Mathematical details of the GENEFOLD algorithm

K. Van Deun
Catholic University of Leuven, Belgium

1 The Loss Function

Consider n genes $i = 1, ..., n$ whose expression was measured in m conditions $j = 1, ..., m$. In the unfolding model, both are represented as points in low dimensional space such that the order of the distances of a gene to the different conditions reflects the order of the expression levels in the nonmetric case, or that the differences between the distances reflect the differences between the expression levels (metric case). Such a configuration can be obtained by minimizing the following loss function,

$$K(\mathbf{X}, \mathbf{Y}, \mathbf{\Gamma}) = n^{-1} \sum_{i=1}^{n} K_i(\mathbf{x}_i, \mathbf{Y}, \boldsymbol{\gamma}_i) = n^{-1} \sum_{i=1}^{n} \left[1 - r^2(\boldsymbol{\gamma}_i, \mathbf{d}_i)\right], \qquad (1)$$

with $\mathbf{d}_i$ the vector of distances for gene i and $\boldsymbol{\gamma}_i$ the transformed expression levels (with an inverse relation between the γ_{ij}'s and the expression levels: higher expression levels correspond to lower γ_{ij}'s). Here, we will use Euclidean distances such that the distance between gene i and condition j is given by $d_{ij} = [(\mathbf{x}_i - \mathbf{y}_j)'(\mathbf{x}_i - \mathbf{y}_j)]^{1/2}$ where $\mathbf{x}_i$ and $\mathbf{y}_j$ are the coordinate vectors for gene i and condition j. As shown by Kruskal and Carroll (1969) and by Van Deun, Groenen, Heiser, Busing, and Delbeke (2005), minimizing (1) or, equivalently, maximizing the squared correlation between the transformed data and the distances is equivalent to minimizing Stress-2. Van Deun et al. (2005) showed that under the condition $\boldsymbol{\gamma}_i' \mathbf{J} \boldsymbol{\gamma}_i = 1$, loss function (1) is equivalent to

$$K(\mathbf{X}, \mathbf{Y}, \mathbf{\Gamma}, \mathbf{a}) = n^{-1} \sum_{i=1}^{n} K_i(\mathbf{x}_i, \mathbf{Y}, \boldsymbol{\gamma}_i, a_i) = n^{-1} \sum_{i=1}^{n} \|\boldsymbol{\gamma}_i - a_i \mathbf{d}_i\|_{\mathbf{J}}^2, \qquad (2)$$

where a_i is a scale parameter and $\mathbf{J} = \mathbf{I} - m^{-1}\mathbf{1}\mathbf{1}'$ a centering operator. The a_i have to be restricted to be nonnegative in order to fit positive correlations between the distances and the transformed data as we want the distance from a gene to a condition to increase when the gene has a lower expression in that condition. An important aspect of the loss is that it is invariant to centering of $\mathbf{Y}$ because the distances $\mathbf{d}_i$ do not change.

We will minimize (2) by alternating least squares and iterative majorization using the algorithm presented here and named GENEFOLD.

1. Choose initial $\mathbf{X}^{(0)}$, $\mathbf{Y}^{(0)}$, and $\mathbf{a}^{(0)}$.
2. $l := 0$.

3. $l := l + 1$.
4. Compute the update of the gene coordinates $\mathbf{X}$ given $\mathbf{Y}^{(l-1)}$, $\mathbf{\Gamma}^{(l-1)}$, and $\mathbf{a}^{(l-1)}$.
5. Compute the update of the condition coordinates $\mathbf{Y}$ given $\mathbf{X}^{(l)}$, $\mathbf{\Gamma}^{(l-1)}$, and $\mathbf{a}^{(l-1)}$.
6. Compute the update of the transformed data $\mathbf{\Gamma}$ given $\mathbf{X}^{(l)}$, $\mathbf{Y}^{(l)}$, and $\mathbf{a}^{(l-1)}$.
7. Compute the update of the weights $\mathbf{a}$ based on $\mathbf{X}^{(l)}$, $\mathbf{Y}^{(l)}$, and $\mathbf{\Gamma}^{(l)}$.
8. If $L^{(l)} - L^{(l-1)} < \epsilon$ or $l = l_{max}$ stop. Else return to 3.

Suitable updates can be easily found for Steps 7 and also for 6 in the metric unfolding case. The more difficult parts are the steps 4, 5, and, in the nonmetric unfolding case, also 6. The update of the coordinates (steps 4 and 5) will be based on majorization procedures. Below we develop each update.

1.1 Updating the coordinates

In this section, we will derive expressions that update the coordinates $\mathbf{X}$ and $\mathbf{Y}$. As the loss function is complicated, we will replace it by majorizing functions that are easy to minimize (see De Leeuw, 1994, Heiser, 1995, and Lange, Hunter, & Yang, 2000 for a general introduction to iterative majorization). These functions are never lower than the original function and touch the original function at a supporting point. By taking the previous estimate as the supporting point in the current iteration, it is guaranteed that the loss function is nonincreasing with each new update. In practice, the values of the loss function converge to a local minimum. When the original function has a bounded Hessian, it can be majorized by a quadratic function. Kiers (2002) introduced one that can be easily updated both in constrained and unconstrained cases. This function is of the following form

$$g(\mathbf{Z}) = k + \text{tr}\mathbf{A}\mathbf{Z} + \sum_q \text{tr}\mathbf{B}_q\mathbf{Z}\mathbf{C}_q\mathbf{Z}', \tag{3}$$

where k includes all terms that are constant in $\mathbf{Z}$ with $\mathbf{Z}$ denoting the matrix we want to update (this is, in our case, $\mathbf{x}_i$ or $\mathbf{Y}$). So, we will look for a function that majorizes (2) and that is of the form (3).

As explained by Kiers (2002), when the $\mathbf{C}_q$ are positive semi-definite, the minimization of (3) is itself based on the majorizing function

$$g(\mathbf{Z}) \leq k + \text{tr}\mathbf{F}\mathbf{Z} + \sum_q \lambda_q \text{tr}\mathbf{C}_q\mathbf{Z}'\mathbf{Z}, \tag{4}$$

with λ_q the largest eigenvalue of $\left(2^{-1}\mathbf{B}_q + 2^{-1}\mathbf{B}_q'\right)$ and with $\mathbf{F} = \mathbf{A} + \sum_q \left(\mathbf{C}_q\mathbf{Z}_0'(\mathbf{B}_q + \mathbf{B}_q') - 2\lambda_q\mathbf{C}_q\mathbf{Z}_0'\right)$. Without constraints, the optimal $\mathbf{Z}$ is given by

$$\mathbf{Z}^+ = -\frac{1}{2}\mathbf{F}'\left(\sum_q \lambda_q\mathbf{C}_q\right)^{-1}. \tag{5}$$

The function that we want to minimize is given by (2), that is,

$$L(\mathbf{X}, \mathbf{Y}) = n^{-1}\sum_{i=1}^{n}\left(1 + a_i^2\mathbf{d}_i'\mathbf{J}\mathbf{d}_i - 2a_i\mathbf{d}_i'\mathbf{J}\boldsymbol{\gamma}_i\right). \tag{6}$$

To minimize this function, we will construct a quadratic majorizing function of the basic form (3). As majorization is closed under summation, we will majorize each of the terms that are not a linear or quadratic function of the coordinates. These are both terms involving distances, namely $a_i^2\mathbf{d}_i'\mathbf{J}\mathbf{d}_i$ and $a_i\mathbf{d}_i'\mathbf{J}\boldsymbol{\gamma}_i$.

Note that we have to update two sets of coordinates $\mathbf{X}$ and $\mathbf{Y}$ where the genes enter the loss function by summing over all n genes. This allows us to update the coordinates $\mathbf{x}_i$ for each of the genes separately (the sum is minimized by minimizing each of its components). Next, we will show how each of the terms can be expressed in the standard form (3), and this both for the condition coordinate matrix $\mathbf{Y}$ and the gene coordinate vector $\mathbf{x}_i$. For the sake of clarity, we will use a notation that follows closely the notation in (3) with numbered and lettered subscripts: we will use a subscript x and a subscript y when the term is expressed in function of respectively $\mathbf{x}_i$ and $\mathbf{Y}$. The numbers count the different terms that are in the same form. For example, k_{2x} would be the second term that is constant in $\mathbf{x}_i$. How the terms $a_i^2\mathbf{d}_i'\mathbf{J}\mathbf{d}_i$ and $-2a_i\mathbf{d}_i'\mathbf{J}\boldsymbol{\gamma}_i$ can be majorized by functions that are quadratic in the coordinates, is explained in appendix.

1.2 Updating X

We replace the second and third terms of equation (6) by their majorizing functions consisting of the sum of (21), (25), (29), and (32) developed in

appendix. The following majorizing function is then obtained for the $\mathbf{x}_i$

$$g(\mathbf{x}_i) = k_x + \text{tr}\mathbf{A}_x\mathbf{x}_i + \sum_{q=1}^{2} \text{tr}\mathbf{B}_{qx}\mathbf{x}_i\mathbf{C}_{qx}\mathbf{x}_i' \tag{7}$$

with $k_x = k_{1x} + k_{2x} + k_{3x} + k_{4x}$ and $\mathbf{A}_x = -2(\mathbf{A}_{1x} + \mathbf{A}_{2x} + \mathbf{A}_{3x} + \mathbf{A}_{4x})$. Note that because of the centered $\mathbf{Y}$, $\mathbf{A}_{1x} = \mathbf{0}$. This majorizing function can in turn be majorized as done by Kiers (2002) (see equation (4)),

$$g(\mathbf{x}_i) \leq k_x + \text{tr}\mathbf{F}_x\mathbf{x}_i + \sum_{q=1}^{2} \lambda_{qx}\text{tr}\mathbf{C}_{qx}\mathbf{x}_i'\mathbf{x}_i, \tag{8}$$

with λ_{qx} the highest eigenvalue of $\left(2^{-1}\mathbf{B}_{qx} + 2^{-1}\mathbf{B}_{qx}'\right)$ and with $\mathbf{F}_x = \mathbf{A}_x + \sum_q \left[\mathbf{C}_{qx}\mathbf{x}_{i0}'(\mathbf{B}_{qx} + \mathbf{B}_{qx}') - 2\lambda_{qx}\mathbf{C}_{qx}\mathbf{x}_{i0}'\right]$. As $\mathbf{B}_{1x}$ and $\mathbf{B}_{2x}$ are scalars, λ_{1x} and λ_{2x} equal these. This also implies that $\mathbf{C}_{qx}\mathbf{x}_{i0}'(\mathbf{B}_{qx} + \mathbf{B}_{qx}') = 2\lambda_{qx}\mathbf{C}_{qx}\mathbf{x}_{i0}'$ so that $\mathbf{F}_x = \mathbf{A}_x$. When there are no constraints, the optimal $\mathbf{x}_i$ is given by

$$\mathbf{x}_i^+ = -\frac{1}{2}\mathbf{F}_x' \left(\sum_q \lambda_{qx}\mathbf{C}_{qx}\right)^{-1}, \tag{9}$$

where $\sum_q \lambda_{qx}\mathbf{C}_{qx}$ is a scalar of which the inverse is easily obtained. A relaxed update $\mathbf{x}_i^{++}$ can be used to accelerate convergence (see De Leeuw & Heiser, 1980 and Heiser, 1995),

$$\mathbf{x}_i^{++} = 2\mathbf{x}_i^+ - \mathbf{x}_i. \tag{10}$$

The relaxed update still guarantees a nonincreasing sequence of loss values. Because the formulas connected to the update of $\mathbf{x}_i$ do not involve another gene i', the order in which the genes are updated does not influence the outcome of the unfolding analysis.

1.3 Updating Y

The majorizing function for the $\mathbf{Y}$ coordinates is given by replacing the second and third terms of equation (6) by the sum of (22), (26), (30), and (33) of the appendix yielding

$$g(\mathbf{Y}) = k_y + \text{tr}\mathbf{A}_y\mathbf{Y} + \sum_{q=1}^{2} \text{tr}\mathbf{B}_{qy}\mathbf{Y}\mathbf{C}_{qy}\mathbf{Y}', \tag{11}$$

with $k_y = k_{1y} + k_{2y} + k_{3y} + k_{4y}$ and $\mathbf{A}_y = -2(\mathbf{A}_{1y} + \mathbf{A}_{2y} - \mathbf{A}_{3y} + \mathbf{A}_{4y})$. The update is then given by

$$\mathbf{Y}^{+} = \frac{1}{2}\mathbf{F}_y \left(\sum_q \lambda_{qy} \mathbf{C}_{qy} \right)^{-1}, \tag{12}$$

with $\mathbf{F}_y = \mathbf{A}_y + \sum_q \left(\mathbf{C}_{qy}\mathbf{Y}_0'(\mathbf{B}_{qy} + \mathbf{B}_{qy}') - 2\lambda_{qy}\mathbf{C}_{qy}\mathbf{Y}_0' \right)$ and λ_{qy} the highest eigenvalue of $\left(2^{-1}\mathbf{B}_{qy} + 2^{-1}\mathbf{B}_{qy}'\right)$. Given the special form of the $\mathbf{B}_{qy}$ matrices, the eigenvalues can be set equal to respectively 1, $\sum_i c_{1i}$ and the maximal value on the diagonal of $\sum_i c_{3i}\mathbf{D}_{i0}^{-1}$ and it follows that $\mathbf{F}_y = \mathbf{A}_y$. The inverse of $\sum_q \lambda_{qy}\mathbf{C}_{qy}$ can be easily obtained as it is a diagonal matrix. Here, too, the number of iterations can be approximately halved using the relaxed update:

$$\mathbf{Y}^{++} = 2\mathbf{Y}^{+} - \mathbf{Y}. \tag{13}$$

1.4 Updating $\boldsymbol{\Gamma}$

The loss for gene i can be written as,

$$L_i(\boldsymbol{\gamma}_i) = \|\boldsymbol{\gamma}_i - a_i\mathbf{d}_i\|_{\mathbf{J}}^2, \tag{14}$$

with $\boldsymbol{\gamma}_i$ being function of $\boldsymbol{\delta}_i$, e.g. in case of metric unfolding $\gamma_{ij} = b_{0i} + b_{1i}\delta_{ij}$ under the constraint $b_{1i} \leq 0$. This function can be minimized by (monotone) regressing the expression profiles $\boldsymbol{\delta}_i$ on $a_i\mathbf{d}_i$ and this under a nonnegativity constraint $\gamma_{ij} \geq 0$ and a length constraint $\boldsymbol{\gamma}_i'\mathbf{J}\boldsymbol{\gamma}_i = 1$. Let $\boldsymbol{\delta}_i$ be the expression profile for gene i, then fixed $\boldsymbol{\gamma}_i$ of the following form are obtained for data with an interval level of measurement (see also Van Deun et al., 2005):

$$\boldsymbol{\gamma}_i = k - (\boldsymbol{\delta}_i'\mathbf{J}\boldsymbol{\delta}_i)^{-1/2}\boldsymbol{\delta}_i, \tag{15}$$

with k an arbitrary constant chosen such that all $\gamma_{ij} \geq 0$. Equation (15) clearly shows that the update is independent from the coordinate values. Therefore, the $\boldsymbol{\gamma}_i$'s can be fixed right away in the metric unfolding case. In the nonmetric case, as shown by Gifi (1990), the length restriction can be imposed after the monotone regression by dividing the updated disparities by $\boldsymbol{\gamma}_i'\mathbf{J}\boldsymbol{\gamma}_i^{1/2}$. In case all γ_{ij} are equal, Groenen, Os, and Meulman (2000) show how the optimal disparities are given by $\boldsymbol{\gamma}_i = \mathbf{S}\mathbf{b}_{opt}$ with $\mathbf{S} = \mathbf{JLM}$ where $\mathbf{M}$ is a matrix of ones on and below the diagonal and zeroes above it, $\mathbf{L}$ is a matrix composed of rows with zeroes and a one in the r-th column of

row j where r is the rank value of d_{ij}. Let $\mathbf{G} = \mathbf{S}'\mathbf{S}$, then $\mathbf{b}_{opt}$ is a vector of all zeroes except one which equals $(g)_{ii}^{-1}$ (see also Van Deun et al., 2005). The nonzero value is chosen such that $\mathbf{d}'\mathbf{d} + 1 - 2\mathbf{b}'\mathbf{S}'\mathbf{d}$ is minimal. An additive constant can be added without loss of generality and this property can be used to make all disparities nonnegative (since $\mathbf{S} = \mathbf{JLM}$ is centered). Nonnegative γ_{ij}'s are important because the majorizing functions for $\mathbf{X}$ and $\mathbf{Y}$ depend on the assumption of nonnegative γ_{ij}'s.

1.5 Updating a

The optimal value for a_i can be found by setting the partial derivative of (6) with respect to a_i equal to zero:

$$\frac{\partial L}{\partial a_i} = 2g_i a_i \mathbf{d}_i'\mathbf{J}\mathbf{d}_i - 2g_i \mathbf{d}_i'\mathbf{J}\boldsymbol{\gamma}_i = 0. \tag{16}$$

Under the constraint that $a_i \geq 0$, the following update is obtained,

$$a_i = \begin{cases} \mathbf{d}_i'\mathbf{J}\boldsymbol{\gamma}_i(\mathbf{d}_i'\mathbf{J}\mathbf{d}_i)^{-1} & \text{if } \ a_i \geq 0 \ , \\ 0 & \text{if } \ a_i < 0 \ . \end{cases} \tag{17}$$

2 Adding constraints

To avoid degenerate solutions for the unfolding model, we will build in some constraints. The degeneracy shows itself in a lack of variation of the distances from a gene point to the condition points. Imposing a variance constraint in order to obtain a minimal amount of variance seems a straightforward solution but is insufficient given the scale dependency of the variance. In other words, it makes no sense to fix the variance to a certain value if the distances can still become very large; some sort of standardization is needed. We propose to work in a space centered at $\mathbf{Y}$ and with $\|\mathbf{x}_i\| \leq 1$, this is restricting the space for the $\|\mathbf{x}_i\|$ to the unit disk. Within this space, we pull the variance of the distances $\mathbf{d}_i$ towards values that are at least v_m, with v_m the variance of the distances of a point on the unit disk to m other points on the unit disk. v is determined by simulation as we could not find a closed form expression for it. It will become clear that this minimal value is forced only for points that fit perfectly.

As explained by Kiers (2002), the equality constraint $\|\mathbf{x}_i\|^2 = 1$ can be imposed by adapting step 4 (see page 2) as follows. Equation (8) can be

rewritten as a linear function under the constraint $\mathbf{x}_i'\mathbf{x}_i = 1$ as $k_x + \mathrm{tr}\mathbf{F}_x\mathbf{x}_i + \sum_{q=1}^{2}\lambda_q \mathrm{tr}\mathbf{C}_{qx}$. This function can be minimized by maximizing $\mathrm{tr}-\mathbf{F}_x\mathbf{x}_i$ so that the update is given by $\mathbf{x}^+ = \mathbf{V}_x\mathbf{U}_x'$ where $\mathbf{V}_x$ and $\mathbf{U}_x$ come from the singular value decomposition of $-\mathbf{F}_x = \mathbf{U}_x\mathbf{\Sigma}_x\mathbf{V}_x'$. To impose the inequality constraint $\|\mathbf{x}_i\|^2 \leq 1$, first calculate the unconstrained update. If the resulting $\mathbf{x}_i$ has a length greater then 1, compute the constrained update. In case of constraints, the relaxed update is no longer of use as convergence cannot be guaranteed anymore. The variance constraint is imposed implicitly by an upper bound on the a_i's: the optimal a_i can be rewritten as $r(\boldsymbol{\gamma}_i, \mathbf{d}_i)(\mathbf{d}_i'\mathbf{J}\mathbf{d}_i)^{-0.5}$ and thus depends on the correlation and on the variance of the distances such that an upper bound u for the a_i will result in restricting the spread of the distances. Within the unit reference space, setting $u = (mv_m)^{-0.5}$ will, for a maximal correlation $r(\boldsymbol{\gamma}_i, \mathbf{d}_i) = 1$, pull the variance of the distances towards a value v_m or higher (with v_m the variance of the distances on a unit disk with uniformly distributed points). The constrained update for a_i becomes then

$$a_i = \begin{cases} 0 & \text{if } \mathbf{d}_i'\mathbf{J}\boldsymbol{\gamma}_i(\mathbf{d}_i'\mathbf{J}\mathbf{d}_i)^{-1} < 0 \quad , \\ \mathbf{d}_i'\mathbf{J}\boldsymbol{\gamma}_i(\mathbf{d}_i'\mathbf{J}\mathbf{d}_i)^{-1} & \text{if } 0 \leq \mathbf{d}_i'\mathbf{J}\boldsymbol{\gamma}_i(\mathbf{d}_i'\mathbf{J}\mathbf{d}_i)^{-1} \leq u \quad , \\ \mathrm{u} & \text{if } \mathbf{d}_i'\mathbf{J}\boldsymbol{\gamma}_i(\mathbf{d}_i'\mathbf{J}\mathbf{d}_i)^{-1} > u \quad . \end{cases} \tag{18}$$

References

De Leeuw, J. (1994). Block relaxation algorithms in statistics. In H. H. Bock, W. Lenski, & M. M. Richter (Eds.), *Information systems and data analysis* (p. 308-325). Berlin: Springer-Verlag.

De Leeuw, J., & Heiser, W. J. (1980). Multidimensional scaling with restrictions on the configuration. In P. R. Krishnaiah (Ed.), *Multivariate analysis* (Vol. 5, pp. 501–522). Amsterdam, The Netherlands: North-Holland Publishing Company.

Gifi, A. (1990). *Nonlinear multivariate analysis*. Chicester, England: Wiley.

Groenen, P. J. F. (2002). *Iterative majorization algorithms for minimizing loss functions in classification.* (Working paper presented at the 8th conference of the IFCS, Krakow, Poland)

Groenen, P. J. F., Os, B.-J., & Meulman, J. J. (2000). Optimal scaling by alternating length-constrained nonnegative least squares with application to distance-based analysis. *Psychometrika*, *65*, 511–524.

Heiser, W. J. (1995). Convergent computation by iterative majorization: theory and applications in multidimensional data analysis. In W. J. Krzanowski (Ed.), *Recent advances in descriptive multivariate analysis* (pp. 157–189). Oxford: Oxford University Press.

Kiers, H. A. L. (2002). Setting up alternating least squares and iterative majorization algorithms for solving various matrix optimization problems. *Computational Statistics and Data Analysis*, *41*, 157–170.

Kruskal, J. B., & Carroll, J. D. (1969). Geometrical models and badness-of-fit functions. In P. R. Krishnaiah (Ed.), *Multivariate analysis* (Vol. 2, p. 639-671). New York: Academic Press.

Lange, K., Hunter, D. R., & Yang, I. (2000). Optimization transfer using surrogate objective functions. *Journal of computational and graphical statistics*, *9*, 1–20.

Van Deun, K., Groenen, P. J. F., Heiser, W. J., Busing, F. M. T. A., & Delbeke, L. (2005). Interpreting degenerate solutions in unfolding by use of the vector model and the compensatory distance model. *Psychometrika*, *70*, 23-47.

A Majorizing the distance terms

A.1 Majorizing $a_i^2\mathbf{d}_i'\mathbf{J}\mathbf{d}_i$

Here, we show how the first term of (2) involving distances can be majorized by a function of the coordinates that is of the form (3). A good reference for finding majorizing inequalities is Groenen (2002). The centering operator $\mathbf{J}$ equals $\mathbf{I} - m^{-1}\mathbf{1}\mathbf{1}'$, so that

$$a_i^2\mathbf{d}_i'\mathbf{J}\mathbf{d}_i = a_i^2\mathbf{d}_i'\mathbf{d}_i - \frac{a_i^2}{m}\mathbf{d}_i'\mathbf{1}\mathbf{1}'\mathbf{d}_i, \tag{19}$$

The first term in the right hand side of (19) can be rewritten as

$$\begin{aligned} a_i^2\mathbf{d}_i'\mathbf{d}_i &= a_i^2\text{tr}(\mathbf{1}\mathbf{x}_i' - \mathbf{Y})'(\mathbf{1}\mathbf{x}_i' - \mathbf{Y}) \\ &= c_{1i}\text{tr}\mathbf{x}_i\mathbf{1}'\mathbf{1}\mathbf{x}_i' - 2c_{1i}\text{tr}\mathbf{x}_i\mathbf{1}'\mathbf{Y} + c_{1i}\text{tr}\mathbf{Y}'\mathbf{Y} \end{aligned} \tag{20}$$

with $c_{1i} = a_i^2$. Rewriting (20) in function of $\mathbf{x}_i$ gives,

$$a_i^2\mathbf{d}_i'\mathbf{d}_i = k_{1x} + \text{tr}\mathbf{B}_{1x}\mathbf{x}_i\mathbf{C}_{1x}\mathbf{x}_i' - 2\text{tr}\mathbf{A}_{1x}\mathbf{x}_i, \tag{21}$$

with $k_{1x} = c_{1i}\text{tr}\mathbf{Y}'\mathbf{Y}$, $\mathbf{B}_{1x} = c_{1i}$, $\mathbf{C}_{1x} = \mathbf{1}'\mathbf{1}$, and $\mathbf{A}_{1x} = c_{1i}\mathbf{1}'\mathbf{Y}$. Note that for centered $\mathbf{Y}$, $\mathbf{A}_{1x} = [0 \quad 0]$. In function of $\mathbf{Y}$ we get

$$\sum_i g_i a_i^2\mathbf{d}_i'\mathbf{d}_i = k_{1y} + \text{tr}\mathbf{B}_{1y}\mathbf{Y}\mathbf{C}_{1y}\mathbf{Y}' - 2\text{tr}\mathbf{A}_{1y}\mathbf{Y}, \tag{22}$$

with $k_{1y} = \sum_i c_{1i}\text{tr}\mathbf{x}_i\mathbf{1}'\mathbf{1}\mathbf{x}'$, $\mathbf{B}_{1y} = \sum_i c_{1i}$, $\mathbf{C}_{1y} = \mathbf{I}$, and $\mathbf{A}_{1y} = \sum_i c_{1i}\mathbf{x}_i\mathbf{1}'$. Consider the second term in the right hand side of (19): $-\mathbf{d}_i'\mathbf{1}\mathbf{1}'\mathbf{d}_i$ is minus the squared sum of d_{ij}'s which can be majorized using the inequality $-cs^2 \leq cs_0^2 - 2css_0$, with c a positive constant. Take $c = m^{-1}a_i^2$ and $s = \mathbf{d}_i'\mathbf{1}$, we then obtain

$$\begin{aligned} -\frac{a_i^2}{m}\mathbf{d}_i'\mathbf{1}\mathbf{1}'\mathbf{d}_i &\leq \frac{a_i^2}{m}\mathbf{d}_{i0}'\mathbf{1}\mathbf{1}'\mathbf{d}_{i0} - \frac{2a_i^2}{m}\mathbf{d}_i'\mathbf{1}\mathbf{1}'\mathbf{d}_{i0} \\ &= \frac{a_i^2}{m}\mathbf{d}_{i0}'\mathbf{1}\mathbf{1}'\mathbf{d}_{i0} - \frac{2a_i^2}{m}(\mathbf{1}'\mathbf{d}_{i0})\mathbf{d}_i'\mathbf{1}. \end{aligned} \tag{23}$$

The last term of function (23) still contains distances that depend on $\mathbf{x}_i$ and $\mathbf{Y}$. The Cauchy Schwarz inequality can be used to obtain a majorizing

function of the coordinates as $-d = -\|\mathbf{x}-\mathbf{y}\| \leq -d_0^{-1}(\mathbf{x}-\mathbf{y})'(\mathbf{x}_0-\mathbf{y}_0)$. Set $c_{2i} = m^{-1}a_i^2(\mathbf{1}'\mathbf{d}_{i0})$ so that

$$\begin{aligned} -2c_{2i}\mathbf{d}_i'\mathbf{1} &\leq -2c_{2i}\text{tr}(\mathbf{1}\mathbf{x}_i' - \mathbf{Y})'\mathbf{D}_{i0}^{-1}(\mathbf{1}\mathbf{x}_{i0}' - \mathbf{Y}_0) \\ &= -2c_{2i}\text{tr}\mathbf{x}_i\mathbf{1}'\mathbf{D}_{i0}^{-1}(\mathbf{1}\mathbf{x}_{i0}' - \mathbf{Y}_0) + 2c_{2i}\text{tr}\mathbf{Y}'\mathbf{D}_{i0}^{-1}(\mathbf{1}\mathbf{x}_{i0}' - \mathbf{Y}_0) \end{aligned} \quad (24)$$

with $\mathbf{D}_{i0}^{-1}$ the inverse diagonal matrix of $\mathbf{d}_{i0}$'s. When $d_{i0} = 0$, we set the corresponding majorizing term equal to zero. In function of the $\mathbf{x}_i$ we obtain

$$-\frac{a_i^2}{m}\mathbf{d}_i'\mathbf{1}\mathbf{1}'\mathbf{d}_i \leq k_{2x} - 2\text{tr}\mathbf{A}_{2x}\mathbf{x}_i, \quad (25)$$

with $k_{2x} = c_{2i}[\mathbf{1}'\mathbf{d}_{i0} + 2\text{tr}\mathbf{Y}'\mathbf{D}_{i0}^{-1}(\mathbf{1}\mathbf{x}_{i0}' - \mathbf{Y}_0)]$ and $\mathbf{A}_{2x} = c_{2i}\mathbf{1}'\mathbf{D}_{i0}^{-1}(\mathbf{1}\mathbf{x}_{i0}' - \mathbf{Y}_0)$. In function of $\mathbf{Y}$

$$-\sum_i \frac{a_i^2}{m}\mathbf{d}_i'\mathbf{1}\mathbf{1}'\mathbf{d}_i \leq k_{2y} + 2\text{tr}\mathbf{A}_{2y}\mathbf{Y}, \quad (26)$$

with $k_{2y} = \sum_i c_{2i}[\mathbf{1}'\mathbf{d}_{i0} - 2\text{tr}\mathbf{x}_i\mathbf{1}'\mathbf{D}_{i0}^{-1}(\mathbf{1}\mathbf{x}_{i0}' - \mathbf{Y}_0)]$ and $\mathbf{A}_{2y} = \sum_i c_{2i}\mathbf{D}_{i0}^{-1}(\mathbf{1}\mathbf{x}_{i0}' - \mathbf{Y}_0)$.

A.2 Majorizing $-2a_i\mathbf{d}_i'\mathbf{J}\boldsymbol{\gamma}_i$

First we replace $\mathbf{J}$ by it's full expression,

$$-2a_i\mathbf{d}_i'\mathbf{J}\boldsymbol{\gamma}_i = -2a_i\mathbf{d}_i'\boldsymbol{\gamma}_i + \frac{2a_i}{m}\mathbf{d}_i'\mathbf{1}\mathbf{1}'\boldsymbol{\gamma}_i. \quad (27)$$

Both terms in the right hand side can be majorized. For the first term, which is again minus a linear function of the distances provided that the γ_{ij} are nonnegative, we get by the Cauchy Schwarz inequality

$$\begin{aligned} -2a_i\mathbf{d}_i'\boldsymbol{\gamma}_i &\leq -2a_i\text{tr}(\mathbf{1}\mathbf{x}_i' - \mathbf{Y})'\mathbf{D}_{i0}^{-1}\boldsymbol{\Gamma}_i(\mathbf{1}\mathbf{x}_{i0}' - \mathbf{Y}_0) \\ &= -2a_i\text{tr}\mathbf{x}_i\mathbf{1}'\mathbf{D}_{i0}^{-1}\boldsymbol{\Gamma}_i(\mathbf{1}\mathbf{x}_{i0}' - \mathbf{Y}_0) + \\ &\quad 2a_i\text{tr}\mathbf{Y}'\mathbf{D}_{i0}^{-1}\boldsymbol{\Gamma}_i(\mathbf{1}\mathbf{x}_{i0}' - \mathbf{Y}_0), \end{aligned} \quad (28)$$

where $\boldsymbol{\Gamma}_i$ is a diagonal matrix of $\boldsymbol{\gamma}_i$ and all $\boldsymbol{\gamma}_{ij} \geq 0$. In function of the $\mathbf{x}_i$ this is

$$-2a_i\mathbf{d}_i'\boldsymbol{\gamma}_i \leq k_{3x} - 2\text{tr}\mathbf{A}_{3x}\mathbf{x}_i, \quad (29)$$

with $k_{4x} = 2g_i a_i \text{tr}\mathbf{Y}'\mathbf{D}_{i0}^{-1}\mathbf{\Gamma}_i(\mathbf{1}\mathbf{x}_{i0}' - \mathbf{Y}_0)$ and $\mathbf{A}_{4x} = g_i a_i \mathbf{1}'\mathbf{D}_{i0}^{-1}\mathbf{\Gamma}_i(\mathbf{1}\mathbf{x}_{i0}' - \mathbf{Y}_0)$ while in function of $\mathbf{Y}$

$$-2\sum_i a_i \mathbf{d}_i'\boldsymbol{\gamma}_i \leq k_{3y} + 2\text{tr}\mathbf{A}_{3y}\mathbf{Y}. \tag{30}$$

Here, $k_{3y} = -2\sum_i a_i \text{tr}\mathbf{x}_i\mathbf{1}'\mathbf{D}_{i0}^{-1}\mathbf{\Gamma}_i(\mathbf{1}\mathbf{x}_{i0}' - \mathbf{Y}_0)$ and $\mathbf{A}_{3y} = \sum_i a_i \mathbf{D}_{i0}^{-1}\mathbf{\Gamma}_i(\mathbf{1}\mathbf{x}_{i0}' - \mathbf{Y}_0)$. The second term, which is a positive sum of distances, can be majorized by the inequality $cs \leq c(2s_o)^{-1}s^2 + c2^{-1}|s_0|$ with c a positive constant. Take $c_{3i} = m^{-1}a_i(\mathbf{1}'\boldsymbol{\gamma}_i)$

$$\begin{aligned} c_{3i}\mathbf{d}_i'\mathbf{1} &\leq c_{3i}\text{tr}(\mathbf{1}\mathbf{x}_i' - \mathbf{Y})'\mathbf{D}_{i0}^{-1}(\mathbf{1}\mathbf{x}_i' - \mathbf{Y}) + c_{3i}\mathbf{d}_{i0} \\ &= c_{3i}\text{tr}\mathbf{x}_i\mathbf{1}'\mathbf{D}_{i0}^{-1}\mathbf{1}\mathbf{x}_i' + c_{3i}\text{tr}\mathbf{Y}'\mathbf{D}_0^{-1}\mathbf{Y} - 2c_{3i}\text{tr}\mathbf{x}_i\mathbf{1}'\mathbf{D}_{i0}^{-1}\mathbf{Y} + c_{3i}\mathbf{d}_{i0} \end{aligned} \tag{31}$$

We rewrite the majorizing function (31) in function of the $\mathbf{x}_i$,

$$c_{3i}\mathbf{d}_i'\mathbf{1} \leq k_{4x} + \text{tr}\mathbf{B}_{2x}\mathbf{x}_i\mathbf{C}_{2x}\mathbf{x}_i' - 2\text{tr}\mathbf{A}_{4x}\mathbf{x}_i, \tag{32}$$

with $k_{4x} = c_{3i}(\text{tr}\mathbf{Y}'\mathbf{D}_{i0}^{-1}\mathbf{Y} + \mathbf{d}_{i0})$, $\mathbf{B}_{2x} = c_{3i}$, $\mathbf{C}_{2x} = \mathbf{1}'\mathbf{D}_{i0}^{-1}\mathbf{1}$, and $\mathbf{A}_{4x} = c_{3i}\mathbf{1}'\mathbf{D}_{i0}^{-1}\mathbf{Y}$. In function of $\mathbf{Y}$,

$$\sum_i c_{3i}\mathbf{d}_i'\mathbf{1} \leq k_{4y} + \text{tr}\mathbf{B}_{2y}\mathbf{Y}\mathbf{C}_{2y}\mathbf{Y}' - 2\text{tr}\mathbf{A}_{4y}\mathbf{Y}, \tag{33}$$

with $k_{4y} = \sum_i c_{3i}(\text{tr}\mathbf{x}_i\mathbf{1}'\mathbf{D}_0^{-1}\mathbf{1}\mathbf{x}_i' + \mathbf{d}_{i0})$, $\mathbf{B}_{2y} = \sum_i c_{3i}\mathbf{D}_{i0}^{-1}$, $\mathbf{C}_{2y} = \mathbf{I}$, and $\mathbf{A}_{4y} = \sum_i c_{3i}\mathbf{x}_i\mathbf{1}'\mathbf{D}_{i0}^{-1}$.