag{1}$$

where $g = G'$ and $\beta = \mathrm{E}\left\{\mathbf{w}_{opt}^T\mathbf{x}g(\mathbf{w}_{opt}^T\mathbf{x})\right\}$.

We will use the Newton's method for iteration, i.e.

$$\mathbf{w}^+ = \mathbf{w} - (\mathbf{J}(\mathbf{w}))^{-1}f(\mathbf{w}),$$

where $\mathbf{J}(\mathbf{w})$ is the Jacobian of the eq. (1). Now we derive the Jacobian:

$$\mathbf{J}(\mathbf{w}) = \frac{\partial}{\partial\mathbf{w}}[\mathrm{E}\left\{\mathbf{x}g(\mathbf{w}^T\mathbf{x})\right\} - \beta\mathbf{w}] = \mathrm{E}\left\{\mathbf{x}g'(\mathbf{w}^T\mathbf{x})\mathbf{x}^T\right\} - \beta\mathbf{I} = \mathrm{E}\left\{\mathbf{x}\mathbf{x}^Tg'(\mathbf{w}^T\mathbf{x})\right\} - \beta\mathbf{I}$$

Since the data is whitened, $\mathrm{E}\left\{\mathbf{x}\mathbf{x}\right\}^T = \mathbf{I}$. We make some approximation $\mathrm{E}\left\{\mathbf{x}\mathbf{x}^Tg'(\mathbf{w}^T\mathbf{x})\right\} \approx \mathrm{E}\left\{\mathbf{x}\mathbf{x}^T\right\}\mathrm{E}\left\{g'(\mathbf{w}^T\mathbf{x})\right\} = \mathrm{E}\left\{g'(\mathbf{w}^T\mathbf{x})\right\}$. Then the Jacobian has final form:

$$\mathbf{J}(\mathbf{w}) = \mathrm{E}\left\{g'(\mathbf{w}^T\mathbf{x})\right\}\mathbf{I} - \beta\mathbf{I} \tag{2}$$

*CTU FEE, Dept. of Circuit Theory, e-mail: xzemant@feld.cvut.cz

So the inversion of $\mathbf{J}(\mathbf{w})$ is simple and iteration formula is as follows:

$$\mathbf{w}^+ = \mathbf{w} - (\mathbf{J}(\mathbf{w}))^{-1} f(\mathbf{w}) = \mathbf{w} - \frac{\mathrm{E}\left\{\mathbf{x} g(\mathbf{w}^T\mathbf{x})\right\} - \beta\mathbf{w}}{\mathrm{E}\left\{g'(\mathbf{w}^T\mathbf{x})\right\} - \beta} \tag{3}$$

Now let's multiply both sides by the denominator and we obtain:

$$\begin{aligned} \mathbf{w}^+(\beta - \mathrm{E}\left\{g'(\mathbf{w}^T\mathbf{x})\right\}) &= \mathbf{w}(\beta - \mathrm{E}\left\{g'(\mathbf{w}^T\mathbf{x})\right\}) + \mathrm{E}\left\{\mathbf{x} g(\mathbf{w}^T\mathbf{x})\right\} - \beta\mathbf{w} \\ &= \mathrm{E}\left\{\mathbf{x} g(\mathbf{w}^T\mathbf{x})\right\} - \mathrm{E}\left\{g'(\mathbf{w}^T\mathbf{x})\right\}\mathbf{w} \end{aligned}$$

Since the new update is normalized after every iteration, the final form of the FastICA algorithm is:

$$\begin{aligned} \mathbf{w}^+ &= \mathrm{E}\left\{\mathbf{x} g(\mathbf{w}^T\mathbf{x})\right\} - \mathrm{E}\left\{g'(\mathbf{w}^T\mathbf{x})\right\}\mathbf{w} \\ \mathbf{w}_{new} &= \frac{\mathbf{w}^+}{\|\mathbf{w}^+\|} \end{aligned} \tag{4}$$

Sometimes, however, the convergence of the Newton's method may be uncertain. So after slight modification we can make the above algorithm more robust by adding a step size in (3):

$$\begin{aligned} \mathbf{w}^+ &= \mathbf{w} - \mu\frac{\mathrm{E}\left\{\mathbf{x} g(\mathbf{w}^T\mathbf{x})\right\} - \beta\mathbf{w}}{\mathrm{E}\left\{g'(\mathbf{w}^T\mathbf{x})\right\} - \beta} \\ \mathbf{w}_{new} &= \frac{\mathbf{w}^+}{\|\mathbf{w}^+\|} \end{aligned} \tag{5}$$

thus obtaining the *stabilized fastICA algorithm*.

References

[1] A. Hyvärinen. A family of fixed-point algorithms for independent component analysis. *In Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP'97), Munich, Germany*, 1997.

[2] R. Vigario J. Karhunen, A. Hyvärinen, J. Hurri, and E. Oja. Applications of neural blind separation to image and signal processing. *In Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP'97), Munich, Germany*, pages 131–134, 1997.

[3] A. Hyvärinen and E. Oja. Independent component analysis by general non-linear hebbian-like learning rules. *Signal Processing*, 64(3):301–313, 1998.

[4] A. Hyvärinen. Survey on independent component analysis. *Neural Computing Surveys*, 2:94–128, 1999.

[5] E.Oja A.Hyvärinen. Independent component analysis: A tutorial, 1999.

[6] Sejnowski T.J. Bell A.J. An information-maximization approach to blind separation and blind deconvolution. *Neural Computation*, 7:1129–1159, 1995.

[7] Bell A.J. Information theory, ICA & Applications, 1996.