Estimating Parameters of 3D Cell Model using a Bayesian Recursive Global Optimizer (BaRGO)

Pietro Miotti[1 [0000−0002−1163−0599]], Edoardo Filippi-Mazzola[1 [0000−0002−1623−7294]], Ernst C. Wit[1 [0000−0002−3671−9610]], and Igor V. Pivkin[1 [0000−0002−9593−5217]]

Institute of Computing, Università della Svizzera Italiana, Lugano, Switzerland
pietro.miotti@usi.ch
https://www.ci.inf.usi.ch/

Abstract. In the field of Evolutionary Strategy, parameter estimation for functions with multiple minima is a difficult task when interdependencies between parameters have to be investigated. Most of the current routines that are used to estimate such parameters leverage state-of-the-art machine learning approaches to identify the global minimum, ignoring the relevance of the potential local minima. In this paper, we present a novel Evolutionary Strategy routine that uses sampling tools deriving from the Bayesian field to find the best parameters according to a certain loss function. The Bayesian Recursive Global Optimizer (BaRGO) presented in this work explores the parameter space identifying both local and global minima. Applications of BaRGO to 2D minimization problems and to parameter estimation of Red Blood Cell model are reported.

Keywords: Bayesian inference · Markov-Chain-Monte-Carlo · Cell model.

1 Introduction

Estimating parameters of computational cell models from experimental measurements is often a difficult task that may involve handling large number of degrees of freedom, high computational costs, and scarcity of data. The estimation is typically performed by manually changing the values of the parameters, which can become laborious and time consuming.

With the developments in the machine learning field, researchers have adapted optimization routines to provide fast and automatic approaches for parameter estimation [1, 5]. In this context, the idea of finding the global minimum of some objective function through a Bayesian-like approach that relies on Markov-Chain-Monte-Carlo (MCMC) method was introduced in reference [2]. Building on similar assumptions about the probability distribution of the population behavior, we propose a Bayesian Recursive Global Optimizer (BaRGO), a novel Bayesian evolutionary strategy for performing parameter estimation. BaRGO combines the reliability of Markov-Chain-Monte-Carlo (MCMC) approaches to estimate the posterior parameters together with probabilistic cluster analysis to

ICCS Camera Ready Version 2023
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-36024-4_27

find multiple minima by assuming a mixture of probability distributions that are estimated in the routine by an Expectation-Maximization algorithm [4]. In this sense, BaRGO explores all the minima it encounters in the domain, to find the best set of parameters for the objective function that are considered to match the experimental data [13] which can provide deeper insight and understanding of dependencies between parameters.

Following a brief introduction to Evolutionary Strategy using a statistical Bayesian perspective in section 2, we present the algorithmic procedure of BaRGO in section 3. In section 4, we evaluate performance of BaRGO on 2D minimization problems and parameter estimation of Red Blood Cell (RBC) model.

2 Bayesian inference on population parameters in evolutionary optimization

Evolutionary Strategies (ES) are a subfamily of stochastic optimization algorithms that investigate a potential solution space using biologically inspired mechanisms such as selection, mutation, and gene crossover. As a result, in ES, iterations are referred to as generations, and the set of values evaluated at each generation is referred to as the population [6].

The main steps in the optimization process are as follows. Given the loss function f and the solution space S, the goal of the ES algorithm is to to find $\boldsymbol{x}^* \in S$ such that $\boldsymbol{x}^* = \arg\min_{\boldsymbol{x}} f(\boldsymbol{x})$. Assuming that λ different elements are sampled from the solution space (population), only k are saved (selection). These are considered to be the best elements that minimize the loss function f. Such elements are then recombined (via crossover and mutation) to sample additional λ elements that will represent the new population (the next generation).

From a probabilistic perspective, the operations of crossover and mutation can be performed by sequentially updating the parameters of the distributions from which the population is sampled. From the Bayesian perspective of continuously updating prior beliefs, it is possible to design a MCMC approach that infers the parameter's population through sampling using the previous generation parameters as priors (see [9] for an overview on Bayesian inference).

In this regard, let's assume that a population $\mathbf{x}_g$ of size λ, at generation g, for $g = 1, \ldots, G$, is multivariate Normally distributed in d dimensions with mean $\boldsymbol{\beta}_g$ and covariance matrix Σ_g, where

$$\mathbf{x}_g = \begin{pmatrix} x_{1,g} \\ x_{2,g} \\ \vdots \\ x_{\lambda,g} \end{pmatrix} \quad \boldsymbol{\beta}_g = \begin{pmatrix} \beta_{1,g} \\ \beta_{2,g} \\ \vdots \\ \beta_{d,g} \end{pmatrix} \quad \Sigma_g = \begin{pmatrix} \sigma_1^2 & \sigma_{1,2} & \ldots & \sigma_{1,d} \\ \sigma_{1,2} & \sigma_2^2 & \ldots & \sigma_{2,d} \\ \vdots & \vdots & & \vdots \\ \sigma_{1,d} & \sigma_{2,d} & \ldots & \sigma_d^2 \end{pmatrix}.$$

From an ES perspective, we can consider $\boldsymbol{\beta}_g$ to be the candidate value which minimizes f (i.e. $\boldsymbol{\beta}_g \approx \boldsymbol{x}^*$) whereas Σ_g quantifies the uncertainty we have in proposing that candidate.

ICCS Camera Ready Version 2023
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-36024-4_27

Using the previous generation parameters as prior beliefs, it is possible to update such parameters for the current generation $\boldsymbol{\beta}_g$ and Σ_g in a Bayesian fashion through a Gibbs sampler routine that sequentially samples from the full conditional distributions. Indeed, this involves defining semi-conjugate priors for the mean and for the covariance matrix. A convenient one for $\boldsymbol{\beta}_g$ is the multivariate normal distribution $p(\boldsymbol{\beta}_g) \sim \mathcal{N}_d(\boldsymbol{\mu}_{g-1}, \Lambda_{g-1})$. By selecting the $k < \lambda$ population elements which minimize the loss function value within all the generations, the full conditional distribution for the mean will then be a multivariate Normal distribution with the following updated parameters:

$$\begin{aligned} p(\boldsymbol{\beta}_g \mid x_1, \dots, x_k, \Sigma_g) &\sim \mathcal{N}_d(\boldsymbol{\mu}_g, \Lambda_g), \\ \boldsymbol{\mu}_g &= (\Lambda_{g-1}^{-1} + n\Sigma_g^{-1})^{-1}(\Lambda_0^{-1}\boldsymbol{\mu}_g + n\Sigma_g^{-1}\bar{x}), \\ \Lambda_g &= (\Lambda_{g-1}^{-1} + n\Sigma_g^{-1})^{-1}, \end{aligned}$$

where $\bar{x}$ is the sample average of the current generation. Similarly, an appropriate semi-conjugate prior for Σ_g is the the inverse-Wishart distribution $p(\Sigma) \sim \text{inv-Wis}(v_{g-1}, S_{g-1}^{-1})$. Thus, the full conditional distribution is

$$\begin{aligned} p(\Sigma_g \mid x_1, \dots, x_k) &\sim \text{inv-Wis}(v_g, S_g^{-1}), \\ v_g &= v_{g-1} + \lambda, \\ S_g &= S_{g-1} + S_\beta, \end{aligned}$$

where S_β is the residual sum of squares from the population mean $\boldsymbol{\beta}_g$.

From the full conditional distributions, it is possible to construct a Gibbs sampler routine that generates posterior samples for $\boldsymbol{\beta}_g$ and Σ_g. Given a set of starting conditions, namely $\{\boldsymbol{\mu}_{g-1}, \Lambda_{g-1}, v_{g-1}, S_g\}$, the Gibbs sampler generates at iteration t, for $t = 1, \dots, T$, $\{\boldsymbol{\beta}_g^{(t+1)}, \Sigma_g^{(t+1)}\}$ from $\{\boldsymbol{\beta}_g^{(t)}, \Sigma_g^{(t)}\}$ according to the following two steps:

1. sample $\boldsymbol{\beta}_g^{(t+1)}$ from its full conditional distribution:
   - update $\boldsymbol{\mu}_g$ and Λ_g through $x_1, \dots, x_k$ and $\Sigma_g^{(t)}$;
   - sample $\boldsymbol{\beta}_g^{(t+1)} \sim \mathcal{N}_d(\boldsymbol{\mu}_g, \Lambda_g)$;
2. sample $\Sigma_g^{(t+1)}$ from its full conditional distribution:
   - update S_g through $x_1, \dots, x_k$ and $\boldsymbol{\beta}_g^{(t+1)}$;
   - sample $\Sigma_g^{(t+1)} \sim \text{inv-Wis}(v_g, S_g^{-1})$.

Therefore, after a sufficient amount of iterations to ensure the convergence of the chain, it is possible to have estimates for the parameters of the posterior by computing the empirical average across the sampled values:

$$\begin{aligned} \hat{\boldsymbol{\beta}}_g &= \mathbb{E}[\boldsymbol{\beta}_g^{t=1\dots T}], \\ \hat{\Sigma_g} &= \mathbb{E}[\Sigma_g^{t=1\dots T}]. \end{aligned}$$

Thus, having updated the new parameters, it is possible to generate a new population $\boldsymbol{x}_{g+1}$ by sampling from a multivariate Normal with the posterior estimates $\hat{\boldsymbol{\beta}}_g$ and $\hat{\Sigma_g}$.

ICCS Camera Ready Version 2023
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-36024-4_27

3 Bayesian Recursive Global Optimizer (BaRGO)

We propose a Bayesian Recursive Global Optimizer (BaRGO), a routine that leverages on Bayesian MCMC described above, to iteratively converge at $\boldsymbol{x}^*$ that minimizes f. BaRGO provides posterior estimates for each generation g, using previous generation information as prior beliefs.

At every generation, g, $\boldsymbol{\beta_g}$ and Σ_g are updated considering the k-best population elements encountered following the Bayesian inference approach introduced in Section 2. For functions with a unique global minimum, after every generation we can expect a decay in the uncertainty $\Sigma_g \to \mathbf{0}$ as $\boldsymbol{\beta}_g$ converges to $\boldsymbol{x}^*$. However, when functions exhibit multiple minima $\tilde{\boldsymbol{x}}_1^*, \ldots, \tilde{\boldsymbol{x}}_m^*$, the best values selected for computing the posteriors may belong to different minima, which may result in the update of $\boldsymbol{\beta}_g$ as the weighted average of the priors and the minima captured by the k-best elements selected. As a consequence, since there are multiple minima, Σ_g values will never converge to $\mathbf{0}$. To solve this issue, BaRGO splits the non-convex problem into small local convex problems and solves them separately. Specifically, once the KullBack Leiber divergence [10] between the multivariate Normal densities of two subsequent generations is smaller than a predefined threshold, the algorithm realizes that there are multiple minima $\tilde{\boldsymbol{x}}_1^*, \ldots, \tilde{\boldsymbol{x}}_m^*$ which are not allowing BaRGO to fully converge. As a consequence, we can consider this case as the result of population data being generated by a mixture of multivariate Gaussian distributions. The k-best population elements can then be re-grouped into m distinct clusters using an Expectation-Maximization (EM) approach [4, 15]. Once the clustering is deployed, BaRGO is recursively applied to each cluster, producing a local result for each of them $\boldsymbol{\beta}_1, ..., \boldsymbol{\beta}_m$. Local results are then compared, while the best one is returned as global minimum.

The upside of using a model-based approach to cluster data is the ability to use information criteria such as the Bayesian Information Criterion (BIC) to perform model selection on the most suited number of clusters given the current population. Indeed, the EM step is applied multiple times on a grid of a potential number of clusters. On each of these, the BIC is evaluated and the one with the lowest score on average is selected as the most suited one.

From implementation perspective, ES are in general highly parallelizable, which makes them perfectly suitable for the parameter estimation of computationally intensive models. Specifically, the evaluation of the population elements can be performed in parallel, hence the computational time of a single generation can be reduced to the computational time of single population element. Each population element can in turn be evaluated in parallel, exploiting HPC solvers.

4 Applications of BaRGO

4.1 Estimating function parameters in 2D minimization problems

The current state-of-the-art ES method, the Covariance Matrix Adaptation Evolutionary Strategy (CMA-ES) [8], exploits a similar idea as it samples directly from a multivariate Normal distribution [7]. In preliminary studies, we compare

ICCS Camera Ready Version 2023
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-36024-4_27

performance of BaRGO and CMA-ES in a 2D minimization problem setup on three traditional functions: Sphere (also known as the cone), Schwefel and Rastring functions. The peculiarities of these functions are entailed in their number of minima: the Sphere function exhibits a single minimum, whereas Rastring and Schwefel have multiple minima. Obtained results are reported in Figure 1. The evolution of the mean and the standard deviation of the sampling distributions $\mathcal{N}_2(\boldsymbol{X}_g, \Sigma_g)$ with $\boldsymbol{X} = [X_1, X_2]$ and Σ_g with standard deviations σ_1 and σ_2, are reported to compare the speed of convergence of both algorithms. In order to perform proper comparisons, since CMA-ES only search for the global optima, we apply the recursive call of BaRGO only to the cluster which exhibits the best candidate value, which means that we do not explore all the possible minima. In all three cases BaRGO converges with smaller number of function evaluations than CMA-ES, which translates to faster convergence properties, as seen from the decay-rate of standard deviations in Figure 1. These preliminary studies demonstrate interesting properties of BaRGO that deserve further investigation and analysis.

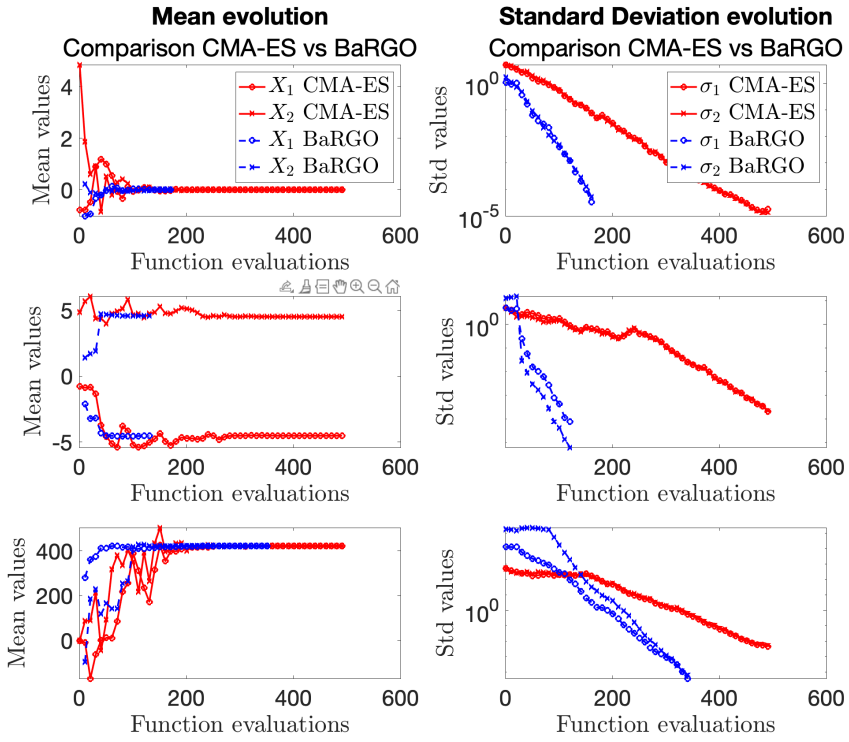

Fig. 1: Performance comparison between CMA-ES and BaRGO on Sphere (upper row), Rastring (middle row) and Schwefel (lower row) functions.

ICCS Camera Ready Version 2023
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-36024-4_27

4.2 Estimating the RBC model parameters using data from optical tweezers experiment

Optical tweezers is an experimental technique based on optical or laser traps that is used to stretch cells in one or more directions by trapping beads that are strategically attached to the cell surface [3]. The different deformations in the axial and transverse diameter of the cell (Figure 2), resulting from the stretching force applied, provide information regarding the elastic properties of the cell. For this application, using experimental data from [14], we tested BaRGO by estimating the parameters of the RBC model from [16]. In this coarse-grained model, which we implemented in LAMMPS [12], the cell membrane is modeled using surface triangulation with $N_v = 5000$ vertices connected with Worm Like Chain (WLC) links. The parameters of the model that we estimate are the WLC persistence length, p, the membrane bending coefficient, k_{bend}, as well as the area and volume conservation coefficients, k_a and k_v. These parameters play a key role in defining the elastic properties of the RBC model. For detailed description of the model and its parameters we refer to reference [16]. In order to speed up convergence the search space of each coefficient is constrained. Specifically, we limit the search for the persistence length between 0 and 0.512 X 10^{-7} (the later value is equal to maximum length of the links in the WLC model, l_{max} [16]), k_{bend} between 0 and 1000, and both k_v and k_a between 0 and 1000000.

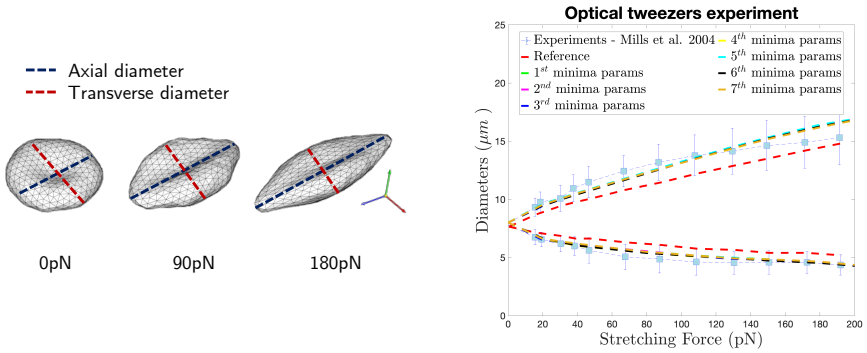

Fig. 2: (Left) Optical tweezers experiment: RBC shape evolution at different stretching forces (0, 90, and 180 pN). (Right) Estimation RBC model's parameters

Similar to experiment, we apply in simulations stretching force and measure the cell deformation by computing the deformation along the axial (x) and transverse diameter (y). With x_F and y_F, we refer to the axial and transverse diameters measured after applying a stretching force F. Six different values of stretching force are considered to perform the fitting of parameters, with corresponding values for axial and transverse diameters taken from experimental data

ICCS Camera Ready Version 2023
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-36024-4_27

from reference [14]. Thus, the loss function to minimize is defined as follows:

$$Loss = \sum_{F \in [16,38,88,108,150,192]} |\tilde{x}_F - x_F| + |\tilde{y}_F - y_F|.$$

where $\tilde{x}_F$ and $\tilde{y}_F$ are the deformations in the axial and transverse diameter computed in simulation.

In the optimization process, BaRGO managed to find multiple minima, reported in Table 1. The values of the loss function in μm are also reported for each minima in the table. The overall cell stiffness is mainly defined by the WLC potential used in the RBC model. Specifically, the values of the persistence length, p are very similar in all minima, while the values of bending coefficient k_{bend} and the coefficients for the constant area and volume, k_a and k_v, can vary significantly without affecting the simulation results (Figure 2). We note that the surface area and volume are well conserved in all cases with the average fluctuations of 0.5% and 0.1%, respectively. The non uniqueness of the set of parameters that minimizes the loss function either might be due to the experiment itself or to the selected measurements, axial and transverse diameters, which may not be not sufficient to capture the detailed behaviour of the RBC membrane. Related discussion can be found in [11], where additional data are proposed to be collected in optical tweezers experiments.

Parameters	1^{st}minima	2^{nd}minima	3^{rd}minima	4^{th}minima	5^{th}minima	6^{th}minima	7^{th}minima	Params from [16].
p	$4.14x10^{-9}$	$4.15x10^{-9}$	$4.17x10^{-9}$	$4.11x10^{-9}$	$4.23x10^{-9}$	$4.13x10^{-9}$	$4.11x10^{-9}$	$3.43x10^{-9}$
k_{bend}	362	438	486	674	602	372	625	200
k_a	752794	597984	525380	663420	722226	513549	743761	6000
k_v	262911	698463	471066	186937	108131	328961	83505	6000
Loss	2.85	2.89	2.9	2.88	2.83	2.89	2.87	

Table 1: Parameters resulted from of the optimization process

5 Conclusions

We propose BaRGO, a novel Evolutionary Strategy algorithm that follows a Bayesian approach and by recursively calling itself is able to find multiple minima of loss function f. Convergence performance of BaRGO was shown to be comparable to the current state-of-the-art CMA-ES model in preliminary studies. Finally, we have applied BaRGO for estimating RBC model parameters using data from the optical tweezers experiment.

One of the drawbacks of BaRGO is its computational intensity, requiring an MCMC cycle at each iteration. Despite this, it can still be a useful tool in cases where complex models require parameter estimation and multiple minima need to be explored to better understand the model and dependencies among parameters.

ICCS Camera Ready Version 2023
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-36024-4_27

Acknowledgements The authors acknowledge support from the Swiss National Science Foundation grants 200021L_204817 and 192549. Simulations were carried out at the Swiss National Supercomputer Center under project u4.

References

1. Balogh, P., Gounley, J., Roychowdhury, S., Randles, A.: A data-driven approach to modeling cancer cell mechanics during microcirculatory transport. Scientific Reports **11**(1), 15232 (Jul 2021)
2. Benhamou, E., Saltiel, D., Vérel, S., Teytaud, F.: BCMA-ES: A bayesian approach to CMA-ES. CoRR **abs/1904.01401** (2019)
3. Dao, M., Lim, C.T., Suresh, S.: Mechanics of the human red blood cell deformed by optical tweezers. Journal of the Mechanics and Physics of Solids **51**(11-12), 2259–2280 (2003)
4. Dempster, A.P., Laird, N.M., Rubin, D.B.: Maximum likelihood from incomplete data via the EM algorithm. Journal of the Royal Statistical Society. Series B (Methodological) **39**(1), 1–38 (1977), http://www.jstor.org/stable/2984875
5. Economides, A., Arampatzis, G., Alexeev, D., Litvinov, S., Amoudruz, L., Kulakova, L., Papadimitriou, C., Koumoutsakos, P.: Hierarchical bayesian uncertainty quantification for a model of the red blood cell. Phys. Rev. Applied **15**, 034062 (Mar 2021)
6. Goldberg, D.E.: Genetic Algorithms in Search, Optimization and Machine Learning. Addison-Wesley Longman Publishing Co., Inc., USA (1989)
7. Hansen, N.: The CMA evolution strategy: A tutorial. arXiv preprint arXiv:1604.00772 (2016)
8. Hansen, N., Müller, S.D., Koumoutsakos, P.: Reducing the time complexity of the derandomized evolution strategy with covariance matrix adaptation (CMA-ES). Evolutionary computation **11**(1), 1–18 (2003)
9. Hoff, P.: A First Course in Bayesian Statistical Methods. Springer Texts in Statistics, Springer New York (2009)
10. Kullback, S., Leibler, R.A.: On information and sufficiency. Annals of Mathematical Statistics **22**, 79–86 (1951)
11. Sigüenza, J., Mendez, S., Nicoud, F.: How should the optical tweezers experiment be used to characterize the red blood cell membrane mechanics?. Biomechanics and Modeling in Mechanobiology, **16**, 1645–1657 (2017)
12. LAMMPS: Lammps (2015), http://lammps.sandia.gov/bench.html
13. Lim, C., Zhou, E., Quek, S.: Mechanical models for living cells—a review. Journal of Biomechanics **39**(2), 195–216 (2006)
14. Mills, J.P., Qie, L., Dao, M., Lim, C.T., Suresh, S.: Nonlinear elastic and viscoelastic deformation of the human red blood cell with optical tweezers. Mech Chem Biosyst **1**(3), 169–80 (2004)
15. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, E.: Scikit-learn: Machine learning in Python. Journal of Machine Learning Research **12**, 2825–2830 (2011)
16. Pivkin, I.V., Karniadakis, G.E.: Accurate coarse-grained modeling of red blood cells. Physical Review Letters **101**(11), 118105 (2008)

ICCS Camera Ready Version 2023
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-36024-4_27